

*Università degli Studi di Napoli Federico II,  
Dipartimento di Agraria*

Corso di Dottorato di Ricerca in Scienze Agrarie ed Agroalimentari,  
XXIX ciclo

Progetto di Tesi di Dottorato  
a.a. 2014-2017

**Pre-processing, classification and semantic  
querying of large-scale Earth observation  
spaceborne/airborne/terrestrial image databases:  
Process and product innovations**

(Innovazione di processo e di prodotto nella pre-elaborazione,  
classificazione ed interrogazione semantica di grandi database di  
immagini della Terra acquisite da piattaforme satellitari, aeree e  
terrestri)

**Tutor**  
*Prof. Francesco Giannino*

**Dottorando con borsa**  
*Ing. Andrea Baraldi*

**Coordinatore del Corso di Dottorato**  
*Prof. Guido D'urso*



## Table of Contents

Acknowledgements .....	13
Abstract .....	14
Acronyms .....	15
Premise .....	17
<b>1 Doctoral Research Objectives and Definitions of Interest .....</b>	<b>18</b>
<b>2 Introduction .....</b>	<b>30</b>
<b>2.1 Taxonomy of EO Imaging Platforms and Sensors .....</b>	<b>30</b>
<b>2.2 Spectral Resolution Specifications of EO Optical Imaging Sensors .....</b>	<b>34</b>
<b>2.3 Quality Assurance Framework for Earth Observation (QA4EO) Guidelines and Calibration/Validation Requirements .....</b>	<b>39</b>
<b>2.4 ESA EO Level 2 Information Product .....</b>	<b>40</b>
<b>2.5 FAO Land Cover Classification System (LCCS) .....</b>	<b>41</b>
<b>2.6 MS Pattern Recognition in 1D Image Analysis: Typical Misuse of Spectral Indexes with Special Emphasis on Vegetation Monitoring .....</b>	<b>43</b>
<b>2.7 1D and 2D Image Analysis Principles .....</b>	<b>47</b>
<b>2.8 References in Chapter 1 to Chapter 2 .....</b>	<b>50</b>
<b>3 <i>Technical report 1 (unpublished): Computational models of human vision - Developments and open challenges in automated detection of multi-scale image-contours, keypoints, texels and texture-boundaries in panchromatic and color images .....</i></b>	<b>55</b>
<b>Motivation and Contributions to the Dissertation .....</b>	<b>55</b>
<b>3.1 Vision is a cognitive task .....</b>	<b>57</b>
<b>3.2 Definitions of interest: The two information theories at the foundation of information technology (IT) - <i>Information-as-thing</i> and <i>Information-as-data-interpretation</i> .....</b>	<b>71</b>
<b>3.3 Definitions of interest: Proposed Minimally Dependent and Maximally Informative (mDMI) set of process and outcome quantitative quality indexes (Q<sup>2</sup>Is) in Big Data analytics .....</b>	<b>72</b>
<b>3.4 Preliminary research and development (R&amp;D) vision project objectives .....</b>	<b>73</b>
<b>3.5 Vision problem background in neuroscience and computer science .....</b>	<b>77</b>
<b>3.5.1 The human visual system .....</b>	<b>77</b>
<b>3.5.1.1 Vision system architecture .....</b>	<b>77</b>
<b>3.5.1.2 The retina .....</b>	<b>79</b>
<b>3.5.1.3 The lateral geniculate nucleus .....</b>	<b>80</b>
<b>3.5.1.4 The visual cortex .....</b>	<b>80</b>
<b>3.5.1.5 Visual information propagation .....</b>	<b>80</b>





3.5.1.6 Visual feature extraction and brightness perception in human vision .....	81
3.5.2 Scale-space representation.....	84
3.5.3 Accounting for ZX pixels through scale: the spatial coincident assumption of Marr.....	85
3.5.4 Yellot’s Theory of Low-Level Vision for Texture Discrimination: The Triple Autocorrelation Uniqueness (TAU) Theorem .....	86
3.6 Original requirements specification for computational models of human vision.....	89
3.7 Texture detection problem requirements specification.....	104
3.8 Test image sets .....	107
3.9 Original automated statistical model-based color constancy algorithm.....	109
3.10 Original expert systems for automated color naming in a calibrated MS reflectance space or in an uncalibrated RGB color space, either true- or false-color .....	111
3.10.1 The SIAM lightweight computer program for MS reflectance space hyperpolyhedralization, superpixel detection and VQ quality assurance.....	116
3.10.2 The RGBIAM lightweight computer program for true- or false-color RGB cube polyhedralization, superpixel detection and VQ quality assurance.....	125
3.11 Original 1D simulations for image analysis and synthesis, including the zero-frequency signal component, image-contour detection and keypoint detection consistent with the Mach bands illusion	128
3.12 Original definition of zero-crossing (ZX) pixels and scale-invariant keypoints in an even-symmetric and odd-symmetric wavelet-based filtered image .....	134
3.13 Original perceptual image-pair quality/ similarity/ dissimilarity index/ metric.....	140
3.14 Generalization in mathematical and linguistic terms of one specific statement by Marr regarding surface discontinuity detection in a 2½D sketch .....	142
3.15 Original design of the 2D multi-scale even- and odd-symmetric Gabor filter bank .....	145
3.16 Original implementation of a stratified multi-scale multi-orientation near-orthogonal image analysis/decomposition and synthesis/reconstruction .....	146
3.16.1 Stratified multi-scale multi-orientation near-orthogonal image analysis/decomposition .....	148
3.16.2 Stratified multi-scale multi-orientation near-orthogonal image synthesis/reconstruction .....	148
3.16.3 Boundary effect removal via predictive mirror padding in the proximity of image/object boundaries .....	148
3.17 Implemented (discrete) quaternary representation of (pos, neg, zero and masked-off) even-symmetric wavelet output values as preliminary ZX segments .....	150
3.18 Implemented continuous and (discrete) binary selection/representation of ZX pixels in an even-symmetric filtered image at either single or multiple scales and orientations .....	153
3.18.1 Continuous selection/representation of ZX pixels in an even-symmetric filtered image .....	154
3.18.2 (Discrete) binary selection/representation of ZX pixels in an even-symmetric filtered image	154
3.18.3 Selection/representation of ZX pixels in a multi-scale multi-orientation even-symmetric filtered image: ZX pixels of a ZX sum across bands, scales and orientations.....	155



3.18.4 Continuous and (discrete) trinary contour representation of continuous ZX pixels detected in an even-symmetric filtered image .....	157
3.18.5 Enhanced (filtered from scale-0 upward) continuous ZX pixels detected in an even-symmetric filtered image .....	163
3.19 Implemented continuous and (discrete) quaternary representation of contrast local extrema in an even- and odd-symmetric filtered image as image keypoints (endpoints, corners, junctions) .....	165
3.20 Original implementation of a raw and full primal sketch consistent with the Marr computational model of human vision .....	168
3.20.1 The Marr’s raw primal sketch in pre-attentional vision .....	168
3.20.2 Original automated implementation of a raw primal sketch consisting of discrete tokens (texels) as ZX segments .....	170
3.20.3 The Marr’s full primal sketch or perceptual grouping of tokens into larger-scale tokens, such as texture contours .....	179
3.20.4 Original automated implementation of a full primal sketch for texture segmentation: multi-scale texture binary profile .....	180
3.21 Original conceptual unifying framework for spatial variance, spatial autocorrelation and the proposed 2D wavelet filter bank .....	182
3.22 Conclusions and open challenges .....	187
References in Chapter 3 .....	191
<i>4 Manuscript 1 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal Remote Sensing of Environment): Systematic Earth Observation Level 2 product generation for semantic querying</i> .....	199
<b>Motivation and Contributions to the Dissertation</b> .....	199
<b>Abstract</b> .....	202
<b>4.1. Introduction</b> .....	202
<b>4.2. Materials and Methods</b> .....	207
4.2.1. System design .....	207
4.2.2. Knowledge/information representation .....	209
4.2.3. “Default” rule base for EO image pre-processing and low-level vision .....	210
4.2.3.1. EO image pre-processing for radiometric enhancement .....	212
4.2.3.2. Pre-attentive vision: Raw and full primal sketch .....	214
4.2.4. GUI .....	228
4.2.5. Array fact base .....	230
4.2.6. Hybrid inference engine .....	232
<b>4.3. Results</b> .....	235
<b>4.4. Discussion</b> .....	240



<b>4.5. Conclusions</b> .....	241
<b>Acknowledgements</b> .....	242
<b>References in Chapter 4</b> .....	242
<b>5 Manuscript 2 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal European Journal of Remote Sensing): Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases</b> .....	247
<b>Motivation and Contributions to the Dissertation</b> .....	247
<b>Abstract</b> .....	248
<b>5.1 Introduction</b> .....	248
<b>5.1.1 Semantic content-based image querying</b> .....	248
<b>5.1.2 Scalable processing</b> .....	249
<b>5.2 EO-IU&amp;SQ system design</b> .....	250
<b>5.2.1 Generic rule base for low-level information layer generation</b> .....	250
<b>5.2.1.2 EO image calibration, physical model-based colour naming and texel detection</b> .....	250
<b>5.2.1.2 Planar shape description</b> .....	250
<b>5.2.2 Spatiotemporal conceptual modelling of real-world objects in the scene domain</b> .....	251
<b>5.2.3 Array fact base</b> .....	251
<b>5.3 EO-SQ subsystem prototypical implementation</b> .....	252
<b>5.4 Use case examples</b> .....	252
<b>5.4.1 User case I – LC change detection through time</b> .....	252
<b>5.4.2 User case II – Planar shape descriptors to infer high-level LC classes from EO Level 2 products</b> .....	252
<b>5.5 Conclusions</b> .....	253
<b>Acknowledgements</b> .....	253
<b>References in Chapter 5</b> .....	253
<b>Figures in Chapter 5</b> .....	255
<b>6 Manuscript 3 (published, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 12, pp. 5560-5575, Dec. 2016): Automated Hierarchical 2D and 3D Object-Based Recognition and Reconstruction of ISO Containers in a Harbor Scene</b> .....	260
<b>Motivation and Contributions to the Dissertation</b> .....	260
<b>Abstract</b> .....	261
<b>6.1 Introduction to Part-B</b> .....	261
<b>6.2 Automated hierarchical 2D and 3D object-based recognition and reconstruction of ISO containers in a harbor scene</b> .....	263
<b>6.2.1 Introduction</b> .....	263



6.2.2 Methods .....	263
6.2.3 Automated RGB Image Pre-processing First Stage .....	264
6.2.4 Second-Stage Classification with Spatial Reasoning in an Heterogeneous 2D and 3D Data Space .....	267
6.2.5 3D Reconstruction of ISO Containers .....	269
6.2.6 Results and Discussion .....	269
6.3 Discussion of the 3D Contest .....	270
6.3.1 Submissions .....	270
6.3.2 The winners .....	271
6.3 Conclusion on the Data Fusion Contest 2015 .....	271
Acknowledgements .....	272
References in Chapter 6 .....	272
7 <i>Manuscript 4 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01930): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 1: Theory</i> .....	276
Motivation and Contributions to the Dissertation .....	276
Abstract .....	278
7.1 Introduction .....	278
7.2 Color naming problem background in cognitive science .....	283
7.3 Related works in static MS reflectance space hyperpolyhedralization .....	286
7.4 Original hybrid eight-step guideline for identification of a categorical variable-pair relationship .....	288
7.5 Original measure of association in a categorical variable-pair relationship .....	289
7.6 Conclusions .....	292
Acknowledgments .....	292
Disclosure statement .....	292
References in Chapter 7 .....	293
Figures and figure captions in Chapter 7 .....	299
Tables and table captions in Chapter 7 .....	309
8 <i>Manuscript 5 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01932): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 2: Validation</i> .....	313
Motivation and Contributions to the Dissertation .....	313
Abstract .....	315
8.1 Introduction .....	315
8.2 Materials .....	317



<b>8.3 Methods</b> .....	321
<b>8.4 Validation session</b> .....	322
<b>8.4.1 Verification of the Co-Registration Requirements for Pixel-based Inter-Map Comparison</b> ....	323
<b>8.4.2 Inter-Annual SIAM-WELD Map Comparisons for Years 2006 to 2009</b> .....	323
<b>8.4.3 Comparison of the SIAM-WELD 2006 and NLCD 2006 Thematic Maps</b> .....	324
<b>8.4.4 Probabilities of the SIAM-WELD Test Labels Conditioned by the NLCD Reference Labels and Vice Versa</b> .....	327
<b>8.4.5 Stratification by Ecoregions</b> .....	328
<b>8.4.6 OP-Q<sup>2</sup>I values of the SIAM application and product</b> .....	329
<b>8.5 Discussion</b> .....	329
<b>8.6 Conclusions</b> .....	331
<b>Acknowledgments</b> .....	332
<b>Disclosure statement</b> .....	332
<b>References in Chapter 8</b> .....	333
<b>Figures and figure captions in Chapter 8</b> .....	337
<b>Tables and table captions in Chapter 8</b> .....	349
<b>9 Manuscript 6 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01940): Automated Near Real-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images</b> .....	355
<b>Motivation and Contributions to the Dissertation</b> .....	355
<b>Abstract</b> .....	356
<b>9.1 Introduction</b> .....	357
<b>9.2 Problem Background</b> .....	360
<b>9.2.1 Human vision</b> .....	360
<b>9.2.2 RGB Cube Partitioning into “Universal” Basic Colors Known in Advance</b> .....	361
<b>9.2.3 Color Constancy</b> .....	361
<b>9.2.4 Deductive and Inductive Vector Quantization</b> .....	362
<b>9.2.5 Two-Pass Connected-Component Multi-Level Image Labeling</b> .....	363
<b>9.2.6 Superpixel Detection Equivalent to Texel Detection in the Pre-Attentional Raw Primal Sketch</b> .....	364
<b>9.3 Methods</b> .....	364
<b>9.3.1 Software Design, Algorithm Selection and Implementation</b> .....	364
<b>9.3.2 Quantitative Quality Assurance (Q<sup>2</sup>A) of the Low-level Vision Software Pipeline</b> .....	367
<b>9.3.3 Comparison with Alternative Approaches</b> .....	368
<b>9.4 Materials</b> .....	368



<b>9.5 Results</b> .....	369
<b>9.6 Discussion</b> .....	370
<b>9.7 Conclusions</b> .....	371
<b>Acknowledgment</b> .....	372
<b>References in Chapter 9</b> .....	373
<b>Figures and figure captions in Chapter 9</b> .....	378
<b>Tables and table captions in Chapter 9</b> .....	386
<b>10 Manuscript 7 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01941): Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis</b> .....	388
<b>Motivation and Contributions to the Dissertation</b> .....	388
<b>Abstract</b> .....	390
<b>10.1 Introduction</b> .....	390
<b>10.2 Terminology</b> .....	393
<b>10.3 Related Works</b> .....	393
<b>10.4 Materials</b> .....	395
<b>10.5 Methods</b> .....	395
<b>10.5.1 Qualitative Maximization of Informativeness of a Feature Set</b> .....	396
<b>10.5.2 Quantitative Minimization of Dependence of a Feature Set</b> .....	396
1) <i>First sufficient not necessary criterion for pairwise feature non-causality: Statistical independence</i> 397	
2) <i>Second sufficient not necessary criterion for pairwise feature non-causality: Inter-feature statistical non-monotonicity (at the population level)</i> .....	398
3) <i>Third sufficient not necessary criterion for pairwise feature non-causality: Inter-feature local non-monotonicity at the individual level</i> .....	398
<b>10.6 Experimental Results</b> .....	399
<b>10.6.1 Original Representation and Implementation of 2D Shape Features whose Q<sup>2</sup>IOs Must Score High</b> .....	399
1) <i>Area-based convexity sensitive to within-segment holes</i> .....	400
2) <i>Polygonal representation of a 2D shape alternative to skeletonization</i> .....	401
3) <i>Contour-based fuzzy rectangularity</i> .....	401
4) <i>Area-based roundness (compactness) sensitive to within-segment holes</i> .....	402
5) <i>Contour-based multiscale straightness of boundaries</i> .....	402
6) <i>Pixel- and segment-based morphological multiscale characteristic</i> .....	402
7) <i>Area-based elongatedness sensitive to within-segment holes</i> .....	406



8) Area-based simple connectivity as a measure of the presence of within-segment holes .....	407
<b>10.6.2 Pairwise Feature Test of Statistical Independence .....</b>	<b>408</b>
<b>10.6.3 Monotonically Increasing or Decreasing Relationship between Pairs of Planar Geometric Features at the Population and Individual Levels of Analysis .....</b>	<b>408</b>
<b>10.7 Discussion .....</b>	<b>409</b>
<b>10.7.1 Qualitative Analysis of 2D Shape Indexes .....</b>	<b>409</b>
<b>10.7.2 Qualitative Interpretation of Quantitative Statistical (In)dependence Results .....</b>	<b>410</b>
<b>10.7.3 Qualitative Interpretation of Quantitative SRCC results.....</b>	<b>410</b>
<b>10.8 Conclusions .....</b>	<b>411</b>
<b>Appendix 1 – OpenCV library .....</b>	<b>412</b>
<b>Appendix 2 – ENVI EX 5.0.....</b>	<b>412</b>
<b>Appendix 3 – QA4EO guidelines.....</b>	<b>413</b>
<b>Acknowledgments.....</b>	<b>413</b>
<b>References in Chapter 10.....</b>	<b>414</b>
<b>Figures and figure captions in Chapter 10 .....</b>	<b>419</b>
<b>Tables and table captions in Chapter 10.....</b>	<b>427</b>
<b>11 Manuscript 8 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01942): Multi-spectral Image Panchromatic Sharpening – Outcome and Process Quality Assessment Protocol .....</b>	<b>430</b>
<b>Motivation and Contributions to the Dissertation.....</b>	<b>430</b>
<b>Abstract .....</b>	<b>431</b>
<b>11.1 Introduction .....</b>	<b>431</b>
<b>11.2 Problem Background .....</b>	<b>433</b>
<b>11.2.1 Quantitative Image Quality Metrics: Signal Fidelity Measures and Perceptual Visual Quality Metrics .....</b>	<b>434</b>
<b>11.2.2 Non-Injective Property of Summary (Gross) Characteristics.....</b>	<b>435</b>
<b>11.2.3 Multi-scale Image Statistics .....</b>	<b>436</b>
<b>11.2.4 Yellot's Theory of Low-Level Vision for Texture Discrimination: The Triple Autocorrelation Uniqueness (TAU) Theorem .....</b>	<b>436</b>
<b>11.2.5 Criteria for Quality Improvement of Existing PAN-Sharpended MS Image Estimation Procedures.....</b>	<b>438</b>
<b>11.3 Critical Review of Existing Procedures for Q<sup>2</sup>A of PAN-Sharpended MS Images .....</b>	<b>439</b>
<b>11.3.1 Abstract Three-Statement Wald's Protocol, where an Ideal Reference Image at Fine Resolution is Available for Comparison Purposes .....</b>	<b>439</b>
<b>11.3.2 Quantitative Analysis with the Sensory MS<sub>i</sub> Image Adopted as Reference: Revised Two-Statement Wald's Protocol and Its One-Statement Simplified Version.....</b>	<b>440</b>



11.3.3 Quantitative Analysis at High Spatial Scale $h$ , Without Reference Image.....	442
11.4 Materials and methods.....	442
11.4.1 Validation Dataset .....	442
11.4.2 MS Image PAN-Sharpener Algorithms Selected for Testing .....	443
11.4.3 State-of-the-Art Multivariate Scalar QIs Selected for Comparison Purposes.....	444
11.4.4 Perceptual Image Quality Assessment.....	444
11.4.5 Expert System in Operating Mode for Prior Knowledge-Based MS Data Space Discretization (Partitioning).....	445
11.4.6 New Protocol for $Q^2A$ of MS Image PAN-sharpening Outcome and Process.....	446
4) First category of product QIs: SPCTRL - Context-insensitive (pixel-based) and Position (row and column)-independent (Rotation invariant).....	446
5) Second category of product QIs: SPCTRL & SPTL1 - Context-insensitive and Position-dependent (Sensitive to Rotation) .....	447
6) Third category of product QIs: SPCTRL & SPTL2 - Context-sensitive Position-independent (Rotation invariant).....	447
7) Fourth category of product QIs: SPCTRL & SPTL1 & SPTL2 - Context-sensitive Position-dependent (Sensitive to rotation).....	448
8) Process QIs .....	448
9) Quantitative QI combination and ranking.....	448
11.5 Results.....	449
11.6 Discussion .....	449
11.7 Conclusions .....	450
References in Chapter 11 .....	453
Figures and figure captions in Chapter 11 .....	460
Tables and table captions in Chapter 11 .....	470
12 <i>Technical report 2 (made available in the public archive arXiv: 1701.04256): Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+</i> .....	477
Motivation and Contributions to the Dissertation.....	477
12.1 Introduction .....	478
12.2 EO-VAS objectives, technical requirements and proposed approach.....	478
12.2.1 EO-VAS aims and degrees of innovation .....	478
12.2.2 EO-VAS architecture and implementation .....	480
12.2.3 System integration, quality assessment and comparison of alternative solutions .....	483
12.2.4 Differences between the proposed solution and alternative existing solutions .....	484
12.2.5 Target user communities.....	486





<b>12.2.6 Expected benefits of the proposed EO-VAS solution .....</b>	<b>486</b>
<b>12.2.7 Future opportunities of the proposed EO-VAS solution.....</b>	<b>486</b>
<b>12.2.8 Proposed approach to the work and first iteration of tasks .....</b>	<b>486</b>
<b>12.3 Potential problem areas .....</b>	<b>486</b>
<b>12.3.1 Identification of the main problem areas likely to be encountered in performing the activity .....</b>	<b>486</b>
<b>12.3.2 Proposed solutions to the problems identified .....</b>	<b>487</b>
<b>12.3.3 Proposed trade-off analyses and identification of possible limitations or non-compliances .....</b>	<b>487</b>
<b>12.4 Technical implementation / Programme of work.....</b>	<b>487</b>
<b>12.4.1 Proposed work logic .....</b>	<b>487</b>
<b>12.4.2 Detailed procurement plan for the EO data.....</b>	<b>487</b>
<b>References in Chapter 12 .....</b>	<b>487</b>
<b>Figures and figure captions in Chapter 12.....</b>	<b>491</b>
<b>Tables and table captions in Chapter 12 .....</b>	<b>497</b>
<b>13 R&amp;D Summary and Future Works .....</b>	<b>500</b>



# Pre-processing, classification and semantic querying of large-scale Earth observation spaceborne/airborne/terrestrial image databases: Process and product innovations

(Innovazione di processo e di prodotto nella pre-elaborazione, classificazione ed interrogazione semantica di grandi database di immagini della Terra acquisite da piattaforme satellitari, aeree e terrestri)

Non domandarci la formula che mondi possa aprirti,  
sì qualche storta sillaba e secca come un ramo.  
Codesto solo oggi possiamo dirti:  
ciò che non siamo, ciò che non vogliamo.

Eugenio Montale  
*Non Chiederci la Parola*, in *Ossi di Seppia*, 1923

There is no semantics in data.

Emanuel Diamant  
*Not only a lack of right definitions: Arguments for a shift in information-processing paradigm*, 2010

With all due respect to the brilliant Geoff Hinton - thought is not a vector, and artificial intelligence is not a problem in statistics.

Oren Etzioni  
*Quora - What shortcomings do you see with deep learning?*, 2017

Keywords: tribute to David Marr, Bayesian inference in vision.

One of David Marr's key is the notion of *constraints*. The idea that the human visual system embodies constraints that reflect properties of the world is foundational. Indeed, this general view seemed (to me) to provide a sensible way of thinking about Bayesian approaches to vision. Accordingly, Bayesian priors are Marr's constraints. The priors/constraints have been incorporated into the human visual system over the course of its evolutionary history (according to the "levels of understanding" manifesto extended by Tomaso Poggio in 2012).

Philip Quinlan  
*Marr's Vision 30 years on: From a personal point of view*, 2012



## Acknowledgements

First and foremost, I wish to thank my advisor, Prof. Francesco Giannino. His openmindedness, enthusiasm, support and guidance were invaluable during my experience as a doctoral student to overcome some professional and personal problems I faced with my academic institution and doctoral committee.

I also wish to acknowledge the willingness to help and scientific expertise of foreign researchers and host institutions I had the privilege to collaborate with during my doctoral experience, specifically, Prof. Josep Lladós, Director of the Computer Vision Center (CVC) of the Universidad Autonoma de Barcelona (UAB), Spain, and Prof. Joost van de Weijer (CVC-UAB); Prof. Josef Strobl, Director of the interdisciplinary Centre of Competence for Geoinformatics (Z-GIS) of the Department of Geoinformatics, University of Salzburg, Austria, Prof. Stefan Lang (Z-GIS), Prof. Dirk Tiede (Z-GIS) and Prof. Thomas Blaschke (Z-GIS); Prof. Arnon Karnieli, Head of the Remote Sensing Laboratory, Ben-Gurion University of the Negev, Sede-Boker Campus, Israel.

I am also thankful to Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, MD, for his friendship and support, and to Prof. Raphael Capurro, Founder of the Capurro Fiek Foundation, for his generosity and open-mindedness.

In 2012 there had been the 30th anniversary of the original publication date of David Marr's *Vision* whose third part reports on discussions he had with Francis Crick, Leslie Orgel and Tomaso Poggio at the Salk Institute in La Jolla in 1979. Since I met Tomaso Poggio in person at the beginning of my research career, Marr and Poggio's influence on my thinking has endured.

Last, but not least, I am grateful to my family and friends for their support, starting from my wife, Lucia. Without her love, patience, recommendations and encouragement, it would have been harder to complete this dissertation.



## Abstract

By definition of Wikipedia, “big data is the term adopted for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The big data challenges typically include capture, curation, storage, search, sharing, transfer, analysis and visualization”.

Proposed by the intergovernmental Group on Earth Observations (GEO), the visionary goal of the Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015 is systematic transformation of multi-source Earth Observation (EO) “big data” into timely, comprehensive and operational EO value-adding products and services, submitted to the GEO Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements. To date the GEOSS mission cannot be considered fulfilled by the remote sensing (RS) community. This is tantamount to saying that past and existing EO image understanding systems (EO-IUSs) have been outpaced by the rate of collection of EO sensory big data, whose quality and quantity are ever-increasing. This true-fact is supported by several observations. For example, no European Space Agency (ESA) EO Level 2 product has ever been systematically generated at the ground segment. By definition, an ESA EO Level 2 product comprises a single-date multi-spectral (MS) image radiometrically calibrated into surface reflectance (SURF) values corrected for geometric, atmospheric, adjacency and topographic effects, stacked with its data-derived scene classification map (SCM), whose thematic legend is general-purpose, user- and application-independent and includes quality layers, such as cloud and cloud-shadow. Since no GEOSS exists to date, present EO content-based image retrieval (CBIR) systems lack EO image understanding capabilities. Hence, no semantic CBIR (SCBIR) system exists to date either, where semantic querying is synonym of semantics-enabled knowledge/information discovery in multi-source big image databases.

In set theory, if set A is a strict superset of (or strictly includes) set B, then  $A \supset B$ . This doctoral project moved from the working hypothesis that  $SCBIR \supset$  computer vision (CV), where vision is synonym of scene-from-image reconstruction and understanding  $\supset$  EO image understanding (EO-IU) in operating mode, synonym of  $GEOSS \supset$  ESA EO Level 2 product  $\supset$  human vision. Meaning that necessary not sufficient pre-condition for SCBIR is CV in operating mode, this working hypothesis has two corollaries. First, human visual perception, encompassing well-known visual illusions such as Mach bands illusion, acts as lower bound of CV within the multi-disciplinary domain of cognitive science, i.e., CV is conditioned to include a computational model of human vision. Second, a necessary not sufficient pre-condition for a yet-unfulfilled GEOSS development is systematic generation at the ground segment of ESA EO Level 2 product.

Starting from this working hypothesis the overarching goal of this doctoral project was to contribute in research and technical development (R&D) toward filling an analytic and pragmatic information gap from EO big sensory data to EO value-adding information products and services. This R&D objective was conceived to be twofold. First, to develop an original EO-IUS in operating mode, synonym of GEOSS, capable of systematic ESA EO Level 2 product generation from multi-source EO imagery. EO imaging sources vary in terms of: (i) platform, either spaceborne, airborne or terrestrial, (ii) imaging sensor, either: (a) optical, encompassing radiometrically calibrated or uncalibrated images, panchromatic or color images, either true- or false color red-green-blue (RGB), multi-spectral (MS), super-spectral (SS) or hyper-spectral (HS) images, featuring spatial resolution from low ( $> 1\text{km}$ ) to very high ( $< 1\text{m}$ ), or (b) synthetic aperture radar (SAR), specifically, bi-temporal RGB SAR imagery.

The second R&D objective was to design and develop a prototypical implementation of an integrated closed-loop EO-IU for semantic querying (EO-IU4SQ) system as a GEOSS proof-of-concept in support of SCBIR. The proposed closed-loop EO-IU4SQ system prototype consists of two subsystems for incremental learning. A primary (dominant, necessary not sufficient) hybrid (combined deductive/top-down/physical model-based and inductive/bottom-up/statistical model-based) feedback EO-IU subsystem in operating mode requires no human-machine interaction to automatically transform in linear time a single-date MS image into an ESA EO Level 2 product as initial condition. A secondary (dependent) hybrid feedback EO Semantic Querying (EO-SQ) subsystem is provided with a graphic user interface (GUI) to streamline human-machine interaction in support of spatiotemporal EO big data analytics and SCBIR operations. EO information products generated as output by the closed-loop EO-IU4SQ system monotonically increase their value-added with closed-loop iterations.



## Acronyms

AI: Artificial Intelligence  
 ANN: Artificial Neural Network  
 AOI: (geographic) Area Of Interest  
 ATCOR: Atmospheric/Topographic Correction commercial software product  
 arXiv: The arXiv (pronounced "archive") is a repository of electronic preprints, known as eprints, of scientific papers in the fields of mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance, which can be accessed online. In many fields of mathematics and physics, almost all scientific papers are self-archived on the arXiv repository. Begun on August 14, 1991.  
 AVHRR: Advanced Very High Resolution Radiometer  
 BIVRTAB: Bivariate frequency Table  
 Cal/Val: Calibration and Validation  
 CBIR: Content-based Image Retrieval  
 CEOS: Committee on Earth Observation Satellites  
 CMTRX: Confusion Matrix  
 CV: Computer Vision  
 CVPAI: Categorical Variable Pair Association Index (in range [0, 1])  
 DCNN: Deep Convolutional Neural Network  
 DLR: Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)  
 DN: Digital Number  
 DP: Dichotomous Phase (in the Land Cover Classification System taxonomy)  
 EO: Earth Observation  
 ESA: European Space Agency  
 ETAU: Enhanced Triple Autocorrelation Uniqueness principle  
 FAO: Food and Agriculture Organization  
 GEO: Intergovernmental Group on Earth Observations  
 GPS: Global Positioning System  
 GEOSS: Global EO System of Systems  
 GIGO: Garbage In, Garbage Out principle of error propagation  
 GIS: Geographic Information System  
 GIScience: Geographic Information Science  
 GUI: Graphic User Interface  
 HR: High spatial Resolution (EO imagery, in range [1 m, 30 m])  
 HS: Hyper-Spectral  
 IQM: Image Quality Metric  
 IT: Information Technology  
 IUS: Image Understanding System  
 IU4SQ: Image Understanding for Semantic Querying system  
 LC: Land Cover  
 LCC: Land Cover Change  
 LCCS: Land Cover Classification System  
 LR: Low spatial Resolution (EO imagery,  $\geq 1$  km)  
 mDMI: Minimally Dependent and Maximally Informative (set of quality indicators)  
 MHP: Modular Hierarchical Phase (in the LCSS taxonomy)  
 MIR: Medium InfraRed  
 MR: Medium spatial Resolution (EO imagery, in range (30 m, 1 km))  
 MS: Multi-Spectral  
 NASA: National Aeronautics and Space Administration  
 NIR: Near InfraRed  
 NLCD: National Land Cover data  
 NOAA: National Oceanic and Atmospheric Administration



OAMTRX: Overlapping Area Matrix  
 OBIA: Object-Based Image Analysis  
 OP-Q<sup>2</sup>I: Outcome and Process Quantitative Quality Index  
 PA: Precision Agriculture  
 PhD: Doctor of Philosophy  
 PVQM: Perceptual Visual Quality Metric  
 QA: Quality Assurance  
 QA4EO: Quality Accuracy Framework for Earth Observation  
 Q<sup>2</sup>I: Quantitative Quality Indicator  
 R&D: Research and technical/technological Development  
 RGB: monitor-typical Red-Green-Blue data cube  
 RGBIAM<sup>TM</sup>: RGB Image Automatic Mapper<sup>TM</sup>  
 RS: Remote Sensing  
 SCBIR: Semantic Content-based Image Retrieval  
 SEN2COR: Sentinel 2 (atmospheric) Correction Prototype Processor  
 SIAM<sup>TM</sup>: Satellite Image Automatic Mapper<sup>TM</sup>  
 SQ: Semantic Querying  
 SS: Super-Spectral  
 STRATCOR: Stratified topographic correction  
 SURF: Surface Reflectance  
 TAU: Triple Autocorrelation Uniqueness principle  
 TIR: Thermal InfraRed  
 TM: (non-registered) Trademark  
 TOA: Top-Of-Atmosphere  
 TOARD: TOA Radiance  
 TOARF: TOA Reflectance  
 TOC: Topographic Correction  
 TPM: Topology-Preserving (feature) Map  
 UAV: Unmanned Aerial Vehicle  
 UAS: Unmanned Aircraft System  
 USGS: US Geological Survey  
 VHR: Very High spatial Resolution (EO imagery, < 1 m)  
 VIS: Visible (wavelengths)  
 VQ: Vector Quantization  
 WELD: Web-Enabled Landsat Data set  
 WGCV: Working Group on Calibration and Validation  
 ZC: Zero-Concavity (image-segment)  
 ZX: Zero-Crossing (pixel in an image-contour or image-segment), point where a function's value changes its sign.

## Premise

This doctoral dissertation was written to comply with guidelines for a successful Doctor of Philosophy (PhD) experience proposed in the engineering/computer science literature {1}, {2}. For the sake of completeness these guidelines are revisited and summarized as follows.

- (1) A PhD award is the highest college or university degree awarded for original research undertaken to increase a field-specific or interdisciplinary stock of knowledge of humans, culture and society, and the use of this stock of knowledge to devise new applications (developments). A PhD contribution in research and development (R&D) to the human knowledge is realistically expected to be as that shown in Fig. A.

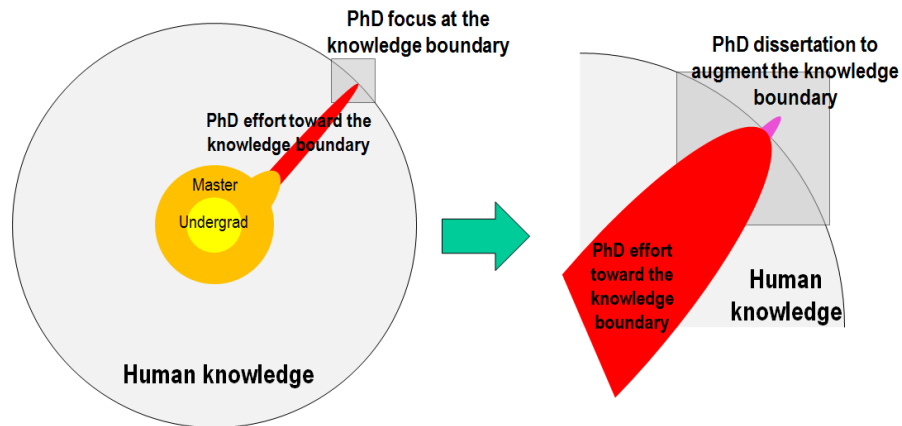


Fig. A. Adapted from {1}, {2}. The contribution to human knowledge expected by a PhD dissertation.

- (2) A scientific dissertation adopts the scientific method whose phases are: Observe the physical world, Identify a problem to solve; Formulate hypotheses; Develop solutions; Run experiments; Draw analytic and experimental conclusions to substantiate or refute those hypotheses; Go back to world observation. Noteworthy, the essence of a dissertation is critical thinking, not experimental data. Analysis and concepts are key. A dissertation concentrates on principles: it states the lessons learned, and not merely the facts behind them. Merits of a scientific dissertation are always independent of its commercial impact.
- (3) A scientific dissertation is not a personal voyage of self-discovery. It should contain 2-3 key research hypotheses, either field-specific or interdisciplinary, that make a difference in the scientific community, and provide original, rigorous, experimental and/or formal arguments capable of convincing fellow scientists to substantiate or refute those hypotheses. Every statement must be supported by a reference to the scientific literature or by original work. The entire field, including the very smartest people who practice in it, should learn something significant in reading the dissertation.
- (4) To make a mark in the chosen scientific field or interdisciplinary domain, a dissertation in engineering/computer science can introduce, first, a new theory, concept or system architecture to formalize a known problem or a new problem that has never been considered before. Second, it can present and discuss new algorithms and/or implementations to be reproduced by others and whose breaking points and failure modes are clearly documented together with their advantages.
- (5) Since all scientists need to communicate their discoveries, an ideal dissertation should be accessible to everyone in engineering/computer science, not just to specialists.

The impact of a dissertation can be measured by the extent its generated knowledge: (i) is acknowledged by the scientific community in terms of citations. This also means that substantial portions of a dissertation are expected to be published or accepted for publication by the time you defend your thesis. (ii) Leads to a technology transfer, e.g., in terms of software licensing to companies which make profit from this knowledge. Nevertheless, in a scientific dissertation conclusions should never regard the economic viability or commercial impact of an idea. (iii) It is acknowledged by society; society pays for scientific research and can provide generated knowledge with reputation in terms of popularization.



# 1 Doctoral Research Objectives and Definitions of Interest

To comply with guidelines for a successful Doctor of Philosophy (PhD) experience proposed in the engineering/computer science literature {1}, {2}, this doctoral dissertation focuses on vision, with special emphasis on computer vision (CV) applied to Earth observation (EO) image understanding, where EO images, either single-date or time-series, are acquired by terrestrial, airborne or spaceborne optical imaging sensors.

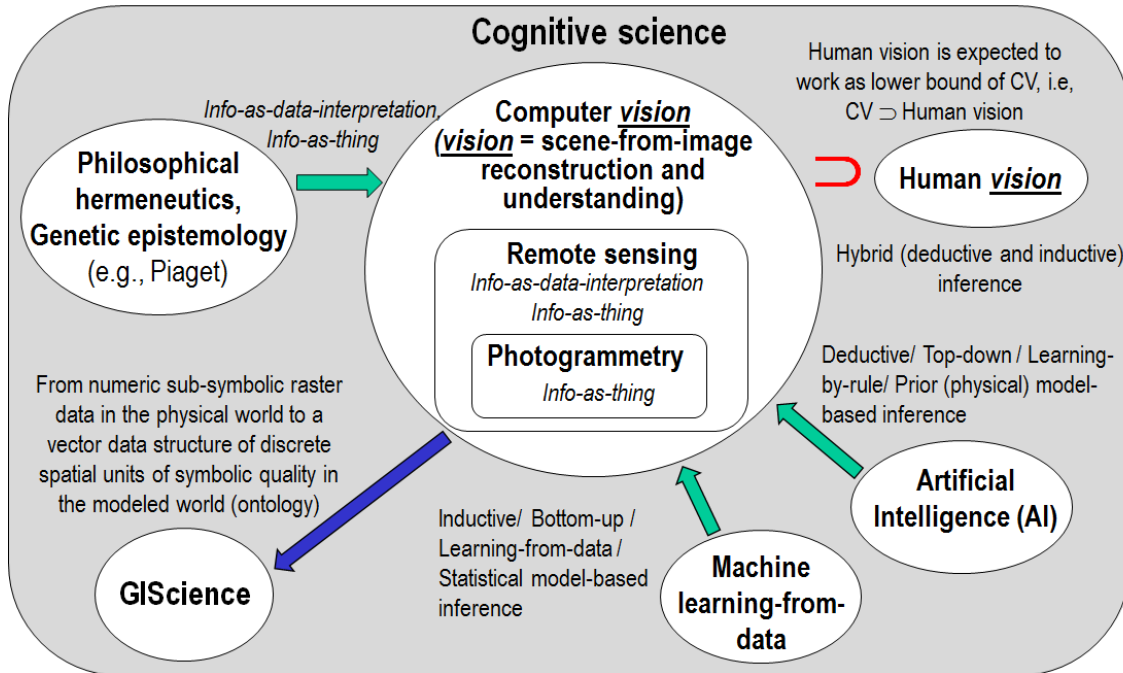


Fig. 1-1. Multi-disciplinary cognitive science {7}, {8}, {9}, {10}, {11}. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning {118}, adaptation, self-organization) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred either in the neuro-cerebral apparatus of living organisms or in machines, e.g., computers. Neurophysiology studies the neuro-cerebral apparatus of living organisms. Neural network (NN) is synonym of distributed processing system consisting of neurons as elementary processing elements and synapses as lateral connections. Is it convenient and even possible to mimic biological mental functions, e.g., human reasoning, by an artificial mind whose physical support is not an electronic brain implemented as an artificial NN (ANN)? The answer is no according to the “connectionists approach” promoted by traditional cybernetics, where a complex system always comprises an “artificial mind-electronic brain” combination. This is alternative to traditional artificial intelligence (AI) whose symbolic approach investigates an artificial mind independently of its physical support {18}.

Vision is a cognitive task, synonym of scene-from-image reconstruction and understanding, see Fig. 1-1 {3}, {4}, {5}, {6}, {7}, {8}, {9}, {10}, {11}. Cognitive tasks pursue an inherently equivocal (subjective) interpretation of sub-symbolic sensory data (observables) in the real (physical) world, equivalent to numeric/quantitative variables typically known as ever-varying sensations, into categorical/nominal/qualitative variables of semantic (symbolic) quality in the modeled world (mental world, world ontology, world model {3}, {4}), equivalent to stable percepts. Hence, cognitive tasks lead to the qualitative information theory known as *information-as-data-interpretation* {5}. Any *information-as-data-interpretation* problem, synonym of cognitive task, is inherently ill-posed {3}, {12} in the Hadamard sense {13}. As such, it is very difficult to solve {12} and requires *a priori* knowledge in addition to sensory data to become better posed for numerical solution {12}. For example, no biological cognitive system starts from an absolute beginning (tabula rasa) {16}; rather, biological cognitive systems start from a deductive/top-down *a priori* genotype as initial condition of a phenotypic inductive/bottom-up learning-from-examples inference system, capable of exploring the neighborhood of its initial condition in a solution space {17}. In the words of Poggio “the problem of learning is at the core of the problem of



intelligence and of understanding the brain... From the point of view of statistical learning theory, it seems that the incredible effectiveness with which humans (and many animals) learn from and perform in the world *cannot* result only from superior learning algorithms, but also from a huge platform of knowledge and priors. This is right in the spirit of Marr’s computational approach: constraints, “discovered” by evolution, allow the solution of the typically ill-posed problems of intelligence. Thus evolution is responsible for intelligence, and should be at the top of our levels of understanding” {118}.

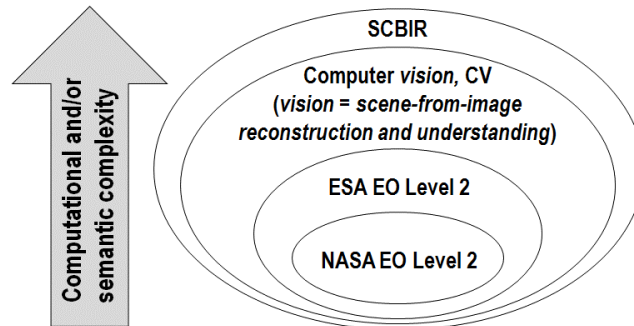


Fig. 1-2. Our working hypothesis was that existing EO content-based image retrieval (CBIR) systems support no semantic querying because they lack EO image understanding capabilities. This conjecture implies that to solve the semantic CBIR (SCBIR) open problem, a necessary not sufficient pre-condition is the computational solution of the cognitive problem of vision, understood as synonym of scene-from-image reconstruction and understanding. If relationship  $SCBIR \supset CV$  in operating mode holds true, then the complexity of SCBIR is not inferior to the complexity of vision, acknowledged to be a cognitive problem very difficult to solve because inherently ill-posed in the Hadamard sense, non-polynomial (NP)-hard in computational complexity. Vision is inherently ill-posed because affected by: (i) a 4D-to-2D data dimensionality reduction from the 4D spatiotemporal scene-domain to the (2D) image-domain, e.g., responsible of occlusion phenomena, and (ii) a semantic information gap from sub-symbolic (observable) data (ever-varying sensations) in the (2D) image-domain to stable percepts of symbolic quality in the modeled world (world ontology, world model), consisting of 4D spatiotemporal entities (classes of real-world objects) and inter-class relationships. Within the  $CV \supset EO$  image understanding (EO-IU) domain, a National Aeronautics and Space Administration (NASA) EO Level 2 product, defined as “a data-derived geophysical variable at the same resolution and location as Level 1 source data” {115}, is a special case of ESA EO Level 2 product defined as {30}: (i) a single-date multi-spectral (MS) image whose digital numbers (DNs) are radiometrically corrected into surface reflectance (SURF) values for atmospheric, adjacency and topographic effects, stacked with (ii) its data-derived general-purpose, user- and application-independent scene classification map (SCM), whose thematic map legend includes quality layers such as cloud and cloud-shadow.

In the interdisciplinary domain of cognitive science, term “vision” encompasses both human vision and computer vision (CV), see Fig. 1-1. In the words of Iqbal and Aggarwal: “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception... This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” {19}. In the words of Pessoa, “if we require that a computational vision model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published models is surprisingly small” {20}. In Fig. 1-1, constraint  $CV \supset$  human vision means that inherently ill-posed CV is conditioned by human vision to become better posed for numerical solution. If CV is a superset of human vision, then CV must include a computational model of human vision when light in the visible spectrum is reflected by objects in the environment, which means that CV complies with human visual perception, including perceptual visual illusions. The study of optical illusions in visual perception, occurring when humans mentally see (perceive) what is not either in the observed scene (e.g., Mach bands illusion {20}) or in their retinal stimuli (e.g., the well-known Kanizsa Triangle illusion, first described in 1955 by an Italian psychologist named Gaetano Kanizsa), has yielded much insight into what assumptions (prior knowledge) the visual system requires in addition to sensory data to become better conditioned for achieving plausible solution(s) in scene-from-image reconstruction and understanding.

In greater detail, vision is a cognitive (*information-as-data-interpretation*) problem, whose goal is 4D spatiotemporal scene reconstruction and understanding from (2D) imagery, which is inherently ill-posed in the Hadamard sense and whose computational complexity is non-polynomial (NP), i.e., vision is NP-hard {14}, {15}. Vision is inherently ill-posed because it has to cope with:

- A 4D to 2D data dimensionality reduction from the scene-domain (space-time domain) to the image-domain.
- A semantic information gap from numeric sensory data (sub-symbolic sensations, observables) in the (2D) image-domain to categorical variables of semantic quality (symbolic percepts belonging to a world ontology, mental world or “world model”) in the 4D spatio-temporal scene-domain {3}.

In both human vision and CV, the following terminology is adopted {4}, {21}, {22}, {23}.

- ✓ Low-level (pre-attentive) vision is synonym of image pre-processing, where numeric (quantitative) variables are transformed into numeric variables of enhanced sub-symbolic quality (unequivocal *information-as-thing* {5}), e.g., pixel values of improved geometric or radiometric quality, such as dimensionless digital numbers (DNs) provided with a physical unit of radiometric measure following radiometric calibration data pre-processing. According to the Marr computational model of vision, it consists of two phases {4}.
  - Raw primal sketch, where sub-symbolic tokens are detected in the (2D) image-domain. In the Marr terminology tokens are image-contours (zero-crossing pixels), located by an image-contour detector, perceptually uniform image-objects (zero-crossing segments, blobs, closed-contours), identified by an image segmentation (partition) phase, and keypoints consisting of corners, endpoints, T-junctions and X-junctions.
  - Full primal sketch. Synonym of texture segmentation or perceptual spatial grouping of texture elements (texels). The 2D spatial organization of tokens is investigated in the image-domain to detect perceptually uniform textured areas of sub-symbolic quality.

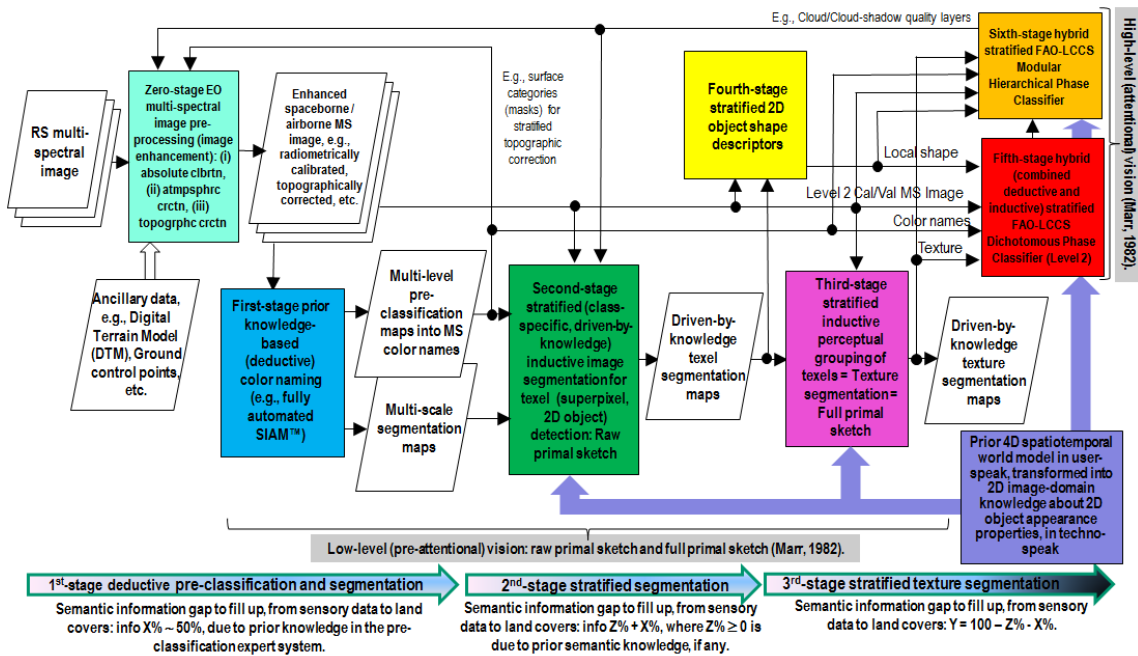


Fig. 1-3. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) modular design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed closed-loop EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to the feedforward inductive learning-from-data inference approach adopted by a large majority of the CV and the remote sensing (RS) literature.



- ✓ High-level (attentional) vision is expected to fill the semantic information gap from sub-symbolic numeric (quantitative) or categorical (qualitative) variables (information primitives) in the (2D) image-domain to symbolic categorical variables (stable percepts) in the 4D spatiotemporal scene-domain, which is an inherently equivocal *information-as-data-interpretation* process {5}. Percepts are classes of real-world objects, provided with sub-symbolic spatiotemporal relationships (e.g., adjacency) and/or symbolic (conceptual) relationships (e.g., part-of, subset-of, etc.), belonging to a world model available *a priori*, i.e., available in addition to sensory data {3}, {24}.

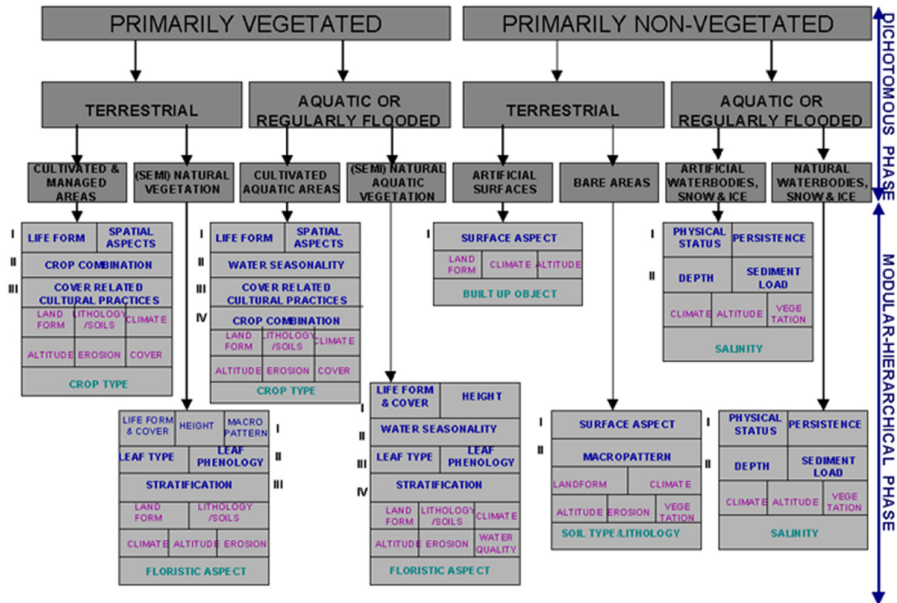


Fig. 1-4. The “fully nested” FAO Land Cover Classification System (LCCS) taxonomy comprises an application-independent three-level eight-class Dichotomous Phase (DP), followed by a subsequent so-called Modular Hierarchical Phase (MHP) {41}. The nested 3-level LCCS-DP layers are: (i) vegetation versus non-vegetation, (ii) terrestrial versus aquatic, and (iii) managed versus natural or semi-natural. They deliver as output the following 3-level 8-class LCCS-DP taxonomy. (A11) Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas. (A12) Natural and Semi-Natural Terrestrial Vegetation. (A23) Cultivated Aquatic or Regularly Flooded Vegetated Areas. (A24) Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation. (B35) Artificial Surfaces and Associated Areas. (B36) Bare Areas. (B47) Artificial Waterbodies, Snow and Ice. (B48) Natural Waterbodies, Snow and Ice. The general-purpose user- and application-independent 8-class LCCS-DP taxonomy is preliminary to a user- and application-specific LCCS-MHP taxonomy consisting of one-class LC classifiers.

In agreement with the Group on Earth Observations (GEO) Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements {25}, an information processing system, such as a CV system, can be considered in operating mode when it scores (fuzzy) “high” in each quantitative quality indicator (Q<sup>2</sup>I) of a minimally dependent and maximally informative (mDMI) set of outcome and process Q<sup>2</sup>Is (OP-Q<sup>2</sup>Is), to be community-agreed upon to be used by members of the community. In the remote sensing (RS) literature, a proposed mDMI set of EO OP-Q<sup>2</sup>Is includes the following observable (non-latent) indexes to be directly measured {26}, {27}.

- Degree of automation, inversely related to the required degree of human-machine interaction. For example, the system’s degree of automation is monotonically decreasing with the number of system’s free-parameters to be user-defined based on heuristics.
- Effectiveness, e.g., thematic mapping accuracy.
- Degree of semantics, if any.
- Efficiency in computation time.
- Efficiency in memory occupation.
- Robustness (vice versa, sensitivity) to changes in input data.
- Robustness to changes in input parameters to be user-defined, if any.
- Scalability to changes in user requirements and in sensor specifications.



- Timeliness from data acquisition to information product generation.
- Costs in manpower and computer power.

In this doctoral project the primary working hypothesis was that existing CV systems  $\supset$  EO image understanding systems (EO-IUSs) are expected to be a superset of human vision, i.e., relationship  $CV \supset EO-IUS$  in operating mode  $\supset$  human vision should hold. Unfortunately, in operating mode existing EO-IUSs typically score “low”, i.e., they are unable to transform EO sensory big data into EO value-adding products and services, in compliance with the GEO’s visionary goal of a Global Earth Observation System of Systems (GEOSS) {98} conditioned by the QA4EO *Call/Val* requirements {25}. This is tantamount to saying that past and present EO-IUSs have been outpaced by the rate of collection of EO sensory big data, whose quantity and quality are ever-increasing {28}. Several undisputable observations (true-facts) support this working hypothesis.

First, in recent decades a kind of Moore’s law of productivity appears applicable to the technological growth of imaging sensors, leading to the first information theory of quantitative *information-as-thing*, inherently “easy” because unequivocal: acquisition and transmission of sensory data is independent of the meaning of the transmitted message {5}. Unfortunately, the same increase of productivity does not appear to apply to EO-IUSs, leading to the second information theory of qualitative *information-as-data-interpretation*, inherently “difficult” because equivocal: since there is no semantics in sensory data, the message receiver has a pro-active role in providing sensory data with meanings {5}.

Second, in 2002 the percentage of EO data ever downloaded from the European Space Agency (ESA) databases was estimated at about 10% or less {29}.




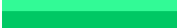







		Pseudocolor
A11	1. Cultivated and Managed Terrestrial (non-aquatic) Non-vegetated Areas	
A12	2. Natural and Semi-Natural Terrestrial Vegetation	
A23	3. Cultivated Aquatic or Regularly Flooded Vegetated Areas	
A24	4. Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation	
B35	5. Artificial Surfaces and Associated Areas	
B36	6. Bare Areas	
B47	7. Artificial Waterbodies, Snow and Ice	
B48	8. Natural Waterbodies, Snow and Ice.	
	9. Quality layer: Cloud	
	10. Quality layer: Cloud-shadow	
	11. Others	

Fig. 1-5. “Ideal” 3-level 8-class FAO Land Cover Classification System (LCCS) Dichotomous Phase (DP) legend augmented with cloud and cloud-shadow quality layers. Proposed as target ESA EO Level 2 scene classification map (SCM) legend, to be accomplished by an EO image understanding (EO-IU) subsystem integrated in a closed-loop EO-IU4SQ system prototype.

Third, no ESA EO data-derived Level 2 product has ever been generated systematically at the ground segment {30}. The ESA definition of EO Level 2 product comprises {30}: (i) a multi-spectral (MS) image corrected for atmospheric, adjacency and topographic effects {31}, {32}, {33}, stacked with (ii) its data-derived general-purpose, user- and application-independent scene classification map (SCM), whose legend includes quality layers such as cloud and cloud-shadow {30}, {34}, {35}. For example, the Sentinel-2 Level 2 product is not delivered at the ESA ground segment but generated on user side through the Sentinel-2 software Toolbox {30}, {34}, {35}. Noteworthy, within the  $CV \supset EO-IU$  domain, a National Aeronautics and Space Administration (NASA) EO Level 2 product, defined as “a data-derived geophysical variable at the same resolution and location as Level 1 source data” {115}, is a subset of ESA EO Level 2 product, i.e., the complexity of the latter is superior or equal (not inferior) to the complexity of the former.

Fourth, no existing EO content-based image retrieval (CBIR) system supports semantic querying, for example, “retrieve all images from sensor X not necessarily cloud-free where wetlands are visible and located adjacent to a highway near a coast in the eastern part of country Y” {36}, {37}, {38}, {39}. EO semantic CBIR (SCBIR) is synonym of “semantics-enabled information/knowledge discovery” in large-scale multi-source EO image databases {37}. In practice, no EO SCBIR system in operating mode exists to date because existing EO-CBIR systems lack EO image understanding capabilities.



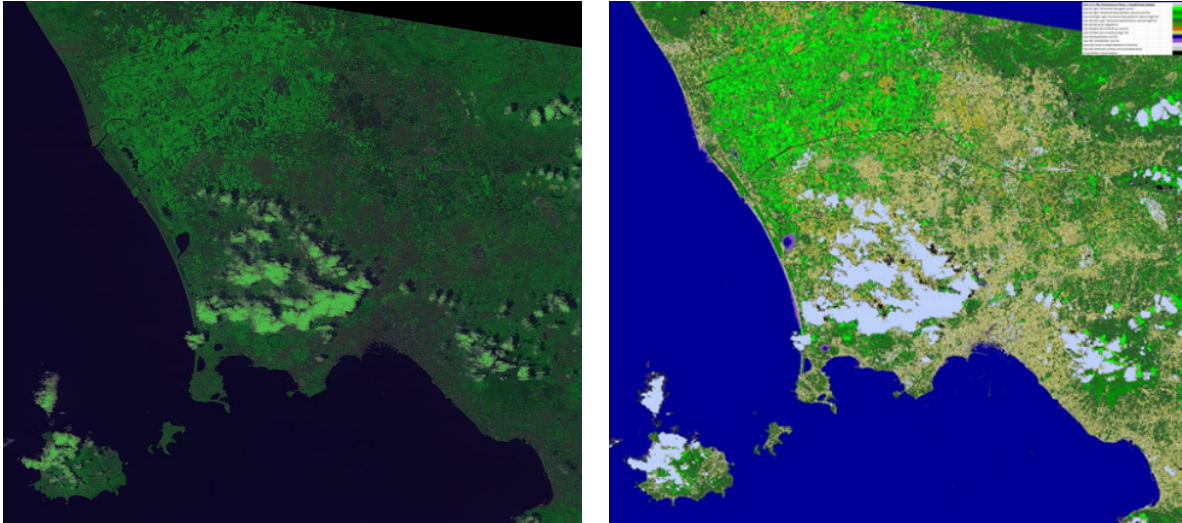


Fig. 1-6. Left: 4-band (visible blue - B, visible green - G, visible red - R, Near InfraRed - NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere (TOA) reflectance (TOARF) values and depicted in false colors (channel R = band R, channel G = band NIR, channel B = band B), 10 m resolution. Acquired on 2004-13-06. No histogram stretching is applied for visualization purposes. Right: 12-class classification map based on a convergence-of-evidence approach, in compliance with the FAO LCCS 3-level 8-class Dichotomous Phase (DP) legend, plus Cloud, Cloud-shadow and Unknown quality layer detection. Map legend: shown in Fig. 1-7.

	Pseudocolor
<b>Class #1 Vgtn Terrestrial Managed LowTxtr</b>	
<b>Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxtr</b>	
<b>Class #3 Bright Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #5 Residual Vegetation</b>	
<b>Class #6 Bare Soil Or Built-up LowTxtr</b>	
<b>Class #7 Bare Soil or BuiltUp High Txtr</b>	
<b>Class #8 DeepWater LowTxtr</b>	
<b>Class #9 TurbidWater LowTxtr</b>	
<b>Class #10 Cloud or Bright BareSoil Or BuiltUp</b>	
<b>Class #11 Artificial surfaces and associated areas</b>	
<b>Unclassified / cloud-shadow</b>	

Fig. 1-7. Implemented ESA EO Level 2 scene classification map (SCM) legend, consisting of 12 classes, approximately equivalent to a FAO Land Cover Classification System (LCCS) Dichotomous Phase (DP)-like 1<sup>st</sup>-level (veg/non-veg) and 2<sup>nd</sup>-level (water/terrestrial) + quality layers cloud and cloud-shadow. To be compared against the “ideal” 3-level 8-class FAO LCCS-DP legend augmented with cloud and cloud-shadow quality layers shown in Fig. 1-5. Visual features input to the implemented ESA EO Level 2 SCM classifier are: MS color names detected by the Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain {26}, {27}, {33}, {91}, {92}, {105}, and texture segmentation, automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range {0, 7}. Visual features available, but not yet employed as input by the implemented ESA EO Level 2 SCM classifier are: local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships. Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size. No human-machine interaction is required by the implemented ESA EO Level 2 product generator, neither for supplying training data nor for defining system’s free-parameters.

Fifth, EO-IUSs presented in the RS literature are typically assessed and compared based on their sole mapping accuracy, which means their mDMI set of OP-Q<sup>2</sup>Is remains largely unknown to date. For example, when a large-scale EO data-derived thematic map featuring “high” accuracy was generated by a supervised data learning EO-IUS, the most limiting



factors turned out to be the cost, timeliness, quality and availability of adequate supervised (labeled) data samples, collected from field sites, existing maps or geospatial data archives in tabular form {40}.

A secondary working hypothesis adopted in this doctoral project was that relationship  $SCBIR \supset CV \supset EO-IUS$  in operating mode  $\supset$  human vision holds true. It means that the complexity of SCBIR is not inferior to the complexity of vision, acknowledged to be an inherently ill-posed cognitive problem NP-hard in computational complexity, see Fig. 1-1. This is tantamount to saying that a necessary not sufficient pre-condition to tackle the SCBIR problem is computational solution of the cognitive problem of vision. Hence, the development of EO-IUSs in operating mode pre-dates the development of EO-SCBIR systems. In the inter-disciplinary framework of cognitive science (see Fig. 1-1), CV and SCBIR are considered two sides of the same cognitive problem, where solution of the latter depends on the solution of the former, see Fig. 1-2. Once an EO-IU subsystem is developed in operating mode as necessary not sufficient pre-condition, it can be connected in closed-loop with an EO semantic quering (EO-SQ) subsystem to form an integrated EO-IU for semantic querying (EO-IU4SQ) system architecture, capable of semantic querying large-scale multi-source EO image databases for semantics-enabled knowledge/information discovery {37}, {42}. According to this working hypothesis, claimed EO-SCBIR system prototypes cannot be considered in operating mode if they incorporate an EO-IU subsystem which falls short in operating mode {36}, {37}, {38}, {39}.

To accomplish the overarching goal of developing a closed-loop EO-IU4SQ system prototype capable of filling the information gap from EO big sensory data to EO value-adding information products and services as a GEOSS proof-of-concept in support of SCBIR, the primary goal of this doctoral research was to develop an automated near real-time multi-source EO-IU subsystem in operating mode. Provided with an innovative hybrid (combined deductive and inductive) feedback modular design, see Fig. 1-3, it must be capable of systematic ESA EO Level 2 product generation as necessary not sufficient pre-condition to initialize an EO-SQ subsystem. Selected from the well-known two-phase Food and Agriculture Organization of the United Nations (FAO) Land Cover Classification System (LCCS) {41}, see Fig. 1-4, the standard “fully nested” 3-level 8-class Dichotomous Phase (DP) legend was augmented with quality layers cloud and cloud-shadow to provide a general-purpose, user- and application-independent ESA EO Level 2 SCM taxonomy, see Fig. 1.5.

An example of ESA EO Level 2 SCM product automatically generated by the implemented EO-IU subsystem is depicted in Fig. 1-6, whose map legend is shown in Fig. 1-7. This present instantiation of ESA EO Level 2 SCM legend should be compared with the “ideal” realization of a 3-level 8-class FAO LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow, shown in Fig. 1-5, to be accomplished by future developments of the EO-IU subsystem implemented in the closed-loop EO-IU4SQ system prototype. Visual features input to the implemented ESA EO Level 2 SCM classifier are MS color names, automatically detected by the Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain {26}, {27}, {33}, {91}, {92}, {105}, and texture segmentation, automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range {0, 7}. Although they have been computed, visual information primitives not yet employed as input by the implemented ESA EO Level 2 SCM classifier are local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships. Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size, which means near real-time in practice. No human-machine interaction is required by the implemented ESA EO Level 2 product generator, neither for supplying training data nor for defining system’s free-parameters.

The analytic description of the automated near real-time multi-source EO-IU subsystem architecture shown in Fig. 1-3 can be summarized as follows. In a CV system where image understanding is based on convergence-of-evidence {3}, any spatial unit  $x$ , with  $x$  either (0D) point, (1D) line or (2D) polygon in the image-domain according to the Open Geospatial Consortium (OGC) Simple Feature Specification nomenclature {108}, is described by the following well-known visual information primitives {43}.

- Color, including panchromatic luminance (intensity) {44}, {45}, {46}.
- Local shape {47}, {48}, {49} and size of image-objects.
- Texture, defined as perceptual spatial grouping of texture elements (texels) {49}, {50}, {51}, {52}, {53}.
- Inter-object spatial relationships, either topological or non-topological {3}, {24}, {43}.

These visual features of a spatial unit  $x$  in the (2D) image-plane are modeled as numeric variables  $ColorValue(x)$ ,  $localShapeValue(x)$ ,  $TextureValue(x)$  and  $SpatialRelationships(x, Neigh(x))$ , where  $ColorValue(x)$  is a vector data belonging to a multi-spectral (MS) data space  $\mathfrak{R}^{MS}$ , i.e.,  $ColorValue(x) \in \mathfrak{R}^{MS}$ , and  $Neigh(x)$  is a generic 2D spatial neighborhood of



spatial unit  $x$  in the image-domain. Irrespective of their Pearson's cross-correlation, if any, it is easy to prove these visual attributes are statistically independent, because cross-correlation does not mean causation {114}. Let's consider a taxonomy of real-world objects in the 4D scene-domain as categorical and semantic variable  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ . The Bayesian law provides a principled way of combining new evidence stemming from data with prior beliefs available in addition to and before looking at data. According to the Bayesian law, if priors are ignored because considered equiprobable in a maximum *a posteriori* optimization criterion and if the multiple sources of evidence satisfy a "naïve" hypothesis of statistical independence, then the so-called naïve Bayes classifier formulation becomes

$$\begin{aligned}
 & p(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\
 & p(\text{ColorValue}(x) | c) \cdot p(\text{ShapeValue}(x) | c) \cdot p(\text{TextureValue}(x) | c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c) = \\
 & \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} p(\text{ColorValue}(x) | \text{ColorName}) p(\text{ColorName} | c) \cdot p(\text{ShapeValue}(x) | c) \cdot \\
 & p(\text{TextureValue}(x) | c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c), \quad c = 1, \dots, \text{ObjectClassLegendCardinality}, \quad (1-1)
 \end{aligned}$$

with probability  $p(\cdot) \in [0, 1]$  and where color space  $\mathfrak{R}^{\text{MS}}$  has been partitioned into a totally exhaustive and mutually exclusive set of hyperpolyhedra, equivalent to a static Vector Quantization (VQ) codebook of codewords {12}. This set of hyperpolyhedra or VQ codebook is associated with a discrete and finite dictionary of static (non-adaptive-to-data) color names, equivalent to a latent/hidden variable which links input data (observables) in the physical world with output classes in the modeled world, where  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$  {45}, {54}, {55}, see Fig. 1-8.

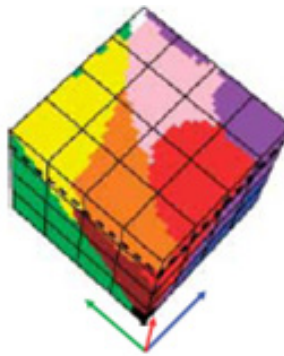


Fig. 1-8. Reproduced with permission, courtesy of {45}. Monitor-typical red-green-blue (RGB) cube partitioned by Griffin into perceptual polyhedra {54}, corresponding to a discrete and finite dictionary of basic color (BC) names proposed by Berlin and Kay {45}, to be community-agreed upon in advance to be employed by members of the community. The mutually exclusive and totally exhaustive polyhedra are neither necessarily convex nor connected. In practice BC names belonging to a finite and discrete color dictionary are equivalent to Vector Quantization (VQ) levels belonging to a VQ codebook {12}.

To further simplify Eq. (1-1), its canonical interpretation based on frequentist statistics can be relaxed by fuzzy logic {56}, so that the logical-AND operator is replaced by a fuzzy-AND (min) operator, inductive class-conditional probability  $p(\text{ColorValue}(x) | c) \in [0, 1]$ , where  $\sum_{c=1}^{\text{ObjectClassLegendCardinality}} p(\text{ColorValue}(x) | c) \geq 0$ , is replaced by a deductive crisp (binary) membership (compatibility) function  $m(\text{ColorValue}(x) | c) \in \{0, 1\}$ , where  $\sum_{c=1}^{\text{ObjectClassLegendCardinality}} m(\text{ColorValue}(x) | c) = 1$ , with color space hyperpolyhedra considered mutually exclusive and totally exhaustive. If these simplifications are adopted, then Eq. (1-1) becomes

$$\begin{aligned}
 & m(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\
 & \min \left\{ \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x) | \text{ColorName}) m(\text{ColorName} | c), m(\text{ShapeValue}(x) | c), \right. \\
 & \left. m(\text{TextureValue}(x) | c), m(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c) \right\} =
 \end{aligned}$$



$$\min \{m(\text{ColorName}^* | c), m(\text{ShapeValue}(x) | c), m(\text{TextureValue}(x) | c), m(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c)\}, c = 1, \dots, \text{ObjectClassLegendCardinality}, \text{ where } \text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}, \text{ such that } m(\text{ColorValue}(x) | \text{ColorName}^*) = 1 \text{ and } m(\text{ColorName}^* | c) \in \{0, 1\}. \quad (1-2)$$

In Eq. (1-2), the following considerations hold.

- Any numeric  $\text{ColorValue}(x)$  in color space  $\mathfrak{R}^{\text{MS}}$  belongs to a single hyperpolyhedron, identified by  $\text{ColorName}^*$  in the static color dictionary, i.e.,  $\forall \text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , then  $\text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}$  exists such that  $\sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x) | \text{ColorName}) = m(\text{ColorValue}(x) | \text{ColorName}^*) = 1$  holds, where  $m(\text{ColorValue}(x) | \text{ColorName}) \in \{0, 1\}$ ,  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ , see Fig. 1-8.
- Color names, physically equivalent to color hyperpolyhedra in a numeric color space  $\mathfrak{R}^{\text{MS}}$ , are conceptually equivalent to a latent/hidden variable linking observable (“sub-symbolic”) data in the real-world (physical world) to categorical variables of semantic (symbolic) quality in the modeled world.

		Target classes of individuals (entities in a conceptual model for knowledge representation built upon an ontology language)			
		Class 1, Water body	Class 2, Tulip flower	Class 3, Italian tile roof	
Color names	black			√	
	blue		√	√	
	brown		√	√	
	grey				
	green		√	√	
	orange			√	
	pink			√	
	purple			√	
	red			√	√
	white			√	
	yellow			√	

Table 1-1. Example of a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{DictionaryOfColorNames}$ , with cardinality  $|A| = a = \text{ColorDictionaryCardinality} = 11$ , and the set  $B = \text{LegendOfObjectClassNames}$ , with cardinality  $|B| = b = \text{ObjectClassLegendCardinality} = 3$ . The latter dictionary is a superset of the typical taxonomy of land cover (LC) classes adopted by the RS community. “Correct” entry-pairs (marked with  $\checkmark$ ) must be: (i) selected by domain experts based on a hybrid combination of deductive prior beliefs with inductive evidence from data and (ii) community-agreed upon.

- Set  $A = \text{DictionaryOfColorNames}$ , with cardinality  $|A| = a = \text{ColorDictionaryCardinality}$ , and set  $B = \text{LegendOfObjectClassNames}$ , with cardinality  $|B| = b = \text{ObjectClassLegendCardinality}$ , can be considered a bivariate categorical random variable where two univariate categorical variables  $A$  and  $B$  are generated from a single population. A binary relationship from set  $A$  to set  $B$ ,  $R: A \Rightarrow B$ , is a subset of the 2-fold Cartesian product (product set)  $A \times B$ , whose size is rows  $\times$  columns =  $a \times b$ . The Cartesian product of two sets  $A \times B$  is a set whose elements are ordered pairs. Hence, the Cartesian product is non-commutative,  $A \times B \neq B \times A$ . In agreement with common sense, see Table 1-1,  $R: \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  is a set of ordered pairs where each  $\text{ColorName}$  can be assigned to none, one or several target classes  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  of observed scene-objects, whereas each class of observed objects can be assigned with none, one or several color names as instances of a class-specific color attribute. Binary membership values  $m(\text{ColorName} | c) \in \{0, 1\}$  and  $m(c | \text{ColorName}) \in \{0, 1\}$ , with  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  and  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ , can be community-agreed upon based on various kinds of evidence, whether viewed all at once or over time, such as a combination of



prior beliefs with additional evidence inferred from new data in agreement with a Bayesian updating rule (Bayesian inference), largely applied in artificial intelligence (AI) and expert systems. A binary relationship  $R: A \Rightarrow B \subseteq A \times B$  where sets  $A$  and  $B$  are categorical variables generated from a single population guides the interpretation process of a two-way *contingency table* (also known as association matrix, cross tabulation, bivariate table or frequency table) {57}, {58}, {59}, {60},  $BIVRTAB = \text{FrequencyCount}(A \times B)$ . In the conventional domain of frequentist inference with no reference to prior beliefs, a BIVRTAB is the 2-fold Cartesian product  $A \times B$  instantiated by the bivariate frequency counts of the two univariate categorical variables  $A$  and  $B$  generated from a single population. For any BIVRTAB instance, either square or non-square, there is a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  that guides the interpretation process, where “correct” entry-pair cells of the 2-fold Cartesian product  $A \times B$  can be either off-diagonal (scattered) or on-diagonal, if a main diagonal exists. When a BIVRTAB is estimated from a geospatial population without sampling, it is called *overlapping area matrix* (OAMTRX) {61}, {62}. When the binary relationship  $R: A \Rightarrow B$  is a bijective function (both 1-1 and onto), i.e., when the two categorical variables  $A$  and  $B$  estimated from a single population coincide, then the BIVRTAB is square and sorted and typically called confusion matrix (CMTRX) or error matrix {57}, {58}, {59}, {60}. In a (square and sorted) CMTRX the main diagonal guides the interpretation process. For example, a square OAMTRX =  $\text{FrequencyCount}(A \times B)$ , where  $A$  = test thematic map legend,  $B$  = reference thematic map legend and cardinality  $a = b$ , is a CMTRX if and only if  $A = B$ , i.e., if the test and reference codebooks are the same sorted set of concepts. In general the class of (square and sorted) CMTRX instances is a special case of the class of OAMTRX instances, either square or non-square, i.e.,  $OAMTRX \supseteq CMTRX$ . A similar consideration holds about summary  $Q^2$ Is generated from an OAMTRX or a CMTRX, i.e.,  $Q^2I(OAMTRX) \supseteq Q^2I(CMTRX)$ .

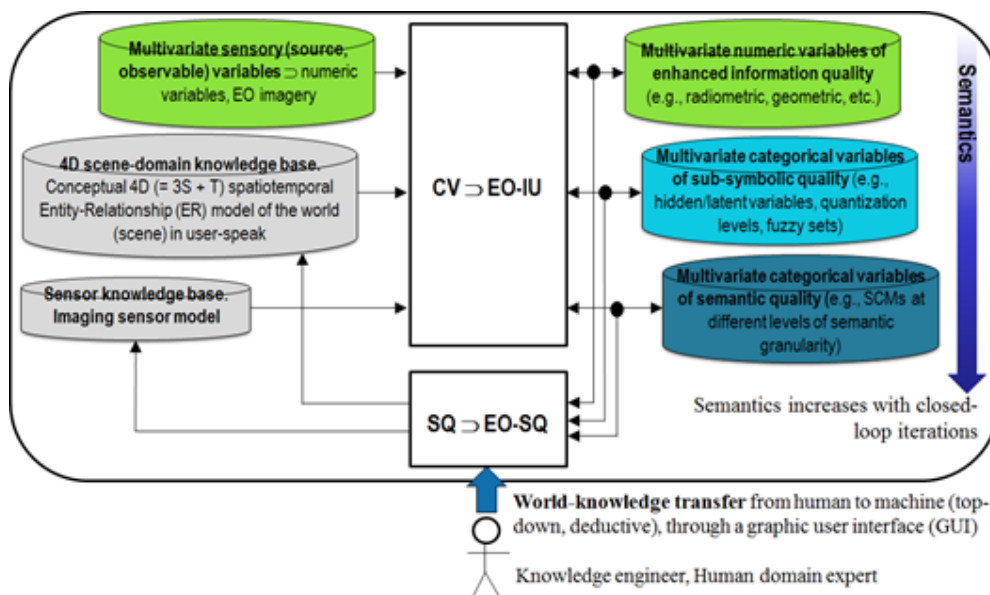


Fig. 1-9. Top-level modular design of a closed-loop EO image understanding (EO-IU) for semantic querying (EO-IU4SQ) system architecture, suitable for incremental learning. It comprises a primary (dominant, necessary not sufficient) hybrid (combined deductive and inductive) EO-IU subsystem in closed-loop with a secondary (dominated) hybrid EO-SQ subsystem. The EO-IU subsystem must be automatic (requiring no human-machine interaction) and near real-time to provide the EO-SQ subsystem with useful information products, including Scene Classification Maps (SCMs), as initial necessary not sufficient pre-condition of symbolic quality for semantic querying and semantics-enabled information/knowledge discovery. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline high-level user- and application-specific EO image interpretation and SCBIR operations. Output products generated by the closed-loop EO-IU4SQ system monotonically increase their value-added with closed-loop iterations.

In simple analytic terms Eq. (1-2) shows that for any spatial unit  $x$  in the image-domain and any class label  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ , when a hierarchical CV classification approach estimates posterior  $m(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x)))$  starting from a near real-time context-insensitive color naming first stage {45}, {54}, {55}, where condition  $m(\text{ColorValue}(x) | \text{ColorName}^*) = 1$  holds, see Fig. 1-8, if

condition  $m(\text{ColorName}^* | c) = 0$  is true according to a static community-agreed binary relationship  $R$ :  $\text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  known *a priori*, see Table 1-1, then  $m(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) = 0$  irrespective of any second-stage assessment of spatial terms  $\text{ShapeValue}(x)$ ,  $\text{TextureValue}(x)$  and  $\text{SpatialRelationships}(x, \text{Neigh}(x))$ , whose computational model is typically difficult to find and computationally expensive. Intuitively Eq. (1-2) shows that static color naming {45}, {54}, {55} allows the stratification of unconditional multivariate spatial variables into color class-conditional data distributions, in agreement with the statistic stratification principle {63} and the divide-and-conquer problem solving approach {12}, {64}. Well known in statistics, the principle of statistic stratification guarantees that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” {63}.

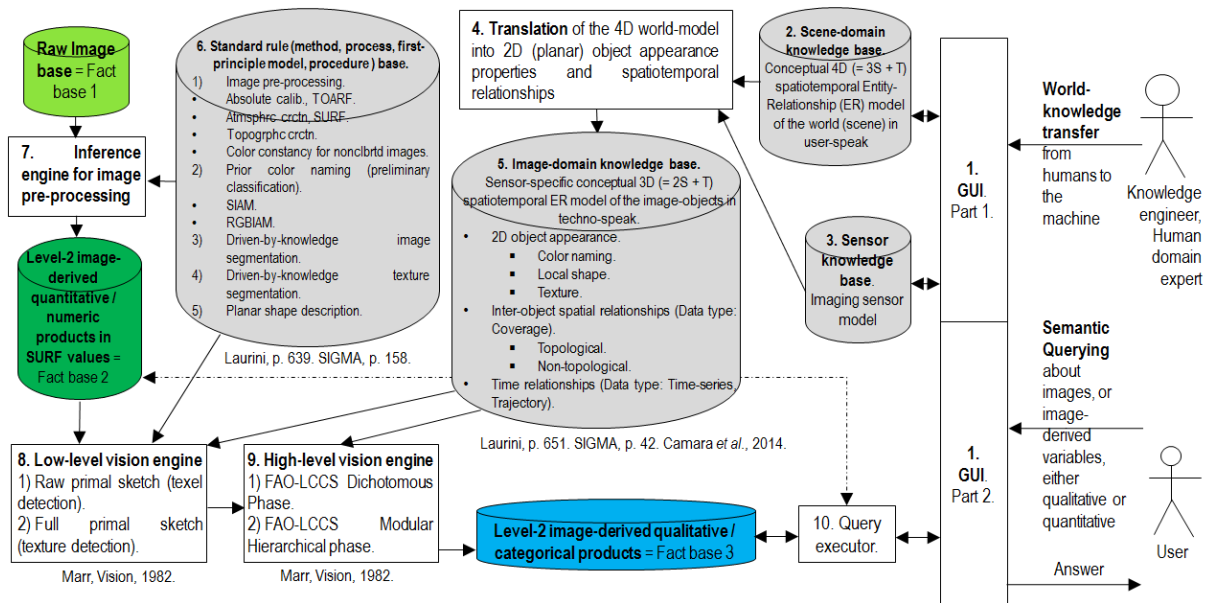


Fig. 1-10. EO-IU4SQ system architecture proposed at a finer level of detail than the top-level modular design shown in Fig. 1-9. Processing modules are shown as rectangles and databases as cylinders. Acronyms shown in this figure means the following: graphic user- interface (GUI), surface reflectance (SURF), top-of-atmosphere reflectance (TOARF), Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain {26}, {27}, {33}, {91}, {92}, {105}, RGB Image Automatic Mapper (RGBIAM) lightweight computer program for RGB space polyhedralization into color names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain {116}, FAO Land Cover Classification System (LCCS) taxonomy {41}.

The secondary objective of this doctoral project was to develop an integrated EO-IU4SQ system prototype as a GEOSS proof-of-concept in support of SCBIR, see Fig. 1-9 and Fig. 1-10, where a secondary (dependent) EO-SQ subsystem was connected in closed-loop to a primary (dominant, necessary not sufficient) EO-IU subsystem in operating mode. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline human-machine interaction in support of spatiotemporal EO big data analytics and SCBIR operations, see Fig. 1-11 {42}.

This dissertation reports on the doctoral research and technical development (R&D) of each information processing block identified by the EO-IU subsystem architecture sketched in Fig. 1-3 and on their integration in a prototypical implementation of a closed-loop EO-IU4SQ system whose design agrees with Fig. 1-9 and Fig. 1-10.

The rest of this document is organized as follows. To make this dissertation self-contained, Chapter 2 reports on topics and definitions of interests considered preliminary to the rest of this doctoral dissertation. Provided with a relevant survey value, Chapter 3 presents and discusses novel computational models of human vision at the core of the proposed EO-IU4SQ system. An integrated view of the whole EO-IU4SQ system prototype design and implementation is provided in Chapter 4 and Chapter 5. An application of methods and algorithms adopted as standard rule base by the EO-IU4SQ system (see Fig. 1-10) is described in Chapter 6. Chapter 7 to Chapter 12 consist of original working papers where individual information processing blocks identified by the EO-IU subsystem architecture shown in Fig. 1-3 are discussed in detail.

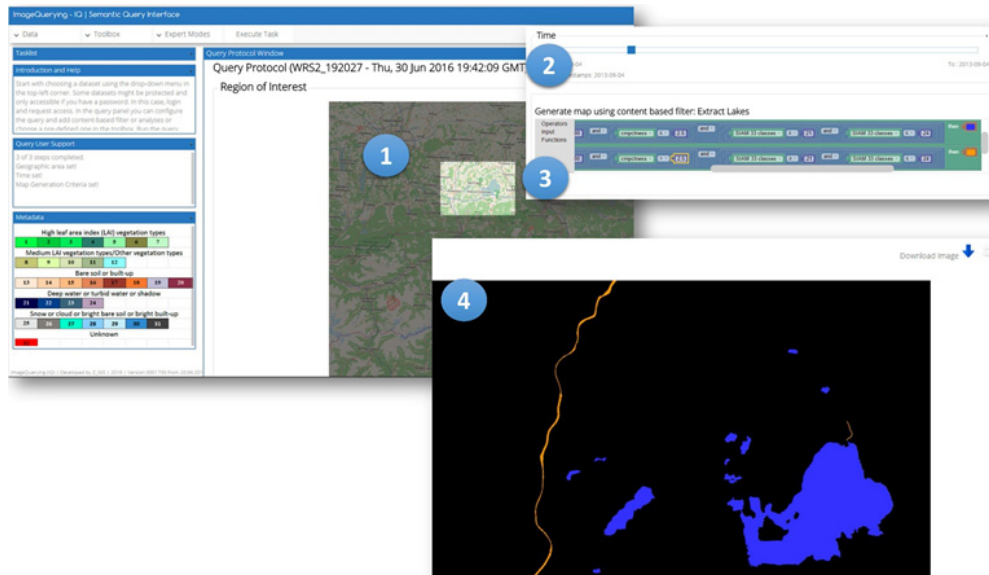


Fig. 1-11. Instantiation of the graphic user interface (GUI) implemented in a secondary (dependent) EO-SQ subsystem connected in closed-loop to a primary (dominant, necessary not sufficient) EO-IU subsystem, to form an integrated EO-IU4SQ system prototype. Example of a semantic querying to infer new information layers from the fact base (1). A Landsat-8 image time-series is analysed by a spatiotemporal semantic query (2), (3), to distinguish between lakes and rivers as water areas (spectral information) featuring different size and planar shape attributes. Target values of these numeric attributes expressed in user-speak in the scene-domain, including physical units of measure, are provided by the world model. They are mapped onto the image-plane in techno-speak, e.g., in pixel units, by the imaging sensor transfer function. In the image-domain, each image-object stored in the fact base is always provided with at least one symbolic label, together with sub-symbolic categorical and numeric attributes as visual information primitives color by name, texture, shape indexes and size, in addition to inter-object spatial and non-spatial relationships. For example, image-objects provided with the LCCS-DP class label B48, Natural Waterbodies, can be easily discriminated into the LCCS Modular Hierarchical Phase (MHP) classes Lake and River based on planar shape and size properties (4).



## 2 Introduction

To comply with the dissertation requirements specification proposed in {1}, {2}, where an ideal thesis as Doctor of Philosophy (PhD) is required to be accessible to everyone in engineering/computer science, not just to specialists, this Chapter reports on the remote sensing (RS) problem background and definitions of interest adopted by the rest of this doctoral dissertation.

### 2.1 Taxonomy of EO Imaging Platforms and Sensors

EO images can be acquired by multiple platforms categorized as follows.

- Terrestrial, including portable devices, such as cellular phones, typically mounting uncalibrated color cameras, e.g., either true- or false-color red-green-blue (RGB) color cameras, including low-cost consumer-level RGB cameras. Noticeably, terrestrial light detection and ranging (LiDAR) 3D data point clouds are acquired in combination with intensity images {65} or MS images {66}. The global LiDAR market, including terrestrial and airborne, is anticipated to expand at a 15.0% rate from 2014 to 2020, growing from a value of US\$225 Mn in 2013 to US\$605 Mn in 2020 {67}.
- Unmanned Aircraft Systems (UASs). For a taxonomy of UASs, refer to {68}, {117}. Typically, the unmanned aerial vehicles (UAVs) market is segmented on the basis of type into commercial and military drones. UAVs are segmented on the basis of payload into commercial drones (up to 25 Kg) and military drones (up to 150 Kg, up to 600 Kg, and above 600 Kg). Prominent segmentations involved with the non-military commercial and consumer UAS market are the following. Type: Fixed Wing, Rotary Wing. Property: Commercial UAS by endurance and altitude (LowEndurance LowAltitude, AverageEndurance LowAltitude, AverageEndurance AverageAltitude, HighEndurance HighAltitude), Consumer UAS by range type (LowRange, AverageRange). The overall UAS market displays two distinct characteristics, that is, the military drones is a mature market which growing at a low rate. On the other hand, commercial and civilian sector UAVs is a booming market and are expected to grow rapidly in the next few years. Americas is the major market for military as well as commercial UASs. Europe is the second largest market contributing to the overall revenue. According to a report from Information Gatekeepers, Inc. (IGI) Consulting Inc. the U.S. market for UAVs is estimated to grow from \$5 billion in 2013 to \$15 billion in 2020 {69}. The drivers for this market are rapid technological advancements in UAVs and the growing demand for drone-generated data in commercial applications. The main restraint for this market is social issues such as privacy and nuisance concerns. Worldwide, UAS-based agricultural services, which means precision agriculture (PA) applications for commercial UAVs, are estimated to grow at the highest rate {69}, {70}. Over the next several decades, the global population is expected to grow rapidly. The Food and Agriculture Organization (FAO) of the United Nations expects the global population by 2050 to require a 70% increase in food production while arable land is expected to increase by less than 5%. One of the ways to offset this disparity will be to apply technological innovations to the farming industry to increase peracre crop yields while preserving resources. In recent years, as inexpensive time-saving technologies have come to market, there has been a growing interest on PA. "Precision agriculture (PA) or satellite farming or site-specific crop management is a farming management concept based on observing, measuring and responding to inter and intra-field variability in crops" {69}. The final goal of PA research is to define a decision support system capable of transforming sensory data into useful information to take decisions about optimization of returns while preserving resources at the farm- and crop-specific spatial extent. The practice of PA has been enabled by the advent of global navigation satellite systems, such as the U.S. Global Positioning System (GPS). The farmer's and/or researcher's ability to locate their precise position in a field allows for the creation of maps of the spatial variability of as many variables as can be measured, e.g. crop yield, terrain features/topography, organic matter content, moisture levels, nitrogen levels, pH, etc. In recent years, instead of having to inspect crops by walking through them, farmers and crop scouts have increasingly turned to spaceborne and airborne imaging to evaluate crop health. UAVs are useful for agricultural planners as they greatly reduce the time and cost required to conduct an accurate survey and offer real-time data collection, high resolution imagery of farmland, and are low cost in comparison with other related techniques, e.g., spaceborne and aerial {69}, {70}. Multispectral (MS) and thermal infra-red (TIR) sensors were until recently too heavy and bulky for small UAV platforms, even though their potential was demonstrated almost a decade ago {71}. Nowadays, however, lightweight multispectral and thermal sensors on small UAVs are commercially available and



are used in multiple applications including PA to assess vitality of the plants, crop yield estimations and other vegetation monitoring applications {71}, {72}, {73}, {74}.

- Airborne, excluding UAVs. For technical details, refer to {75}, {76}. “The global aerial imaging market, valued at USD 0.97 billion in 2013, is expected to see strong growth of 13.4% from 2014 – 2020” {77}.

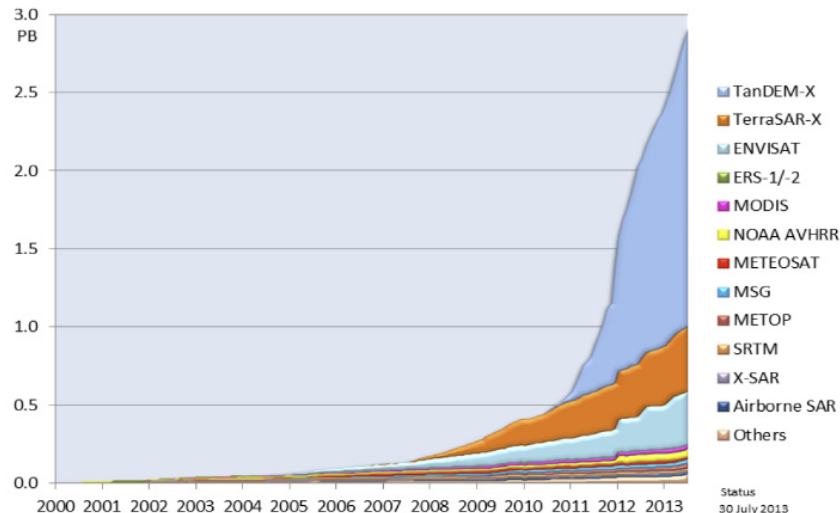


Fig. 2-1. Reproduced with permission, courtesy of {82}. DLR archive size in petabytes (PB) of optical and SAR sensory data. Status on July 30, 2013.

- Spaceborne. For technical details, refer to {75}, {76}. In recent years, continuous improvements in the spatial, spectral and temporal resolution of satellite imaging sensors fostered a dramatic increase in the quantity and quality of spaceborne EO sensory data potentially available to the general public. For example, from year 1997 to year 2003 the size of the U.S. Geological Survey (USGS) active Landsat archive increased exponentially from less than 10 to 434 terabytes, made up of 31 years of Landsat 1–5 acquisitions accounting for 165 terabytes, plus four years of Landsat 7 acquisitions accounting for 269 terabytes {78}. According to the National Aeronautics and Space Administration (NASA)’s Earth Observing System Data and Information System (EOSDIS) metrics for 2014, the EOSDIS manages more than 9 petabytes (PB) of data {79}. EO sensory data ever archived by ESA since 1975 range from 3 to 10 PB {80}. In Feb. 2017, DigitalGlobe’s 17-year, time-lapse library of high-resolution commercial satellite imagery amounts to more than 100 PB {81}. Similar considerations hold for the German Aerospace Center (DLR), whose EO big data repositories encompass spaceborne imaging sensors both optical and synthetic aperture radar (SAR), see Fig. 2-1 {82}. Fig. 2-1 shows there is a kind of Moore’s law of technological growth and productivity applicable to imaging sensors (submitted to the first information theory, quantitative *information-as-thing*, inherently “easy” because unequivocal: acquisition and transmission of sensory data is independent of the meaning of the transmitted message {5}). “Globally, the satellite imaging market was valued at USD 2,054.5 million in 2012 and is forecast to grow at 13.9% from 2013 - 2019. The data collected by satellites images has commercial value across industries, including commercial enterprises, civil engineering, military, forestry & agriculture, energy sectors and insurance, among others. Global commercial satellite imaging market in 2012 was dominated by the military segment, which accounted for 29.2% revenue share. This technology is mainly used in the energy sector, geospatial technology, conservation & natural resources management, construction & development, media & entertainment, disaster response management, defense & intelligence among others. Geospatial technology, energy and natural resource management together accounted for approximately 41.8% of market revenue share in 2012. Governments purchase commercial satellite imagery in order to support national security reconnaissance activities, climate change research, weather prediction, and land management activities. Growth of the commercial satellite imagery is driven by increasing demand from defense sector, predominantly by countries with large imagery intelligence requirements. Currently, due to rising terrorism concerns, defense and intelligence departments all over the world are seeking ways to support their security initiatives using satellite imagery. Geographically, North America is expected to remain largest market for commercial satellite imagery followed by Europe. North America and Europe collective had revenue share of 70.7% in 2012.





Market participants include renowned companies such as DigitalGlobe Inc. (including GeoEye Inc.), Astrium Geo, who are currently dominating the market space. For example, GeoEye and DigitalGlobe represented approximately 65.1% of commercial satellite imagery market in 2012” {83}. According to the Euroconsult’s latest report on Prospects for the Small Satellite Market {84}, we are on the cusp of a major revolution for the space sector and overall space ecosystem, as more than 3,600 smallsats are expected to be launched through 2025, a significant increase from the previous decade. The total market value of these satellites is anticipated to be \$22 billion (manufacture and launch), a 76% increase over that of 2006-2015. This rate of growth is unprecedented for the space sector and will bring about fundamental changes as both new and established industry players attempt to increase their capabilities in order to gain market share.

Imaging sensors are categorized as either active or passive.

- ✓ Active sensors.
  - Synthetic Aperture Radar (SAR) imagery. For technical details, refer to {75}, {76}. In the present dissertation, bi-temporal SAR images represented in a monitor-typical Red-Green-Blue (RGB) cube as proposed in {85}, called bi-temporal RGB-SAR images, are considered of potential interest.
  - Light Detection and Ranging (LiDAR) {67}. For more details, refer to {65}. LiDAR is a RS method that uses light in the form of a pulsed laser to measure ranges (variable distances) of 3D scene-objects from the LiDAR source. 3D point cloud data acquired by LiDAR provide an either dense or sparse sampling of the physical surface in a real-world scene. 3D point cloud data are typically unstructured, irregularly distributed and not carrying any semantics of the depicted 3D objects. 3D points require data structures and processing methods different from those developed for (2D) imagery and merit attraction for a variety of applications, ranging from 3D scene reconstruction to 3D scene interpretation, from 3D object detection to 3D object recognition. Modern scanning systems allow to acquire a regular grid of intensity (luminance) information together with sparse range information acquired on the discrete scan grid. These (2D) intensity images provide an information source complementary to range data. In recent years the first MS airborne laser scanners have been launched, where MS information is for the first time directly available for 3D airborne laser scanning point clouds {66}. Typically, keypoints are detected as feature correspondences in multi-viewing intensity image sequences. For each feature correspondence, the respective 2D keypoints may be projected to the 3D space by considering the respective range information. This yields sparse sets of corresponding 3D points for 3D point cloud registration {65}.
- ✓ Passive (optical) imaging sensors, either radiometrically calibrated or uncalibrated, see Table 2-1. For technical details, refer to {75}, {76}. Radiometric calibration (*Cal*) is the transformation of dimensionless digital numbers into a community-agreed physical unit of radiometric measure, e.g., top-of-atmosphere (TOA) radiance (TOARD) in range  $[0, \infty)$ , TOA reflectance (TOARF) in range  $[0, 1]$  or surface reflectance (SURF) in range  $[0, 1]$  {25}. Although overlooked in the RS common practice, EO data *Cal* is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” {86} (p. 29), {87}, {88}. For example, *Cal* is considered mandatory by the GEO QA4EO *Cal/Val* requirements {25}. Non-calibrated panchromatic or MS pixels are equivalent to, respectively, a scalar or a multivariate numeric variable provided with no physical meaning.

In general, EO imaging sensor specifications encompass the following.

- (i) Spectral resolution, described by a sensor sensitivity curve as a function of electromagnetic wavelength, see Fig. 2-2. With regard to the EO imaging sensor’s spectral resolution, the rest of this work adopts the terminology shown in Table 2-1.
- (ii) Spatial resolution per pixel. With regard to the EO imaging sensor’s spatial resolution, the rest of this work adopts the following terminology.
  - Very high spatial resolution (VHR):  $< 1$  m per pixel side.
  - High spatial resolution (HR):  $[1$  m,  $30$  m].
  - Medium spatial resolution (MR):  $(30$  m,  $1$  km).
  - Low spatial resolution (LR):  $\geq 1$  km.

- (iii) Temporal resolution (revisit time), e.g. daily, weekly, number of times per year.
- (iv) Radiometric resolution per pixel, e.g., 1 byte equivalent to 256 gray levels.
- (v) Instantaneous coverage, specifically, swath width and acquisition length in km.
- (vi) If optical imagery, is EO data radiometric *Cal* supported in compliance with the GEO QA4EO *Cal/Val* requirements {25}? If *Cal* is supported, are radiometric *Cal* metadata parameter file specifications available?
- (vii) Costs per full scene or km<sup>2</sup>.

According to Chapter 1, in this PhD work EO data sources considered of potential interest to be transformed into timely, operational and comprehensive EO valued-added products and services in compliance with the GEO QA4EO *Cal/Val* requirements {25} are summarized as follows.

- Spaceborne bi-temporal RGB-SAR imagery {85}.
- Passive (optical) EO images, either panchromatic or color images, specifically, RGB imagery either true- or false-color, MS, SS or HS, see Table 2-1, either calibrated or uncalibrated, acquired from imaging sensors mounted on terrestrial, UAV, airborne and spaceborne platforms.
- Fused (2D) image and 3D point cloud data, with special emphasis on 2D keypoint detection and description for keypoint-based 3D point cloud registration {65}.
- EO images of any possible spatial resolution, from LR to VHR.

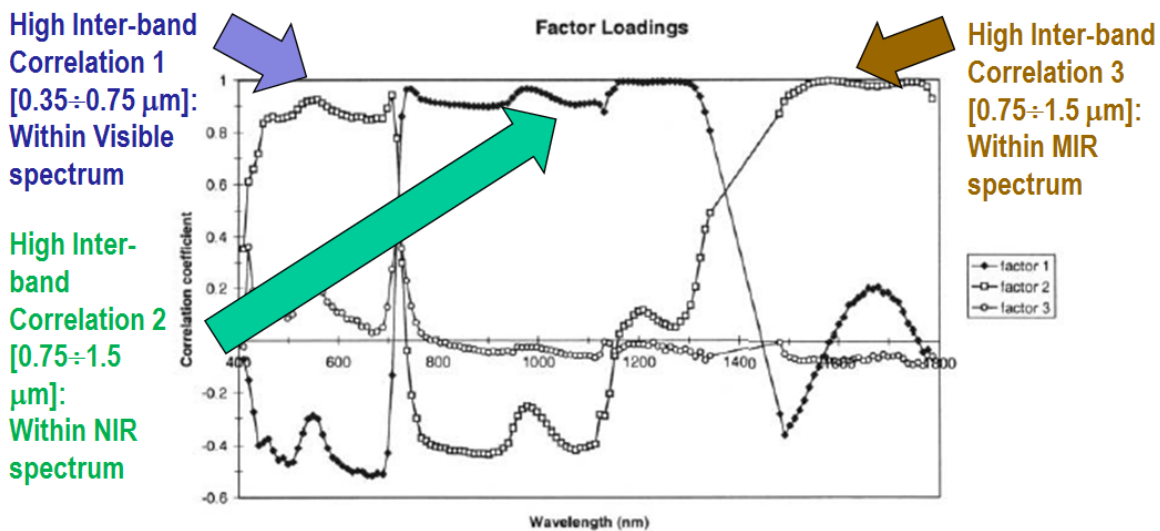


Fig. 2-2. Derived from {89}. Pearson’s cross-correlation (CC) coefficients for the main factors resulting from a principal component analysis and factor rotation for an agricultural data set based on spectral bands of the AVIRIS hyper-spectral (HS) spectrometers 1, 2 and 3. Flevoland test site, July 5th 1991. Inter-band CC values are “high” (> 0.8) within the visible spectral range, the Near Infra-Red (NIR) wavelengths and the Medium IR (MIR) wavelengths. The general conclusion is that, irrespective of non-stationary local information, the global (image-wide) information content of a multi-spectral (MS) image whose number  $N$  of spectral channels  $\in \{2, 9\}$ , a super-spectral (SS) image with  $N \in \{10, 20\}$ , or an hyperspectral (HS) image with  $N > 20$ , can be preserved by selecting one visible, one NIR, one MIR and one thermal IR (TIR) band, such as in the spectral resolution of the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) imaging sensor series in operating mode from 1978 to date.

Terminology / Acronym	No. of spectral bands
Multi-spectral (MS)	{2, 9}
Super-spectral (SS)	{10, 20}
Hyper-spectral (HS)	> 20

Table 2-1. Terminology: multi-spectral (MS), super-spectral (SS) and hyper-spectral (HS) optical imagery.

## 2.2 Spectral Resolution Specifications of EO Optical Imaging Sensors

According to Chapter 2.1, different EO imaging sensors are parameterized by different sensor specifications, whose core information encompasses spatial, spectral and temporal resolution. The present Chapter focuses on EO spectral resolution specifications, with special emphasis on spaceborne and airborne optical imaging sensors, including UAV's, see Table 2-2 to Table 2-4.

In {89}, {90}, a continuous spectral signature approximated by an HS imaging sensor shows that inter-band Pearson's cross-correlation (CC) coefficients are "high" (> 0.8) within the visible spectral range, the near infrared (NIR) wavelengths and the medium infrared (MIR) wavelengths, see Fig. 2-2. The important conceptual observation stemming from Fig. 2-2 is that, in general, the global (image-wide) information content of an HS image, irrespective of non-stationary local information, can be compressed into one visible, one NIR, one MIR and one thermal infrared (TIR) band. Noteworthy, this is the four-band spectral resolution of the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) imaging sensor series in operating mode from 1978 to date {75}, refer to Table 2-4.

Atmo-spheric window	Spectral region ( $\mu\text{m}$ )	Electromagnetic spectrum in optical wavelengths ( $\mu\text{m}$ )	Spectral region conventional name
1	0.3-1.3	Violet, V: 0.35-0.45. Blue, B: 0.45-0.50. Green, G: 0.50-0.57. Yellow, Y: 0.57-0.60. Orange, O: 0.60-0.65. Red, R: 0.65-0.72. 0.72-1.3	Visible Near IR (NIR)
2	1.5-1.8	1.3-3.0 (actually, 1.3-7.0)	Middle IR (MIR)
3	2.0-2.6		
4	3.0-3.6		
5	4.2-5.0		
6	7.0-15.0	7.0-15.0 (excluding TIR) 8.0-12.0	Far IR (FIR) Thermal IR (TIR)

Table 2-2. The optical spectrum and the atmospheric windows {75}. Also refer to Fig. 2-3.

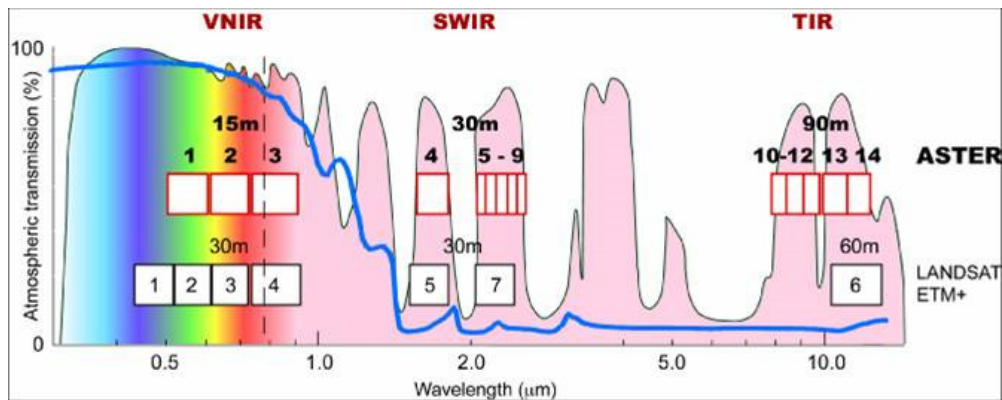


Fig. 2-3. Spaceborne ASTER super-spectral (SS) resolution (14 bands) and Landsat-7 ETM+ multi-spectral (MS) resolution (7 bands).

The Landsat program is the longest-running enterprise for acquisition of satellite imagery of the Earth and it is likely to be considered the most successful EO data acquisition mission by a large majority of the RS community. The success of the Landsat 7-band Enhanced Thematic Mapper (ETM)+ imaging sensor series is due to several factors, including a combination of spectral bands specifically tailored to EO data applications. Let us identify the seven Landsat-like spectral channels as follows: ETM1 for visible blue [(B) - 0.45–0.52  $\mu\text{m}$ ], ETM2 for visible green [(G); 0.52–0.60  $\mu\text{m}$ ], ETM3 for visible red [(R); 0.63–0.69  $\mu\text{m}$ ], ETM4 for near IR [(NIR); 0.76–0.90  $\mu\text{m}$ ], ETM5 for medium IR1, MIR1 (1.55–1.75  $\mu\text{m}$ ), ETM7 for medium IR2, MIR2 (2.08–2.35  $\mu\text{m}$ ), and ETM6 for thermal IR [(TIR); 10.4–12.5  $\mu\text{m}$ ]. The Landsat ETM+ imaging sensor series' spectral resolution can be adopted as a reference baseline because almost any of the existing or





future planned satellite optical imaging sensors features a spectral resolution that overlaps with Landsat's, refer to Table 2-2 to Table 2-4, see Fig. 2-3 and Fig. 2-4. According to {90}, {91}, {92}, the following considerations hold.

- The loss of the B channel is expected to decrease the capability of detecting haze, smoke plumes, and water types {93}, {94}.
- The loss of the MIR1 channel is expected to decrease the accuracy and reliability of the separation of: (i) snow from cloud {75}, and (ii) vegetation and rangeland from bare soil types. Since its wavelengths are sensitive to water absorption, MIR1 is sensitive to both vegetation moisture content and soil moisture {95}, {96}. For example, in {96} the channel MIR1 is described as the best Landsat band overall. This is in contrast with the RS common knowledge where the NIR band is typically considered the single channel most useful for separating spectral signatures of vegetated LC classes from bare soil {95}.
- The loss of the MIR2 channel is expected to decrease the capability of discriminating bare soil types, particularly burned areas. For example, several burned-area indexes employ MIR2 in comparison with the NIR channel {97}.
- The loss of the TIR channel is expected to increase the spectral confusion between light-toned (highly reflective) soil types, particularly in mountainous (and cold) areas, with classes thick and thin clouds and either snow or ice spectral classes {75}.

Properties of lightweight MS and thermal imaging sensors suitable for small UAVs platforms are summarized in Table 2-5, while their spectral sensitivity profiles are shown in Fig. 2-5 to Fig. 2-7 {74}.

Landsat-4/-5 TM and Landsat-7 ETM+		LDCM/Landsat-8 OLI and TIRS		ASTER		MODIS		SPOT-4 HRVIR, SPOT-5 HRG, and SPOT-4/5 VMI		NASA-NOAA National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) - Visible Infrared Imaging Radiometer Suite (VIIRS)	
Band	Spectral region ( $\mu\text{m}$ )	Band	Spectral region ( $\mu\text{m}$ )	Band	Spectral region ( $\mu\text{m}$ )	Band	Spectral region ( $\mu\text{m}$ )	Band	Spectral region ( $\mu\text{m}$ )	Band	Spectral region ( $\mu\text{m}$ )
		1-coastal/aerosol	0.433–0.453								
1 (B)	0.45-0.52	2	0.450–0.515	-	-	(3+10) OR 3	[(0.459 – 0.479) + (0.483 – 0.493)] OR (0.459 – 0.479)	1 (VMI)	0.43-0.47	M3 (750 m)	0.478-0.498
2 (G)	0.52-0.60	3	0.525–0.600	1	0.52-0.60	(11+4) OR 4	[(0.526 – 0.536) + (0.545 – 0.565)] OR (0.545 – 0.565)	1 (HRVIR, HRG)	0.50-0.59	M4 (750 m)	0.545-0.565
3 (R)	0.63-0.69	4	0.630–0.680	2	0.63-0.69	(1+14) OR 1	[(0.620 – 0.670) + (0.673 – 0.683)] OR (0.620 – 0.670)	2	0.61-0.68	I1 (375 m)	0.600-0.680
4 (NIR)	0.76-0.90	5	0.845–0.885	3	0.76-0.86	(15+2+17) OR (15+1+17) OR 2	[(0.743 – 0.753) + (0.841 – 0.876) + (0.890 – 0.920)] OR	3	0.79-0.89	I2 (375 m)	0.846-0.885



							[(0.743 – 0.753) + (0.862 – 0.877) + (0.890 – 0.920)] OR (0.841 – 0.876)				
		9-cirrus	1.360–1.390								
5 (MIR1)	1.55-1.75	6	1.560–1.660	4	1.600-1.700	6	1.628 – 1.652	4	1.58-1.75	I3 (375 m) OR M10 (750 m)	I3: 1.580-1.640, M10: 1.580-1.640
7 (MIR2)	2.08-2.35	7	2.100–2.300	(5 + 6 + 7 + 8)	[(2.145-2.185) + (2.185-2.225) + (2.235-2.285) + (2.295-2.365)]	7	2.105 – 2.155	-	-	I4 (375 m) OR M11 (750 m)	I4: 3.550-3.93, M11: 2.225-2.275
6 (TIR)	10.4-12.5	10-T1R1 + 11-T1R2	[(10.620-11.200) + (11.500-12.500)]	(13+14)	[(10.25-10.95) + (10.95-11.65)]	(31+32) OR 31	[(10.780 – 11.280) + (11.770 – 12.270)]	-	-	I5 (375 m) OR M15 (750 m)	I5: 10.500-12.400, M15: 10.263-11.263

Table 2-3. Spectral resolutions of Landsat-4/-5 TM and Landsat-7 ETM+ compared with those of Landsat-8, SPOT, ASTER, and MODIS satellite sensors. Also refer to Fig. 2-3.

Landsat-4/-5 TM and Landsat-7 ETM+		ENVISAT AATSR, ERS-2 ATSR-2, SENTINEL-3 SLSTR (bands S1 to S9)		ENVISAT MERIS, SENTINEL-3 OLCI (bands O1 to O23)		SENTINEL-2 Multi-Spectral Instrument (MSI)		AVHRR		MSG	
Band	Spectral region (μm)	Band	Spectral region (μm)	Band	Spectral region (nm)	Band	Spectral region (nm)	Band	Spectral region (μm)	Band	Spectral region (μm)
1 (B)	0.45-0.52	-	-	(3+4), (O4+O5)	(490±10/2 + 510±10/2)	2 (10 m)	490±65/2	-	-	-	-
2 (G)	0.52-0.60	1, S1	0.545-0.565	5, O6	560±10/2	3 (10 m)	560±35/2	-	-	-	-
3 (R)	0.63-0.69	2, S2	0.649-0.669	(7+8), (O8+O9)	(665±10/2 + 681.25±7.5/2)	4 (10 m)	665±30/2	1	0.58-0.68	1	0.6
4 (NIR)	0.76-0.90	3, S3	0.855-0.875	(11+ 12+ 13+ 14+ 15), (O12+ O13 (new)+ O23(new)+ O14/15 + O16/17 + O18 + O19)	MERIS: (760.625±3.75/2 + 778.65±15/2 + 865±20/2 + 885±10/2 + 900±10/2), OLCI: (760.625±3.75/2 + 764.375±3.75/2 + 767.5±2.5/2 + 778.65±15/2 + 865±20/2 + 885±10/2 + 900±10/2),	8 (10 m) OR 7 (20 m) + 8a (20 m)	842±115/2 OR 783±20/2 + 865±20/2	2	0.725-1.10	2	0.8
		S4 - cirrus	1.375								
5 (MIR1)	1.55-1.75	4, S5	1.46-1.76	-	-	11 (20 m)	1610±90/2	3 [3(A)]	1.58-1.64	3	1.6
7 (MIR2)	2.08-2.35	S6	2.25	-	-	12 (20 m)	2190±180/2	-	-	-	-
6 (TIR)	10.4-12.5	6 OR (6 + 7), S8 OR (S8 + S9)	(10.35-11.35) OR [(10.35-	-	-	-	-	(5+6) [4+5]	[(10.30-11.30) +	8 OR 9	(9.80-11.80) OR (11.00-13.00)



(central wavelength 11.45)		11.35) + (11.50-12.50)]						(11.50-12.50)]		(central wavelength 10.8 OR 12.0)
----------------------------	--	-------------------------	--	--	--	--	--	----------------	--	-----------------------------------

Table 2-4. Spectral resolutions of Landsat-4/-5 TM and Landsat-7 ETM+ compared with those of the ENVISAT AATSR, ERS-2 ATSR-2, SENTINEL-3 SLSTR, ENVISAT MERIS, SENTINEL-3 OLCI, SENTINEL-2 MSI, NOAA AVHRR and Meteosat 2nd Generation (MSG) satellite sensors. Also refer to Fig. 2-3 and Fig. 2-4.

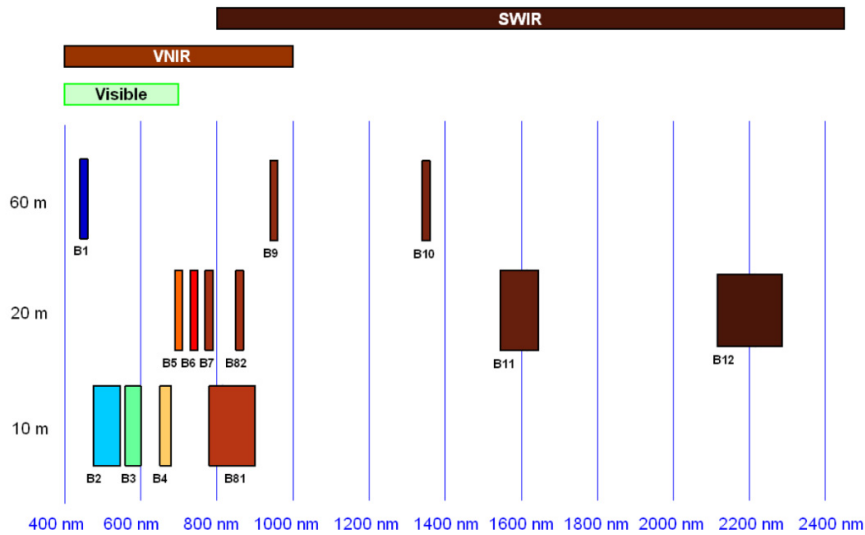


Fig. 2-4. Spaceborne Sentinel-2 MSI super-spectral (SS) resolution (13 bands).

Properties of lightweight MS and thermal imaging sensors suitable for small UAVs platforms	Canon S110 NIR (derived from the Canon S110 RGB by means of modified Bayer color filters)	multiSPEC 4C Prototype	multiSPEC 4C Commercial	ThermoMAP
Pixels per sensor	12MP (Bayer pattern)	4 sensors, 0.4MP each	4 sensors, 1.2MP each	0.3MP (640 x 512)
Sensor size [mm]	7.44 x 5.58	4.51 x 2.88 (per sensor)	4.8 x 3.6 (per sensor)	10.88 x 8.70
Pixel size [µm]	1.33	3.75		17.0
Ground sampling distance (GSD) at 100m AGL [cm]	3.5	20	10	18.5
Spectral channels (central-frequency / opt. Band width) [nm]	G (550) R (625) NIR (850)	G (550 ± 20) R (660 ± 20) Red Edge (RE) (735 ± 5) NIR (790 ± 20)		7,000-16,000
Approx. price [EUR]	900	Prototype	8,000	10,000

Table 2-5. Properties of lightweight MS and thermal imaging sensors suitable for small UAVs platforms {74}. The Canon S110 RGB is a low-cost consumer-level colour camera, the Canon S110 NIR and S110 RE are low-cost MS cameras while the multiSPEC 4C is a high-end lightweight MS imaging sensor. The Canon MS cameras are equipped with modified Bayer colour filters: instead of recording blue, green and red, the green (G), red (R) and near infrared (NIR) bands are captured. Just one lens is needed resulting in precisely co-registered spectral channels with overlapping spectral sensitivities (Fig. 2-5). To be compared with the spectral sensitivity profile of the low-cost false-color 3-band VisNIR (Visible to Near-InfraRed) camera shown in Fig. 2-6. In contrast, the multiSPEC 4C has four lenses and four monochromatic CCD sensors; the colour separation takes place at the optical units via band-pass interference filters with well-defined central frequencies and bandwidths (Fig. 2-7). A zenith-looking panchromatic sensor enables the images to be normalised.

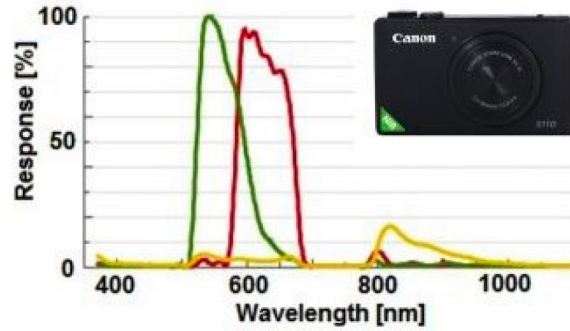


Fig. 2-5. Spectral sensitivity profile of the low-cost Canon S110 NIR camera suitable for small UAVs platforms. This is a lightweight 3-band G-R-NIR camera, with spectral channels G (550), R (625) and NIR (850), derived from a consumer-level Canon S110 RGB camera by means of modified Bayer color filters {74}. To be compared with the spectral sensitivity profile of the low-cost false-color 3-band VisNIR (Visible to Near-InfraRed) camera shown in Fig. 2-6.

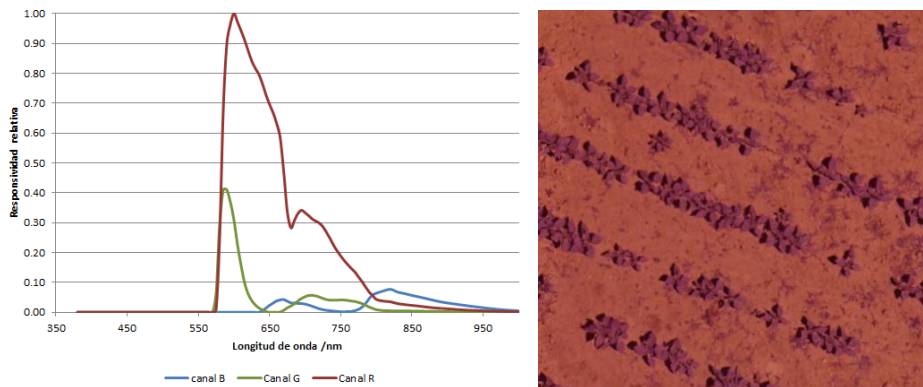


Fig. 2-6. Left: Spectral sensitivity profile of the low-cost false-color 3-band VisNIR (Visible to Near-InfraRed) camera suitable for small UAVs platforms, developed by the IMAPING Group - Remote Sensing for Precision Agriculture Crop Protection Department, Institute for Sustainable Agriculture, CSIC, Spanish National Research Council, Cordoba (Spain). Spectral channels RGB have spectral sensitivities ranging from band Y (yellow) (0.570-0.600  $\mu\text{m}$ ) as channel G (green), bands Y to R (Y: 0.57-0.60, Orange O: 0.60-0.65, Red R: 0.65-0.72.) as channel R (Red) and band NIR (0.820) as channel B (blue). Right: Example of an uncalibrated false-color VisNIR image of a field of sunflowers located in Spain. Acquired from a small UAV platform for precision agriculture (PA) applications, e.g., weed detection, RGB channels as described above, 3 cm spatial resolution, no histogram stretching is applied for visualization purposes.

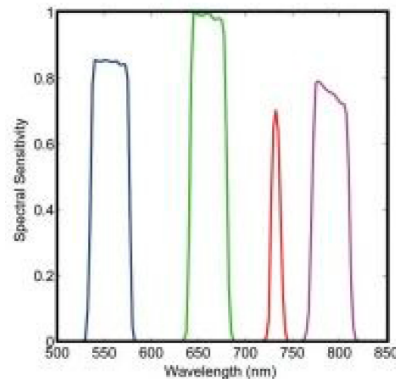


Fig. 2-7. Spectral sensitivity profile of the high-end 4-band multiSPEC 4C camera suitable for small UAVs platforms {74}. Spectral channels are G ( $550 \pm 20$ ), R ( $660 \pm 20$ ), Red Edge (RE) ( $735 \pm 5$ ) and NIR ( $790 \pm 20$ ). To be compared with the spectral sensitivity profiles of the low-cost false-color uncalibrated RGB cameras shown in Fig. 2-5 and Fig. 2-6.



### 2.3 Quality Assurance Framework for Earth Observation (QA4EO) Guidelines and Calibration/Validation Requirements

A visionary objective of the intergovernmental GEO is the development of a Global Earth Observation System of Systems (GEOSS), capable of transforming multi-source EO images into operational, timely and comprehensive EO value-added information products and services “to allow the access to the Right Information, in the Right Format, at the Right Time, to the Right People, to Make the Right Decisions”, in compliance with the GEOSS implementation plan for years 2005-2015 {98} and with the QA4EO *Cal/Val* requirements {25}, refer to Chapter 1.

According to the GEO QA4EO guidelines, the yet-unfulfilled development of a GEOSS requires EO sensory data and EO data-derived information products and services capable of satisfying two quality criteria.

- Accessibility. Related to the quantitative concept of *information-as-thing* {5}. In recent years the cost-free access to large-scale spaceborne, airborne and terrestrial EO image databases has become a reality.
- Suitability, submitted to the QA4EO *Cal/Val* requirements.
  - Radiometric *Cal* is the transformation of dimensionless digital numbers into a radiometric unit of measure to be community-agreed upon, e.g., top-of-atmosphere (TOA) radiance (TOARD) in range  $[0, \infty)$ , TOA reflectance (TOARF) in range  $[0, 1]$  or surface reflectance (SURF) in range  $[0, 1]$ , etc. {25}. EO data *Cal* is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” {86} (p. 29), {87}, {88}. In particular, EO data *Cal* is considered mandatory by the GEO QA4EO *Cal/Val* requirements {25}. In general, all sensory data we deal with on a daily basis are provided with a community-agreed physical unit of measure. Whereas physical variables can be investigated by physical, statistical or hybrid (combined deductive and inductive) data models, uncalibrated sensory data must be investigated by statistical models exclusively. Although statistical models do not require physical variables as input, they can benefit from data *Cal* in terms of augmented robustness to changes in the input data set acquired through time, space and sensors. Irrespective of these unquestionable facts, the QA4EO *Cal* requirements remain largely neglected by the RS community. For example, the large majority of papers published in the RS literature cope with uncalibrated EO data provided with no physical unit of radiometric measure. One consequence is that, to date, statistical model-based EO-IUSs dominate the RS literature as well as commercial EO image processing software toolboxes, consisting of overly complicated collections of inductive algorithms to choose from based on heuristics.
  - According to a GEO quality assurance (QA) and *Val* policy, each processing step in an EO information processing system must be provided with outcome and process (OP) Quantitative Quality Indicators ( $Q^2I$ ), refer to Chapter 1. To have a statistical meaning, each OP- $Q^2I$  must be provided with a degree of uncertainty in measurement  $\pm\delta$ , with  $\delta \in [0\%, 100\%]$ . By definition a GEO Stage 3 *Val* requires that “spatial and temporal consistency of the product with similar products are evaluated by independent means over multiple locations and time periods representing global conditions. In Stage 4 *Val*, results for Stage 3 are systematically updated when new product versions are released and as the time-series expands” {99}.

It is important and not at all obvious to stress that, to date, neither the GEOSS development has been accomplished by the RS community nor the GEO QA4EO *Cal/Val* requirements are enforced in the RS common practice. The following consideration holds.

- A minimally dependent and maximally informative (mDMI) set of EO OP- $Q^2I$ s was proposed in Chapter 1. According to the Pareto formal analysis of multi-objective optimization problems, optimization of an mDMI set of OP- $Q^2I$ s is an inherently-ill posed problem, where many Pareto optimal solutions lying on the Pareto efficient frontier can be considered equally good {100}. EO-IUSs presented in the RS literature are typically assessed and compared based on the sole mapping accuracy, which means their mDMI set of OP- $Q^2I$ s remains largely unknown to date. For example, when a large-scale EO data-derived thematic map product was generated by a supervised data learning EO-IUS at “high” accuracy, the most limiting factors turned out to be the cost, timeliness, quality and availability of adequate supervised (labeled) data samples, collected from field sites, existing maps or geospatial data archives in tabular form {40}.



- In the RS common practice, typical OP-Q<sup>2</sup>Is are thematic mapping accuracy, e.g., overall accuracy (OA)  $\in [0\%, 100\%]$ , and computation time. In general,  $OA \pm \delta$  values are not published. It means that most published EO data classification systems feature an OA value provided with no statistical meaning.
- Irrespective of the fact that EO data *Cal* is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” {86} (p. 29), {87}, {88}, the QA4EO *Cal* requirements remain largely overlooked by the RS community. One consequence is that the RS literature is dominated by statistical model-based (inductive) EO-IUS solutions, because physical model-based (deductive) EO data analysis cannot be applied to EO data provided with no physical unit of radiometric measure.

According to the original experience of the present author, radiometric calibration of digital numbers (DNs) into TOARF or SURF values has a triple advantage {88}.

- (1) It is mandatory to guarantee inter-image harmonization and inter-sensor operability by providing dimensionless DN<sub>s</sub> with a community-agreed physical unit of measure, in agreement with the QA4EO guidelines.
- (2) It is beneficial because DN<sub>s</sub> provided with a physical unit of measure can be input to physical, statistical as well as hybrid (combined deductive and inductive) data models. On the contrary, DN<sub>s</sub> provided with no physical unit of measure can be input to statistical data models exclusively.
- (3) It is beneficial because images radiometrically calibrated into TOARF or SURF values in range [0, 1] and coded as a 4-byte data float can be compressed by a factor of 4 into a 1-byte char in range {0, 255} with a negligible quantization error = 0.2%. If a 4-byte data float in range [0, 1] is quantized into a 1-byte integer in range {0, 255}, the float-to-byte quantization error is

$$(\text{input value max} - \text{input value min}) / \text{number of quantization levels} / 2 \text{ (because of the rounding error to the closest integer, either above or below)} = (1 - 0) / 256 / 2 = 0.00195 = 0.2\% \quad (2-1)$$

Eq. (2-1) shows that radiometrically calibrated TOARF or SURF values  $\in [0, 1]$  can be byte-coded (with data type: char), i.e., their memory occupation can be reduced to a minimum size, with a quantization error  $\leq 0.2\%$ , typically considered negligible. For example, when metadata files of radiometric parameters are available to transform TOARD  $\geq 0$  values into TOARF values  $\in [0, 1]$ , approximation of the sun-Earth distance to 1 irrespective of the image acquisition date typically causes rounding errors of about 5% {91}, {92}, {88}.

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	BARE_SOILS
6	WATER
7	CLOUD_LOW_PROBABILITY
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

Fig. 2-8. General-purpose, user- and application-independent ESA Level 2 SCM’s legend adopted by the Sentinel-2 software Toolbox, known as Sentinel 2 (atmospheric) Correction (SEN2COR) Prototype Processor {30}, {34}, {35}, to be run on user side. Although it supports the SEN2COR software development and distribution to end users, the ESA data provider does not employ SEN2COR at the ground segment to avoid accountability for Level 2 outcome.

#### 2.4 ESA EO Level 2 Information Product

An EO Level 2 product is defined by ESA as a multi-spectral (MS) image corrected for geometric, atmospheric, adjacency and topographic effects {31}, {32}, {33}, stacked with its data-derived scene classification map (SCM) whose general-purpose, user- and application-independent legend includes quality layers, such as cloud and cloud-shadow {30},



{34}, {35}. A National Aeronautics and Space Administration (NASA) EO Level 2 product, defined as “a data-derived geophysical variable at the same resolution and location as Level 1 source data” {115}, is a special case of ESA EO Level 2 product, see Fig. 1-2. No ESA EO Level 2 product has ever been systematically generated at the ground segment {30}, refer to Chapter 1. For example, developed and distributed by ESA, the Sentinel 2 (atmospheric) Correction (SEN2COR) Prototype Processor must be run on user side {30}, {34}, {35}. The SEN2COR’s Level 2 SCM legend is shown in Fig. 2-8. The SEN2COR work flow {34}, {35} is the same as Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software product’s {32}. The SEN2COR cloud/cloud-shadow detector is a spatial context-sensitive hybrid (combined physical model-based and statistical model-based) inference system {34}, {35}. All other target classes in the SEN2COR’s Level 2 SCM legend employ a pixel-based static decision tree for 1D spectral pattern analysis where spatial topological and non-topological information components in the image-domain are totally ignored.

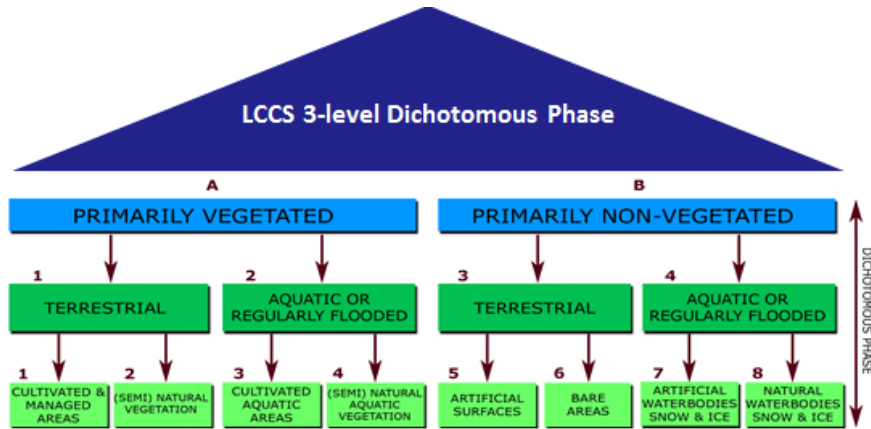


Fig. 2-9. FAO’s general-purpose, user- and application-independent 3-level 8-class LCCS-DP taxonomy followed by a user- and application-specific LCCS–MHP taxonomy {41}.

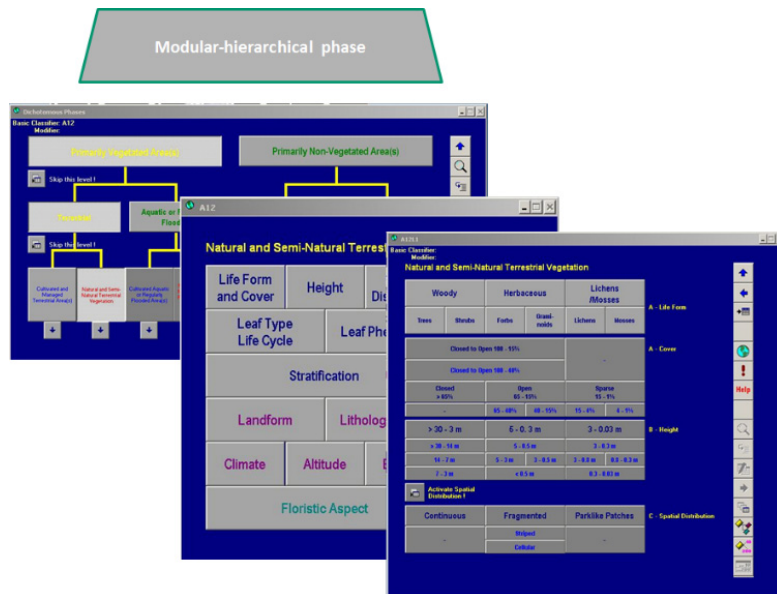


Fig. 2-10. Example of a user- and application-specific LCCS–MHP taxonomy instantiation, supported by the open source FAO LCCS software toolbox {41}.

## 2.5 FAO Land Cover Classification System (LCCS)

The “fully nested” FAO LCCS taxonomy consists of two phases {41}, see Fig. 1-4.



- (1) An initial general-purpose, user- and application-independent 3-level 8-class LCCS Dichotomous Phase (DP), see Fig. 2-9. Eight LC types are distinguished as a combination of three-dichotomous mapping criteria: (I) Vegetation versus non-vegetation, (II) Terrestrial versus aquatic, and (III) Managed versus Natural or semi-natural. The eight dichotomous LC types are: A11, Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas. A12, Natural and Semi-Natural Terrestrial Vegetation. A23, Cultivated Aquatic or Regularly Flooded Vegetated Areas. B35, Artificial Surfaces and Associated Areas. B36, Bare Areas. B47, Artificial Waterbodies, Snow and Ice. B48, Natural Waterbodies, Snow and Ice.
- (2) A so-called second-stage LCCS Modular Hierarchical Phase (MHP), consisting of a hierarchical battery of user- and application-specific one-class LC classifiers equivalent to one-class LC grammars {49}, see Fig. 1-4 and Fig. 2-10.

Level 1	Level 2	Level 3	
1. Artificial surfaces	1.1. Urban fabric	1.1.1. Continuous urban fabric	
		1.1.2. Discontinuous urban fabric	
		1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units
			1.2.2. Road and rail networks and associated land
	1.2.3. Port areas		
	1.2.4. Airports		
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites	
		1.3.2. Dump sites	
		1.3.3. Construction sites	
	1.4. Artificial non-agricultural vegetated areas	1.4.1. Green urban areas	
		1.4.2. Sport and leisure facilities	
	2. Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land
			2.1.2. Permanently irrigated land
			2.1.3. Rice fields
2.2. Permanent crops		2.2.1. Vineyards	
		2.2.2. Fruit trees and berry plantations	
		2.2.3. Olive groves	
2.3. Pastures		2.3.1. Pastures	
2.4. Heterogeneous agricultural areas		2.4.1. Annual crops associated with permanent crops	
		2.4.2. Complex cultivation	
		2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation	
		2.4.4. Agro-forestry areas	
3. Forests and semi-natural areas		3.1. Forests	3.1.1. Broad-leaved forest
	3.1.2. Coniferous forest		
	3.1.3. Mixed forest		
	3.2. Shrub and/or herbaceous vegetation association	3.2.1. Natural grassland	
		3.2.2. Moors and heathland	
		3.2.3. Sclerophyllous vegetation	
		3.2.4. Transitional woodland shrub	
	3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes, and sand plains	
		3.3.2. Bare rock	
3.3.3. Sparsely vegetated areas			
3.3.4. Burnt areas			
3.3.5. Glaciers and perpetual snow			
4. Wetlands	4.1. inland wetlands	4.1.1. Inland marshes	
		4.1.2. Peatbogs	
	4.2. Coastal wetlands	4.2.1. Salt marshes	
		4.2.2. Salines	
		4.2.3. Intertidal flats	

Table 2-6. CORINE Land Cover (CLC) taxonomy {102}, whose Level 1 is four-class.

In recent years the FAO two-phase LCCS-DP and LCCS-MHP taxonomies have gained increasing popularity {101}. One reason of popularity is that, unlike alternative hierarchical LC class taxonomies whose Level 1 is multi-class, such as the CORINE Land Cover {102}, see Table 2-6, the two-phase LCCS-DP and LCCS-MHP taxonomies are “nested” end-to-end, starting from the first-level LCCS-DP layer vegetation/non-vegetation, whose quality assurance (QA) becomes paramount for all subsequent LCCS layers according to an intuitive garbage in garbage out quality principle. This multi-level dependence is neither trivial nor obvious to underline. For example, vegetation/non-vegetation discrimination is acknowledged to be very challenging when pursued in EO image composites at continental or global scale by means of traditional supervised data learning EO-IUSs {40}, which are inherently semi-automatic and training data-specific {95}.

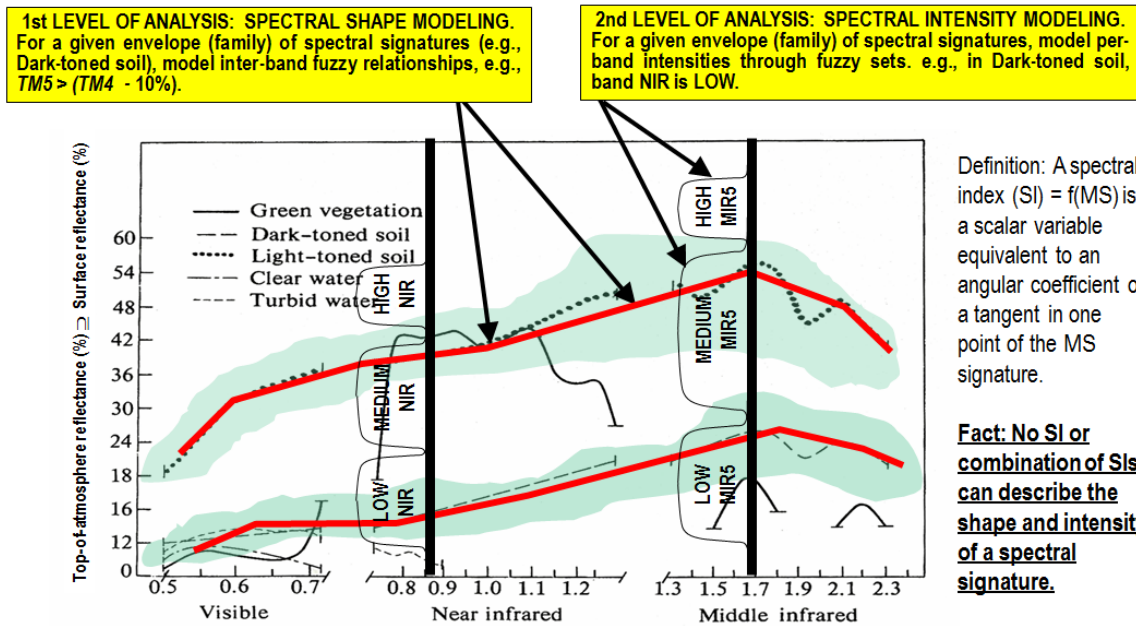


Fig. 2-11. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values which include surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF values forms a buffer zone (support area, envelope). Each target family of spectral signatures should be modeled in terms of multivariate shape and multivariate intensity as a viable alternative to multivariate analysis of spectral indexes. A typical spectral index is a scalar band ratio equivalent to an angular coefficient of a tangent in one point of the spectral signature. Infinite functions can feature the same tangent value in one point. In practice, no spectral index or combination of spectral indexes can reconstruct the multivariate shape and multivariate intensity of a spectral signature.

## 2.6 MS Pattern Recognition in 1D Image Analysis: Typical Misuse of Spectral Indexes with Special Emphasis on Vegetation Monitoring

According to a garbage in garbage out information principle, the FAO 3-level 8-class LCCS-DP taxonomy {41}, see Fig. 1-4, highlights the fact that the quality assurance of the 1-level 2-class dichotomous vegetation/nonvegetation classification problem is of paramount importance for all subsequent LCCS layers, see Fig. 2-9 and Fig. 2-10. In contrast with the RS common knowledge, EO data-derived vegetation/non-vegetation discrimination cannot be considered trivial to accomplish in operating mode, which means automatically (without human-machine interaction) and at large-spatial scale, where robustness to changes in input EO images acquired across time and space must be guaranteed, refer to Chapter 1. For example, vegetation/non-vegetation discrimination is acknowledged to be very challenging when pursued in EO image composites at continental or global scale by means of traditional supervised data learning EO-IUSs {40}, which are inherently semi-automatic and training data-specific {95}, refer to Chapter 2.5.

Traditional EO data classifiers based on a 1D image analysis approach, capable of secondary MS pattern discrimination, including vegetation/non-vegetation MS signature recognition, while primary spatial topological information in the image-domain is ignored, tend to score “low” in OP-Q<sup>2</sup>Is such as degree of automation, scalability to changes in sensor specifications and robustness to changes in input EO data acquired across time and space {26}, {27} (refer to Chapter 1). According to the present author MS pattern recognition algorithms typically score low in operating mode because of the large “misuse” of spectral indexes affecting the RS community to date. This thesis is supported by the following considerations on spectral information representation.

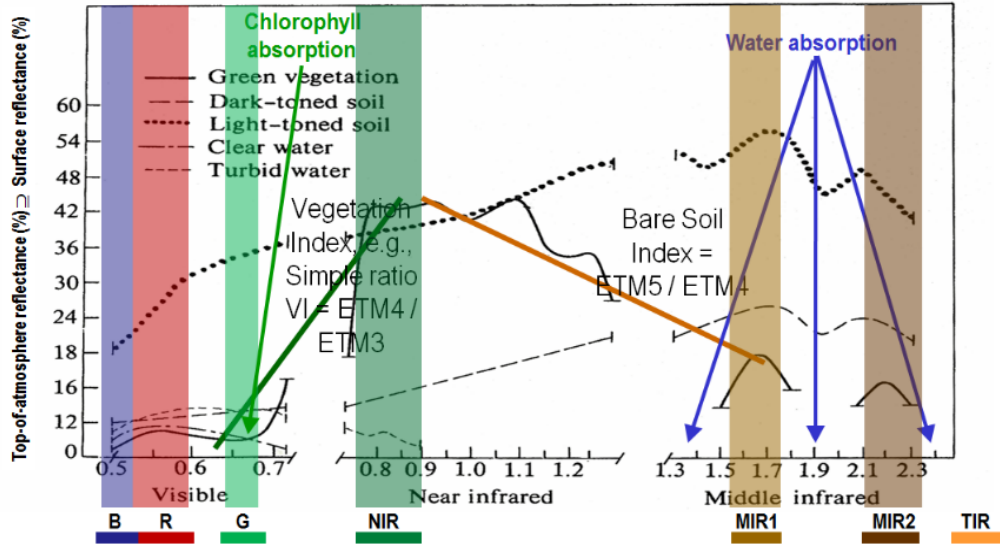


Fig. 2-12. An intuitive strategy to remove soil effects from a spectral vegetation index (VI) defined as a vegetation ratio index,  $VRI = NIR/R$ , is to replace VRI, equivalent to a 2-band first-order derivative, with a 3-band second-order derivative Greenness Index  $(R, NIR, MIR) \propto$  physical model-based parameter Leaf Area Index (LAI), where a second-order derivative is intuitively equal to the difference between two first-order derivatives and is capable of estimating local concavity, centered on the NIR waveband and required to be monotonically increasing with VRI {103}. It can be computed as follows:  $Greenness(R, NIR, MIR) = \text{local concavity centered on NIR} = \text{difference of first-order derivatives, monotonically increasing with } NIR/R \propto [(MIR - NIR) - (NIR - R)] \propto [MIR/NIR - NIR/R] \propto VRI - \text{Bare soil index} \propto NIR/R + NIR/MIR$ , with Bare soil index =  $MIR/NIR$ , where bare soil detection is the dual problem of vegetation detection {91}.

In a MS reflectance space, any target family of LC class-specific spectral signatures is a multivariate data distribution (envelope, hyperpolyhedron, manifold) accounting for within-class spectral variance. Like a vector quantity has two characteristics, a magnitude and a direction, any MS manifold is characterized by a multivariate shape and a multivariate intensity, see Fig. 2-11. Hence, spectral information redundancy, required to gain robustness to changes in spectral resolution specifications, can regard the modelling of both the MS shape and MS intensity information components of a target MS manifold. It is an unquestionable fact that neither a scalar spectral index nor a multivariate spectral index can model both the multivariate shape and the multivariate intensity information components of a target family of LC class-specific spectral signatures. Any scalar spectral index, either a normalized difference, e.g., the popular normalized difference vegetation index,  $NDVI = (NIR - Red) / (NIR + Red)$ , with  $NDVI \in [-1, 1]$ , where NIR = Landsat band 4 and Red = Landsat band 3 (see Table 2-3) or an unbounded band ratio, e.g., a vegetation ratio index ( $VRI = NIR/Red \geq 0$ ), is conceptually equivalent to the slope of a tangent to the spectral signature in one point, see Fig. 2-12. This spectral slope is a MS shape descriptor independent of the MS intensity, i.e., infinite functions with different intensity values can feature the same tangent value in one point. Although appealing due to its conceptual and numerical simplicity {95}, any scalar spectral index is unable *per se* to represent either the multivariate shape information or the multivariate intensity information component of a spectral signature. For example, it is well known that any vegetation spectral index, such as NDVI or VRI, can score “high” in shadow areas or water areas {90}. In the RS common practice, scalar spectral indexes are ever-increasing in number and variety {91}, {92}, {95}. This is a case of vicious cycle. The endless search for yet-another spectral index, supposedly more informative in MS pattern recognition tasks, is affected by the true-fact that, in general, neither a scalar (univariate) spectral index nor a multivariate spectral index is representative of the multivariate shape and multivariate intensity information components of a target MS manifold (hyperpolyhedron), see Fig. 2-11. For example, as shown in Fig. 2-12, an intuitive strategy to remove soil effects from  $VRI = NIR/R$  is to replace VRI, equivalent to a first-order derivative capable of estimating the local gradient, with a second-order derivative capable of estimating the local concavity, centered on the NIR waveband as follows {91}, {92}, {95}, {103}.

$$Greenness(R, NIR, MIR) = VRI - \text{Bare soil index} = VRI - BSI = NIR/R + NIR/MIR. \quad (2-2)$$



Intuitively, LC bare soil detection is the dual problem of LC vegetation detection. A typical bare soil index is  $BSI = MIR/NIR \{91\}, \{92\}, \{95\}$ . Hence,  $(-BSI) \approx \text{inverse of } BSI = 1/BSI = NIR/MIR$ . The proposed Greenness(R, NIR, MIR) index is a function of three spectral bands centered on the NIR channel, specifically, it is a second-order derivative equal to the difference between two first-order derivatives,  $VRI(R, NIR)$  and  $BSI(NIR, MIR)$ , each of which is a well-known spectral index defined as a function of two spectral channels  $\{103\}$ . The conclusion is that to discriminate vegetation from soil effects based on MS pattern properties, a scalar 3-variable local concavity estimator, such as Greenness(R, NIR, MIR), is more informative than any scalar 2-variable local gradient estimator, such as NDVI, VRI or BSI.

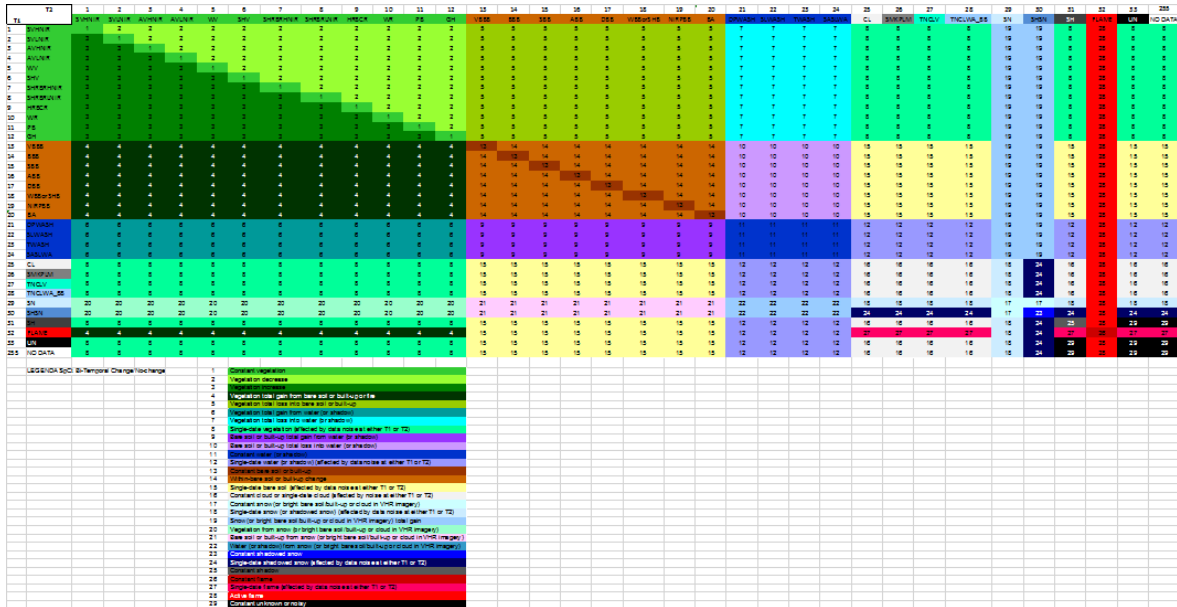


Fig. 2-13. Bi-temporal post-classification LC change/no-change detection. Implemented transition matrix between two Satellite Image Automatic Mapper (SIAM) maps generated at time T1 (rows) < T2 (columns) featuring the same so-called shared spectral legend, consisting of 33 spectral categories (MS color names). The output bi-temporal post-classification LC change/no-change SIAM map legend consists of 29 spectral categories (MS color names) through time  $\{26\}, \{27\}, \{33\}, \{91\}, \{92\}, \{105\}$ , see Fig. 2-14.

In practice a scalar spectral index reduces an  $N$ -channel MS image to a panchromatic image affected by an  $N$ -to-1 data dimensionality reduction  $\{104\}$ . No photointerpreter would typically consider a panchromatic image as informative as a MS image or simple enough to be submitted to a binary thresholding criterion for one-class LC class detection, e.g., vegetation versus non-vegetation discrimination.

An intuitive understanding of the inherent information loss stemming from MS signature compression into a scalar spectral index or a multivariate spectral index, see Fig. 2-12, is well grounded in statistics. Multivariate data statistics are known to be more informative than a sequence of univariate data statistics. For example, maximum likelihood data classification, accounting for multivariate data correlation and variance (covariance), is typically more accurate than parallelepiped data classification whose rectangular decision regions, equivalent to a concatenation of univariate data constraints, poorly fit multivariate data in the presence of bivariate cross-correlation  $\{76\}$ .

Fortunately, there are viable alternatives to the inefficient modelling of spectral signatures adopted by a large portion of the RS community, where the multivariate shape and multivariate intensity information components of a target LC class-specific hyperpolyhedron (manifold) in a MS reflectance space, see Fig. 2-11, are typically represented by either a scalar spectral index or a multivariate spectral index, see Fig. 2-12. One of these alternative solutions in operating mode is the static (non-adaptive-to-data) spectral decision tree for color naming, capable of prior knowledge-based MS reflectance space hyperpolyhedralization into a community-agreed dictionary of color names, proposed in pseudo-code in  $\{105\}$ . This static decision tree for MS reflectance space hyperpolyhedralization represents the multivariate shape and multivariate intensity information components of a target MS hyperpolyhedron, neither necessarily convex nor connected, as a converging combination of many independent  $j$ th-variable functions, with  $j \in \{1, \text{total number } N \text{ of spectral channels}\} \{26\}, \{27\}, \{91\}, \{92\}, \{105\}$ .





1	Constant vegetation
2	Vegetation decrease
3	Vegetation increase
4	Vegetation total gain from bare soil or built-up or fire
5	Vegetation total loss into bare soil or built-up
6	Vegetation total gain from water (or shadow)
7	Vegetation total loss into water (or shadow)
8	Single-date vegetation (affected by data noise at either T1 or T2)
9	Bare soil or built-up total gain from water (or shadow)
10	Bare soil or built-up total loss into water (or shadow)
11	Constant water (or shadow)
12	Single-date water (or shadow) (affected by data noise at either T1 or T2)
13	Constant bare soil or built-up
14	Within-bare soil or built-up change
15	Single-date bare soil (affected by data noise at either T1 or T2)
16	Constant cloud or single-date cloud (affected by noise at either T1 or T2)
17	Constant snow (or bright bare soil/built-up or cloud in VHR imagery)
18	Single-date snow (or shadowed snow) (affected by data noise at either T1 or T2)
19	Snow (or bright bare soil/built-up or cloud in VHR imagery) total gain
20	Vegetation from snow (or bright bare soil/built-up or cloud in VHR imagery)
21	Bare soil or built-up from snow (or bright bare soil/built-up or cloud in VHR imagery)
22	Water (or shadow) from snow (or bright bare soil/built-up or cloud in VHR imagery)
23	Constant shadowed snow
24	Single-date shadowed snow (affected by data noise at either T1 or T2)
25	Constant shadow
26	Constant flame
27	Single-date flame (affected by data noise at either T1 or T2)
28	Active flame
29	Constant unknown or noisy

Fig. 2-14. A Satellite Image Automatic Mapper (SIAM)-based bi-temporal post-classification LC change/no-change detection map legend consists of 29 spectral categories (MS color names) through time {26}, {27}, {33}, {91}, {92}, {105}. See Fig. 2-13.

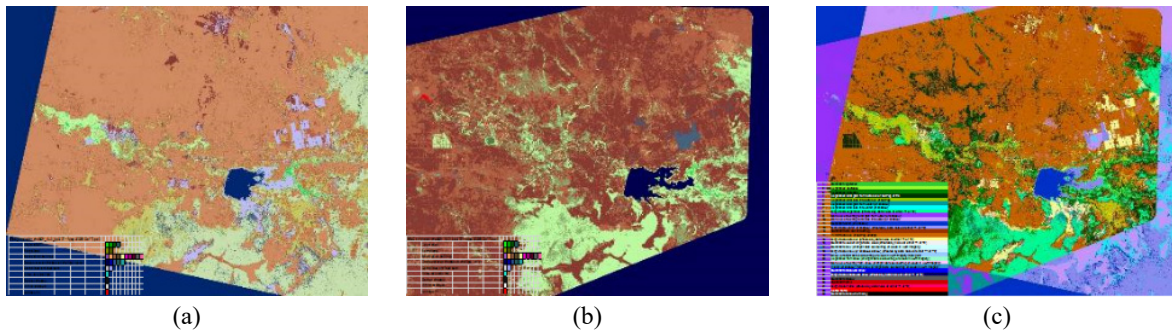


Fig. 2-15. Satellite Image Automatic Mapper (SIAM)-based bi-temporal inter-sensor post-classification land cover (LC) change/no-change detection. Example: Gambella, Ethiopia. Sensor 1 image (not shown): SPOT-5 (MIR, NIR, R, G), 10 m resolution. Sensor 2 (not shown): RapidEye (NIR, R, G, B), 5 m resolution. Both input images were radiometrically calibrated into TOARF values and mapped into static color names by the Satellite Image Automatic Mapper (SIAM), an expert system for MS reflectance space hyperpolyhedralization {26}, {27}, {33}, {91}, {92}, {105}. (a) Prior knowledge-based color map depicted in pseudo colors, generated from the SPOT-5 inter-image overlapping area, upscaled to 5 m. SPOT-like SIAM (S-SIAM) subsystem's map legend consisting of 33 "shared" spectral categories. (b) Prior knowledge-based color map depicted in pseudo colors, generated from the RapidEye inter-image overlapping area, at 5 m resolution. QuickBird-like SIAM (Q-SIAM) subsystem's map legend consisting of 33 "shared" spectral categories. (c) Bi-temporal inter-sensor post-classification LC change/no-change detection generated from the prior knowledge-based color map pair shown in Fig. (a) and Fig. (b). Post-classification map's legend consisting of 29 spectral categories, see Fig. 2-13 and Fig. 2-14.

The same consideration reported above for single-date EO image classification holds true for multi-temporal EO image classification. MS patterns of vegetation-through-time are typically discriminated by means of a multi-temporal scalar vegetation index recognition approach. In practice, each single-date MS image is compressed into a single-date scalar spectral index, equivalent to a panchromatic image; next, single-date scalar images are stacked through time to be analyzed.



According to the superposition principle, this multi-temporal image classification approach is affected by a spectral information loss occurring at each single-date of the time-series, due to the single-date MS-to-panchromatic information compression. In other words, multi-temporal scalar vegetation index analysis is affected by a spectral information loss monotonically increasing with the length of the temporal series. To avoid MS information loss in the analysis of MS image time-series, an existing alternative is an automated post-classification LC change/no-change detection in operating mode, whose inputs are a time-series of multi-level color maps automatically generated by, for example, a static decision tree for color naming applied in sequence to each single-date MS image in the time-series {26}, {27}, {33}, {91}, {92}, {105}, see Fig. 2-13 to Fig. 2-15. This approach requires no single-date MS data dimensionality reduction. In a post-classification approach, the final classification accuracy  $\in [0, 1]$  is not superior to, i.e., it is equal to or lower than the product of the single-date classification accuracies in the time-series {59}. For example,

$$\text{Bi-temporal post-classification LC change/no-change (LCC) detection overall accuracy (OA), } OA-LCC_{1,2} \in [0, 1], \\ OA-LCC_{1,2} = f(\text{OA of the LC maps at time T1 and time T2, identified as OA-LC1 and OA-LC2 respectively}) \leq OA-LC1 \times OA-LC2, \text{ where } OA-LC1 \in [0, 1] \text{ and } OA-LC2 \in [0, 1]. \quad (2-3)$$

For example, if  $OA-LC1 = 0.90$ ,  $OA-LC2 = 0.90$ , then  $OA-LCC_{1,2} = 0.81$ . Hence, post-classification analysis is recommended for its simplicity if and only if the single-date classification accuracies are “high” {26}, {27}, {33}, {91}, {92}, {105}. In other words, a necessary not sufficient pre-condition for multi-temporal image analysis to score “high” in accuracy according to a post-classification LC change/no-change detection approach is that single-date image classification scores “high” in accuracy.

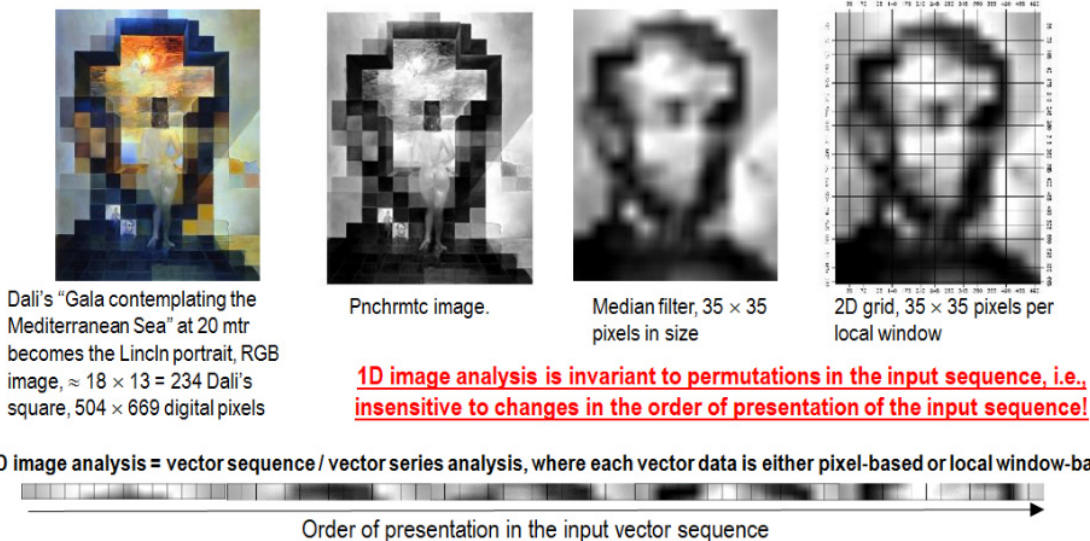


Fig. 2-16. Example of 1D image analysis. The (2D) image at left is transformed into the 1D vector data stream shown at bottom, where vector data are either pixel-based or spatial context-sensitive, e.g., local window-based. This 1D vector data stream means nothing to a human photointerpreter. When it is input to a traditional inductive data learning classifier, it is what the inductive classifier actually sees when watching the (2D) image at left. Undoubtedly, computers are more successful than humans in 1D image analysis. Nonetheless, humans are still far more successful than computers in (2D) image analysis.

## 2.7 1D and 2D Image Analysis Principles

In vision (refer to Chapter 1), spatial topological and spatial non-topological information components typically dominate color information {3}. This thesis is proved by the undisputable fact that achromatic (panchromatic) human vision, familiar to everybody when wearing sunglasses, is nearly as effective as chromatic vision in scene-from-image reconstruction and understanding. It means that the following necessary and sufficient condition holds for a CV system.

*A necessary and sufficient condition for a CV system to fully exploit spatial topological and spatial non-topological information components in addition to color is to perform nearly as well when input with either panchromatic or color imagery.*

Neglecting the fact that spatial topological and non-topological information components typically dominate color information in both the (2D) image-domain and the 4D spatiotemporal scene-domain involved with vision {3}, traditional EO-IUSs adopt a 1D image analysis approach, see Fig. 2-16 and Fig. 2-17. In 1D image analysis, a 1D streamline of vector data, either spatial context-sensitive (e.g., window-based or image object-based) or context-insensitive (pixel-based), is processed irrespective of the order of presentation of the input sequence. In practice 1D image analysis is invariant to permutations, such as in orderless encoders {106}. If vector data are spatial context-sensitive then 1D image analysis ignores spatial topological information. This consideration applies to the object-based image analysis (OBIA) paradigm {107} when it adopts a 1D image analysis approach, see Fig. 2-17. If vector data are pixel-based (context-insensitive) then 1D image analysis ignores both spatial topological and spatial non-topological information components. Noteworthy, prior knowledge-based color naming of a spatial unit  $x$  in the image-domain, where  $x$  is either (0D) point, (1D) line or (2D) polygon defined according to the OGC nomenclature {108}, is a special case of 1D image analysis, either pixel-based or image object-based where, respectively, spatial topological and non-topological information or spatial topological information is ignored, refer to Eq. (1-2) in Chapter 1.

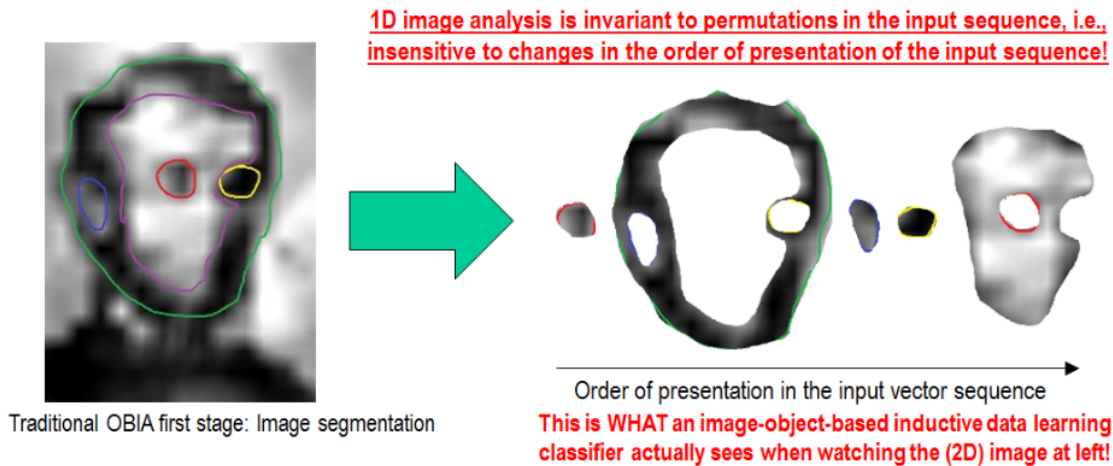


Fig. 2-17. 1D image analysis consistent with the context-sensitive object-based image analysis (OBIA) paradigm {107}. The (2D) image at left (corresponding to the median-filtered image,  $35 \times 35$  pixels in size, shown in Fig. 2-16) is transformed into the 1D vector data stream shown at left, where vector data are sub-symbolic image-objects generated by a traditional OBIA first stage for image segmentation. This 1D vector data stream means nothing to a human photointerpreter. When it is input to a traditional inductive data learning classifier, this 1D vector data stream is what the inductive classifier actually sees when watching the (2D) image at left. Undoubtedly, computers are more successful than humans when dealing with 1D image analysis consistent with the context-sensitive OBIA paradigm. Nonetheless, humans are still far more successful than computers in (2D) image analysis.

Alternative to 1D image analysis, 2D image analysis relies on a sparse (distributed) 2D array (2D regular grid) of local spatial filters {14}, synonym of retinotopic/topology-preserving feature mapping {109}, {110}, see Fig. 2-18. The human brain's organizing principle is topology-preserving feature mapping {111}. In the biological vision system, topology-preserving feature maps are primarily spatial, where activation domains of physically adjacent processing units in the 2D array of convolutional filters are spatially adjacent regions in the 2D visual field. Provided with a superior degree of biological plausibility in modelling 2D spatial topological and non-topological information, distributed processing systems capable of 2D image analysis, such as deep convolutional neural networks (DCNNs), outperform 1D image analysis approaches {106}. This apparently trivial consideration is at odd with a relevant portion of the RS literature, where pixel-based 1D image analysis is mainstream followed by context-sensitive 1D image analysis implemented within the OBIA paradigm {107}. As a consequence, the recent revamp of DCNN in CV applications {106} is progressively affecting the RS community {112}, {113}.

Since traditional EO-IUSs adopt a 1D image analysis approach where dominant spatial information is neglected in favour of secondary color information, it is useful to turn attention to the multidisciplinary framework of cognitive science to shed light on how humans deal with color information. According to cognitive science, see Fig. 1-1, which includes linguistics, the study of languages, humans discretize (fuzzify) ever-varying quantitative (numeric) photometric and spatiotemporal sensations into stable qualitative (categorical, nominal) percepts, eligible for use in symbolic human reasoning based on a

convergence-of-evidence approach {3}. In their seminal work, Berlin and Kay proved that 20 human languages, spoken across space and time in the real-world, partition quantitative color sensations collected in the visible portion of the electromagnetic spectrum (see Fig. 2-1) onto the same “universal” dictionary of eleven basic color (BC) names {45}: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown. In a 3D monitor-typical red-green-blue (RGB) cube, BC names are intuitive to think of and easy to visualize. They provide a mutually exclusive and totally exhaustive partition of the RGB cube into RGB polyhedra neither necessarily connected nor convex, see Fig. 1-8 {54}, {55}. Since they are community-agreed upon to be used by members of the community, RGB BC polyhedra are prior knowledge-based, i.e., stereotyped, non-adaptive-to-data (static), general-purpose, application- and data-independent. Multivariate measurement space hyperpolyhedralization is the transformation of a numeric variable into a categorical variable. This is a typical problem in many scientific disciplines, such as inductive vector quantization (VQ) in machine learning-from-data {12} and deductive numeric variable fuzzification into discrete fuzzy sets in fuzzy logic {56}.

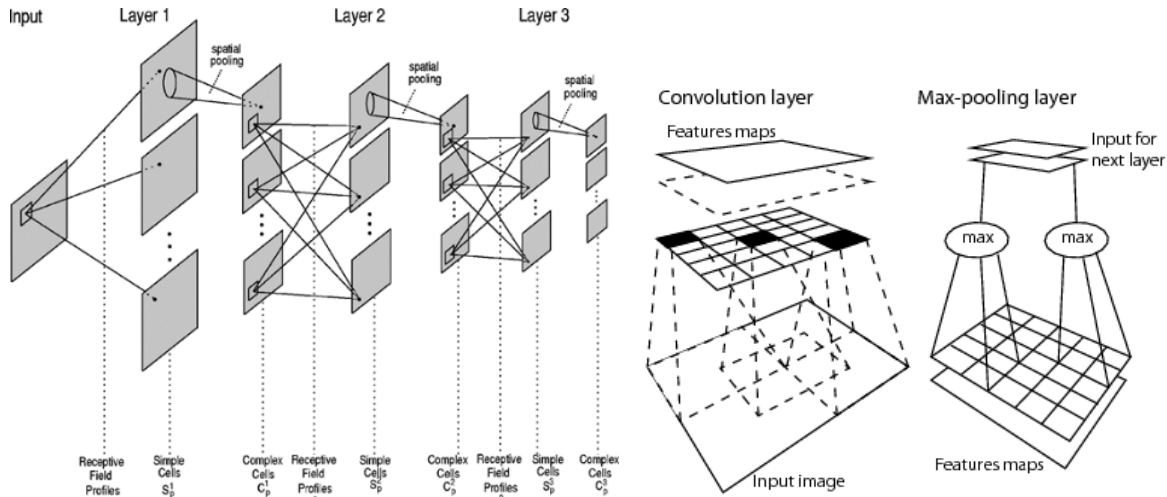


Fig. 2-18. 2D image analysis as synonym of retinotopic/topology-preserving feature mapping in a (2D) image-domain {109}, {110}. Activation domains of physically adjacent processing units in the 2D array of convolutional filters are spatially adjacent regions in the 2D visual field. Provided with a superior degree of biological plausibility in modelling 2D spatial topological and non-topological information, distributed processing systems capable of 2D image analysis, such as deep convolutional neural networks (DCNNs), typically outperform traditional 1D image analysis approaches. Will computers become as good as humans in 2D image analysis?

In Chapter 1, an analytic model of vision based on a convergence-of-evidence approach was proposed, see Eq. (1-2). In Eq. (1-2), a first-stage (preliminary) prior knowledge-based color naming {45}, {54}, {55} of a spatial unit  $x$ , either (0D) point, (1D) line or (2D) polygon {108}, is a special case of 1D image analysis, either pixel-based or image object-based, where spatial topological and non-topological information or spatial topological information are ignored respectively. According to Eq. (1-2), if condition  $m(\text{ColorValue}(x) | \text{ColorName}^*) = 1$  holds true according to a 1D static color naming first stage, where  $\text{ColorName}^* \in \{1, \text{DictionaryOfColorNames}\}$ , then spatial unit  $x$ -specific spatial information components  $\text{ShapeValue}(x)$ ,  $\text{TextureValue}(x)$ , and  $\text{SpatialRelationships}(x, \text{Neigh}(x))$  are to be further investigated in a second-stage 2D image analysis focused on (conditioned by, stratified by) the image subset (mask, stratum, candidate area) where condition  $\text{ColorName} = \text{ColorName}^*$  holds. Hence, Eq. (1-2) intuitively shows that, for a given discrete spatial unit  $x$  in the image-domain, either point, line or polygon {108}, a 1D static color naming first stage {45}, {54}, {55} allows the preliminary stratification of unconditional multivariate 2D spatial variables, such as  $\text{ShapeValue}(x)$ ,  $\text{TextureValue}(x)$ , and  $\text{SpatialRelationships}(x, \text{Neigh}(x))$ , into color class-conditional 2D data distributions, in agreement with the statistic stratification principle {63} and the divide-and-conquer problem solving approach {12}, {64}. To summarize, first-stage 1D image analysis by color naming can be employed to initialize with stratification masks a 2D image analysis second stage, in agreement with a divide-and-conquer problem solving approach and convergence of first-stage color evidence with second-stage spatial evidence.



## 2.8 References in Chapter 1 to Chapter 2

- {1} Prof. Dan Lopresti, "Advice for a successful PhD experience," Sept. 18, 2011, Computer Science and Engineering Dept., Lehigh University Bethlehem, PA, USA. Online available: [http://www.rfai.li.univ-tours.fr/Documents/Autres\\_PDF/Advice\\_for\\_a\\_Successful\\_PhD\\_Experience.pdf](http://www.rfai.li.univ-tours.fr/Documents/Autres_PDF/Advice_for_a_Successful_PhD_Experience.pdf)
- {2} Prof. Josep Lladós, "10 musings for a successful PhD Experience", Computer Vision Center (CVC) of the Universidad Autònoma de Barcelona (UAB), July 22, 2016.
- {3} T. Matsuyama and V. S. Hwang, SIGMA – A Knowledge-based Aerial Image Understanding System. New York, NY: Plenum Press, 1990.
- {4} D. Marr, Vision. New York, NY: Freeman and C., 1982.
- {5} R. Capurro and B. Hjørland, The concept of information, Annual Review of Information Science and Technology, B. Cronin, Ed., Medford, NJ, USA: Information Today Inc., 2003, vol. 37, ch. 8, pp. 343–411. [Online]. Available: <http://www.capurro.de/infoconcept.html>
- {6} C. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, 1948.
- {7} Cognitive Science. [Online]. Available: [https://en.wikipedia.org/wiki/Cognitive\\_science](https://en.wikipedia.org/wiki/Cognitive_science). Retrieved 2016-09-05.
- {8} D. Parisi, "La scienza cognitive tra intelligenza artificiale e vita artificiale," in Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza, Bologna, Italy: Patron Editore, 1991.
- {9} G. A. Miller, "The cognitive revolution: a historical perspective", in Trends in Cognitive Sciences, vol. 7, pp. 141-144, 2003.
- {10} F. J. Varela, E. Thompson, and E. Rosch, The Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA: MIT Press, 1991.
- {11} F. Capra and P. L. Luisi, The Systems View of Life: A Unifying Vision. Cambridge, UK: Cambridge University Press, 2014.
- {12} V. Cherkassky and F. Mulier. Learning from Data: Concepts, Theory, and Methods. New York, NY: Wiley, 1998.
- {13} J. Hadamard, "Sur les problemes aux derivees partielles et leur signification physique," Princeton University Bulletin, vol. 13, pp. 49–52, 1902.
- {14} J. K. Tsotsos, "Analyzing vision at the complexity level," Behavioral and Brain Sciences, vol. 13, pp. 423-469, 1990.
- {15} S. Frintrop, "Computational visual attention," in Computer Analysis of Human Behavior, Advances in Pattern Recognition, A. A. Salah and T. Gevers, Eds., Springer, 2011.
- {16} J. Piaget, Genetic Epistemology. New York: Columbia University Press, 1970.
- {17} D. Parisi, "La scienza cognitive tra intelligenza artificiale e vita artificiale," in Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza. Bologna, Italy: Patron Editore, 1991.
- {18} R. Serra and G. Zanarini, Complex Systems and Cognitive Processes, Berlin: Springer-Verlag, 1990.
- {19} Q. Iqbal and J. K. Aggarwal, "Image retrieval via isotropic and anisotropic mappings," in Proc. IAPR Workshop Pattern Recognit. Inf. Syst., Setubal, Portugal, Jul. 2001, pp. 34–49.
- {20} L. Pessoa, "Mach Bands: How Many Models are Possible? Recent Experimental Findings and Modeling Attempts", Vision Res., Vol. 36, No. 19, pp. 3205–3227, 1996.
- {21} C. Mason and E.R. Kandel, "Central Visual Pathways," in Principles of Neural Science; Kandel, E., Schwartz, J., Eds.; Norwalk, CT, USA: Appleton and Lange, pp. 420–439, 1991.
- {22} P. Gouras, "Color Vision," in Principles of Neural Science; Kandel, E., Schwartz, J., Eds.; Norwalk, CT, USA: Appleton and Lange, pp. 467–479, 1991.
- {23} E. R. Kandel, "Perception of Motion, Depth and Form," in Principles of Neural Science; Kandel, E., Schwartz, J., Eds.; Appleton and Lange: Norwalk, CT, USA; pp. 441–466, 1991.
- {24} S. Grove, "Knowledge-based interpretation of multisensor and multitemporal remote sensing images," Int. Archives of Photogrammetry and Remote Sensing, vol. 32, Part 7–4–3 W6, Valladolid, Spain, 3–4 June, 1999.
- {25} A Quality Assurance Framework for Earth Observation, version 4.0, Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS), 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- {26} A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOI) - Part 1: Introduction," Remote Sens., vol. 4, no. 9, pp. 2694-2735, 2012. doi:10.3390/rs4092694.





- {27} A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation," *Remote Sens.*, vol. 4, no. 9, pp. 2768-2817, 2012.
- {28} Big data, 2016. [Online]. Available: [en.wikipedia.org/wiki/Big\\_data](http://en.wikipedia.org/wiki/Big_data)
- {29} European Space Agency, S. D'Elia, Personal communication, 2002.
- {30} European Space Agency, Sentinel-2 User Handbook, Standard Document. Issue 1 Rev 2, 2015.
- {31} M. P. Bishop and J. D. Colby, "Anisotropic reflectance correction of SPOT-3 HRV imagery," *Int. J. Remote Sens.*, vol. 23, no. 10, pp. 2125–2131, May 2002.
- {32} R. Richter and D. Schlöpfer, Atmospheric / Topographic correction for airborne imagery – ATCOR-4 User Guide, Version 6.3.01, February 2014. Retrieved February 1, 2014, from [http://www.rese.ch/pdf/atcor4\\_manual.pdf](http://www.rese.ch/pdf/atcor4_manual.pdf)
- {33} A. Baraldi, D. Simonetti, and M. Girona, "Operational two-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, pp. 112-146, 2010.
- {34} Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) and Telespazio VEGA Technologies, "Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document." Document S2PAD-ATBD-0001, European Space Agency, 2011.
- {35} Telespazio VEGA Technologies, "Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual." Document S2PAD-VEGA-SUM-0001, European Space Agency, 2016.
- {36} A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1349-1380, 2000.
- {37} C. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, "GeoIRIS: Geospatial Information Retrieval and Indexing System—Content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, 2007.
- {38} S. S. Dhurba and R. L. King, "Semantics-enabled framework for knowledge discovery from Earth observation data archives," *IEEE Trans. Geosci. Remote Sens.*, col. 43, no. 11, pp. 2563-2572, 2005.
- {39} C. Dumitru, S. Cui, G. Schwarz, and M. Datcu, M. "Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies," *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.*, vol. 8, no. 4, pp. 1635 – 1650, 2015.
- {40} Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D., Turner, B.L., Cochrane, M.A., Eds.; Land Change Science; Kluwer: Dordrecht, The Netherlands, 2004.
- {41} A. Di Gregorio and L. Jansen, Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO: Rome, Italy, FAO Corporate Document Repository, 2000. [Online]. Available: <http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm>
- {42} A. Baraldi, D. Tiede, M. Sudmanns, M. Belgiu, and S. Lang, "Automated near real-time Earth observation Level 2 product generation for semantic querying," GEOBIA 2016, 14-16 Sept., University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands.
- {43} A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
- {44} T. Gevers, A. Gijzenij, J. van de Weijer, and J-M. Geusebroek, *Color in Computer Vision*, Hoboken, New Jersey, USA: Wiley, 2012.
- {45} B. Berlin and P. Kay, *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.
- {46} A. Baraldi and F. Parmiggiani, "Combined detection of intensity and chromatic contours in color images," *Optical Engineering*, vol. 35, no. 5, pp. 1413-1439, May 1996.
- {47} J. Soares and A. Baraldi, "Operational Estimation of a Comprehensive Set of Complementary Shape, Size, and Photometric Attributes of Image-Objects," submitted for consideration for publication, *IEEE Trans. Image Proc.*, TIP-12445-2014, 2014.
- {48} Wenwen Li, M. F. Goodchild and R. L. Church, "An efficient measure of compactness for 2D shapes and its application in regionalization problems," *Int. J. of Geographical Info. Science*, pp. 1-24, 2013.
- {49} M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall, 1994.
- {50} J. I. Yellott, "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," *Optical Society of America*, vol. 10, no. 5, pp. 777-793, May 1993.



- {51} J. Victor, "Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: Comment," *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1680-1684, May 1994.
- {52} B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception*, vol. 2, pp. 391-405, 1973.
- {53} B. Julesz, "Texton gradients: The texton theory revisited," in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, vol. 54, no. 4-5, Aug. 1986.
- {54} L. D. Griffin, "Optimality of the basic color categories for classification," *J. R. Soc. Interface*, vol. 3, pp. 71–85, 2006.
- {55} R. Benavente, M. Vanrell, and R. Baldrich, "Parametric fuzzy sets for automatic color naming," *J. Optical Society of America A*, vol. 25, pp. 2582-2593, 2008.
- {56} L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–353, 1965.
- {57} K. Kuzera and R. G. Pontius Jr., "Importance of matrix construction for multiple-resolution categorical map comparison," *GIScience and Remote Sens.*, vol. 45, pp. 249–274, 2008.
- {58} R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, Boca Raton, FL, USA: Lewis Publishers, 1999.
- {59} R. Lunetta and D. Elvidge, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, London, UK: Taylor & Francis, 1999.
- {60} S. V. Stehman and R. L. Czaplewski, "Design and analysis for thematic map accuracy assessment: Fundamental principles," *Remote Sens. Environ.*, vol. 64, pp. 331-344, 1998.
- {61} M. Beauchemin and K. Thomson, "The evaluation of segmentation results and the overlapping area matrix," *Int. J. Remote Sens.*, vol. 18, pp. 3895–3899, 1997.
- {62} A. Ortiz and G. Oliver, "On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures," *Pattern Recognition Letters*, vol. 27, pp. 1916-1926, 2006.
- {63} N. Hunt and S. Tyrrell, *Stratified Sampling*. Coventry University, 2012. [Online] Available: <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>
- {64} C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- {65} M. Weinmann, *Reconstruction and Analysis of 3D Scenes*, Switzerland: Springer, 2016.
- {66} L. Matikainen, J. Hyypä, and P. Litkey, "Multiplespectral airborne laser scanning for automated map updating," *Int. Archives of the Photogrammetry, Remote Sens. and Spatial Information Sciences*, vol. XLI-B3, 2016, XXIII ISPRS Congress, Prague, Czech Republic, 12–19 July 2016.
- {67} Global LiDAR Market to Surpass US\$605 Mn by 2020 Fueled by Increased Demand From Coastal Applications: Transparency Market Research, GIScafe', 2016. [Online]. Available: [www10.giscafe.com/nbc/articles/view\\_article.php?section=CorpNews&articleid=1397792](http://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1397792)
- {68} A. C. Watts, V. G. Ambrosia and E. A. Hinkle, "Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use," *Remote Sens.*, vol. 4, pp. 1671-1692, 2012.
- {69} Special Coverage: UAS: Disruption in the Skies, GIScafe', 2014. [Online]. Available: <http://www10.giscafe.com/blogs/gissusan/2014/10/23/special-coverage-uas-disruption-in-the-skies/>
- {70} P. K. Freeman and R. S. Freeland, "Agricultural UAVsintheU.S.: potential, policy, and hype," *Remote Sensing Applications: Society and Environment*, vol. 2, pp. 35–43, 2015.
- {71} S. Nebiker, A. Annen, M. Scherrer, D. Oesch, "A Light-weight Multiplespectral Sensor for Micro UAV – Opportunities for Very High Resolution Airborne Remote Sensing," *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Beijing, China: ISPRS, pp. 1193–1200, 2008.
- {72} S. Nebiker, N. Lack., M. Abächerli, S. Läderach, "Multiplespectral and Multitemporal Imagery from UAV for Predicting Grain Yield and Detecting Plant Diseases," to appear in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XIII ISPRS Congress, Prague, Czech Republic, 2016.
- {73} 5 Drones for Precision Agriculture, GIScafe'. [Online] Available: [http://www10.giscafe.com/nbc/articles/view\\_article.php?section=CorpNews&articleid=1478149](http://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1478149). Accessed on 14 Jan. 2017.
- {74} S. Nebiker and N. Lack, "Capabilities for precision farming and heat mapping - Multiplespectral and Thermal Sensors on UAVs," *GIM International*, 02/08/2016. [Online] Available: <https://www.gim-international.com/content/article/multiplespectral-and-thermal-sensors-on-uavs>. Accessed on 14 Jan. 2017.
- {75} P. H. Swain and S.M. Davis, *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill, 1978.
- {76} T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation*, New York: John Wiley and Sons, 1979.





- {77} Research and Markets: Aerial Imaging Market - 2014 Global Industry Analysis, GIS Cafe', 2014. [Online]. Available: [http://www10.giscafe.com/nbc/articles/view\\_article.php?section=CorpNews&articleid=1317266](http://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1317266)
- {78} S. S. Durbha and Roger L. King, "Semantics-enabled framework for knowledge discovery from Earth observation data archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2563-2572, Nov. 2005.
- {79} National Aeronautics and Space Administration (NASA), "Getting Petabytes to People: How the EOSDIS Facilitates Earth Observing Data Discovery and Use," 2016. [Online]. Available: <https://earthdata.nasa.gov/getting-petabytes-to-people-how-the-eosdis-facilitates-earth-observing-data-discovery-and-use>. Accessed on December 29, 2016.
- {80} G. M. Pinna and F. Ferrante, "The ESA Earth Observation Payload Data Long Term Storage Activities," European Space Agency, 2009. Retrieved December 26, 2016, from [http://www.cosmos.esa.int/documents/946106/991257/13\\_Pinna-Ferrante\\_ESALongTermStorageActivities.pdf/813babe0-58db-4e23-b710-3bd9d6b58b12](http://www.cosmos.esa.int/documents/946106/991257/13_Pinna-Ferrante_ESALongTermStorageActivities.pdf/813babe0-58db-4e23-b710-3bd9d6b58b12)
- {81} DigitalGlobe, Esri, and Harris Corporation Partner to Bring the Power of Geospatial Big Data Analytics and Deep Learning Technology to ArcGIS Users, GIS Cafe', Feb. 2017. [Online]. Available: [https://www10.giscafe.com/nbc/articles/view\\_article.php?section=CorpNews&articleid=1484459](https://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1484459)
- {82} V. Manilici, S. Kiemle, C. Reck, and M. Winkler, "EOLib Architecture Concept for an Information Mining System for Earth Observation Data," PV2013, ESRI, Frascati, 2013-11-04.
- {83} Commercial Satellite Imaging Market to Reach USD 5018.6 million by 2019, Globally: Transparency Market Research, GIS Cafe', 2014. [Online]. Available: [www10.giscafe.com/nbc/articles/view\\_article.php?section=CorpNews&articleid=1248945](http://www10.giscafe.com/nbc/articles/view_article.php?section=CorpNews&articleid=1248945)
- {84} Euroconsult, Prospects for the Small Satellite Market, 2016. [Online] Available: <http://www.euroconsult-ec.com/research/smallsats-2016-brochure.pdf>
- {85} D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A new framework for SAR multitemporal data RGB representation: Rationale and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117-133, 2015.
- {86} G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik, "Reflectance quantities in optical remote sensing - Definitions and case studies," *Remote Sens. Environ.*, vol. 103, pp. 27-42, 2006.
- {87} F. Pacifici, N. Longbotham, and W. J. Emery, "The Importance of Physical Quantities for the Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images," *IEEE Trans. Geosci. Remote Sens.*, in press, 2014.
- {88} A. Baraldi, "Impact of radiometric calibration and specifications of spaceborne optical imaging sensors on the development of operational automatic remote sensing image understanding systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 2, pp. 104-134, June 2009.
- {89} F. van der Meer and S.M. De John, Eds., *Imaging Spectrometry*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000.
- {90} *Imaging Spectrometry*, Eds. F. Van der Meer and S. De Jong, Dordrecht, The Netherlands: Springer 2011.
- {91} A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1299 - 1325, March 2010.
- {92} A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1326 - 1354, March 2010.
- {93} P. S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 459-479, Apr. 1988.
- {94} R. R. Irish, "Landsat 7 automatic cloud cover assessment (ACCA)," in *Proc. SPIE—Algorithms Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, S. S. Shen and M. R. Descour, Eds., 2000, vol. 4049, pp. 348-355. [Online]. Available: [http://www.gsfc.nasa.gov/IAS/handbook/pdfs/ACCA\\_SPIE\\_paper.pdf](http://www.gsfc.nasa.gov/IAS/handbook/pdfs/ACCA_SPIE_paper.pdf)
- {95} S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ: Wiley, 2004.
- {96} L. A. Dupigny-Giroux and J. E. Lewis, "A moisture index for surface characterization over a semiarid area," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 8, pp. 937-945, Aug. 1999.
- {97} C. H. Key and N. C. Benson, "Landscape assessment," USDA, Forest Service, Fort Collins, CO, 2006.



- {98} GEO, “The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan,” adopted Feb. 16, 2005. [Online]. Available: <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>
- {99} Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS) - Working Group on Calibration and Validation (WGCV), 2015. Land Product Validation (LPV). Accessed March 20, 2015. <http://lpvs.gsfc.nasa.gov/>
- {100} L. Boschetti, S.P. Flasse, and P.A. Brivio, “Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto boundary,” *Remote Sens. Environ.*, vol. 91, pp. 280–292, 2004.
- {101} O. Ahlqvist, “In search of classification that supports the dynamics of science: the FAO Land Cover Classification System and proposed modifications,” *Environment and Planning B: Planning and Design*, vol. 35, pp. 169–186, 2008.
- {102} M. Bossard, J. Feranec and J. Othel, “CORINE land cover technical guide – Addendum 2000, Technical report No 40,” European Environment Agency, 2000.
- {103} Y. Li, et al., Use of second derivatives of canopy reflectance for monitoring prairie vegetation over different soil backgrounds, *Remote Sens. Environment*, vol. 44, no. 1, pp. 81–87, 1993.
- {104} D. Riaño, E. Chuvieco, J. Salas, and I. Aguado, “Assessment of different topographic corrections in Landsat TM data for mapping vegetation types,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1056–1061, May 2003.
- {105} A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, “Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 2563–2586, 2006.
- {106} M. Cimpoi, S. Maji, I. Kokkinos I., and A. Vedaldi, “Deep filter banks for texture recognition, description, and segmentation,” *CoRR*, abs/1411.6836, 2014.
- {107} T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, “Geographic object-based image analysis - towards a new paradigm,” *ISPRS J. Photogram. Remote Sens.*, vol. 87, pp. 180–191, 2014.
- {108} Open source computer vision library (OpenCV), [Online]. Available: <http://opencv.org/>. Accessed on March 20, 2015.
- {109} T. Martinez, G. Berkovich, and K. Schulten, “Topology representing networks,” *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
- {110} B. Fritzsche, “Some Competitive Learning Methods,” 1997. Accessed 17 March 2015: <http://www.demogng.de/JavaPaper/t.html>
- {111} J. Feldman, “The neural binding problem(s),” *Cogn. Neurodyn.*, vol. 7, pp. 1–11, 2016.
- {112} A. Romero, C. Gatta, and G. Camps-Valls, “Unsupervised deep feature extraction for remote sensing image classification,” *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 3, pp. 1349–1362, Oct. 2015.
- {113} M. Volpi and D. Tuia, “Dense semantic labeling of sub-decimeter resolution images with convolutional neural networks,” arXiv:1608.00775v1 [cs.CV], 2 Aug 2016.
- {114} A. Baraldi, F. Despini, and S. Teggi, “Multi-spectral Image Panchromatic Sharpening, Outcome and Process Quality Assessment Protocol,” *Subjects: Computer Vision and Pattern Recognition (cs.CV)*, arXiv:1701.01942. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01942.pdf>
- {115} National Aeronautics and Space Administration (NASA) (2016b). Data Processing Levels. [Online]. Available: <https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products>. Accessed on December 20, 2016.
- {116} Andrea Baraldi, Dirk Tiede, Stefan Lang, “Automated Linear-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images,” *Subjects: Computer Vision and Pattern Recognition (cs.CV)*, arXiv:1701.01940. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01940.pdf>
- {117} K. J. Hayhurst, J. M. Maddalon, A. T. Morris, N. Neogi, H. A. Verstynen, A Review of Current and Prospective Factors for Classification of Civil Unmanned Aircraft Systems, National Aeronautics and Space Administration (NASA), August 2014. [Online] Available: [https://pdfs.semanticscholar.org/81d5/5f2e8b8d783c2303b1f028905a581f45ba5b.pdf?\\_ga=1.129579724.1353430899.1491025448](https://pdfs.semanticscholar.org/81d5/5f2e8b8d783c2303b1f028905a581f45ba5b.pdf?_ga=1.129579724.1353430899.1491025448). Accessed on April 1, 2017.
- {118} T. Poggio, “The Levels of Understanding framework, revised,” *Computer Science and Artificial Intelligence Laboratory, Technical Report, MIT-CSAIL-TR-2012-014, CBCL-308*, May 31, 2012.



### ***3 Technical report 1 (unpublished): Computational models of human vision - Developments and open challenges in automated detection of multi-scale image-contours, keypoints, texels and texture-boundaries in panchromatic and color images***

#### **Motivation and Contributions to the Dissertation**

At the core of this doctoral dissertation, the original contribution of this technical report (Chapter 3) is sixfold.

- (1) To provide computer vision (CV) applications with a scientific background well grounded in human vision and cognitive science, refer to Chapter 3.1 to Chapter 3.5. In the words of Iqbal and Aggarwal: “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception... This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” {19}.
- (2) To specify an original set of requirements to develop computational models of human vision whose goal is not only to mimic the processing of visual information in primates, but to match human performance in complex visual tasks, refer to Chapter 3.6. For example, consistency with perceptual visual illusions is considered mandatory. In the words of Pessoa, “if we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published computational vision models becomes surprisingly small” {20}.
- (3) To provide low-level vision tasks, defined in agreement with the Marr terminology {4}, with original solutions in operating mode in compliance with the proposed requirements specification for computational models of human vision, refer to Chapter 3.6. Original CV algorithms encompass the following.
  - (i) Automated statistical model-based color constancy for color image pre-processing, refer to Chapter 3.9.
  - (ii) Raw primal sketch: Automated zero-crossing image-contour detection consistent with the Mach bands illusion for ramp-edge detection. Refer to Chapter 3.11 to Chapter 3.19.
  - (iii) Raw primal sketch: Automated keypoint detection. Refer to Chapter 3.11, Chapter 3.15 and Chapter 3.16.
  - (iv) Raw primal sketch: Automated zero-crossing image-segment detection from zero-crossing image-contours. Refer to Chapter 3.20.
  - (v) Full primal sketch: perceptual spatial grouping of texture elements (texels). Refer to Chapter 3.20.
- (4) To present and discuss an original pair of expert systems (prior knowledge-based decision trees) for color naming in a calibrated multi-spectral (MS) reflectance space or in an uncalibrated RGB color space, either true- or false-color. Color naming transforms a numeric variable (color value, colorimetric sensation) into a categorical variable, specifically, into color names belonging to a pre-defined dictionary of color names, equivalent to a latent/hidden variable and eligible for use in symbolic human reasoning. Refer to Chapter 3.10.
- (5) To present and discuss an original perceptual image-pair quality/ similarity/ dissimilarity index/ metric. To date, no “universal” image-pair perceptual visual quality metric exists. Refer to Chapter 3.13.
- (6) To present and discuss an original conceptual unifying framework between spatial variance, spatial autocorrelation and the proposed 2D wavelet-based filter bank. Refer to Chapter 3.21.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 3 (Technical report 1) are color filled.

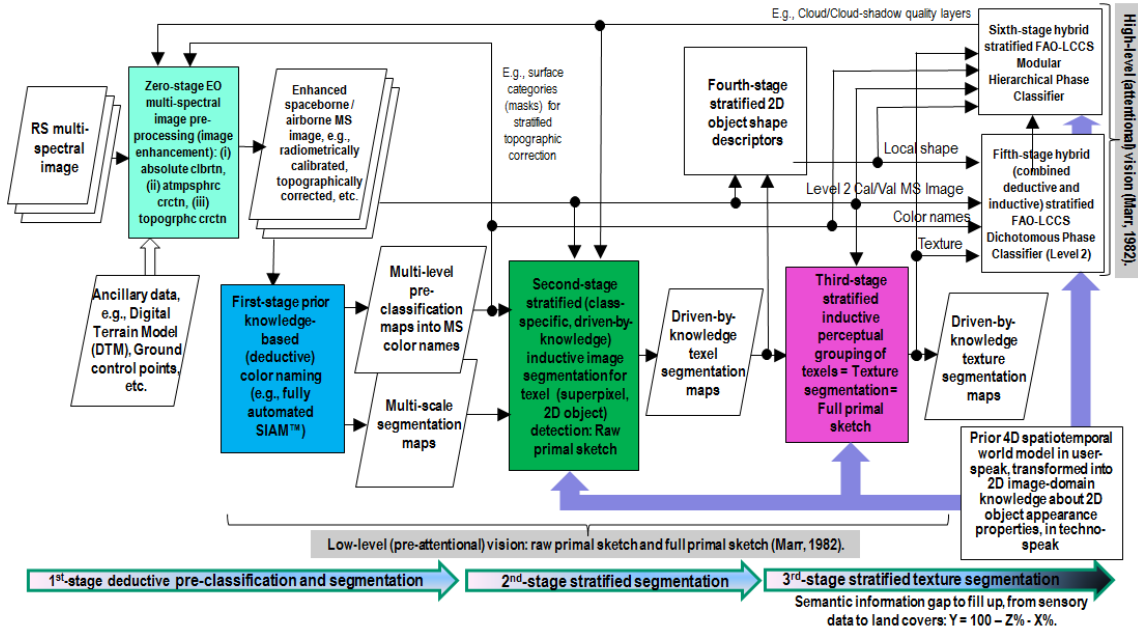


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 3 (Technical report 1).



## - Technical report No. 1, version 10.1 -

April. 9, 2017

# Computational models of human vision: Developments and open challenges in automated detection of multi-scale image-contours, keypoints, texels and texture-boundaries in panchromatic and color images

*Andrea Baraldi*

*Università degli Studi di Napoli Federico II, Dipartimento di Agraria  
Corso di Dottorato di Ricerca, XXIX ciclo*

### 3.1 Vision is a cognitive task

Cognitive science is the interdisciplinary scientific study of the mind and its processes [81], [84], [85], [86], [138], [169]. Cognitive science examines what cognition is, what it does and how it works [169]. Cognitive scientists study behavior and cognition. Mental faculties of concern to cognitive scientists include perception, language, memory, attention, reasoning, and emotion. Cognition is synonym of (biological) intelligence, learning, adaptation, self-organization [81], [138], [169]. Cognitive science especially focuses on how information/knowledge is represented, acquired, processed and transferred either in the neuro-cerebral apparatus of living organisms or in machines, e.g., computers. Neural systems are distributed processing systems consisting of neurons as elementary processing elements and synapses as lateral connections. Distributed processing systems are also called “complex systems” or neural networks (NNs) [169]. Neurophysiology studies the neuro-cerebral apparatus of living organisms at the micro scale of analysis. The so-called “connectionist approach” promoted by traditional cybernetics for artificial reasoning (intelligence) in machines is that an “artificial mind” (software) cannot be pursued independently of an “electronic brain” as its physical support (hardware). Electronic brain is synonym of artificial distributed processing system, typically called artificial NNs (ANNs) [76], [79]. ANNs are made of “simple” processing units and inter-node connections at the micro (local) scale of analysis, but are capable of complex network-wide behaviors at the macro (global) scale of analysis largely unpredictable based on analytic considerations and theoretical knowledge; hence, ANNs typically require computational simulations to be investigated [76], [79], [81], [169]. In cybernetics “complex system” is synonym of “artificial mind-electronic brain” encompassing the complementary not alternative macro and micro scales of analysis of cognitive functions [169] (p. 8). Is it convenient and even possible to mimic biological mental functions, e.g., human reasoning, by an artificial mind whose physical support is not an electronic brain? The answer is no according to the “connectionists approach” adopted by neuromorphic engineering. Neuromorphic engineering is an interdisciplinary subject that takes inspiration from biology, physics, mathematics, computer science, and electronic engineering to design artificial neural systems, such as vision systems, head-eye systems, auditory processors, and autonomous robots, whose physical architecture and design principles are based on those of biological nervous systems. The connectionist approach of neuromorphic engineering is alternative to the “symbolic approach” promoted by traditional artificial intelligence (AI) where an artificial mind is investigated independently of its physical support [169]. In recent years new developments of neuroscience and computer science have been relaunching the connectionist approach traditionally promoted by cybernetics. Fig. 3.1-1 illustrates the scientific fields that contributed to the birth of cognitive science as the interdisciplinary study of the mind and its processes, including linguistics, philosophy, anthropology, psychology (focused on the study of the mind at a macro level of analysis independent of the brain at a micro scale of analysis), neuroscience (focused on the study of the neuro-cerebral apparatus in living organisms), deductive (top-down) artificial intelligence whose symbolic approach investigates artificial mind functions independently of their physical support, and inductive (bottom-up) machine learning-from-data where the connectionist approach of traditional cybernetics is adopted in the study of ANNs [81], [138], [169].

In [81], intelligence by a cognitive agent or human decision maker (HDM) provided with sensory data (stimula, sub-symbolic numeric/quantitative variables) generated from his/her interaction with an external spatiotemporal environment is defined as the capability of forecasting the sensorial consequences of his/her actions and/or decisions. Hence, at the basis

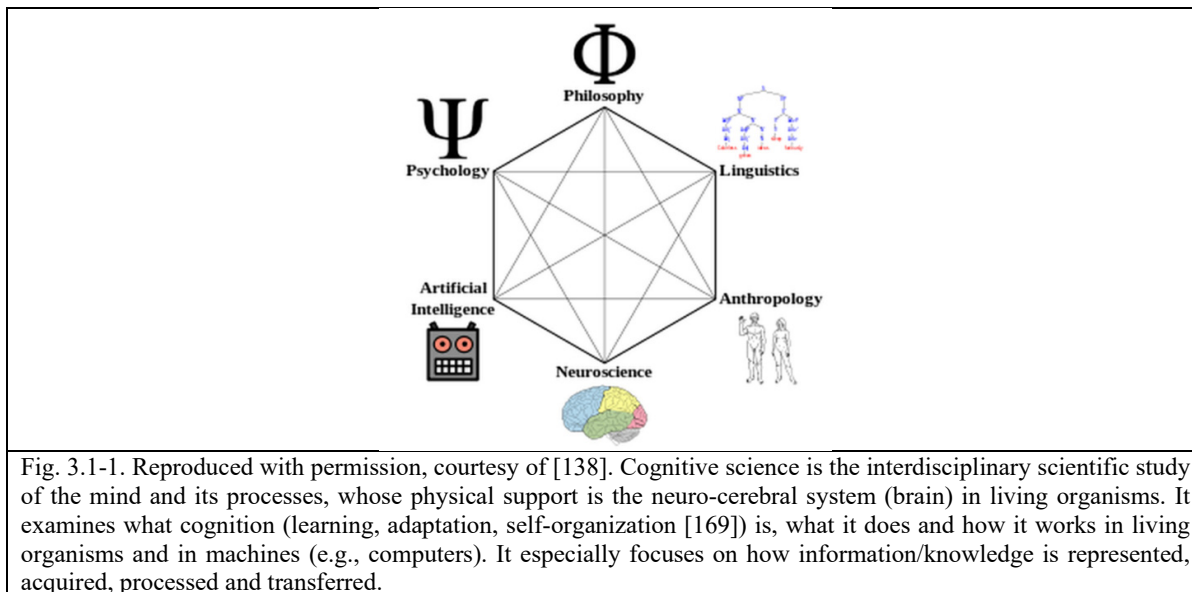




of cognition (intelligence, decision-making) there is the capability of interpreting sensory data into symbolic categorical variables of semantic quality. According to philosophical hermeneutics, quantitative data interpretation is inherently equivocal (ill-posed, subjective) because it requires a pro-active role by the message receiver to provide the (sub-symbolic) data message with a (symbolic) meaning, since there is no semantics in numeric variables [31], [32].

In the words of Poggio “the problem of learning is at the core of the problem of intelligence and of understanding the brain... From the point of view of statistical learning theory, it seems that the incredible effectiveness with which humans (and many animals) learn from and perform in the world cannot result only from superior learning algorithms, but also from a huge platform of knowledge and priors. This is right in the spirit of Marr’s computational approach: constraints, “discovered” by evolution, allow the solution of the typically ill-posed problems of intelligence. Thus evolution is responsible for intelligence, and should be at the top of our levels of understanding” [189].

In his tribute to David Marr’s contribution in Bayesian inference in vision, Quinlan states: “one of David Marr’s key is the notion of constraints. The idea that the human visual system embodies constraints that reflect properties of the world is foundational. Indeed, this general view seemed (to me) to provide a sensible way of thinking about Bayesian approaches to vision. Accordingly, Bayesian priors are Marr’s constraints. The priors/constraints have been incorporated into the human visual system over the course of its evolutionary history (according to the “levels of understanding” manifesto extended by Tomaso Poggio in 2012 [189])” [190].



The transformation of data into information and knowledge is at the foundation of information technology (IT). Unfortunately, as clearly pointed out by philosophical hermeneutics, word “information” has a double meaning [31], [32]. In practice the double meaning of word “information” works as a major cause of equivocality at the root of IT. In greater detail, there are two complementary not alternative concepts (theories, paradigms) of information: (I) the unequivocal quantitative concept of *information-as-thing*, related to the popular Shannon’s mathematical theory of data communication/transmission [125], irrespective of the meaning of the transmitted message, e.g., bits or documents are transmitted through a communication channel independent of their meaning, to be correctly received by a message receiver, and (II) the inherently equivocal qualitative concept of *information-as-data-interpretation*, where the message receiver has a pro-active role in providing (sub-symbolic) quantitative data with a (symbolic) meaning, because there is no semantics in numeric variables [31], [32].

Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. RS encompasses Earth observation (EO) image understanding. EO image understanding is a subset of computer vision (CV). CV is a subset of vision, encompassing both biological and artificial vision. Vision is intended as synonym of scene-from-image



representation (low-level or pre-attentive vision) and understanding (high-level or attentive vision). Vision is an inherently ill-posed cognitive (inference) process dealing with equivocal information-as-data-interpretation. Hence, vision is a subset of cognitive science, see Fig. 3.1-2.

To summarize, Fig. 3.1-2 shows that vision, which includes EO image understanding as a special case of CV whose lower bound is human vision, is a multi-disciplinary cognitive task dealing with the inherently ill-posed interpretation of 2D sensory data (images) for 4D spatiotemporal scene reconstruction and understanding. In the words of Iqbal and Aggarwal: “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception... This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [49].

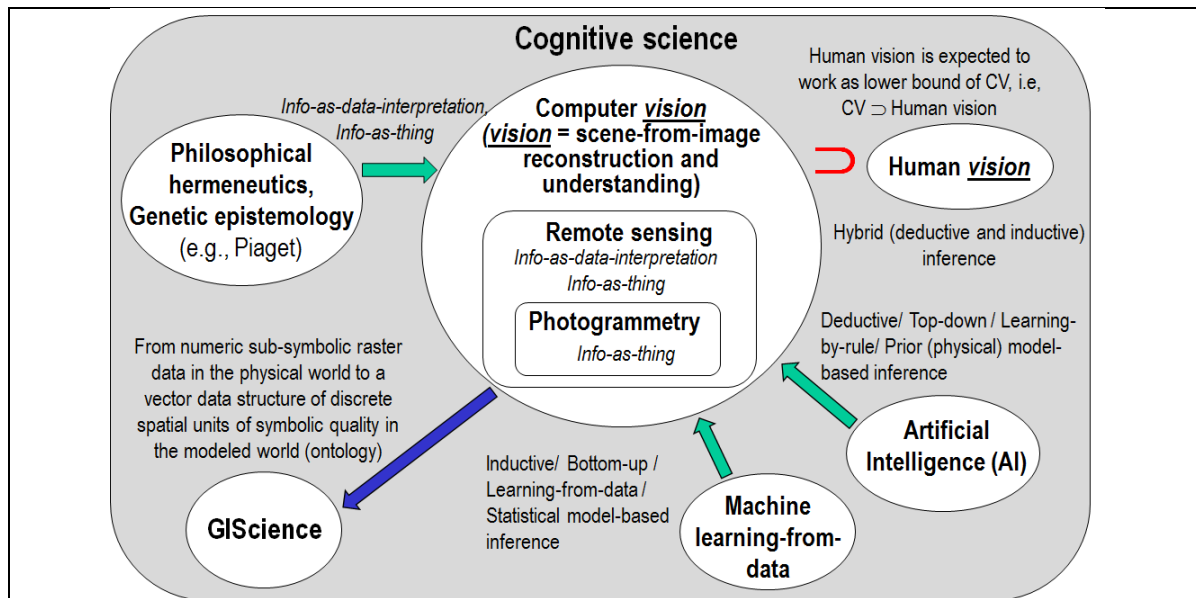


Fig. 3.1-2. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. Cognitive science is the interdisciplinary scientific study of the mind and its processes [84], [85], [86], [138]. It examines what cognition (learning [189], adaptation, self-organization [169]) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred either in the neuro-cerebral apparatus of living organisms or in machines, e.g., computers. Neuroscience, in particular neurophysiology, studies the neuro-cerebral apparatus of living organisms. Neural network (NN) is synonym of distributed processing system consisting of neurons as elementary processing elements and synapses as lateral connections. Is it convenient and even possible to mimic biological mental functions, e.g., human reasoning, by an artificial mind whose physical support is not an electronic brain implemented as an artificial NN (ANN)? The answer is no according to the “connectionists approach” promoted by traditional cybernetics, where a complex system always comprises an “artificial mind-electronic brain” combination. This is alternative to traditional artificial intelligence (AI) whose symbolic approach investigates an artificial mind independently of its physical support [169].

Formally, a finite image  $I$  is a function that assigns colors (coded here by real numbers belonging to set  $\mathfrak{R}$ ) to a finite, two-dimensional (2D, planar) rectangular array of locations in space, coded here by ordered pairs of integers, row and column (col). It is convenient to fix two integers  $N$  and  $M$  as maximum numbers of rows and columns and let two 1D spatial coordinates  $X$  and  $Y$  range as  $X = \{0, 1, \dots, N\}$  and  $Y = \{0, 1, \dots, M\}$ . Then, a one-dimensional (1D) signal is simply a function  $I: X \rightarrow \mathfrak{R}$ , equivalent to a 1D vector stream (sequence). For any spatial unit  $x \in X$ ,  $I[x]$  denotes the value assigned by  $I$  to  $x$ . A 2D image is a function  $I$  mapping the Cartesian product  $X \times Y$  into  $\mathfrak{R}$ . For any pair (row, col) of image coordinates  $(x, y) \in X \times Y$ ,  $I[x, y]$  denotes the value assigned by  $I$  to  $(x, y)$ . Regardless of whether  $I$  is a 1D data stream or a 2D vector sequence, discrete spatial points in the domain of function  $I$  will be called picture elements, known as pixels [52].

The main role of any biological or artificial visual system is to back-project the information from the (2D) image domain to the 4D spatiotemporal world domain [45], see Fig. 3.1-3. In greater detail, the inherently ill-posed objective of an image understanding system (IUS) is to provide one or more plausible symbolic description(s) of a 3D scene belonging to a 4D spatiotemporal world-through-time domain, viewed by the imaging sensor and projected onto a (2D) image domain (planar array) at a given acquisition time, by finding one or multiple plausible associations between sub-symbolic numeric variables in the image domain (sensations, e.g., image contours, pixel color values, etc.) and symbolic classes (stable percepts) of 4D objects in the spatiotemporal world-through-time domain [45]. A world model, also called 4D spatiotemporal ontology of the world-through-time, can be graphically represented as a semantic (concept) network [56], [57]. In practice, the human visual system is able to construct on the basis of a brief glance a complete scene-from-image representation in our brain: from global gist to spatial layout of objects [143], from local syntax of individual objects to multiple plausible semantic scene interpretations even provided with emotions [24].

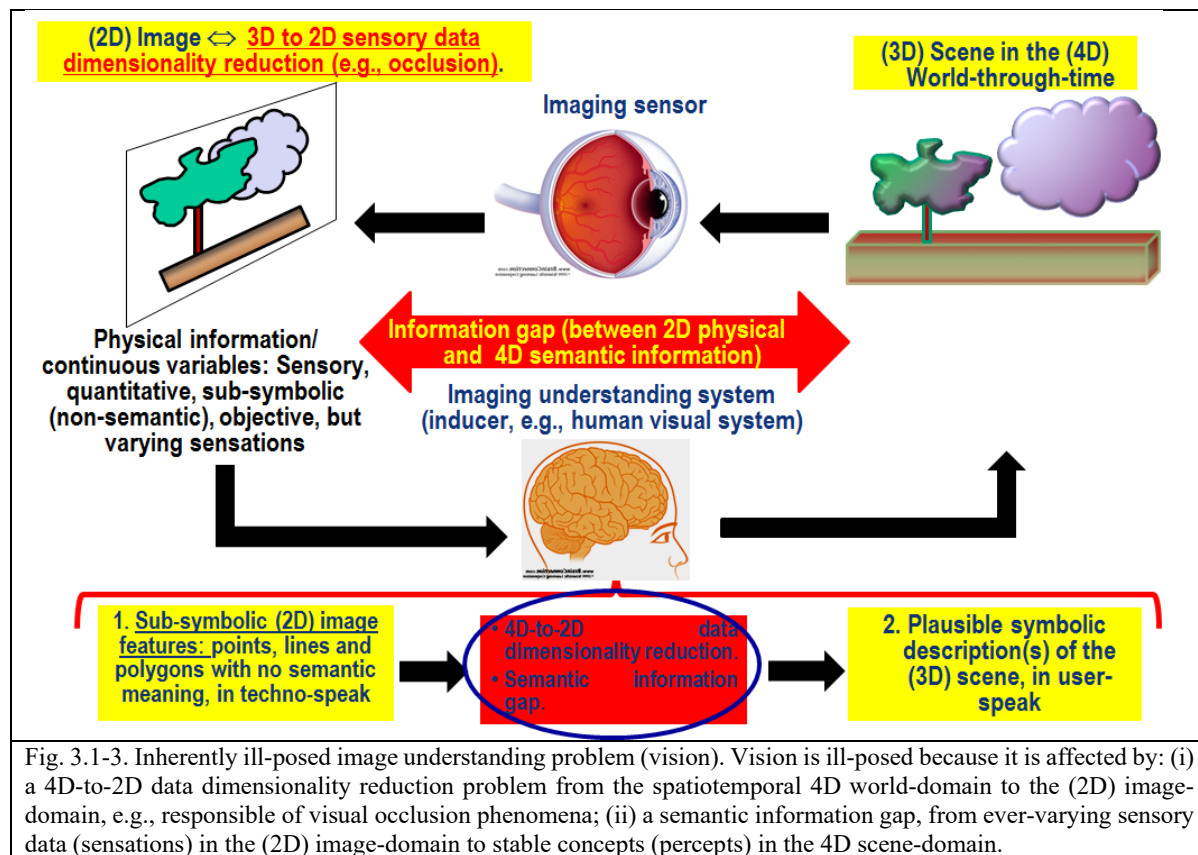


Fig. 3.1-3. Inherently ill-posed image understanding problem (vision). Vision is ill-posed because it is affected by: (i) a 4D-to-2D data dimensionality reduction problem from the spatiotemporal 4D world-domain to the (2D) image-domain, e.g., responsible of visual occlusion phenomena; (ii) a semantic information gap, from ever-varying sensory data (sensations) in the (2D) image-domain to stable concepts (percepts) in the 4D scene-domain.

Noteworthy, there are about 40 types of biological vision systems in nature, e.g., foveated and non-foveated visual systems [72], [74], [109], [110]. The former employs spatial filters whose resolution is eccentricity dependent, i.e., different spatial filters span different visual fields and the eye moves across fixation points selected in a scene. In the latter, the whole visual field is processed the same and the eye does not move across the target scene, like in bees. Hence, eccentricity-dependent foveated vision must rely on a focus-of-visual attention (Focus-of-Attention, FoA) [28], [41], [43], [66] and eye fixation [109], [110] mechanism, which is synonym of saliency map extraction [71], [73], [74], [142], also called selective window search [184] in deep convolutional neural networks (DCNNs) [95], [96], [114], [115], [116], [117], [118], [119], [120], [127], [128].

In the terminology proposed by the Open Geospatial Consortium (OGC) Simple Feature Specification [55], sub-symbolic (2D) image features in the image-domain are (0D) points, (1D) lines, (2D) polygons, or multi-part polygons (strata). In the CV literature, image plane entities are also called image-polygons, image-objects, 2D segments, 2D regions, patches, parcels, blobs, tokens [5], or texture elements [5], known as texels [2], [3], [58], [59], [60], considered as inputs to an intermediate-level vision phase known as full primal sketch [5], perceptual grouping [6] or texture segmentation [58]. For



example, in [155], [188] texels are considered equivalent to superpixels, where a superpixel is defined as a connected set of pixels featuring a “homogeneous” numeric color value or the same categorical color name belonging to a finite and discrete dictionary (codebook) of color names as codewords. A pre-defined dictionary of color names must be community-agreed upon to be used by members of the community. It provides a mutually exclusive and totally exhaustive partition of a color space into hyperpolyhedra neither necessarily convex nor connected [78]. Well-known visual features (numeric variables) in the image domain are color, local shape of image-objects, texture (perceptual grouping of texels [2], [3], [5], [58], [59], [60]) and inter-object spatial relationships, either topological (e.g., adjacency, inclusion, etc.) or non-topological (e.g., spatial distance, angle measure) [45], [80], [167].

In synthesis, vision is synonym of scene-from-image representation and understanding. It is an inference problem inherently ill-posed [8], [26], [44], [45], [46], [47], [48], [68], [71] in the Hadamard sense [75]. As such, scene-from-image representation is extremely difficult to solve because it is eligible for one, none or multiple solutions [75], i.e., it is an NP-hard problem whose computational complexity is non-polynomial (NP) [68], [71]. Vision is inherently ill-posed because, first, it is affected by a 4D to 2D data dimensionality reduction phenomenon. For example, scene-from-image data dimensionality reduction is responsible of occlusion phenomena. Second, vision is affected by a semantic information gap, from quantitative sub-symbolic ever-varying 2D sensory data (sensations, numeric variables) in the (2D) image-domain, to stable symbolic percepts (nominal/categorical variables provided with semantics) in the 4D spatiotemporal scene-domain. Since it is inherently ill-posed, vision requires *a priori* knowledge in addition to sensory data to make the inference problem better conditioned for solution [76].

According to Hadamard, analytic or statistical models of physical phenomena are defined as well-posed (respectively, ill-posed) when they satisfy (respectively, do not satisfy at least one of) the following requirements [75].

- A solution exists.
- The solution is unique. And
- The solution’s behavior hardly changes when there is a slight change in the initial condition.

In perceptual phenomena, including vision understood as synonym of scene-from-image reconstruction and understanding, input ever-varying sensations are equivalent to observable numeric/quantitative variables of “sub-symbolic” quality, i.e., input sensations are equivalent to sensory data directly measured in the real world and provided with no semantic content. Stable percepts are output nominal/categorical/qualitative variables of “symbolic” quality, i.e., they are categorical variables provided with a semantic content in a modeled world, also known as world ontology or “world model” available *a priori* in addition to sensory data [45]. In statistics, latent/hidden variables are not directly measured, but rather inferred from observable numeric variables to link sensory data in the real world to categorical variables of semantic quality in the modeled world. The terms “hypothetical variable” or hypothetical construct may be used when latent variables correspond to abstract concepts, like perceptual categories or mental states. Hence, to fill the semantic gap from low-level numeric variables of sub-symbolic quality to high-level categorical variables of semantic quality, statistical models of human vision, i.e., computer vision (CV) systems, are expected to rely on hypothetical variables, equivalent to mid-level categorical variables of “semi-symbolic” quality, i.e., qualitative variables provided with some degree of semantic content.

In the terminology of computational complexity problem analysis, let us express the amount (length) of input data needed to describe a problem instance as  $N$ .

(Start long quote of Tsotsos [68], p. 425)

“If the number of (primitive) operations required to solve a problem is an exponential function of  $N$ , then the problem has exponential time complexity”, which is a special case of non-polynomial (NP) computational complexity. “If the number of required operations can be represented by a polynomial function in  $N$ , the problem has polynomial (P) time complexity. Similarly, space complexity is defined as a function for an algorithm that expresses its space or memory requirements. Algorithm complexity is the cost of a particular algorithm. This should be contrasted with problem complexity, which is the minimal cost over all possible algorithms. These two forms of complexity are often confused. The notion of a good algorithm and an intractable problem was developed in the mid-to-late 1960s. A good (tractable) algorithm is one whose time requirements (number of processing steps) can be expressed as a polynomial function of

input length  $N$ ,  $P(N)$ ... The class  $P$  of tractable computational problems consists of all those problems that can be solved in polynomial time. If we accept the premise that a computational problem is not tractable unless there is a polynomial-time algorithm to solve it, then all tractable problems belong in  $P$ ... An intractable problem is one whose time requirements are exponential functions of problem length, or in other words, a non-polynomial problem,  $NP$ , that cannot be solved by any polynomial time algorithm for all instances... If a problem is in the class  $NP$ , then the problem can be solved by an exponential algorithm having time complexity  $O(2^{P(N)})$ . A problem is  $NP$ -Complete if it is in the class  $NP$ , and it polynomially reduces to an already proven  $NP$ -Complete problem. These problems form an equivalence class. Clearly, there must have been a "first"  $NP$ -Complete problem. There are hundreds of  $NP$ -Complete problems - Knapsack is one of them. If any  $NP$ -Complete problem can be solved in polynomial time, then they can all be. Most computer scientists are pessimistic about the possibility that non-exponential algorithms for these problems will ever be found, so proving a problem to be  $NP$ -Complete is now regarded as strong evidence that the problem is intrinsically intractable. If an efficient algorithm can be found for anyone (and hence all)  $NP$ -Complete problems, however, this would be a major intellectual breakthrough with immense practical implications. What does a computer scientist do when confronted with an  $NP$ -Complete problem? A variety of approaches have been taken.

1. Develop an algorithm that is fast enough for small problems but would take too long with larger problems. This approach is often used when the anticipated problems are all small.
2. Develop a fast algorithm that solves a special case of the problem, but does not solve the general problem. This approach is often used when the special case is of practical importance.
3. Develop an algorithm that quickly solves a large proportion of the cases that come up in practice, but in the worst case may run for a long time. This approach is often used when the problems occurring in practice tend to have special features that can be exploited to speed up the computation.
4. For an optimization problem, develop an algorithm that always runs quickly but produces an answer that is not necessarily optimal. Sometimes a worst-case bound can be obtained on how much the answer produced may differ from the optimum, so that a reasonably close answer is assured. This is an area of active research, with suboptimal algorithms for a variety of important problems being developed and analyzed.
5. Use natural parameters to guide the search for approximate algorithms. There are a number of ways a problem can be exponential. Consider the natural parameters of a problem rather than a constructed problem length and first attempt to reduce the exponential effect of the largest valued parameters.

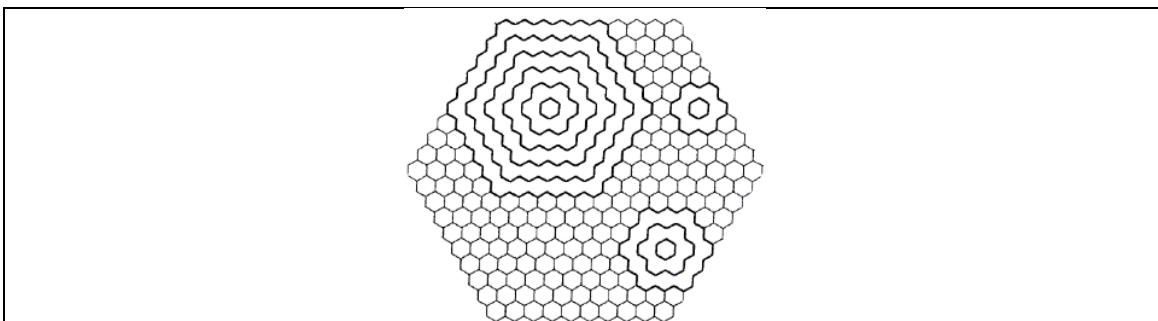


Fig. 3.1-4. Reproduced with permission, courtesy of [68]. The hexagonal retinotopic stimulus representation. The hexagon is of order  $N$ , that is,  $N$  elements per side. The diameter of the hexagon is  $2N-1$  elements. For three of the elements, the corresponding complete set of receptive fields that can be centered on those elements is shown. In this way each element in the hexagon can be the center of a number of hexagonal receptive fields. There are  $N^3$  receptive fields in all.

$NP$ -completeness effectively eliminates the possibility of developing a completely satisfactory algorithm. Once a problem is seen to be  $NP$ -complete, it is appropriate to direct efforts toward a more achievable goal. In most cases, a direct understanding of the size of the problems of interest and the size of the processing machinery is of tremendous help in determining which are the appropriate approximations... Note that the boundary between "good" and "bad" problems is not precise. A polynomial time complexity of  $N^{1000}$  is surely not very practical whereas an exponential time complexity of  $2^N$  with  $N = 0.001$  is perfectly realizable. Yet empirical evidence seems to point to the fact that natural problems simply do not have such running times, and that the distinction between tractable  $P$  and intractable  $NP$  problem complexity in input length  $N$  is a useful one..."



(End long quote of Tsotsos [68], p. 425)

To recapitulate, according to [68] there are only a few options available when one is faced with a problem that is NP-Complete. The unbounded visual search problem, where VP is the number of target visual prototypes and M is the number of visual quantitative features (feature maps) extracted from an image of P pixels, is exponential in the number of image pixels, P. It means that:

$$\text{Matching operations required in the worst case of the unbounded visual search problem} = VP \times 2^{P \times M} \quad (1-1)$$

In a visual system each visual processor (hypercolumn) has a receptive field defined as a subset of pixel locations, equivalent to a subimage. Each visual prototype must be matched against each processing element whose receptive field is a possible subimage. Assuming a hexagonal image of order N, we only consider hexagonal contiguous regions of whole array elements as processor receptive fields. Simple geometry yields  $N^3$  receptive fields of the type described above over the whole image, or in pixels, approximately,  $[P^{1.5} / (3 \times \sqrt{3})] + P/2 + 5 \times \sqrt{P/3} / 8$ , see Fig. 3.1-4. This  $N^3$  estimate, which gives the total number of hexagonal, contiguous receptive fields of all sizes and centered at all locations in the image array, replaces the number P of pixels in the visual field. This reduces Eq. (1-1) to become equal to

$$\text{Matching operations required in the worst case of the unbounded visual search problem} = VP \times N^3 \times 2^M \quad (1-2)$$

The powerset of maps,  $2^M$ , still remains in the expression because it is not known a priori which subset of maps is the correct one for the best image-to-prototype match; hence in the worst case all subsets must be examined. To cope with this NP-hard problem, one strategy is to look for optimizations and approximations by a computational model of human vision constrained by the few relevant estimates of natural visual parameters including the amount of input data and the number of target visual prototypes in memory. These natural visual parameters are the following [68].

(Start long quote of Tsotsos [68])

- (i) "A stimulus array or retina with P elements (pixels). This is a retinotopic representation, that is, one whose physically adjacent elements represent spatially adjacent regions in the visual scene (synonym of topology-preserving feature mapping). P is the number of locations in the retinotopic representation; an upper, middle, and lower value will be used. The number of receptors in the retina (130,000,000) is the upper value, the number of retinal ganglion cells (approximately 1,000,000 and roughly the same as the number of pixels in a 1K X 1K image) is the middle value, and the size of a 256 X 256 image is the lower values (65,536 pixels).
- (ii) At each array element, one or more quantitative/numeric visual parameters of the scene may be computed. In practice, a map is defined as a retinotopic representation of only one type of visual parameter (feature). Each feature map represents at least a portion of visual space and each has its own distinct characteristics. Maps are logical abstractions, not necessarily physically separable entities. For example, Marr uses six different quantities in his primal sketch, from which all other required visual information can be derived [5]: relative depth; local changes in depth; discontinuities in depth; local surface orientation; local changes in surface orientation; and discontinuities in surface orientation. There may be 30 visual areas or so in primates, but not all are organized retinotopically, and, even then, with varying degrees of retinotopy. Because many areas have more than one population of neurons, there can be more logical maps than physical ones. The logical retinotopic map is the unit of the parameter M discussed in [68]. The areas commonly accepted as being retinotopic include VI, V2, V3, MT, and V4, whereas the nonretinotopic ones include IT, posterior parietal cortex, and the frontal eye fields. The division between retinotopic and nonretinotopic areas, although fuzzy in general, may be placed after areas MT and V4 and before IT, area 7, and the frontal eye fields. Feature maps seem to be organized hierarchically, as a partial ordering, so that the greater the distance from the retina, the smaller the maps are, and the larger the receptive fields of their neurons. There is also more than one pathway from the retina to higher levels of processing. In [68], Tsotsos adopts an upper estimate of M as 12. For example, let's consider the computational model of human vision proposed in Chapter 3.7, where number of even- and odd-symmetric filter types = 2, spatial scales = 4, spatial orientations = 2, capable of modeling end-stopped cells as keypoint detectors. Then physical maps are 16 belonging to 4 hierarchical levels. If logical feature maps are considered one image-contour (e.g., zero-crossing pixel map [5]) per scale across orientations and one keypoint map across scales and orientations, then logical feature maps are 5. If logical feature maps are considered one image-contour (e.g., zero-crossing pixel map or





zero-crossing segment map [5]) across scale and orientations and one keypoint map across scales and orientations, then logical feature maps are 2.

- (iii) A knowledge base of target visual prototypes, each representing a particular visual object, event, scene, or episode. Let VP represent the number of prototypes. There are 30,000 readily identifiable individual objects in the world, excluding whole scenes or collections of objects. If these were included a very large number of visual prototypes VP would presumably result. Thus, a conservative lower estimate for the number of target visual prototypes is  $VP = 100,000$ . A large but arbitrary upper estimate would be  $VP = 10,000,000$ .
- (iv) A large layer of identical processors (hypercolumns), each able to choose a subset of the stimulus array locations, fetching a subset of the tokens representing physical characteristics at each location, accessing one visual prototype, and then matching the token set to the prototype, see Fig. 3.1-5 and Fig. 3.1-6. Hubel and Wiesel [17], [19] discovered that primary visual cortex (also called area 17 or VI in mammals) exhibited a distinct columnar architecture with some apparent functional significance: the hypercolumn. They proposed that the hypercolumn is the basic processing unit and that each contain a complete collection of neurons sensitive to and selective for all the basic visual properties (color, motion, orientation, binocular disparity, luminance). The receptive fields within a hypercolumn were all overlapping and specific for a given region of visual space, called receptive field. Crossing into a neighboring hypercolumn reveals the same collection of neural sensitivities, but for an adjacent region of visual space. Thus, the representation is retinotopic. Thus, a receptive field is defined as the area of the visual scene in which a change in the visual stimulus causes a change in the output of the processor to which it is connected. The area of each hemisphere of the primary visual cortex in humans, V1, is 1500-3700 mm<sup>2</sup>, which comprises the set of "units of output" in V1, called hypercolumns, where each hypercolumn is approximately 1 mm<sup>2</sup> in area and that there are therefore 1500-3700 hypercolumns or 2100 on average in V1. The matching process is the basic visual operation. Matching here means that the processor determines whether or not the collection of tokens over the selection of locations optimally represents an image-specific projection of the prototype. The output of a processor is matching success or failure with perhaps an indication of response strength. Each processor completes this operation in S seconds. The final output of the system is also available in S seconds; thus the actual time required for this process does not matter. The effective speed-up due to parallelism will be denoted by the variable II. No difficulty is posed for determining first-order complexity by not specifying exactly what each processor does. As long as the complexity of each is polynomial rather than exponential, there is no change in complexity class".

(End long quote of Tsotsos [68])

It is worth mentioning here that term "retinotopic stimulus representation" [68] adopted in neurophysiology, psychophysics and neuroengineering [187] is synonym of topology-preserving feature mapping in computer science [69], [145], [178], [179], [180], which is synonym of 2D image analysis in CV versus 1D image analysis, synonym of non-retinotopic representation in the visual system [187]. For example, a quote in neuroengineering is "the geometry of early visual processing is relatively well understood. Three dimensional stimuli are imaged on the retina through the optics of the eye following the laws of optics. This geometric transformation maps neighboring points in the environment to neighboring photoreceptors in the retina. The projections of neurons from retina to early visual cortical areas preserve these neighborhood relations, a property known as retinotopy. Thus, scenes create two-dimensional images like images created by a camera" [187].

A computational architecture of human vision that satisfies computational complexity-level analysis, including an input abstraction hierarchy, logically segregated visual maps, a layer of parallel processors, a hierarchically organized set of visual prototypes, and a spatially contiguous definition of receptive fields is shown in Fig. 3.1-5 and Fig. 3.1-6.



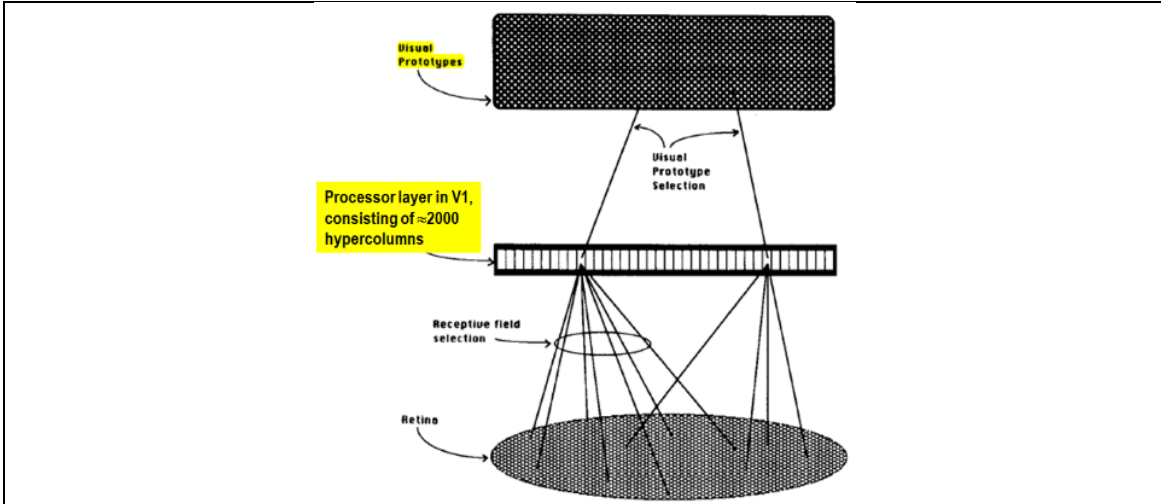


Fig. 3.1-5. Reproduced with permission, courtesy of [68]. The idealized architecture of vision that corresponds to the NP-Complete problem of unbounded visual search. Each processor within the processor layer matches one subset of retinal locations and measurements with one visual prototype from memory.

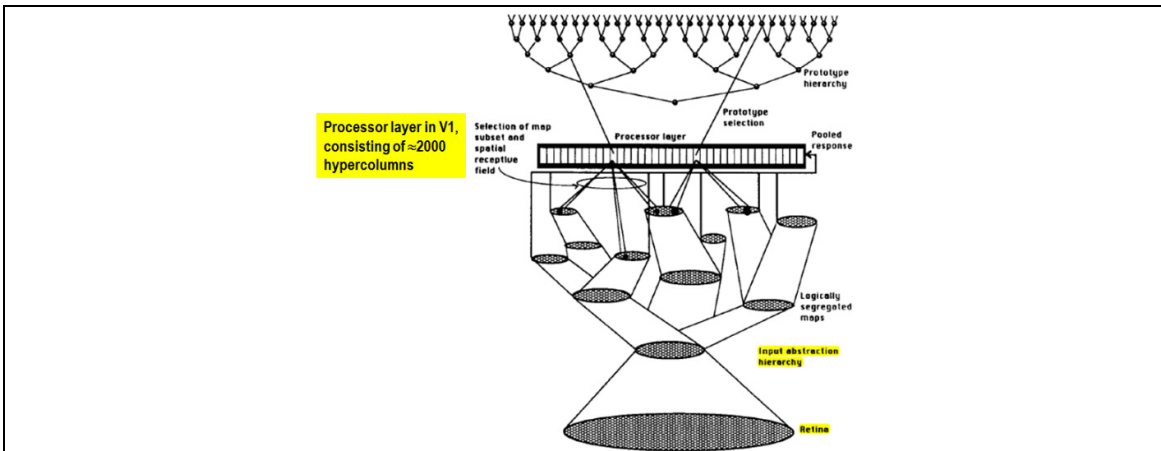


Fig. 3.1-6. Reproduced with permission, courtesy of [68]. The architecture that satisfies computational complexity-level analysis. It includes an input abstraction hierarchy, logically segregated visual maps, a layer of parallel processors, a hierarchically organized set of visual prototypes, and a spatially contiguous definition of receptive fields.

There is a long-standing, sometimes contentious debate in artificial intelligence (AI) concerning the relative merits of a symbolic, top-down approach vs. a neural, bottom-up approach to engineering intelligent machine behaviors adoptive to sensory data. While neurocomputational methods excel at lower-level cognitive tasks (incremental learning for pattern classification, low-level sensorimotor control, fault tolerance and processing of noisy data, etc.), they are largely non-competitive with top-down symbolic methods for tasks involving high-level cognitive problem solving (goal-directed reasoning, metacognition, planning, etc.) [93].

It is clearly acknowledged by the machine learning community that “any inductive learning-from-data problem (e.g., supervised data classification or function regression, unsupervised data clustering, quantization or density function estimation) is inherently ill-posed (difficult to solve) in the Hadamard sense and requires *a priori* knowledge in addition to data to become better posed for numerical solution” [76] (p. 26).

To take advantage of the unique features of each and overcome their shortcomings, statistical/ bottom-up/ inductive/ learning-from-examples inference models and physical/ top-down/ learning-by-rule/ deductive inference models have been



typically combined into hybrid inference systems [77]. This complies with the observation that, in nature no cognitive system, consisting of phenotype and genotype, starts from scratch (*tabula rasa*). In particular, any biological inductive learning-from-examples phenotype explores the neighborhood of its initial conditions, provided by a deductive genotype, in a solution space [81]. Intuitively, we can say that if an entity/agent is born “square” in the solution space, then there is a longer path in the solution space to be covered during life-time to die “round”. In practice, a hybrid inference system can alternate deductive and inductive inference stages, starting from a deductive (prior knowledge-based) first stage required to initialize the hybrid inference process.

Hence, the inherently ill-posed vision process, which requires *a priori* knowledge in addition to sensory data to become better posed for numerical solution, must rely on a *hybrid* (combined deductive/top-down and inductive/bottom-up) inference system [77]. This (unquestionable) true-fact is proved by optical illusions in visual perception. The study of perceptual visual illusion, occurring when humans mentally see (perceive) what is not either in the observed scene (e.g., Mach bands illusion [25]) or in their retinal stimuli (e.g., the well-known Kanizsa Triangle illusion, first described in 1955 by an Italian psychologist named Gaetano Kanizsa, see Fig. 3.1-7), has yielded much insight into what assumptions (prior knowledge) the visual system requires in addition to sensory data to become better conditioned for achieving plausible solution(s) in scene-from-image reconstruction and understanding. One popular visual illusion is the Kanizsa Triangle illusion, see Fig. 3.1-7. In this optical illusion, a white equilateral triangle can be seen in the image even though there is not actually a triangle there. The effect is caused by illusory or subject contours provided by some *a priori* knowledge, belief, perceptual spatial grouping strategy or spatial pattern matching mechanism available in addition to sensory data. Gestalt psychologists use this illusion to describe the law of closure, one of the gestalt laws of perceptual organization. According to this principle, objects that are grouped together tend to be seen as being part of a whole. We tend to ignore gaps and perceive the contour lines in order to make the image appear as a cohesive whole.

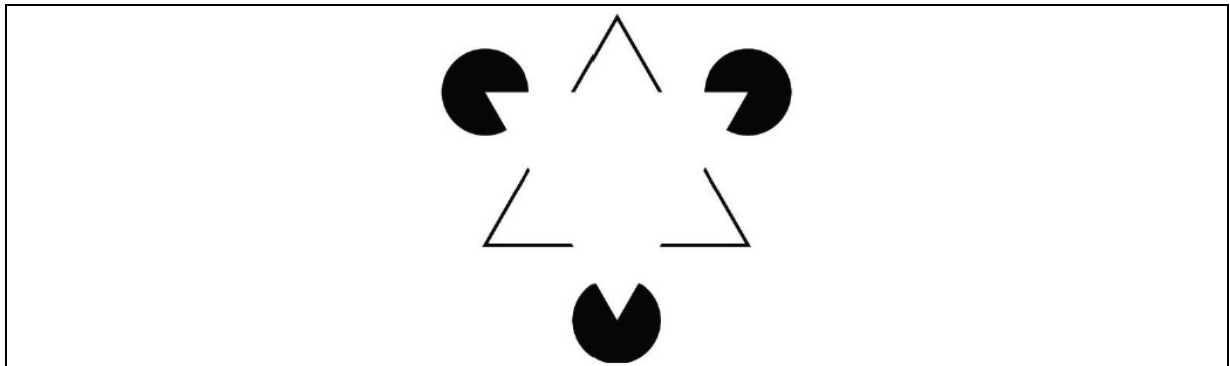


Fig. 3.1-7. Visual perception is a hybrid inference process, combining bottom-up sensory data with prior knowledge available in addition to data. This is proved by visual illusions. In the popular Kanizsa Triangle illusion, a white equilateral triangle can be seen in the image even though there is not actually a triangle there.

To date the human visual system can still be seen as a huge puzzle with a lot of missing pieces. Even in the first processing layers in the primary visual cortex (PVC, or area V1 of the visual cortex) there remain many gaps, despite all knowledge already compiled. Nevertheless, some of the gaps are being filled by developing and studying computational models of human vision. Models of simple, complex and end-stopped cells have been developed more than 10 years ago [28], [41], [43], [66]. It is an undisputable fact that the brain’s organizing principle is topology-preserving feature mapping and that in the visual system these topology-preserving feature maps [69], [145], [178], [179], [180], [187] are primarily spatial [94]. In [68], Tsotsos provides the following definitions.

- (1) A retinotopic representation of a visual feature (feature map) is a stimulus 2D array (2D regular gridded data set) with  $P$  elements. In this retinotopic representation, physically adjacent elements represent spatially adjacent regions in the visual field. Hence, retinotopic representation of a visual feature and topology-preserving feature mapping are synonyms [69], [145], [178], [179], [180], [187].
- (2) A map is defined as a retinotopic representation of only one type of visual parameter (visual feature).

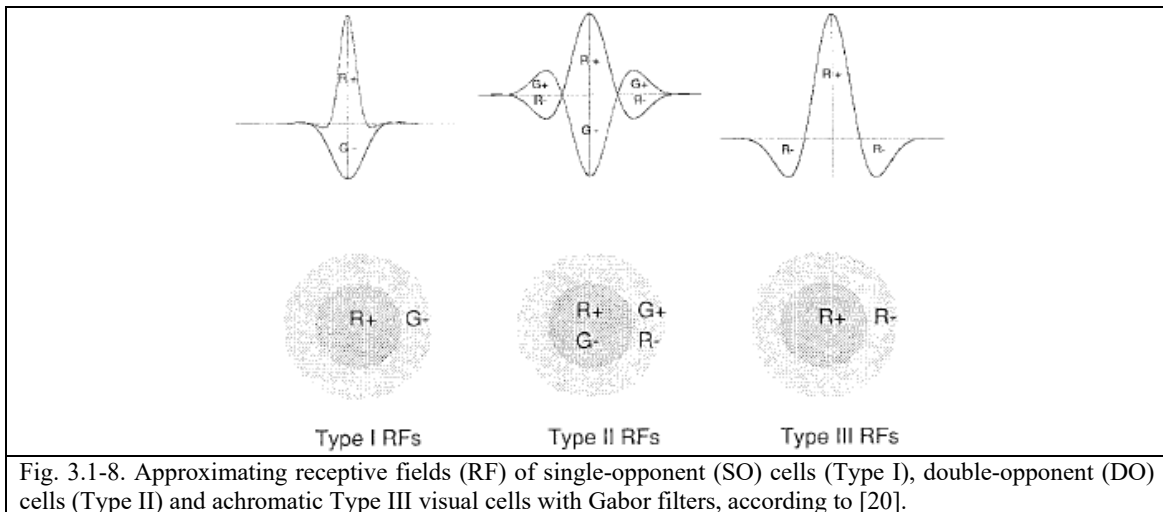
There may be 30 visual areas or so in primates, e.g., V1, V2, V3, MT and V4, with various degree of retinotopy [68]. Because many visual areas have more than one population of neurons, there are more logical maps than physical ones. The areas commonly accepted as being retinotopic include V1, V2, V3, MT, and V4, whereas the nonretinotopic ones include IT, posterior parietal cortex, and the frontal eye fields. According to [68], the division between retinotopic and nonretinotopic areas, although fuzzy in general, may be placed after areas MT and V4 and before IT, area 7, and the frontal eye fields. Maps seem to be organized hierarchically, as a partial ordering, so that the greater the distance from the retina, the smaller the maps are, and the larger the receptive fields of their neurons.

There is also more than one pathway from the retina to higher levels of processing, e.g., magno (M-) and parvo (P)-subsystems, also called the “what” and “where” systems or pathways [24]. Because at any level of the map hierarchy there are no more than a handful or so of maps, the number of maps at the output of early vision is on the order of a handful, e.g., in [68] the number of visual parameters is 8.

Another unquestionable observation (true-fact) is that human vision is a feedback system [8], [24], [82], [83], [93], [96], [126], [142]. Perception involves a complex interaction between feedforward sensory-driven information and feedback attentional, memory, and executive processes that modulate such feedforward processing. For example, visual mental imagery is known to induce retinotopically organized activation of early visual areas via feedback connections, which is tantamount to saying that mental images in the mind's eye can alter the way we see things in the retina [82], [83]. A feedforward system could, in principle, be used as the front-end of a visual system as part of a feedforward prediction - feedbackward verification loop. The feedforward path would provide an initial hypothesis about what object is presented in the visual field, yet to be verified through feedback loops. For example, Gestalt-like properties, such as continuity, symmetry and parallelism, are likely to involve lateral and feedback connections, yet to be inserted in traditional CV models [126].

Hence, vision must consist of a *hybrid* inference system provided with feedback loops [77], [126].

Another true-fact (observation) based on our everyday cognitive experience occurring when humans wear sunglasses is that human achromatic vision is nearly as affective as chromatic vision in scene-from-image representation. It means that spatial information, either topological (e.g., adjacency, inclusion, etc.) or non-topological (e.g., spatial distance, angle measure), typically dominates color information both in the 4D spatiotemporal scene-domain and in the (2D) image-domain [45]. This evidence forms the very foundation of the object-based image analysis (OBIA) paradigm [48], [87], proposed as a viable alternative to pixel-based image analysis insensitive to spatial information, either topological or non-topological [45], [80], [167].

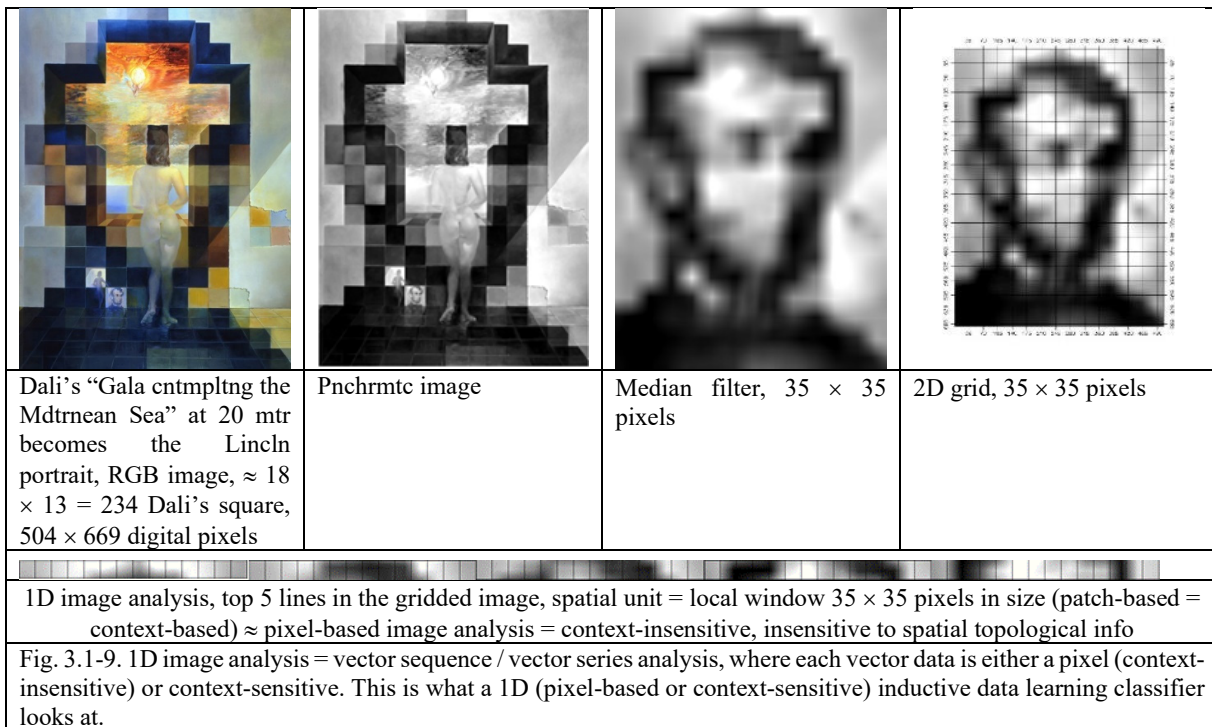


Color/gray-value is the sole visual property available at the pixel resolution. Visual features (numeric variables) in the image domain are known to be color, local shape of image-objects, texture (spatial perceptual grouping of texels) and inter-object spatial relationships, either topological or non-topological [45], [80], [167]. All visual features but pixel-based color are spatial properties. The elementary spatial unit of information, a picture element (pixel), explicitly represented in any



image is purely *locational* [52]. In itself, however, such locational information is of little visual relevance. If an image is investigated in terms of *locational* information exclusively, i.e., if an image is examined *pixel-based (context-insensitive)*, then 2D pixel values are converted into a 1D string (stream, sequence) of concatenated pixel values, where spatial topological information is fully lost. Sensitivity to any sort of visual structure or pattern requires extraction of *relational* image information, also called *structural* image information [53], [54], i.e., *context-sensitive information* about the spatial distribution of colors occurring jointly throughout the 2D (planar) image domain [52].

Any 1D image analysis algorithm is an orderless encoder invariant to permutations in the input vector sequence [120], where spatial topological information in the (2D) image domain is lost. For example, in 1D image analysis (classification) where vector data in the 1D vector data stream is context-sensitive, e.g., each vector data is the convolutional value between the image and a 2D spatial filter, then spatial topological information is lost. Pixel-based image analysis is a special case of 1D image analysis where spatial information in the image domain, either topological or non-topological, is totally lost. In contrast with traditional 1D image interpretation approaches widely adopted by the remote sensing (RS) community, combined exploitation of local spatial filters with image topology-preserving mapping functions explains the increasing popularity of deep convolutional neural networks (DCNNs) in computer vision applications [95], [96], [114], [115], [116], [117], [118], [119], [120], [127], [128], [184].



Noteworthy, DCNNs whose spatial filter weights are inductively learned-from-supervised data end-to-end by error backpropagation require as input mean-subtracted or median-subtracted input images [127], [128]. For example, adding a constant to the input image can have a dramatic effect on the performance of the optimisation algorithm, e.g., it can fail to reach convergence. This is due to a fairly general principle: centring the data usually makes learning problems much better conditioned. In practice, mean-subtracted input images feature a zero direct current (DC)-component, i.e., they are pre-processed to be independent of their average color/intensity value.

The same consideration holds for simple cells in the primate visual cortex, which are zero DC-component spatial filters, see Fig. 3.1-8, i.e., their response is zero for any constant signal, irrespective of its intensity, whereas they are sensitive to specific spatial changes in local intensity, e.g., change in sign of local concavity [1].

The conclusion is that spatial filters employed by the primate visual cortex and in artificial DCNNs are indeed sensitive to spatial information as local changes in the 2D signal while they are insensitive to the local mean color / intensity / amplitude

information, which agrees with the observation that, since chromatic and achromatic primate visions are nearly as effective in scene-from-image representation, then it must hold true that spatial information dominates color information in the 2D image-domain and in the 4D spatiotemporal scene-domain.

Whereas human vision is extremely efficient in the interpretation of a (2D) image, the human capacity of interpreting an image converted onto a 1D string (stream) of numbers is almost null, see Fig. 3.1-9. On the contrary, 1D image analysis, including pixel-based image analysis as a special case, accomplished by traditional computer vision (CV) systems consisting of a feedforward image pre-processing zero stage (optional) followed by an inductive supervised data learning classification first stage, see Fig. 3.1-10, has provided remarkable results, especially if compared to the human capability of interpreting a 1D string of values, see Fig. 3.1-9.

The same consideration holds when a 1D sequence of image-objects is visually interpreted by humans in comparison with 1D context-sensitive inductive feedforward CV systems, including OBIA approaches [87], where a 1D sequence of image-objects is input to a statistical vector-based classifier, see Fig. 3.1-11 and Fig. 3.1-12. In recent years, inductive feedforward CV systems adopting the OBIA paradigm [87], consisting of an image pre-processing (optional) zero stage followed by an inductive unsupervised data learning image segmentation first stage whose output image-objects are input to an inductive supervised data learning classification second stage, see Fig. 3.1-12, have provided remarkable results, especially if compared to the human capability of interpreting a 1D string of image-objects, see Fig. 3.1-11.

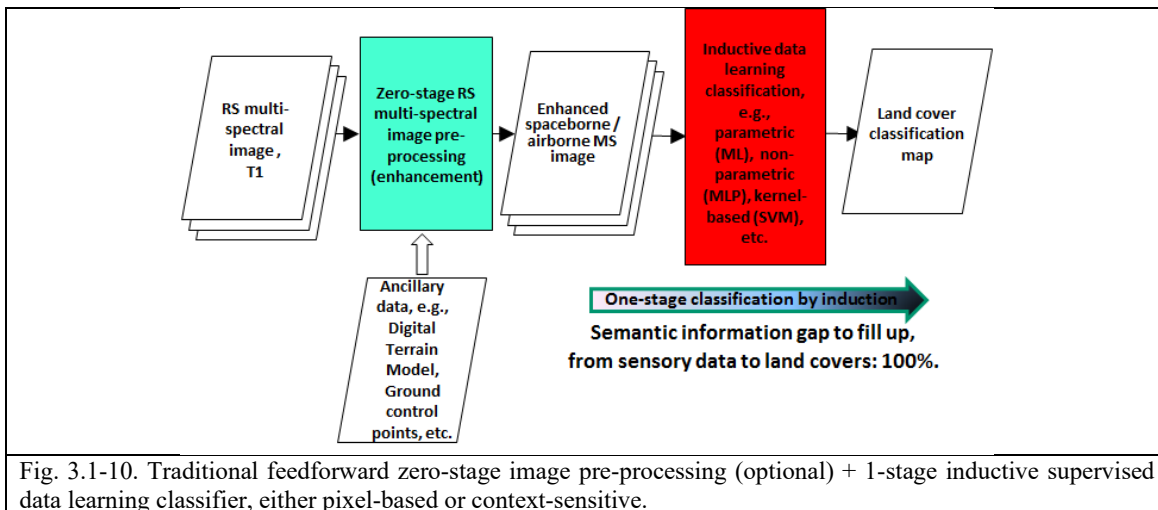


Fig. 3.1-10. Traditional feedforward zero-stage image pre-processing (optional) + 1-stage inductive supervised data learning classifier, either pixel-based or context-sensitive.

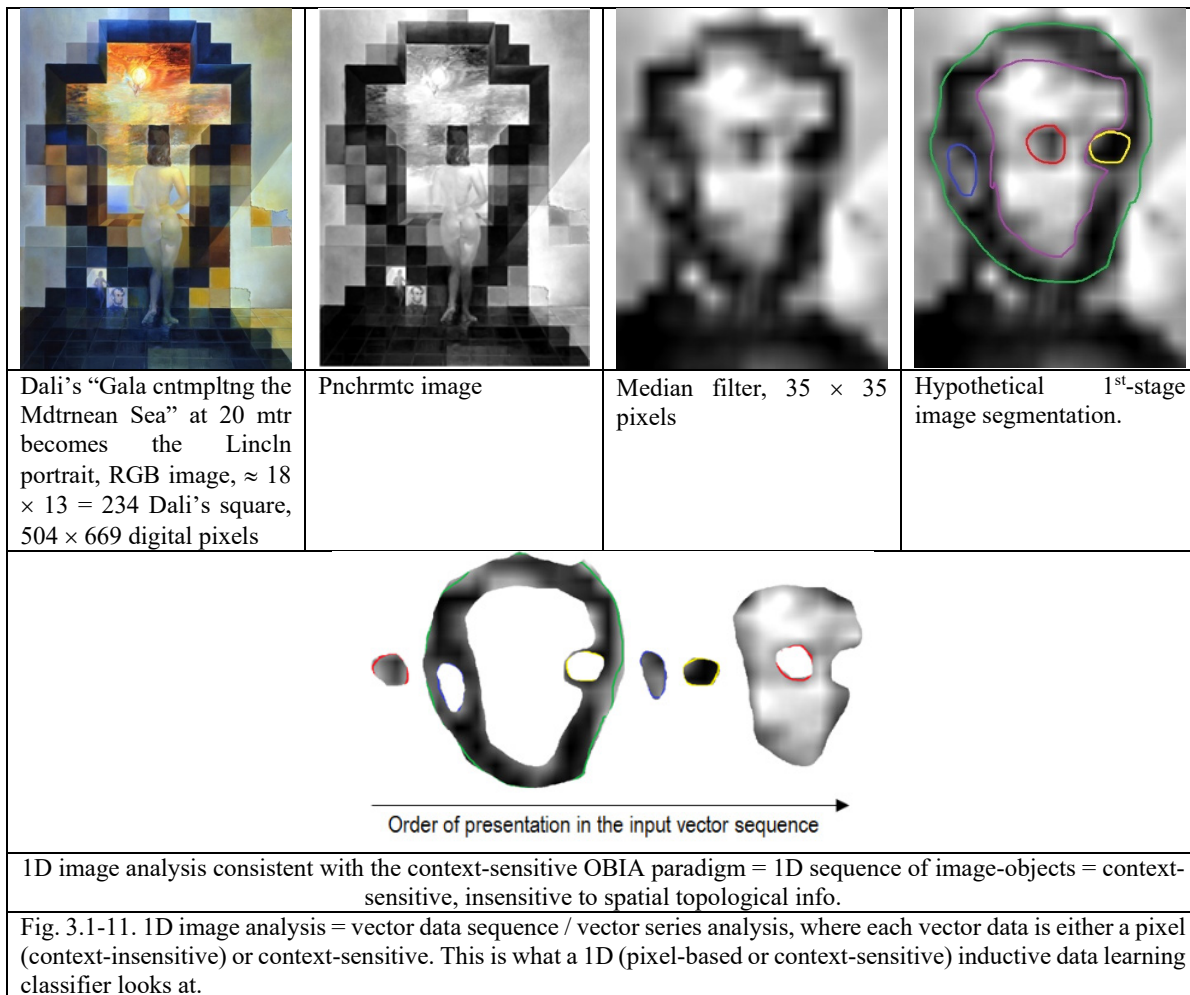
It is well known, but rarely acknowledged by the RS and OBIA communities, that in low-level vision the following observations hold true simultaneously. First, image segmentation is inherently ill-posed in the Hadamard sense [8], [26], [44], [45], [46], [47], [48], [68], [71]; second, image-contour detection is inherently ill-posed too [44]. In fact, image segmentation is the dual problem of image-contour detection where image-objects consist of closed contours [5]. It is an indisputable fact that "the computer vision community has tried to solve the low-level vision problem of image segmentation for decades without a satisfactory solution. The crux of the problem is the dilemma that to segment the image area for an object it helps to detect/recognize it first, while recognizing this object requires segmenting its image area first... Such a problem stems ultimately from the dilemma that segmentation requires classification and classification requires segmentation" [6].

It means that the OBIA system architecture shown in Fig. 3.1-10, where an inherently ill-posed inductive learning-from-data algorithm for image segmentation is adopted as first stage, is semi-automatic and site-specific [77].

The conclusion is that since 1D context-sensitive or context-insensitive inductive feedforward CV systems, such as those shown in Fig. 3.1-10 and Fig. 3.1-12, outperform human vision in 1D data stream analysis whereas they are typically outperformed, in terms of both low-level visual feature extraction and high-level visual feature interpretation, by the 2D image analysis approach adopted by human vision (retinotopic/topology-preserving representation in the human visual



system), then they are biologically implausible. In 2D image analysis, spatial topological information (e.g., adjacency, inclusion) and spatial non-topological information (e.g., spatial distance, angle measure) are thoroughly investigated because spatial information typically dominates color information in both the scene domain and the image domain [45], [80], [167]. Since color information is the sole visual feature available at the pixel resolution, it should be considered context-insensitive in nature. Since spatial information dominates color information in vision, chromatic vision systems pursuing 2D image analysis/retinotopic/topology-preserving visual feature representation are expected to down-scale to achromatic (panchromatic) vision problems with near lossless image understanding accuracy. Vice versa, if a chromatic vision system does not down-scale to achromatic vision successfully, then it tends to ignore the paramount spatial information in favor of subordinate (secondary) context-insensitive color information, such as class-specific spectral signatures typically investigated in pixel-based single-date or multi-temporal Earth observation (EO) image understanding systems (EO-IUSs).



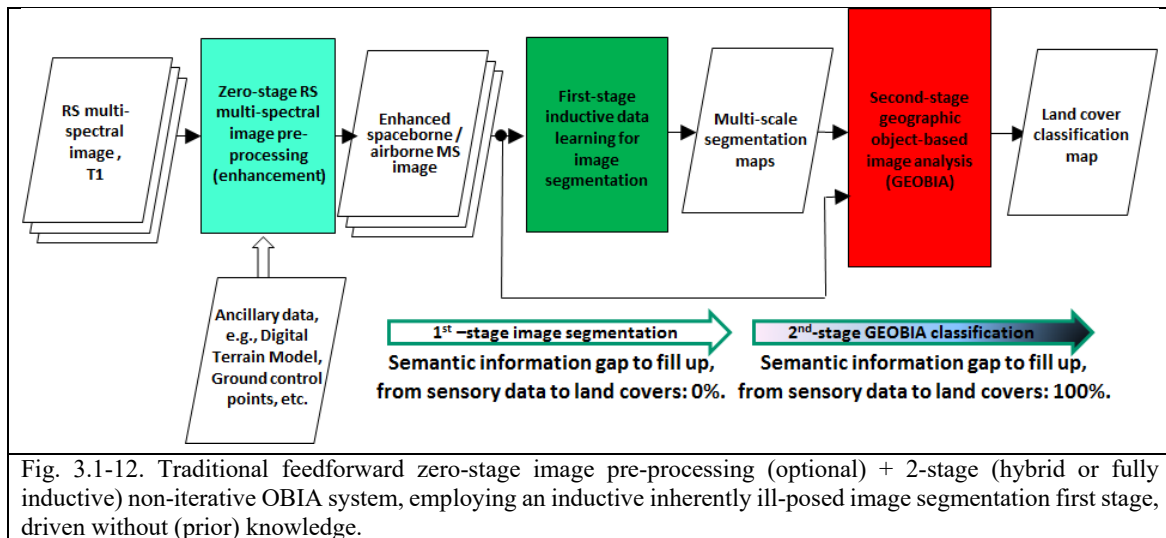
As a direct consequence of the aforementioned considerations, there is a necessary and sufficient experimental proof of system scalability required to prove that a CV system thoroughly investigates spatial topological and spatial non-topological information, which typically dominate color information in vision tasks.

*If a chromatic CV system does not down-scale seamlessly to achromatic image analysis, then it tends to ignore the paramount spatial topological and non-topological information components in favor of subordinate (secondary) context-insensitive (pixel-based) color information in the image-domain.*



Unfortunately, most CV algorithms have no direct biological background, and those with a clear biological background are often limited to one spatial scale [43], [66]. On the one hand, in computer vision, “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception. . . This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [49].

On the other hand, the human visual system can still be seen as a huge puzzle with a lot of missing pieces. Even in the first processing layers in the primary visual cortex (PVC, or area V1 of the visual cortex) there remain many gaps, despite all knowledge already compiled. Nevertheless, some of the gaps are being filled by developing and studying computational models of human vision. Models of simple, complex and end-stopped cells have been developed more than 10 years ago [28], [41], [43], [66].



### 3.2 Definitions of interest: The two information theories at the foundation of information technology (IT) - *Information-as-thing* and *Information-as-data-interpretation*

The transformation of data into information and knowledge is at the foundation of IT. Unfortunately, there is an inherent ambiguity of meanings in the single noun “information” adopted by acronym IT. As clearly pointed out by philosophical hermeneutics [31], [32], there are two complementary not alternative concepts of information, refer to Chapter 3.1.

The popular concept of *information-as-thing* is quantitative and unequivocal (“easy”); it is related to the Shannon’s mathematical theory of data communication, equivalent to data coding/ transmission/ decoding through a communication channel, where numeric or categorical variables, such as bits, documents, etc., are transmitted through a communication channel independent of their meaning, to be correctly received by a message receiver [125].

The concept of *information-as-data-interpretation* is qualitative and inherently equivocal (difficult, ill-posed). It pertains to the domain of cognitive science (see Fig. 3.1-2): since there is no semantics in quantitative/ sensory data, the data receiver has a pro-active role in providing the transmitted data with a meaning.

Vision (see Fig. 3.1-3) is an inherently qualitative and equivocal cognitive process (see Fig. 3.1-2) pertaining to the inherently difficult (ill-posed) domain of *information-as-data-interpretation*.



### 3.3 Definitions of interest: Proposed Minimally Dependent and Maximally Informative (mDMI) set of process and outcome quantitative quality indexes (Q<sup>2</sup>Is) in Big Data analytics

According to the Quality Assurance Framework for Earth Observation (QA4EO) guidelines [111] proposed by the intergovernmental Group on Earth Observations (GEO), an Earth observation (EO) image understanding system (IUS) is considered in operating mode if it systematically transforms multi-source EO “big data” [98], [135], [136], [177] into timely, comprehensive and operational information products, such as quantitative/numeric variables, e.g., enhanced EO images of augmented geometric and radiometric quality, or qualitative/nominal/categorical variables, e.g., thematic maps, at local to global geographic extents and spatial resolutions ranging from coarse (> 1 km) to very high (< 1 m).

“Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis and visualization... What is considered “big data” varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data set in its domain... “Big data” has increased the demand of information management specialists in software firms only specializing in data management and analytics... Big data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times. Real or near-real time information delivery is one of the defining characteristics of big data analytics. Latency is therefore avoided whenever and wherever possible” [135].

The lesson to be learned from big data analytics requirements is summarized as follows.

There are potentially many algorithms whose outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is) score “high” in “small data” analytics, equivalent to toy problems. Hence, there are many algorithms eligible for consideration in operating mode in “small data” analytics. Among these only few algorithms feature OP-Q<sup>2</sup>Is that score “high” when scaled to “big data” analytics. Similarly, if an algorithm’s OP-Q<sup>2</sup>Is score “high” in “big data” analytics (e.g., pixel-based classification of massive image time-series), any of these OP-Q<sup>2</sup>Is can score “low” in “small data” analytics (e.g., pixel-based classification of single-date imagery), see Fig. 3.3-1.

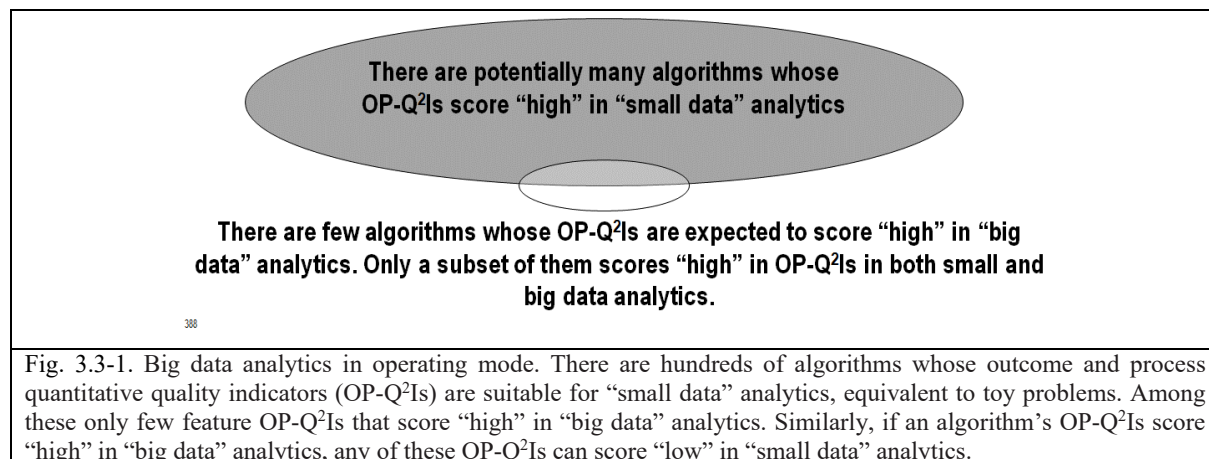


Fig. 3.3-1. Big data analytics in operating mode. There are hundreds of algorithms whose outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is) are suitable for “small data” analytics, equivalent to toy problems. Among these only few feature OP-Q<sup>2</sup>Is that score “high” in “big data” analytics. Similarly, if an algorithm’s OP-Q<sup>2</sup>Is score “high” in “big data” analytics, any of these OP-Q<sup>2</sup>Is can score “low” in “small data” analytics.

By definition, “big data” are characterized by four V’s [172].

- (i) (Big) volume. For example, according to the National Aeronautics and Space Administration (NASA)’s Earth Observing System Data and Information System (EOSDIS) metrics for 2014, the EOSDIS manages more than 9 petabytes (PB) of data [181]. EO sensory data ever archived by ESA since 1975 range from 3 to 10 PB [182]. Similar considerations hold for the German Aerospace Center (DLR), whose EO big data repositories encompass spaceborne imaging sensors both optical and synthetic aperture radar (SAR) [183]. The rate of increase in size through time of EO big data archives shows there is a kind of Moore’s law of technological growth and productivity applicable to imaging sensors submitted to the first information theory, quantitative *information-as-thing*, inherently “easy” because unequivocal: acquisition and transmission of sensory data is independent of the meaning of the transmitted



message [31]. To cope with big data volumes, an information processing system is required to be robust to changes in the input data set.

- (ii) (Large) variety (different sources of heterogeneous data). To cope with heterogeneous data sources, one viable strategy is to require sensory data radiometric calibration; it provides dimensionless data with an unequivocal physical unit of measure, which guarantees harmonization of data acquired across time, space and sensors.
- (iii) (High) velocity (on-line real-time analysis of data streams). “Real or near-real time information delivery is one of the defining characteristics of big data analytics. Latency is therefore avoided whenever and wherever possible” [135].
- (iv) (Unknown) veracity (uncertainty/reliability of data). In general, preliminary to data fusion, uncertainty/reliability of multi-source data must be accounted for, e.g., in compliance with a convergence-of-evidence approach typical of human reasoning, where “weak” but independent sources of evidence allow to infer “strong” (reliable) conjectures [45].

About “big data” [135], the European Commission (EC) funded a 2-year-long Big Data Public Private Forum (<http://big-project.eu>) through their Seventh Framework Program (FP7) to engage companies, academics and other stakeholders in discussing big data issues. The project aimed to define a strategy in terms of research and innovation to guide supporting actions from the European Commission in the successful implementation of the Big Data economy. Outcomes of this project were used as input for Horizon 2020 ([http://ec.europa.eu/research/horizon2020/index\\_en.cfm?pg=h2020](http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=h2020)), the EC’s next framework program [136].

We propose the following definitions, which are adopted in the rest of this technical work.

Hereafter, “big data” is intended as synonym of *central limit theorem* (CLT). In probability theory, the CLT states that the sum or arithmetic mean of a sufficiently large number of independent and identically distributed (i.i.d.) random variables, whose variance is finite, will be approximately normally distributed. Due to the CLT, the global sum (average) distribution of multi-source i.i.d. “big data” tends to be a Gaussian function, where independent distributions of “local” data (like basis functions) become indistinguishable from the whole.

To be considered in operating mode, an information processing system input with big data is required to score “high” in a minimally dependent maximally informative (mDMI) set [123], [124] of outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is), to be community-agreed upon in advance in agreement with the QA4EO Validation (*Val*) guidelines [111]. The proposed mDMI set of OP-Q<sup>2</sup>Is includes the following [56], [57].

1. Outcome Q<sup>2</sup>Is (O-Q<sup>2</sup>Is).
  - (i) Effectiveness, provided with a degree of uncertainty in measurement  $\pm\delta$ . For example, for a thematic geospatial data map, (a) thematic Q<sup>2</sup>Is (TQ<sup>2</sup>Is)  $\pm\delta$  and (b) spatial Q<sup>2</sup>Is (SQ<sup>2</sup>Is)  $\pm\delta$  [137].
  - (ii) Timeliness, defined as the time interval between data acquisition and product generation. It increases with manpower and computing power.
  - (iii) Costs in manpower and computer power, including costs required to collect labeled data for supervised data learning, if any.
2. Process Q<sup>2</sup>Is (P-Q<sup>2</sup>Is).
  - (i) Degree of automation, inversely related to user-machine interaction. It is monotonically decreasing with the number of system’s free-parameters to be user-defined based on heuristics. If full automation is required, no user-machine interaction is allowed to provide either system’s free-parameters or training data.
  - (ii) Efficiency, in computation time and memory occupation.
  - (iii) Robustness to changes in input data.
  - (iv) Robustness to changes in input parameters, if any.
  - (v) Scalability to changes in sensor and user specifications.

### 3.4 Preliminary research and development (R&D) vision project objectives

Vision is an open problem to date. On the one hand, the human visual system can still be seen as a huge puzzle with a lot of missing pieces [28], [41], [43], [66]. On the other hand, existing CV solutions typically score low in operational terms



including degree of automation, efficiency, accuracy, robustness to changes in input data, robustness to changes in input parameters, scalability to changes in user's requirements and sensor specifications, timeliness and costs in manpower and computer power [55], [56].

In CV, “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception. . . This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [49].

In accordance with this observation, the objective of the present research and development (R&D) project is twofold.

- To provide a survey of the existing state-of-the-art in CV system architecture and implementation solutions, including computational models of human vision [28], [41], [43], [66].
- To provide a substantial body of original, ground-breaking work in the CV domain as part of the interdisciplinary cognitive science domain, see Fig. 3.1-2. To become better conditioned for numerical solution an inherently ill-posed CV system is constrained to comply with human visual perception, i.e., the proposed CV system design and implementation aim at becoming a computational model of human vision [28], [41], [43], [66], whose goal is not only to mimic the processing of visual information in primates, but to match human perception and human performance in complex visual tasks as a lower bound. Hence, the proposed novel CV system's design and implementation solutions are constrained as follows.
  - Consistent with the Mach bands illusion, see Fig. 3.4-1. In the words of Pessoa, “if we require that a computational vision model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published models is surprisingly small” [25]. There is one original lesson of fundamental importance to be learned for 2D signal processing purposes from the physical model-based Mach bands illusion. Largely ignored by the CV and RS communities, this original lesson on statistical 2D signal processing is proposed as follows.

*Image-contour detection is the dual problem of image segmentation and they are both inherently ill-posed problems [8], [26], [44], [45], [46], [47], [48], [68], [71] in the Hadamard sense [75] (refer to Chapter 3.1). According to human perception, one ramp-edge is perceived as a single “homogeneous” region, whose boundaries are localized where the ramp starts from a low-level plateau and ends at a high-level plateau, irrespective of the slope (gradient) of the ramp. Depending on the ramp's slope, the within-ramp variance and the within-ramp contrast (gradient) belong to range  $(0, +\infty)$ . Along a ramp, no image-contour is perceived by human vision, irrespective of the ramp's local contrast in range  $(0, +\infty)$ . Supported by this perceptual counter-example, the conclusion is straightforward: computed by a 2D spatial operator (e.g., a 2D spatial filter, a local window or image-object) in the (2D) image-domain, local variance, local contrast and local first-order derivative (gradient) are statistical features (data-derived numeric variables) NOT suitable to detect image-objects (segments), or vice versa image-contours, required to be perceptually “uniform” (“homogeneous”). In other words, local variance, local contrast and local first-order derivative (gradient) estimated in the (2D) image-domain are unsuitable low-level visual features if detected image-segments/image-contours are required to be consistent with human visual perception, including ramp-edge detection where ramp edges (ramp boundaries) must be located where a ramp meets a plateau irrespective of the ramp slope.*

This original observation is straightforward (obvious, directly inferred from the perceptual Mach bands visual illusion), but not trivial. In practice it goes against the large majority of published image segmentation algorithms, such as the popular image-region growing algorithms proposed by Baatz et al. [173] and by Espindola et al. [174], [175], adopted by respectively the well-known eCognition commercial software product [176] and the freeware SPRING software for RS image processing [175], as well as image-contour detection algorithms based on spatial filter banks, e.g., the popular Canny edge



detector [27], whose visual feature detection criterion is local variance/contrast thresholding, with numeric threshold  $\in (0, +\infty)$  to be user-defined based on heuristics. For example, in the eCognition commercial software product [176], image segmentation is implemented by the region growing algorithm by Baatz et al. [173], where within-segment variance is thresholded by a so-called “scale parameter” to be user-defined in range  $(0, +\infty)$  based on empirical criteria: if the within-region variance threshold gets looser, then image-regions grow larger (at a coarser spatial scale). The corollary is that the infamous “(spatial) scale parameter” in eCognition has nothing to do with spatial scale in the image-domain, but with within-segment variance thresholding, which in turn has nothing to do with perceptual detection of image boundaries, including ramp-edges. Noteworthy, local variance is monotonically increasing with local contrast, which is synonym of local spatial first-order derivative (gradient) in the (2-D) image-domain.

- Multi-source, in particular the same CV approach must be able to process:
  - Panchromatic and color images.

*A necessary not sufficient condition for a CV system to fully exploit spatial topological and spatial non-topological information components in addition to color is to perform nearly as well when input with panchromatic or color imagery [185], [186].*

- Multi-spectral (MS), Super-spectral (SS) and Hyper-spectral (HS) images, whose number of spectral channels  $N$  is  $\{2, 9\}$ ,  $\{10, 20\}$  and  $> 20$  respectively.
  - Optical images and bi-temporal Red-Green-Blue (RGB) synthetic aperture radar (SAR) images [168].
- In operating mode, where full automation is required, i.e., no user-machine interaction is allowed, while robustness to changes in the input dataset must be guaranteed to cope with “big data”.

At the low-level CV system stage, encompassing the raw and full primal sketch proposed by Marr [5], original contributions must cope with:

- Raw primal sketch for token extraction [5] (p. 73): Automatic stratified (driven-by-prior knowledge across image masks) multi-scale zero-crossing (ZX) pixel detection as image-contours, according to the Marr’s terminology [5] and in compliance with the Mach bands illusion [25], see Fig. 3.4-1.
- Raw primal sketch for token extraction [5] (p. 73): Automatic stratified image-object (closed contours [5]) extraction as ZX segments from ZX contour pixels, according to the Marr’s terminology [5] (p. 70).
- Raw primal sketch for token extraction [5] (p. 73): Automatic stratified keypoint detection, related to saliency map extraction [71] and foveated imaging with a fixation point [109], [110].

In the words of Poggio [110]: “The initial goal of the Center for Brains Minds and Machines (CBMM) project titled ‘The computational role of eccentricity dependent resolution in the retina: consequences for hierarchical models of object recognition’ is to understand the consequences for object recognition of foveated imaging: how it affects performance in different situations (e.g., controlled vs. cluttered scenes), the number of fixations required for learning and inference, and the effect of other architectural choices (e.g., local vs. global pooling over space or scale). Of longer term interest is the observation that nested signatures necessarily encode information about the hierarchical structure of scenes, and might therefore be elements of a shared representation for object recognition as well as higher level reasoning about a scene, in the direction of the CBMM challenge at the MIT”.

- Full primal sketch [5] (p. 91) / perceptual spatial grouping of texels [2], [3], [58]-[60]: Automatic stratified texture-boundary detection and texture segmentation into closed texture contours.

Well known in statistics, the principle of statistic stratification states that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” [122].



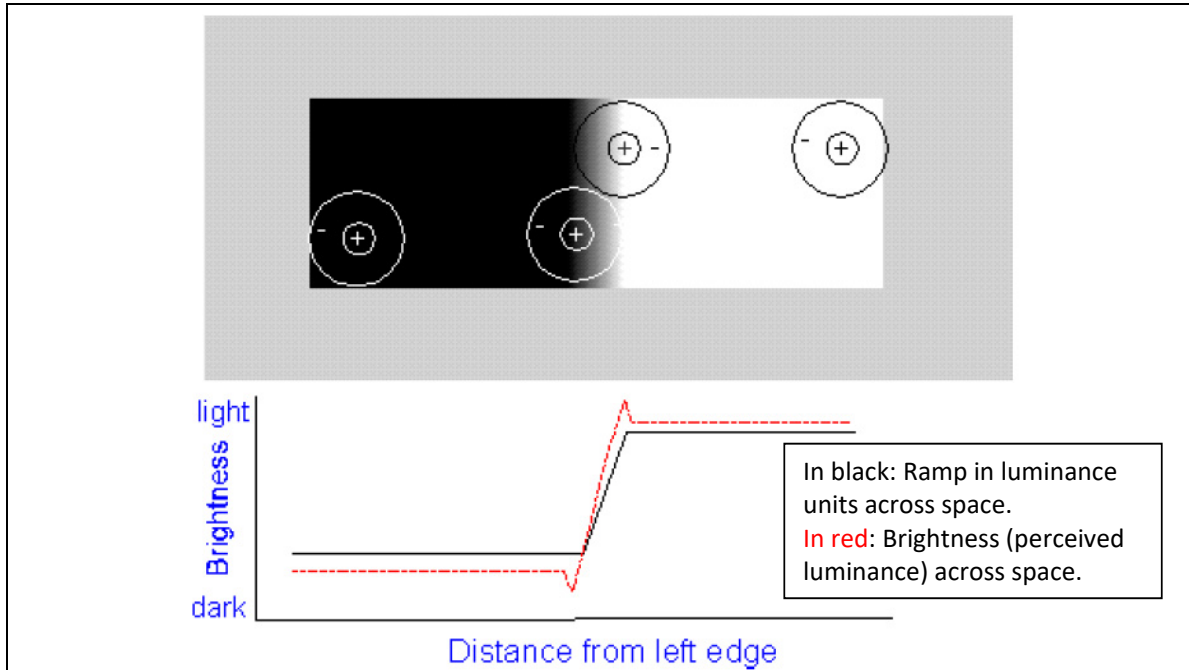


Fig. 3.4-1. Mach bands illusion. In black: Ramp in luminance units across space. In red: Brightness (perceived luminance) across space. One of the best-known brightness illusions (where brightness is defined as a subjective aspect of vision, i.e., brightness is the perceived luminance of a surface) is the psychophysical phenomenon of the Mach bands: where a luminance (radiance, intensity) ramp meets a plateau, there are spikes of brightness, although there is no discontinuity in the luminance profile. Hence, human vision detects two ramp boundaries, one at the beginning and one at the end of the ramp in luminance. Since there is no discontinuity in luminance where brightness is spiking, the Mach bands effect is called a visual “illusion”. Along a ramp, no image-contour is perceived by human vision, irrespective of the ramp’s local contrast (gradient) in range  $(0, +\infty)$ . In the words of Pessoa, “if we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published models is surprisingly small” [25]. In statistical 2D signal processing the lesson to be learned from the Mach bands illusion is that local variance, local contrast, and local first-order derivative (gradient) are statistical features (data-derived numeric variables) computed locally in the (2D) image-domain, e.g., within a moving window or within an image-object or by means of a 2D spatial filter with a finite support, NOT suitable to detect image-objects (segments, closed contours) required to be perceptually “uniform” (“homogeneous”). In other words, these popular local statistics in the (2D) image-domain are unsuitable if detected image-segments/image-contours are required to be consistent with human visual perception, including ramp-edge detection. This original observation can be considered straightforward (obvious), but not trivial. It is in contrast with a large portion of existing literature, where many image segmentation/image-contour extraction algorithms detect image-segments/image-contours by thresholding either a local estimate of the first-order derivative (gradient), for example a local gradient estimated by means of an odd-symmetric spatial filter such as the Canny edge detector [27], or by thresholding the local variance or local contrast estimated within a moving window or within an image-region, such as in image-region growing algorithms proposed by Baatz et al. [173] and by Espindola et al. [174], [175], adopted respectively by the well-known eCognition commercial software product [176] and the freeware SPRING software for RS image processing [175]. Inconsistent with the Mach bands illusion, these traditional edge detectors and image-region growing algorithms are semi-automatic because they depend on a system-free local variance, contrast or gradient threshold  $\in (0, +\infty)$  to be user-defined based on heuristics. For example, in the eCognition commercial software [175], the infamous spatial “scale parameter” to be user-defined is nothing else than a system-free within-segment variance threshold  $\in (0, +\infty)$ . If it is relaxed, then image-regions grow larger (at a coarser spatial scale). Their perceptual inconsistency with the Mach bands illusion explains why image-segments detected by semi-automatic image-region growing algorithms based on heuristics [173], [174], [175], [176], as well as image-contours detected by semi-automatic edge detectors based on heuristics [27], can be counter-intuitive, i.e., inconsistent with human vision to a varying degree depending on the combination of two random variables, the ever-varying complexity of an image as a peculiar combination of four spatial primitives, specifically flat area, step-edge, line and ramp-edge, with a user-defined local variance, contrast or gradient threshold value  $\in (0, +\infty)$ .





According to a well-known divide-and-conquer (*dividi-et-impera*) problem solving approach [79], coincident with the statistic stratification principle [122], a preliminary image stratification (masking) first stage can be adopted to make the inherently ill-posed (NP-hard [68], [71]) image vision problem (from sub-symbolic sensory 2D image to symbolic description(s) of the viewed 3D scene [44], [45]) better conditioned for numerical treatment by an attentional vision second stage. For example, a prior knowledge-based image stratification first stage can be implemented as a spectral prior knowledge-based Satellite Image Automatic Mapper™ (SIAM™) [38], [56], [57], [148], [149], [150], [151], [154], an expert system for continuous color space discretization (partitioning) into color names to be community-agreed upon in advance. In practice, in series with a first-stage prior knowledge-based pre-classifier for color naming (such as SIAM), employing per-pixel (context-insensitive) color properties to provide as output a color space discretization, equivalent to color naming [22], [36], [37], second-stage texture properties can be extracted from a panchromatic spaceborne/airborne optical image within a target image stratum (symbolic mask), generated as output by the expert system for color naming. For example, different types of forest stands can be discriminated and separated from grassland based on texture in a second-stage context-sensitive class-specific classification software module in series with the SIAM™ pre-attentive vision first stage which provides as output a semi-symbolic stratum (image mask) called ‘MS color green-as-vegetation’.

SIAM is a lightweight computer program capable of MS reflectance space hyperpolyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time (computation time increasingly linearly with the image size in pixels) [38], [56], [57], [148], [149], [150], [151], [154]. SIAM is eligible for use in a mobile application software, designed to run on web browsers and mobile devices, such as tablet computers and smartphones, eventually provided with a mobile user interface. According to the OGC terminology [55], a SIAM-derived color name-specific stratum is a semi-symbolic multi-part polygon, i.e., an image-wide collection of mutually disjoint color-labeled polygons (image-objects, 2D region, 2D segment) where each color-labeled polygon is a connected set of pixels featuring the same color name. Each color-labeled polygon can be considered a texel (texture element [2], [3], [58]-[60]) generated as output by the raw primal sketch, to be input to a texture detector at the full primal sketch [5].

To accomplish a prior knowledge-based partition of an either true- or false-color RGB cube into a static (non-adaptive to data) dictionary of color names, first, an automated statistical model-based algorithm for color constancy [112], [113], [144] was developed to comply with non-calibrated RGB images. Second an original RGB Image Automatic Mapper™ (RGBIAM™) lightweight computer program, capable of RGB cube polyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time, was proposed as a down-scaled version of SIAM [155], [188].

### 3.5 Vision problem background in neuroscience and computer science

This short section provides a vision problem background in neuroscience and in computational models of human vision.

#### 3.5.1 The human visual system

The goal of the human visual system, in combination with the other senses, is to recognize objects, to infer from 2D imagery a 3D spatial layout (one or more possible descriptions and understanding) of our 4D spatiotemporal environment (world-through-time), and to prepare for actions, see Fig. 3.1-3. All this is done automatically and very fast, e.g., refer to [143]. However, how all this is done is still a mystery. Despite the tremendous progress in research during the past decades, there still remain many open questions although our view of basic processes has become more clear [24], [41], [43], [66].

##### 3.5.1.1 Vision system architecture

In mammals, with special regard to primates, a vision system is comprised of a pre-attentive vision first phase and an attentive vision second phase summarized as follows.

- (i) Pre-attentive (low-level) vision extracts picture primitives based on general-purpose image processing criteria independent of the scene under analysis. It acts in parallel on the entire image as a rapid (<50 ms) scanning system



to detect variations in simple visual properties [13]–[15]. It is known that the human visual system employs at least four spatial scales of analysis [16], where cells in visual cortex feature gradations of orientation much finer than  $45^\circ$  [17], e.g., around  $15^\circ$  [18]. Single-opponent (SO) and double-opponent (DO) color cells have been called Type I and Type II respectively by Wiesel and Hubel [19] (examples of Type I and Type II receptive fields can be found in [20]). Receptive fields that are spatially opponent but not color opponent have been termed Type III [20], see Fig. 3.1-8.

According to Marr [5] (p. 73), low-level vision consists of two phases.

- (a) The raw primal sketch, which employs as input the zero-crossing (ZX) pixels and generates as output a discrete and finite set of multi-scale tokens (discrete sub-symbolic image plane entities), equivalent to texture elements (texels) [2], [3], [58], [59], [60].

According to Vecera and Farah: "we have demonstrated that image segmentation can be influenced by the familiarity of the shape being segmented", "these results are consistent with the hypothesis that image segmentation is an interactive (hybrid inference) process" "in which top-down knowledge partly guides lower level processing". "If an unambiguous, yet unfamiliar, shape is presented, top-down influences are unable to overcome powerful bottom-up cues. Some degree of ambiguity is required to overcome bottom-up cues in such situations. The main conclusion from these simulation studies is that while bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization" [8] (p. 1294).

According to Li Zhaoping, "the computer vision community has tried to solve the problem of image segmentation for decades without a satisfactory solution" [6]. In practice, automated generation of a raw primal sketch is still an open challenge in the CV and RS literature.

- (b) The full primal sketch [5] (p. 91, Figure 2-7 in p. 53), also called perceptual organization (p. 91) or perceptual grouping [6], to form larger-scale tokens that reflect larger-scale spatial structures (distributions) of elementary tokens in the image (p. 79), e.g., texture boundaries / texture segmentation [5] (p. 96). To investigate the spatial organization of tokens in an image, discontinuities (to be intended as synonyms of "abrupt changes" or "singularities") in six image parameters (quantitative properties, numerical features) must be investigated. Three of them are intrinsic to a token (token-specific) and three pertain to the spatial arrangement of tokens.

(A) Token-specific metrological attributes, affecting perceptual grouping of tokens.

(I) Average achromatic intensity (brightness, where brightness is defined as perceived luminance [1]) or (chromatic) color.

(II) Geometric properties, i.e., shape properties (e.g., length, width, compactness, rectangularity, roughness/straightness of boundaries, simple connectivity, etc.) and orientation. For example, refer to [29].

(III) Size.

(B) Spatial arrangement of tokens. If there is a discontinuity in any of the following attributes, then there is a change in perceived texture, i.e., there is a texture boundary.

(I) Local density of tokens.

(II) Distance apart of tokens. Estimated by the so-called Steven's algorithm for recovering the local orientation of tokens [30], based on an information primitive called, by Marr, inter-token virtual line (p. 82).

(III) Local orientation of tokens, also estimated by the Steven's inter-token virtual line detection algorithm [30].

Unfortunately, in his seminal work [5] Marr proposes no well-defined algorithm to generate the full primal sketch from the discrete token description table in a constructive (iterative, bottom-up) way. In practice, automated generation of a full primal sketch is still an open challenge in the CV and RS literature.



- (ii) Attentive (high-level) vision operates as a careful scanning system employing a focus of attention mechanism. A human or computational attention system computes and combines two attention principles capable of providing a retinotopic (topology-preserving) representation of what is interesting, i.e., salient, in the visual field: an *inductive (bottom-up) attention (saliency) mechanism*, which is derived solely from sensory data, and a *deductive (top-down) attention mechanism*, available *a priori* in addition to sensory data, such as pre-knowledge, expectations and current goals [71], [142]. Scene subsets, corresponding to a narrow aperture of attention, are observed in sequence and each step is examined quickly (20–80 ms) [13]–[19].

Hence, there is strong scientific evidence to consider both pre-attentive and attentive vision as hybrid (combined deductive/top-down and inductive/bottom-up) inference processes, refer to [8] and [71] respectively.

Noteworthy, there are about 40 types of biological vision systems in nature, e.g., foveated and non-foveated visual systems [72], [74]. Foveated imaging is a digital image processing technique in which the image resolution, or amount of detail, varies across the image according to one or more "fixation points." A fixation point indicates the highest resolution region of the image and corresponds to the center of the eye's retina, the fovea [73]. Unlike human foveated vision, capable of learning where to look sequentially in images, machine vision is currently more similar to the non-foveated visual system of a bee, whose eyes do not move but process every part of the image the same way [74].

### 3.5.1.2 The retina

The projected image on the retina is pre-processed there. Rods and cones, the basic photoreceptors, are connected by horizontal cells with excitative and inhibitory synapses, a first indication for spatial (or spatio-temporal) filtering. They are also connected to bipolar cells which connect to amacrine and ganglion cells. Already 12 types of bipolar cells have been identified, with at least 4 types of ON and OFF cone-connected cells. Cones play a role in daylight colour vision whereas rods are for black-white vision when the light level is low. ON and OFF refer to light increments and decrements on a background, for example white and black spots or bars on a grey background. Amacrine cells are inhibitory interneurons of ganglion cells, and as many as 50 morphological types exist. At least 10–15 types of retinal ganglion cells have been identified. These code ON and OFF signals for spatial, temporal, brightness and colour processing, and their outputs, the axons, connect to the lateral geniculate nucleus (LGN) and other brain areas (the LGN is a relay station between the retina and the visual cortex, input area V1) [24].

Most important here is that receptive fields of ON and OFF retinal ganglion cells can be seen as isotropic spatial bandpass filters, i.e. without a preferred orientation and therefore with a circularly-symmetric point spread function, often modelled by means of a "Mexican hat" function with a positive centre and a negative surround. Such filters only respond to transitions like dark-bright edges, and responses in homogeneous regions are zero or very small. The size of the receptive fields is a function of the retinal eccentricity: the fields are small in the centre (fovea) and they are increasingly bigger towards the periphery. Related to the field size is the notion of scale representation: at the point that we fixate fine-scale information is available, whereas the surround is blurred because only medium-and coarse-scale information is available there [24].

Also important is the fact that one very specific type of retinal ganglion cell is not connected, directly nor indirectly, to rods and cones; their own dendrites act as photoreceptors, they have very big receptive fields, and they connect to central brain areas for controlling the circadian clock (day-night rhythm) and, via a feedback loop, the eye's iris (pupil size). These special cells also connect to at least the ventral area of the LGN (LGNv); hence, in principle they can play a role in brightness perception, for obtaining a global background brightness on which lines and edges etc. are projected. This is still speculative and far from trivial, but we need to keep in mind that (a) pure bandpass filters, both retinal ganglion cells and cortical simple cells (see below), cannot convey a global (lowpass) background brightness level, (b) colour information is related to brightness and processed in the cytochrome-oxidase (CO, chromatic) blobs embedded in the cortical hypercolumns, colour being more related to homogeneous image (object) regions than to lines and edges extracted on the basis of simple cells etc. in the hypercolumns and not in the CO blobs, (c) colour constancy, an effect that leads to the same perception of object colours when the colour of the light source (illumination spectrum) changes, is intrinsically related to brightness, i.e. in a more global sense rather than object edges etc., and (d) very fine dot patterns, for example a random pattern composed of tiny black dots on a white kitchen table, are difficult to code with normal retinal ganglion cells or



cortical simple cells. Colour and dot-pattern processing suggest that there are more “pathways” from the retina to the visual cortex, although the availability of a cone-sampled image in the cortex is speculative [24].

### 3.5.1.3 The lateral geniculate nucleus

The traditional view of the lateral geniculate nucleus (LGN) is a passive relay station between the retina and V1, the cortical input layer that connects to higher areas V2, V4 etc. The more recent view is that the LGN plays an active role in visual attention: perhaps only 10% of its input stems from the retina and all other input it receives by means of feedback loops from inferior-temporal (IT) and prefrontal (PF) cortex, where short-term memory is thought to reside, via V4, V2 and V1. This implies that the magno (M-) and parvo (P)-subsystems, also called the “what” and “where” systems or pathways in ventral and dorsal areas throughout the visual cortex, already exist at LGN level: LGNv and LGNd [24]. The names what and where stem from the functionality of the system in testing hypotheses in the interpretation of the coded input information, i.e. what there is (object categorisation and recognition) and where it is (Focus-of-Attention and eye fixations). However, it should be stressed that the LGN is not involved in object recognition. Feedback from the visual cortex only modulates information passing through the LGN.

Understanding the ventral “what” and dorsal “where” parallel visual pathway organization of the mammalian brain and how these pathways provide integrated information to prefrontal cortical regions (via the temporal and parietal cortical regions) is a long-standing debate. Binding the general location information provided by the dorsal pathway with the appropriate detailed object-specific information provided by the ventral pathway is a significant challenge. In the context of engineering artificial neural systems for high-level vision problem solving [93], the dorsal “where” pathway (“Location” subsystem) is responsible for broad but low-resolution vision, identifying that there is an object at a particular location, but not the particular features of that object. The ventral “what” pathway (“Object” subsystem) provides a detailed but narrow view, thus being able to discern details of the object, but remaining ignorant of its location. The two visual pathways influence each other, with the Location module helping to guide the attention of the Object module, and the Object module providing detailed information about the visual field that the Location module lacks. This inter-pathway influence is similar in spirit to past more-biologically realistic models of interacting dorsal and ventral pathways in the brain that have been studied in isolation (i.e., not in the context of problem solving as we do here) in order to better understand the neurobiology of visual processing [93].

### 3.5.1.4 The visual cortex

The what and where pathways lead to V1 and via V2 and V4 to higher areas IT and PP (posterior-parietal). The bottom-up (visual input code) and top-down (expected object and position) data streams are necessary for obtaining size, rotation and translation invariance in object detection and recognition: object templates in memory are thought to represent a few canonical object views, probably normalised (if we close our eyes and imagine a few objects like a cup, a bottle, a cat and a house, one after the other, they all have more or less the same size). Invariance is obtained by dynamic routing in V2 and V4 etc., such that cells at higher levels (a) have bigger receptive fields until they cover the entire visual field, (b) perform more complex tasks, for example a face detector at a high level can combine outputs of eye and mouth detectors at a lower level, the eye and mouth detectors combining feature detectors at yet lower levels, and (c) can control attention and adapt/optimize local detection processes at the lower levels. A nice example of feature extraction is the multi-scale keypoint representation in V1 and beyond for face detection: the use of keypoints (singularities like line and edge crossings and end points) for detecting eyes etc. until a face is detected [24].

### 3.5.1.5 Visual information propagation

The human visual system is able to construct on the basis of a brief glance a complete scene-from-image representation in our brain: from global gist to spatial layout of objects [143], from local syntax of individual objects to multiple plausible semantic scene interpretations even provided with emotions [24].

(Start long quote of [24]) “Although we can detect and recognise objects very fast, almost instantaneously as it seems, processing in the different cortical areas and the information propagation, both bottom-up and top-down, take time. When seeing an image for a split second, we are able to extract the gist and detect specific objects. What happens is that the



flashed image enters the system and, after the computer screen goes blank again, the information propagates through the different levels (the same occurs between fixations, during saccadic eye movements when the image is not stable and the input is inhibited). Typically, objects are recognised within 150–200 ms, and first category-specific activation of PF cortex starts after about 100 ms. In addition, instead of all information propagating at the same time, or in parallel, it is known that coarse-scale information propagates faster than fine-scale information to IT cortex. This suggests that object segregation, categorisation and recognition are sequential but probably overlapping processes: the system starts with coarse scales for a first test to select possible object templates, then employs medium scales in order to refine the categorisation, until finest scales are available for final confirmation of the recognition result...Above, we did not address other issues, like motion and disparity. But some general questions remain: if things are quite complicated, with still many gaps in our knowledge, how is the image created that we perceive? Where in our brain is it created? Well, nobody knows exactly, but researchers who are developing computational models of human vision should have an idea” (End long quote of [24]). For example, in the words of Pessoa, “if we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, then the number of published models is surprisingly small” [25], see Fig. 3.4-1.

### 3.5.1.6 Visual feature extraction and brightness perception in human vision

The primary visual cortex (V1) is the input layer of the visual cortex in both left and right hemispheres of the brain. It is organised in so-called cortical hypercolumns, with neighbouring left-right regions which receive input—via the optic chiasm and one of the two LGNs—from the left and right eyes, with small “islands,” the CO blobs. In the hypercolumns there are simple, complex and end-stopped cells. Simple and complex cells are thought to serve line and edge extraction, whereas end-stopped cells respond to singularities (line/edge crossings, vertices, end points) [28], [43], [66].

There seems to be no sharp distinction between end-stopped and not end-stopped cell populations. Furthermore, apart from the sensitivity to length, end-stopped cells show the well-known characteristics of either simple or complex cells. All this suggests that end-stopping is an attribute added to the simple and complex types. Whether other receptive fields with 2D feature selectivity exist is still an open question [41].

For a complex cell (C-cell) operator, following Adelson and Bergen [42] and Burr and Morrone [26], we then combine the responses of S-operator pairs to a “local energy” representation defined by

$$C_i = \sqrt{O_{i,\text{odd}}^2 + O_{i,\text{even}}^2} \quad (5-1)$$

where  $O_{i,\text{odd}}$  and  $O_{i,\text{even}}$  are the convolutions of the image with the odd and even simple cell (S)-operators of orientation  $\theta_i$  ( $i=1, \dots, 6$ ) in the same scale band. We use the square root in order to obtain the same (linear) contrast response as for the S-operators. We call this the “C-operator” in analogy to complex receptive fields.

To recapitulate, according to [26], [42], the major difference between simple- and complex-cells is that the former are quasilinear while the latter exhibit a clear second-degree (squaring) nonlinearity. According to [126], complex cells tend to have larger receptive fields (twice as large as simple cells), respond to oriented bars or edges anywhere within their receptive fields (tolerance to position), and tend to be more broadly tuned than simple cells (tolerance to size). C-cell units pool over retinotopically organized afferent S-cell units from the previous layer with the same orientation and from the same scale band. In practice, C-cell units are implemented to pool their afferent S-cell units from the previous layer through a maximum (MAX) operation, thereby increasing invariance to position and scale [126].

According to [24], [28], [43], in principle, the end-stopped operators produce the first and second derivatives (with rectification) of the C-operator representations in the direction of their orientation. We define anti-symmetrical and symmetrical end-stopped operators and refer to these as single-stopped and double-stopped operators. For horizontal orientation, we define

$$E_s(x, y) = [C(x - d, y) - C(x + d, y)]^+ \quad (5-2)$$

for the single-stopped and





$$E_D(x, Y) = \{C(x, y) - \frac{1}{2} * [C(x - 2d, y) + C(x + 2d, y)]\}^+ \quad (5-3)$$

for the double-stopped operator, where  $C$  is the  $C$ -operator of horizontal orientation,  $d$  is a constant and  $\{\cdot\}^+$  denotes clipping of negative values. Thus each operator consists of two or three “subunits” of the same size and orientation preference, displaced along the preferred orientation by the distance  $2d$ . Other orientations were obtained by rotating these operators about the origin. Steps of  $30^\circ$  were implemented yielding 12 orientations of single- and 6 orientations of double-stopped operators (because the latter have  $180^\circ$  rotational symmetry). The separation constant  $d$  was chosen according to the “length” of the  $C$ -operator (which in turn depended on the space constant and width of angular function of the  $S$ -operators). It was chosen so that the response of the single-stopped operator to a line-end was three-quarters that of the  $C$ -operator to a line.

There are many cells tuned to different scales, i.e. with receptive fields which range from very small to very big. If we penetrate the surface of the cortex perpendicularly, we find cells tuned to different orientations. Many cells are also disparity-tuned, which indicates that stereo processing starts in V1, if not already in the LGN. It is likely that stereo processing involves simple cells with non-zero phase characteristics. V1 is composed of at least nine major layers, but the processing in those layers is not yet well understood. Apart from simple, complex and end-stopped cells there also are bar and grating cells. These are specialised for extracting aperiodic bars and periodic gratings. In contrast to simple and complex cells that respond to all patterns, bar and grating cells are highly nonlinear: a bar cell does not respond to bright or dark bars in a periodic grating and a grating cell does not respond to isolated bars. There also are cells that respond to illusory contours, e.g. gaps in edges, for example caused by occluding objects like tree branches in front of other branches. Without doubt, there remain cells with other specific functions that will be discovered in the near future.

The tuning of cells to different frequencies (scales), orientations and disparities, together with the existence of e.g. bar cells, points at a multi-scale image representation: lines, edges, keypoints, gratings etc. It is even possible that disparity is attributed to extracted lines and edges, i.e., in principle it is possible to construct a 3D “wireframe” model of objects, like the solid models used in computer graphics, but this is still speculative. However, it is likely that there are at least three (interconnected) data streams within the what and where streams.

(1) The multi-scale line/edge representation, accomplished via simple and complex cells, serves object segregation, categorisation and recognition, with coarse-to-fine-scale processing, the latter also being applied to disparity in order to solve the correspondence problem. We may assume that this stream is responsible for line/edge-related brightness perception (see below).

(2) The multi-scale keypoint representation, accomplished via end-stopped cells, serves as Focus-of-Attention (FoA), a process that directs our eyes—and mental attention—to points with a certain complexity: it does not make much sense to fixate points in homogeneous image regions where there are no structures to be analysed. In combination with motion and other cues, like colour contrast, this stream could be the basic cornerstone of the where stream. In [41], the “keypoints” of an image as the peaks (local maxima in a  $3 \times 3$  neighbourhood) in the summed end-stopped representation. The idea is that the peaks of this (scalar) representation provide an accurate localization of the relevant 2D features of the image, namely, terminations, comers and junctions, so that the further stages of processing can rely on the values of the end-stopped operators at these points. This is analogous to the role of the maxima (ridges) of the complex cell ( $C$ )-operator representation in the analysis of contrast borders. The knowledge of the keypoints of an image reduces the computational burden of the further analysis enormously because their number is generally low, even in complex images. Another important aspect of the use of keypoints concerns the evaluation of the different end-stopped operator signals. The pattern of activity across the different channels contains important information about the 2D signal variation. On end-points and corners, for example, we find stronger responses of single- than double-stopped operators, whereas on curved segments the reverse holds. The activity also varies characteristically across the orientations and directions. This information can be used to classify the local configuration. We found, however, that valid patterns of endstopped operator responses are generally obtained only at the singular points of comers, line-ends, etc. Therefore, these points have to be localised correctly first. The map of keypoints provides this localisation. The information represented at the keypoints complements the edge representation [41]. The edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of this map difficult. One can see that the representation of keypoints indicates





precisely these critical locations, like terminations, corners and junctions. Many of the keypoints are located on occluding contours. It is important to distinguish between 2D signal variations that correspond to intrinsic features of an object and those that are generated by occlusion. The map of keypoints with their adjunct information can be used in two ways to complement the edge representation: (1) The patterns of both single- and double-stopped operators can be used in an obvious fashion to link unconnected contour segments at corners and points of strong curvature (intrinsic keypoints). This would be indicated if a keypoint is located on the extrapolations of two neighbouring edges and the orientations represented at the keypoint match approximately the orientations of those edges. (2) The occlusion keypoints also aid the definition of contours. First, they indicate their location. Second, the single-stopped operator indicates the direction of termination, which can be used to recover the depth order [41].

(3) Colour and texture are surface properties of objects, normally in homogeneous regions but also with global modulations like shading due to light sources (shape-from-shading) and/or the shape of 3D objects (shape-from-texture) [5]. This shape information complements disparity information. Since lines and edges are 1D transitions (1D singularities; keypoints are 2D singularities) without colour, colour is supposed to be “sampled” and represented in the colour blobs.

This is an over-simplification, of course, because FoA in textured regions can direct attention for scrutinising detail, i.e., a conscious action that may complement an unconscious process, like automatic texture segregation, and global modulations (shape-from-X) can invoke different analyses. It is therefore important to stay focused on the main themes: basic processing serves (a) 2D image object structure, (b)  $2\frac{1}{2}$  surface structure, and (c) 3D scene structure.

Coming back to brightness processing, the computational model of vision in primates proposed in [24] was conceived from three rather simple—not trivial—observations which are not so easy to explain to non-specialists.

(Start long quote, [24])

(I) “Simple cells are often modelled by complex Gabor (wavelet) functions, or quadrature filters with a real cosine and an imaginary sine component, both with a Gaussian envelope, see Fig. 3.1-8. Such filters have a bandpass characteristic: the integral over the sine component is zero and the integral over the cosine component is very small or residual. Wavelets are also being used in image coding: the use of a complete set of bandpass filters tuned to all frequencies and orientations, plus one isotropic low-pass filter, which sum up to an all-pass filter (a linear filter that passes all frequency components), allows to reconstruct the input image. Therefore, in principle the brain could use the same strategy: sum the activities of all simple cells plus one “low-pass channel,” for example from the special retinal ganglion cells with photoreceptive dendritic fields, if available in the colour (CO) blobs, into a retinotopic projection map in some neural layer. However, this leads to a paradox: it would be necessary to construct “yet another observer” of this map in our brain. Therefore, we assume that brightness is related to the multi-scale line/edge representation which is necessary for object recognition.

(II) Basic line and edge detection involves simple cells in phase quadrature: positive and negative lines and edges (1D cross sections) can be detected and classified by combining detectors of ZX pixels and extrema (positive or negative) of the sine and cosine components, in combination with (positive) extrema of activities of complex cells. In [24], the computational model is based on simple and complex cells that are multi-scale, since many spatial patterns cannot be described using only one or few scales. However, there is one complication: at ramp edges, where a linear ramp meets a plateau, for example in trapezoidal bars or gratings, the system will detect positive and negative lines. Responses of filters in quadrature do not allow to distinguish between lines and ramp edges, which explains Mach bands at ramp edges, see Fig. 3.4-1.

(III) The implicit, multi-scale line and edge representation must provide information for brightness construction by means of an interpretation. In other words, instead of a re-construction the system builds a virtual impression on the basis of a learned interpretation of responding line and edge cells, perhaps much like a trained neural network. In [24] it is “simply” assumed that a responding line cell (at a certain position, tuned to a scale and orientation) is interpreted as having a Gaussian cross-profile there, with a certain amplitude (the response of the complex cell) and width (the scale of the underlying simple and complex cells). The same way responding edge cells are interpreted, but with a bipolar (positive-negative) cross-profile and modelled by a Gaussian-windowed error function.

This model provides a completely new way for image (re)construction, not like coding based on wavelets or simple cells. An additional observation is that there is a lot of neural noise in the system and we do not know whether there exist simple

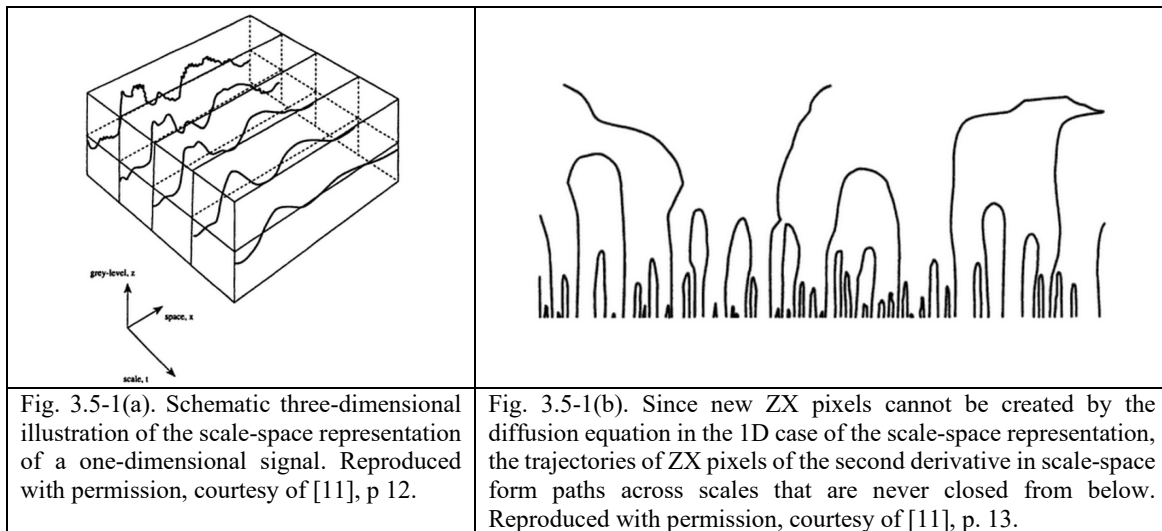


and complex cells etc. at all retinotopic positions and tuned to all scales and all orientations (representation noise and completeness). Stained maps of hypercolumns and dendritic/axonal fields of most if not all cells look rather random. Nevertheless, the image that we perceive looks rather stable and complete. It is very simple to simulate what happens when we suppress information, both in the brightness model as described above and in wavelet coding, the latter being modelled by considering the summation of responses of simple-cells. For example, we can suppress one entire scale channel, or 50% of all information by a random selection.

There is no precise region in our brain where the image that we perceive is created. Our model is limited to feature extractions in V1 and beyond, but this information must propagate to higher brain regions, eventually leading to consciousness, at the least being aware of our position in our actual environment. In other words, we may say that our perceived image, and therefore also at least part of our consciousness, are constructed by the entire brain, perhaps with an emphasis on the visual cortex. This is a holistic view.

Above we wrote that colour is represented in the CO blobs in V1, possibly in the form of sampled values that represent homogeneous object regions. However, recently it was found that many colour cells in V1 are orientation tuned [24]. This probably means that such oriented edge (contour) cells also contribute to colour perception and not only to achromatic brightness as exploited in our brightness model. In addition, contour processing may play an important role in colour constancy. We are far away from a unified framework. The same can be said about object categorisation and recognition. Change blindness, the fact that we do not notice things at positions where we are not looking, points at an interpretational filling-in process. If we do not perceive a specific object, we do not perceive that object's brightness and colour. In such a case our brain may be guessing what the most obvious solution might be, probably on the basis of prior experience with similar images".

(End long quote, [24])



### 3.5.2 Scale-space representation

According to [23], "a well-known property of the scale-space representation is that the amplitude of spatial derivatives in general decreases with scale, i.e., if a signal is subjected to scale-space smoothing, then the numerical values of spatial derivatives computed from the smoothed data can be expected to decrease. This is a direct consequence of the non-enhancement property of local extrema, which state that the value at a local maximum cannot increase and the value at a local minimum cannot decrease".

In addition, it is well known that "the number of ZX pixels in the second derivative of a scale-space representation (e.g., ZX pixels in a multi-scale DOG function) decreases monotonically with scale. This property holds for derivatives of



arbitrary order and also implies that the number of local extrema in any derivative of the signal cannot increase with scale" [11], p. 13). This phenomenon is shown in Fig. 3.5-1, taken from [11].

### 3.5.3 Accounting for ZX pixels through scale: the spatial coincident assumption of Marr

As reported in Chapter 3.2, according to Marr [5] (Figure 2-21, p. 73), the raw primal sketch consists of two steps. First, it detects (selects) *ZX pixels in the  $\nabla^2G$ -filtered image at multiple spatial scales, where a ZX pixel is located where the value of the  $\nabla^2G$ -filtered image passes from positive to negative* [5], (p. 54).

Second, it employs as input the ZX pixels to generate as output the following intermediate products.

- An intermediate information primitive called *ZX segment*, defined as "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" [5] (p. 60). Hence, it is at the level of detection of ZX segments that ZX pixels and contours turn into sub-symbolic discrete image-objects (polygons). According to Marr, ZX segments must be "accounted for" through scale in compliance with the *spatial coincident assumption* which is quoted below [5] (pp. 70, 71).
- A discrete and finite set of multi-scale *tokens* (discrete sub-symbolic image plane entities), namely ([5], caption of Figure 2-21, p. 73):
  - ✓ *edges*,
  - ✓ *blobs (closed contours*, [5], p. 61)
  - ✓ *bars*,
  - ✓ *discontinuities (terminations*, [5], p. 71, caption of Figure 2-22, p. 74), namely, corners, endpoints, T-junctions (e.g., due to occlusion) and X-junctions.
- A discrete and finite 1D list of tokens as data records in tabular form, equivalent to a worksheet/spreadsheet or segment description table (SDT, also refer to Nagao and Matsuyama in their seminal work [140]), where token attributes are: position, orientation, contrast (average intensity slope [5], p. 61), size, length, width, etc. ([5], caption of Figure 2-21, p. 73). Of course, the total number of table entries coincides with the total number of tokens detected in the raw primal sketch. Attention: an SDT is a 1D list in a tabular form [5]. This is one-to-one related to the vector data model, complementary not alternative to the raster data model in the (2D) image-domain. Implemented on a traditional paper sheet as 2D physical support, presentation, organization, analysis and storage of data in tabular form is traditionally called worksheet. "A spreadsheet is an interactive computer application for presentation, organization, analysis and storage of data in tabular form. Spreadsheets are developed as computerized simulations of paper accounting worksheets. In a spreadsheet, each cell may contain either numeric or text data, or the results of formulas that automatically calculate and display a value based on the contents of other cells. A spreadsheet or worksheet may also refer to one such electronic document" [157]. Noteworthy, a 1D list of data records is invariant to permutations, i.e., it is invariant to changes in the order of presentation of the list entries. Hence, a 1D list of data records cannot preserve inter-object spatial topological (e.g., adjacency) and spatial non-topological (e.g., Euclidean distance) relationships [45], [80], [167], when discrete and finite elementary objects belong to an (2D) image domain, equivalent to a 2D Euclidean space.
- A "2D literal bit map of the image" to represent positional information of tokens and allow efficient computation of inter-token spatial relationships at local scale. A "2D literal bit map of the image" is defined as a 2D array [5] (p. 79): (i) the size of the image, (ii) for each descriptive element (token) of the raw primal sketch, corresponding to one entry to the 1D list (worksheet, SDT) of token attributes, the literal bit map has a 1 value at the corresponding (x,y) coordinates / pixel positions, and (iii) each 1 in the bit map is also associated with a pointer to the element's actual description in tabular form. A bit map of the image to is one-to-one related to the raster data model, complementary not alternative to the vector data model equivalent to a tabular form. Noteworthy, in a "2D literal bit map of an image", spatial topological information (e.g., adjacency, inclusion) and non-topological information (e.g., spatial distance,



angle measure) of image-objects are preserved. It means that a “literal bit map” representation of an image is also a retinotopic [68], [94] / topology-preserving feature map [69], [145], [178], [179], [180], [187]. To investigate the spatial arrangement of (2D) tokens/planar objects in an image domain, the “literal bit map” is an efficient knowledge/information representation of tokens when inter-object spatial relationships are investigated at local rather than global (image-wide) scale. For example, to search a local spatial neighborhood of interest, “the literal bit map saves the trouble of searching through the whole 1D list of primal sketch descriptors checking each coordinate to see whether it falls within the specified neighborhood. The underlying reason why using a literal bit map representation of an image is more efficient for spatial searches than a 1D list of token descriptors is that most of the inter-token spatial relationships that must be examined in low-level vision are rather local. If we had to examine” image-wide (global) spatial properties and/or random, “arbitrary, scattered, salt-and-pepper-like spatial configurations, then a bit map would probably be no more efficient than a list” ([5], p. 80).

Unfortunately, in his seminal work, Marr proposes no algorithm to extract ZX pixels, ZX segments and tokens from ZX segments.

About ZX segments, Marr stresses they must be “accounted for” through scale in compliance with the *spatial coincident assumption*, which is quoted as follows [5] (pp. 70, 71).

“If a ZX segment” (note: this information primitive is not a ZX pixel, refer to the further Chapter 3.6) is present in a set of independent (multi-scale) channels (see [5], Figure 2-21, pp. 72-73) over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such ZX segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, illumination, surface orientation, etc.)”, i.e., it corresponds to a “true” image contours. “In other words, provided that the ZXs from independent (multi-scale, see [5], Figure 2-21, pp. 72-73) channels of adjacent sizes (scales) coincide, they can be taken together. If the ZXs do not coincide, they probably arise from distinct surfaces or physical phenomena. It follows that: (1) the minimum number of (multi-scale) channels required to establish physical reality (a physical boundary) is two and (2) if there is a range of channel sizes (scales) and orientations, rules must be derived for combining their ZXs into a description whose discrete primitives (namely, a discrete-token description table where each image primitive, called *token*, is either a discrete *blob* (closed contour), *edge*, *bar* or *termination*, described by token-specific features like: average intensity slope [5] (p. 61), position, size, contrast, orientation, length, width, etc., see Figure 2-21, p. 73 in [5], also refer to Chapter 3.2) are meaningful. The actual details of the rules are quite complicated because a number of special cases have to be taken into account, but the general idea is straightforward. Provided the ZXs in the larger channels are “accounted for” by what the smaller channels are seeing, either because they are in one-to-one correspondence with the ZXs in the smaller channels or because they are blurred, averaged copies of them, then all the evidence points to a physical reality that is roughly what the smaller channels are seeing, perhaps modified and smoothed a little by the noise-reducing, averaging effects of the large ones... If the larger channels' ZXs cannot be accounted for by what the smaller channels are seeing, then new descriptive elements, namely, discrete tokens, must be developed, because the larger channels are recording different physical phenomena... From the caption of Figure 2-21, pp. 72-73: because there are no ZXs in the larger channel that do not correspond to ZXs in the smaller channel, the locations of the edges in the combined description also correspond to that at the smallest spatial scale.”

Attention: no optimized ZX pixels spatial-coincident-through-scale rule set has been proposed explicitly or implemented in [5]. It means that the implementation of such a prior knowledge-based rule set (deductive inference system) is an open challenge to be faced by the present work.

### 3.5.4 Yellot’s Theory of Low-Level Vision for Texture Discrimination: The Triple Autocorrelation Uniqueness (TAU) Theorem

The long-disproved Julesz conjecture concerning texture discrimination in biological vision states that pre-attentive discrimination of textures is possible only for textures that have different 2<sup>nd</sup>-order autocorrelation statistics (univariate statistics of the 2<sup>nd</sup>-order in the spatial domain). Many counter-examples to this theorem have subsequently been discovered by Julesz and co-workers as well as by other independent researchers [2], [3], [59], [60]. In other words, it is possible to construct pairs of physically distinct texture images whose 2<sup>nd</sup>-order univariate statistics are exactly identical. This simple

background knowledge found in existing literature has an important practical consequence: it implies that popular 2<sup>nd</sup>-order spatial statistics, extracted from a gray-level co-occurrence matrix (GLCM) implemented in nearly all existing RS image processing software toolboxes, are inadequate for texture assessment and comparison purposes [62], [63]. Actually, in a more recent paper Yellott appeared to reintroduce the validity of 2<sup>nd</sup>-order spatial statistics, by proving that every discrete, finite image is uniquely determined by its two-dimensional dipole histogram [64].

In the context of more recent re-thinking on this subject, Julesz synthesized his studies of pre-attentive texture discrimination as follows: “In essence, we found that texture segmentation is not governed by global (statistical) rules, but rather depends on local, nonlinear features (textons).” As a consequence, “contrary to common belief, texture segmentation cannot be explained by differences in power spectra” (which are image-wide statistics, rather than local statistics). In other words, in biological vision, the neural computations are inherently local in the 2D spatial domain; next, a spatial average is superimposed on the local computational processes. For example, the overall amount of contrast is a visually salient feature which survives this averaging process, although the precise position of each contrast element does not survive the averaging process [3].

In a more recent paper than [3], Yellott stated the following [2].

- Given a discrete image (2D) array,  $I(c, r)$ ,  $c = 1, \dots, C$ ,  $r = 1, \dots, R$ , consisting of  $C$  columns and  $R$  rows, the discrete image-wide 1<sup>st</sup>-, 2<sup>nd</sup>-, and 3<sup>rd</sup>-order spatial statistics are defined respectively as:

$$a_{1,l} = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r), \quad (5-4)$$

$$a_{2,l}(n, m) = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r)I(c+n, r+m) = \frac{1}{C \cdot R} \cdot \text{AutocrltnFunctn}, \quad (5-5)$$

$$a_{3,l}(n_1, m_1, n_2, m_2) = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r)I(c+n_1, r+m_1)I(c+n_2, r+m_2) = \frac{1}{C \cdot R} \text{TripleAutocrltnFunctn} \quad (5-6)$$

where Eq. (5-5) is the so-called continuous autocorrelation function (up to a multiplicative factor), while Eq. (5-6) is known as the third-order continuous autocorrelation function (up to a multiplicative factor).

- In a black and white (binary) image of finite size, the image-wide third-order statistics are equivalent to the image-wide triple autocorrelation function, which is a generalization of the ordinary image-wide autocorrelation function.
- In a black and white (binary) image of finite size, the image-wide second-order statistics are equivalent to its image-wide autocorrelation function. For images with more than two gray levels, this equivalence breaks down, i.e., two images can have the same autocorrelation function, but different 2<sup>nd</sup>-order statistics.
- Discrimination between textured images of finite size becomes increasingly difficult as their image-wide third-order statistics become more similar.
- The Yellott’s Triple Autocorrelation Uniqueness (TAU) theorem states that every panchromatic (one-channel multi-gray leveled) image of finite size is uniquely determined (up to spatial translation) by its image-wide third-order statistics [2]. Let’s consider the two following statements.
  - ✓ Statement A: Two panchromatic images of finite size are visually identical (up to spatial translation).
  - ✓ Statement B: Two panchromatic images feature identical image-wide third-order statistics.

The TAU principle affirms that statement B is a necessary condition of statement A, i.e., statement A implies statement B. On the other hand, statement B is a sufficient condition of statement A, i.e., statement B implies statement A.

- Identical image-wide third-order statistics imply identical image-wide 2<sup>nd</sup>-order statistics.

In commenting Yellott’s work [2], Victor observes the following [3].

- The TAU theorem is computed image-wide, i.e., it applies to images of finite size, while the Julesz conjecture applies to textures conceived as a single infinite image or as an infinite ensemble of finite images (which relates to the property of ergodic textures, such that averages performed over the infinite ensemble of textures can be replaced by spatial averages over a single spatially infinite image extracted from the ensemble). Thus, the TAU theorem does not apply to texture ensembles, i.e., it does not trivialize the Julesz conjecture based on local, rather than global statistics. In practice, TAU, which refers to image-wide third-order statistics in images of finite size, does not hold true.
- Biological vision consists of a set of ill-posed problems, such as shape from shading, shape from texture, structure from motion, etc. [46]. Due to the inherent ill-posedness of the (3D) scene reconstruction from (2D) imagery, the visual system necessarily makes inferences from partial (incomplete) information, and the discovery of how these inferences are made is what the study of biological vision is all about [45].

By combining the TAU theorem with the inherently ill-posed problem of texture segmentation in pre-attentive vision whose



neural computations are inherently local [3], [13], [14], [15], [16], [20], a new version of the Julesz conjecture, hereafter referred to as the Enhanced TAU (ETAU) theorem, is formulated as follows.

*“Two images of either finite or infinite size are visually identical (up to spatial translation) if their local, non-linear, non-specific elements (textons) of texture perception (“tokens” in the Marr’s terminology [5], where tokens are detected in the raw primal sketch of early vision) have identical third-order spatial statistics; if this occurs, it means that two different textures (homogeneous spatial distributions of tokens, detected in the full primal sketch of early vision [5]) are the same texture.”*

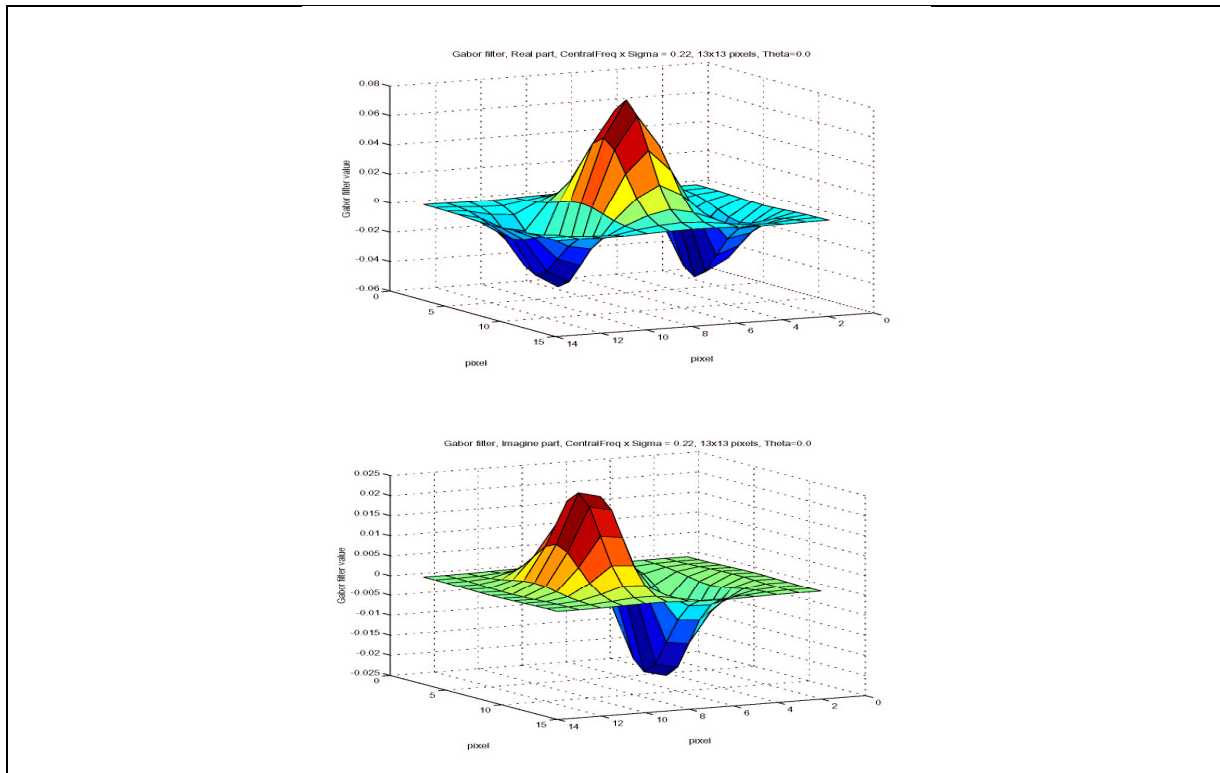


Fig. 3.5-2. Best-fitting 2-D Gabor elementary function for a simple cell of the cat visual cortex [1]. Multi-scale multi-orientation filter bank consisting of Gabor wavelets such that:

- A Gabor filter is a Gaussian function with spread  $\sigma$  modulated by a complex sinusoid with central frequency  $F$ . To approximate a simple cell of the cat visual cortex, a physical model-based relationship between these two parameters was found to be  $F \times \sigma = 0.22$  [1].
- The real part of an oriented Gabor mother-wavelet with  $F \times \sigma = 0.22$  can be considered equivalent to an even-symmetric 2<sup>nd</sup>-order derivative of a Gaussian function. According to [1], [5], this local filter is necessary and sufficient to detect any sort of image contours, namely, step edge, roof, line (ridge) and ramps (according to the Mach bands illusion [25], refer to Fig. 3.4-1).
- The imaginary part of an oriented Gabor mother-wavelet with  $F \times \sigma = 0.22$  can be considered equivalent to an odd-symmetric 1<sup>st</sup>-order derivative of a Gaussian function. According to [1], [5], it is not further employed in image contour detection.
- The real part of an oriented Gabor mother-wavelet provides a 3<sup>rd</sup>-order statistic in line with the works by Yellott [2] and Victor [3].
- The oriented Gabor mother-wavelet is designed with a zero DC-component (to be insensitive to ramps and constant offsets), in line with [1].
- To provide the best compromise between computation time and the quality of the image decomposition/synthesis the following filter bank design is selected.
  - Four dyadic spatial scales (one octave apart), with filter size in pixel units equals to  $3 = 6 \times \sigma$ , then  $\sigma = 0.5$ ,  $6 \times 2\sigma = 6 \times 1 \approx 7$ ,  $6 \times 4\sigma = 6 \times 2 \approx 13$ ,  $6 \times 8\sigma = 6 \times 4 \approx 25$  pixels, in agreement with [16].
  - Two spatial orientations: 0 and 90 degrees.

In this case the test image decomposition is accomplished in few tens of seconds (with a mildly optimized software).



In this latter statement, concepts like texture element/ texon/ token and texture, where texture is defined as the visual effect generated by a spatial distribution of tokens, i.e., texture detection is known as perceptual spatial grouping of texels [2], [3], [58]-[60], are necessarily vague (fuzzy), to account for the inherent ill-posedness of pre-attentive vision [5], [44]. Analogously, the same vagueness holds in the inherently ill-posed early-vision process of texture detection (texture segmentation), dealt with by the pre-attentive visual second stage, known as full primal sketch [5], [44].

A simple relationship between the aforementioned ETAU thesis and biological vision reinforces the former speculation. To date, the human visual system can be seen as a huge puzzle with a lot of missing pieces. Even in the first processing layers of the primary visual cortex (PVC, area V1 of the visual cortex, striate cortex) there remain many gaps, in spite of knowledge acquired by neuroscience [13], [14], [15], [16], [17], [18], [19], [65]. In part, these information gaps are being filled by developing and studying computational models. For example, models of simple, complex and end-stopped cells have been implemented in the last 10 years [28], [43], [66]. However, if we require that a computational model of vision should be able to predict perceptual effects, like the Mach bands illusion, where bright and dark bands are seen at ramp edges, then the number of published vision models becomes surprisingly small [25], see Fig. 3.4-1. In a rather schematic summary, V1 is the input layer of the visual cortex in both left and right hemispheres of the brain. It is organized in so-called cortical hypercolumns, with neighboring left-right regions which receive input—via the optic chiasm and the lateral geniculate nucleus (NGL)—from the left and right eyes, with small “islands,” called the “chromatic” blobs [13], [14], [65]. Traditionally, blobs are believed to consist of color-sensitive cells, called double-opponent cells, (apparently) non-oriented, but sensitive to colors [49]. More recent studies found that many color cells in V1 are also orientation tuned [24]. Differently from double-opponent cells in blob areas, most cells in the large interblob areas are (apparently) selective for orientation, but are not chromatic. In the interblob hypercolumns there are simple (S-)cells, complex (C-)cells and end-stopped cells. Complex cells are thought to receive convergent excitatory connections from several simple cells [13]. A major difference between S- and C-cells is that the former are quasilinear while the latter exhibit a clear second-order nonlinearity [26], [42]. There is general agreement that S- and C-cells serve for line and edge extraction, to accomplish object segregation, categorization and recognition [28], [43]. Unfortunately, there are tens of different computational models trying to explain how S- and C-cells interact for line and edge extraction, e.g., refer to [25], [26], [28], [42], [43], [66].

About end-stopping, there seems to be no sharp distinction between end-stopped and not end-stopped cell populations. Furthermore, end-stopped cells show the well-known characteristics of either simple or complex cells. All this suggests that end-stopping is an attribute added to the simple and complex types [28], [43]. End-stopped cells respond to singularities, like line/edge crossings, vertices and end points. The so-called multi-scale keypoint representation [86], accomplished via end-stopped cells, serves as Focus-of-Attention (FoA) [28], [43], [66]. The information represented at the keypoints complements the edge representation [41]. The edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of an edge map difficult. One can see that the representation of keypoints indicates precisely these critical locations, like terminations, corners and junctions. Typically, many of the keypoints are located on occluding contours [28], [41], [43], [66].

Since they have a proactive role in contour detection, where they are claimed to be sufficient for edge detection by zero-crossing [5], [62], [89], and since they provide inputs to both C-cells (whatever this cell type does) and end-stopped cells suitable for keypoint detection, S-cells are of key relevance in pre-attentive vision. Typically, they are modelled by complex Gabor (wavelet) functions, or quadrature filters with a real cosine and an imaginary sine component, both with a Gaussian envelope, see Fig. 3.5-2. If the even-symmetric (real) part of a Gabor local filter is implemented like a second-order derivative of an oriented Gaussian shape, like that shown in Fig. 3.5-2(a), then it is: (i) suitable for detecting image contours as zero-crossings of the even-symmetric filtered image, in agreement with the Marr’s theory of early vision [5], [62], [89], and (ii) eligible for collecting 3<sup>rd</sup>-order spatial statistics, like those envisaged by the Yellott’s ETAU principle [2].

To conclude, the ETAU speculation finds a physical justification in the multi-scale model of even-symmetric S-cells found in the interblob hypercolumns, as those shown in Fig. 3.5-2(a), in agreement with the Marr’s theory of early vision [5], [62]. Noteworthy, the odd-symmetric (imaginary) part of the same Gabor filter, which is equivalent to a first-order derivative of an oriented Gaussian shape, shown in Fig. 3.5-2(b), would be eligible for collecting 2<sup>nd</sup>-order spatial statistics, like those envisaged by the long-disproved Julesz conjecture about texture discrimination.

### 3.6 Original requirements specification for computational models of human vision



In computer vision, “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception. . . This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [49].

In accordance with this observation, the present R&D project aims at developing an innovative CV system whose goal is not only to mimic the processing of visual information in primates, but to match human performance in complex visual tasks.

A computational complexity-level analysis of the inherently ill-posed (difficult, NP-hard [71]) vision problem, such as that proposed by Tsotsos [68], aims at designing a novel generation of CV systems whose degree of biological plausibility should score high. In the words of Tsotsos, the following considerations hold [68] (pp. 423, 424).

“(Start long quote of Tsotsos, [68], pp. 423, 424). One of the key problems with artificial intelligence is that the solutions proposed are fragile with respect to the question of “scaling up” with problem size: Theoretical solutions are usually derived without regard to the amount of computation required and then if an implementation is produced, it is tried out only on a few small examples. The standard claim is that if faster or parallel hardware were available, a real-time solution would be obtained. There is something very unsatisfying about this type of claim. In particular, parallel solutions, such as those proposed by the connectionist community, although motivated by complexity considerations, typically fail to demonstrate the computational sufficiency of their approaches... If one is committed to realizing systems and proving that they behave in the required manner, the first prerequisite would seem to be that the candidate system be computationally tractable. The problem vision researchers from the many relevant disciplines face is that experiment results and explanatory theories from these disparate fields are not immediately compatible, and often appear contradictory. There has been very little work on «the big picture” which the individual results may fit... There is no test that can be applied to a theory to determine whether or not basic considerations are satisfied. Satisfying complexity constraints is one test that new theories of visual perception must pass... The computational complexity of the perceptual visual task are critical and lead directly to “hard” constraints on the architecture of visual systems, both biological and computational... If the task is an intractable one, as vision in its most general form seems to be, complexity satisfaction is not simply a detail to contend with during implementation, just as discretization and sampling effects or numerical stability are not simply implementational details. Complexity satisfaction is a major constraint on the possible solutions of the problem. It can distinguish between solutions that are realizable and those that are not. It is important to specify exactly what is meant by computational complexity-level analysis: Given a task, a set of performance specifications, a fixed amount of input, and a fixed set of resources with which to accomplish the task, two related questions can be asked. The first is, “How much computation is required to accomplish the task?”; the second is, “Are the given resources sufficient to accomplish the task?” In general, the resources specified in a problem description do not necessarily match the required computation; there could be a mismatch between problem complexity (the answer to the first question) and the resources. This does not mean that no realization is possible – it means that further analysis is required to reshape the task or to optimize the resources so as to attain a satisfactory match. Note that reshaping the problem often means making approximations or being content with suboptimal solutions – aspects of the full generality of the problem must be sacrificed to obtain a realizable solution. This process of optimization toward matching the computational requirements of a problem with a given resource I call “analysis at the (computational) complexity level.” The result of the analysis will show how much computation will actually be performed, what the nature of the actual problem solved is, and what the first-order performance characteristics of the realization are. This analysis will not provide answers to “how” questions – how the computation is actually carried out. Nevertheless, ascertaining how much computation can be performed will strongly constrain which computations are chosen to actually solve the problem. As with many aspects of science, this analysis points to an iterative methodology. This complexity analysis will deal only with first-order principle model complexity, analogous to building a house starting with the internal wood frame. Once the frame defines the skeleton of the house, one can begin to add detail. So too with this analysis fine-scale considerations are not dealt with. When a problem is inherently intractable, one must reduce the intractability at the large scale before worrying about detailed considerations at finer scales. The process of design does not depend on only one type of building material but on many. We will consider



only two types of material: complexity satisfaction and minimization of cost” (in compliance with the Occam’s razor principle in problem solving [76], [79]). “The constraints we derive will be termed “sufficient” in one sense only: They are sufficient to satisfy the first-order computational complexity-level analysis. It must also be noted that the constraints are not formal necessary conditions. An engineer is provided with a set of design specifications a new apparatus must meet. We are faced with an inverse problem and much more difficult one: discovering the specifications and design principles of an existing system whose performance and composition is still far from being understood. This lack of understanding would appear to make such an inverse analysis impossible. However, there is a core of well-accepted facts about the ventral stream in the primate visual cortex eligible for use as quantitative constraints in a computational model of vision. They are summarized as follows [5], [68], [126].

- There are 30,000 readily identifiable individual objects in the world, excluding whole scenes or collections of objects. If these were included a very large number of visual prototypes VP would presumably result. Thus, a conservative lower estimate for the number of target visual prototypes is  $VP = 100,000$ . A large but arbitrary upper estimate would be  $VP = 10,000,000$ .
- The average connectivities among neurons is about 1000 for both fan-out and fan-in.
- There are 30 visual areas or so in primates, e.g., V1, V2, V3, MT and V4, with various degree of retinotopy.
- There is a hierarchical organization connecting these areas. Hierarchical visual processing aims to build invariance to position and scale first and then to viewpoint and other transformations [126].
- The area of each hemisphere of the primary visual cortex, V1, in humans is 1500-3700 mm<sup>2</sup>, which comprises the set of “units of output” in V1, called hypercolumns, where each hypercolumn is approximately 1 mm<sup>2</sup> in area and that there are therefore 1500-3700 hypercolumns in V1 or 2100 on average.
- Responses of individual neurons may be affected by a spectrum of stimuli rather than a single one. Along the hierarchy, the receptive fields of the neurons (i.e., the part of the visual field that could potentially elicit a response from the neuron) as well as the complexity of their optimal stimuli (i.e., the set of stimuli that elicit a response of the neuron) increases [126]. The initial processing of information is feedforward and pre-attentional for immediate recognition tasks, i.e., when the image presentation is rapid and there is no time for eye movements or shifts of attention [5], [126].
- Performance of a given neuron may be profoundly affected by attentional influences related to the task at hand” (End long quote of Tsotsos, [68], pp. 423, 424).

Inspired by Tsotsos, to accomplish a CV system development whose goal is not only to mimic the processing of visual information in human beings, but to match human performance in complex visual tasks, an original specification of a biologically plausible CV system’s design and implementation requirements is proposed as follows.

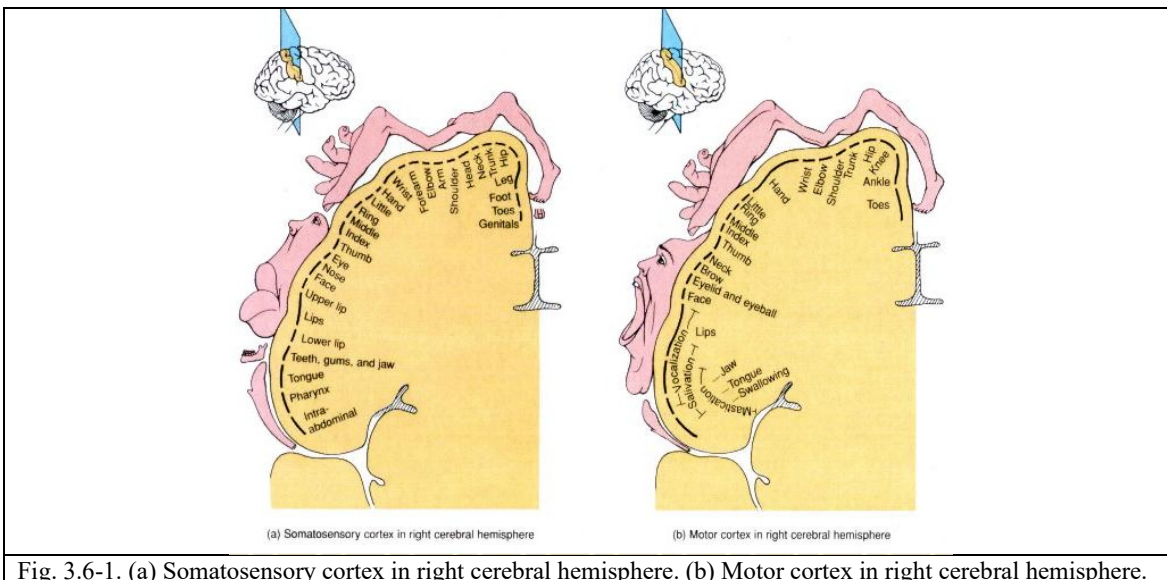
1. **Mandatory image enhancement (pre-processing) for data harmonization across time, space and sensors.** Physical model-based and hybrid inference systems require as input sensory data provided with a physical meaning, i.e., provided with a physical unit of measure. In Earth observation (EO) image understanding systems (IUSs), in agreement with the QA4EO guidelines [111], radiometric calibration of digital numbers into top-of-atmosphere or surface reflectance values becomes mandatory. In statistical systems, radiometric calibration is not mandatory, but highly recommended to guarantee inter-image and inter-sensor image harmonization. When no radiometric calibration parameter is available, statistical color constancy becomes mandatory. In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions [144], so that they appear identical to a “canonical” (reference) image, subject to a “canonical” (known) light source (of controlled quality), e.g., under a white light source [112]. In short, solution of the color constancy problem is the recovery “of an illuminant-independent representation of the reflectance values in a scene” [113].
2. **Complex system = distributed processing system = artificial neural network (ANN) paradigm**, see Chapter 3.1. The CV system must comply with the cybernetics paradigm of an artificial mind in an electronic brain, see Fig. 3.1-2. Hence, the CV system is implemented as an ANN which comprises: (i) elementary processing units (PUs) belonging to different families of PUs, where each family of PUs is provided with a specific input-output data (signal) transfer function, (ii) inter-node (lateral) connections (synapses), provided with (I) a local weight (transmission coefficient) to be modulated by an application-specific mechanism of cooperation (coupling), and (II) an inter-node dynamic switching on/off mechanism to be activated by an application-specific mechanism of competition (independence), e.g.,

refer to the Competitive Hebbian learning Rule (CHR) for dynamic generation of a node in a network based on local properties proposed by Martinetz et al. [69]. When inter-node competition is introduced, one ANN tends to form specialized sub-networks, individually provided with intra-network connections, but mutually independent because provided with no inter-network connection [69], [145], [178], [179], [180], [187]. It is worth recalling that ANN applications for CV tasks have been recently revamped by DCNNs [45], [80], [95], [96], [114], [115], [116], [117], [118], [119], [120], [127], [128], [167], [184]. However, in DCNNs spatial filter convolutional coefficients are learned from supervised training data by error backpropagation applied end-to-end, i.e., throughout the DCNN multi-layer depth. Once initial conditions of a DCNN are fixed *a priori* based on heuristics in terms of system architecture (number of layers, number of filters per layer, spatial filter size, inter-filter spatial stride, local filter size for spatial pooling, inter-filter spatial stride for spatial pooling, etc.), the statistical model-based approach to DCNN parameterization, specifically learning-from-data of convolutional coefficients, is considered alternative to traditional so-called hand-crafted (physical model-based) spatial filters, such as shift invariant feature transform (SIFT) [21]. No physical model-based (hand-crafted) spatial filter bank is employed by statistical model-based DCNNs at any hierarchical filter level. For example, it would be impossible to answer questions such as: where and how a statistical model-based DCNN detects image-contours? In other words, what is an operational definition of image-contour provided by a statistical model-based DCNN? The answer would be: it is impractical/impossible to know. Alternative to statistical model-based DCNNs, hybrid model-based spatial filter banks should combine statistical model-based and physical model-based inference approaches.

3. **The distributed processing system is capable of topology-preserving feature mapping** [69], [145], [178], [179], [180], [187], see Fig. 3.1-5 and Fig. 3.1-6, where topology-preserving feature mapping is synonym of 2D image analysis in CV and retinotopic visual feature representation in biological vision [187]. The brain's organizing principle is topology-preserving feature mapping and in the visual system topology-preserving maps are primarily spatial [94]. There may be 30 visual areas or so in primates, e.g., V1, V2, V3, MT and V4, with various degree of retinotopy [68].

Well-known biological examples of topology-preserving maps (TPMs) are:

- Somatosensory cortex (omuncolo) and motor cortex in the right cerebral hemisphere, see Fig. 3.6-1.
- Retinotopic maps in the visual cortex.



In 2D arrays of spatial filters, such as deep convolutional neural networks (DCNNs), both spatial topological and spatial non-topological information are investigated [45], [80], [95], [96], [114], [115], [116], [117], [118], [119], [120], [127], [128], [167], [184]. Hence, DCNNs are instances of the TPM class.





A formal definition of topology-preserving feature mapping was proposed Martinetz et al. in their seminal work [69]. Let us consider a mapping  $\phi$  from an input data manifold  $X \subseteq R^D$ , where  $D$  is the input data dimensionality, onto the units (neural vertices, processing elements)  $i, i=1, \dots, M$ , of a graph (network)  $G$  defined as a set of interconnected vertices, where inter-node arcs are also called *lateral connections*. In a physical world, a real-world network  $G$  of processing elements and lateral connections can belong to a 1D, 2D or 3D Euclidean space. Each vertex in  $G$  has a pointer (codeword)  $\mu_i \in X$  associated to it, such that pointer set (codebook)  $C = \{\mu_1, \dots, \mu_M\}$ , where  $M$  is the total number of nodes in graph  $G$ . Each pointer  $\mu_i \in X$  is the “center” of the (Voronoi) receptive field (hypervolume of attraction, domain of activation) in the data-domain  $X$  of vertex  $i \in G$ . A feature vector  $x \in X$  is mapped to vertex  $w1(x) \in \{1, M\}$  the pointer  $\mu_{w1(x)}$  of which is closest (in the Euclidean sense) to  $x$ ; that is,  $x$  is mapped to the vertex  $w1(x) \in \{1, M\}$  whose Voronoi polyhedron  $V_{w1(x)}$  encloses a  $D$ -dimensional data vector  $x$ .

$$\phi_C: X \rightarrow G, x \in X \rightarrow w1(x) \in G, \quad (6-1)$$

$$index\ w1(x) = \underset{j \in \{1, M\}, j \in G}{\operatorname{argmin}} \|x - \mu_j\|, \mu_j \in C. \quad (6-2)$$

Forward mapping  $\phi_C$  from data manifold  $X$  to graph  $G$  is *neighborhood preserving* if vectors  $x$ s that are close in  $X$  are mapped to vertices  $i$ s that are close within  $G$ . This requires that pointers  $\mu$ s that are neighboring on  $X$ , i.e., pointers whose Voronoi polyhedra,  $V$ s, are adjacent on  $X$ , are assigned to vertices that are adjacent in  $G$ .

Inverse (backward) mapping  $\phi_C^{-1}$  from graph  $G$  onto manifold  $X$  is determined by

$$\phi_C^{-1}: G \rightarrow X, i \in G \rightarrow \mu_i \in X, \quad (6-3)$$

Analog to neighborhood preservation of  $\phi_C$ , neighborhood preservation of inverse mapping  $\phi_C^{-1}$  is given if pointers  $\mu$ s belonging to adjacent vertices in  $G$  are neighboring on  $X$ , i.e., if their Voronoi polyhedra  $V$ s are adjacent in  $X$ .

A Topology-preserving Map (TPM) of input data manifold  $X$  is a graph  $G$  whose vertices  $i$ s are assigned to pointers  $\mu$ s lying on the input data manifold  $X$  such that the mapping  $\phi_C$  from manifold  $X$  to graph  $G$  as well as inverse mapping  $\phi_C^{-1}$  from graph  $G$  onto manifold  $X$  is *neighborhood preserving*. Only then  $G$  forms a map on which adjacent vertices  $i$ s correspond to pointers  $\mu$ s whose Voronoi polyhedra  $V$ s are adjacent in the input manifold  $X$ , and vice versa [69] (pp. 512, 515), see Fig. 3.6-2 to Fig. 3.6-4.

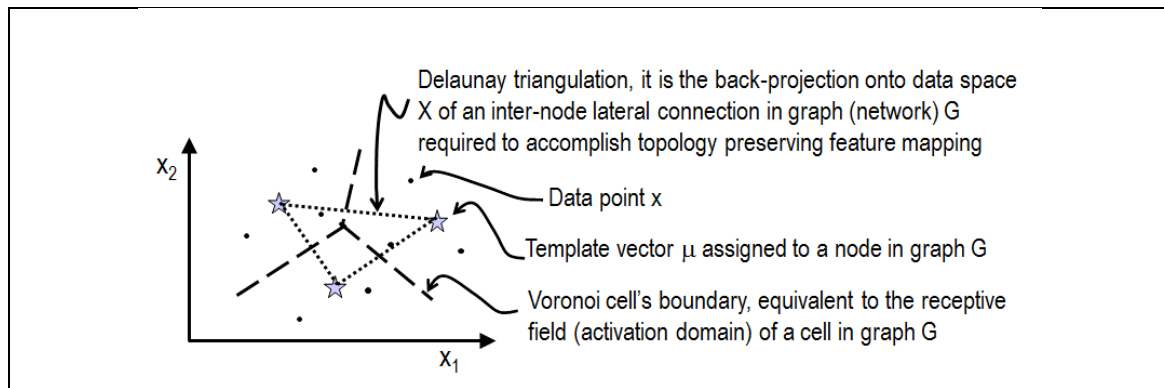


Fig. 3.6-2. A powerful structure from computational geometry that solves or, at least, yields a starting point for efficiently solving several proximity problems is the so-called Voronoi diagram and its dual, the Delaunay triangulation. The *(induced) Delaunay triangulation* is the graph that connects pointers  $\mu$ s whose Voronoi polyhedra,  $V$ s, are adjacent on the given input manifold  $X$ , i.e., Voronoi cells that share an edge [69] (p. 515).

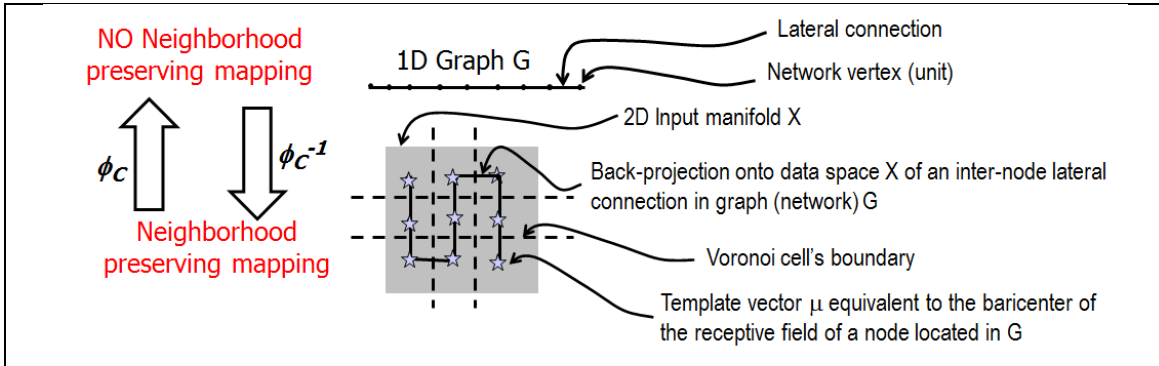


Fig. 3.6-3. TPM is possible if and only if (iff) the topology and dimensional structure of network  $G$  matches or is superior to the topology of input manifold  $X$ . This is the typical problem encountered with the Self-Organizing Map (SOM), whose 1- or 2D topology is fixed on an a priori basis, thus SOM cannot match the topology of any input manifold whose dimensionality  $D > 2$ , i.e., SOM cannot guarantee for topology-preserving feature mapping [69].

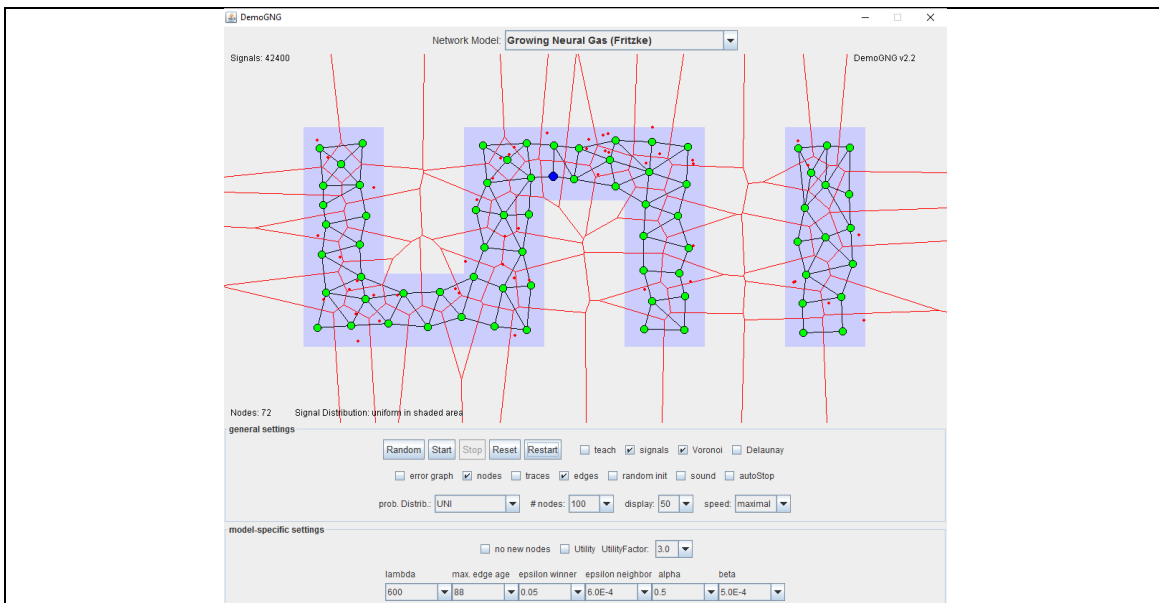


Fig. 3.6-4. Example of two mutually independent TPMs dynamically generated on the web by the self-organizing Growing Neural Gas simulator (<http://www.demogng.de/>) [145]. Light blue: input 2D data manifold  $X$ . Red line: Voronoi's cell boundaries. Red circles: Data samples in  $X$ . Green circles: Codewords in manifold  $X$  corresponding to nodes in network  $G$ . Each codeword is a receptive field's baricenter. Black lines: Induced Delaunay triangulation. Each line between a pair of codewords in  $X$  is the back-projection in  $X$  of a lateral inter-node connection in network  $G$ , where  $G$  is expected to belong to a physical 2D or 3D Euclidean space. Blue circle: last first best-match center of receptive field where a data sample occurs.

Self-organizing TPM algorithms combine cooperation and competition mechanisms among processing elements and lateral connection (synapses) in  $G$  to accomplish specialization of multiple TPMs mutually independent because they do not share lateral connections, see Fig. 3.6-4. Local rules of self-organizing TPM algorithms encompass four degrees of freedom: (i) dynamic (on-the-fly) generation of a node in  $G$  based on local properties, (ii) dynamic removal of a node in  $G$  based on local properties, (iii) dynamic generation of a lateral connection based on the so-called Competitive Hebbian learning rule (CHR) proposed by Martinez et al. [69], (ii) dynamic removal of a lateral connection in  $G$  based on local properties [69], [145], [178], [179], [180], [187]. The elegant, but powerful CHR states that given a data sample  $x$  in manifold  $X$ , determine the first best-matching node  $w_1(x)$  and the second best-matching node  $w_2(x)$  according to Eq. (6-2). If between these two nodes a lateral connection already exists, then “refresh” (reinforce) the



weight associated with this lateral connection. Otherwise, generate a new lateral connection between these two nodes [69].

4. **Hybrid (combined deductive/top-down/physical model-based and inductive/bottom-up/statistical model-based) inference** [45], [77]. There is strong evidence that both low- and high-level vision are hybrid inference processes, e.g., refer to the work by Vecera and Farah on low-level image segmentation [8] and to the work by Frintrop on visual attention [71]. In general, no cognitive system in nature starts from scratch (*tabula rasa*), but from initial conditions provided by genotype [81], equivalent to prior knowledge available in addition to sensory data. By starting from prior knowledge, a hybrid inference system can require no user interaction to initialize any system's free-parameter. Noteworthy, like physical models, hybrid inference systems where deductive and inductive stages alternate, starting from a deductive initialization first stage, consider sensory data radiometric calibration (*Cal*) as mandatory, to provide dimensionless data with a physical unit of measure [111]. Unlike physical and hybrid data processing systems, statistical data models do not require as input numeric variables provided with a physical meaning. Although statistical models do not require physical variables as input, they can benefit from data harmonization accomplished through data *Cal*, which augments their robustness to changes in the input data set acquired across time, space and sensors.
5. **Feedback loops** [8], [24], [82], [83], [93], [96], [126], [142]. Perception involves a complex interaction between feedforward sensory-driven information and feedback attentional, memory, and executive processes that modulate such feedforward processing. For example, mental imagery is known to induce retinotopically organized activation of early visual areas via feedback connections, which is tantamount to saying that mental images in the mind's eye can alter the way we see things in the retina [82], [83].
6. **Hierarchical submodular approach to form neural modularity systems, equivalent to a network of sub-networks according to a divide-and-conquer problem solving approach** [81], [159], [169]. According to the simple "power of two" rule at the basis of the Theory of Connectivity, an operational principle can serve as a unified wiring and computational logic for organizing and constructing cell assemblies into a pre-configured building block, termed functional connectivity motif (FCM).

“(Start long quote of Tsien [159]) Defined by the power-of-two based equation,  $N = 2^i - 1$ , each FCM consists of the principal projection neuron cliques ( $N$ ), ranging from those specific cliques receiving information inputs ( $i$ ) to those general and sub-general cliques receiving various combinatorial convergent inputs... Instead of using a single neuron as the computational unit in some extremely simple brains, in the most of the brains a group of neurons exhibiting the similar tuning properties, termed as a neural clique, should serve as the basic computing unit. This allows the system to avoid a catastrophic failure in the event of losing a single neuron (in engineering, the term for this phenomenon is “graceful degradation” or, simply, “robustness”). Here,  $N$  is the number of distinct neural cliques connected in different possible ways;  $i$  is the information types (e.g., sensory scalar variables) this FCM is dealing with. According to this equation, each FCM is predicted to consist of a full range of neural cliques that extract and process a variety of inputs in a combinatorial manner. Thus, depending on how many distinct information input variables a microcircuit is designed to handle, this equation,  $N = 2^i - 1$ , defines how many neural cliques are needed for a particular FCM.” For example, test animals were given different combinations of four different foods, such as usual rodent biscuits as well as sugar pellets, rice and milk, and as the Theory of Connectivity would predict, the scientists could identify all 15 different cliques, or groupings of neurons, that responded to the potential variety of food combinations. The neuronal cliques appear prewired during brain development because they showed up immediately when the food choices did. Due to exponential growth in input numbers  $i$ , the cost in terms of cell (neuron) resources can quickly become prohibitive. For instance, in order to cover all possible patterns for processing 2, 3, 4, 10, 20, 30, 40 distinct perceptual inputs, an FCM would require 3, 7, 15, 1023, 1048675, 1073741823, 1 099 511 627 775 neural cliques, respectively. Even the human brain, which is estimated to have  $8.6 \times 10^{10}$  neurons, can quickly run out of cells to be able to afford this exponential coverage. The best and necessary solution is to employ modular approaches, or a divide-and-conquer strategy, to segregate or stream information inputs through distinct sensory domains or submodular pathways. For example, “if a central node in a small neural circuit needs to cover all possible connectivity wiring patterns to represent eight distinct types of inputs or information, a total of 255 binary elementary neurons would be required ( $N = 2^8 - 1 = 255$ ) by this node or FCM. However, when a sub-modular approach is employed, e.g., by dividing the eight distinct types of inputs into the 16 possible combinations of four inputs per sub-node or FCM, where each sub-node consists of 15 binary elementary neurons, the same original set of 255 binary



elementary neuron can increase its combinatorial processing capacity by a factor of  $N_2 = 255$  total neurons / 15 neurons per sub-node = 17 times. Similarly, if a sub-node or FCM was structured to process only three distinct information, with  $N_2 = 2^3 - 1 = 7$  out of eight types of inputs whose  $N_1 = 255$ , then 255 neurons can be used for  $N_1 / N_2 = 255 / 7 = 36$  assemblies, i.e., the same number of 255 neurons required to process 8 distinct inputs can be employed in 36 different FCMs, each one dealing with three distinct inputs.” “In order to extract an increasing amount of relational patterns across distinct sensory modalities, scaling up this power-of-two based microarchitecture is necessary, but can be a major challenge from an engineering perspective. For example, the classic three- or six-layered cortex is the evolutionary solution to execute the FCM logic in a replicable large-scale fashion, as the brain evolves from small-scale circuits (sub-networks) to larger circuits (networks of sub-networks). In other words, input cortical layers should consist of most of the specific cliques and simple sub-general cliques, e.g., 2-event combinatorial cliques whose inputs are sensory variables; whereas output layers would host most of the more complex sub-general cliques and general cliques, e.g., 3 event- or 4 event-combinatorial cells whose inputs stem from lower-level cliques. In fact, this is in general agreement that layer 4 neurons are simple cells and predominantly project to layers 2 and 3 from which neurons then descend to layers 5 and 6. Although a majority of these output cell cliques should be sub-general and general cliques, one should also expect a certain percentage of cells in these layers to be specific due to direct inputs from layer 4. It should be noted that layers 2/3 are also the output layers to other cortical regions. As a whole, this arrangement can enable the discovery and broadcasting of general and combinatorial patterns in the output layers while still being capable of retaining the ability for pattern discrimination. (End long quote of Tsien [159])”

- 7. Capability to fully exploit spatial topological and spatial non-topological information components in the (2D) image-domain and in the 4D spatiotemporal scene-domain** [45], [80], [167]. Based on daily human experience it is an undisputable fact that biological chromatic and achromatic visual systems are nearly as effective [45]. It means that spatial information, either topological (e.g., adjacency) or non-topological (e.g., spatial distance), typically dominates color information in both the image domain and the scene domain. In common practice, chromatic and achromatic computational models of human vision must prove to be nearly as effective, see Fig. 3.1-5 and Fig. 3.1-6. A necessary and sufficient experimental proof of scalability is required to prove that the CV system at hand thoroughly investigates spatial topological and spatial non-topological information which typically dominate color information in vision tasks.

If a chromatic CV system does not down-scale seamlessly to achromatic image analysis, then it tends to ignore the paramount spatial information in favor of subordinate (secondary) context-insensitive color information, such as class-specific spectral signatures typically investigated in traditional pixel-based single-date or multi-temporal EO-IUSs, see Fig. 3.1-10. In other words, a *necessary and sufficient condition for a CV system to fully exploit spatial topological and spatial non-topological information components in addition to color is to perform nearly as well when input with either panchromatic or color imagery.*

- 8. Foveated vision** [109], [110]. Human vision is provided with eccentricity-dependent resolution in the retina, so that spatial filters with different spatial scales span different visual fields (window sizes), see Fig. 3.6-5. Hence, human foveated vision implies that fixation point selection based on topology-preserving saliency maps becomes mandatory [71], [73], [74], [142]. When they process every part of the image the same way, eyes do not move, such as in bees, differently from humans [72], [74]. For example, DCNNs are typically implemented in a non-foveated imaging framework, where spatial filters at different spatial scales span the same window size [95], [96], [114], [115], [116], [118], [119], [120]. An exception is found in [117], where several DCNNs employ in parallel one filter size empirically related to one window size across several filter sizes, see Fig. 3.6-7. In [184], a selective window search is proposed for an eccentricity-independent DCNN.

In the words of Poggio [110]: “The initial goal of the Center for Brains Minds and Machines (CBMM) project titled ‘The computational role of eccentricity dependent resolution in the retina: consequences for hierarchical models of object recognition’ is to understand the consequences for object recognition of foveated imaging: how it affects performance in different situations (e.g., controlled vs. cluttered scenes), the number of fixations required for learning and inference, and the effect of other architectural choices (e.g., local vs. global pooling over space or scale). Of longer term interest is the observation that nested signatures necessarily encode information about the hierarchical structure

of scenes, and might therefore be elements of a shared representation for object recognition as well as higher level reasoning about a scene, in the direction of the CBMM challenge at the MIT”.

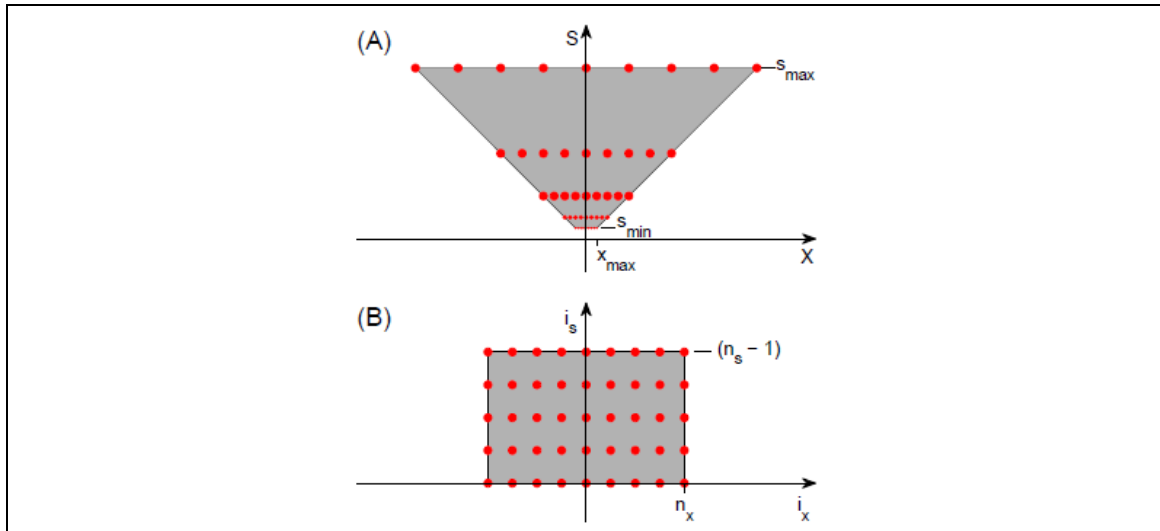


Fig. 3.6-5. Reproduced with permission, courtesy of [109]. An inverted truncated pyramid (A) has the same number of sample points at every scale. It maps perfectly onto a square array (B) when  $x$  is replaced by  $i_x = x = s$ , i.e., the number of samples from the center.  $i_s$  is the scale band number. This inverted truncated pyramid is called the “magic map”. Noteworthy, in the (2D) image plane, a “magic map” is 3D, whose space-scale dimensions are  $[(I_x, I_y), I_s]$ . This retinotopic model of  $\approx 40 \times 40$  hypercolumns is different from Tsotsos’, shown in Fig. 3.1-5 and Fig. 3.1-6, where  $\approx 2000$  hypercolumns are organized in a non-retinotopic 1D array.

9. **Consistency with human visual perception, featuring visual illusions.** For example, perceptual consistency with the Mach bands illusion, where a bright and a dark band are seen at ramp edges, implies that statistical variables such as local variance, local contrast and first-order derivative (gradient) computed in moving windows, local spatial filters or within image-segments (in OBIA approaches) are NOT suitable for perceptual detection of image boundaries, see Fig. 3.4-1. Noteworthy, local variance is monotonically increasing with local contrast, which is synonym of local intensity of a spatial first-order derivative (gradient) in the (2-D) image-domain. Stemming from the Mach bands illusion in human vision well known by cognitive science, this perceptual fact is in contrast with a large portion of the CV and RS literature, where mainstream image-contour detection [27] and image segmentation algorithms [173], [174], [175] employ local contrast (gradient) or local variance thresholding criteria. In the words of Pessoa, “if we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published computational vision models becomes surprisingly small [25], refer to Chapter 3.4. For example, the popular region growing algorithms proposed by Baatz et al. [173] and by Espindola et al. [174], [175], adopted by respectively the well-known eCognition commercial software product [176] and the freeware SPRING software for RS image processing [175], as well as image-contour detection algorithms based on spatial filter banks, e.g., the popular Canny edge detector [27], employ a visual feature detection criterion based on local variance or local contrast thresholding, with numeric threshold  $\in (0, +\infty)$ . In particular, in the eCognition commercial software product [176], where image segmentation is implemented by the region growing algorithm by Baatz et al. [173], within-segment variance is thresholded by a so-called “(spatial) scale parameter” to be user-defined in range  $(0, +\infty)$  based on heuristic criteria: if the within-region variance threshold gets looser, then image-regions grow larger (at a coarser spatial scale). The corollary is that the infamous “scale parameter” in eCognition has nothing to do with spatial scale in the image-domain, but with within-segment variance thresholding, which in turn has nothing to do with perceptual detection of image boundaries, which is independent of local contrast (gradient) and local variance.
10. **Multi-source image scalability to varying imaging sensor specifications,** i.e., capability to process multi-source imagery, including:
  - Uncalibrated panchromatic and RGB images.





- Radiometrically calibrated multi-spectral (MS), Super-spectral (SS) and Hyper-spectral (HS) images, whose number of spectral channels  $N$  is  $\{2, 9\}$ ,  $\{10, 20\}$  and  $> 20$  respectively.
  - Bi-temporal Red-Green-Blue (RGB) synthetic aperture radar (SAR) images [168].
11. **Convergence-of-evidence approach** [45], see Fig. 3.6-6. It is somehow related to the Feldman’s so-called “neural binding problem” (NBD) [94]. Typical of human symbolic reasoning traditionally mimicked by fuzzy logic [97], a convergence-of-evidence problem solving approach allows to infer strong conjectures from multiple independent sources of eventually weak evidence. By definition of conditional probability,

$$p(x|y) = \frac{p(x,y)}{p(y)}.$$

According to the Bayesian law, which provides a principled way of combining new evidence stemming from new data with prior beliefs, the following equation holds true.

$$p(c|x) = \frac{p(x|c)p(c)}{\sum_{c1=1}^{Classes} p(x|c1)p(c1)} \propto p(x|c)p(c), \quad c = 1, \dots, Classes.$$

If the “naive” conditional independence of features  $F_i, i = 1, \dots, I$ , is assumed to hold true, by using the chain rule for repeated applications of the definition of conditional probability then the popular formulation of a so-called “naive” Bayes classifier becomes

$$p(c|F_1, \dots, F_I) = p(c) \prod_{i=1}^I p(F_i|c). \quad (6-4)$$

Eq. (6-4) shows that any convergence-of-evidence approach is more selective than each individual source of evidence, in line with a focus-of-attention mechanism [71].

Let’s consider a convergence-of-evidence approach to CV for image understanding where any spatial unit  $x$ , where  $x$  is either a (0D) point, (1D) line or (2D) polygon in the image-domain according to the OGC nomenclature [55], is described by several well-known sources of visual evidence such as [80]:

- color [22], [132],
- local shape [29], [166],
- texture [2], [3], [58], [59], [60],
- spatial relationships [45], [80],

specifically, ColorValue(x), local ShapeValue(x), TextureValue(x) and SpatialRelationships(x, Neigh(x)) where ColorValue(x) belongs to a MS data space  $\mathcal{R}^{MS}$ , i.e., ColorValue(x)  $\in \mathcal{R}^{MS}$ , and Neigh(x) is a generic 2D spatial neighborhood of spatial unit  $x$  in the image-domain.

Irrespective of their Pearson’s cross-correlation, if any, it is easy to prove these visual properties are statistically independent (because cross-correlation does not mean causation) [29].

If priors are ignored because considered equiprobable in a maximum class-conditional likelihood inference approach alternative to a maximum *a posteriori* optimization criterion, then the naive Bayes classifier becomes

$$\begin{aligned} & p(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\ & p(\text{ColorValue}(x)|c) \cdot p(\text{ShapeValue}(x)|c) \cdot p(\text{TextureValue}(x)|c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x))|c) = \\ & \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} p(\text{ColorValue}(x)|\text{ColorName})p(\text{ColorName}|c) \cdot p(\text{ShapeValue}(x)|c) \cdot \\ & p(\text{TextureValue}(x)|c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x))|c), \\ & c = 1, \dots, \text{ObjectClassLegendCardinality}, \end{aligned} \quad (6-5)$$

where color space  $\mathcal{R}^{MS}$  is partitioned into hyperpolyhedra, equivalent to a discrete and finite dictionary of static color names, with ColorName = 1, ..., ColorDictionaryCardinality. To further simplify Eq. (6-5), its canonical interpretation based on frequentist statistics can be relaxed by fuzzy logic [97], so that the logical-AND operator is replaced by a fuzzy-AND (min) operator, inductive class-conditional probability  $p(\text{ColorValue}(x)|c) \in [0, 1]$ , where

$\sum_{c=1}^{\text{ObjectClassLegendCardinality}} p(\text{ColorValue}(x)|c) \geq 0$ , is replaced by a deductive membership (compatibility) function  $m(\text{ColorValue}(x)|c) \in [0, 1]$ , where  $\sum_{c=1}^{\text{ObjectClassLegendCardinality}} m(\text{ColorValue}(x)|c) \geq 0$ , and color space hyperpolyhedra are considered mutually exclusive and totally exhaustive. If these simplifications are adopted, then Eq. (6-5) becomes

$$m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \min\left\{ \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x)|\text{ColorName})m(\text{ColorName}|c), m(\text{ShapeValue}(x)|c), m(\text{TextureValue}(x)|c), m(\text{SpatialRelationships}(x, \text{Neigh}(x))|c) \right\} = \min\{m(\text{ColorName}^*|c), m(\text{ShapeValue}(x)|c), m(\text{TextureValue}(x)|c), m(\text{SpatialRelationships}(x, \text{Neigh}(x))|c)\},$$

$c = 1, \dots, \text{ObjectClassLegendCardinality}$ , where  $\text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}$ , such that  $m(\text{ColorValue}(x)|\text{ColorName}^*) = 1$  and  $m(\text{ColorName}^*|c) \in \{0, 1\}$ . (6-6)

Eq. (6-6) shows that for any spatial unit  $x$  in the image-domain, when a hierarchical CV classification approach estimates posterior  $m(c| \text{ColorValue}(x), \text{ShapeValue}(x), p(\text{TextureValue}(x)|c), \text{SpatialRelationships}(x, \text{Neigh}(x)))$  starting from a near real-time context-insensitive color naming first stage where condition  $m(\text{ColorValue}(x)|\text{ColorName}^*) = 1$  holds, if condition  $m(\text{ColorName}^*|c) = 0$  is true according to a static community-agreed binary relationship  $R: \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  (and vice versa) known *a priori*, see Table 3.6-1, then  $m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) = 0$  irrespective of any second-stage assessment of spatial terms  $\text{ShapeValue}(x)$ ,  $\text{TextureValue}(x)$  and  $\text{SpatialRelationships}(x, \text{Neigh}(x))$ , whose computational model is typically difficult to find and computationally expensive. Intuitively Eq. (6-6) shows that static color naming allows the stratification of unconditional multivariate spatial variables into color class-conditional data distributions, in agreement with the statistic stratification principle [122] and the divide-and-conquer problem solving approach [76], [79]. Well known in statistics, the principle of statistic stratification guarantees that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” [122].

		Target classes of individuals (entities in a conceptual model for knowledge representation built upon an ontology language)		
		Class 1, Water body	Class 2, Tulip flower	Class 3, Italian tile roof
Color names	black		√	
	blue	√	√	
	brown	√	√	√
	grey			
	green	√	√	
	orange		√	
	pink		√	
	purple		√	
	red		√	√
	white			√
yellow			√	

Table 3.6-1. Example of a matching function between a dictionary of color names and a dictionary of classes of individual entities in the real-world. The latter dictionary is a superset of the typical taxonomy of land cover (LC) classes adopted by the RS community. “Correct” table entries (marked as √) must be: (i) selected by domain experts and (ii) community-agreed upon.

In Eq. (6-6), the following considerations hold.

- Any numeric  $\text{ColorValue}(x)$  in color space  $\mathfrak{R}^{\text{MS}}$  belongs to a single hyperpolyhedron, identified by  $\text{ColorName}^*$  in the static color dictionary, i.e.,  $\forall \text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , then  $\text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}$

exists such that  $\sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x)|\text{ColorName}) = m(\text{ColorValue}(x)|\text{ColorName}^*) = 1$  holds, where  $m(\text{ColorValue}(x)|\text{ColorName}) \in \{0, 1\}$ ,  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ .

- Color names, physically equivalent to color hyperpolyhedra in a numeric color space  $\mathfrak{R}^{\text{MS}}$ , are conceptually equivalent to latent variables linking observable (“sub-symbolic”) data in the real world to categorical variables of semantic (symbolic) quality in the modeled world.
- The set  $A = \text{DictionaryOfColorNames}$ , with cardinality  $|A| = a = \text{ColorDictionaryCardinality}$ , and the set  $B = \text{LegendOfObjectClassNames}$ , with cardinality  $|B| = b = \text{ObjectClassLegendCardinality}$ , can be considered a bivariate categorical random variable where two univariate categorical variables  $A$  and  $B$  are generated from a single population. A binary relationship (product set) from set  $A$  to set  $B$ ,  $R: A \Rightarrow B$ , is a subset of the 2-fold Cartesian product  $A \times B$ , whose size is rows  $\times$  columns =  $a \times b$ . The Cartesian product of two sets  $A \times B$  is a set whose elements are ordered pairs. Hence, the Cartesian product is non-commutative,  $A \times B \neq B \times A$ . In agreement with common sense, see Table 3.6-1,  $R: \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  is a set of ordered pairs where each  $\text{ColorName}$  can be assigned to none, one or several classes  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  of observed scene-objects, whereas each class of observed objects can be assigned with none, one or several color names as the class-specific color attribute. Binary membership values  $m(\text{ColorName}|c) \in \{0, 1\}$  and  $m(c|\text{ColorName}) \in \{0, 1\}$ , with  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  and  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ , can be community-agreed upon based on various kinds of evidence, whether viewed all at once or over time, such as a combination of prior beliefs with additional evidence inferred from new data in agreement with a Bayesian updating rule (Bayesian inference), largely applied in artificial intelligence and expert systems. A binary relationship  $R: A \Rightarrow B \subseteq A \times B$  where sets  $A$  and  $B$  are categorical variables generated from a single population guides the interpretation process of a two-way *contingency table* (also known as association matrix, cross tabulation, bivariate table or frequency table) [160],  $\text{BIVRTAB} = \text{FrequencyCount}(A \times B)$ . In the conventional domain of frequentist inference with no reference to prior beliefs, a BIVRTAB is the 2-fold Cartesian product  $A \times B$  instantiated by the bivariate frequency counts of the two univariate categorical variables  $A$  and  $B$  generated from a single population. For any BIVRTAB instance, either square or non-square, there is a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  that guides the interpretation process, where “correct” entry-pair cells of the 2-fold Cartesian product  $A \times B$  can be either off-diagonal (scattered) or on-diagonal, if a main diagonal exists. When a BIVRTAB is estimated from a geospatial population without sampling, it is called *overlapping area matrix* (OAMTRX) [163], [164]. When the binary relationship  $R: A \Rightarrow B$  is a bijective function (both 1-1 and onto), i.e., when the two categorical variables  $A$  and  $B$  estimated from a single population coincide, then the BIVRTAB is square and sorted and typically called confusion matrix (CMTRX) or error matrix [161], [162], [165]. In a CMTRX the main diagonal guides the interpretation process. For example, a square  $\text{OAMTRX} = \text{FrequencyCount}(A \times B)$ , where  $A = \text{test thematic map legend}$ ,  $B = \text{reference thematic map legend}$  and cardinality  $a = b$ , is a CMTRX if and only if  $A = B$ , i.e., if the test and reference codebooks are the same sorted set of semantic concepts. In general the class of (square and sorted) CMTRX instances is a special case of the class of OAMTRX instances, either square or non-square, i.e.,  $\text{OAMTRX} \supset \text{CMTRX}$ . A similar consideration holds about summary  $Q^2$ Is generated from an OAMTRX or a CMTRX, i.e.,  $Q^2\text{I}(\text{OAMTRX}) \supset Q^2\text{I}(\text{CMTRX})$ .

A convergence-of-evidence inference approach is consistent with a hierarchical multi-map organization of the human brain, where each logically segregated map is a retinotopic representation of only one type of visual parameter [68]. Typical visual features in the image domain are color, local shape, texture (perceptual spatial grouping of texels [2], [3], [58], [59], [60]) and inter-object spatial relationships, either topological or non-topological [45], [80], [167]. The hierarchical hybrid feedback CV system design proposed in Fig. 3.6-6 is alternative to 1D inductive feedforward CV system architectures, either pixel-based or context-sensitive, shown in Fig. 3.1-10 and Fig. 3.1-12 respectively.

12. **Operating mode.** To be considered in operating mode, a CV system is required to score “high” in a minimally dependent maximally informative (mDMI) set [123], [124] of outcome and process quantitative quality indicators (OP- $Q^2$ Is), to be community-agreed upon in advance in agreement with the QA4EO Validation (*Val*) guidelines [111]. The proposed mDMI set of OP- $Q^2$ Is includes the following [56], [57], refer to Chapter 3.3.

- Degree of automation, inversely related to user-machine interaction. It is monotonically decreasing with the number of system's free-parameters to be user-defined based on heuristics. In the present R&D project requirements specification, full automation is required, i.e., no user-machine interaction is allowed to provide system's free-parameters and/or training data. By definition, an automated system requires no human-machine interaction.
- Accuracy, e.g., mapping accuracy.
- Efficiency, in computation time and memory occupation.
- Robustness to changes in input data.
- Robustness to changes in input parameters, if any.
- Scalability to changes in sensor and user specifications.
- Timeliness, defined as the time interval between data acquisition and product generation.
- Costs in manpower and computer power, including costs required to collect labeled data for supervised data learning, if any. For example, the class of polynomial (P) computational problems, consisting of all those problems that can be solved in polynomial time, is considered tractable, whereas the NP-hard class of problems is considered intractable [68], refer to Chapter 3.1.

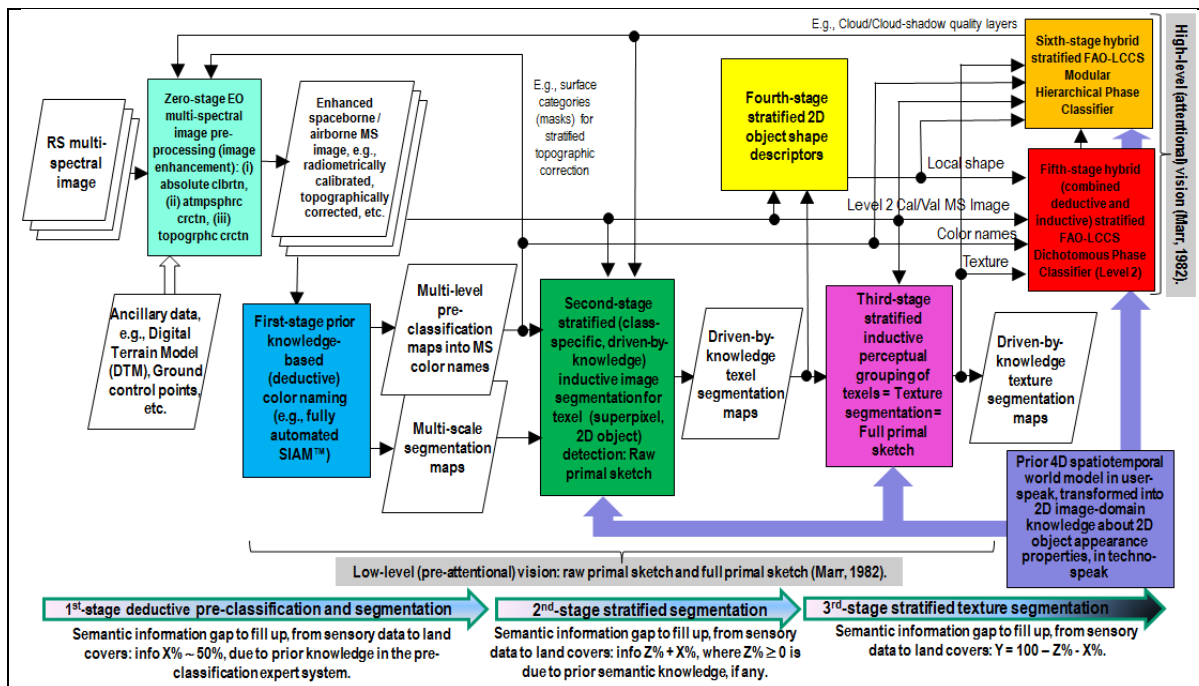
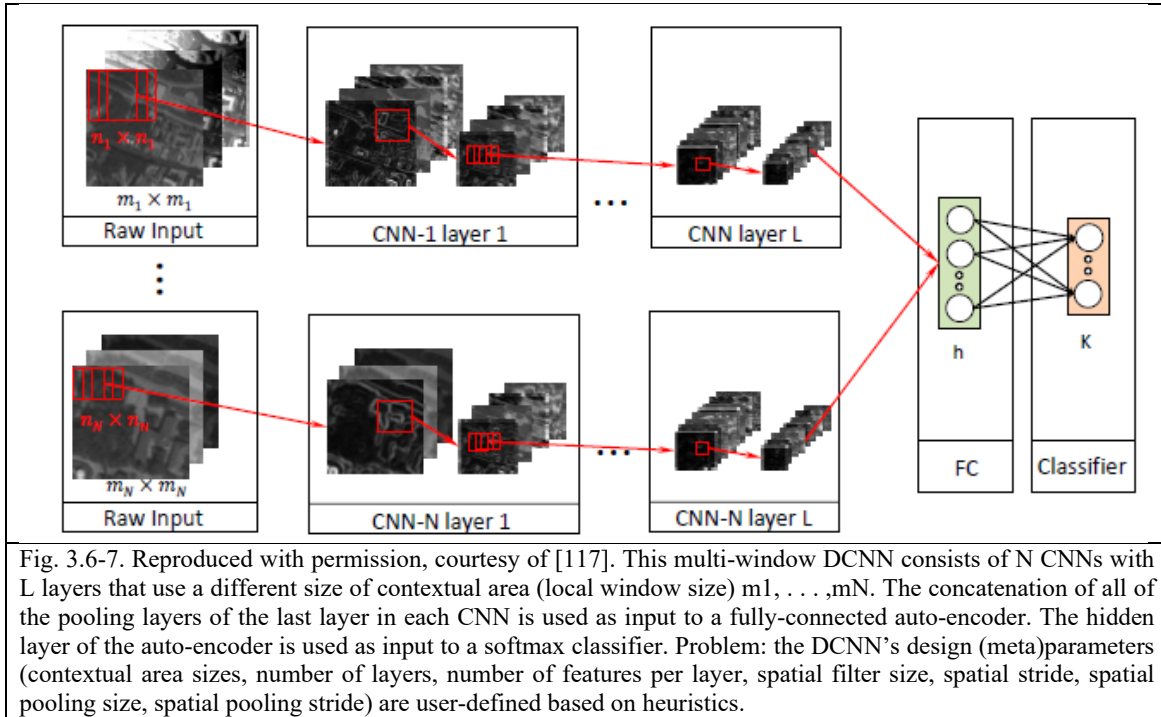


Fig. 3.6-6. Six-stage hybrid feedback EO image understanding system (EO-IUS) based on a convergence-of-evidence approach to low- and high-level EO image understanding (classification). For the sake of visualization each of the six data processing stages plus stage-zero for EO data pre-processing is depicted as a rectangle with a different color fill. Visual evidence stems from color, local shape, texture and inter-object spatial relationships. An example of preliminary general-purpose, user- and application-independent EO image classification taxonomy, such as that required by an ESA EO Level 2 Scene Classification Map (SCM), is the 3-level 8-class FAO Land Cover Classification System (LCCS) Dichotomous Phase (DP) map legend in addition to quality layers such as cloud and cloud-shadow. High-level EO image classification is user- and application-specific, such as an EO SCM product provided with an LCCS Modular Hierarchical Phase (MHP) taxonomy. To provide each EO image stored in a large-scale EO image database with EO value-adding products, including semantics, this EO-IUS is adopted by an innovative EO image understanding for semantic querying (EO-IU4SQ) system [98], [177].



According to these CV system's OP-Q<sup>2</sup>Is, traditional 1D inductive feedforward image analysis approaches, including pixel-based image classifiers (see Fig. 3.1-9) and OBIA approaches (see Fig. 3.1-11), score very low or low respectively. With regard to 2D inductive feedforward image analysis approaches, such as increasingly popular inductive DCNNs, typical drawback are twofold. First, DCNN design (meta)parameters (number of layers, number of features per layer, spatial filter size, spatial stride, spatial pooling size, spatial pooling stride) are user-defined based on heuristics. Second, no foveated imaging framework is pursued, because all multi-scale spatial filters span the same visual field [95], [96], [114], [115], [116], [118], [119], [120]. An exception is found in [117], where one filter size is empirically related to one window size for several filter sizes, see Fig. 3.6-7.

Unlike Fig. 3.6-7, Fig. 3.6-8 shows an original foveated imaging system, i.e., an inverted truncated pyramid called the "magic map" in [109], whose design parameters are fixed, based on a physical model of human vision, such as that discussed by Tsotsos [68] and Baraldi and Parmiggiani [1], and further discussed in Chapter 3.7 (titled "Original 1D simulations for image analysis and synthesis, including the zero-frequency signal component, image-contour detection and keypoint detection consistent with the Mach bands illusion", see Fig. 3.4-1), to be considered the original core of the present Technical Document. The original set of foveated vision parameters suitable to accomplish a wavelet-based eccentricity dependent resolution is the following.

- Number of hypercolumns  $\approx 2000$  [68]  $\approx 40 \times 40$  spatial cells in the high-resolution 102oveolar, in agreement with [109].

Attention: in the (2D) image plane, the Poggio's "magic map" sketched in Fig. 3.6-5 is 3D, whose space-scale dimensions are  $[(l_x, l_y), l_s]$ . This retinotopic model of  $\approx 40 \times 40$  hypercolumns is different from Tsotsos', shown in Fig. 3.1-5 and Fig. 3.1-6, where  $\approx 2000$  hypercolumns are organized in a non-retinotopic 1D array.

- Dyadic spatial scales (one octave apart), in agreement with [16], equivalent to window sizes, required for near lossless (lossy) image analysis/synthesis: from 4 up to 7.
- 2D spatial filter size in pixel units equals to; Scale 0,  $3 = 6 \times \sigma$ , then  $\sigma = 0.5$  pixel units; Scale 1,  $6 \times 2\sigma = 6 \times 1 \approx 7$  pixel units; Scale 2,  $6 \times 4\sigma = 6 \times 2 \approx 13$  pixel units; Scale 3,  $6 \times 8\sigma = 6 \times 4 \approx 25$  pixel units, in agreement with [16].



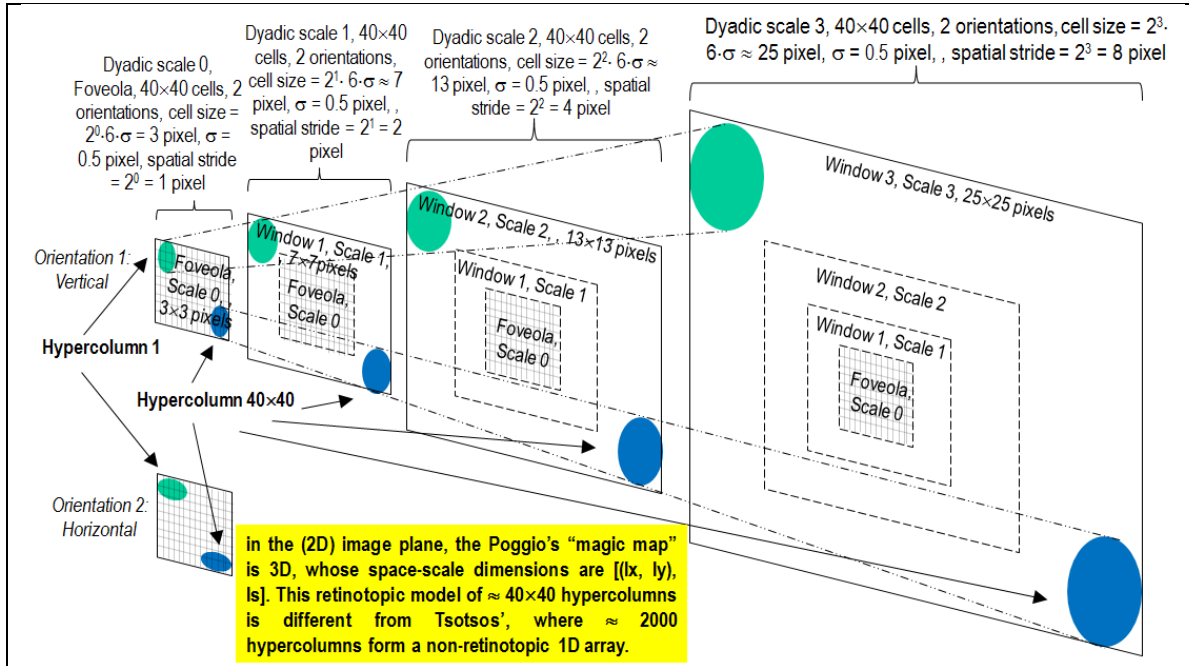


Fig. 3.6-8. Original foveated imaging system, featuring an eccentricity dependent resolution based on a physical model-based parameterization:  $40 \times 40$  cells [109]  $\approx 2000$  hypercolumns [68], 4 dyadic spatial scales, 2 orientations [1]. In this low-level visual system model, the same hypercolumn copes with different receptive fields in different visual fields. Receptive fields of the same hypercolumn are not centered upon the same visual position and do not necessarily overlap one another. Noteworthy, in the (2D) image plane, the Poggio's "magic map" sketched in Fig. 3.6-5 is 3D, whose space-scale dimensions are  $[(lx, ly), ls]$ . This retinotopic model of  $\approx 40 \times 40$  hypercolumns is different from Tsotsos', shown in Fig. 3.1-5 and Fig. 3.1-6, where  $\approx 2000$  hypercolumns are organized in a non-retinotopic 1D array.

- Orientations: from 2 (horizontal and vertical) up to 4 [16]. As shown in Chapter 3.16, two physical model-based convolutional spatial filters are recommended within-scale per-orientation: a band-pass even-symmetric filter  $\partial^2 G / \partial x^2$  and a low-pass Gaussian filter  $G$ .
- Spatial stride (inter-filter distance, from filter center to the neighboring filter center): Scale 0,  $3 \times 3$  filter size in pixel units, inter-filter distance =  $2^0 = 1$  pixel units; Scale 1,  $7 \times 7$  filter size in pixel units, inter-filter distance =  $2^1 = 2$  pixel units; Scale 2,  $13 \times 13$  filter size in pixel units, inter-filter distance =  $2^2 = 4$  pixel units; Scale 3,  $25 \times 25$  filter size in pixel units, inter-filter distance =  $2^3 = 8$  pixel units.

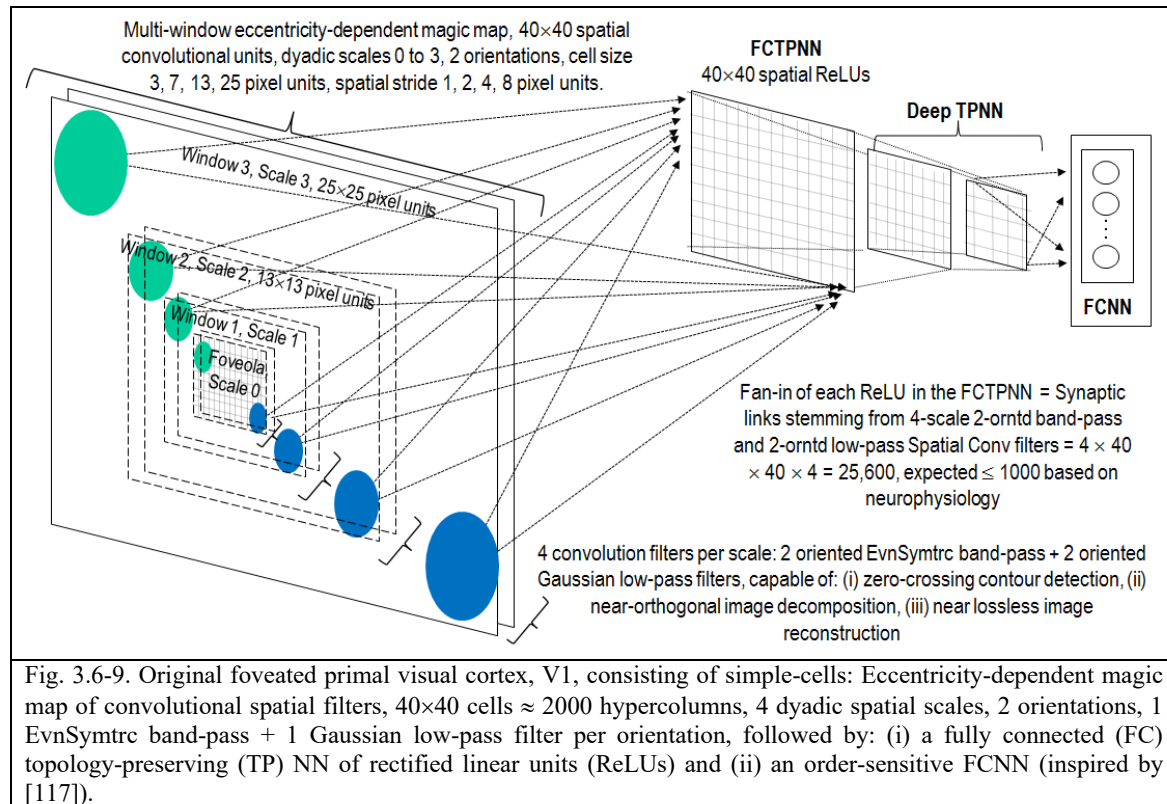
In this original low-level visual system model, the same hypercolumn copes with different receptive fields in different visual fields. Receptive fields of the same hypercolumn are not centered upon the same visual position and do not necessarily overlap one another. The original physical model-based "magic map" shown in Fig. 3.6-8 is equivalent to a multi-window eccentricity-dependent primary visual cortex (V1), consisting of multi-scale multi-orientation simple-cells organized into hypercolumns. It can be proved (refer to Chapter 3.16) that an eccentricity-dependent magic map V1 of convolutional spatial filters, such as that shown in Fig. 3.6-8, consisting of  $40 \times 40$  cells  $\approx 2000$  hypercolumns, where each hypercolumn features 4 dyadic spatial scales, 2 orientations per scale, 2 spatial filters per orientation, specifically a band-pass even-symmetric filter  $\partial^2 G / \partial x^2$  and a low-pass Gaussian filter  $G$ , is maximally informative and minimally redundant because it is capable of:

- (i) zero-crossing (ZX) contour detection in the Marr sense.
- (ii) Near-orthogonal image decomposition.
- (iii) Near lossless image reconstruction.

How can the original physical model-based "magic map" V1 shown in Fig. 3.6-8 connect to higher-level visual network layers, e.g., V2 and beyond? Inspired by the multi-window DCNN shown in Fig. 3.6-7 [117], the following original neural

network (NN, distributed processing system) architecture is proposed as shown in Fig. 3.6-9. The proposed “magic map” V1 is followed by a sequence of: (i) a fully connected (FC) topology-preserving (TP) NN consisting of  $40 \times 40$  rectified linear units (ReLUs), like the first layer superior to the zero-level convolutional layer whose input is the image in a traditional DCNN, and (ii) an order-sensitive FCNN (inspired by [117]). Each ReLU in the FCTPNN features a fan-in, equal to the number of synaptic links stemming from the V1’s filters, specifically, 4-scale 2-oriented even-symmetric band-pass  $\partial^2 G/\partial x^2$  and 2-oriented Gaussian low-pass  $G$  spatial convolutional filters =  $4 \times 40 \times 40 \times 4 = 25,600$ , which is unfortunately superior to an expected fan-in  $\approx 1000$  based on neurophysiological evidence [68].

By the way, the traditional NBD states refers to any large parallel/distributed system where a lot of local information is stored at every node, but this information cannot be fully accessible as a whole to make ideal decisions/actions based on all available information, because this is combinatorially impossible – the system architecture needs to privilege certain combinations [94].



### 3.7 Texture detection problem requirements specification

The requirements specification of a stratified texture detector to be developed is summarized below.

- (1) Provide a stratified (i.e., within-stratum/within-layer) near-orthogonal image analysis/decomposition and synthesis/composition, where an input stratum (mask, layer, multi-part polygon) can be as large as the whole image and as small as a single pixel, such that the image analysis/synthesis is not affected by boundary effects (artifacts) in the proximity of the stratum boundaries.
- (2) In operating mode, refer to Chapter 3.3 and Chapter 3.6.
- (3) Fully automated, i.e., without human-machine interaction, refer to Chapter 3.3 and Chapter 3.6.
- (4) Relate to existing literature as follows. The pre-attentive vision first stage, also called *primal sketch*, consists of a *raw primal sketch* and a *full primal sketch* described below [5]. Also refer to Chapter 3.5.1.1.
  - (I) *Raw primal sketch* [5]. According to Marr [5] (Figure 2-21, p. 73), the raw primal sketch employs as input the zero-crossing (ZX) pixels and generates as output a discrete and finite set of multi-scale *tokens* (discrete

sub-symbolic image plane entities). In general, *a zero-crossing is defined as a place where the value of a function passes from positive to negative* [5], (p. 54). In particular, a zero-crossing in the second derivative of an intensity function (e.g., a Laplacian operator) is located where a (positive) peak or a (negative) trough (solco) in the first derivative of an intensity function occurs due to a sudden intensity change. In 1D functions, ZX pixels in the  $n$ -th derivative identify local extrema in the  $n$ -th - 1 derivative. However, "in two and higher dimensions there is no absolute relationship between locations of the Laplacian ZX curves and the local extrema of a signal. A Laplacian ZX curve may enclose either no extremum, one extremum, or more than one local extremum. Only in the one-dimensional case it holds that there is exactly one local extremum point between two ZXs of the second derivative" [11] (p. 213). In 2D intensity functions  $I(x,y)$ , since intensity changes occur at different spatial scales in an image, then their optimal detection requires the use of operators of different sizes. The most satisfying local operator according to several criteria is the filter  $\nabla^2 G$ , where  $\nabla^2$  is the Laplacian operator ( $\partial^2/\partial x^2 + \partial^2/\partial y^2$ ) and  $G(x,y)$  is the 2D Gaussian function [5] (p. 54). Hence, ZX pixels are intended in the  $\nabla^2 G$ -filtered image, such that  $\nabla^2(G*I) = (\nabla^2 G)*I$  [5] (pp. 57, 58), where  $\nabla^2$  is the Laplacian operator ( $\partial^2/\partial x^2 + \partial^2/\partial y^2$ ) [5] (p. 54),  $G$  is the 2D Gaussian function  $G(x,y)$ ,  $I$  is a 2D image intensity function  $I(x,y)$ ,  $(G*I)$  is a blurred image intensity function and  $\nabla^2 G$  is a circular (isotropic) even-symmetric Mexican-hat-shaped operator. This local filter features a center-surround configuration such that it is called *on-cell*, i.e., this even-symmetric operator is excited by stimulation in the central part of the receptive field and inhibited by stimulation in the outer part of the receptive field surrounding the center [22] (p. 17).

Noteworthy, because the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution [88]. Therefore, the filtered image  $(\partial^2 G/\partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  applied to the low-pass image adopted by both Canny [27] and Bertero, Torre and Poggio [44].

$$(\partial^2 G/\partial n^2 * I) \neq \partial^2/\partial n^2 (G*I) \quad (7-1)$$

In Marr's words, the raw primal sketch consists of the following two steps.

- (i) Detect (select) ZX pixels in the  $\nabla^2 G$ -filtered image at multiple spatial scales, where a ZX pixel is located where the value of the  $\nabla^2 G$ -filtered image passes from positive to negative [5] (p. 54). This must be intended, in a more general meaning, that a ZX pixel is located where the value of the  $\nabla^2 G$ -filtered image passes from positive to negative or vice versa, or from positive to zero or vice versa, or from negative to zero or vice versa. According to Marr, ZX pixels through scale must be dealt with according to the *spatial coincident assumption* [5] (p. 70), refer to Chapter 3.5.3. In his view, ZX pixels are not physical image contours (edges, [5], p. 68), but candidate pixels for the presence of image contours that must correspond to "physical contours". In the computer vision (CV) and remote sensing (RS) literature this phase is often called *image contour detection*, but it is important to remark that, at the raw primal sketch level of information processing, ZX pixels may have no physical meaning, i.e., they may belong to no physical boundary [5] (p. 68), e.g., refer to the *spatial coincidence assumption* in Chapter 3.5.3 [5] (p. 70). To accomplish ZX (pixel) detection (selection), Marr selects the isotropic even-symmetric Laplacian operator of a Gaussian filter,  $\nabla^2 G$ , as scalable second-order differential operator. Instead, in [1], the oriented even-symmetric real part of a complex Gabor filter is designed as non-isotropic even-symmetric local filter.
- (ii) According to Marr [5] (Figure 2-21, p. 73), the raw primal sketch employs as input the ZX pixels to generate as output the following intermediate products.
  - An intermediate information primitive called *ZX segment*, defined as "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" [5] (p. 60). Hence, it is at the level of detection of ZX segments that ZX pixels and contours turn into sub-symbolic discrete image-objects (polygons).
  - ZX segments must be "accounted for" through scale in compliance with the *spatial coincident assumption* [5] (pp. 70, 71), refer to Chapter 3.5.3.



- A discrete and finite set of multi-scale *tokens* (discrete sub-symbolic image plane entities), namely, *edges*, *blobs* (closed contours), *bars* and *terminators* (*discontinuities*), to be described in a discrete and finite multi-scale token description table, where token attributes are: position, orientation, contrast, length, width, etc., refer to Chapter 3.5.3.
- A bit map of the image to represent basic positional information of the tokens, where a map is a retinotopic representation of a single visual parameter according to Tsotos [68]. This bit map is 1 at the corresponding position of a token described in the token description table. This bit map is used for search of inter-token spatial relations, that are rather local in the perceptual grouping phase (e.g., Gestalt's law of proximity) [6], without the trouble of searching through the whole list of primal sketch descriptors for inter-token spatial relations [5] (p. 79), refer to Chapter 3.5.3.

In the CV and RS literature, the raw primal sketch is often called *image segmentation* (image-object detection, image partitioning into a discrete and finite set of image-objects). For example, at this stage *textons* (texture elements) are expected to be extracted as tokens. This means that at the information processing level of the raw primal sketch, texture boundaries are not detected, but only *textons* are identified (refer to [5], Figure 2-7, p. 53). Unfortunately, in his seminal work, Marr proposes no algorithm to extract *ZX* pixels, *ZX* segments and tokens from *ZX* segments. According to Li Zhaoping [6], "the computer vision community has tried to solve the problem of image segmentation for decades without a satisfactory solution." To recapitulate, generation of a raw primal sketch is still an open challenge in the CV and RS literature.

In recent years the computer vision community has increasingly adopted semi-automatic inductive data learning algorithms for color-homogeneous *superpixel* detection as a prior step in many high-level computer vision tasks [61]. In the computer vision literature, superpixel detection is an image pre-processing first stage employed for image simplification as input to high-level vision tasks. It is required to be fast to compute, memory efficient, simple to use by featuring few and intuitive input parameters to be user-defined, and capable of increasing the speed and quality of the higher-level vision tasks [61]. To the best of these authors' knowledge, the concept of superpixel is not novel, but highly related to the Julesz's *texton/textel* theory of pre-attentive human vision developed back in the 1970's [2], [3], [58], [59], [60]. If this conjecture holds, terms superpixel and *textel/texton* are synonyms and related to the Marr's raw primal sketch in pre-attentive vision [4].

To summarize, Marr's *ZX* segments / tokens, providing the raw primal sketch generated as output by the low-level vision first-of-two phases, are equivalent to so-called *textons* or *texels* (texture elements), investigated by neurophysiology and psychophysics [2], [3], [58], [59], [60], otherwise called *superpixels* by computer vision literature in recent years [61].

- (II) *Full primal sketch* [5] (p. 91, Figure 2-7 in p. 53), also called *perceptual organization* or *perceptual grouping* [6], to form larger-scale tokens that reflect larger-scale spatial structures (distributions) of elementary tokens in the image, e.g., texture boundaries / texture segmentation [5] (p. 96). Unfortunately, in his seminal work Marr proposed no explicit (complete, well-defined) algorithm to generate the full primal sketch from the discrete token description table in a constructive (iterative, bottom-up) way [5] (p. 52). On the other hand, Marr states explicitly that, to investigate the *spatial organization* of tokens in an image, elsewhere called *spatial distribution* of tokens (p. 79), *perceptual grouping* (p. 91) or *texture discrimination* (p. 96), *discontinuities* (to be intended as synonyms of "abrupt changes" or "singularities") in six image parameters (quantitative properties, numerical features) must be investigated. Three of them are intrinsic to a token (token-specific) and three pertain to the spatial arrangement of tokens.
- (i) Token-specific metrological attributes, affecting perceptual grouping of tokens.
- i. Average achromatic intensity (brightness, where brightness is defined as perceived luminance [1]) or (chromatic) color.



- ii. Geometric properties, i.e., shape properties (e.g., length, width, compactness, rectangularity, roughness/straightness of boundaries, simple connectivity, etc.) and orientation. For example, refer to [29].
- iii. Size.
- (ii) Spatial arrangement of tokens. If there is a discontinuity in any of the following attributes, then there is a change in perceived texture, i.e., there is a texture boundary.
  - ii. Local density of tokens.
  - iii. Distance apart of tokens. Estimated by the so-called Steven's algorithm for recovering the local orientation of tokens [30], based on an information primitive called, by Marr, inter-token *virtual line* (p. 82).
  - iv. Local orientation of tokens, also estimated by the Steven's inter-token *virtual line* detection algorithm [30].

To recapitulate, generation of a full primal sketch is still an open challenge in the CV and RS literature. In his seminal work, Marr proposes no well-defined algorithm to generate the full primal sketch from the discrete token description table in a constructive (iterative, bottom-up) way [5] (p. 52). In more recent works, such as [6], a primary visual cortex (V1) hypothesis for creating a bottom-up saliency map for pre-attentive selection and segmentation capable of detecting texture boundaries is proposed.

- (5) Provide stratified (masked) multi-scale texture features useful for symbolic per-pixel context-sensitive image understanding. In CV, pixel-based classification (interpretation) is called semantic segmentation [95] or image parsing [96]. An image mask is equivalent to a candidate area. Stratified multi-scale texture features must be insensitive to “artificial” mask boundaries between data/no-data image subsets.

### 3.8 Test image sets

The selected test data set consists of both natural and synthetic panchromatic images suitable for testing texture boundary detection and image contour detection algorithms.

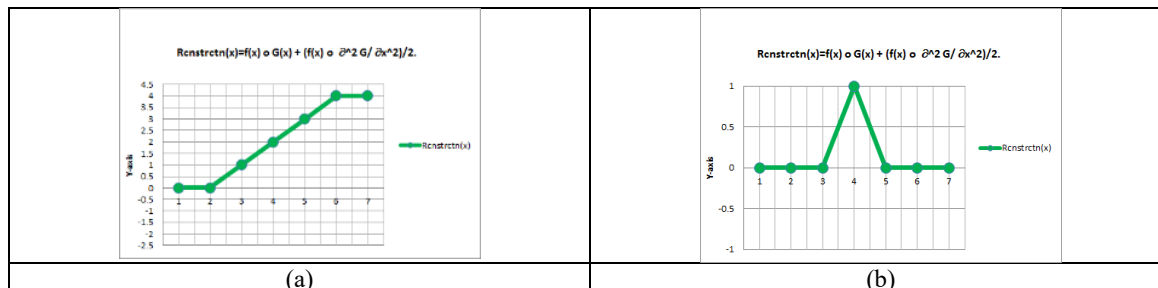
Any image is a 2D regularly gridded function where a combination of four spatial shapes occurs [1].

1. Flat areas.
2. Ramps. Mandatory to test the Mach bands illusion [25].
3. Step edges.
4. Lines.

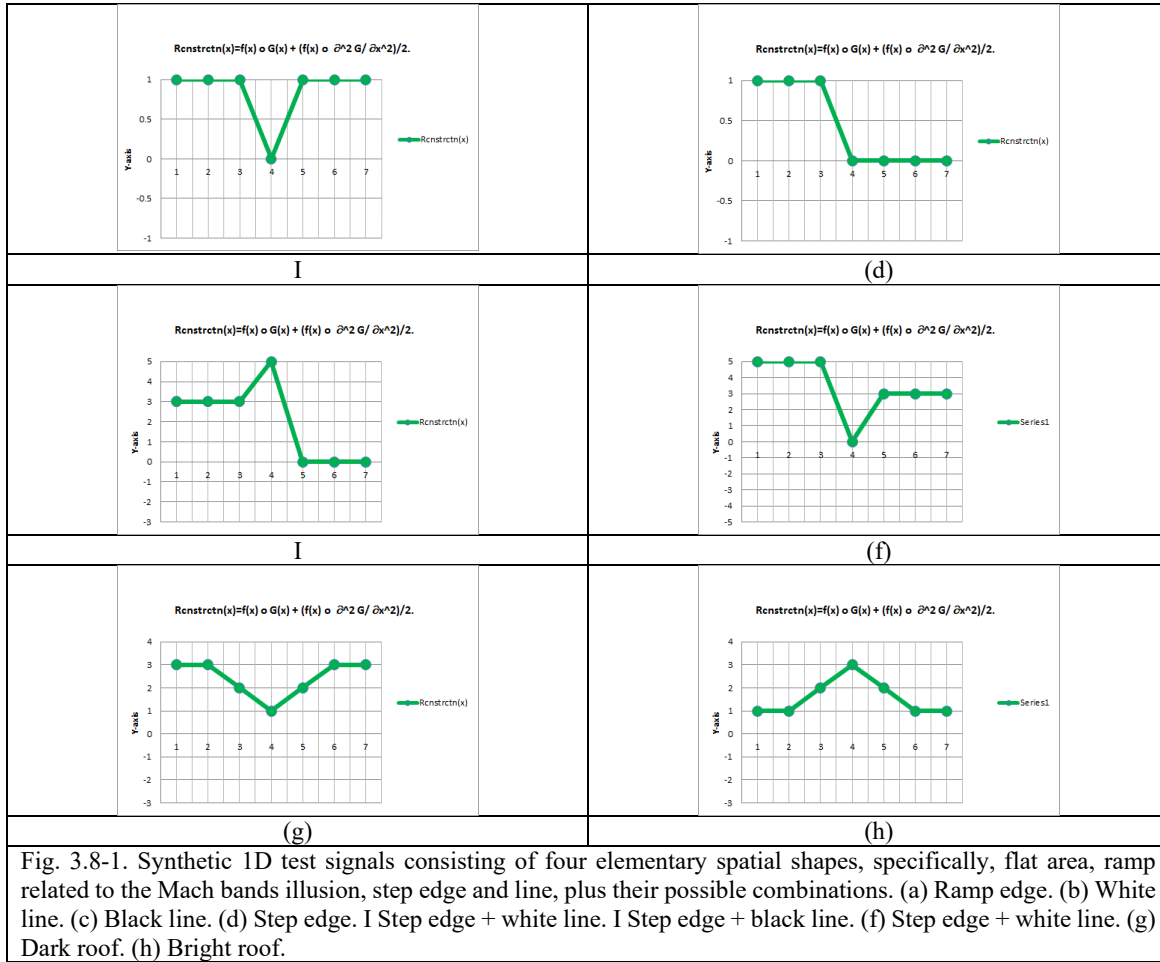
A synthetic image of controlled complexity is “simpler” than a natural image of unknown (unbounded above) complexity. A panchromatic image is “simpler” (less informative) than a color image.

According to the CV project requirements proposed in Chapter 3.4 and Chapter 3.6, no CV system should be tested on complex color images of natural scenes if it fails on simpler test cases of increasing signal complexity, specifically, 1D synthetic signals, synthetic panchromatic (2D) images and synthetic color images of controlled (known) complexity.

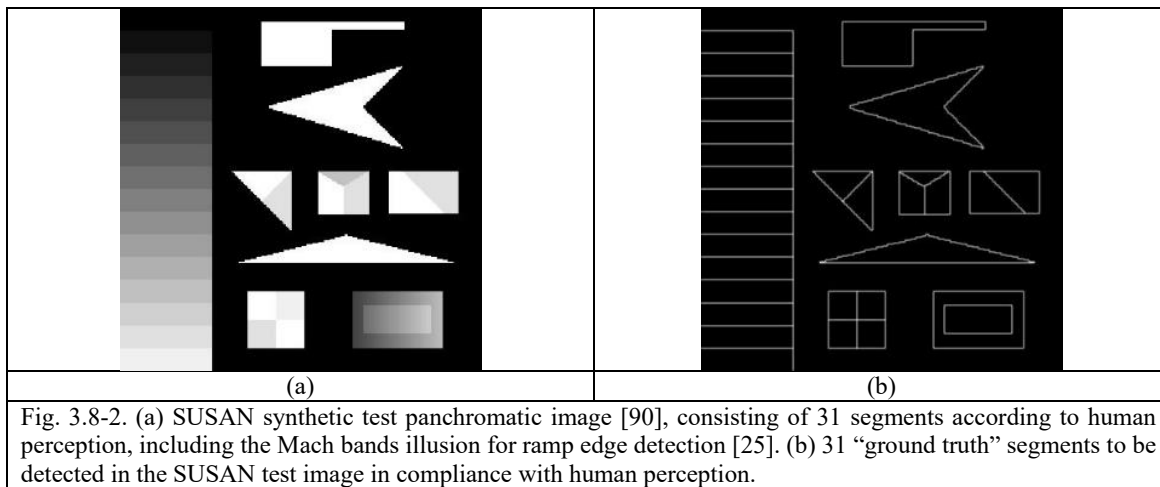
- Synthetic 1D test signals consisting of four elementary spatial shapes, (i) flat area, (ii) ramp related to the Mach bands illusion, (iii) step edge, and (iv) line, plus their possible combinations, see Fig. 3.8-1.





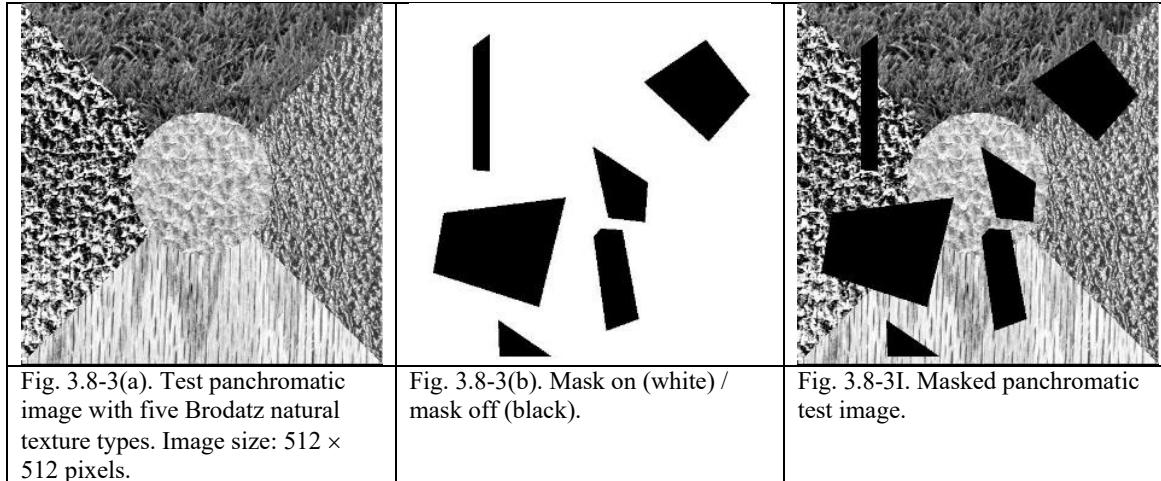


- Synthetic (2D) images for testing image-contour detection/image segmentation algorithms, such as the SUSAN synthetic panchromatic image where there are 31 segments accounting for the ramp-edge effect (Mach bands illusion), see Fig. 3.8-2.

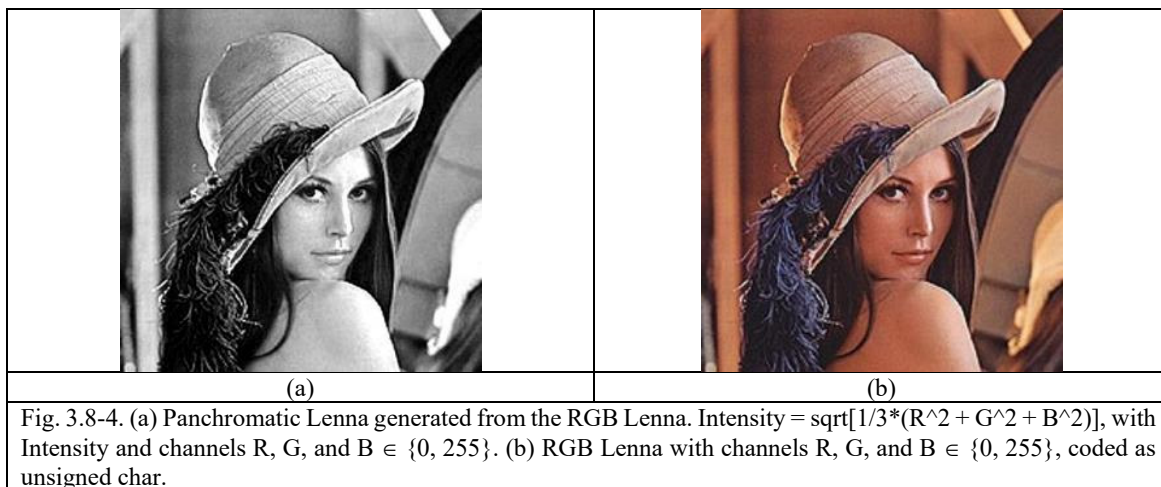


- ATTENTION. Any image-contour detection/image segmentation implementation is required to be submitted to the following test benchmark (necessary not sufficient condition).

- Detect the exact 31 segments in the SUSAN test image without any user interaction, i.e., automatically. If and only if this CV system requirement is satisfied, then move on in testing the CV system with other test images, including natural images.
- An artificial mosaic of natural panchromatic images consisting of five Brodatz natural texture types is shown in Fig. 3.8-3(a). Natural images are typically intuitive to interpret by a human photointerpreter, i.e., their “ground truth” is “objective” and agreed-upon by all possible photointerpreters.



- Additional panchromatic natural (real-world) images for testing contour detection algorithms, e.g., panchromatic Lenna, see Fig. 3.8-4(a). Natural images are typically intuitive to interpret by a human photointerpreter, i.e., their “ground truth” is “objective” and agreed-upon by all possible photointerpreters.
- Natural color images, e.g., RGB Lenna, see Fig. 3.8-4(b), whose “ground truth” (semantic meaning as a result of an *information-as-data-interpretation* process) is “objective” (ultimate, unquestionable, unambiguous) and agreed-upon by (familiar to) all possible photointerpreters.



### 3.9 Original automated statistical model-based color constancy algorithm

In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions, so that they appear identical to a “canonical” (reference) image, subject to a “canonical” (known) light source (of controlled quality), e.g., under a white light source [112]. In short, solution of the color constancy problem is the recovery “of an illuminant-independent representation of the reflectance values in a scene” [112]. Color constancy



is intrinsically related to brightness in a more global (large scale) sense than, say, image-contour detection, which depends on local non-stationary image properties. Although the biological mechanisms involved with the color constancy ability are not yet fully understood [22], [112], [113], [129], [131], [144], it is speculated that a special type of retinal ganglion cells can play a role in the estimation of a global “background” brightness, on which lines, ramps and step edges [1] can be projected [24]. In principle, this special type of retinal ganglion cells can be involved with brightness perception because: (i) it features a very large receptive field. (ii) It is not connected to either rods or cones. (iii) It is connected to central brain areas for controlling the circadian clock (day-night rhythm). (iv) Via a feedback loop, it is connected to the eye’s iris (pupil size). (vi) These special retinal cells also connect to at least the ventral area of the lateral geniculate nucleus [24].

In agreement with the QA4EO *Cal* requirements [111], considered mandatory when radiometric calibration metadata parameters are acquired onboard together with the image, color constancy should be considered mandatory to guarantee uncalibrated image harmonization and interoperability when no calibration metadata parameters are available. In common practice, computational color constancy is a fundamental prerequisite of many CV applications to guarantee harmonization of images acquired under varying illumination conditions.

Also because biophysical mechanisms of color constancy remain largely unknown, computational color constancy algorithms are typically unable to simulate color constancy effects observed in humans [49]. Computational color constancy is an under-constrained problem in the Hadamard sense [75]. Since it does not have a unique solution, it requires *a priori* knowledge in addition to data for numerical treatment [76]. For these reasons there has been a large number of alternative color constancy algorithms proposed in the computer vision literature in the last 30 years [129]. Early computational models were derived from works on human perceptual theory, resulted in the Retinex perceptual theory by Land [130], considered inadequate by now. In survey works such as [22], [112] and [129], computational color constancy approaches are divided into three categories. (1) Low-level statistical model-based methods. (2) Physical model-based methods. (3) Gamut-based methods. In statistical models, the best-known statistical assumption about color distributions is the so-called *Grey-World* assumption: the average reflectance in a scene under a neutral light source is achromatic, which means that the color of the light source can be estimated by computing the average color in the image. Another well-known assumption is the *White-Patch* assumption: the maximum response in the RGB channels is assumed to be caused by a perfect reflector. The assumption of perfect reflectance is alleviated by considering the color channels separately, resulting in the *max-RGB* assumption, where the illuminant is estimated as the maximum response in each color channel separately. In common practice, if an image contains few edges (corresponding to a “low” signal-to-noise ratio), then pixel-based (context-insensitive, 1<sup>st</sup>-order spatial distribution) methods, like *Grey-World* and *White-Patch*, are preferred. When the signal-to-noise ratio is “medium” or “high”, context-sensitive (edge-based) methods are preferred: for example, 1<sup>st</sup>- and 2<sup>nd</sup>-order local spatial derivatives are adopted in the so-called *Grey-Edge* method [131]. In [129], the eleven “universal” human basic color (BC) categories, identified by Berlin and Kay in their cross-cultural survey of color names [132], were adopted as a form of *a priori* knowledge, available in addition to and independent of data, to improve the computational efficiency of the color-by-correlation algorithm proposed in [112]. In the statistical color constancy algorithm implemented in the Environment for Visualizing Images (ENVI) commercial software toolbox [133], “ENVI does a special (‘ultimate’) stretch for the display case, which really can’t be reproduced using the ENVI’s default 2% stretch. If the histogram features more than three bins, the special stretch will calculate the left hand percent stretch on `hist[1:*]` and the right hand percent stretch on `hist[0:n_elements(hist)-2]`. If the histogram is a Gaussian (normal) shaped curve, then the difference between this and the ‘full’ histogram is negligible. However, if there is a large saturation of min or max values (such as an image with a lot of background, e.g., water pixels, and/or foreground, e.g., cloud pixels), then ENVI’s default stretch will ignore the spike(s) and calculate the percent linear stretch, e.g., a 2% stretch, based on the rest of the “real” histogram. This allows ENVI to display, by default, many images which otherwise would not stretch well with a 2% linear stretch since they contain more than 2% background and/or foreground” [134].

For a complete survey of computational color constancy methods, the interested reader can refer to the existing literature, e.g., [22], [112], [113], [129], [131], [144].

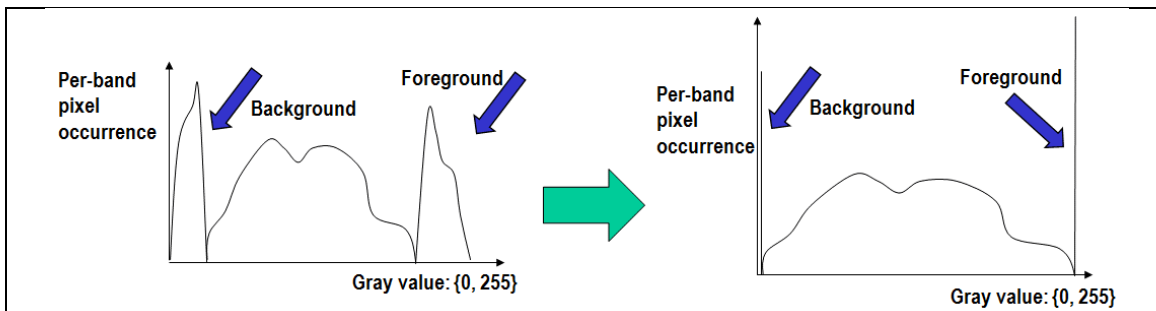


Fig. 3.9-1. Example of a gray-value univariate (one-channel) 1<sup>st</sup>-order distribution, to be stretched for color constancy. Four categories of univariate distributions can be considered: (i) neither a background nor a foreground mode is present in addition to a central mode; (ii) only a background mode with a right tail can be identified, (iii) only a foreground mode with a left tail can be identified, and (iv) both background and foreground modes are present in addition to a central mode.

The present author designed and implemented an original self-organizing statistical algorithm (never published, patent pending) for 1<sup>st</sup>-order histogram-based (non-contextual) image color constancy in linear time  $\leq O(I \times N \times B)$ , where the number of learning-from-data iterations  $I$  is  $\leq 3$ . It was inspired by the ENVI “ultimate” image stretching algorithm summarized as above [134]. Our solution analyzes each single channel to detect one-of-four 1<sup>st</sup>-order histogram distributions, described as follows, see Fig. 3.9-1. (i) Neither a background nor a foreground mode is present in addition to a central mode. (ii) A background mode with a long right tail can be identified. (iii) A foreground mode with a long left tail can be identified. (iv) One background and one foreground mode can be identified in addition to a central mode. Once background and foreground modes are detected, if any, they are mapped onto the minimum output gray value,  $\text{hist}[0]$ , and the maximum output gray value,  $\text{hist}[255]$ , respectively. Next, a standard linear stretching algorithm is applied per channel, to fill the histogram bins  $\text{hist}[1:254]$ , according to a traditional max-RGB criterion, see Fig. 3.9-2.

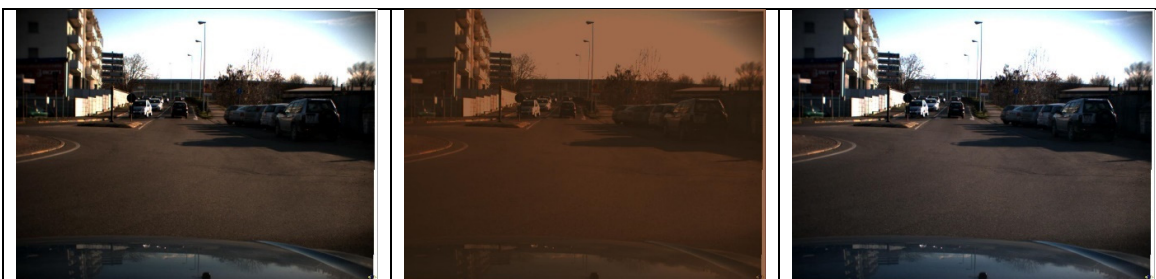


Fig. 3.9-2. Left: True color quick-look RGB (3-band, in the .tif file format). Courtesy of VisLab srl, c/o Dipartimento di Ingegneria dell’Informazione, Autonomous driving vehicle control systems, U. of Parma, Italy. Center: left image subject to additive and multiplicative distortion. Specifically:  $\text{byte}((R/2 + 30) < 255)$ ,  $\text{byte}((G/3 + 20) < 255)$ ,  $\text{byte}((B/4 + 10) < 255)$ . Right: center image subject to the proposed original automated color constancy algorithm.

### 3.10 Original expert systems for automated color naming in a calibrated MS reflectance space or in an uncalibrated RGB color space, either true- or false-color

Bridging independent studies on color naming conducted by linguistics and computer vision [78], the seminal work by Griffin proved the hypothesis that the best partition of a monitor-typical red-green-blue (RGB) data cube into color categories for pragmatic purposes coincides with human basic colors (BCs) identified by Berlin and Kay [132], see Fig. 3.10-1. Central to this consideration is Berlin and Kay’s landmark cross-cultural survey of color names in 20 human languages. On the basis of that study they claimed that the basic color terms of any given language are always drawn from a universal inventory of eleven color names: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown [132]. These perceptual BC categories are expected to be “universal”, i.e., general-purpose, user- and application-independent. In common practice, users are not required to learn a new color representation for ever-varying sensory data sets, but they can apply the same universal color representation independently of the image understanding problem at hand [144].

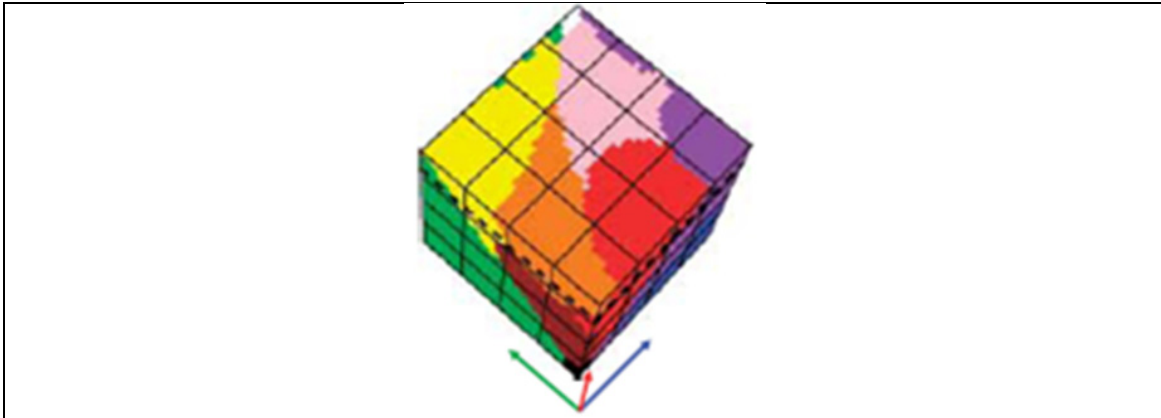


Fig. 3.10-1. Reproduced with permission, courtesy of [78]. Perceptual continuous color space discretization (quantization) into a prior knowledge-based (deductive, top-down) dictionary of color names. Example of RGB cube polyhedralization proposed by Griffin: prior 112 polyhedral, corresponding to RGB color names, can be either convex or not, either connected or not [78].

As shown in Fig. 3.10-1, color naming is synonym of static (prior knowledge-based, not adaptive to data) color space polyhedralization. If a 3D to 2D color space dimensionality reduction is accomplished, for example by transforming a monitor-typical RGB cube into a 3D Munsell color space, whose dimensions are (i) Intensity  $\in [0$  (black), 10 (white)], (ii) Hue  $\in [0-360^\circ]$ , and (iii) Chroma = Saturation  $\in [0$ , no upper bound), see Fig. 3.10-2, and, next, by projecting the 3D Munsell color space onto a so-called 2D Munsell grid, then alternative color name dictionaries for color space partitioning become easier to visualize and more intuitive to think of, see Fig. 3.10-3 [78], [132], [171], than in the 3D RGB cube shown in Fig. 3.10-1.

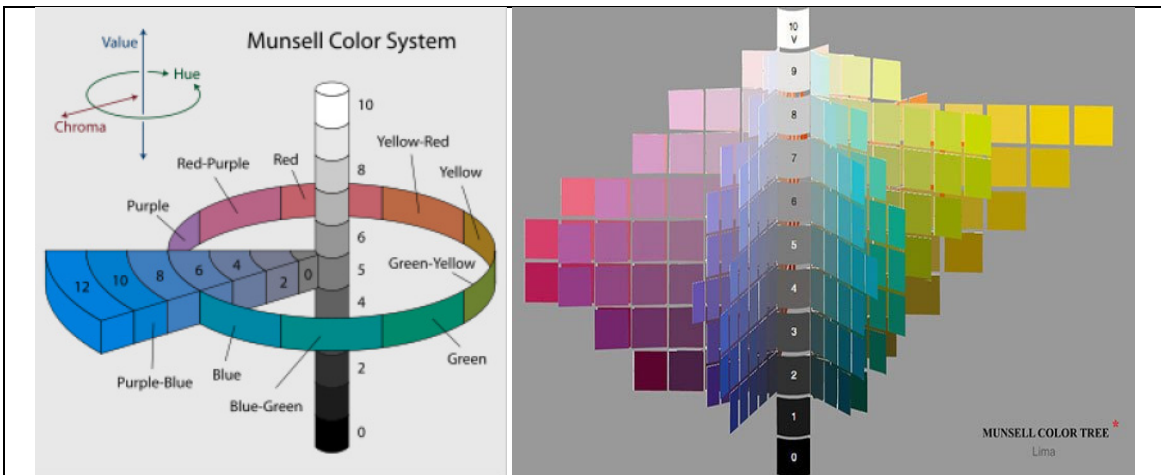


Fig. 3.10-2. The 3D Munsell color system also known as Munsell color tree [78], [132], [171]. Color dimensions are: (i) Intensity  $\in [0$  (black), 10 (white)]. (ii) Hue  $\in [0-360^\circ]$ . Munsell divided each horizontal hue circle  $\in [0-360^\circ]$  into five principal hues: Red, Yellow, Green, Blue, and Purple, along with 5 intermediate hues (e.g., YR) halfway between adjacent principal hues. (iii) Chroma = Saturation  $\in [0$ , no upper bound). It represents the “purity” of a color (related to saturation), with lower chroma being less pure (more washed out, as in pastels). Note that there is no intrinsic upper limit to chroma. Different areas of the color space have different maximal chroma coordinates. This led to a wide range of possible chroma levels—up to the high 30s for some hue–chroma value-pair combinations. For example, vivid bare soil colors are in the chroma range of approximately 8.



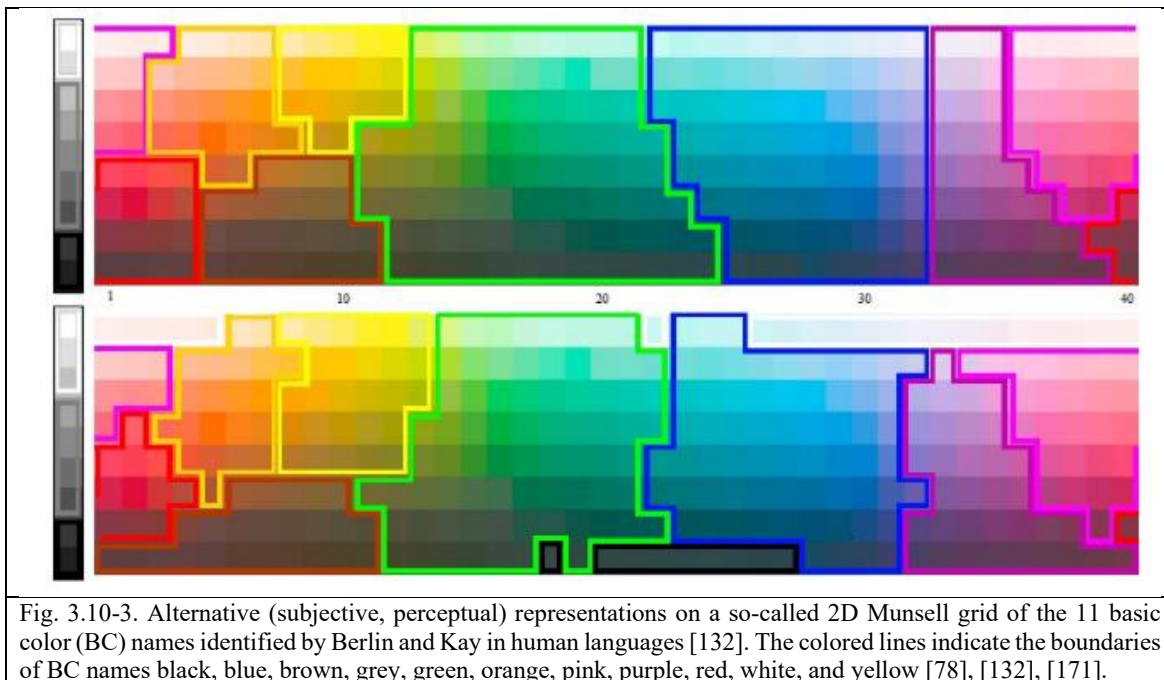


Fig. 3.10-3. Alternative (subjective, perceptual) representations on a so-called 2D Munsell grid of the 11 basic color (BC) names identified by Berlin and Kay in human languages [132]. The colored lines indicate the boundaries of BC names black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow [78], [132], [171].

Intended as synonym of static color space polyhedralization, prior knowledge-based color space partitioning is the automatic deductive counterpart of inherently ill-posed, semi-automatic and site-specific [42] inductive data learning algorithms for vector quantization (VQ) [76], [79], such as the popular k-means algorithm, also known as Lloyd's or Linde-Buzo-Gray's VQ algorithm [145], [146], or the Iterative Self-Organizing Data Analysis Technique (ISODATA) algorithm [147], which is one of the most widely used VQ heuristics in RS and computer vision applications. When they adopt a Euclidean metric distance minimization criterion and they reach convergence, inductive VQ algorithms accomplish a convex Voronoi tessellation of the input vector space, which is a special case of convex polyhedralization [145]. Because they detect convex hyperpolyhedra, inductive VQ algorithms are unsuitable for detecting unlabeled data clusters when their number, shape and size can be any [69], [145]. Since they are statistical data models, inductive VQ algorithms can be input with unlabeled data provided with no physical meaning. Available for selection in the most popular open source and commercial EO image processing software toolboxes, inductive VQ algorithms, such as k-means and ISODATA, are widely employed by the RS community. Unfortunately, they are typically input with uncalibrated EO data, in disagreement with the QA4EO Cal requirements [111].

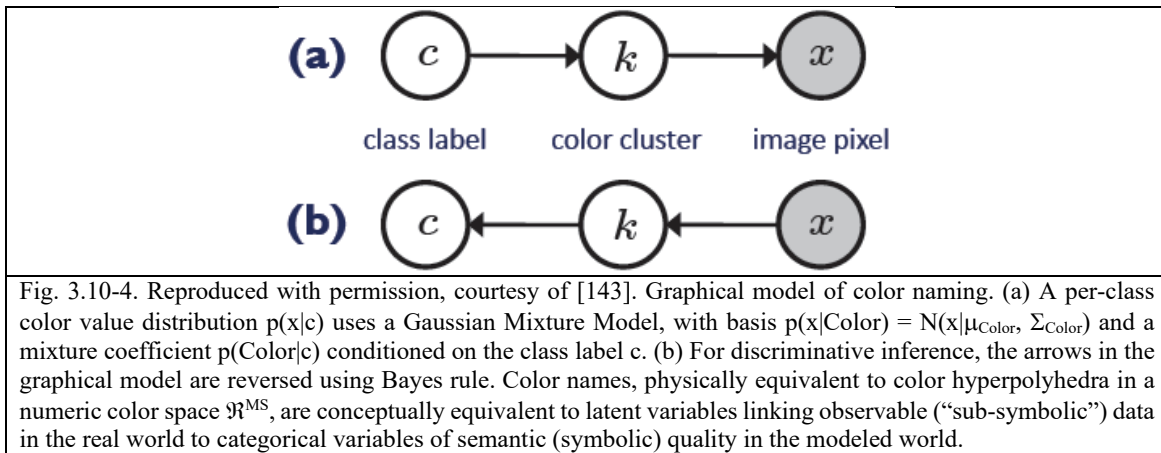
Color naming is an ongoing field of research by the CV community [22], [36], [37]. In CV algorithms, it is clearly acknowledged that individual objects tend to have a relatively compact color distribution, so that exploitation of a numeric color discretization model for color naming, capable of describing image objects by community-agreed discrete and finite color names, can dramatically improve image segmentation and classification where color non-contextual properties can be combined with contextual spatial topological and non-topological visual properties according to a convergence-of-evidence approach, e.g., refer to Fig. 3.6-6. For example, in [143] per-class color models  $p(x|Color)$  are represented as Gaussian Mixture Models (GMMs) in a CIELab color space  $x$  where the Gaussian mixture coefficients  $p(Color|c)$  depend on the class label  $c$ , with  $c = 1, \dots, ObjectClassLegendCardinality$ . In particular, according to a maximum a posteriori (MAP) classification approach of a Bayesian theorem (Bayesian law):

$$p(c|x) = \frac{p(x|c)p(c)}{\sum_{c1=1}^{Classes} p(x|c1)p(c1)} \propto p(x|c)p(c) = \left[ \sum_{Color=1}^{ColorDictionary} p(x|Color)p(Color|c) \right] p(c), \quad c = 1, \dots, ObjectClassLegendCardinality, \quad (10-1)$$

where  $p(x|Color) = N(x|\mu_{Color}, \Sigma_{Color}) \in [0, 1]$  is modeled as a fuzzy convex Gaussian distribution with mean =  $\mu_{Color}$  and variance =  $\Sigma_{Color}$ . According to this equation where the same color cluster can be shared between different classes while

the same class can feature several colors in the color dictionary, only term  $P(\text{Color}|c)$  depends on the class label, which makes this color model more efficient to learn than a separate GMM for each class. The graphical model depicted in Fig. 3.10-4 shows (a) a per-class color value distribution  $p(x|c)$  modeled as a Gaussian Mixture with basis  $p(x|\text{Color}) = N(x|\mu_{\text{Color}}, \Sigma_{\text{Color}}) \in [0, 1]$  and a mixture coefficients  $P(\text{Color}|c) \in [0, 1]$  conditioned on the class label  $c$ . (b) For discriminative (classification) inference, to assess  $p(c|x)$ , the arrows in the graphical model are reversed using Bayes' rule.

“Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. The critical point about Bayesian inference is that it provides a principled way of combining new evidence with prior beliefs, through the application of Bayes' rule  $p(c|x) = \frac{p(x|c)p(c)}{\sum_{c_1=1}^{\text{ObjectClassLegendCardinality}} p(x|c_1)p(c_1)}$ ,  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ . This is in contrast with frequentist inference, which relies only on the evidence  $p(x|c)$  as a whole, with no reference to prior beliefs. Furthermore, Bayes' rule can be applied iteratively: after observing some evidence, the resulting posterior probability can then be treated as a prior probability, and a new posterior probability computed from new evidence. This allows for Bayesian principles to be applied to various kinds of evidence, whether viewed all at once or over time. This procedure is termed ‘Bayesian updating’” [153].

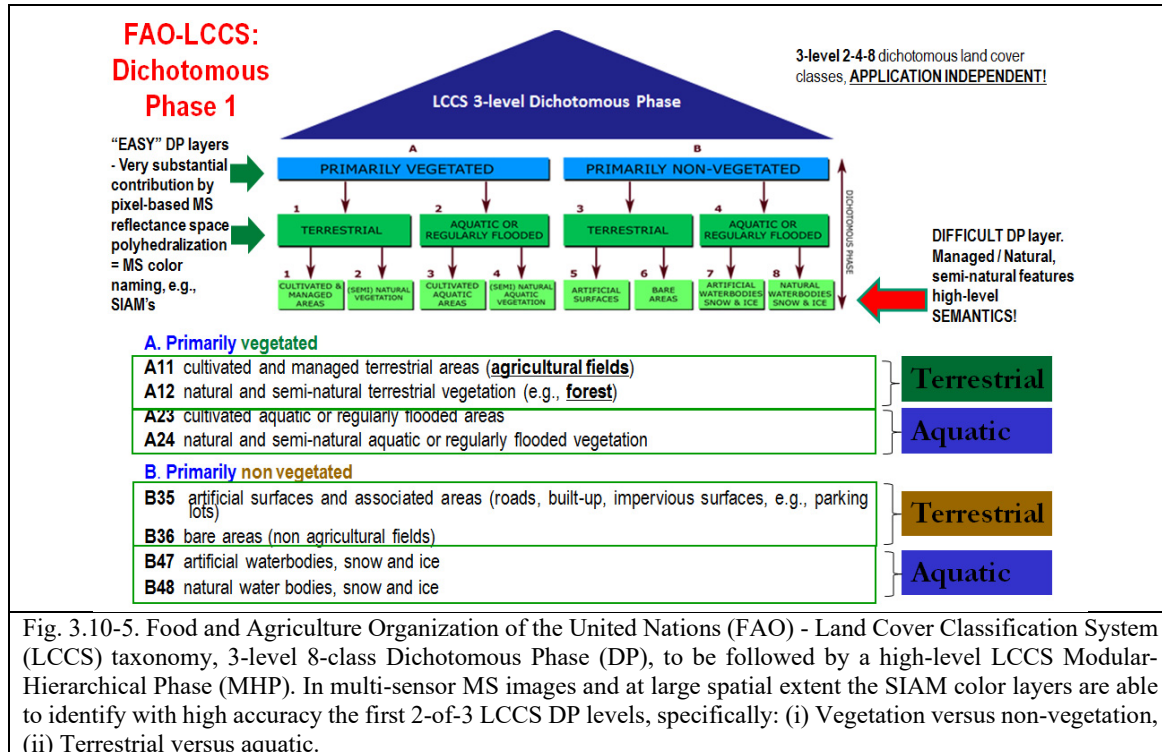


Unlike the fuzzy GMM adopted in [143], given a spatial unit  $x$ , where  $x$  is either (0D) pixel, (1D) line or (2D) polygon, a prior knowledge-based color space polyhedralization is equivalent to a crisp membership function  $m(\text{ColorValue}(x)|\text{ColorName}) \in \{0, 1\}$  with  $\text{ColorName} \in \{1, \text{ColorDictionaryCardinality}\}$ , where the following considerations hold.

- A numeric color value belonging to a MS data space,  $\text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , is a numeric attribute associated with spatial unit  $x$ .
- A numeric variable  $\text{ColorValue}(x)$  is mapped onto a single color name  $\text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}$ , corresponding to one hyperpolyhedron neither necessarily convex nor connected in the color space  $\mathfrak{R}^{\text{MS}}$ , such that  $m(\text{ColorValue}(x)|\text{ColorName}^*) = 1 = p(\text{ColorName}^*|\text{ColorValue}(x))$ .
- As a consequence,  $\forall \text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}, \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x)|\text{ColorName}) = m(\text{ColorValue}(x)|\text{ColorName}^*) = 1$  holds, where  $m(\text{ColorValue}(x)|\text{ColorName}) \in \{0, 1\}$ ,  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ .
- Noteworthy, equality  $m(\text{ColorValue}(x)|\text{ColorName}^*) = 1 = p(\text{ColorName}^*|\text{ColorValue}(x))$  always holds true together with membership function  $m(\text{ColorName}^*|c) \in \{0, 1\}$ , with  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ .
- Eq. (10-1) shows that color names, physically equivalent to color hyperpolyhedra in a numeric color space  $\mathfrak{R}^{\text{MS}}$ , are levels of a “semi-symbolic” categorical variable conceptually equivalent to a latent/hidden variable linking observable (“sub-symbolic”) data in the real world, specifically, univariate or multivariate  $\text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , to a categorical variable of semantic (symbolic) quality in the modeled world,  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ .

If these simplifications are adopted, then the following Eq. (6-6) holds in a naïve Bayes classification approach, refer to Chapter 3.6.

$$\begin{aligned}
 & m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\
 & \min \left\{ \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x)|\text{ColorName})m(\text{ColorName}|c), m(\text{ShapeValue}(x)|c), \right. \\
 & \left. m(\text{TextureValue}(x)|c), m(\text{SpatialRelationships}(x, \text{Neigh}(x))|c) \right\} = \\
 & \min \{m(\text{ColorName}^*|c), m(\text{ShapeValue}(x)|c), m(\text{TextureValue}(x)|c), m(\text{SpatialRelationships}(x, \text{Neigh}(x))|c)\}, \\
 & c = 1, \dots, \text{ObjectClassLegendCardinality}, \text{ where } \text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}, \text{ such that} \\
 & m(\text{ColorValue}(x)|\text{ColorName}^*) = 1 \text{ and } m(\text{ColorName}^*|c) \in \{0, 1\}. \tag{6-6}
 \end{aligned}$$



As reported in Chapter 3.6, Eq. (6-6) shows that for any spatial unit  $x$  in the image-domain, when a hierarchical CV classification approach estimates posterior  $m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x)))$  starting from a near real-time context-insensitive color naming first stage where condition  $m(\text{ColorValue}(x)|\text{ColorName}^*) = 1$  holds, if condition  $m(\text{ColorName}^*|c) = 0$  is true according to a static community-agreed binary relationship  $R: \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  from set  $A = \text{DictionaryOfColorNames}$  to set  $B = \text{LegendOfObjectClassNames}$  (and vice versa), where static relationship  $R: A \Rightarrow B$  is known *a priori*, see Table 3.6-1, i.e., it is equivalent to *prior belief* employed in the assessment of posterior  $m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x)))$ , then posterior  $m(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) = 0$  irrespective of any second-stage assessment of spatial terms  $\text{ShapeValue}(x)$ ,  $\text{TextureValue}(x)$  and  $\text{SpatialRelationships}(x, \text{Neigh}(x))$ , whose computational model is typically difficult to find and computationally expensive. Intuitively Eq. (6-6) shows that static color naming allows the stratification of unconditional multivariate spatial variables into color class-conditional data distributions, in agreement with the statistic stratification principle [122] and the divide-and-conquer problem solving approach [79], [76]. Well known in statistics, the principle of statistic stratification guarantees that "stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest" [122].

As shown in Fig. 3.10-1, a true- or false-color RGB color space polyhedralization is intuitive to think of and easy to visualize. On the contrary, a MS reflectance space polyhedralization, with a number of spectral bands  $MS > 3$ , is very difficult to think of and impossible to visualize.

Two static decision-tree lightweight computer programs in operating mode capable of static RGB monitor-typical color space polyhedralization and static MS reflectance space polyhedralization called, respectively, RGB Image Automatic Mapper (RGBIAM™) and Satellite Image Automatic Mapper™ (SIAM™) [38], [56], [57], [148], [149], [150], [151], [154], are summarized hereafter.

### 3.10.1 The SIAM lightweight computer program for MS reflectance space hyperpolyhedralization, superpixel detection and VQ quality assurance

Presented in the RS literature in recent years [38], [56], [57], [148], [149], [150], [151], the SIAM software toolbox in operating mode is an expert system (prior knowledge-based decision tree) for physical model-based/deductive/top-down VQ and VQ quality assessment in a MS reflectance space, capable of automated (no user-machine interaction is required) color naming [22], [36], [37], superpixel detection in the color map-domain and per-pixel VQ quality assessment in near real-time (computational complexity increases linearly with image size).

According to the existing literature [154], SIAM provides a very substantial contribution to the detection of the first 2-of-3 levels of the FAO Land Cover Classification System (LCCS) Dichotomous Phase (DP) taxonomy, specifically: (i) vegetation/non-vegetation and (ii) terrestrial/aquatic, based on context-insensitive spectral properties exclusively, see Fig. 3.10-5.

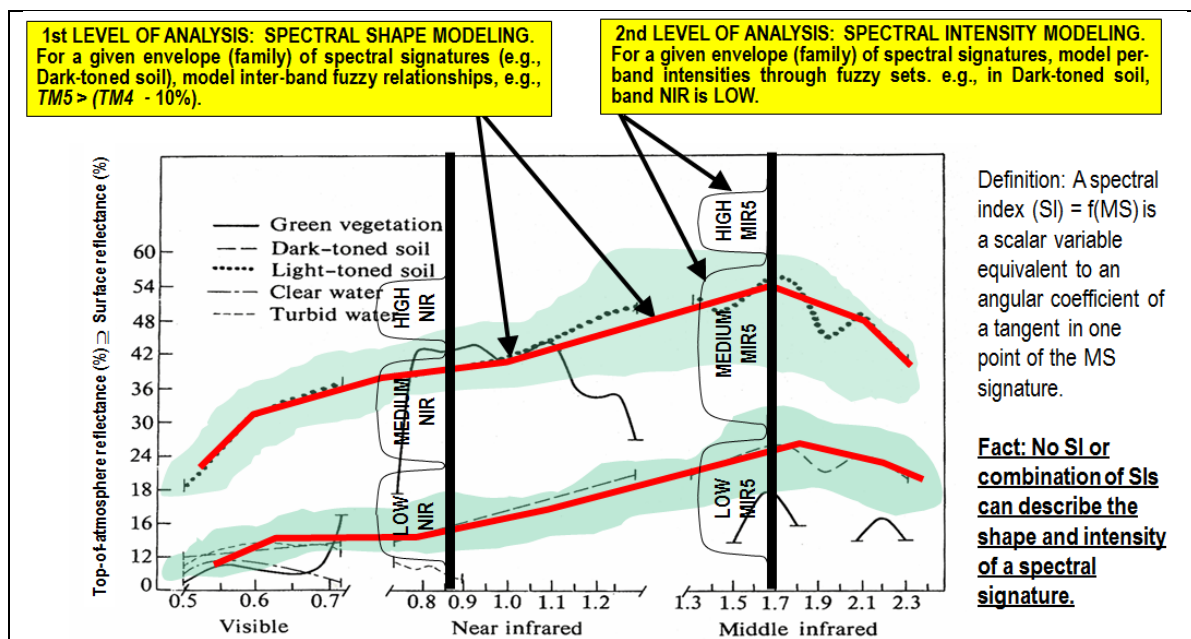


Fig. 3.10-6. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values which include surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions [158]. A within-class family of spectral signatures (e.g., Dark-toned soil) in TOARF values forms a buffer zone (support area, envelope). SIAM models each target family of spectral signatures in terms of multivariate shape and multivariate intensity as a viable alternative to univariate analysis of spectral indexes (SIs). Since a typical SI is a scalar band ratio equivalent to an angular coefficient of a tangent in one point of the spectral signature, it is an unquestionable fact that no univariate SI or multivariate combination of SIs can reconstruct the shape and intensity of a spectral signature. For example, based on common sense, infinite functions can feature the same tangent value in one point.

To improve LC classification accuracy, reliability and informativeness, spectral-based LC classification evidence provided by SIAM should not be employed on a standalone basis, but must be combined with additional evidence stemming from





spatial information, either topological (e.g., adjacency, inclusion) or non-topological (e.g., spatial distance, angle measure), which typically dominates color information, according to a convergence-of-visual-evidence approach, see Fig. 3.6-6.

Since it is a physical data model, SIAM requires as input Earth observation (EO) data provided with a physical meaning, specifically, EO data radiometrically calibrated (*Cal*) into either top-of-atmosphere reflectance (TOARF) or surface reflectance (SURF) values where, for every land cover (LC) class in the real-world domain, SURF values are a special case of TOARF values in clear sky and flat terrain conditions [158], i.e.,  $TOARF \supseteq SURF$  values, see Fig. 3.10-6.

To provide a partition of a MS reflectance space into a static dictionary of color names, SIAM models each target family of spectral signatures in terms of multivariate shape and multivariate intensity as a viable alternative to the univariate or multivariate analysis of spectral indexes typically adopted by a large majority of the RS community in single-date or multi-temporal MS image classification problems, see Fig. 3.10-6.

In a MS reflectance space, any target family of LC class-specific spectral signatures is a multivariate data distribution (envelope, hyperpolyhedron, manifold), see Fig. 3.10-6. Like a vector quantity has two characteristics, a magnitude and a direction, any MS manifold is characterized by a multivariate shape and a multivariate intensity, see Fig. 3.10-6. It is an unquestionable fact that no multivariate spectral index can model both the multivariate shape and the multivariate intensity of a target family of LC class-specific spectral signatures. Any scalar spectral index, either a normalized difference, e.g., the popular normalized difference vegetation index,  $NDVI = (NIR - Red) / (NIR + Red)$ , with  $NDVI \in [-1, 1]$ , where  $NIR = \text{Landsat band 4}$  and  $Red = \text{Landsat band 3}$  or an unbounded band ratio, e.g., a vegetation ratio index ( $VRI = NIR/Red \geq 0$ ), is conceptually equivalent to the slope of a tangent to the spectral signature in one point, see Fig. 3.10-7. This spectral slope is a MS shape descriptor independent of the MS intensity, i.e., infinite functions with different intensity values can feature the same tangent value in one point. Although appealing due to its conceptual and numerical simplicity [77], any scalar spectral index is unable *per se* to represent either the multivariate shape information or the multivariate intensity information of a spectral signature. For example, it is well known that any vegetation spectral index, such as NDVI or VRI, can score “high” in shadow areas or water areas [170]. In spectral pattern recognition, one consequence of their multivariate shape and multivariate intensity information loss is that scalar spectral indexes are ever-increasing in number and variety [77]. In practice no multivariate spectral index is representative of the multivariate shape and multivariate intensity information components of a MS hyperpolyhedron, see Fig. 3.10-6 and Fig. 3.10-7. This unquestionable fact surprisingly agrees with only a minor portion of the RS community. According to some existing literature [77], [152], it is not the first derivative, but the second derivative (local concavity) of canopies centered on the red (R) and near-infrared (NIR) wavebands to be highly related to a vegetation physical variable, such as the Leaf Area Index (LAI), regardless of the different backgrounds, e.g., burned and unburned. It is computed using three wavebands centered around the max first derivative of a vegetation reflectance red edge (Red Edge Inflection Point, REIP) [77], [152], [170]. For example, in [149] an intuitive strategy to remove soil effects from a vegetation ratio index,  $VRI = NIR / R$ , is to replace VRI, equivalent to a first-order derivative, with a second-order derivative (local concavity) centered on the NIR waveband as follows (obviously, three bands are needed), in agreement with [152], [170].

$$\text{(Down-ward) Concavity centered on the NIR waveband} = \text{Greenness}(R, NIR, MIR) = \max\{0, (NIR / R) + (NIR / MIR_{1.55-1.75}) - (R / MIR_{1.55-1.75})\} \geq 0, \text{ i.e., Greenness} \in [0, \infty). \quad (10-2)$$

$\text{Greenness}(R, NIR, MIR)$  should be considered an original combination of three well-known spectral indices whose physical meaning is sometimes fuzzy in existing RS literature, namely: (i) A canopy chlorophyll absorption index,  $VRI = NIR / R$ , (ii) a canopy water absorption index, (iii) a snow/water ratio index. In line with [152], this sum was proved to be linearly related (much better than VRI and NDVI) to LAI, irrespective of canopy background, e.g., burned and unburned [149].

By definition, an expert system, such as SIAM, relies exclusively on *a priori* knowledge available in addition to data; hence, any expert system is fully automated, i.e., it requires neither user-defined parameters nor training data to run, and one-pass, which means near real-time.



The SIAM software toolbox for prior knowledge-based MS reflectance space hyperpolyhedralization, equivalent to color naming, superpixel detection in the image-domain and VQ quality assessment consists of six subsystems, see Fig. 3.10-8. In the VQ encoding phase, SIAM partitions the MS reflectance space into static (non-adaptive to data) hyperpolyhedra, neither necessarily convex nor connected, equivalent to a dictionary (codebook) of MS color names (codewords), see Table 3.10-1. Each MS pixel is mapped onto one MS hyperpolyhedron associated with a MS color name, see Fig. 3.10-9. Unfortunately, when the MS space dimensionality is greater than three, a prior dictionary of mutually exclusive and totally exhaustive hyperpolyhedra is difficult to think of and impossible to visualize. Next, for image analysis purposes a 2D multi-level VQ map (multi-level color map) is automatically generated in linear time with the image size, see Fig. 3.10-10 and Fig. 3.10-11, together with its image-segmentation map, see Fig. 3.10-12, and image-contour map, see Fig. 3.10-13. In practice, a well-posed (deterministic) two-pass connected-component multi-level image labeling algorithm [58], [156] is input with a multi-level color map. Connect-components automatically detected in the multi-level color map-domain consist of connected sets of pixels featuring the same color label. They are typically known as (homogeneous) segments or image-objects in the OBIA literature [87], superpixels in the CV literature [61] or texels in human vision [2], [3], [58], [59], [60]. To visualize contours of image-segments detected in the multi-level color map-domain, an automatic 4- or 8-adjacency cross-aura measure is estimated in linear time, see Fig. 3.10-13. In the VQ decoding phase, equivalent to image synthesis generated via a superpixelwise-constant MS image reconstruction, also known as “image-object mean view”, a per-pixel VQ quality assessment is accomplished to guarantee VQ quality assurance, in compliance with the QA4EO validation (*Val*) requirements [111]. In VQ problems, a typical community-agreed Q<sup>2</sup>I is the per-vector encoding-decoding Euclidean distance [76], see Fig. 3.10-11.

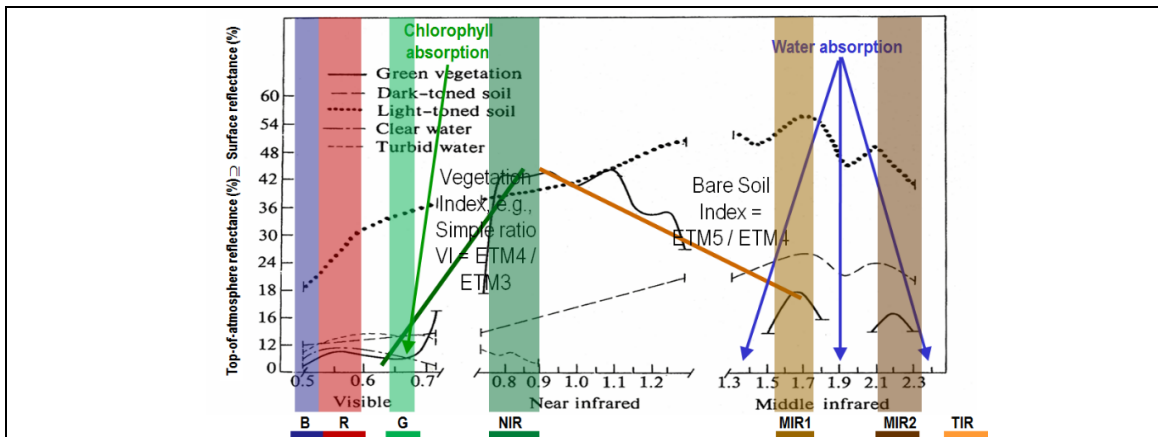


Fig. 3.10-7. 3-band Greenness Index(R, NIR, MIR)  $\propto$  Leaf Area Index (LAI). The second-order derivative (concavity) of canopies centered on the red and NIR wavebands are highly related to LAI regardless of the different backgrounds, e.g., burned and unburned. In [152], [170] it is computed using three wavebands centered around the max first-order derivative of a vegetation reflectance red edge (Red Edge Inflection Point, REIP). A 3-band second-order derivative, intuitively equal to the difference between two first-order derivatives, capable of estimating a local concavity centered on the NIR waveband and required to be monotonically increasing with the vegetation ratio index  $VRI = NIR/R$ , can be computed as follows: Greenness(R, NIR, MIR), = local concavity centered on NIR = difference of first-order derivatives, monotonically increasing with  $NIR/R \propto -[(MIR - NIR) - (NIR - R)] \propto -[MIR/NIR - NIR/R] \propto VRI - \text{Bare soil index} \propto NIR/R + NIR/MIR$ , with Bare soil index =  $MIR/NIR$ , where bare soil detection is the dual problem of vegetation detection [149].

To summarize, the SIAM expert system automatically detects in near real-time texels as connected sets of pixels featuring the same color name [59]. In CV applications, traditional texels have been recently renamed superpixels, to be detected by semi-automatic inductive data learning algorithms where at least two free-parameters are user-defined based on heuristics [61]. Texels automatically detected by SIAM can be input to the full primal sketch for texture detection in low-level vision, according to the Marr’s computational model of human vision [5].

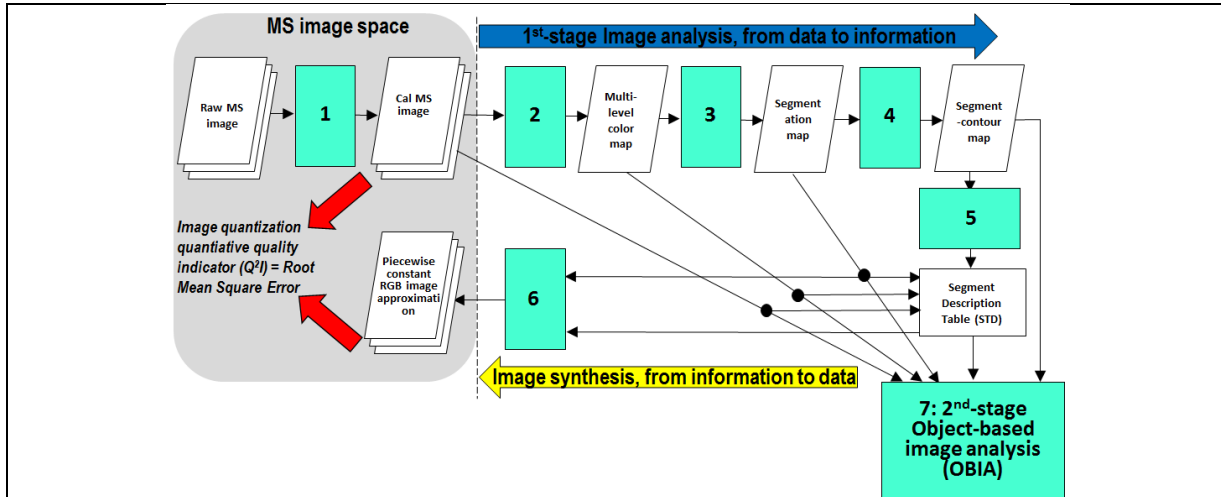


Fig. 3.10-8. The SIAM lightweight computer program for automated prior knowledge-based MS reflectance space hyperpolyhedralization, superpixel detection and vector quantization (VQ) quality assessment, consisting of boxes 1 to 6. Phase 1-of-2 = Encoding phase/Image analysis - Stage 1: sensor-specific MS data calibration into TOARF or SURF values. Stage 2: Prior knowledge-based SIAM reflectance space partitioning. Stage 3: Well-posed two-pass connected-component multi-level image labeling [58], [156]. Connected-components in the image-domain are connected sets of pixels featuring the same color label. They are also called image-objects, segments or superpixels. Stage 4: Well-posed segment-contour extraction. Stage 5: Well-posed Superpixel Description Table 3.(STD) allocation and initialization. Phase 2-of-2 = Decoding phase/Image synthesis - Stage 6: Superpixelwise-constant input image approximation ('Object-mean view') and per-pixel VQ error estimation. (Stage 7: Object-based image analysis (OBIA), in cascade to the SIAM color naming).

SIAM, r88v6	Input bands	Preliminary classification map output products: Number of output spectral categories			
		Fine discretization levels	Intermediate discretization levels	Coarse discretization levels	Inter-sensor discretization levels (*)
L-SIAM	7 – B, G, R, NIR, MIR1, MIR2, TIR	96	48	18	33
S-SIAM	4 – G, R, NIR, MIR1	68	40	15	* employed for inter-sensor post-classification change/no-change detection
AV-SIAM	4 – R, NIR, MIR1, TIR	83	43	17	
Q-SIAM	4 – B, G, R, NIR	61	28	12	

Table 3.10-1. SIAM is an EO system of systems, in compliance with the GEOSS guidelines (Group on Earth Observation, 2005). SIAM consists of six subsystems, to be input with MS images of different spectral resolution acquired by any past, present or future MS imaging sensor radiometrically calibrated into TOARF or SURF values. The 7-band Landsat-like SIAM (L-SIAM) is the “master” SIAM decision tree. “Slave” SIAM decision trees are derived from the “master” L-SIAM when the MS sensor’s spectral resolution overlaps with Landsat’s, but is inferior to Landsat’s. “Slave” SIAM implementations are the 4-band SPOT-like SIAM (S-SIAM), 4-band AVHRR-like SIAM (AV-SIAM), and 4-band QuickBird-like SIAM. Depending on the informativeness of the MS sensor’s spectral resolution, three different levels of polyhedralization of the MS reflectance space are implemented at, respectively, fine, intermediate and coarse granularity.

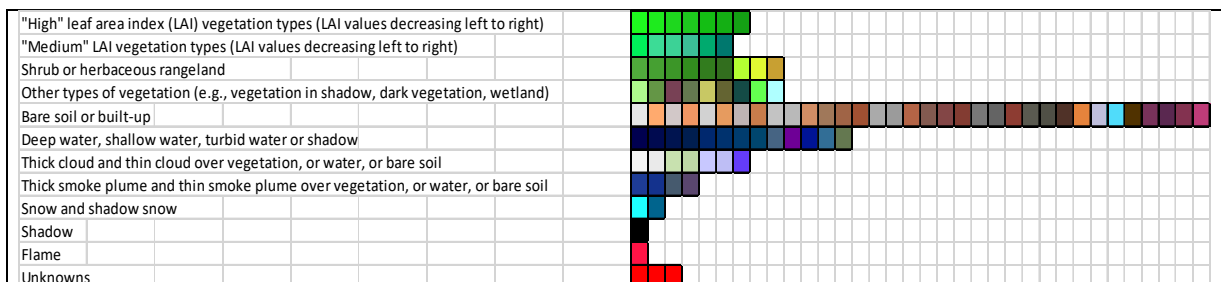




Fig. 3.10-9(a). L-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at fine granularity. Pseudocolors of 96 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" leaf area index vegetation). The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that MS hyperpolyhedron.



Fig. 3.10-9(b). L-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at coarse granularity. The coarse vector quantization (VQ) is a mutually exclusive and totally exhaustive combination of the fine VQ. Pseudocolors of the 18 spectral categories, equivalent to parent spectral categories, are chosen as to mimic natural colors of pixels belonging to each MS hyperpolyhedron.

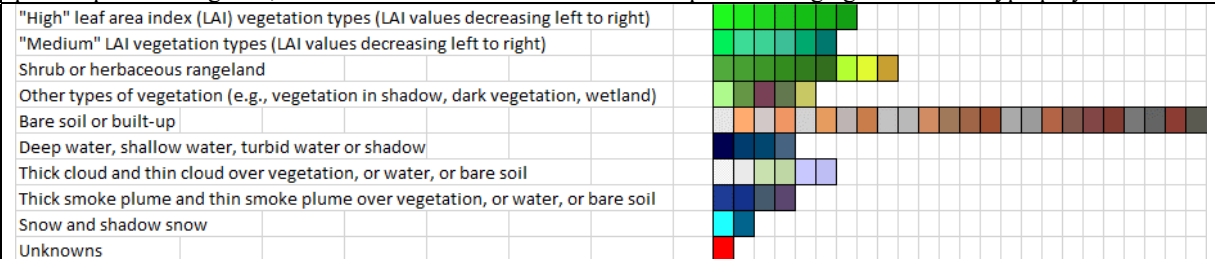


Fig. 3.10-9(c). S-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at fine granularity. Pseudocolors of 68 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" leaf area index vegetation). The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that MS hyperpolyhedron.

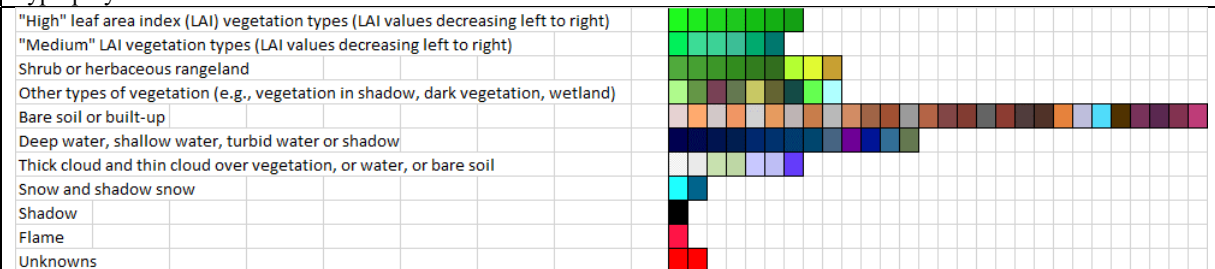


Fig. 3.10-9(d). AV-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at fine granularity. Pseudocolors of 83 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" leaf area index vegetation). The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that MS hyperpolyhedron.

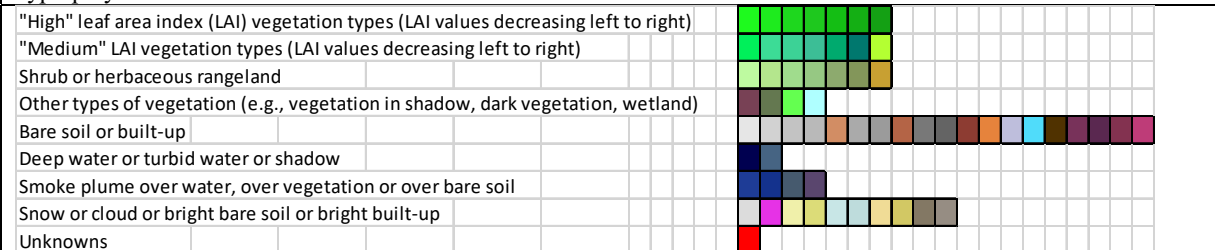




Fig. 3.10-9(e). Q-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at fine granularity. Pseudocolors of 61 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" leaf area index vegetation). The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that MS hyperpolyhedron.

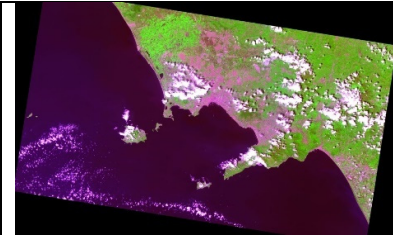


Fig. 3.10-10(a). ALOS AVNIR-2 image of the gulf of Naples, Italy (acquisition date: 2004-13-06), depicted in false colors (R: band CH3, G: band CH4, B: band CH1), 10 m resolution, calibrated into TOA reflectance. Default ENVI 2% linear histogram stretching.

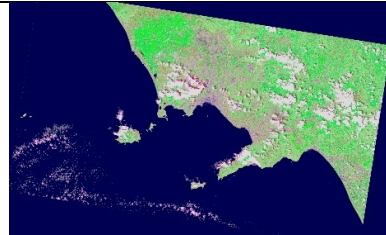


Fig. 3.10-10(b). Q-SIAM™ preliminary classification map generated from Fig. 3.10-10(a), consisting of 61 spectral categories depicted in pseudo colors (refer to the map legend in Fig. 3.10-9).



Fig. 3.10-10(c). 4-adjacency cross-aura measure in range {0, 4} extracted from the SIAM™ preliminary classification map generated from Fig. 3.10-10(a), consisting of 61 spectral categories and shown in Fig. 3.10-10(b).

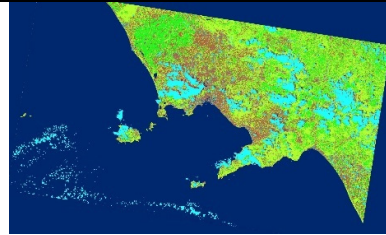
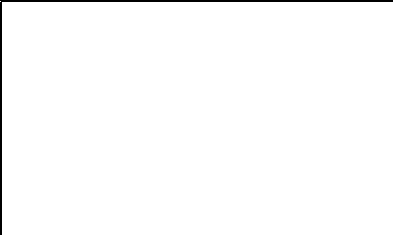


Fig. 3.10-10(d). Q-SIAM™ preliminary classification map generated from Fig. 3.10-10(a), consisting of 12 spectral categories depicted in pseudo colors (refer to the map legend in Fig. 3.10-9).

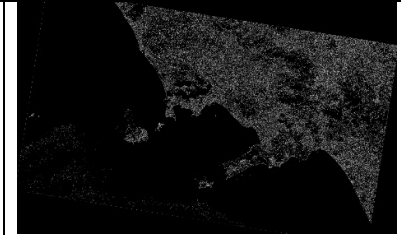


Fig. 3.10-10(e). 4-adjacency cross-aura measure in range {0, 4} extracted from the SIAM™ preliminary classification map generated from Fig. 3.10-10(a), consisting of 12 spectral categories and shown in Fig. 3.10-10(d).



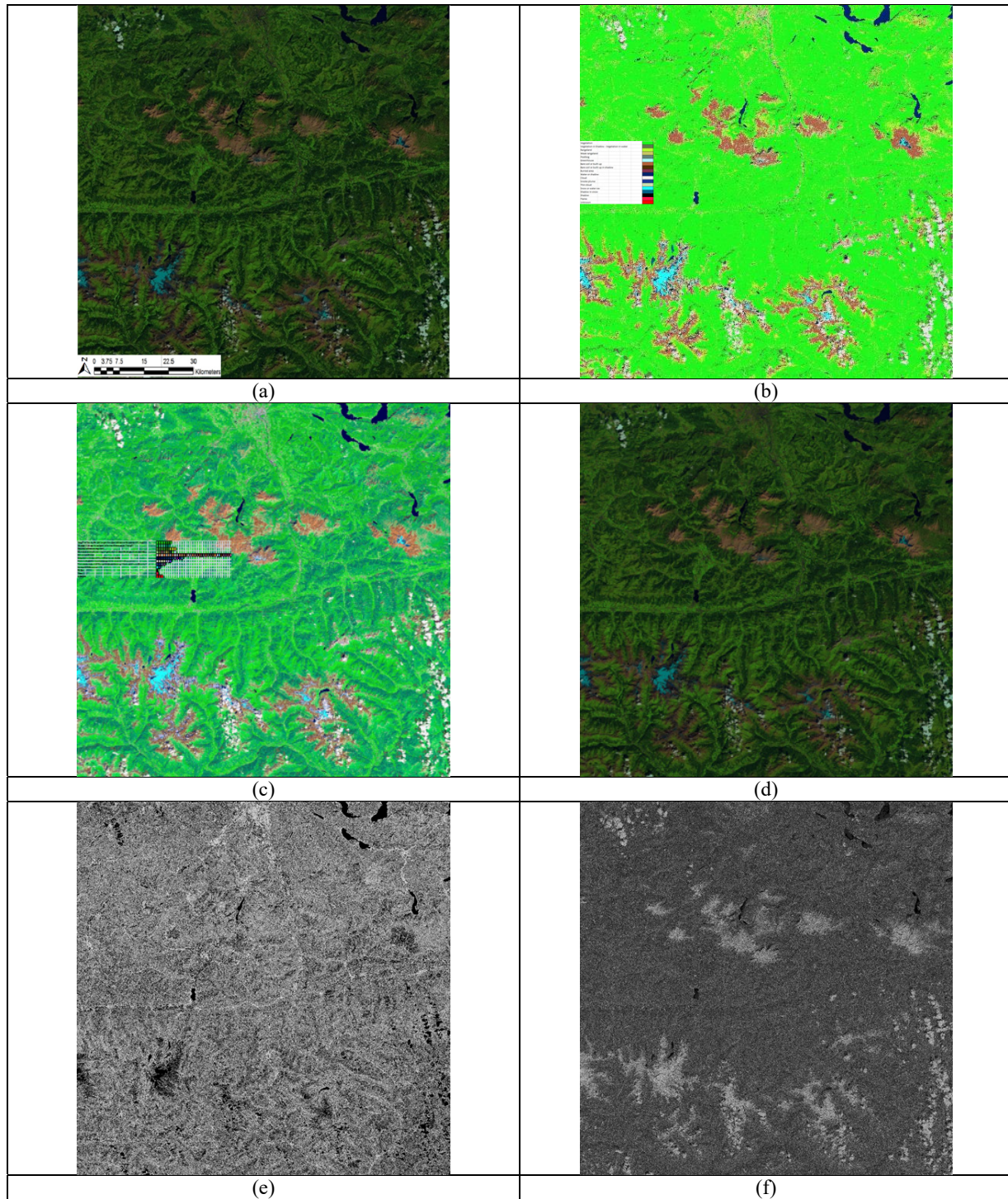


Fig. 3.10-11. (a) Sentinel-2A MSI Level-1C image of the Earth surface, located south of the city of Salzburg, Austria. The city area is visible around the middle of the image upper boundary (Lat-long coordinates: 47°48'25.0"N 13°02'43.6"E). Acquired on 2015-08-13. Spatial resolution: 10 m. Image size: 110×110 km. Radiometrically calibrated into TOARF values in range {0, 255}, it is depicted as a false color RGB image, where: R = Medium InfraRed (MIR) = Band 11, G = Near IR (NIR) = Band 8, B = Blue = Band 2. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at coarse color granularity, consisting of 18 spectral categories depicted in pseudo colors, refer to the map legend in Fig. 3.10-9. Coarse-granularity color categories are generated by merging color hyperpolyhedra at fine color granularity, according to pre-defined parent-child relationships, refer to Table 3.10-1. (c) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, refer to the





map legend in Fig. 3.10-9. (d) Superpixelwise-constant approximation of the input image (“image-object mean view”) generated from the L-SIAM’s 96 color map at fine granularity. Depicted in false colors: R = MIR = Band 11, G = NIR = Band 8, B = Blue = Band 2. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (e) 8-adjacency cross-aura contour map in range  $\{0, 8\}$  automatically generated from the L-SIAM’s 96 color map at fine granularity. It shows contours of connected sets of pixels featuring the same color label. These connected-components are also called image-objects, segments or superpixels. (f) Per-pixel scalar difference between the input MS image shown in (a) and the superpixelwise-constant MS image reconstruction shown in (d). This scalar difference is computed as the per-pixel Root Mean Square Error (RMSE) in range  $\{0, 255\}$ . The RMSE is a well-known vector quantization (VQ) error. Image-wide basic statistics: Min = 0, Max = 130, Mean = 2.60, Stdev = 3.45. Histogram stretching is applied for visualization purposes.

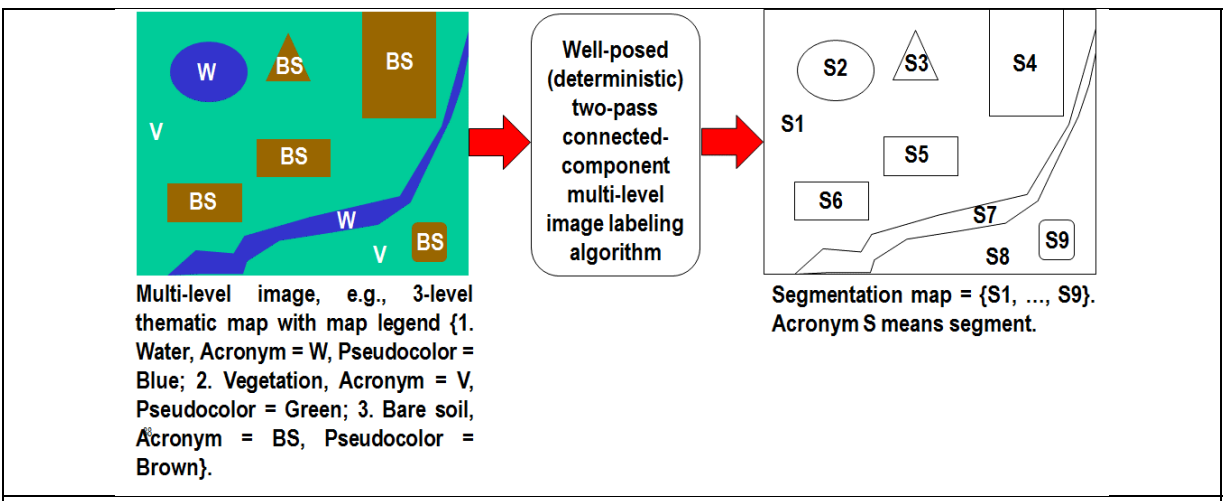


Fig. 3.10-12. One segmentation map is deterministically generated from one multi-level image, such as a thematic map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. In this example, nine image-objects/segments S1 to S9 can be detected in the 3-level thematic map shown at left. Each segment consists of a connected set of pixels sharing the same thematic map label [58], [156]. Each stratum/layer/level consists of one or more segments, e.g., stratum Vegetation (V) consists of the two disjoint segments S1 and S8. In any multi-level (categorical, nominal) image domain, three spatial primitives co-exist and are provided with parent-child relationships: pixel (row-column coordinate pair) with a parent segment identifier (ID) and a super-parent level ID, segment (polygon) with a segment ID and a parent level ID, and a level/stratum (multi-part polygon) with a level ID.

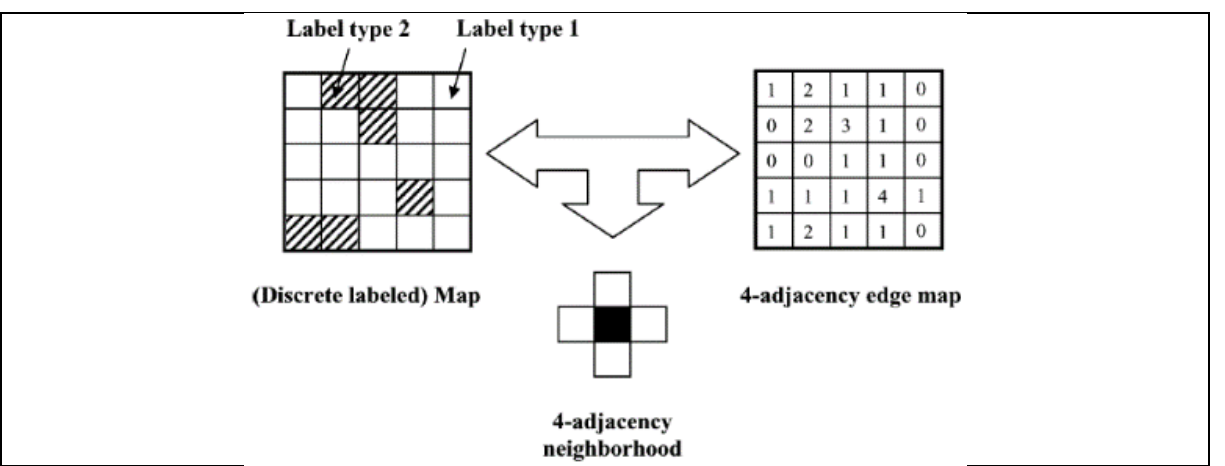
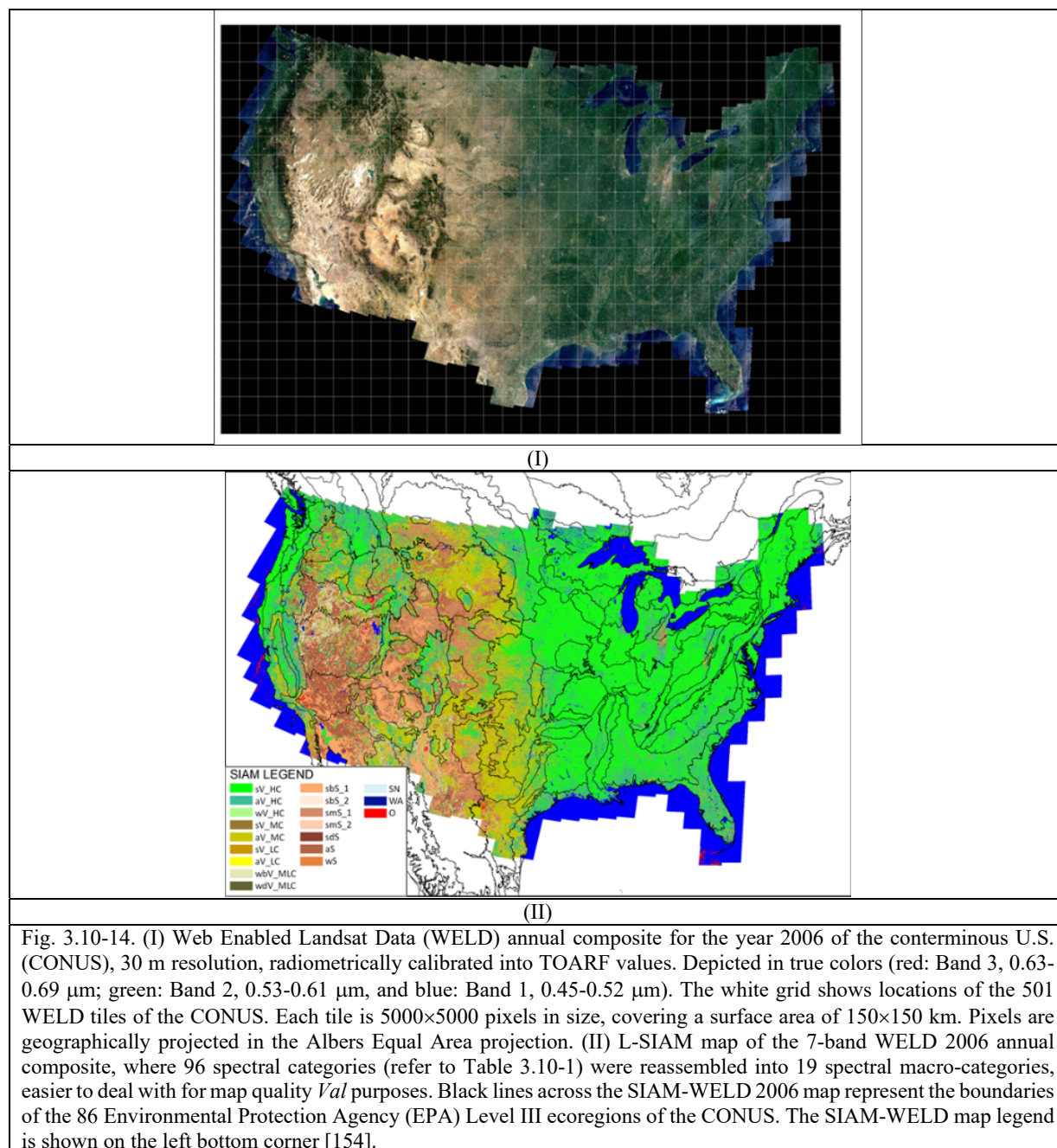


Fig. 3.10-13. Example of a 4-adjacency cross-aura map generated from a multi-level image, such as a binary (2-level) classification map.



To prove the SIAM capability of mapping EO “big data” into MS color names automatically and in near real-time in compliance with the QA4EO’s *Cal/Val* requirements [111], the SIAM lightweight computer program was tested at the continental U.S. (CONUS) spatial extent covered by the annual Web Landsat Data Set (WELD) in comparison with the USGS NLCD 2006 reference map, see Fig. 3.10-14, and at the European spatial extent covered by the Image 2006 Coverage 1 mosaic, 10 m resolution, consisting of two thousands 4-band IRS-P6 LISSIII, SPOT-4, and SPOT-5 images, mostly acquired during 2006, radiometrically calibrated into TOARF values, see Fig. 3.10-15.



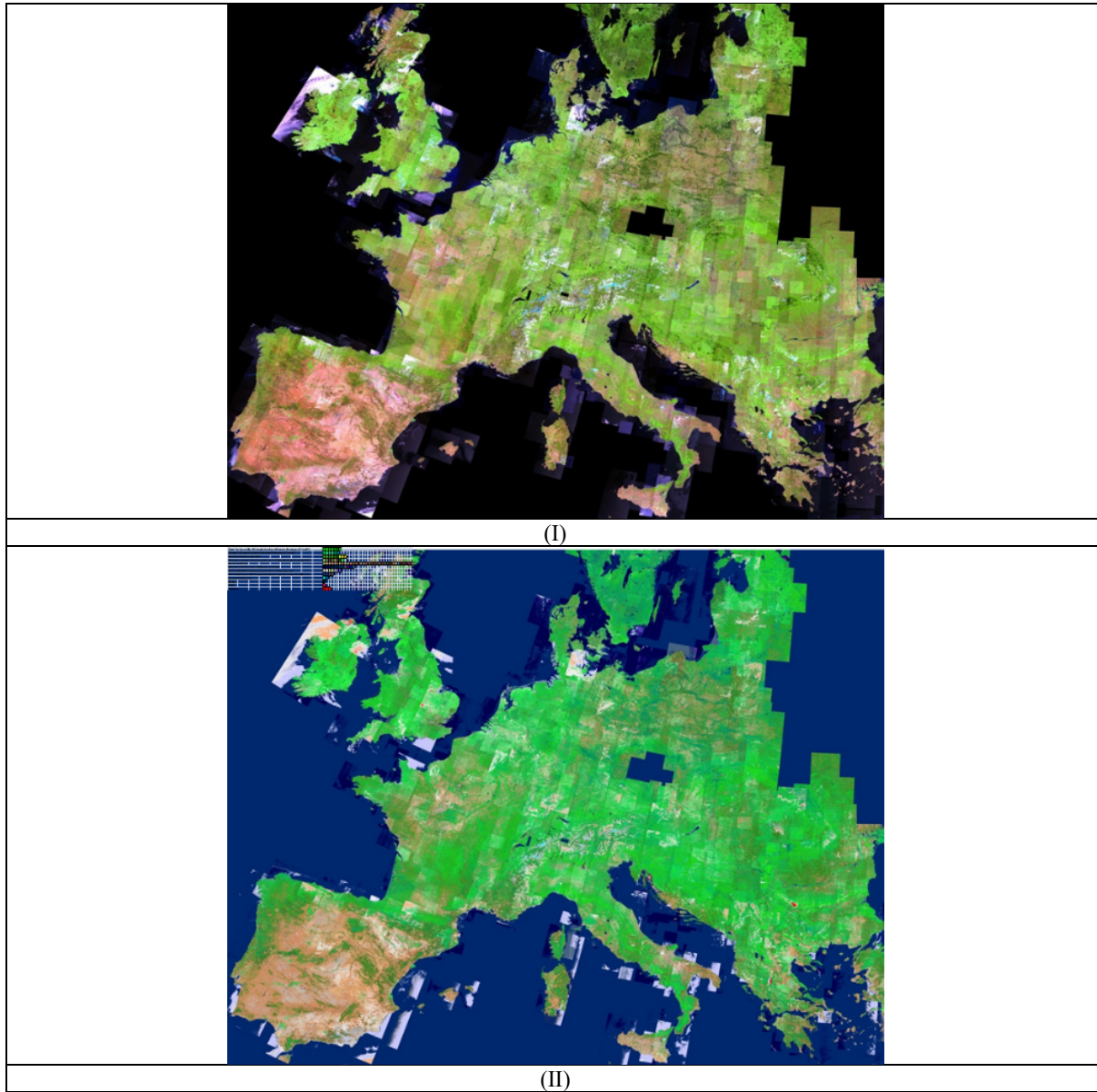


Fig. 3.10-15. (I) Image 2006 Coverage 1 mosaic, 10 m resolution, consisting of two thousands 4-band IRS-P6 LISSIII, SPOT-4, and SPOT-5 images, mostly acquired during 2006, radiometrically calibrated into TOARF values, geometrically orthorectified and depicted in false colors: Red – Band 4 (Short Wave InfraRed, SWIR), Green – Band 3 (Near IR, NIR), Blue – Band 1 (Visible Green). (II) S-SIAM map of the 4-band Image 2006 Coverage 1 mosaic. The S-SIAM map legend, consisting of 68 spectral categories (refer to Table 3.10-1) depicted in pseudocolors, is shown in the left top corner [149], [150].

### 3.10.2 The RGBIAM lightweight computer program for true- or false-color RGB cube polyhedralization, superpixel detection and VQ quality assurance

In recent years the quality and quantity of consumer-level color cameras mounted in mobile electronic devices, e.g., smartphones, and/or on light-weight unmanned aerial vehicles (UAVs) [139], have been ever increasing, so is the number of acquired uncalibrated true- or false-color RGB images.



To cope with top-down/deductive color naming in uncalibrated true- or false-color RGB images, the RGBIAM software toolbox in operating mode was developed as a down-scaled version of the SIAM software toolbox, consisting of six subsystems as shown in Fig. 3.10-8.

Adopted in [155], [188], RGBIAM accomplishes uncalibrated true- or false-color RGB cube polyhedralization, where mutually exclusive and totally exhaustive polyhedra are neither necessarily convex nor connected, see Fig. 3.10-1.

To comply with the QA4EO Cal/Val requirements [111] when radiometric calibration metadata are not available, RGB color constancy is considered mandatory to guarantee harmonization and interoperability of uncalibrated RGB images acquired across time, space, sensors and varying illumination conditions, refer to Chapter 3.9 and to the existing literature on color constancy [22], [112], [113], [129], [131], [144].

Since it is a physical model-based decision tree non-adaptive to input data, the RGBIAM's spectral decision tree considers color constancy as a mandatory pre-processing first stage of any uncalibrated true- or false-color RGB image. Aside from this difference in the data pre-processing first stage, the SIAM workflow, shown in Fig. 3.10-8, coincides with the RGBIAM pipeline. Textels automatically detected by RGBIAM as connected sets of pixels featuring the same color label can be input to a low-level vision full primal sketch for texture detection.

In RGBIAM two RGB cube discretization levels were implemented: (a) a fine color discretization level, consisting of  $49+1 = 50$  color names, including class "unknown", and (b) a coarse color discretization level, consisting of the 11 human BCs identified by Berlin and Kay in their cross-cultural survey of color names in human languages [132], plus 1 class "unknown" = 12 color names, generated as a fixed parent-child combination of the 50 color names available at the fine discretization level, see Fig. 3.10-16.

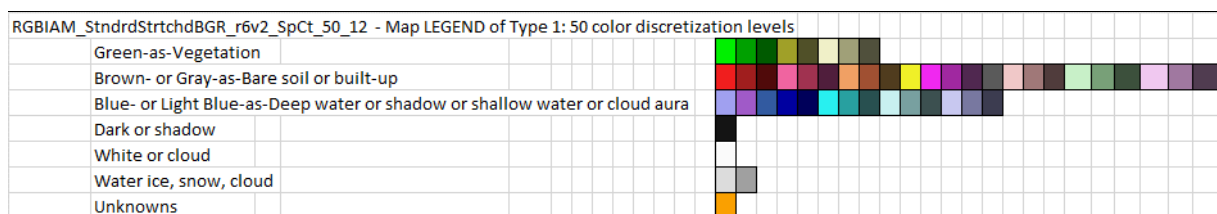


Fig. 3.10-16(a). Two-level RGBIAM's quantization of the RGB cube. Color map's legend at (a) fine (49 + 1 class unknown).



Fig. 3.10-16(b). Two-level RGBIAM's quantization of the RGB cube. Color map's legend at (b) coarse (11 + 1 class unknown) quantization levels, where the coarse VQ is a mutually exclusive and totally exhaustive combination of the fine VQ.

The RGBIAM is near real-time because its computational complexity is linear, specifically,  $\leq O(C1 \cdot N \cdot B + C2 \cdot N)$ , with  $N$  = image size in pixels,  $B$  = number of spectral bands = 3, where  $C1 = 50$  = cardinality of the fine-granularity color dictionary,  $C2 = 12$  = cardinality of the coarse-granularity color dictionary, generated as an aggregation of the  $C1$  color names, i.e., inequality  $C2 < C1$  must hold.

Examples of RGBIAM outcomes on an EO false-color RGB images and on a true-color RGB natural terrestrial image are shown in Fig. 3.10-17 and Fig. 3.10-18.



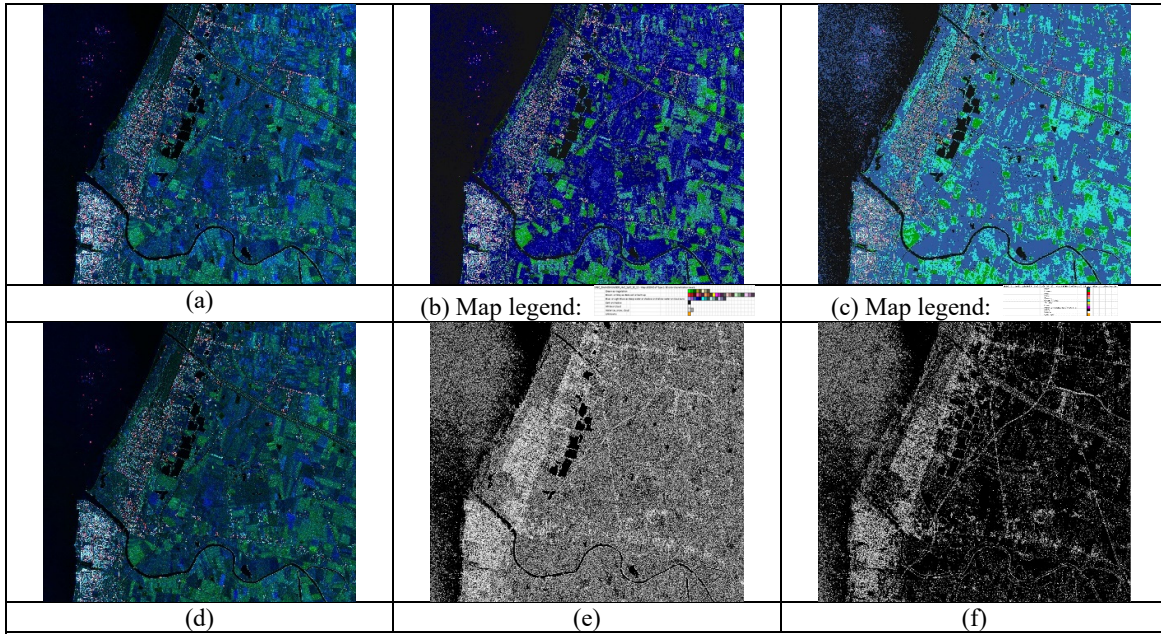


Fig. 3.10-17. EO false-color RGB image. Spaceborne bi-temporal RGB-SAR image of the Campania region, Italy (courtesy of University of Naples Federico II, Italy) [168]. Image size: RW = 4480, CL = 5012. (a) RGB image subject to color constancy. (b) RGBIAM color map in pseudocolors, fine 49+1 color quantization levels. (c) RGBIAM color map in pseudocolors, coarse 11+1 color quantization levels. (d) Original RGB-SAR image. (e) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM color map at fine 49+1 color quantization levels. (f) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM color map at coarse 11+1 color quantization levels.

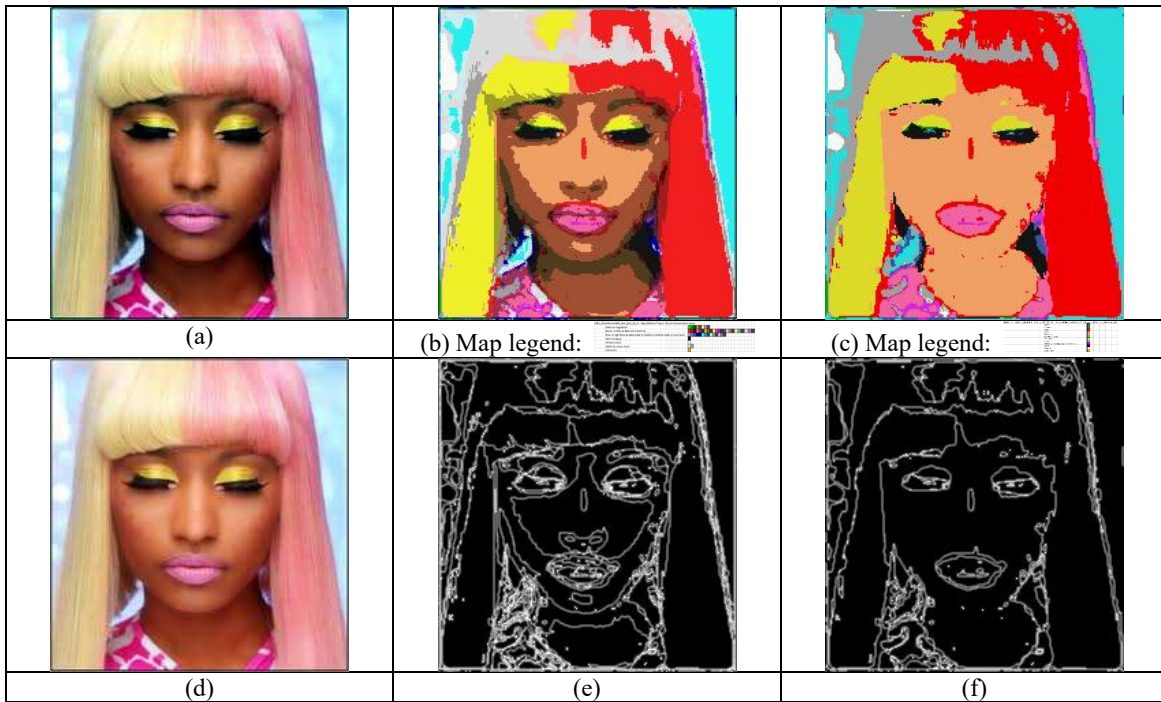


Fig. 3.10-18. True-color RGB terrestrial image of a natural subject. Spaceborne bi-temporal RGB-SAR image of the Campania region, Italy (courtesy of University of Naples Federico II, Italy). Image size: RW = 230, CL = 219. (a) RGB image subject to color constancy. (b) RGBIAM color map in pseudocolors, fine 49+1 color quantization levels. (c) RGBIAM color map in pseudocolors, coarse 11+1 color quantization levels. (d) Original RGB-SAR



image. (e) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM color map at fine 49+1 color quantization levels. (f) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM color map at coarse 11+1 color quantization levels.

### 3.11 Original 1D simulations for image analysis and synthesis, including the zero-frequency signal component, image-contour detection and keypoint detection consistent with the Mach bands illusion

A simple but not trivial image model, violated in practice by all image segmentation algorithms inconsistent with the Mach bands illusion, considers an image (i.e., a 2D regular grid of numbers) a combination of four spatial shapes [1].

5. Flat areas.
6. Ramps.
7. Step edges.
8. Lines.

Starting from preliminary results proposed in [1], Fig. 3.11-1 shows a 1D function  $f(x)$  and its single-scale wavelet decomposition, composition and contrast estimation suitable for function partitioning (segmentation). For 1D simulation purposes,

- 3 pixel-wide 1D odd-symmetric filter:  $(+1, 0, -1)$ .
- 3 pixel-wide 1D even-symmetric filter:  $(-0.5, 1, -0.5)$ .
- 3 pixel-wide 1D Gabor filter:  $(+0.25, 0.5, +0.25)$ . This low-pass filter provides a local estimate of the signal zero-frequency (direct current, DC) component.

The original “*perceptual contrast*” (PrcptlCntrst2) and function reconstruction (Rcnstrctn) expressions adopted in these simulations are defined as follows.

$$\checkmark \text{ PrcptlCntrst2}(x) = \text{abs}[f(x) \circ \partial^2 G/\partial x^2] + \text{abs}[f(x) \circ \partial G/\partial x]/2, \text{ hence PrcptlCntrst2}(x) \geq 0. \quad (11-1)$$

$$\checkmark \text{ Rcnstrct}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G/\partial x^2]/2. \quad (11-2)$$

Noteworthy, first, PrcptlCntrst2(x) is a non-negative combination of even- and odd-symmetric simple cells alternative to complex cells featuring a second-degree (squaring) nonlinearity proposed in Eq. (5-1) by Adelson and Bergen [42] and Burr and Morrone [26].

Second, Rcnstrct(x) is a function of the Gaussian function and even-symmetric filter,  $\partial^2 G/\partial x^2$ , exclusively, i.e., no odd-symmetric filter is involved. This agrees with the Adelson and Bergen’s image reconstruction by a Gaussian pyramid plus a Laplacian pyramid, where the Laplacian pyramid is generated as the residual (difference) between the low-pass Gaussian filtered image at scale  $s$  and the up-scale of the low-pass Gaussian filtered image at scale  $s+1$ , such that this inter-scale difference of Gaussian (DOG) is approximately equivalent to the even-symmetric Laplacian of Gaussian,  $\nabla^2 G = \partial^2 G/\partial x^2 + \partial^2 G/\partial y^2$  [42].

It is important to remind that, because the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution [88], see Eq. (7-1). Therefore, the even-symmetric filtered image  $(\partial^2 G/\partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  applied to the low-pass image adopted by both Canny [27] and Bertero, Torre and Poggio [44].

$$(\partial^2 G/\partial n^2 * I) \neq \partial^2/\partial n^2 (G * I) \quad (11-3)$$

These 1D experiments with a single-scale even-, odd- and Gaussian filter show the following, see Fig. 3.11-1.

- The even-symmetric filter shown in Fig. 3.11-1(I) is necessary and sufficient to detect all kinds of 1D edges shown in Fig. 3.11-1(III) to Fig. 3.11-1(XIII), where the image boundary position is localized by ZX pixels where the value of the output function, OEvenSymtrc(x), passes from positive to non-positive, i.e., from positive to either zero or negative values, or vice versa.
- The even-symmetric filter shown in Fig. 3.11-1(I), which is necessary and sufficient to detect all kinds of 1D edges shown in Fig. 3.11-1(III) to Fig. 3.11-1(XIII), complies with one of the best-known brightness illusions, namely, the Mach bands illusion, when dealing with the ramp edge, see Fig. 3.4-1. This means that this even-symmetric filter features biological plausibility as a value added.



- $Rnstrct(x)$  of the 1D function  $f(x)$  is perfect, i.e., lossless. Noteworthy, to compute the proposed function  $Rnstrct(x)$ , Gaussian filters are combined with even-symmetric filters.
- The proposed function  $PrcptlCntrst2(x) \geq 0$  is an original combination of odd-symmetric filters with even-symmetric filters, alternative to the definition of complex cells in the PVC adopted by a great section of the existing computer vision literature, including Canny [27], Burr and Morrone [26], Rodrigues and du Buf's [28], [43], Heitger et al. [41], Adelson and Bergen [42], Smith and Brady [90] and many others, e.g., refer to [25].
- $PrcptlCntrst2(x)$  allows to partition the 1D space  $x$  into (connected) segments featuring  $PrcptlCntrst2(x) > 0$  and  $PrcptlCntrst2(x) = 0$ . In greater detail:

ATTENTION:  $PrcptlCntrst2(x)$  allows to partition (discriminate) zero-concavity (ZC) segments, where  $[f(x) \circ \partial^2 G/\partial x^2] = 0$ , into either ramps or flat areas: in any ramp  $[f(x) \circ \partial^2 G/\partial x^2] = 0$  AND  $PrcptlCntrst2(x) > 0$ , in any flat area  $[f(x) \circ \partial^2 G/\partial x^2] = 0$  AND  $PrcptlCntrst2(x) = 0$ .

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 1.

Any zero-crossing (ZX) pixel should be merged with the neighboring pixel whose  $PrcptlCntrst2(x) = 0$  or "low", if any. This is equivalent to requiring the neighboring pixel to belong to a flat area. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 2.

Any pair of neighboring pixels, either ZX or not, should be merged with the neighboring pixel whose  $\Delta GrayValue = 0$  or "low", if any. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 3.

Any pair of neighboring pixels, either ZX or not, both featuring  $PrcptlCntrst2(x) = 0$  or "low", should be merged into the same zero-concavity (ZC) segment, either new or pre-existing. This region growing rule applies to image areas where there are smooth (low) concavity values, encompassing changes in the sign of concavity. Example: the smooth skin effect in the shoulder or cheeks of Lenna.

- The assignment of boundary pixels to segments featuring condition  $PrcptlCntrst2(x) > 0$  must be carefully scrutinized. In particular, the local extrema (local maxima and local minima) of  $PrcptlCntrst2(x) \geq 0$  feature the following properties.
  - They represent a small set of the set of pixels featuring  $PrcptlCntrst2(x) > 0$ , hence their scrutiny should be easier to accomplish.
  - According to Fig. 3.11-1, the local extrema (local maxima and local minima) of  $PrcptlCntrst2(x) \geq 0$  can either or not coincide with ZX pixels.
  - They appear of particular interest for:
    - ✓ perceptual contour detection, in combination with ZX pixels detected by the even-symmetric filtering operator, and simultaneously
    - ✓ for image saliency perception [142], including end-point, T-junction, X-junction and corner detection (called terminations by Marr ([5], p. 71), see Fig. 3.11-2. In other words, the local extrema (local maxima and local minima) of  $PrcptlCntrst2(x) \geq 0$  appear related to keypoints by Lowe and/or end-stopped cell's outputs by Rodrigues and du Buf's [28], [43], Heitger et al. [41].
      - For example, in [21], Lowe searches for scale-invariant keypoints (scale invariant feature transform, SIFT) as local extrema in the difference of Gaussian (DOG), equivalent to a  $\nabla^2 G$ -filtered image, that are not, simultaneously, ZX pixels, i.e., pixels whose second derivative, estimated by means of a Hessian matrix (refer to this text above), is strong because they lie along edges, refer to the further Chapter 3.12.
      - In [41], [43], the computational model of end-stopped cells allows to detect "keypoints" as the peaks (local maxima in a  $3 \times 3$  neighbourhood) in the summed end-stopped representation. These keypoints are one-to-one related to, i.e., they are the biological counterpart of, the Lowe SIFT operators, although keypoint estimation through simple- and double-stopped operators,

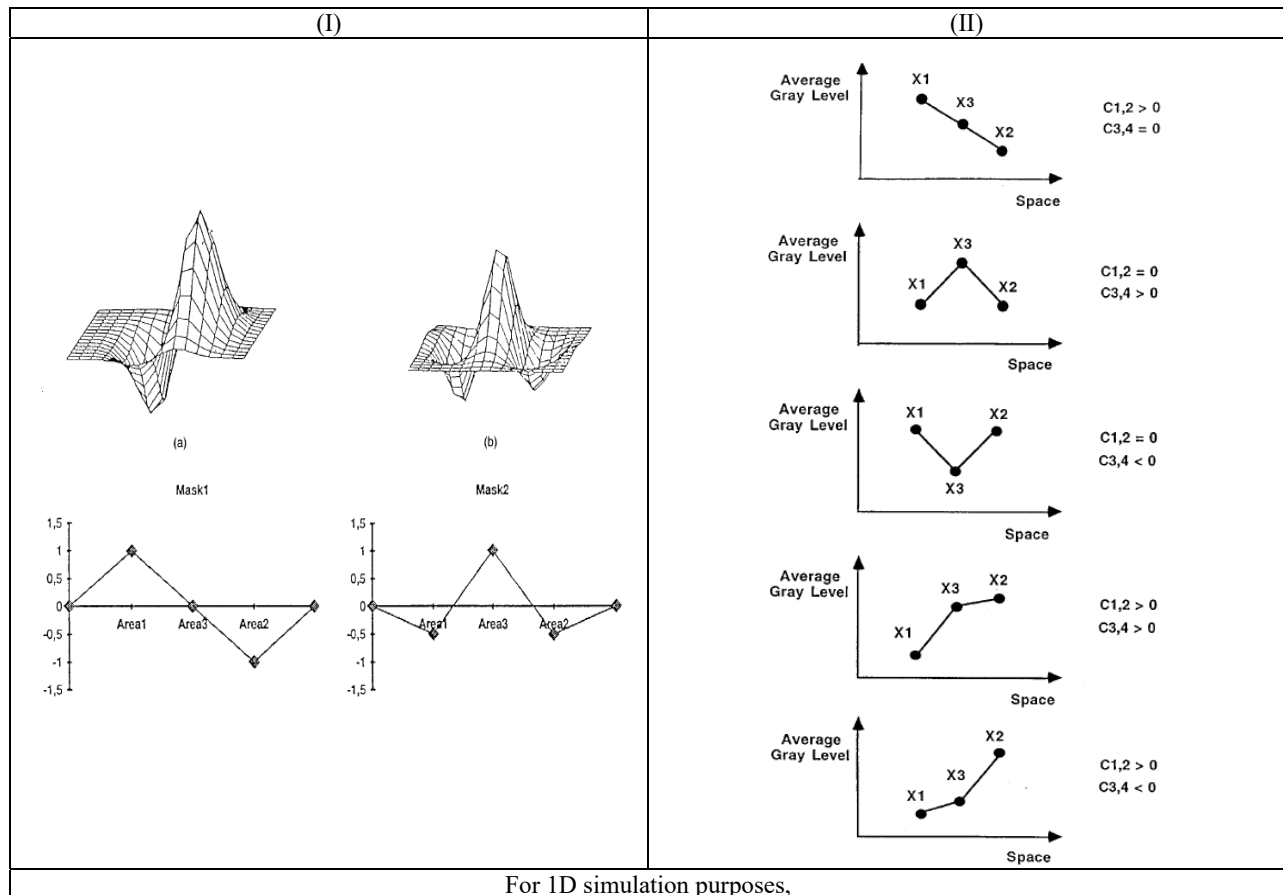
see Eq. (5-2) and Eq. (5-3), adopted in [41] and [43] is not computationally equivalent to Lowe's [21].

According to the existing literature [41], [43], the information represented at the keypoints complements the edge representation. The edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of this map difficult. One can see that the representation of keypoints indicates precisely these critical locations, like terminations, corners and junctions. Many of the keypoints are located on occluding contours.

According to [24], if we require that a computer vision model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, see Fig. 3.4-1, the number of published models is surprisingly small, as proved in [25].

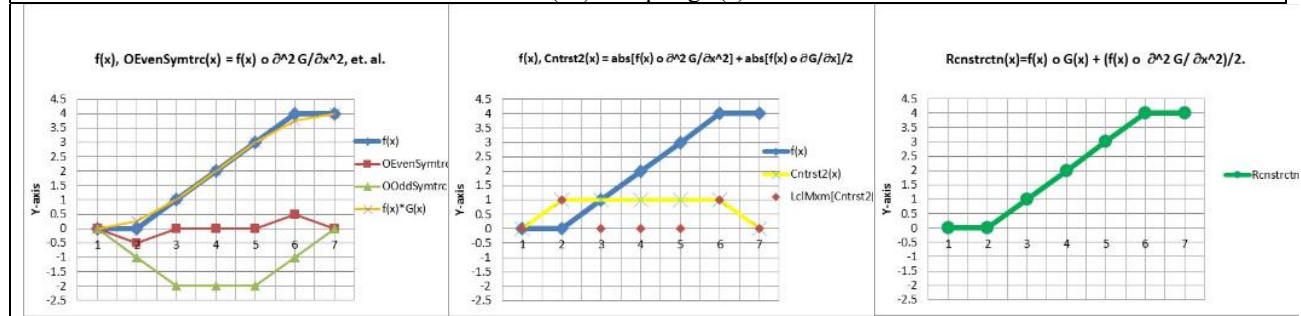
Based on conclusions of survey papers, like [24] and [25], the proposed combinations of Gaussian, even- and odd-symmetric Gabor filters in functions  $Rnstrct(x)$  and  $PrcptlCntrst2(x)$  can be considered:

- (I) consistent with the Mach bands illusion, see Fig. 3.4-1.
- (II) Completely novel, i.e., different from existing combinations of even- and odd-symmetric filters for image analysis/synthesis and/or image-contour detection, like those proposed by Canny [27], Burr and Morrone [26], Rodrigues and du Buf's [28], [43], Heitger et al. [41], Adelson and Bergen [42], Smith and Brady [90] and many others, e.g., refer to [25].
- (III) Capable of near-orthogonal image decomposition.
- (IV) Capable of lossless image reconstruction.
- (V) Capable of image-contour detection. Contour pixels are ZX pixels localized in the image-domain where a change in sign of the local concavity  $[I(n) \circ \partial^2 G / \partial n^2]$  of the 2D image function  $I(n)$  with pixel  $n = (x, y)$  occurs with respect to the local concavity of pixels belonging to the 8-adjacency neighborhood of pixel  $n$ .
- (VI) Capable of keypoint (end-point, T-junction, X-junction and corner) detection as local extrema (local maxima and local minima) of  $PrcptlCntrst2(x) \geq 0$ .

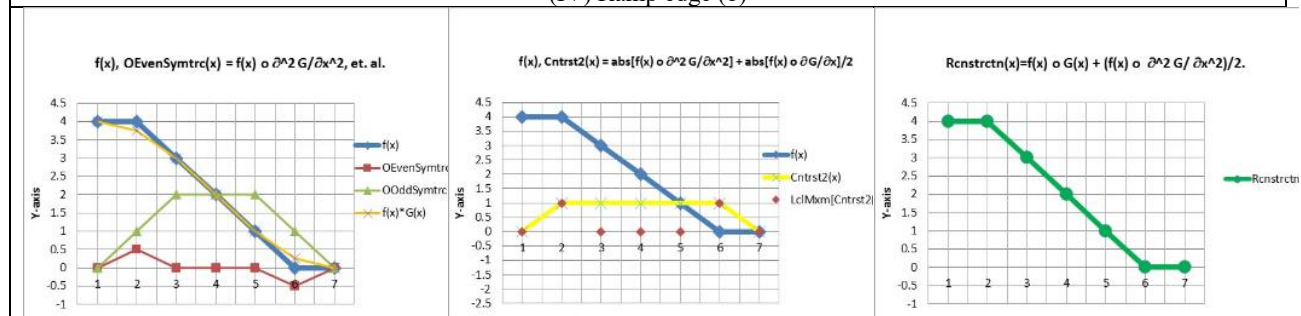


- 3 pixel-wide 1D odd-symmetric filter: (+1, 0, -1).
- 3 pixel-wide 1D even-symmetric filter: (-0.5, 1, -0.5).
- 3 pixel-wide Gaussian (0.25, 0.50, 0.25)

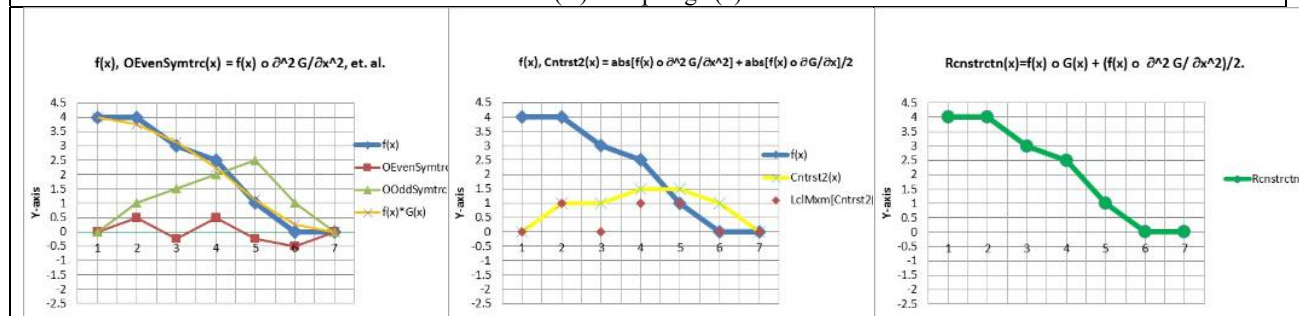
(III) Ramp edge (a)



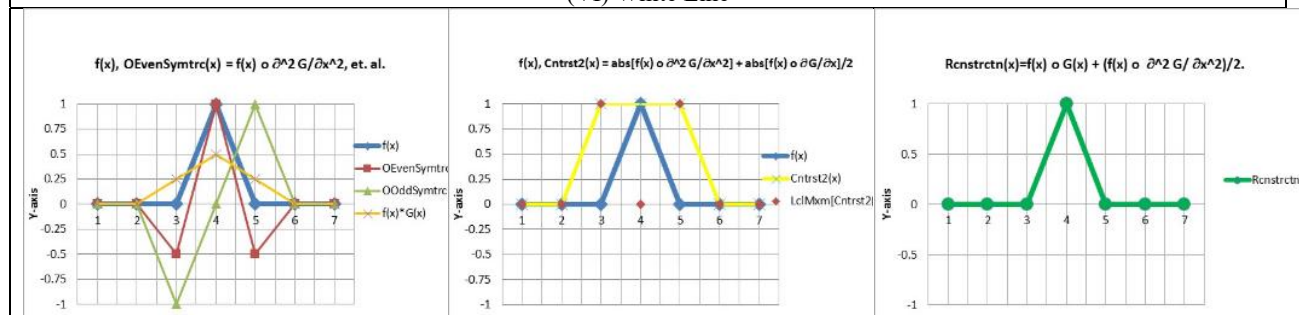
(IV) Ramp edge (b)



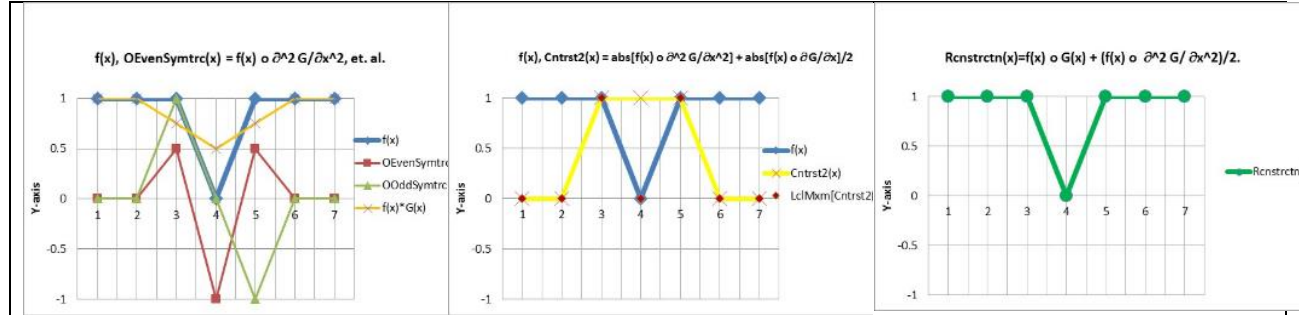
(V) Ramp edge (c)



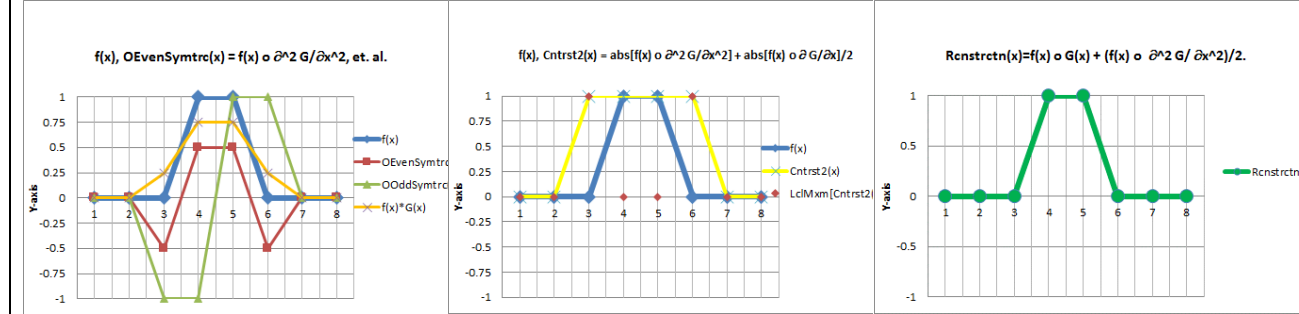
(VI) White Line



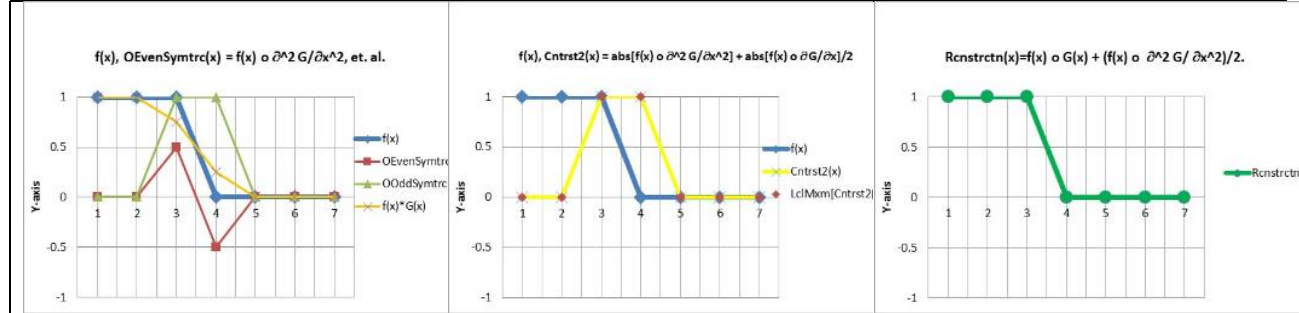
(VII) Black Line



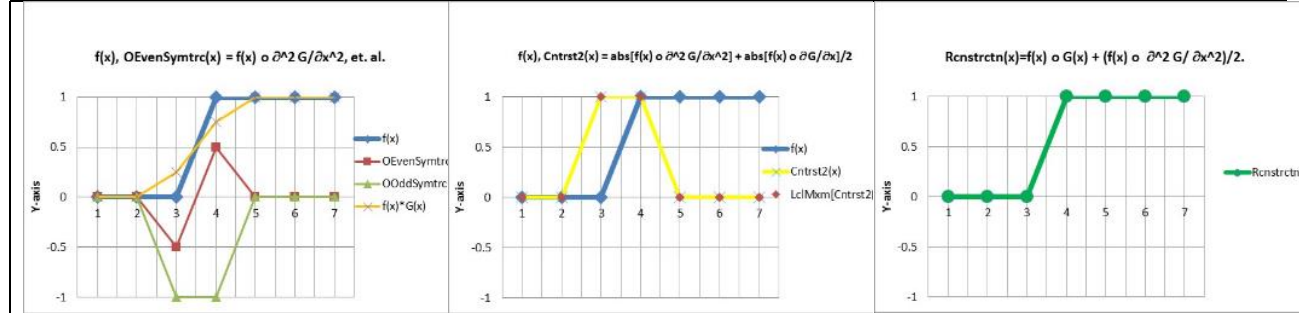
(VIII) White Line, two-pixel wide



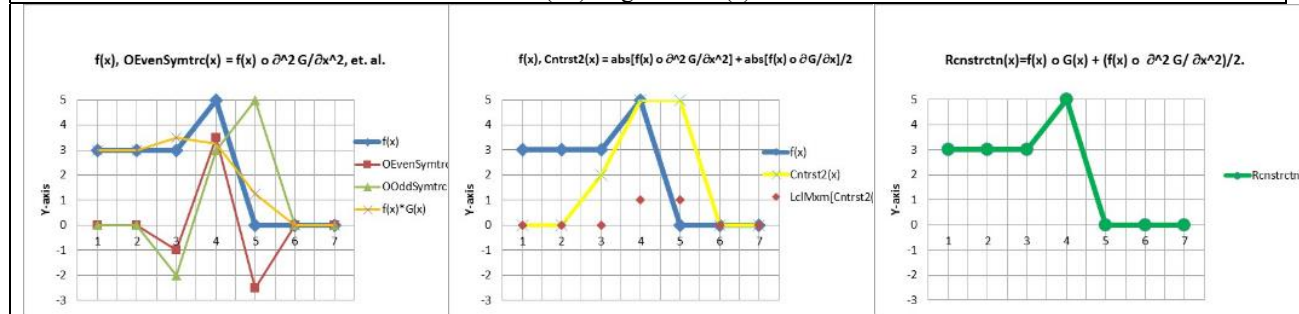
(IX) Step Edge (Z)



(X) Step Edge (S)



(XI) Edge + Line (a)



(XII) Edge + Line (b)





Fig. 3.11-1. An image is a combination of four spatial shapes: flat areas, ramps, step edges and lines. Inspired to the even- and odd-symmetric filter outputs proposed in [1]. 1D function  $f(x)$  and its wavelet decomposition, composition (reconstruction,  $Rcnstrctn$ ) and original “perceptual contrast” ( $PrcptlCntrst2$ ) estimation suitable for function partitioning (segmentation). For 1D simulation purposes,

- 3 pixel-wide 1D odd-symmetric filter: (+1, 0, -1).
- 3 pixel-wide 1D even-symmetric filter: (-0.5, 1, -0.5).
- 3 pixel-wide 1D Gabor filter: (+0.25, 0.5, +0.25).

$OOddSymtrc(x)$  is the output of the odd-symmetric filter shown in Fig. 3.11-1(I),  $OEvenSymtrc(x)$  is the output of the even-symmetric filter shown in Fig. 3.11-1(I). The even-symmetric filter in Fig. 3.11-1(I) computes as output an estimate of the local curvature (concavity) of the image intensity. In particular:

Even-symmetric filter output values can be  $< 0$  (local concavity up),  $= 0$  (no local concavity, i.e., the local image intensity is a straight line, either horizontal or sloped; this is called zero-concavity (ZC) image-segment) or  $> 0$  (local concavity down).

It is easy to verify that the even-symmetric filter shown in Fig. 3.11-1(I), computing as output the value  $OEvenSymtrc(x)$ , is necessary and sufficient to detect all kinds of 1D edges shown in Fig. 3.11-1(III) to Fig. 3.11-1(XIII), in line with one of the best-known brightness illusions, namely, the Mach bands illusion, when dealing with the ramp edge, see Fig. 3.4-1.

The original “perceptual contrast” ( $PrcptlCntrst2$ ) and function reconstruction ( $Rcnstrctn$ ) expressions adopted in these simulations are:

$$\checkmark \text{PrcptlCntrst2}(x) = \text{abs}[f(x) \circ \partial^2 G/\partial x^2] + \text{abs}[f(x) \circ \partial G/\partial x]/2. \quad (11-1)$$

$$\checkmark \text{Rcnstrct}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G/\partial x^2]/2. \quad (11-2)$$

Noteworthy:

- $PrcptlCntrst2(x)$  allows to partition (discriminate) zero-concavity (ZC) regions, where  $[f(x) \circ \partial^2 G/\partial x^2] = 0$ , into either ramps or flat areas: in any ramp  $[f(x) \circ \partial^2 G/\partial x^2] = 0$  AND  $PrcptlCntrst2(x) = \text{Constant} > 0$ , in any flat area  $[f(x) \circ \partial^2 G/\partial x^2] = 0$  AND  $PrcptlCntrst2(x) = 0$ .

- Any zero-crossing (ZX) pixel should be merged with the neighboring pixel whose  $\text{PrcptlCntrst2}(x) = 0$  or “low”, if any. This is equivalent to requiring the neighboring pixel to belong to a flat area. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.
- Any pair of neighboring pixels, either ZX or not, should be merged with the neighboring pixel whose  $\text{DeltaGrayValue} = 0$  or “low”, if any. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.
- Any pair of neighboring pixels, either ZX or not, both featuring  $\text{PrcptlCntrst2}(x) = 0$  or “low”, should be merged into the same zero-concavity (ZC) segment, either new or pre-existing.



Fig. 3.11-2(a). Reproduced with permission, courtesy of [43]. Keypoint (endpoint, junction, corner) detection according to the end-stopped cell implementation, as a function of simple- and complex-cells, proposed in [43].

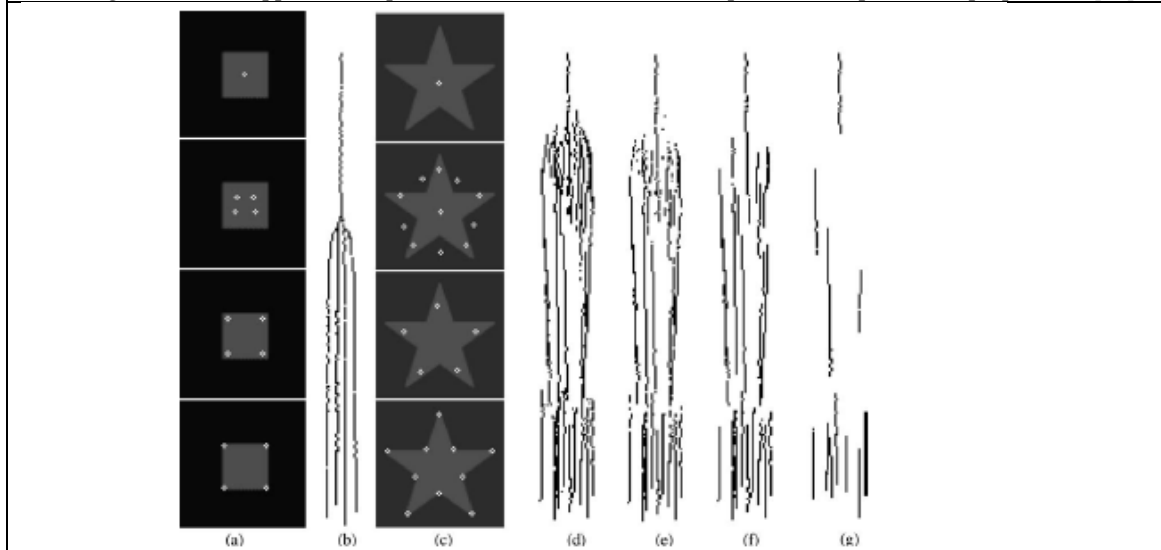


Fig. 4. Keypoint scale space, with finest scale at the bottom: (a) square, (b) projected 3D keypoint trajectories of square, (c) and (d) star and projected trajectories, (e) micro-scale stability, (f) and (g) stability over at least 10 and 40 scales, respectively.

Fig. 3.11-2(b). Reproduced with permission, courtesy of [43]. Keypoint scale space representation. Linking keypoints from coarse to fine scales can contribute to object segregation because keypoint trajectories converge from the contours at fine scales to the centres of objects at coarse scales.

### 3.12 Original definition of zero-crossing (ZX) pixels and scale-invariant keypoints in an even-symmetric and odd-symmetric wavelet-based filtered image

According to Marr [5] (Figure 2-21, p. 73), the raw primal sketch employs as input ZX pixels of the 2D intensity function  $I(x,y)$  to generate as output a discrete and finite set of multi-scale *tokens*, defined as discrete sub-symbolic image plane



entities: edges, blobs (closed contours), bars and terminators (discontinuities). In Marr's work, the following definitions hold, also refer to Chapter 3.5.3.

- In general, a ZX pixel is defined as a place (spatial unit, namely, pixel) where the value of a function passes from positive to negative (or vice versa) [5] (p. 54). In a broader sense, it means:
  - transitions from positive to non-positive, i.e., from positive to either zero or negative, or vice versa, or
  - transitions from negative to non-negative, i.e., from negative to either zero or positive, or vice versa.

For example, according to this definition Marr visualizes ZX lying on the bright side of the image contours exclusively [5] (Figure 2-12, 2-13, 2-14, 2-15, pp. 58-61).

- In the particular context of Marr's work, ZX pixels are intended as ZX pixels of the  $\nabla^2G$ -filtered image, such that  $\nabla^2(G*I) = (\nabla^2G)*I$  [5] (pp. 57, 58), where  $\nabla^2$  is the Laplacian operator ( $\partial^2/\partial x^2 + \partial^2/\partial y^2$ ) [5] (p. 54) and  $G$  is the 2D Gaussian function  $G(x,y)$ .

It is well known that  $\nabla^2G$  is a circularly symmetric (isotropic) even-symmetric Mexican-hat-shaped operator. It is also known that, in mathematics, the Hessian matrix (or simply the Hessian), defined as the square matrix of second-order partial derivatives of a function  $f$ , describes the local curvature (also called *concavity*) of a function of many variables (e.g., refer to [http://en.wikipedia.org/wiki/Hessian\\_matrix](http://en.wikipedia.org/wiki/Hessian_matrix)). For example, the Hessian matrix  $H$  of a function  $f(x,y)$  is:

$$H(f(x,y)) = \begin{vmatrix} \partial^2 f / \partial x^2 & \partial^2 f / \partial x \partial y \\ \partial^2 f / \partial x \partial y & \partial^2 f / \partial y^2 \end{vmatrix}$$

Hence, the following original definitions are proposed.

- DEFINITION 1: Based on the simulation proposed in [1] and extended in Fig. 3.11-1, the  $\nabla^2G$ -filtered image,  $(\nabla^2G)*I$ , is an estimate of the *local curvature*, also called *local concavity*, of the 2D function  $I(x,y)$ . The term “*local curvature*” = “*local concavity*” is synonym of “*second-order derivative*”, “*change in first-order derivative*”, “*change in local slope*” or “*change in gradient*”. For example, in a ramp, the gradient is constant, but the local curvature  $(\nabla^2G)*I$  is zero. In a flat image area, the gradient is zero and the local curvature  $(\nabla^2G)*I$  is zero. In particular,
  - If  $(\nabla^2G)*I > 0$  (positive value), then the estimated local concavity is down.
  - If  $(\nabla^2G)*I < 0$  (negative value), then the estimated local concavity is up.

For example, a horizontal line or a ramp, has no local curvature, then  $(\nabla^2G)*I = 0$ . Based on the simulation proposed in [1], the output value of an even-symmetric filter, like  $\nabla^2G$  or the real part of the Gabor wavelet employed in [1], which are both even-symmetric on-center and off-surround local filters (on-cell in the nomenclature of [22], p. 17) estimates the inverse of the local curvature of the (2D) image function, equivalent to the inverse of the local concavity, such that if local concavity is up, then output value is negative, while if local concavity is down then output value is positive. Therefore:

- (I) if across the **on-center (excitatory center) and off-surround (inhibitory surround)** portions of the domain of activation of the even-symmetric wavelet the intensity change (intensity first-order derivative) is zero or constant (i.e., the local intensity is a straight line, either horizontal or sloping), then there is zero-concavity (ZC) (local curvature is zero) and the wavelet output = 0. See Fig. 3.11-1.
- (II) if across the on-center (excitatory center) and off-surround (inhibitory surround) domain of activation of the even-symmetric wavelet there is a change in the intensity change (intensity first-order derivative  $\neq 0$ ) with



concavity down (up), i.e., there is a local curvature (concavity) either down or up because the local intensity is not a straight line, either horizontal or sloping, then wavelet output  $> 0$  ( $< 0$ ). The sign of the wavelet output ( $> 0$  if concavity is down,  $< 0$  if concavity is up) is the inverse of the sign of the second-order derivative ( $> 0$  if concavity is up,  $< 0$  if concavity is down) because the shape of the selected wavelet is even-symmetric on-cell [22], whereas the second derivative of a Gaussian is even-symmetric off-cell [22], in practice, adopted on-center off-surround even-symmetric wavelet = (1 - second derivative of a Gaussian). See Fig. 3.11-1.

This is in line with [11] (p. 213): "in two and higher dimensions there is no absolute relationship between locations of the Laplacian ZX curves and the local extrema of a signal. A Laplacian ZX curve may enclose either no extremum, one extremum, or more than one local extremum. Only in the one-dimensional case it holds that there is exactly one local extremum point between two ZXs of the second derivative".

To recapitulate, in the Marr's vision model for contour detection [5], at one single spatial scale of the local operator  $\nabla^2 G$ , contour pixels defined as ZX pixels in the  $\nabla^2 G$ -filtered image do not occur where there is a change in the local concavity value  $(\nabla^2 G) * I$  of the image (which would be equivalent to a third-order local derivative value of the image different from zero), but where the local concavity value  $(\nabla^2 G) * I$  of the image changes in sign, from positive to non-positive, i.e., from positive to either negative or zero values, or, vice versa, from negative to non-negative, i.e., from negative to either positive or zero values. Hence, according to Marr:

DEFINITION 2: according to the Marr's vision system model [5]:

Image contour pixels = pixels where there is a change in sign of the local concavity,  $(\nabla^2 G) * I$ .

In more details, the original unequivocal operational definition of an image-contour pixel is the following.

*A pixel  $I(n)$  with pixel coordinates  $n = (x, y)$  in a 2D array is an image-contour pixel if it is a zero-crossing (ZX) pixel, where the image local concavity, equal to  $[I(n) \bullet \partial^2 G / \partial n^2]$ , changes in sign, either from positive to non-positive, i.e., from positive to either zero or negative, or from negative to non-negative, i.e., from negative to either zero or positive, in comparison with the local concavity of any of its 8-adjacency neighboring pixel.*

Noteworthy, because the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution. Therefore, the filtered image  $(\partial^2 G / \partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  applied to the low-pass image adopted by both Canny [27] and Bertero, Torre and Poggio [44].

$$(\partial^2 G / \partial n^2 * I) \neq \partial^2 / \partial n^2 (G * I).$$

To the best of these authors' knowledge, this definition is alternative to those provided by the existing computer vision literature, including Canny [27], Burr and Morrone [26], Rodrigues and du Buf's [28], [43], Heitger et al. [41], Adelson and Bergen [42], and many others, e.g., refer to [25].

According to Marr, ZX pixels defined as one-scale local concavity values,  $(\nabla^2 G) * I$ , where there is a change in sign (namely, from positive to non-positive, i.e., from positive to either negative or zero, or from negative to non-negative, i.e., from negative to either positive or zero) in the local concavity, must be dealt with through scale according to the spatial coincident assumption [5] (p. 70) (refer to Chapter 3.5.3 above), such that if a ZX segment (note: this information primitive is not a ZX pixel, refer to Chapter 3.6) is present in a set of independent (multi-scale) channels (see [5], Figure 2-21, pp. 72-73) over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such ZX segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, illumination, surface orientation, etc.)", i.e., it corresponds to a "true" image contours.

In the present work, given the original aforementioned definition of ZX pixels detected by a single-scale local operator  $\nabla^2G$ , the definition of ZX pixels detected by a multi-scale bank of local filters  $\nabla^2G$  is still considered an open problem. In fact, multiple strategies to combine multi-scale (and multi-orientation) even-symmetric (and odd-symmetric) local filters can be found in the existing literature, e.g., refer to [25]-[28]. In the rest of this work, a novel operational definition to detect ZX pixels in a multi-scale bank of even-symmetric local filters is proposed.

- DEFINITION 3: Based on the simulation proposed in [1] and extended in Fig. 3.11-1, ZX pixels of the  $\nabla^2G$ -filtered image,  $(\nabla^2G)*I$ , are defined hereafter as pixels  $(x1,y1)$  where:

- $(\nabla^2G)*I \neq 0$ , i.e., the per-pixel  $(\nabla^2G)*I(x1,y1)$  value is either  $> 0$  or  $< 0$ , AND
- In the 8-adjacency neighborhood centered on pixel  $(x1,y1)$ , there is a change in sign,
  - from positive to non-positive, i.e., from positive to either negative or zero,
 or, vice versa,
  - from negative to non-negative, i.e., from negative to either positive or zero,

of the concavity (local curvature) of the 2D function,  $I(x,y)$ , estimated for that pixel  $(x1,y1)$ ,  $(\nabla^2G)*I(x1,y1)$ , with respect to the sign of the concavity of the 2D function,  $I(x,y)$ , estimated for any of the 8-adjacency neighboring pixels  $(nx1, ny1)$  of the central  $(x1,y1)$ ,  $n = 1, \dots, 8$ , where  $(\nabla^2G)*I(nx1,ny1)$  may be equal to zero.

In greater detail, based on the analysis of the (either positive or negative) 2ndOrderDerivEvenSymWaveletOutput values generated from step, ramp, line (ridge) and roof edges proposed in Fig. 3.11-1, an original operational definition of discrete (binary or trinary) and continuous ZX pixels detected by an even-symmetric oriented (non-isotropic) filter (like the real part of the complex Gabor wavelet computed in [1], which is different from the Marr isotropic Laplacian of Gaussian proposed in [5]) for every scale  $s = 1, \dots, S$ , is defined as follows:

1. In the direction of the gradient, perpendicular to the orientation  $orntn \in \{1, \dots, Orntn\}$  of the filter (eventually coincident with the direction of the physical contour, if any), if the current pixel  $x$  shows a  $2ndOrderDerivEvenSymWaveletOutput_{s,orntn}(x) > 0$  and the neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient  $coefficient_{s,orntn}(nx) == 0$ , then pixel  $x$  is a ZX on the bright side of the contour. Proof: ramp edge (a) in Fig. 3.11-1(IV) [1].

In practice, to account for the inaccuracy of the local boundary direction estimate in combination with the quantization error due to the discretization of the direction of the gradient, the following implementation is adopted: if the current pixel  $x$  shows a  $2ndOrderDerivEvenSymWaveletOutput_{s,orntn}(x) > 0$  and, at the same scale and orientation of pixel  $x$ , at least one 8-adjacency neighboring pixel  $nx$  shows coefficient  $coefficient_{s,orntn}(nx) == 0$  while all other neighboring pixels, if any, show coefficient  $coefficient_{s,orntn}(nx) > 0$ , then pixel  $x$  is a ZX on the bright side of the contour.

2. In the direction of the gradient, perpendicular to the orientation  $orntn \in \{1, Orntn\}$  of the filter (coincident with the direction of the potential physical contour, if any), if the current pixel  $x$  shows a  $2ndOrderDerivEvenSymWaveletOutput_{s,orntn}(x) < 0$  and the neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient  $coefficient_{s,orntn}(nx) == 0$ , then pixel  $x$  is a ZX on the dark side of the contour. Proof: ramp edge (a) in Fig. 3.11-1(IV) [1].

In practice, to account for the inaccuracy of the local boundary direction estimate in combination with the quantization error due to the discretization of the direction of the gradient, the following implementation is adopted: if the current pixel  $x$  shows a  $2ndOrderDerivEvenSymWaveletOutput_{s,orntn}(x) < 0$  and at least one 8-adjacency neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient  $coefficient_{s,orntn}(nx) == 0$



while all other neighboring pixels, if any, show coefficient $_{s, orntn}(nx) < 0$ , then pixel  $x$  is a ZX on the dark side of the contour.

3. In the direction of the gradient, perpendicular to the orientation  $orntn \in \{1, \dots, Orntn\}$  of the filter (coincident with the direction of the potential physical contour, if any), if the current pixel  $x$  shows a  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x) < 0$  (or  $> 0$ ) and the neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient $_{s, orntn}(nx) > 0$  (or  $< 0$ ), then both pixels  $x$  and  $nx$  are ZX pixels on the dark (or bright) and bright (or dark) side of the contour, respectively. Proofs: white line == ridge (a), dark line == ridge (b) in Fig. 3.11-1(XI) [1] and step edge in Fig. 3.11-1(III) [1].

In practice, to account for the inaccuracy of the local boundary direction estimate in combination with the quantization error due to the discretization of the direction of the gradient, the following implementation is adopted: if the current pixel  $x$  shows a  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x) < 0$  (or  $> 0$ ) and at least one 8-adjacency neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient $_{s, orntn}(nx) > 0$  (or  $< 0$ ), both pixels  $x$  and  $nx$  are ZX pixels on the dark (or bright) and bright (or dark) side of the contour, respectively.

To recapitulate, the OR-combination of the aforementioned ZX conditions 1 to 3 is summarized as follows.

DEFINITION 4: For a given  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x)$  convolution product generated from an input image, if the current pixel  $x$  shows a  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x) < 0$  (or  $> 0$ ) and at least one 8-adjacency neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ , coefficient $_{s, orntn}(nx) \geq 0$  (or  $\leq 0$ ), both pixels  $x$  and  $nx$  are ZX pixels on the dark (or bright) and bright (or dark) side of the contour, respectively.

Noteworthy, because the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution [88]. Therefore, the filtered image ( $\partial^2 G/\partial n^2 * I$ ) is different from the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  applied to the low-pass image adopted by both Canny [27] and Bertero, Torre and Poggio [44].

$$(\partial^2 G/\partial n^2 * I) \neq \partial^2/\partial n^2 (G * I)$$

According to the aforementioned operational definition 4 of ZX pixels, at a given scale  $s$  and orientation  $orntn$ , a ZX pixel never features a  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x) = 0$ , i.e., a  $2ndOrdDerivEvenSymWaveletOutput_{s, orntn}(x)$  of a ZX pixel  $x$  is always  $> 0$  or  $< 0$ .

To the best of this author's knowledge, the definition 4 of ZX pixels is novel and original. For example, after a Google search of keywords: "image contour as change of sign in the image concavity (equivalent to local curvature)", no similar definition was found in the existing literature. This original definition of ZX pixels can be considered an enhancement of that proposed in [1]. The important conclusion of the present work, in line with [1], is that an even-symmetric filter, like the real-part of the oriented Gabor filter shown in Fig. 3.11-1(I), is necessary and sufficient to detect all kinds of 1D edges shown in Fig. 3.11-1(III) to Fig. 3.11-1(XIII), in line with one of the best-known brightness illusions, namely, the Mach bands illusion, when dealing with the ramp edge, see Fig. 3.4-1.

For a ZX pixel, which satisfies the definition 4 proposed above, the following properties hold.

- If the pixel is ZX according to the aforementioned definition 4 and  $(\nabla^2 G) * I > 0$  (positive value), then the estimated local concavity is down and the ZX pixel lies on the bright side of an image boundary (contour).
- If the pixel is ZX according to the aforementioned definition 4 and  $(\nabla^2 G) * I < 0$  (negative value), then the estimated local concavity is up and the ZX pixel lies on the dark side of an image boundary (contour).
- ZX pixels, which satisfy the aforementioned definition 4, can be, but do not have to be, local extrema in the  $\nabla^2 G$ -filtered image,  $(\nabla^2 G) * I$ , i.e., there are ZX pixels that are not local extrema (either local maxima or local minima)

in the  $\nabla^2G$ -filtered image, e.g., see Fig. 3.11-1(XII). In addition, not all local extrema in the  $\nabla^2G$ -filtered image,  $(\nabla^2G)*I$ , are ZX pixels. For example, in [21], Lowe searches for scale-invariant keypoints (scale invariant feature transform, SIFT) as local extrema in the difference of Gaussian (DOG), equivalent to a  $\nabla^2G$ -filtered image, that are not, simultaneously, ZX pixels, i.e., pixels whose second derivative, estimated by means of a Hessian matrix (refer to this text above), is strong because they lie along edges, refer to the paragraph below.

- It is noteworthy that, in [21], Lowe defines scale-invariant keypoints (scale invariant feature transform, SIFT) as local extrema in the DOG function, equivalent to a  $\nabla^2G$ -filtered image, where local extrema in the DOG pyramid are detected by comparing a pixel at one scale of the DOG representation to its 26 neighbors in 3x3 regions at the current and adjacent scales. According to [23], "a well-known property of the scale-space representation (refer to Chapter 3.4.1) is that the amplitude of spatial derivatives in general decreases with scale, i.e., if a signal is subjected to scale-space smoothing, then the numerical values of spatial derivatives computed from the smoothed data can be expected to decrease. This is a direct consequence of the non-enhancement property of local extrema, which state that the value at a local maximum cannot increase and the value at a local minimum cannot decrease". In addition to being defined as local extrema through three adjacent scales of the DOG function, there is one more difference between SIFT points and ZX pixels. For stability purposes, the SIFT keypoints are further constrained to remove points equivalent to unstable extrema in DOG affected by:
  - low contrast (i.e., low value of the first-order derivative of the intensity function), and
  - their spatial vicinity to image-edges, where the location of a local extremum in DOG is poorly determined, i.e., it is unstable to small amounts of noise, although the DOG function has a strong response along edges. In [21], to detect local extrema of the DOG function along edges, a procedure based on a Hessian matrix (refer to this text above) consisting of second derivatives estimated in neighboring sample points is implemented. On the contrary, it would be easier to identify these points along edges based on the ZX pixel selection strategy inspired to Marr's work [5].

In practice, local extrema in the  $\nabla^2G$ -filtered image which are also ZX pixels (lying along image contours) are rejected as SIFT points. As a consequence, according to Lowe [21], SIFT and ZX pixels are complementary, i.e., their inter-set overlap is zero. Noteworthy, the OR-combination of SIFT pixels, which are necessarily local extrema of the  $\nabla^2G$ -filtered image, and ZX pixels, which may or may not be local extrema of the  $\nabla^2G$ -filtered image, always overlaps with the set of local extrema in the  $\nabla^2G$ -filtered image, i.e.,  $(\text{Lowe's SIFT} \cup \text{Marr's ZX pixels}) \supseteq \text{local extrema in the } \nabla^2G\text{-filtered image}$  (also refer to Chapter 3.5).

- In functional terms, end-stopped cells in [41], [43], where "keypoints" of an image are defined as the peaks (local maxima in a 3 x 3 neighbourhood) in the summed end-stopped representation, are one-to-one related to, i.e., they are the biological counterpart of, the Lowe SIFT operators, although keypoint estimation through simple- and double-stopped operators, see Eq. (5-2) and Eq. (5-3), adopted in [41] and [43] is not computationally equivalent to Lowe's [21].

To recapitulate, in agreement with [1], the present work considers an even-symmetric filter, like the real-part of the oriented Gabor filter shown in Fig. 3.11-1(I), as necessary and sufficient to detect all kinds of 1D edges shown in Fig. 3.11-1(III) to Fig. 3.11-1(XIII), in compliance with one of the best-known brightness illusions, namely, the Mach bands illusion, see Fig. 3.4-1.

It is worth mentioning that, according to [1], odd-symmetric filters (e.g., an imaginary part of a complex Gaussian filter), such as that depicted in Fig. 3.11-1(I), are found to become useful in combination with even-symmetric filters in the detection of 2D contours in (2D) imagery. In the present work (see Fig. 3.11-1), the original "perceptual contrast"  $\text{PrcptlCntrst2}(x)$  expression proposed in Chapter 3.11,

$$\checkmark \quad \text{PrcptlCntrst2}(x) = \text{abs}[f(x) \circ \partial^2 G/\partial x^2] + \text{abs}[f(x) \circ \partial G/\partial x]/2, \text{ hence } \text{PrcptlCntrst2}(x) \geq 0. \quad (11-1)$$



combines even- and odd-symmetric filter responses, such that  $\text{PrcptlCntrst2}(x) > 0$  values are a superset of both contour pixels (ZX pixels, in line with the Marr's model of vision [5]), see Fig. 3.11-1, and scale-invariant keypoints (as those defined as SIFT by Lowe [21] and those detected by the computational model of end-stopped cells proposed in [41], [43], see Fig. 3.11-2).

### 3.13 Original perceptual image-pair quality/ similarity/ dissimilarity index/ metric

In the words of Iqbal and Aggarwal: “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception... This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [49].

Objective (quantitative) quality evaluation for images and video, also called objective image quality assessment (IQA), can be classified into two board types: signal fidelity measures and perceptual visual quality metrics (PVQMs) [20], [99], [108].

Objective IQA metrics can be classified as: (i) full reference (FR), when an original (distortion-free) image is available for comparison with a distorted image. Most of the existing approaches are known as full-reference, meaning that a complete reference image is assumed to be known. (ii) No-reference (NR). In many practical applications, a no-reference or “blind” quality assessment approach is desirable and the reference image is not available. (iii) Reduced-reference (RR) methods, when the reference image is only partially available, in the form of a set of extracted features made available as side information to help evaluate the quality of the distorted image [108].

The signal fidelity measures refer to the traditional MAE (mean absolute error), MSE (mean square error), SNR (signal-to-noise ratio), PSNR (peak SNR), etc. Although they are simple, well defined, with clear physical meanings and widely accepted, signal fidelity measures can be a poor predictor of perceived visual quality, especially when the noise is not additive. For example, MAE and MSE are pixel-by-pixel differences, i.e., these statistics are non-contextual and position-dependent. Since they consider a (2D) image as a 1D string/sequence of pixel-specific vectors, i.e., they ignore any spatial contextual information, either topological or non-topological, therefore they are inconsistent with visual perception. In addition, being image position-dependent, they are sensitive to image rotations.

According to a relevant portion of the computer vision literature, the primary use of image quality metrics is to quantitatively measure an image quality that correlates with perceptual visual quality. So-called perceptual visual quality metrics, PVQMs, are objective models for predicting subjective visual quality scores, like the resultant mean opinion score (MOS) obtained by many observers through repeated viewing sessions [99]. In spite of the recent progress in related fields, objective evaluation of picture quality in line with human perception is still a long and difficult odyssey due to the complex, multi-disciplinary nature of the problem (related to physiology, psychology, vision research and computer science) [99]. For example, cognitive understanding, prior knowledge and interactive visual processing (e.g., eye movements) influence the perceived quality of images; this is the so-called cognitive interaction problem [100]. A human observer will give different quality scores to the same image if s/he is provided with different instructions. Prior information regarding the image content, or attention and fixation, may also affect the evaluation of the image quality. But most image quality metrics do not consider these effects, they are difficult to quantify and not well understood [100]. It is clear that, unlike so-called signal fidelity measures, PVQMs have to quantify the spatial difference (e.g., position difference in image contours) together with the spectral difference (e.g., image-wide difference in spectral means) between a reference and test image pair [101], [102], [103]. There are two major categories of PVQMs with regard to reference requirements: double-ended and single-ended. Double-ended metrics require both the reference (original) signal and the test (processed) signal, and can be further divided into two subclasses: reduced-reference (RR) metrics that need only part of the reference signal and full-reference (FR) ones that need the complete reference signal. Single-ended metrics use only the processed signal, and are therefore also called no-reference (NR) ones. Most existing PVQMs are FR ones [100], e.g., the popular univariate (one-channel) “universal” (scalar) image quality index (UIQI), or Q index for brevity [104], which was further generalized into the so-called structural similarity (SSIM) index [99], [100]. A multi-scale implementation of the single-scale SSIM index was proposed by the same authors [121]. Noteworthy, although SSIM is considered a PVQM, it does not appear to be provided with a perceptual relevance on a strong theoretical ground, in fact SSIM bears certain similarities with traditional

signal fidelity measures, such as the MSE [99]. This is clearly explained in [105] whose conclusions are quoted as follows: “In both an empirical study and a formal analysis, evidence of a relationship between the increasingly popular structural similarity index and the conventional mean squared error is uncovered. This research is perhaps the first to uncover a statistical link of this nature and likely the only in which a formal connection is established... Collectively, these findings suggest that the performance of the SSIM is perhaps much closer to that of the MSE than some might claim. Consequently, one is left to question the legitimacy of many of the applications of the SSIM. Ultimately, this investigation once again illustrates the enormous gap that continues to exist between an automated measure of quality and that of the human mind. Until a more radical approach is considered, this problem will likely continue to confound researchers in the field.”

To recapitulate, in a PVQM, quantitative spatial and spectral (2-D) image QIs must to be estimated jointly, to be validated by the MOS collected from a group of human subjects [99], e.g., refer to [106] for a detailed description of a visual analysis of PAN-sharpened MS images.

Differently from the SSIM, one of the same co-authors, Eero Simoncelli, recently proposed in [107] a PVQM based on a normalized Laplacian pyramid for image analysis and synthesis. The proposed perceptual metric is given by:

$$D(I_R, I_T) = \frac{1}{S} \sum_{s=0}^{S-1} \frac{1}{\sqrt{SF_s}} \|\hat{I}_{R,s} - \hat{I}_{T,s}\|^2, \quad (13-1)$$

where  $I_R$  and  $I_T$  are the reference and the test image respectively,  $\hat{I}_{R,s}$  and  $\hat{I}_{T,s}$  denote vectors containing the transformed reference and distorted image data at scale  $s = 0, \dots, S-1$ , respectively, and where  $SF_s$  is the number of spatial filters in the subband at scale  $s$ . In this equation a root mean squared error is computed for each scale, and then averaged over these scales giving larger weight to the lower frequency coefficients (which are fewer in number, due to subsampling). Per se, this scale-dependent weighting policy is not supported by any perceptual plausibility.

In [108], a so-called non-shift edges (NES) index is proposed as an image quality assessment metric based on early vision features. NES is proposed as:

$$NSE(A, B) = F_A \cap F_B, \quad (13-2)$$

where  $A$  and  $B$  denote the reference and distorted images whose edge maps are  $F_A$  and  $F_B$ , respectively. An edge map is a binary image, where “1” denotes an edge point and “0” denote a non-edge point. Obviously, the NSE map can be calculated by the “AND” operation of the two binary edge maps, or other combinations of omission and commission errors with respect to the reference edge map.

Alternative to the normalized Laplacian pyramid proposed in [107], the single-scale even-symmetric and Gaussian filter bank proposed in Chapter 3.11 provides a 1D signal decomposition into high- and low-frequency components and a zero-crossing (ZX) signal-contour detection consistent with the Mach bands illusion in human visual perception. It is such that:

$$\checkmark \text{ Renstrct}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G / \partial x^2] / 2, \quad (11-2)$$

where  $\{f(x) \circ G(x)\} \in [0, \text{MaxGrayValue}]$  and  $\{[f(x) \circ \partial^2 G / \partial x^2] / 2\} \in [-\text{MaxGrayValue} / 2, \text{MaxGrayValue} / 2]$ . According to these properties, the proposed perceptual visual dissimilarity metric (PVDM) is (attention: for the time being, this index ignores multiple spatial scales and orientations of 2D spatial filter banks involved with near-orthogonal image decomposition and analysis):

$$PVDM(I_R, I_T) = |I_R(x) \circ G(x) - I_T(x) \circ G(x)| + \left| \frac{[I_R(x) \circ \frac{\partial^2 G}{\partial x^2}]}{2} - \frac{[I_T(x) \circ \frac{\partial^2 G}{\partial x^2}]}{2} \right|, \quad (13-3)$$

where  $|I_R(x) \circ G(x) - I_T(x) \circ G(x)| \in [0, \text{MaxGrayValue}]$  and  $|[I_R(x) \circ \partial^2 G / \partial x^2] / 2 - [I_T(x) \circ \partial^2 G / \partial x^2] / 2| \in [0, \text{MaxGrayValue}]$ . In mathematical terms, this is a Minkowski distance with degree  $d$  equal to 1. If appropriate,  $d$  can be set equal 2 (to apply to a Euclidean distance) or superior. Since it is a perceptual visual quality distance rather than a perceptual visual similarity indicator,  $PVDM(I_R, I_T) \in 2 * [0, \text{MaxGrayValue}]$  is best when minimized.



The difference between the two PVQMs proposed in Eq. (13-3) and Eq. (13-1) is twofold. First, the two PVQM formulations are quite different, where Eq. (13-3) provides analysis and synthesis of two separate low- and high-spatial frequency components. Second, spatial information primitives employed in Eq. (11-2) at the basis of Eq. (13-3), with special regard to even-symmetric spatial filters consistent with the Mach bands illusion in human visual perception, are different from those implemented in Eq. (13-1) proposed in [107] whose consistency with the Mach bands illusion is unknown or questionable.

### 3.14 Generalization in mathematical and linguistic terms of one specific statement by Marr regarding surface discontinuity detection in a 2½D sketch

From his seminal book [5], the following three quotes by Marr are highlighted.

1. “The two ideas underlying the detection of (image) intensity changes are (1) that intensity changes occur at different spatial scales in an image (2D array function), and so their optimal detection requires the use of operators of different sizes; and (2) that a sudden intensity change will give rise to a peak in the first derivative or, equivalently, to a ZX in the second derivative. A ZX (of a function) is a place where the value of a function passes from positive to negative” (p. 54).
2. “(As part of the business of) the 2½D sketch, discontinuities in the distance of visible surfaces from the viewer must be made explicit” (p. 81) “in a viewer-centered coordinate frame” (p. 37).
3. “The second type of clue to surface discontinuity consists of discontinuities in various parameters that describe the spatial organization (of tokens in an image)... We isolated six image properties that are useful to measure, three of them intrinsic to a token - average brightness, size (perhaps length and width), and orientation – and three pertaining to the spatial arrangement of tokens – their local density, distance apart, and the orientation structure.” (p. 93).

Once the aforementioned first statement is rephrased according to the previous Chapter 3.5 and Chapter 3.6, the following considerations stem from these quotes.

First, according to Marr, in a function (signal, variable, sensory data stream), either univariate or multivariate, where sensory data are related to the concept of quantitative unequivocal information-as-thing [31], [32], information (rather related to the concept of qualitative equivocal information-as-interpretation, which is complementary to the concept of information-as-thing [31], [32]) is conveyed where function discontinuities occur.

In the vocabulary of a natural language, proposed synonyms of discontinuities in a univariate or multivariate function (signal, variable, sensory dataflow) are:

- singularities,
- abrupt changes (different from slowly varying or smooth changes).

Although it may appear trivial, the proposed definition of function discontinuities implies the following non-trivial effects, also refer to Fig. 3.11-1. First, in a ramp function, where there is a constant change (constant gradient or function first-order derivative), irrespective of its slope (including zero slope), there is no along-ramp function discontinuity (in accordance with the Mach bands illusion in vision [25]), i.e., along a ramp function there is no *information* (defined as above). Noteworthy, along a ramp function, spatial autocorrelation is maximum in absolute values (either 1 or -1), e.g., refer to Moran’s I [39] and Geary’s C [40] spatial autocorrelation functions. It means that, in a ramp function, information is located exclusively at the beginning and termination of the ramp, e.g., see Fig. 3.11-1-(III) and Fig. 3.11-1-(IV).

Driven from the previous paragraph, the following conjecture is proposed: in a 1- to 3D spatial function, information, i.e., function discontinuities in the space domain, are not where the spatial autocorrelation is equal to  $\pm 1$ , like intensity homogeneous areas (whose spatial autocorrelation is +1) or ramp-like areas (where spatial autocorrelation can be  $\pm 1$ ).

Neither a function discontinuity defined as above is located anywhere there is a change in gradient/slope/first-order derivative, i.e., anywhere the function second-order derivative (concavity estimate) is different from zero, or where there





is a local maximum or minimum of the first-order derivative, such that the second-order derivative is equal to zero, unless the change in gradient implies a change in the sign of the concavity, e.g., see Fig. 3.11-1-(V)(c).

Rather, an operational definition of function discontinuities is the following.

Given any function (signal, variable, sensory dataflow, etc.), either one or multivariate, e.g. a function in the 1D time, 2D or 3D spatial space or 4D time-space domain (for example, consider a multi-layer data set  $H(x,y)$ , defined in a regular 2D grid  $(x, y)$  or in a scattered 2D domain of points  $(x, y)$ , such that the single-layer (scalar) data set  $H1 = I(x,y)$  is the intensity value of each point  $(x, y)$ , while the single-layer (scalar) data set  $H2 = Z(x,y)$  is the range value (distance-from-the-viewer), collected for each (2D) point coordinate  $(x, y)$ , by a Light Detection And Ranging (LiDAR) system), the function discontinuities (singularities) are located where ZX pixels of the function's second-order derivative, i.e., ZX pixels of the function's local concavity estimate, occur.

According to the previous Chapter 3.5 and Chapter 3.6, ZX pixels of a generic function occur where the function changes its values:

- from positive to non-positive, i.e., from positive to zero or negative values, or vice versa, or
- from negative to non-negative, i.e., from negative to zero or positive values, or vice versa.

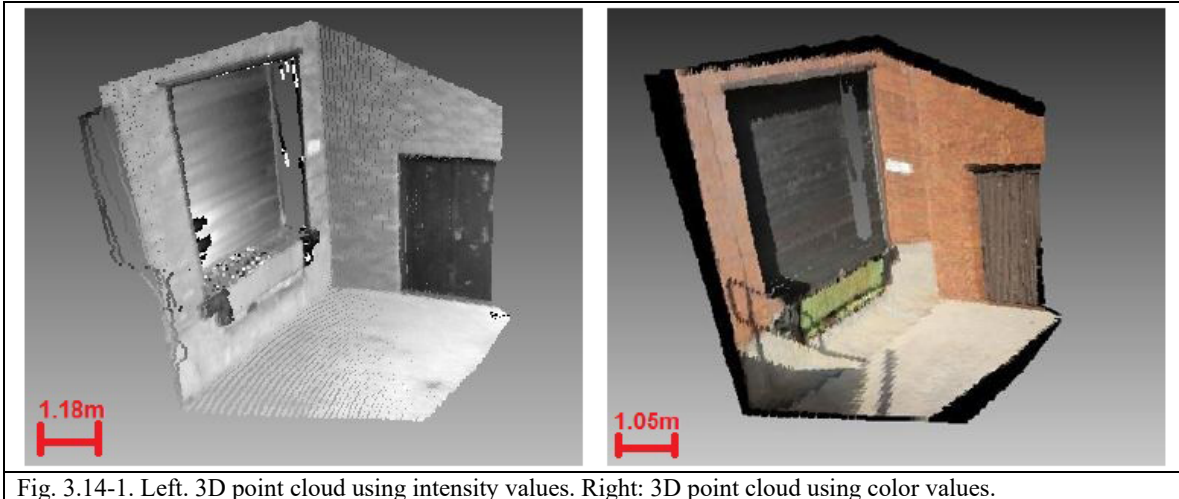
In common practice, the simplest 1D and 2D implementations of an isotropic second-order derivative operator are the 1D  $\partial^2 G/\partial x^2$  and the 2D Laplacian operator,  $\nabla^2 G(x,y) = \partial^2 G/\partial x^2 + \partial^2 G/\partial y^2$  [5] (pp. 55-57).

The Gaussian function is a well-known separable kernel, i.e., the Gaussian kernel for dimensions higher than one, say  $N$ , can be described as a regular product of  $N$  1-dimensional Gaussian kernels, for example:  $g_{2D}(x, y; \sigma_1^2 + \sigma_2^2) = g_{1D}(x; \sigma_1^2) \times g_{1D}(y; \sigma_2^2)$ , where symbol  $\times$  identifies the (regular) product operator (not to be confused with the dot (scalar) product, vector product, etc.) [33].

Hence, for example, an  $N$ -band ( $N$ -layer) function of two variables,  $H(x, y)$ , defined in a 2D  $(x, y)$  domain, can be investigated for layer-specific discontinuities by a multi-scale bank of (separable) 1D  $\partial^2 G/\partial x^2$  and  $\partial^2 G/\partial y^2$  operators, such that, in sequence:

- layer-specific filter outputs in the  $x$ - and  $y$ -direction must be combined per scale by means a regular product for each domain point  $(x, y)$ ,
- layer-specific single-scale ZX pixels are detected,
- layer-specific single-scale ZX pixels are combined across scales and, finally,
- layer-specific multi-scale ZX pixels are combined across layers.

For example, a LiDAR imaging system is an active remote sensing technique that uses a laser and internal mirrors to scan an area. The time it takes a laser beam to reach an object, reflect, and return to a detector, is used to find the range (distance) of the object from the active laser source. A discrete LiDAR stores a point for each surface the laser reflects off, generating a 3D point cloud. Each point in the cloud contains an  $(x; y; z)$  spatial coordinate (made possible through the knowledge of angular scans, or use of a GPS and inertial measurement unit, depending on whether the unit is airborne or ground-based) and an intensity scalar (panchromatic) value that represents the strength of the reaction of the laser off a given surface. If an intensity image is created from the 3D point cloud and a color image is aligned with the intensity image, then the aligned intensity and color images can be registered and output as a 3D point cloud. Each pixel in the intensity image, for which there is an associated  $(x; y; z)$  point cloud coordinate, is assumed to match to the pixel in the color image at the same location. The red, green, and blue pixel values from the color image are merged with the  $(x; y; z)$  values from the intensity image to create a six element vector for each pixel. These 4- or 6D vectors are then written out to a Stanford polygon file (.ply format) so that they can be viewed as either panchromatic or chromatic 3D point clouds. Fig. 3.14-1 shows two 3D point clouds for the same scene, one colored with intensity values from the LiDAR, and the other colored with values from a color image acquired by a color camera (whether or not it is mounted onboard the LiDAR system) [34].



Instances of 3D point clouds  $(x; y; z)$ , colored with panchromatic or chromatic values, like those shown in Fig. 3.14-1, are expected to be a valuable test case to be input to a  $2\frac{1}{2}$ D sketch, required to be capable of detecting surface discontinuities. As an example, let us identify as  $Z(x, y)$  the range data set and  $I(x, y)$  the intensity value of each point  $Z(x, y)$ . In compliance with the previous paragraphs, in a 3D point cloud provided with intensity values, discontinuities can be detected automatically (i.e., without user interactions, required to define the system's free-parameters, if any, based on heuristics), by means of a  $2\frac{1}{2}$ D sketch whose system's architecture and implementation can be summarized as follows.

- (i) First, let us consider function  $Z(x, y)$ , exclusively. Design and implement an S-scale bank of 1D (separable)  $\partial^2 Gs/\partial x^2 = g_{1D}(x; \sigma_1^2)$  and  $\partial^2 Gs/\partial y^2 = g_{1D}(y; \sigma_2^2)$  operators,  $s = 1, \dots, S$ , such that the following property holds:  $g_{2D}(x, y; \sigma_1^2 + \sigma_2^2) = g_{1D}(x; \sigma_1^2) \times g_{1D}(y; \sigma_2^2)$ , where symbol  $\times$  identifies the (regular) product operator, e.g., refer to separable filter design and implementation principles proposed in [20], [33], [35], [92].
- (ii) Across the 2D  $(x, y)$  domain of function  $z(x, y)$ , run along the x-dimension an S-scale bank of 1D  $\partial^2 Gs/\partial x^2$  operators,  $s = 1, \dots, S$ , such that  $Evnsymtrcfltrd\_Zx\_s(x,y) = [(\partial^2 Gs/\partial x^2) * z(x, y)]$ ,  $s = 1, \dots, S$ ,  $x = 1, \dots, MaxX$ ,  $y = 1, \dots, MaxY$ .
- (iii) Like in (2), along direction y, to compute  $Evnsymtrcfltrd\_Zy\_s(x,y) = (\partial^2 Gs/\partial y^2 * z(x, y))$ ,  $s = 1, \dots, S$ ,  $x = 1, \dots, MaxX$ ,  $y = 1, \dots, MaxY$ .
- (iv) For each scale  $s = 1, \dots, S$ , compute the (regular) product  $Evnsymtrcfltrd\_Z\_s(x,y) = Evnsymtrcfltrd\_Zx\_s(x,y) \times Evnsymtrcfltrd\_Zy\_s(x,y)$ ,  $x = 1, \dots, MaxX$ ,  $y = 1, \dots, MaxY$ .
- (v) For each scale  $s = 1, \dots, S$ , compute ZX pixels of function  $Evnsymtrcfltrd\_Z\_s(x,y)$ ,  $x = 1, \dots, MaxX$ ,  $y = 1, \dots, MaxY$ .
- (vi) Combine across scale ZX pixels of function  $Evnsymtrcfltrd\_Z\_s(x,y)$ . This step is kept vague, although it is accomplished by means of an original solution in the rest of this document.
- (vii) Do the same as steps (1) to (7) for  $I(x, y)$ .
- (viii) Combine either continuous or binary discontinuities (ZX pixels) collected at steps (6) and (7) from the two-variable functions  $Z(x, y)$  and  $I(x, y)$  by means of, respectively, a fuzzy OR operator (MAX) or a logical OR operator.

Noteworthy, in [20], multi-scale filter banks for color image analysis (decomposition), image synthesis (reconstruction) and image-contour detection are proposed.

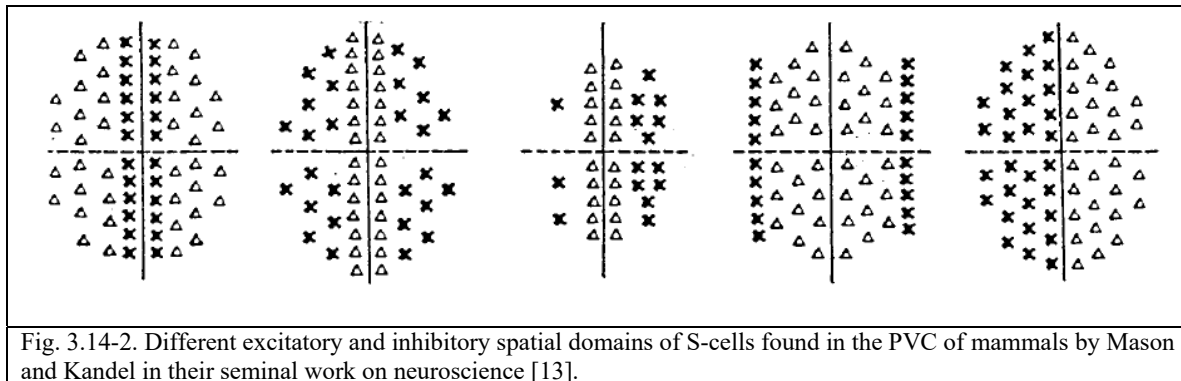


Fig. 3.14-2. Different excitatory and inhibitory spatial domains of S-cells found in the PVC of mammals by Mason and Kandel in their seminal work on neuroscience [13].

### 3.15 Original design of the 2D multi-scale even- and odd-symmetric Gabor filter bank

About the design of the filter bank, refer to Fig. 3.5-2, supported by Fig. 3.14-2 and Fig. 3.1-8.

- Because the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution [88]. Therefore, the filtered image  $(\partial^2 G/\partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  applied to the low-pass image adopted by both Canny [27] and Bertero, Torre and Poggio [44], refer to Chapter 3.11.

$$(\partial^2 G/\partial n^2 * I) \neq \partial^2/\partial n^2 (G * I) \quad (11-3)$$

- A Gabor filter is a Gaussian function with spread  $\sigma$  modulated by a complex sinusoid with central frequency  $F$ . To approximate a simple cell of the cat visual cortex, a physical model-based relationship between these two parameters was found to be  $F \cdot \sigma = 0.22$  [1].
- The real part of an oriented Gabor mother-wavelet with  $F \cdot \sigma = 0.22$  can be considered equivalent to an even-symmetric 2<sup>nd</sup>-order derivative of a Gaussian function. According to [1], [5], this local filter is necessary and sufficient to detect any sort of image contours, namely, step edge, roof, line (ridge) and ramps (according to the Mach bands illusion [25], refer to Fig. 3.4-1).
- The imaginary part of an oriented Gabor mother-wavelet with  $F \cdot \sigma = 0.22$  can be considered equivalent to an odd-symmetric 1<sup>st</sup>-order derivative of a Gaussian function. According to [1], [5], it is not further employed in image contour detection.
- The real part of an oriented Gabor mother-wavelet with  $F \cdot \sigma = 0.22$  provides a 3<sup>rd</sup>-order statistic in line with the works by Yellott [2] and Victor [3].
- To be convolved with a (2D) image, a quasicomplete 2D Gabor filter bank must be optimized into 1D separable filters [20], [33], [35], [92].

In a Gaussian distribution:

- $\mu \pm \sigma$  covers 68.27% of the area underneath the curve,
- $\mu \pm 2 \times \sigma$  covers 95.45% of the area underneath the curve,
- $\mu \pm 3 \times \sigma$  covers 99.73% of the area underneath the curve.

The size of the receptive field of a Gabor filter's Gaussian shape modulated by a complex sinusoid is approximated by a number of  $\sigma$ s before truncation equal to `NO_OF_GAUSSIAN_SIGMA_BEFORE_TRUNCATION = 6`.

Thus,  $\sigma_{\min} = \text{GABOR\_FILTER\_MIN\_SIZE\_IN\_PIXELS} / \text{NO\_OF\_GAUSSIAN\_SIGMA\_BEFORE\_TRUNCATION} = 3 / 6 = 0.5$ .



Two spatial orientations, 0 and 90 degrees, and four dyadic spatial scales are computed one octave apart, in agreement with [16]. Hence, the multi-scale filter size is:

- Scale 1 (finest),  $\sigma_{\min} = \sigma_1 = 0.5$ , thus filter size in pixels  $\text{NO\_OF\_GAUSSIAN\_SIGMA\_BEFORE\_TRUNCATION} \times \sigma_1 = 6 \times 0.5 = 3$  pixels. Spatial stride (inter-filter center distance) =  $2^0 = 1$  pixel unit.
- Scale 2,  $\sigma_2 = 2 \times \sigma_1 = 2^1 \times \sigma_{\min} = 1$ , thus filter size in pixels  $\text{NO\_OF\_GAUSSIAN\_SIGMA\_BEFORE\_TRUNCATION} \times \sigma_2 = 6 \times 1 \approx 7$  pixels. Spatial stride (inferred) =  $2^1 = 2$  pixel units.
- Scale 3,  $\sigma_3 = 2 \times \sigma_2 = 2^2 \times \sigma_1 = 2$ , thus filter size in pixels  $\text{NO\_OF\_GAUSSIAN\_SIGMA\_BEFORE\_TRUNCATION} \times \sigma_3 = 6 \times 2 \approx 13$  pixels. Spatial stride (inferred) =  $2^2 = 4$  pixel units.
- Scale 4 (coarsest),  $\sigma_4 = 2 \times \sigma_3 = 2^3 \times \sigma_1 = 4$ , thus filter size in pixels  $\text{NO\_OF\_GAUSSIAN\_SIGMA\_BEFORE\_TRUNCATION} \times \sigma_4 = 6 \times 4 \approx 25$  pixels. Spatial stride (inferred) =  $2^3 = 8$  pixel units.

Only Type III cells [20], see Fig. 3.1-8, that are spatially opponent but not color opponent, are implemented.

These wavelet have no zero frequency component, i.e., they respond zero at the zero frequency, hence the mean of these filter coefficients must equal zero.

Noteworthy, the real part (even-symmetric filter) of the proposed oriented Gabor mother-wavelet shown in Fig. 3.5-2 provides a 3rd-order statistic in line with the works by Yellott [2] and Victor [3]. The (E)TAU theorem proposed by Yellott states that if two panchromatic (multi-gray leveled) images feature identical image-wide third-order statistics, then this is a sufficient condition to state that those two images are visually identical (up to spatial translation) [2], refer to Chapter 3.2.4.

In contradiction with his own TAU theorem, in more recent years Yellott wrote that “very discrete, finite image is uniquely determined by its (two-dimensional) dipole histogram” [52], defined as follows (p. 487).

“A one-dimensional dipole is a triple, (d, a, b), with d an integer-valued displacement, and a and b real numbers. We shall say that a (one-dimensional) dipole (d, a, b) bridges a pair (x1, x2) of pixels in I if  $x_2 - x_1 = d$ ,  $I[x_1] = a$ , and  $I[x_2] = b$ . The dipole histogram DI assigns to each dipole (d, a, b) the number of distinct pairs in I bridged by (d, a, b). Thus, if  $DI(4, 0, 2) = 16$ , then there are 16 pixels x of I such that  $I[x] = 0$ , and  $I[x+4] = 2$ . (Note that this definition of DI requires that the image I be finite.) A two-dimensional dipole is a triple, (d, a, b), with d=(dx, dy), for dx a horizontal, and dy a vertical, integer-valued displacement, and a and b real numbers. For any two-dimensional image I, a (two-dimensional) dipole (d, a, b) is said to bridge a pixel pair ((x1, y1), (x2, y2)) in I if  $x_2 - x_1 = dx$ ,  $y_2 - y_1 = dy$ ,  $I[x_1, y_1] = a$ , and  $I[x_2, y_2] = b$ . As in the one-dimensional case, the dipole histogram DI assigns to each dipole (d, a, b) the number of pixel pairs in I bridged by (d, a, b). For any dipole (d, a, b), regardless of the dipole’s dimensionality, d is called the dipole’s displacement; a is called its a-value, and b is called its b-value.”

If the (E)TAU conjecture holds, then the proposed oriented even-symmetric filter design is sufficient to assess whether two images are perceptually identical, when they share the same image-wide combinations of local statistics, up to third-order statistics.

**3.16 Original implementation of a stratified multi-scale multi-orientation near-orthogonal image analysis/decomposition and synthesis/reconstruction**

In mammals, a vision system is comprised of a pre-attentive vision first phase and an attentive vision second phase, refer to Chapter 3.4.

	Scale 0	Scale 1	Scale 2	Scale 3	Scale 4
--	---------	---------	---------	---------	---------

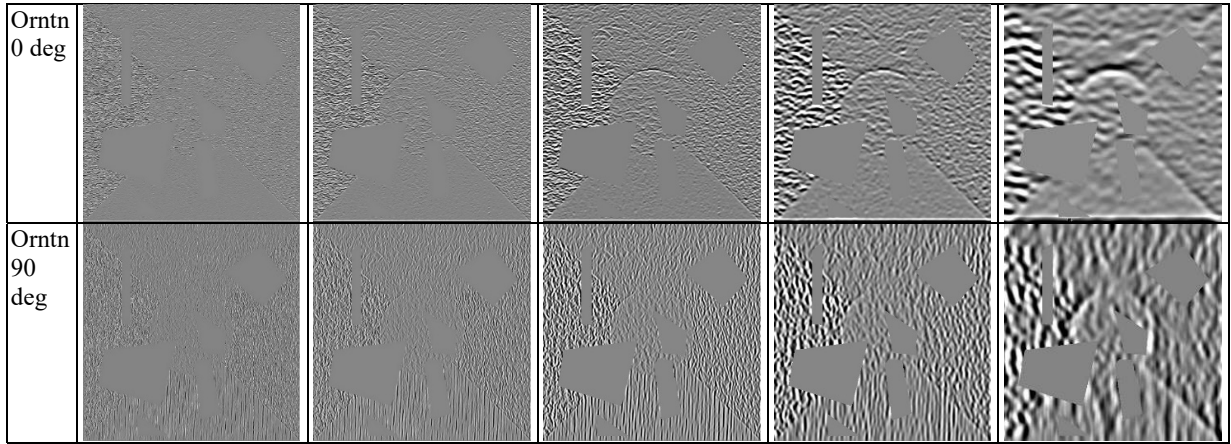


Fig. 3.16-1. Five-scale two-orientation Gabor wavelet-based image near-orthogonal decomposition. Even-symmetric filter output values can be  $< 0$  (local concavity up),  $=0$  (no local concavity, i.e., the local image intensity is a straight line, either horizontal or sloped) or  $> 0$  (local concavity down), also refer to Fig. 3.11-1.

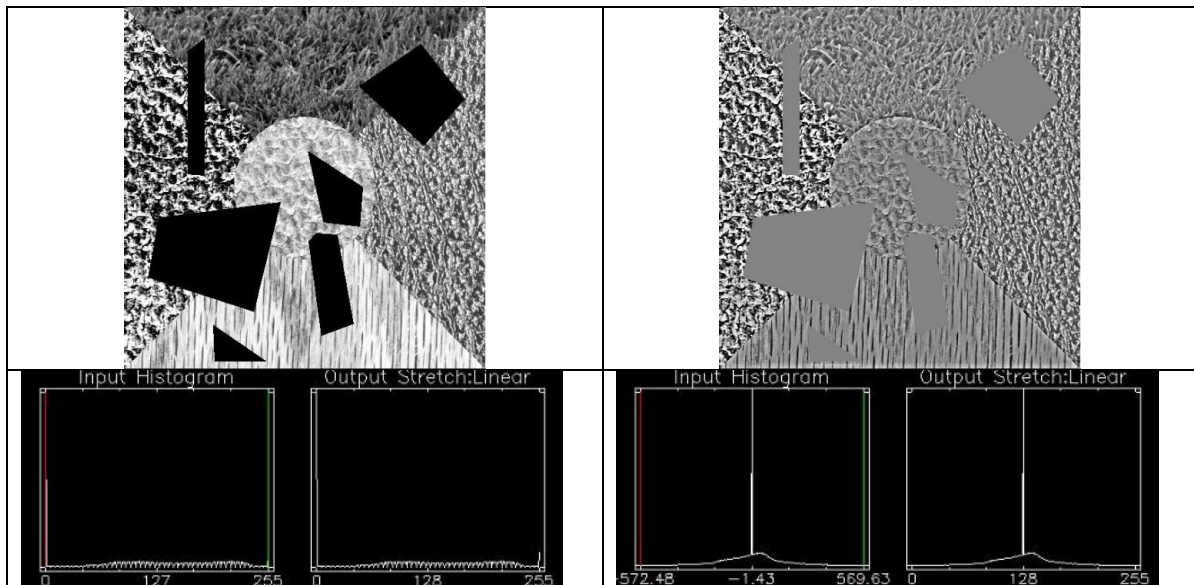
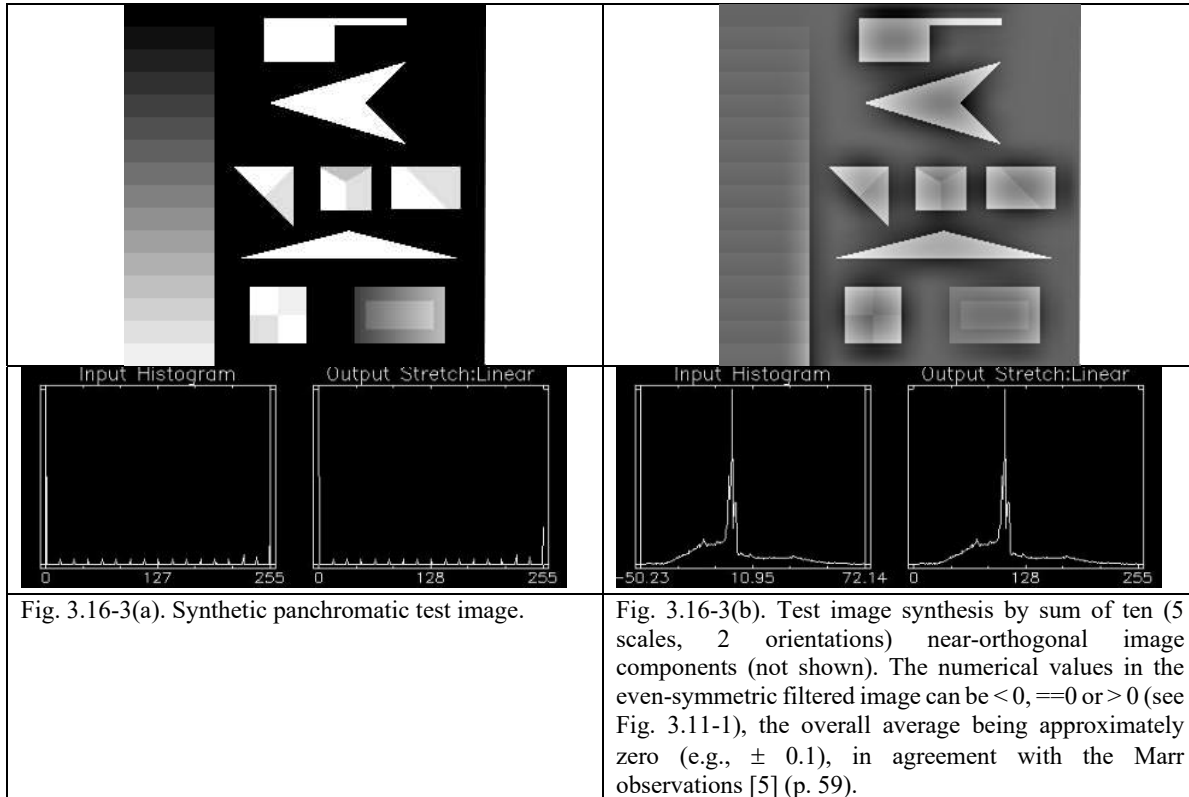


Fig. 3.16-2(a). Masked panchromatic test image, also shown in Fig. 3.8-3(a).

Fig. 3.16-2(b). Test image synthesis (reconstruction) by sum of ten (5 scales, 2 orientations) near-orthogonal image components shown in Fig. 3.16-1. The numerical values in the even-symmetric filtered image can be  $< 0$ ,  $=0$  or  $> 0$  (see Fig. 3.11-1), the overall average being approximately zero (e.g.,  $\pm 0.1$ ), in agreement with the Marr observations [5] (p. 59).





### 3.16.1 Stratified multi-scale multi-orientation near-orthogonal image analysis/decomposition

About the filter bank outputs in the test case, refer to Fig. 3.16-1.

For the sake of simplicity, the DC-component of the 2D signal, to be analyzed by 2D Gaussian filters, is ignored, refer to Chapter 3.4.

### 3.16.2 Stratified multi-scale multi-orientation near-orthogonal image synthesis/reconstruction

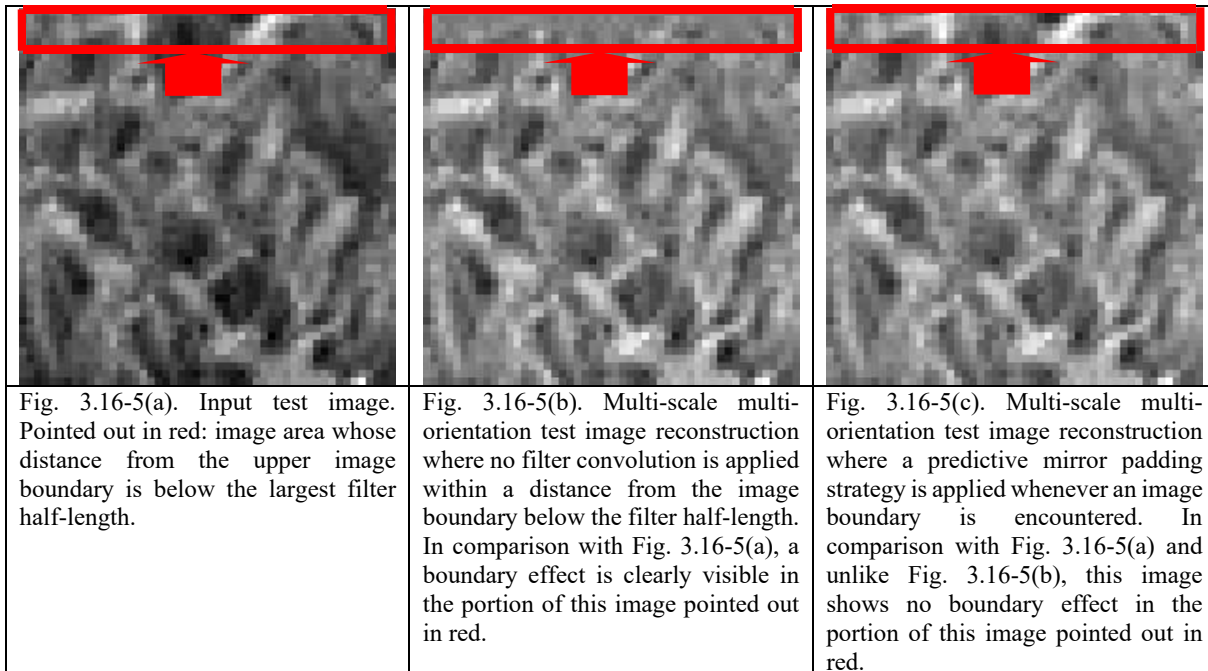
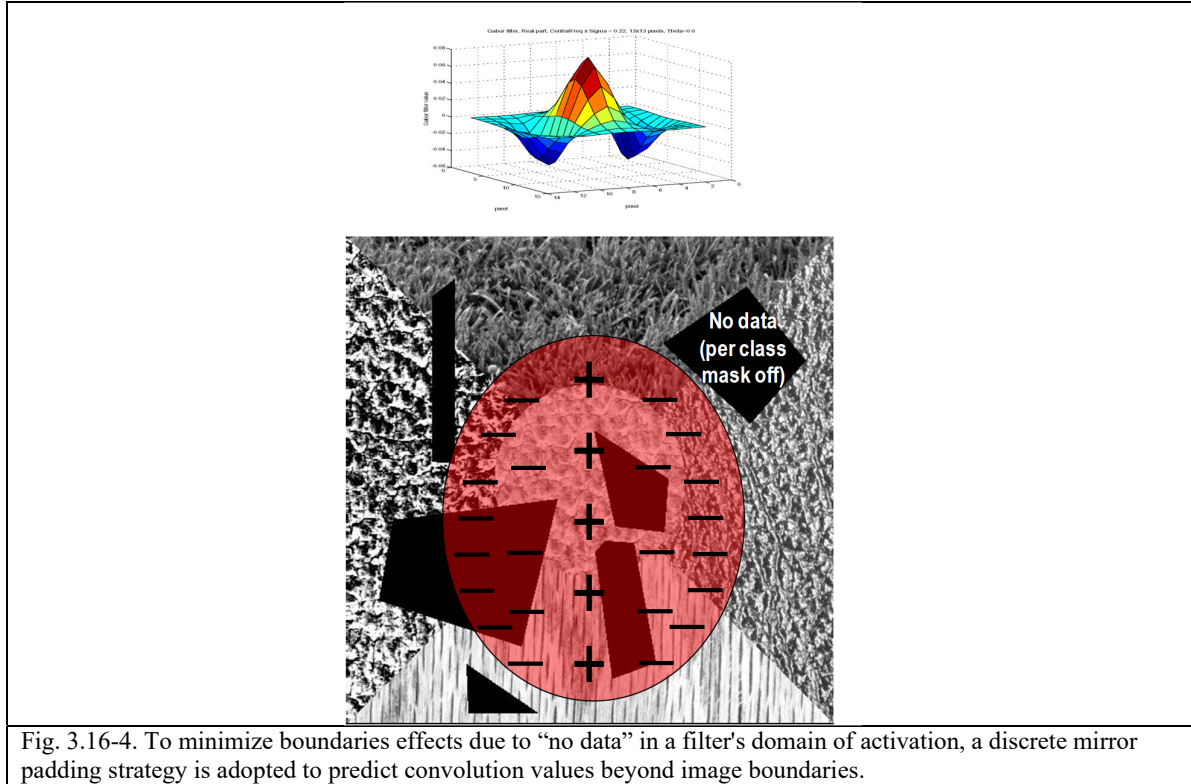
The effectiveness of the proposed filter bank is proved qualitatively by image synthesis when the image near-orthogonal components, shown in Fig. 3.16-1, are summed together. The sum of the near-orthogonal scalar products is shown in Fig. 3.16-2 and Fig. 3.16-3.

For the sake of simplicity, the DC-component of the 2D signal, to be synthesized from 2D Gaussian filter outputs, is ignored, refer to Chapter 3.4.

Qualitatively, Fig. 3.16-2 and Fig. 3.16-3 show that, during the multi-scale near-orthogonal image decomposition, all relevant image features, at both low and high spatial frequencies, are correctly captured by the even-symmetric filter bank, in agreement with 1D simulations proposed in Fig. 3.11-1.

### 3.16.3 Boundary effect removal via predictive mirror padding in the proximity of image/object boundaries

To accomplish the project requirements stated in Chapter 3.2, stratified texture detection must be capable of assessing within-object texture without being affected by artifacts in the proximity of the image/object boundaries, see Fig. 3.16-4.



When a predictive mirror padding strategy is applied whenever an image/object boundary is encountered, a multi-scale multi-orientation test image reconstruction is no longer affected by boundary effects, see Fig. 3.16-5.

To conclude, the stratified multi-scale multi-orientation near-orthogonal image analysis/synthesis, insensitive to stratum/image boundary effects, can be considered successfully accomplished.



### 3.17 Implemented (discrete) quaternary representation of (pos, neg, zero and masked-off) even-symmetric wavelet output values as preliminary ZX segments

According to Marr (refer to Chapter 3.2 and Chapter 3.4.2), stratified *ZX pixels*, namely, *local maxima in the image first-order derivative* (maxima in intensity changes), equivalent to *ZX pixels in the second derivative* [5] (pp. 52, 54, 71-73), must be dealt with through scale according to the *spatial coincident assumption* by Marr [5] (p. 70), refer to Chapter 3.5.3.

In Chapter 3.6, the following original concept is stated: *ZX pixels* of the  $\nabla^2 G$ -filtered image,  $(\nabla^2 G) * I$ , are pixels  $(x, y)$  where a change in sign (from positive to negative, or vice versa) of the concavity (local curvature) of the 2D function  $I(x, y)$  occurs. To the best of this author's knowledge, this definition is novel and original.

In [5] (p. 60), Marr writes: "*ZX pixels* can be represented in various ways. I choose to represent them by a set of oriented primitives called *ZX segments*, each describing a piece of the contour whose intensity slope (first-order derivative of the within-segment intensity, rate at which the convolution changes across the segment) and local orientation are roughly uniform. Because of their eventual physical significance, it is also important to make explicit those places at which the orientation of a *ZX* changes "discontinuously" ... one can construct a practical definition of discontinuity. In addition (to *ZX segments*) small, closed contours are represented as blobs, each also with an associated orientation, average intensity slope and size defined by its extent, along a major and minor axis. Finally, several sizes of operators will be needed to cover the range of scales over which intensity changes occur."

Based on the rather fuzzy (linguistic) definition of *ZX segment* provided by Marr, a natural question would be: does the image primitive called *ZX segment* belong to the list of primitives called by Marr as tokens, which comprise: blob (closed contours, in the terminology of Marr), edge, bar (see [5], caption of Figure 2-21, p. 73) and discontinuity (termination) (see [5], p. 71 and caption of Figure 2-22, p. 74)?

In Chapter 3.12 an original *ZX pixel* definition (selection strategy) was proposed. In this section, the concept of *ZX segment*, introduced by Marr [5] (p. 60), is discussed and investigated.

According to Marr, a *ZX segment* is "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" [5] (p. 60). Hence, it is at the level of *ZX segment* detection that sub-symbolic *ZX contours* turn into sub-symbolic discrete image-objects.

In his work, Marr shows the image's convolution with the  $\nabla^2 G$  operator as a trinary (white/gray/ black) representation, where color white depicts positive output values, black depicts negative values and an intermediate gray depicts zero output values. Hence, the even-symmetric filtered image is partitioned into "black" segments, i.e., connected set of pixels in the (2D) image domain featuring a negative even-symmetric on-center filtered value ( $\partial^2 G / \partial n^2 * I$ ), corresponding to an upward concavity /positive second-order derivative ( $\partial^2 G / \partial n^2 * I$ ), "white" segments, i.e., connected set of pixels in the image domain featuring a positive even-symmetric on-center filtered value ( $\partial^2 G / \partial n^2 * I$ ), corresponding to a downward concavity /negative second-order derivative ( $\partial^2 G / \partial n^2 * I$ ), in addition to non-informative intermediate "gray" segments featuring zero concavity, such that *ZX pixels* are located at the boundary between these "black" and "white" segments with the rest of the world (including "gray" segments) [5] (Figures 2-12, 13, 14, p. 58).

In the present work, the trinary representation of the even-symmetric filtered image adopted by Marr is replaced by a quaternary representation, where positive, zero, negative and mask-off even-symmetric filtered image values are represented as High (250), Intermediate (150), Low (50) and Very Low (5) gray levels, respectively, see Fig. 3.17-1.

The conclusion of potential interest for the transformation of continuous *ZX pixels* into blobs (closed contours, in the terminology of Marr), i.e., image-objects (polygons) hereafter called *ZX segments*, to be generated as output of the raw primal sketch is that, in the aforementioned quaternary representation of the even-symmetric wavelet output values, the following indicators of *ZX segments* can be identified. This is called *preliminary ZX segment map*.

- Downward (negative) concavity (NC) segments = Positive 2<sup>nd</sup>-order derivative segments = Connected sets of pixels where  $(\partial^2 G / \partial n^2 * I) > 0$  = "White" (High = 250) segments (featuring 2ndOrderDerivEvenSymWaveletOutput



values  $> 0$ , equivalent to the bright side of boundaries). *ZX* pixels on the bright side of boundaries lie on the whole perimeter of each “white” segment.

- Zero-concavity (*ZC*) segments = “Gray” (Intermediate = 150) segments (featuring *2ndOrdrDerivEvenSymWaveletOutput* values ( $\partial^2 G / \partial n^2 * I$ ) = 0): these masked-on pixels featuring a zero even-symmetric filter value are never *ZX* pixels, i.e., they are pixels interior to a either “white” or “dark” segment.
- Upward (positive) concavity (*PC*) segments = Negative 2<sup>nd</sup>-order derivative segments = Connected sets of pixels where  $(\partial^2 G / \partial n^2 * I) < 0$  = “Dark” (Low = 50) segments (featuring *2ndOrdrDerivEvenSymWaveletOutput* values  $< 0$ , equivalent to the dark side of boundaries). *ZX* pixels on the dark side of boundaries lie on the whole perimeter of each “dark” segment.
- “Black” (Very Low = 5) segments equivalent to masked-out regions. These pixels are omitted from texture analysis.

Noteworthy, based on Fig. 3.11-1, it is intuitive to understand that *ZX* pixels, each one assigned to one polygon in the *preliminary ZX segment map*, may have to be re-assigned to a neighboring polygon according to the following three criteria.

- (1) Any zero-crossing (*ZX*) pixel should be merged with the neighboring pixel whose *PrcptlCntrst2(x)* == 0 or “low”, if any. This is equivalent to requiring the neighboring pixel to belong to a flat area. Hence, these two pixels belong to the same zero-concavity (*ZC*) segment, either new or pre-existing.
- (2) Any pair of neighboring pixels, either *ZX* or not, should be merged with the neighboring pixel whose *DeltaGrayValue* == 0 or “low”, if any. Hence, these two pixels belong to the same zero-concavity (*ZC*) segment, either new or pre-existing.
- (3) Any pair of neighboring pixels, either *ZX* or not, both featuring *PrcptlCntrst2(x)* == 0 or “low”, should be merged into the same zero-concavity (*ZC*) segment, either new or pre-existing.

Only once the aforementioned *ZX* pixel re-assignment criteria are applied to a *preliminary ZX segment map*, then a *final ZX segment map* is generated as output.

The Weber–Fechner sensitivity law [67] states that a physical entity (e.g., a weight) seems to have to increase by 5% for someone to be able to reliably detect (sense) the increase, and this minimum required fractional increase (of 5/100 of the original weight) is referred to as the “Weber (sensitivity) fraction” for detecting changes in weight. Other discrimination tasks, such as detecting changes in brightness, or in tone height (pure tone frequency), or in the length of a line shown on a screen, may have different Weber sensitivity fractions, but they all obey Weber's law in that observed values need to change by at least some small but constant proportion of the current value to ensure human observers will reliably be able to detect (sense) that change.

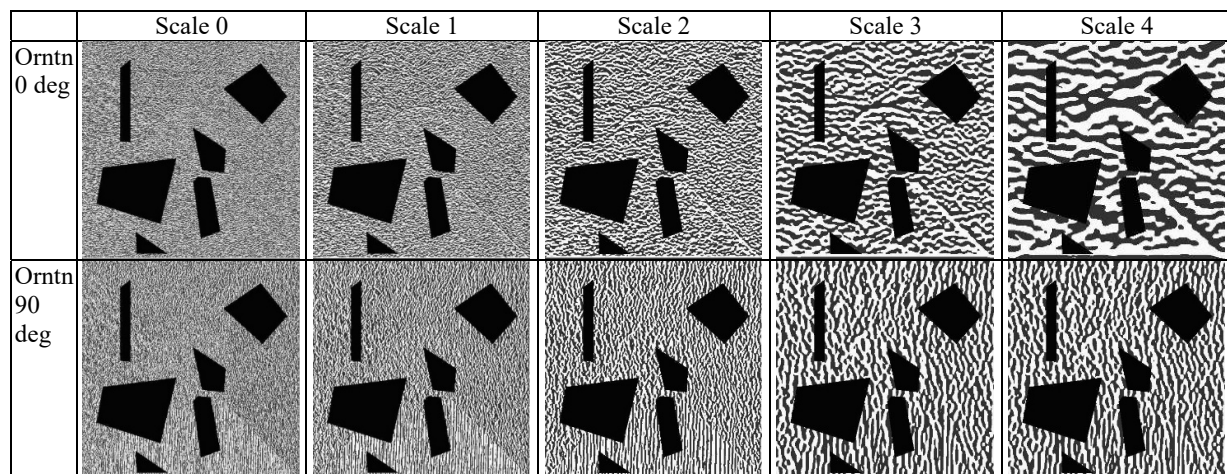


Fig. 3.17-1. Simplified quaternary visualization of an even-symmetric filtered image, inspired to [5].



- Pixels with positive values of the 2ndOrderDerivEvenSymWaveletOutput (which identify the brighter side of physical contours) are shown as High (250).
- Pixels with zero values of the 2ndOrderDerivEvenSymWaveletOutput are shown as Intermediate (150).
- Pixels with negative values of the 2ndOrderDerivEvenSymWaveletOutput (which identify the darker side of physical contours) are shown as Low (50).
- Masked-out pixels are depicted as Very Low (5).

Unfortunately, “white”, “gray” and “dark” connected segments shown in the (discrete) quaternary representation of 2ndOrderDerivEvenSymWaveletOutput values, called *preliminary ZX segment map*, do not coincide with textons/texels, i.e., these white and black segments cross texture boundaries. If adopted, the equivalence between “white”, “gray” and “dark” (connected) segments with textons would lead to undersegmentation problems that should be recovered from at the level of the full primal sketch (perceptual grouping). As an example, consider Fig. 3.17-2 where the same white (black) segment overlaps with two different texture types. In other words, it is not true that “white”, “gray” (ZC segments) and “dark” connected segments in the (discrete) quaternary even-symmetric wavelet output representation, called *preliminary ZX segment map*, identifies textons (texture elements).

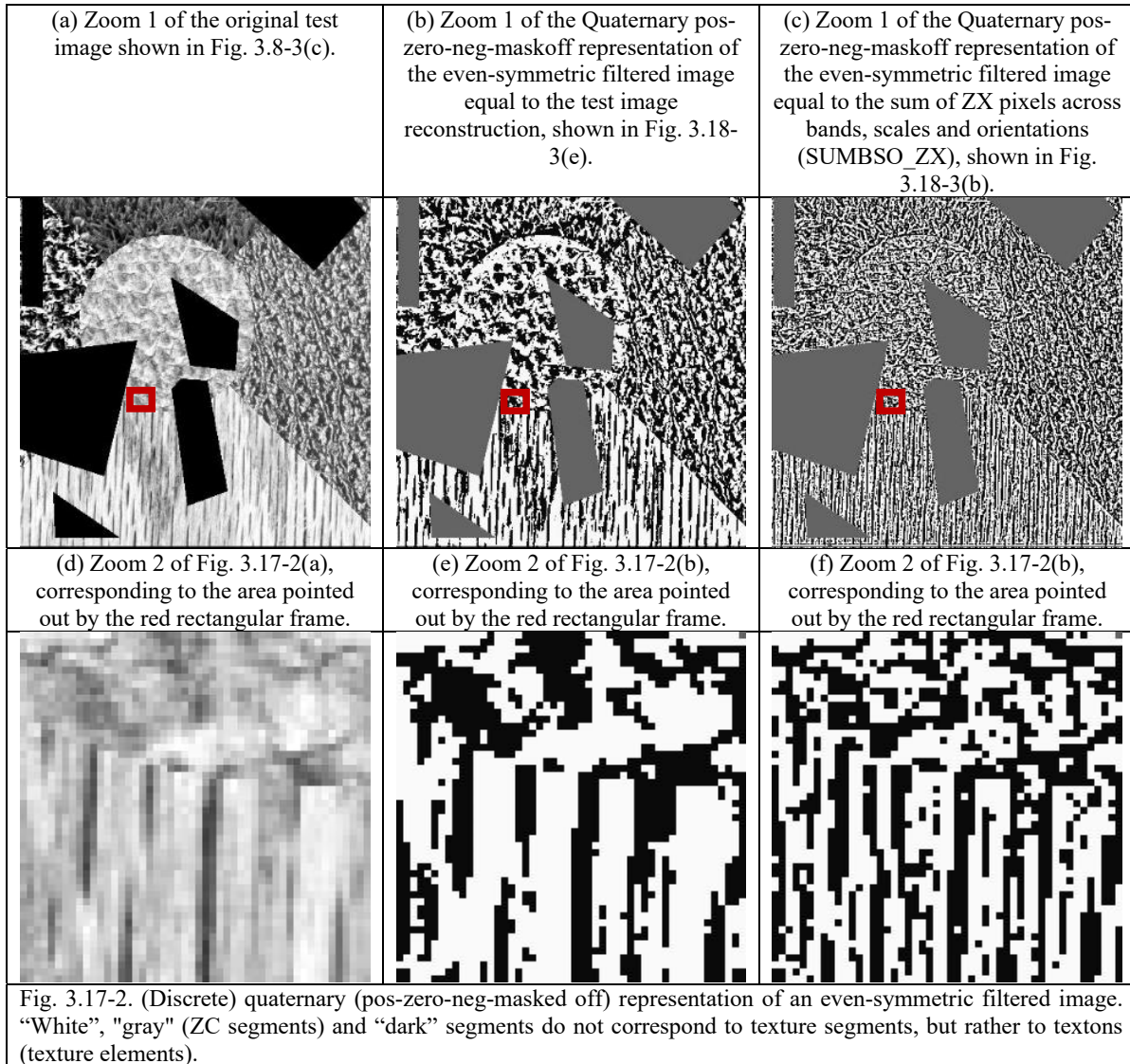
The observation that “white”, “gray” (ZC segments) and “dark” connected segments belonging to the *preliminary ZX segment map* do not coincide with textons is tantamount to saying that “white”, “gray” and “dark” segments do not comply with the Marr definition of a ZX segment as “a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform” [5] (p. 60).

To move from the aforementioned concepts of “white”, “gray” and “dark” segments in the *preliminary ZX segment map* to a *final ZX segment map* that agrees with the Marr’s definition of ZX segments, it is important to consider how a continuous ZX pixel can be assigned to a ZX segment. The following considerations hold.

- Note 2. Ramp edge (a) in Fig. 3.11-1(IV) [1] reveals that the current ZX pixel  $x$  showing a wavelet 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(x) > 0$  can be surrounded by neighboring pixels  $nx$  whose 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(nx) = 0, \forall nx$  in the 8-adjacency neighborhood. In this case pixel  $x$  belongs to the segment of the neighboring pixel  $nx$  such that  $I(x) = I(nx)$ , where  $I(x)$  is the image intensity in pixel  $x$ .
- Note 3. Ramp edge (a) in Fig. 3.11-1(IV) [1] reveals that the current ZX pixel  $x$  showing a wavelet 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(x) < 0$  can be surrounded by neighboring pixels  $nx$  whose 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(nx) = 0, \forall nx$  in the 8-adjacency neighborhood. In this case pixel  $x$  belongs to the segment of the neighboring pixel  $nx$  such that  $I(x) = I(nx)$ , where  $I(x)$  is the image intensity in pixel  $x$ .
- Note 4. Step edge (a) in Fig. 3.11-1(III) [1] reveals that the current ZX pixel  $x$  showing a 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(x) < 0 (> 0)$  can be surrounded by neighboring pixels  $nx$  whose 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(nx) = 0$  OR  $> 0 (< 0), \forall nx$  in the 8-adjacency neighborhood. In this case pixel  $x$  belongs to the segment of the neighboring pixel  $nx$  such that 2ndOrderDerivEvenSymWaveletOutput value  $s_{\text{orn}}(nx) = 0$  AND  $I(x) = I(nx)$ , where  $I(x)$  is the image intensity in pixel  $x$ .
- Note 5. Step edge (a) in Fig. 3.11-1(III) [1], white line == ridge (a) and dark line == ridge (b) in Fig. 3.11-1(VI) [1], reveal that the current ZX pixel  $x$  showing a 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(x) < 0 (> 0)$  never belongs to the same segment of a neighboring pixel  $nx$ , where  $nx$  belongs to the 8-adjacency neighborhood of pixel  $x$ , if 2ndOrderDerivEvenSymWaveletOutput  $s_{\text{orn}}(nx) > 0 (< 0)$ .

To our best knowledge, no Marr’s ZX segment generation from the detected set of ZX pixels, i.e., no *final ZX segment map* generation from the *preliminary ZX segment map* of “white”, “gray” (ZC segments) and “dark” connected segments, has ever been implemented in the existing literature. This accomplishment can be considered challenging, as it may deal with texture boundary detection at the level of the full primal sketch [5] (Figure 2-7, p. 52).





### 3.18 Implemented continuous and (discrete) binary selection/representation of ZX pixels in an even-symmetric filtered image at either single or multiple scales and orientations

Based on the simulation proposed in [1] and extended in Fig. 3.11-1, Chapter 3.6 proposes an original definition of ZX pixels as the locus of points where the  $\nabla^2 G$ -filtered image,  $(\nabla^2 G) * I$ , which is an estimate of the *local curvature*, also called *concavity*, of the 2D function  $I(x,y)$ , passes from positive to non-positive (zero or negative) values (or vice versa). According to Chapter 3.6, the following definition of ZX pixel holds.

If the current pixel  $x$  shows a  $2ndOrdDerivEvenSymWaveletOutput_{s, orn}(x) < 0$  (or  $> 0$ ) and at least one 8-adjacency neighboring pixel  $nx$  shows, at the same scale and orientation of pixel  $x$ ,  $coeff_{s, orn}(nx) \geq 0$  (or  $\leq 0$ ), both pixels  $x$  and  $nx$  are ZX pixels on the dark (or bright) and bright (or dark) side of the contour, respectively. Hence, the following relation always holds true.

Due to the original definition of ZX pixel proposed in Chapter 3.6, set of ZX pixels in an even-symmetric filtered image, whose wavelet output value is  $\neq 0$  (either  $> 0$  or  $< 0$ )

⊆

Set of pixels in the even-symmetric filtered image whose wavelet output value is  $\neq 0$  (either  $> 0$  or  $< 0$ ).

At a given spatial scale  $s$  and orientation  $orntn$ , a set of ZX pixels ( $\nabla^2 G * I(x, y)$ ), capable of satisfying the aforementioned definition, can be selected and represented as a continuous or (discrete, simplified) binary image subset.

Next, ZX pixels can be selected in the multi-scale multi-orientation sum across bands, scales and orientations of scale- and orientation-specific ZX pixels (SUMBSO\_ZX).

### 3.18.1 Continuous selection/representation of ZX pixels in an even-symmetric filtered image

In the output continuous image of ZX pixels generated from a generic ( $\nabla^2 G * I(x, y)$ ) (e.g., generated at a given scale  $s$  and orientation  $orntn$ ), ZX pixels feature a value either  $< 0$  or  $> 0$  and equal to their wavelet output value, such that:

1. Pixel value  $> 0$  (contour bright side) if this is a mask-on pixel, its value is  $> 0$  and at least one of its neighbors is  $\leq 0$ , in compliance with the original definition proposed in Chapter 3.6.
2. Pixel value  $< 0$  (contour dark side) if this is a mask-on pixel, its value is  $< 0$  and at least one of its neighbors is  $\geq 0$ , in compliance with the original definition proposed in Chapter 3.6.
3. Pixel value  $= 0$  otherwise, including mask-off pixels.

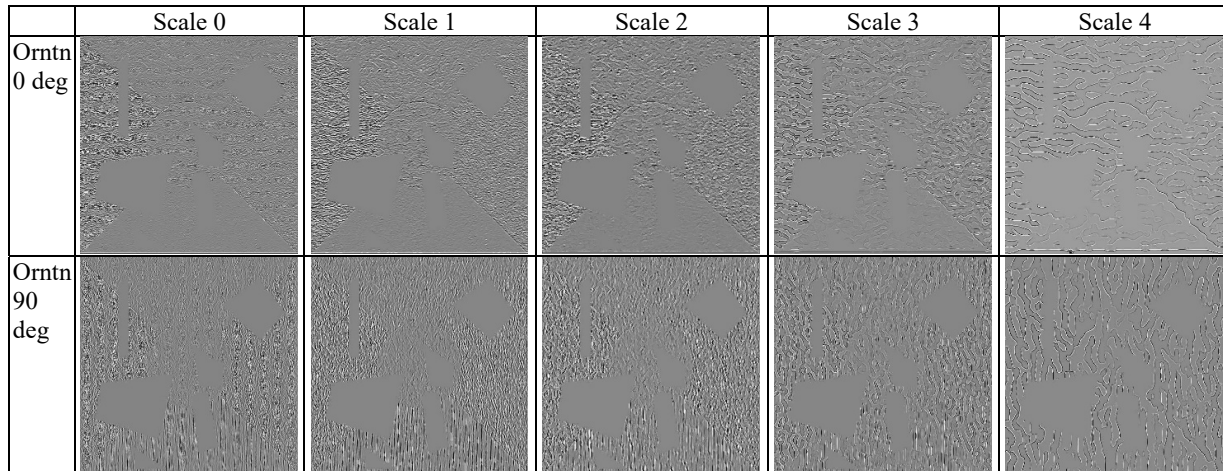


Fig. 3.18-1. Continuous selection/representation of ZX pixels in an even-symmetric filtered image, like the near-orthogonal image components shown in Fig. 3.17-1. Any selected ZX pixel, which satisfies the original definition proposed in Chapter 3.12, features a 2ndOrderDerivEvenSymWaveletOutput value  $> 0$  (which identify the brighter side of physical contours) or  $< 0$  (which identify the darker side of physical contours). Non ZX pixels OR masked-out pixels feature a zero value (depicted as an "intermediate" gray value).

### 3.18.2 (Discrete) binary selection/representation of ZX pixels in an even-symmetric filtered image

In the output (discrete) binary image of ZX pixels generated from a generic ( $\nabla^2 G * I(x, y)$ ) (e.g., at a given scale  $s$  and orientation  $orntn$ ), ZX pixels feature a value either  $< 0$  or  $> 0$  and equal to their wavelet output value, such that:

1. TRUE if pixel value  $\neq 0$  in the output continuous image of ZX pixels (refer to Chapter 3.10.1.1).
2. FALSE otherwise.



### 3.18.3 Selection/representation of ZX pixels in a multi-scale multi-orientation even-symmetric filtered image: ZX pixels of a ZX sum across bands, scales and orientations

The sums of the continuous ZX output images collected across bands, scales and orientations are shown in Fig. 3.18-2 and Fig. 3.18-3. Based on these results, the conclusion is the following.

The multi-scale multi-orientation sum (SUMBSO\_ZX) of the continuous ZX pixels detected in each scale- and orientation-specific near-orthogonal image component appears able to preserve (irrespective of a constant dc component) the original small but genuine image details better than the continuous ZX pixels selected in the near-orthogonal image reconstruction.

This result is in line with the Marr quote [5] (p.67): "ZXs provide a natural way of moving from an analogue or continuous representation like the two-dimensional image intensity values to a discrete representation (into discrete tokens, namely, blobs, edges, bars and termination points through so-called ZX segments [5], p. 60). A fascinating thing about this transformation is that it probably incurs no loss of information... (in practice), a one-octave band-pass signal can be completely reconstructed (up to an overall multiplicative constant) from its ZXs".

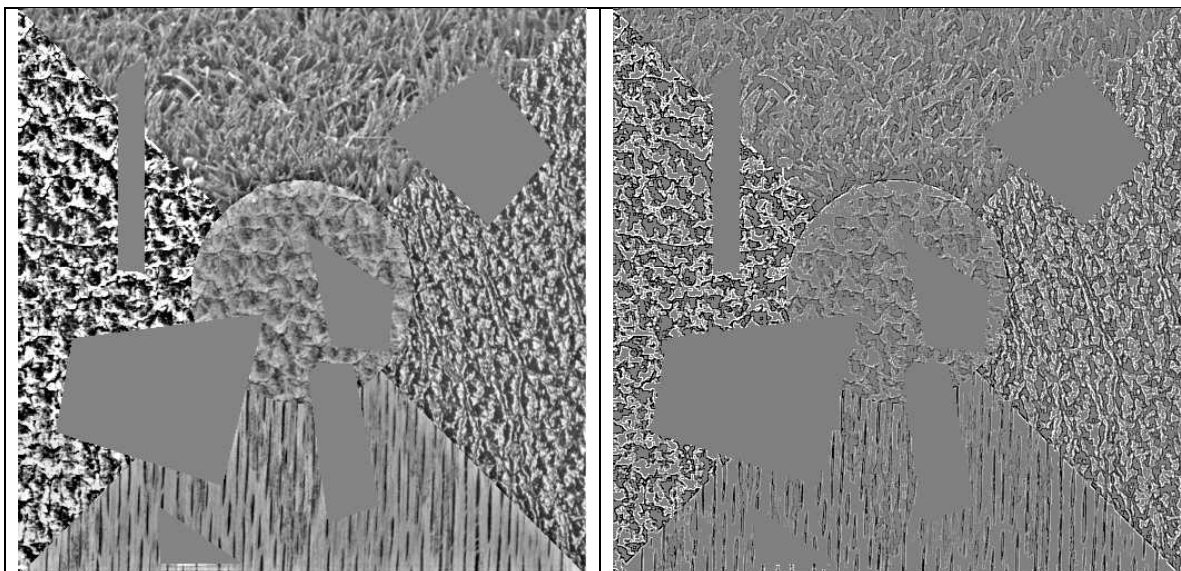


Fig. 3.18-2(a). Wavelet-based input image reconstruction: Sum of wavelet-based near-orthogonal image components, shown in Fig. 3.16-2. Identified as sum across bands, scales and orientations of the even-symmetric filters, SUMBSO\_EF. Basic Stats: Min = -62.4806274, Max = -49.4454773, Mean = -0.060112, Stdev = 13.1206412.

Fig. 3.18-2(b). ZX pixels selected in the wavelet-based input image reconstruction shown in Fig. 3.18-2(a). Identified as ZX(SUMBSO\_EF). Basic Stats: Min = -61.087627, Max = 46.051975, Mean = -0.037284, Stdev = 8.966838.



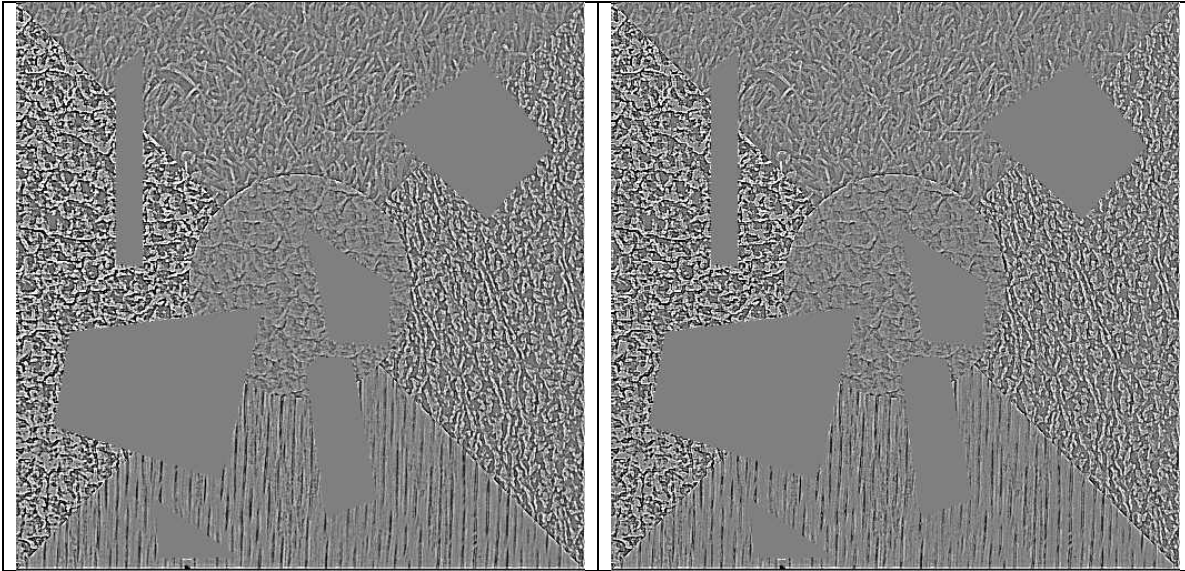


Fig. 3.18-2(c). Sum across bands, scales and orientations of ZX pixels (SUMBSO\_ZX) selected in each scale- and orientation-specific wavelet-based near-orthogonal image component. This image shows more genuine but small details of the original image, depicted in Fig. 3.16-2, than Fig. 3.18-2(b). Basic Stats: Min = -47.2526520, Max = 47.5248535, Mean = 0.02070, Stdev = 7.497319.

Fig. 3.18-2(d). ZX pixels selected in the SUMBSO\_ZX continuous image shown in Fig. 3.18-2(c), identified as ZX(SUMBSO\_ZX). This ZX(SUMBSO\_ZX) image can look similar to Fig. 3.18-2(c), but it is not the same as Fig. 3.18-2(c): internal pixels in “white” or “dark” ZX segments are set to zero. Basic Stats: Min = -47.2526520, Max = 47.5248535, Mean = 0.016498, Stdev = 7.345815. As expected, in practice, Fig. 3.18-2(a) and Fig. 3.18-2(c) are almost the same image. Fig. 3.18-2(d), ZX(SUMBSO\_ZX), looks “better” (more detailed) than its simpler counterpart shown in Fig. 3.18-2(b), ZX(SUMBSO\_EF).

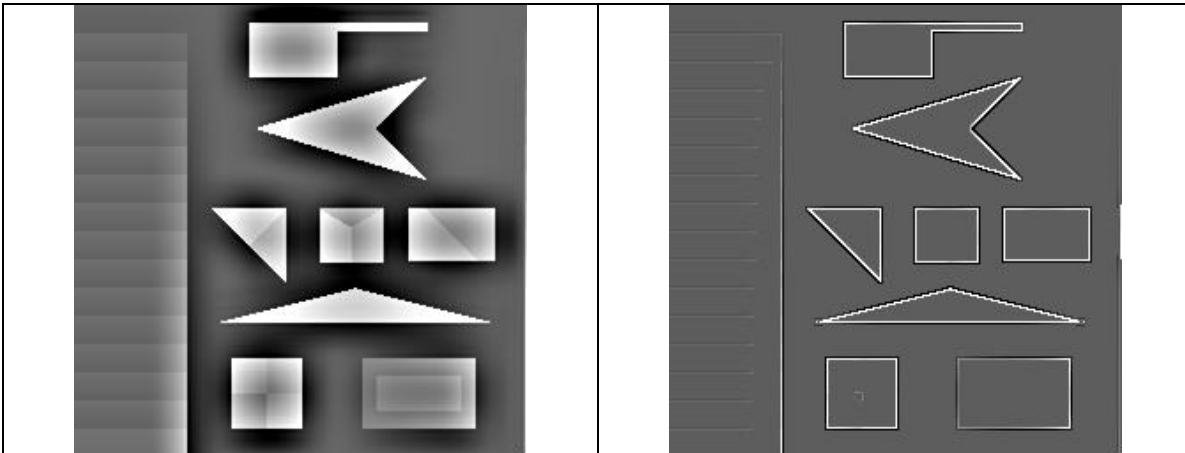


Fig. 3.18-3(a). Wavelet-based input image reconstruction: Sum of wavelet-based near-orthogonal image components, shown in Fig. 3.16-3. Identified as (SUMBSO\_EF). Basic Stats: Min = -50.230831, Max = 72.136589, Mean = -0.280402, Stdev = 18.077869.

Fig. 3.18-3(b). ZX pixels selected in the wavelet-based input image reconstruction shown in Fig. 3.18-3(a). Identified as ZX(SUMBSO\_EF). Basic Stats: Min = -50.230831, Max = 68.869339, Mean = 0.450975, Stdev = 7.841392.

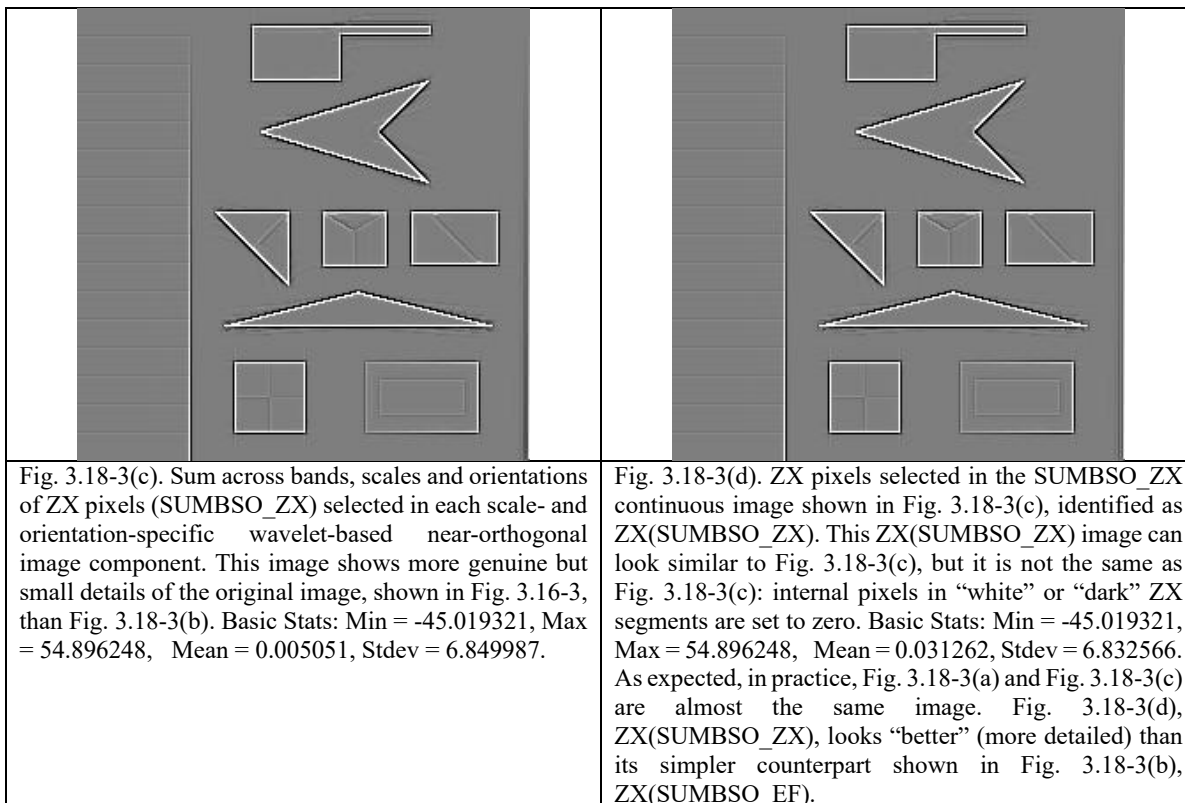


Fig. 3.18-3(c). Sum across bands, scales and orientations of ZX pixels (SUMBSO\_ZX) selected in each scale- and orientation-specific wavelet-based near-orthogonal image component. This image shows more genuine but small details of the original image, shown in Fig. 3.16-3, than Fig. 3.18-3(b). Basic Stats: Min = -45.019321, Max = 54.896248, Mean = 0.005051, Stdev = 6.849987.

Fig. 3.18-3(d). ZX pixels selected in the SUMBSO\_ZX continuous image shown in Fig. 3.18-3(c), identified as ZX(SUMBSO\_ZX). This ZX(SUMBSO\_ZX) image can look similar to Fig. 3.18-3(c), but it is not the same as Fig. 3.18-3(c): internal pixels in “white” or “dark” ZX segments are set to zero. Basic Stats: Min = -45.019321, Max = 54.896248, Mean = 0.031262, Stdev = 6.832566. As expected, in practice, Fig. 3.18-3(a) and Fig. 3.18-3(c) are almost the same image. Fig. 3.18-3(d), ZX(SUMBSO\_ZX), looks “better” (more detailed) than its simpler counterpart shown in Fig. 3.18-3(b), ZX(SUMBSO\_EF).

Based on Fig. 3.18-2, Fig. 3.18-3 and other experiments (also refer to the next sections), the original conclusion of this subsection is that the detection of continuous ZX pixels in the sum of continuous ZX pixels across bands, scales and orientations (SUMBSO\_ZX), identified as ZX(SUMBSO\_ZX), is considered superior to the detection of continuous ZX pixels in the near orthogonal image reconstruction, identified as ZX(SUMBSO\_EF), where acronym SUMBSO\_EF means: sum across bands, scales and orientations of the even-symmetric filter output values.

This difference in spatial details is made somewhat more intuitive in Chapter 3.18.4, where a simplified continuous and (discrete) ternary contour representation of continuous ZX pixels is discussed.

To the best of this author’s knowledge, based on survey papers like [24] and [25], the proposed multi-scale multi-orientation combinations of even-symmetric Gabor filters for image-contour detection is completely novel. On the one hand, it can be considered a possible implementation of the multi-scale filter combination strategy verbosely sketched by Marr [5] (p.67). On the other hand, it is completely different from the multi-scale multi-orientation filter combination strategies proposed for contour detection by Burr and Morrone’s [26], Canny’s [27], Rodrigues and du Buf’s [28], [43], and many others, e.g., refer to [25].

### 3.18.4 Continuous and (discrete) ternary contour representation of continuous ZX pixels detected in an even-symmetric filtered image

ZX pixels, where the sign of the local concavity changes (refer to Chapter 3.6), are located at the boundary of “white”, “gray” or “dark” ZX segments (refer to Chapter 3.9). Hence, ZX pixels represent both sides (bright and dark sides) of image contours. In other words, image contours have two sides, the dark and the bright one. For contour representation purposes, when one side is depicted, e.g., the bright side to comply with human perception, the other side, which is adjacent, e.g., the dark side, can be omitted.

Original simplified contour-like (discrete) ternary and continuous representations of continuous ZX pixels detected in an even-symmetric filtered image are generated as follows.





1. If pixel  $x$  features  $2ndOrdDerivEvenSymWaveletOutput(x) > 0$  (potential bright side of the boundary) and at least one of its 8-adjacency neighbors  $nx$  features  $2ndOrdDerivEvenSymWaveletOutput(nx) < 0$  (dark side of the boundary) OR  $2ndOrdDerivEvenSymWaveletOutput(nx) == 0$ , then pixel  $x$  is a ZX pixel belonging to the bright side of an image contour. Refer to: (i) step edge (a) in Fig. 3.11-1(III) [1], (ii) ramp edge (a) in Fig. 3.11-1(IV) (see Fig. 3.21-2 in [1]) and (iii) white line == ridge (a) and dark line == ridge (b) in Fig. 3.11-1(VI) [1]. In practice, this ZX pixel is selected for visualization purposes of the bright side of image contours. In this case:
  - ✓ The output continuous value of the contour pixel is set equal to the  $2ndOrdDerivEvenSymWaveletOutput(x) > 0$ .
  - ✓ The output trinary value of the contour pixel is set equal to HIGH\_VALUE\_UCHAR.
2. If a ZX pixel  $x$  features  $2ndOrdDerivEvenSymWaveletOutput(x) < 0$  (dark side of the boundary) and among its 8-adjacency neighbors  $nx$  there is at least one  $2ndOrdDerivEvenSymWaveletOutput(nx) == 0$  (bright side of the boundary), but not a single  $2ndOrdDerivEvenSymWaveletOutput(nx) > 0$ , i.e., there is no ZX neighboring pixel  $nx$  to be selected as belonging to the bright side of an image contour, then the ZX pixel  $x$  is marked as a contour pixel. Refer to: (i) ramp edge (a) in Fig. 3.11-1(IV) [1]. In addition to the bright side of boundaries, this is the sole case a dark side of boundaries must be selected for contour visualization purposes. In this case:
  - ✓ The output continuous value of the contour pixel is set equal to the  $abs(2ndOrdDerivEvenSymWaveletOutput(x) < 0)$ , which is a positive value.
  - ✓ The output trinary value of the contour pixel is set equal to HIGH\_VALUE\_UCHAR.
3. Masked-out pixel, depicted in gray. In this case:
  - The output continuous value of the contour pixel remains equal to 0.
  - The output trinary value of the contour pixel is set equal to VERY\_LOW\_VALUE\_UCHAR.
4. Not case 1 and not case 2, depicted in black. In this case:
  - The output continuous value of the contour pixel is set equal to 0.
  - The output trinary value of the simplified ZX pixel is set equal to LOW\_VALUE\_UCHAR.

Examples of the continuous and (discrete) trinary contour representation of continuous ZX pixels are shown in Fig. 3.18-4 and Fig. 3.18-5.

For a more intuitive understanding of the accomplished results, a 5-scale 2-orientation even-symmetric Gabor filter-based image analysis/synthesis is applied to a test set of natural images, see Fig. 3.18-6, Fig. 3.18-7 and Fig. 3.18-8.

Farther insight about the proposed ZX detection (selection) and representation strategies can stem from the analysis of the two synthetic test images featuring ramp, roof and step edges shown in Fig. 3.16-3 and Fig. 3.18-9.

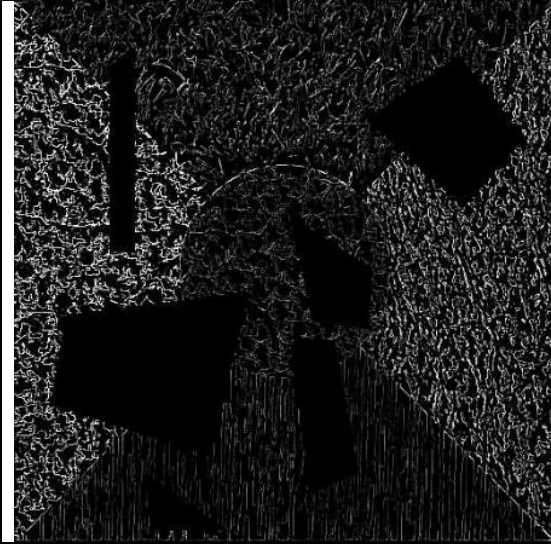


Fig. 3.18-4(a). Continuous contour representation of continuous ZX pixels selected from the near-orthogonal image reconstruction shown in Fig. 3.18-3(d).

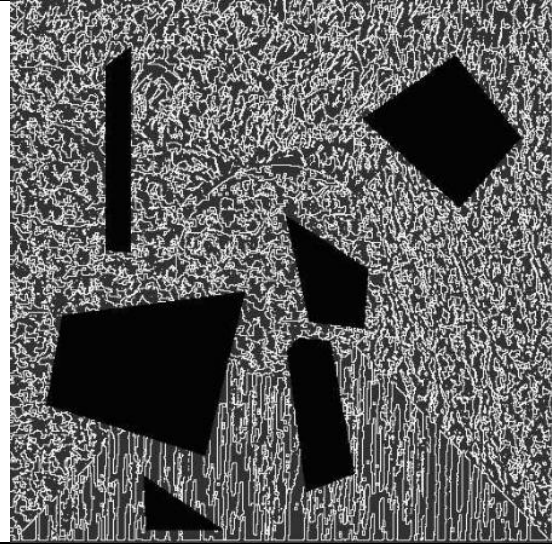


Fig. 3.18-4(b). (Discrete) trinary contour representation of continuous contours shown in Fig. 3.18-4(a).

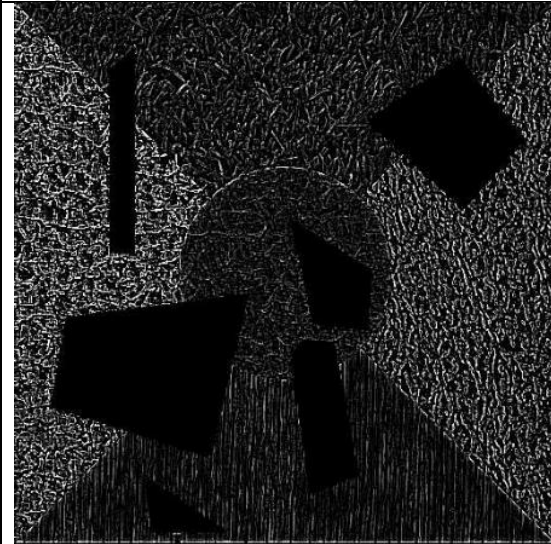


Fig. 3.18-4(c). Continuous contour representation of continuous ZX pixels selected from the continuous ZX sum across bands, scales and orientations shown in Fig. 3.18-3(a). As expected, Fig. 3.18-4(c) shows more genuine but small image details than Fig. 3.18-4(a).

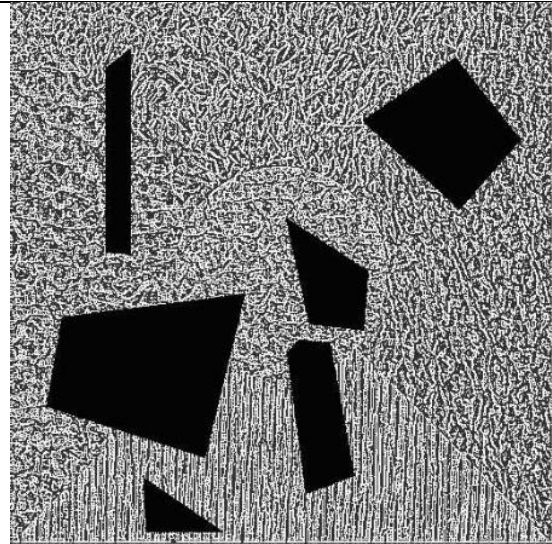


Fig. 3.18-4(d). (Discrete) trinary contour representation of continuous contours shown in Fig. 3.18-4(c).



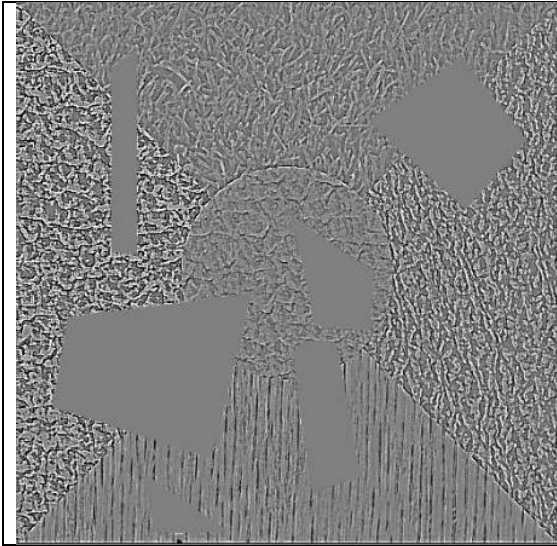


Fig. 3.18-5(a). Enhanced (filtered from scale-0 upward) continuous ZX pixels selected from the continuous ZX sum across bands, scales and orientations shown in Fig. 3.16-2(b). Basic Stats: Min = -47. 252651, Max = 47.524853, Mean = 0.015746, Stdev = 7. 345312.

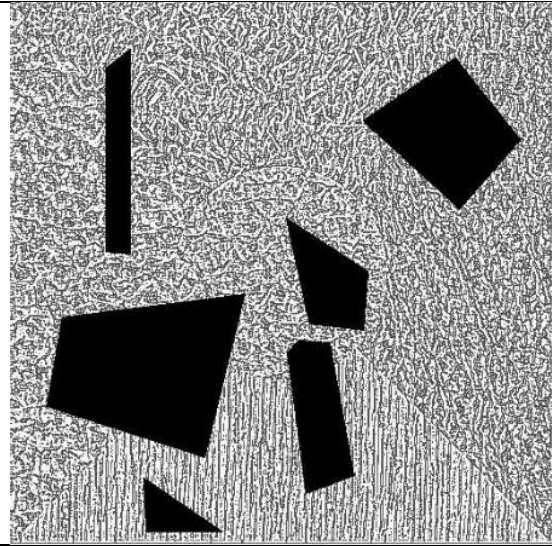


Fig. 3.18-5(b). (Discrete) quaternary representation of the continuous enhanced ZX pixels shown in Fig. 3.18-5(a).

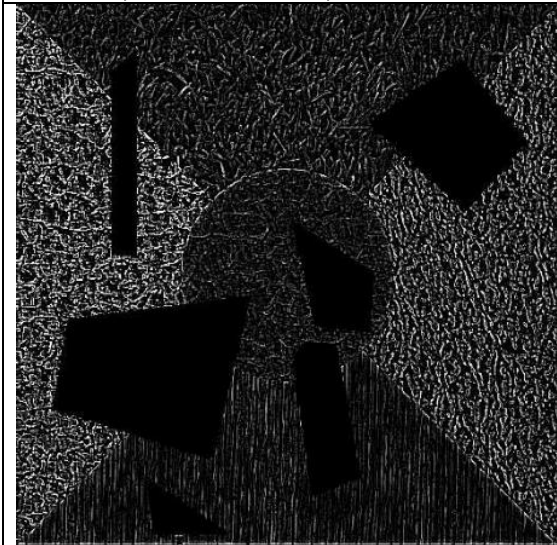


Fig. 3.18-5(c). Continuous contour representation of the continuous enhanced ZX pixels shown in Fig. 3.18-5(a).

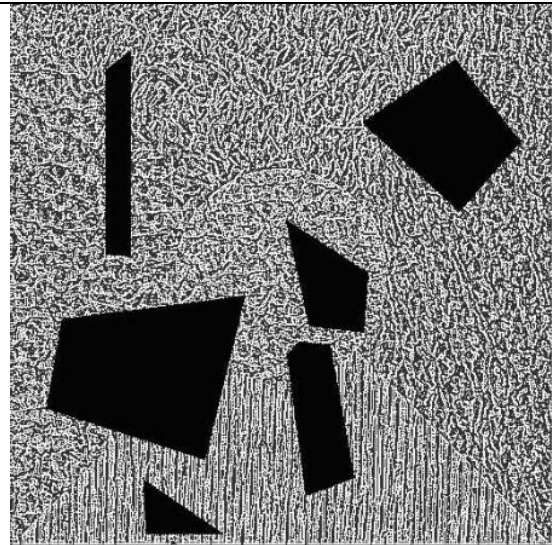
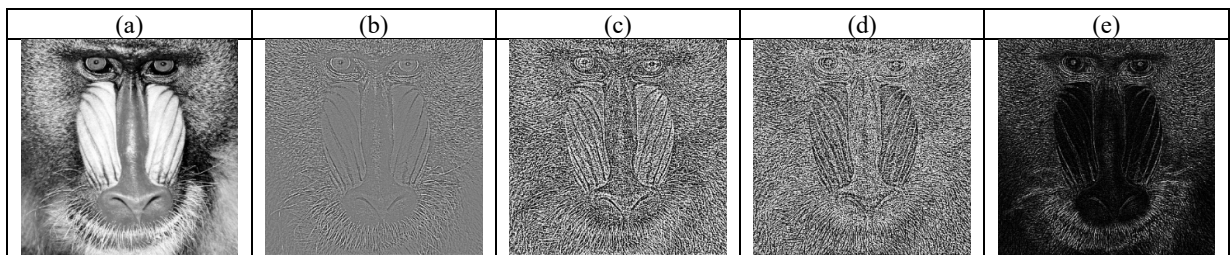


Fig. 3.18-5(d) (Discrete) trinary contour representation of the continuous enhanced ZX pixels shown in Fig. 3.18-5(a).





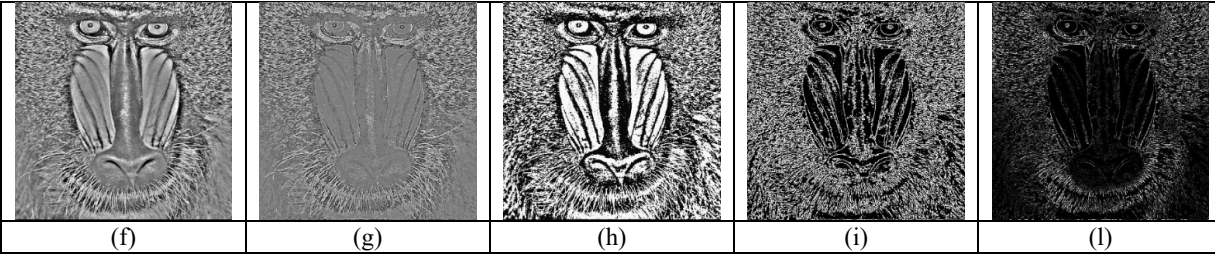


Fig. 3.18-6(a) to Fig. 3.18-6(l).

- (a) Original panchromatic image. Image size:  $512 \times 512$  pixels.
- (b) Sum across bands, scales and orientations (SUMBSO\_ZX) of continuous ZX pixels detected per scale and orientation in the even-symmetric image decomposition.
- (c) Quaternary pos-zero-neg-masked off representation of the SUMBSO\_ZX continuous image depicted in Fig. 3.18-6(b).
- (d) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-6(b).
- (e) Continuous contour representation of the ZX(SUMBSO\_ZX) continuous image generated from the SUMBSO\_ZX image depicted in Fig. 3.18-6(b). This ZX(SUMBSO\_ZX) image is practically indistinguishable from the continuous contour representation of enhanced continuous ZX pixels selected according to Chapter 3.10.5.
- (f) Test image reconstruction from an even-symmetric image decomposition, 5 scales, 2 orientations.
- (g) Continuous ZX pixels of the image reconstruction shown in Fig. 3.18-6(f).
- (h) Quaternary pos-zero-neg-maskoff representation of the continuous ZX pixels depicted in Fig. 3.18-6(g).
- (i) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-6(g).
- (l) Continuous contour representation of continuous ZX pixels depicted in Fig. 3.18-6(g).
- Conclusion of the comparison of Fig. 3.18-6(a) to Fig. 3.18-6(l): in line with previous results shown in Chapter 3.10.3, Fig. 3.18-6(e) looks perceptually better (i.e., it features less false negative contours and less false positive contours) than Fig. 3.18-6(l).

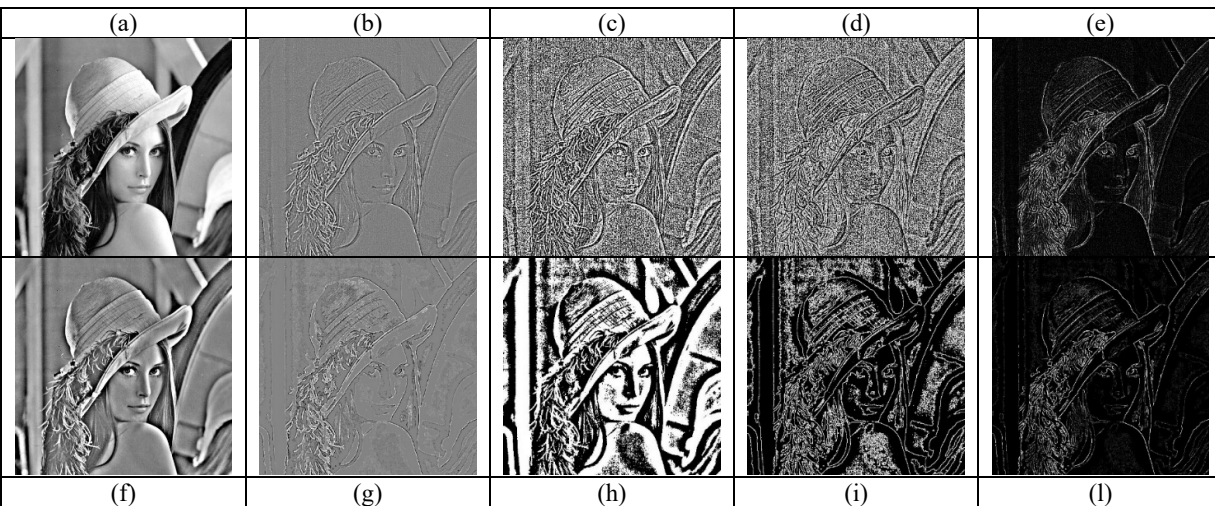


Fig. 3.18-7(a) to Fig. 3.18-7(l).

- (a) Original panchromatic image. Image size:  $512 \times 512$  pixels.
- (b) Sum across bands, scales and orientations of continuous ZX pixels detected per scale and orientation in the even-symmetric image decomposition, identified as the SUMBSO\_ZX continuous image.
- (c) Quaternary pos-zero-neg-masked off representation of the SUMBSO\_ZX continuous image depicted in Fig. 3.18-7(b).
- (d) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-7(b).
- (e) Continuous contour representation of the ZX(SUMBSO\_ZX) continuous image generated from the SUMBSO\_ZX image depicted in Fig. 3.18-7(b). This ZX(SUMBSO\_ZX) image is practically indistinguishable from the continuous contour representation of enhanced continuous ZX pixels selected according to Chapter 3.10.5.
- (f) Test image reconstruction from an even-symmetric image decomposition, 5 scales, 2 orientations.
- (g) Continuous ZX pixels of the image reconstruction shown in Fig. 3.18-7(f).
- (h) Quaternary pos-zero-neg-maskoff representation of the continuous ZX pixels depicted in Fig. 3.18-7(g).

- (i) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-7(g).
- (l) Continuous contour representation of continuous ZX pixels depicted in Fig. 3.18-7(g).
- Conclusion of the comparison of Fig. 3.18-7(a) to Fig. 3.18-7(l): in line with previous results shown in Chapter 3.11.3, Fig. 3.18-7(e) looks perceptually better (i.e., it features less false negative contours and less false positive contours) than Fig. 3.18-7(l).

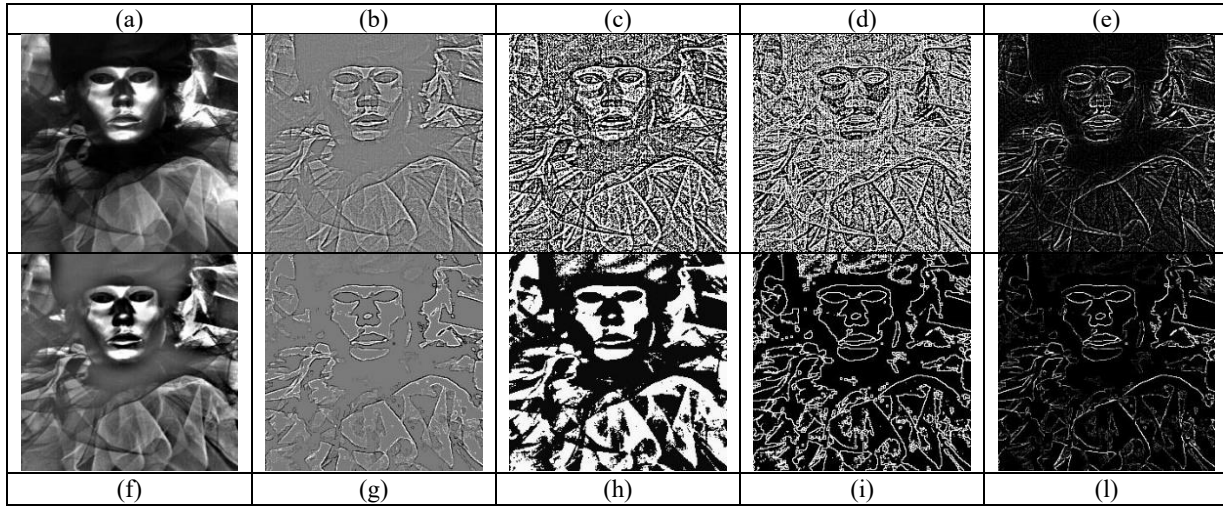


Fig. 3.18-8(a) to Fig. 3.18-8(l).

(a) Original panchromatic image. Image size:  $256 \times 256$  pixels.

(b) Sum across bands, scales and orientations of continuous ZX pixels detected per scale and orientation in the even-symmetric image decomposition, identified as the SUMBSO\_ZX continuous image.

(c) Quaternary pos-zero-neg-masked off representation of the SUMBSO\_ZX continuous image depicted in Fig. 3.18-8(b).

(d) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-8(b).

(e) Continuous contour representation of the ZX(SUMBSO\_ZX) continuous image generated from the SUMBSO\_ZX image depicted in Fig. 3.18-8(b). This ZX(SUMBSO\_ZX) image is practically indistinguishable from the continuous contour representation of enhanced continuous ZX pixels selected according to Chapter 3.11.5.

(f) Test image reconstruction from an even-symmetric image decomposition, 5 scales, 2 orientations.

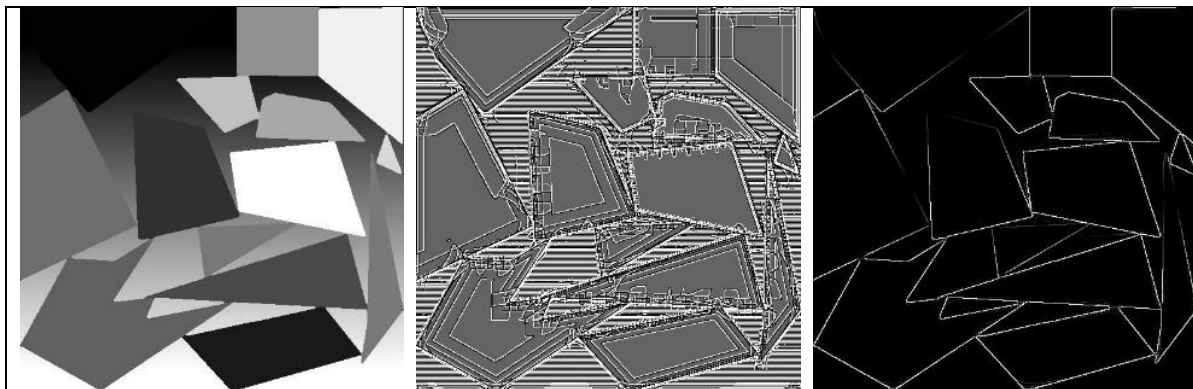
(g) Continuous ZX pixels of the image reconstruction shown in Fig. 3.18-8(f).

(h) Quaternary pos-zero-neg-maskoff representation of the continuous ZX pixels depicted in Fig. 3.18-8(g).

(i) (Discrete) trinary contour representation of continuous ZX pixels depicted in Fig. 3.18-8(g).

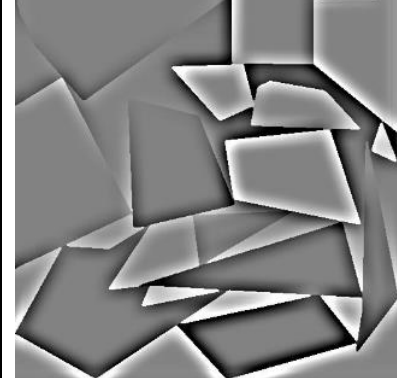
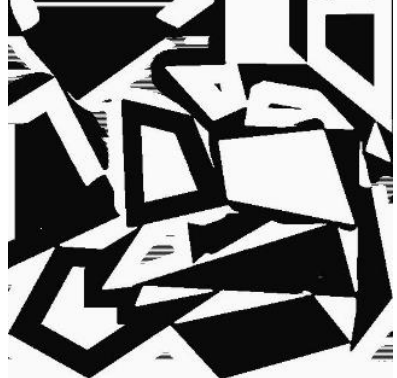
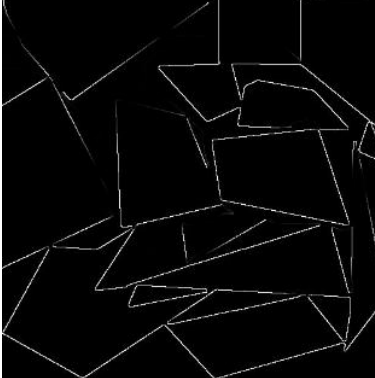
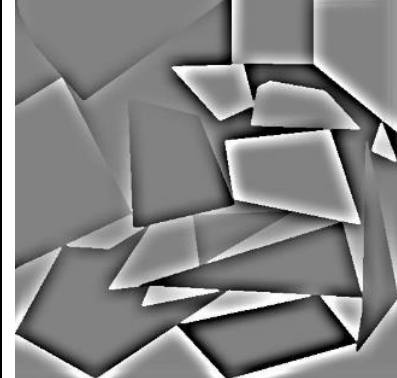
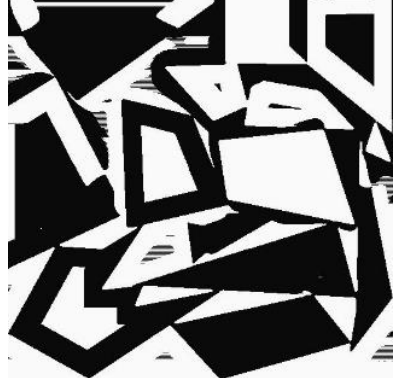
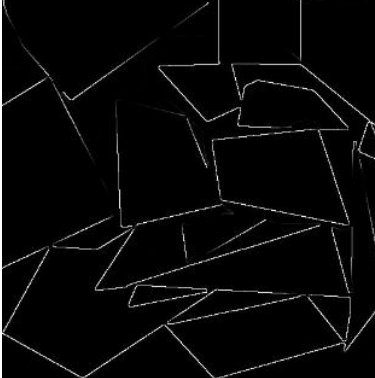
(l) Continuous contour representation of continuous ZX pixels depicted in Fig. 3.18-8(g).

- Conclusion of the comparison of Fig. 3.18-8(a) to Fig. 3.18-8(l): in line with previous results shown in Chapter 3.11.3, Fig. 3.18-8(e) looks perceptually better (i.e., it features less false negative contours and less false positive contours) than Fig. 3.18-8(l).







<p>Fig. 3.18-9(a). Test synthetic image featuring ramp and step edges.</p>	<p>Fig. 3.18-9(b). (Discrete) quaternary (pos-zero-neg-masked off) representation of the even-symmetric wavelet output values, generated from the SUMBSO_ZX continuous image (not shown).</p>	<p>Fig. 3.18-9(c). Continuous contour representation of the ZX(SUMBSO_ZX) continuous image (not shown). None of the "true edges" appears as missing. False positive edges, visible in Fig. 3.18-9(b), become "perceptually" invisible.</p>
		
<p>Fig. 3.18-9(d). Test image reconstruction, 5 scales, 2 orientations. Reconstructed values are <math>&lt; 0</math> or <math>&gt; 0</math> or equal to zero.</p>	<p>Fig. 3.18-9(e). (Discrete) quaternary (pos-zero-neg-masked off) representation of the even-symmetric wavelet output values, generated from the test image reconstruction shown in Fig. 3.18-9(d).</p>	<p>Fig. 3.18-9(f). Continuous contour representation of continuous ZX pixels detected in the test image reconstruction shown in Fig. 3.18-9(d). Some "true edges" are missing. False positive edges, visible in Fig. 3.18-9(e), become "perceptually" invisible.</p>
		

### 3.18.5 Enhanced (filtered from scale-0 upward) continuous ZX pixels detected in an even-symmetric filtered image

Fig. 3.18-10(d) shows that many (weak but) false image contours are visible if continuous ZX pixels are selected from the continuous ZX sum across bands, scales and orientations of an even-symmetric image decomposition.

These false positive contour pixels are removed as follows.

If at the finest spatial scale (scale 0), there is no firing activity in the EvnSymtrc filters irrespective of their orientation, i.e., if the sum of the output values of the EvnSymtrc filters at a given location  $x$  at scale 0 is zero, then the input image at location  $x$  is either constant or ramp-like, i.e., there is no change of local curvature in pixel  $x$ . If at the same spatial location  $x$  there is a ZX pixel detection at any coarser scale, this detection must be considered an artifact.

In Fig. 3.18-11 this so-called enhanced ZX pixel selection strategy in an even-symmetric filtered image is proved to be effective.

The proposed enhanced ZX pixel selection, where possible ZX pixels are filtered out from scale-0 upward, somehow accounts for the spatial coincidence assumption by Marr [5] (pp. 70-71), also refer to Chapter 3.5.3: "Provided the ZXs in the larger channels are "accounted for" by what the smaller channels are seeing, either because they are in one-to-one correspondence with the ZXs in the smaller channels or because they are blurred, averaged copies of them, then all the evidence points to a physical reality that is roughly what the smaller channels are seeing, perhaps modified and smoothed a little by the noise-reducing, averaging effects of the large ones... If the larger channels' ZXs cannot be accounted for by what the smaller channels are seeing, then new descriptive elements, namely, discrete tokens, must be developed, because the larger channels are recording different physical phenomena..."

The two synthetic images depicted in Fig. 3.16-3 and Fig. 3.18-9 show that the detection of continuous ZX pixels in the SUMBSO\_ZX continuous image can be considered superior to the detection of continuous ZX pixels in the near orthogonal image reconstruction.

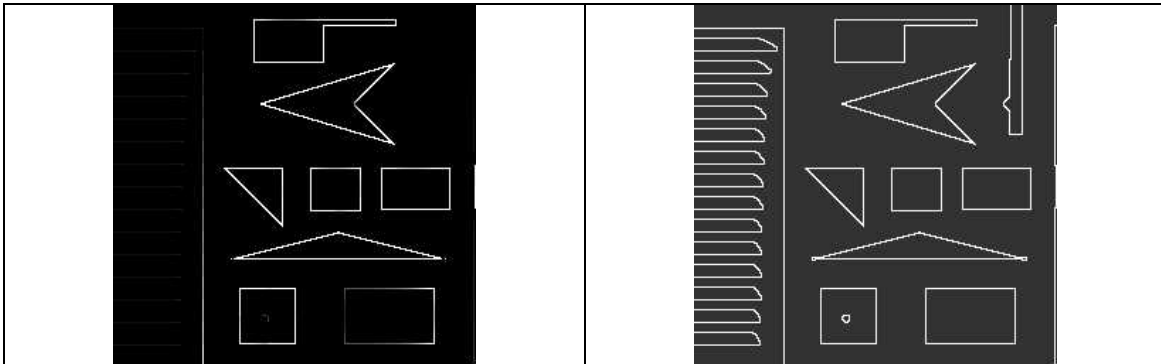


Fig. 3.18-10(a). Continuous contour representation of continuous ZX pixels selected from the near-orthogonal image reconstruction shown in Fig. 3.16-3(b). Basic Stats: Min = 50, Max = 250, Mean = 62.344360, Stdev = 48.130283.

Fig. 3.18-10(b). (Discrete) trinary contour representation of continuous contours shown in Fig. 3.18-10(a). Many genuine contours are missing.

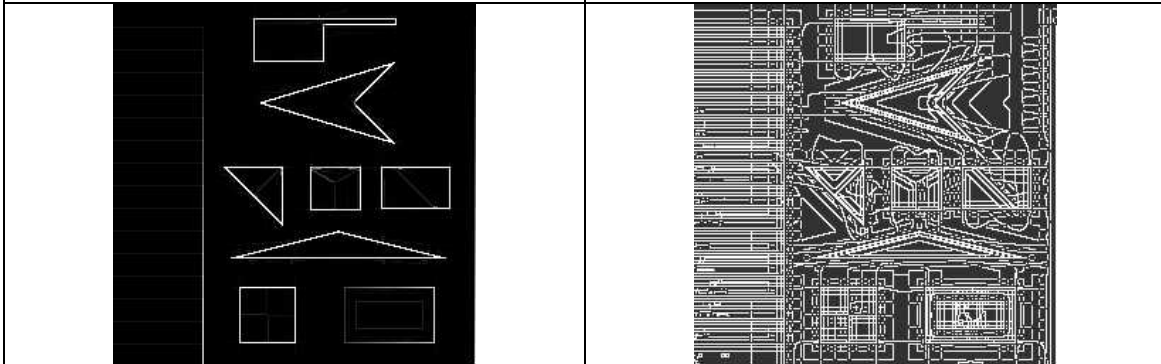


Fig. 3.18-10(c). Continuous contour representation of continuous ZX pixels selected from the continuous ZX sum across bands, scales and orientations of Fig. 3.16-3(a). As expected, Fig. 3.18-10(c) shows more genuine but small image details than Fig. 3.18-10(a). Basic Stats: Min = 0.000000, Max = 54.896248, Mean = 0.978152, Stdev = 4.909817.

Fig. 3.18-10(d). (Discrete) trinary contour representation of continuous contours shown in Fig. 3.18-10(b). Many (weak but) false contours are visible.

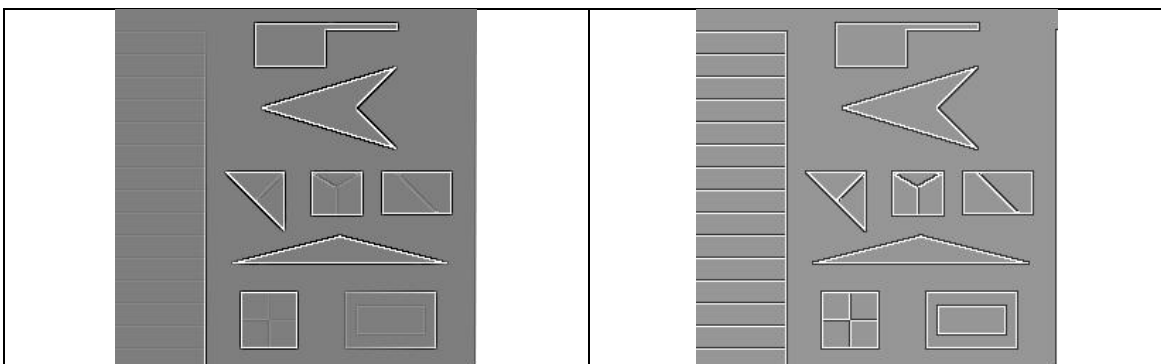
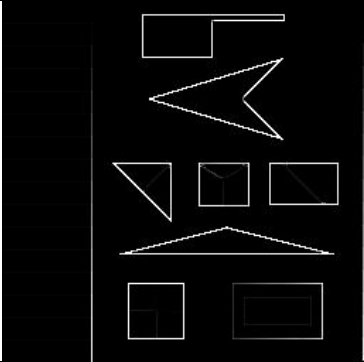
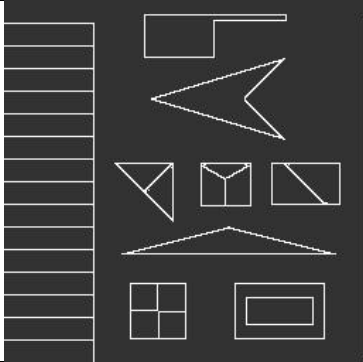


Fig. 3.18-11(a). Enhanced (filtered from scale-0 upward) continuous ZX pixels selected from the

Fig. 3.18-11(b). (Discrete) quaternary representation of the continuous enhanced ZX pixels shown in Fig. 3.18-11(a).

continuous ZX sum across bands, scales and orientations shown in Fig. 3.18-3(a).		
		
Fig. 3.18-11(c). Continuous contour representation of the continuous enhanced ZX pixels shown in Fig. 3.18-11(a). Basic Stats: Min = 0.000000, Max = 54.896248, Mean = 0.913662, Stdev = 4.910434.	Fig. 3.18-11(d). (Discrete) trinary contour representation of the continuous enhanced ZX pixels shown in Fig. 3.18-11(a). This result is better than those shown in Fig. 3.18-10(b) and Fig. 3.18-10(d). False positive and false negative contour pixels are minimized.	

### 3.19 Implemented continuous and (discrete) quaternary representation of contrast local extrema in an even- and odd-symmetric filtered image as image keypoints (endpoints, corners, junctions)

In Chapter 3.11, Fig. 3.11-1 shows a 1D function  $f(x)$  and its wavelet decomposition, composition and contrast estimation suitable for function partitioning (segmentation). For 1D simulation purposes,

- 3 pixel-wide 1D odd-symmetric filter: (+1, 0, -1).
- 3 pixel-wide 1D even-symmetric filter: (-0.5, 1, -0.5).
- 3 pixel-wide 1D Gabor filter: (+0.25, 0.5, +0.25).

The original “perceptual contrast” (PrcptlCntrst2) and function reconstruction (Rcnstrctn) expressions adopted in these simulations are:

$$\checkmark \text{PrcptlCntrst2}(x) = \text{abs}[f(x) \circ \partial^2 G/\partial x^2] + \text{abs}[f(x) \circ \partial G/\partial x]/2 = \text{abs}(\text{EvnSym}) + [\text{abs}(\text{OddSym}) / 2], \text{ hence} \quad (11-1)$$

$$\text{PrcptlCntrst2}(x) \geq 0.$$

$$\checkmark \text{Rcnstrct}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G/\partial x^2]/2. \quad (11-2)$$

Noteworthy, first,  $\text{PrcptlCntrst2}(x)$  is a non-negative combination of even- and odd-symmetric simple cells alternative to complex cells featuring a second-degree (squaring) nonlinearity proposed in Eq. (5-1) by Adelson and Bergen [42] and Burr and Morrone [26].

Second,  $\text{Rcnstrct}(x)$  is a function of the Gaussian function and even-symmetric filter,  $\partial^2 G/\partial x^2$ , exclusively, i.e., no odd-symmetric filter is involved. This agrees with the Adelson and Bergen’s image reconstruction by a Gaussian pyramid plus a Laplacian pyramid, where the Laplacian pyramid is generated as the residual (difference) between the low-pass Gaussian filtered image at scale  $s$  and the up-scale of the low-pass Gaussian filtered image at scale  $s+1$ , such that this inter-scale difference of Gaussian (DOG) is approximately equivalent to the even-symmetric Laplacian of Gaussian,  $\nabla^2 G = \partial^2 G/\partial x^2 + \partial^2 G/\partial y^2$  [42].

In Fig. 3.11-1, the proposed 1D experiments show that:

- $\text{Rcnstrct}(x)$  of the 1D function  $f(x)$  is perfect, i.e., lossless.
- $\text{PrcptlCntrst2}(x)$  allows to partition the 1D space  $x$  into (connected) segments featuring  $\text{PrcptlCntrst2}(x) > 0$  and  $\text{PrcptlCntrst2}(x) = 0$ . It is noteworthy that the assignment of boundary pixels to segments featuring condition  $\text{PrcptlCntrst2}(x) > 0$  must be carefully scrutinized. In particular, the local maxima, rather than the local minima, of  $\text{PrcptlCntrst2}(x) \geq 0$  feature the following:
  - They represent a small set of the set of pixels featuring  $\text{PrcptlCntrst2}(x) > 0$ , hence their scrutiny should be easier to accomplish, and



- They appear of particular interest for multi-scale multi-orientation contour detection and/or keypoint detection, e.g., endpoints, corners and junctions [43], [66], see Fig. 3.19-1 and Fig. 3.19-2.
- The local minima of  $\text{PrcptlCntrst2}(x) \geq 0$  do not appear to be meaningful, falling on flat areas or in the middle of ramps.

In addition to an original formulation of the  $\text{PrcptlCntrst2}(x)$  and  $\text{Rcstrctn}(x)$  equations, an original multi-scale multi-orientation extension of these entities is proposed hereafter as  $\text{PrcptlCntrst2SumSclOrntn}(x)$ .

In a multi-scale image synthesis/reconstruction, the  $\text{PrcptlCntrst2SumSclOrntn}(x)$  value is estimated as the output sum of the  $\text{PrcptlCntrst2}(x)$  values computed across bands, scales and orientations, where per scale and orientation  $\text{PrcptlCntrst2}(x) = \text{abs}(\text{EvnSym}) + [\text{abs}(\text{OddSym}) / 2]$ . Next, the local maxima and local minima of  $\text{PrcptlCntrst2SumSclOrntn}(x)$  are detected in a quaternary output image, such as:

- 1)  $\text{PrcptlCntrst2SumSclOrntn}(x)$  local maxima have value `HIGH_VALUE_UCHAR`,
- 2)  $\text{PrcptlCntrst2SumSclOrntn}(x)$  local minima have value `INTERMEDIATE_VALUE_UCHAR`.
- 3) Non local extrema which are not masked-out pixels have value `LOW_VALUE_UCHAR`.
- 4) Masked-out pixels have value `VERY_LOW_VALUE_UCHAR`.

In the continuous representation, local maxima and local minima of  $\text{PrcptlCntrst2SumSclOrntn}(x)$  have value  $> 0$ , otherwise the output pixel value is set to zero.

Examples of local extrema in  $\text{PrcptlCntrst2SumSclOrntn}(x)$  are shown in Fig. 3.19-1 and Fig. 3.19-2.

$\text{PrcptlCntrst2SumSclOrntn}(x)$  are expected to be related not only to contour points, as shown in Fig. 3.11-1, but also to the scale-invariant keypoints as extrema of: (i) the multi-scale difference of Gaussian (DOG, approximately equivalent to the Laplacian of Gaussian,  $\nabla^2 G$ ), defined by David Lowe [11], [12] and known as Scale Invariant Feature Transform (SIFT), and (ii) keypoints of an image detected as the peaks (local maxima in a  $3 \times 3$  neighbourhood) in the summed end-stopped representation [41], [43]. In [11], [12], it is written that local extrema in the  $\nabla^2 G$ -filtered image which are also ZX pixels (i.e., contour pixels) are removed from the initial set of SIFT (by Lowe), i.e., SIFT (by Lowe) and ZX pixels are complementary. In other words, it is true that the OR-combination of SIFT (by Lowe [11], [12]) and ZX pixels (by Marr, refer to Chapter 3.6) provides a superset of the whole set of local extrema in the  $\nabla^2 G$ -filtered image, i.e., (Lowe's SIFT  $\cup$  Marr's ZX pixels)  $\supseteq$  local extrema in the  $\nabla^2 G$ -filtered image, also refer to Chapter 3.6.

According to [41], the information represented at the keypoints complements the edge representation. In fact, many of the keypoints are located on occluding contours, whereas the edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of this contour map difficult. One can see that the representation of keypoints indicates precisely these critical locations.

In [43], where object (e.g., face) detection is clearly distinguished from object (e.g., face) recognition, it is speculated that keypoints detected on the basis of end-stopped operators, and in particular a few partial saliency maps that cover overlapping scale intervals, provide very important information for object detection (related to the fast “where” data stream or vision subsystem, whose goal is object segregation, consisting of object detection, i.e., separation, and object grouping, e.g., two eyes, one nose and one mouth in a given spatial arrangement form a face object). In addition, “a global saliency map provides ideal information for focus-of-attention (related to the slow “what” data stream or vision subsystem, whose goal is object recognition), because distinct peaks are found at structures with a high complexity”. This global saliency map can also be used for automatic scale selection, such that stable keypoints which are most characteristic for an object can be prepared for a first – but very fast – categorisation.

This understanding of the “where” and “what” vision information workflows agrees with works by other authors, such as in [6] (pp. 105-117), where the primary visual cortex (V1) hypothesis is formulated for creating a bottom-up saliency map for pre-attentive selection and perceptual segmentation without classification, where the saliency of a visual location is defined as the degree to which the spatial location attracts selection by bottom-up mechanisms only, also refer to [71], [142].

Furthermore, in [43] it is shown that linking keypoints from coarse to fine scales can contribute to object segregation because keypoint trajectories converge from the contours at fine scales to the centres of objects at coarse scales. In other words, at the coarsest level, each keypoint corresponds to one object, see Fig. 3.11-2. This implies that object segregation by means of a coarse-to-fine-scale strategy is feasible.



The conclusion is that keypoint detection can be complemented with multi-scale line and edge detection, which is also supposed to occur in V1. It has already been shown that object segregation and categorisation – for example for distinguishing dogs, horses and cows – can also be achieved by only considering the line/edge scale space [28]. This implies that the combination of detected keypoints and detected lines and edges will lead to improved performance, e.g., enabling face recognition, but how all information (stemming from simple-, complex- and end-stopped cells in the V1) can be combined in the best way remains an open question.

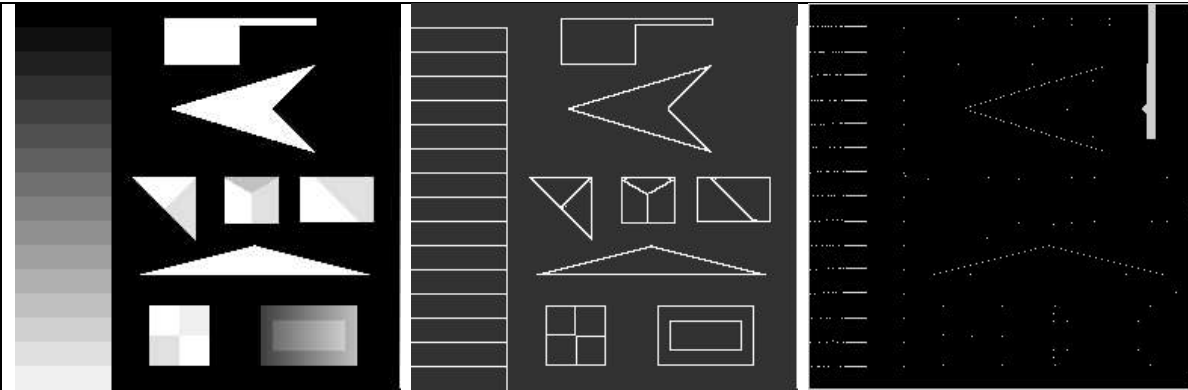


Fig. 3.19-1(a). Test synthetic image shown in Fig. 3.16-3(a).

Fig. 3.19-1(b). (Discrete) trinary contour representation of enhanced continuous ZX pixels generated from the SUMBSO ZX continuous image of Fig. 3.19-1(a).

Fig. 3.19-1(c). Quaternary representation of local extrema in  $\text{PrcptlCntrst2SumSclOrntn}(x)$  values. All local extrema are represented, irrespective of their intensity, refer to the legend described in the text. No user-defined parameter is involved with any multi-scale image processing up to this stage.



Fig. 3.19-2(a). Test natural image shown in Fig. 3.2(a).

Fig. 3.19-2(b). (Discrete) trinary contour representation of enhanced continuous ZX pixels generated from the SUMBSO ZX continuous image of Fig. 3.19-2(a).

Fig. 3.19-2(c). Quaternary representation of local extrema in  $\text{PrcptlCntrst2SumSclOrntn}(x)$  values. All local extrema are represented, irrespective of their intensity, refer to the legend described in the text. No user-defined parameter is involved with any multi-scale image processing up to this stage.





### 3.20 Original implementation of a raw and full primal sketch consistent with the Marr computational model of human vision

According to Marr, the pre-attentive vision first stage, also called *primal sketch*, consists of a *raw primal sketch* and a *full primal sketch* described below [5], refer to Chapter 3.5.3. Starting from the original contributions of Chapter 3.11 and Chapter 3.21, an innovative implementation of the raw primal sketch and full primal sketch is proposed hereafter in agreement with the Marr's computational model of human vision [5].

#### 3.20.1 The Marr's raw primal sketch in pre-attentive vision

According to Marr [5] (Figure 2-21, p. 73), the raw primal sketch employs as input the *ZX* pixels and generates as output a discrete and finite set of multi-scale *tokens* (discrete sub-symbolic image plane entities), see Chapter 3.5.3.

In general, a *zero-crossing (ZX)* is defined as a place where the value of a function passes from positive to negative [5], (p. 54). In particular, a zero-crossing in the second derivative of an intensity function (e.g., a Laplacian operator) is located where a (positive) peak or a (negative) trough (solco) in the first derivative of an intensity function occurs due to a sudden intensity change. In 1D functions, *ZX* pixels in the  $n$ -th derivative identify local extrema in the  $n$ -th - 1 derivative. However, "in two and higher dimensions there is no absolute relationship between locations of the Laplacian *ZX* curves and the local extrema of a signal. A Laplacian *ZX* curve may enclose either no extremum, one extremum, or more than one local extremum. Only in the one-dimensional case it holds that there is exactly one local extremum point between two *ZX*s of the second derivative" [11] (p. 213). In 2D intensity functions  $I(x,y)$ , since intensity changes occur at different spatial scales in an image, then their optimal detection requires the use of operators of different sizes. The most satisfying local operator according to several criteria is the filter  $\nabla^2 G$ , where  $\nabla^2$  is the Laplacian operator ( $\partial^2/\partial x^2 + \partial^2/\partial y^2$ ) and  $G(x,y)$  is the 2D Gaussian function [5] (p. 54). Hence, *ZX* pixels are intended in the  $\nabla^2 G$ -filtered image, such that  $\nabla^2(G*I) = (\nabla^2 G)*I$  [5] (pp. 57, 58), where  $\nabla^2$  is the isotropic Laplacian operator ( $\partial^2/\partial x^2 + \partial^2/\partial y^2$ ) [5] (p. 54),  $G$  is the 2D Gaussian function  $G(x,y)$ ,  $I$  is a 2D image intensity function  $I(x,y)$ ,  $(G*I)$  is a blurred image intensity function and  $\nabla^2 G$  is a circular (isotropic) even-symmetric Mexican-hat-shaped operator. This local filter features a center-surround configuration such that it is called *on-cell*, i.e., this even-symmetric operator is excited by stimulation in the central part of the receptive field and inhibited by stimulation in the outer part of the receptive field surrounding the center [22] (p. 17).

In Marr's words, the raw primal sketch consists of the following two steps.

- (1) Detect (select) *ZX* pixels in the  $\nabla^2 G$ -filtered image at multiple spatial scales, where a *ZX* pixel is located where the value of the  $\nabla^2 G$ -filtered image passes from positive to negative [5] (p. 54). This must be intended, in a more general meaning, that a *ZX* pixel is located where the value of the  $\nabla^2 G$ -filtered image passes from positive to negative or vice versa, or from positive to zero or vice versa, or from negative to zero or vice versa. According to Marr, *ZX* pixels through scale must be dealt with according to the *spatial coincident assumption* [5] (p. 70) (refer to Chapter 3.5.3 below). In his view, *ZX* pixels are not physical image contours (edges, [5], p. 68), but candidate pixels for the presence of image contours that must correspond to "physical contours". In the computer vision (CV) and remote sensing (RS) literature this phase is often called *image contour detection*, but it is important to remark that, at the raw primal sketch level of information processing, *ZX* pixels may have no physical meaning, i.e., they may belong to no physical boundary [5] (p. 68), e.g., refer to the *spatial coincidence assumption* in Chapter 3.5.3 [5] (p. 70). To accomplish *ZX* detection (selection), Marr selects the isotropic even-symmetric Laplacian operator of a Gaussian filter,  $\nabla^2 G$ , as scalable second-order differential operator. Instead, in [1], the oriented even-symmetric real part of a complex Gabor filter is designed as non-isotropic even-symmetric local filter.
- (2) According to Marr [5] (Figure 2-21, p. 73), the raw primal sketch employs as input the *ZX* pixels to generate as output the following intermediate products.
  - (i) An intermediate information primitive called *ZX segment*, defined as "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" [5] (p. 60). Hence, it is at the level of detection of *ZX* segments that *ZX* pixels and contours turn into sub-symbolic discrete image-objects (polygons).



- (ii) ZX segments must be "accounted for" through scale in compliance with the *spatial coincident assumption* [5] (pp. 70, 71), refer to Chapter 3.5.3.
- (iii) A discrete and finite set of multi-scale *tokens* (discrete sub-symbolic image plane entities), namely, *edges*, *blobs* (closed contours), *bars* and *terminators* (*discontinuities*), to be described in a discrete and finite multi-scale token description table, equivalent to a vector data model, where token attributes are: position, orientation, contrast, length, width, etc., refer to Chapter 3.5.3.
- (iv) A (binary) bit map of the image to represent basic positional information of the tokens according to a raster data model, refer to Chapter 3.5.3. According to Feldman [94], the brain's organizing principle is topology-preserving feature mapping and in the visual system these topology-preserving maps [69] are primarily spatial. In [68], Tsotsos provides the following definitions.
  - i. A retinotopic representation of a visual feature (visual parameters) is a stimulus 2D array (2D regular gridded data set) with P elements. In this retinotopic representation, physically adjacent elements represent spatially adjacent regions in the visual scene [68].
  - ii. A map is defined as a retinotopic representation of only one type of visual parameter [68].
  - iii. There may be 30 visual areas or so in primates, but not all are organized retinotopically, and, even then, with varying degrees of retinotopy [68]. The areas commonly accepted as being retinotopic include VI, V2, V3, MT, and V4, whereas the nonretinotopic ones include IT, posterior parietal cortex, and the frontal eye fields. According to van Essen and Maunsell [70], the division between retinotopic and nonretinotopic areas, although fuzzy in general, may be placed after areas MT and V4 and before IT, area 7, and the frontal eye fields. Maps seem to be organized hierarchically, as a partial ordering, so that the greater the distance from the retina, the smaller the maps are, and the larger the receptive fields of their neurons. The Marr's binary bit map is equal to 1 at the corresponding position of a token described in the token description table. This bit map is used for search of inter-token spatial relations, that are rather local in the perceptual grouping phase (e.g., Gestalt's law of proximity), without the trouble of searching through the whole list of primal sketch descriptors for inter-token spatial relations [5] (p. 79), refer to Chapter 3.4.2.

Vice versa, in a traditional 1D image analysis approach the topological information in the (2D) image domain is totally lost, also refer to topology-preserving mapping of a data manifold (multivariate distribution) onto a graph (network of processing elements and lateral connections) and vice versa, from the graph onto the data manifold [69].

To summarize, according to Marr the output of the raw primal sketch consists of two representations of a discrete and finite set of tokens in tabular form (vector data model) and as a bit map (raster data model). Marr pointed out these two vector and raster representations of tokens are complementary not alternative. In Marr's words "the important point is that the bit map saves the trouble of searching through the whole (1D) list of primal sketch descriptors (see [5], p. 67, Figure 2-21) checking each coordinate to see whether it falls within the specified 2D spatial neighborhood in the image-domain. The underlying reason why using a literal bit map representation of an image is more efficient is that most of the inter-token spatial relationships that must be examined in early vision stage are rather local. If we had to examine arbitrary, scattered, salt-and-pepper-like configurations, then a bit map would probably be no more efficient than a list (tabular form)", refer to Chapter 3.5.3.

Unfortunately, in his seminal work [5], Marr proposes no algorithm to extract ZX pixels, ZX segments and tokens from ZX segments.



### 3.20.2 Original automated implementation of a raw primal sketch consisting of discrete tokens (texels) as ZX segments

In the CV and RS literature, the raw primal sketch is often called *image segmentation* (image-object detection, image partitioning into a discrete and finite set of image-objects). For example, at this stage *textons* (texture elements) are expected to be extracted as tokens. This means that at the information processing level of the raw primal sketch, texture boundaries are not detected, but only textons are identified (refer to [5], Figure 2-7, p. 53). Unfortunately, in his seminal work, Marr proposes no algorithm to extract ZX pixels, ZX segments and tokens from ZX segments. According to Li Zhaoping [6], "the computer vision community has tried to solve the problem of image segmentation for decades without a satisfactory solution." According to [24], if we require that a computer vision model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, see Fig. 3.4-1, then the number of published models is surprisingly small, as proved in [25].

To recapitulate, generation of a raw primal sketch is still an open challenge in the CV and RS literature.

Chapter 3.11 and Chapter 3.12 proposed original definitions and implementations of ZX pixels and scale-invariant keypoints, such as cross-points, corners and end-points. To detect ZX segments from ZX pixels automatically (without user-machine interaction), we propose the following.

First, a 2D gridded dataset of ZX pixels is automatically partitioned into a *preliminary ZX segment map*, consisting of "white" connected segments whose pixels feature  $(\partial^2 G / \partial n^2 * I) > 0$  (concavity down), "gray" (ZC segments) and "dark" connected segments whose pixels feature  $(\partial^2 G / \partial n^2 * I) < 0$  (concavity up). Next, a *preliminary ZX segment map* is transformed into a *final ZX segment map* where ZX pixels are eventually re-assigned to neighboring ZX segments (see Fig. 3.11-1 and refer to Chapter 3.10) according to the following original algorithm, that was designed, implemented and tested.

Based on Fig. 3.11-1, it is intuitive to understand that ZX pixels, each one assigned to one polygon in the *preliminary ZX segment map*, may have to be re-assigned to a neighboring polygon according to the following three criteria, described in pseudo-code.

- (1) Any zero-crossing (ZX) pixel should be merged with the neighboring pixel whose  $\text{PrcptlCntrst2}(x) == 0$  or "low", if any. This is equivalent to requiring the neighboring pixel to belong to a flat area. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.
- (2) Any pair of neighboring pixels, either ZX or not, should be merged with the neighboring pixel whose  $\text{DeltaGrayValue} == 0$  or "low", if any. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.
- (3) Any pair of neighboring pixels, either ZX or not, both featuring  $\text{PrcptlCntrst2}(x) == 0$  or "low", should be merged into the same zero-concavity (ZC) segment, either new or pre-existing.

Only once the aforementioned ZX pixel re-assignment criteria are applied to a *preliminary ZX segment map*, then a *final ZX segment map* is generated as output.

This ZX segment detection algorithm depends on two hidden general-purpose user- and application-independent parameters, to be set based on *a priori* psychophysical knowledge: the  $\text{WeberSensitivityFraction} = 0.010 = 1\%$  in range  $[0, 1]$  and the  $\text{NrmlzdPrcptlCntrstActionPotential} = 0.012 = 1.2\%$  in range  $[0, 1]$ .

The Weber–Fechner sensitivity law [67] states that a physical entity (e.g., a weight) seems to have to increase by 5% for someone to be able to reliably detect (sense) the increase, and this minimum required fractional increase (of 5/100 of the original weight) is referred to as the "Weber (sensitivity) fraction" for detecting changes in weight. Other discrimination tasks, such as detecting changes in brightness, or in tone height (pure tone frequency), or in the length of a line shown on a screen, may have different Weber sensitivity fractions, but they all obey Weber's law in that observed values need to change by at least some small but constant proportion of the current value to ensure human observers will reliably be able to detect (sense) that change.



In physiology, an action potential is a shortlasting event in which the electrical membrane potential of a cell rapidly rises and falls, following a consistent trajectory. Action potentials occur in several types of animal cells, called excitable cells, which include neurons, muscle cells, and endocrine cells. Each excitable patch of membrane has two important levels of membrane potential: the resting potential, which is the value the membrane potential maintains as long as nothing perturbs the cell, and a higher value called the threshold potential. At the axon hillock of a typical neuron, the resting potential is around  $-70$  millivolts (mV) and the threshold potential is around  $-55$  mV. Synaptic inputs to a neuron cause the membrane to depolarize or hyperpolarize; that is, they cause the membrane potential to rise or fall. Action potentials are triggered when enough depolarization accumulates to bring the membrane potential up to threshold. When an action potential is triggered, the membrane potential abruptly shoots upward and then equally abruptly shoots back downward, often ending below the resting level, where it remains for some period of time [91].

Hereafter, the original algorithm for ZX segment detection from ZX pixels detected by a local oriented tripole operator, which is 3-pixel wide and single-scale, capable of dealing with both panchromatic and color images consisting of one or more spectral channels, is described in pseudocode.

(Start pseudocode)

```

*****
; Author: Andrea Baraldi
; Last update: March 23, 2016 == day/month/year = 23/03/2016.
*****

; OBJECTIVE: Physical knowledge-based computational model of low-level (pre-attentional)
; human vision capable of fully automated multi-spectral or panchromatic
; zero-crossing (ZX) pixel detection and ZX segment detection, where ZX segments
; are equivalent to texels (texture elements), in compliance with the Marr's raw primal sketch.
;
; For 1-D simulation purposes, a tripole (3-pixel long, 1-pixel wide) operator is adopted.
; It consists of an even-symmetric, an odd-symmetric and a low-pass filter, specifically,
;
; • 3-pixel long, 1-pixel wide 1-D odd-symmetric filter: (+1, 0, -1).
; • 3-pixel long, 1-pixel wide 1-D even-symmetric filter: (-0.5, 1, -0.5).
; • 3-pixel long, 1-pixel wide 1-D low-pass Gaussian filter: (+0.25, 0.5, +0.25).
;
; Nomenclature:
; * Zero-crossing (ZX) segments must be detected from ZX pixels, according to Marr.
; * ZX segment detection = texel/texton/token detection.
; * The combination of quantitative variables generated as output
; by (multi-scale) multi-orientation even-symmetric filter (EF) banks, can be partitioned
; into a 3-level image consisting of:
; (i) Zero-concavity (ZC) segments.
; (ii) Positive-convavity (PC) segments.
; (iii) Negative-concavity (NC) segments.

;-- User-defined parameters

OneBand_PerceptualPixelPairCntrst_WeberSensitivityFraction = 0.010 ;0.01 = 1% in range [0, 1]

;The perceptual parameter fOneBand_PerceptualPixelPairCntrst_WeberSensitivityFraction is introduced as a degree of
tolerance (specifically, as a normalized (dis)similarity threshold) around the concept of two pixels featuring the "same"
gray value (with some degree of tolerance, i.e., DeltaGrayValue is "LOW", if NrmldLocalContrast = abs(Pixel1Value -
Pixel2Value) / (Pixel1Value + Pixel2Value), where Pixel1Value and Pixel2Value are >= 0, such that NrmldLocalContrast
in [0, 1], is <= fOneBand_PerceptualPixelPairCntrst_WeberSensitivityFraction.

fNrmldPrcptlCntrstActionPotential_AndreaNrmldPrcptlCntrstNoisePrmtr = 0.012
;visually assessed on the SUSAN synthetic test image and on the S2A_LargeSet of Austria.

```



```
;--- Allocate and initialize to zero memory arrays

Out_SUMB_EF_OneOrtnMap = FLTARR(iHalf8AdjcnryOrientations, iMaxSample, iMaxLine)

fOut_SUMB_EF_OneOrtnMskdByZxMap = FLTARR(iHalf8AdjcnryOrientations, iMaxSample, iMaxLine)

cZeroXingPixelsBinaryMap = BYTARR(iMaxSample, iMaxLine)

fOut_SUMOB_EF_Map = FLTARR(iMaxSample, iMaxLine)

fOut_SUMOB_EF_MskdByZX_Map = FLTARR(iMaxSample, iMaxLine)

;--- Program start

;-----
; Program 1 of 12. ; Near-orthogonal image analysis and synthesis by means of a per-pixel single-scale multi-orientation
even-symmetric 2nd-order derivative for local concavity estimation.
; (i) IF fOutNrmlzdEvenSymtrc2ndOrderDrvtvForConcavityEstmntMap[iOrientation, iSample, iLine]
; EQ 0 THEN ZERO CONCAVITY (FLAT AREA OR RAMP).
; (ii) IF fOutNrmlzdEvenSymtrc2ndOrderDrvtvForConcavityEstmntMap[iOrientation, iSample, iLine]
; LT 0 THEN CONCAVITY UP.
; (iii) IF fOutNrmlzdEvenSymtrc2ndOrderDrvtvForConcavityEstmntMap[iOrientation, iSample, iLine]
; GT 0 THEN CONCAVITY DOWN.
;-----

For each Ortn = 0 to 3,

  For each Band = 1, MaxBand

    Compute fOut_SUMB_EF_OneOrtnMap[Ortn, iSample, iLine] based on per-pixel
    fEvenSymtrc2ndOrderDrvtvForConcavityEstmnt values, where acronyms SUMB = Sum across bands, EF =
    Even-symmetric filter.

    Compute fOut_SUMOB_EF_OF_NrmlzdPrcptlCntrst_Map[iSample, iLine] where fPrcptlCntrst =
    abs(fEvenSymtrc2ndOrderDrvtvForConcavityEstmnt) + abs(fOutOddSymtrc1stOrderDrvtv / 2.)

  Endfor Band

  Update fOut_SUMOB_EF_Map[iSample, iLine] = fOut_SUMOB_EF_Map[iSample, iLine] +
  fOut_SUMB_EF_OneOrtnMap[Ortn, iSample, iLine]

;--- Compute zero-crossings per fOut_SUMB_EF_OneOrtn

  cZeroXingPixelsBinaryMap = ZXpixel(fOut_SUMB_EF_OneOrtnMap)

;--- Mask fOut_SUMB_EF_OneOrtnMap[Ortn, *, *] with the zero-crossing binary mask

  fOut_SUMB_EF_OneOrtnMskdByZxMap = fOut_SUMB_EF_OneOrtnMap * cZeroXingPixelsBinaryMap

;--- Update fOut_SUMOB_EF_MskdByZX_Map

  Update fOut_SUMOB_EF_MskdByZX_Map = fOut_SUMOB_EF_MskdByZX_Map +
  fOut_SUMB_EF_OneOrtnMskdByZxMap

Endfor Ortn

Linear transform fOut_SUMOB_EF_OF_NrmlzdPrcptlCntrst_Map into range [0, 1]
```





```

;--- Normalize fOut_SUMOB_EF_MskdByZX_Map into range [-iMaxCharValue, iMaxCharValue]

Normalize fOut_SUMOB_EF_MskdByZX_Map into range [-iMaxCharValue, iMaxCharValue] by dividing it by
iTotNearOrthgnlCmpnts = MaxBand * MaxOrntn

cZeroXingPixelsBinaryMap_fOut_SUMOB_EF_MskdByZX_Map = ZXpixel(fOut_SUMOB_EF_MskdByZX_Map)

;--- Overwrite fOut_SUMOB_EF_MskdByZX_Map with its own continuous zero-crossing pixel values

fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map      =      fOut_SUMOB_EF_MskdByZX_Map      *
cZeroXingPixelsBinaryMap_fOut_SUMOB_EF_MskdByZX_Map

;--- Compute cOut_3level_CntnuousZX_SUMOB_EF_MskdByZX_Map

cOut_3level_CntnuousZX_SUMOB_EF_MskdByZX_Map[*,*] = $
((fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[*,*] LT 0) * cEvenSymtricFilterNegValueLabel + $
(fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[*,*] EQ 0) * cEvenSymtricFilterZeroValueLabel + $
(fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[*,*] GT 0) * cEvenSymtricFilterPosValueLabel)

;      ;(1) IF fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[iSample, iLine] EQ 0 THEN
;      ;      ZERO CONCAVITY (FLAT AREA OR RAMP).
;      ;(2) IF fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[iSample, iLine] LT 0 THEN
;      ;      CONCAVITY UP.
;      ;(3) IF fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map[iSample, iLine] GT 0 THEN
;      ;      CONCAVITY DOWN.

;-----
Program stage 2 of 12. Two-pass partitioning of the binary map (iTarget_3level_SUMOB_EF_Map EQ 2) =
(fTarget_SUMOB_EF_Map EQ 0) = zero-concavity (ZC)
;-----

fTarget_SUMOB_EF_OneOrntnMap = fOut_SUMOB_EF_OneOrntnMskdByZxMap

iTarget_3level_SUMOB_EF_Map = cOut_3level_CntnuousZX_SUMOB_EF_MskdByZX_Map

fTarget_SUMOB_EF_Map = fOut_CntnuousZX_SUMOB_EF_MskdByZX_Map

cTarget_ZeroXingPixels_SUMOB_EF_BinaryMap      =
cZeroXingPixelsBinaryMap_fOut_SUMOB_EF_MskdByZX_Map

iTargetBinaryMapToSgmntInTwoPasses[*,*]      =      iTarget_3level_SUMOB_EF_Map[*,*]      EQ
cEvenSymtricFilterZeroValueLabel

iOutSgmntnOfBinaryZeroCncvty_SUMOB_EF_Map      =
TwoPassConnectdComponentMultiLvlImgLablngAlgrthm(iTargetBinaryMapToSgmntInTwoPasses)

;-----
Program stage 3 of 12. Two-pass multiple-criteria assignment of pairs of neighboring pixels to zero-concavity (ZC)
segments, either pre-existing or new. The result is a partition of an augmented (either the same or larger) zero-concavity
binary map.
;-----

```

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 1.



Any zero-crossing (ZX) pixel should be merged with the neighboring pixel whose  $\text{PrcptlCntrst2}(x) == 0$  or “low”, if any. This is equivalent to requiring the neighboring pixel to belong to a flat area. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 2.

Any pair of neighboring pixels, either ZX or not, should be merged with the neighboring pixel whose  $\text{DeltaGrayValue} == 0$  or “low”, if any. Hence, these two pixels belong to the same zero-concavity (ZC) segment, either new or pre-existing.

- ATTENTION, ADJACENT PIXEL-PAIR MERGING into a new or pre-existing Zero-Concavity (ZC) segment, RULE 3.

Any pair of neighboring pixels, either ZX or not, both featuring  $\text{PrcptlCntrst2}(x) == 0$  or “low”, i.e., both neighboring pixels belong to a flat or nearly flat area, should be merged into the same zero-concavity (ZC) segment, either new or pre-existing. This region growing rule applies to image areas where there are smooth (low) concavity values, encompassing changes in the sign of concavity. Example: the smooth skin effect in the shoulder or cheeks of Lenna.

`iTarget_3level_EF_Map` = overwritten, now featuring an “augmented” zero-concavity (ZC) layer.

```

;-----
Program stage 4 of 12. Compute the segmentation map of a binary map (iTarget_3level_SUMOB_EF_Map EQ 1) =
(fTarget_SUMOB_EF_Map LT 0) = concavity up, accounting for the previous augmentation of the zero-concavity (ZC)
binary mask.
;-----

```

```

iTargetBinaryMapToSgmtInTwoPasses[*,*]      =      iTarget_3level_SUMOB_EF_Map[*,*]      EQ
cEvenSymtricFilterNegValueLabel

```

```

iOutSegmentationOfCncvtyUpBinaryMap          =
TwoPassConnectdComponentMultiLvlImgLablngAlgrthm(iTargetBinaryMapToSgmtInTwoPasses)

```

```

;-----
Program stage 5 of 12. Compute the segmentation map of a binary map (iTarget_3level_SUMOB_EF_Map EQ
cEvenSymtricFilterPosValueLabel) = (fTarget_SUMOB_EF_Map GT 0) = concavity down, accounting for the previous
augmentation of the zero-concavity binary mask.
;-----

```

```

iTargetBinaryMapToSgmtInTwoPasses[*,*]      =      iTarget_3level_SUMOB_EF_Map[*,*]      EQ
cEvenSymtricFilterPosValueLabel

```

```

iOutSegmentationOfCncvtyDownBinaryMap       =
TwoPassConnectdComponentMultiLvlImgLablngAlgrthm(iTargetBinaryMapToSgmtInTwoPasses)

```

```

;-----
Program stage 6 of 12. Apply a two-pass partitioning of a new segmentation map iInputFinalSegmentationMap =
iOutSgmtnOfBinaryZeroCncvty_SUMOB_EF_Map[*,*] + (iOutSegmentationOfCncvtyUpBinaryMap[*,*] +
iMaxSgmtAfterEqvIncRemoval_ZeroCncvty) + (iOutSegmentationOfCncvtyDownBinaryMap[*,*] +
iMaxSgmtAfterEqvIncRemoval_ZeroCncvty + iMaxSgmtAfterEqvIncRemoval_CncvtyUp)
;-----

```



```
iOutputFinalSegmentationMap =
TwoPassConnectdComponentMultiLvImgLablngAlgrthm(iInputFinalSegmentationMap)
```

```
-----
Program stage 7 of 12. Provide quantitative quality indicators of the intermediate iInputFinalSegmentationMap product,
specifically: (i) a piecewise-constant approximation of each band of the input image. (ii) an 8-adjacency cross-aura measure
of the iInputFinalSegmentationMap.
-----
```

```
-----
Program stage 8 to Program stage 10 of 12. Once the augmented Zero-concavity (ZC) segments have been detected, which
incorporate some of the ZX pixels, "white" (concavity down,  $(\partial^2 G / \partial n^2 * I) > 0$ ) and "dark" (concavity up,  $(\partial^2 G / \partial n^2 * I) < 0$ )
segments must be detected per orientation, and then apply the superposition principle to pool "white" and "dark"
segments across orientations. It works as follows:
```

```
FOR iMainOrientation = 0L, iHalf8AdjcnctyOrientations - 1,
```

```
-----
Program stage 8 of 12. Two-pass segmentation of an orientation-specific binary map
(fTarget_SUMB_EF_OneOrntnMap[iMainOrientation,*,*] LT 0) = concavity up, where the zero-concavity (ZC)
segments, detected as (iTarget_3level_SUMOB_EF_Map EQ cEvenSymtricFilterZeroValueLabel), must be
masked out (removed).
-----
```

```
-----
Program stage 9 of 12. Two-pass segmentation of an orientation-specific binary map
(fTarget_SUMB_EF_OneOrntnMap[iMainOrientation,*,*] GT 0) = concavity down, where the zero-concavity
(ZC) segments, detected as (iTarget_3level_SUMOB_EF_Map EQ cEvenSymtricFilterZeroValueLabel), must be
masked out (removed).
-----
```

```
-----
Program stage 10 of 12. Two-pass segmentation of the multi-level image iInputFinalSegmentationOneOrntnMap =
iOutSgmntnOfBinaryZeroCncvty_SUMOB_EF_Map[*,*] +
(iOutSegmentationOfCncvtyUpOneOrntnBinaryMap[*,*] + iMaxSgmntAfterEqvlncRemoval_ZeroCncvty) +
(iOutSegmentationOfCncvtyDownOneOrntnBinaryMap[*,*] + iMaxSgmntAfterEqvlncRemoval_ZeroCncvty +
iMaxSgmntAfterEqvlncRemoval_CncvtyUp)
```

```
iOutput_EF_OB_CumulativeNonZeroCncvtySgmntnMap =
TwoPassConnectdComponentMultiLvImgLablngAlgrthm(iInputFinalSegmentationOneOrntnMap)
```

```
Endfor iMainOrientation
```

```
-----
Program stage 11 of 12. Combine iOutSgmntnOfBinaryZeroCncvty_SUMOB_EF_Map with
iOutput_EF_OB_CumulativeNonZeroCncvtySgmntnMap and provide a final two-pass segmentation map of the multi-
level image:
```

-----

```
iInputFinalSegmentationOneOrntnMap[*,*] = (iOutput_EF_OB_CumulativeNonZeroCncvtySgmntnMap[*,*] *
(~(iTarget_3level_SUMOB_EF_Map[*,*] EQ cEvenSymtricFilterZeroValueLabel))) +
(iOutSgmntnOfBinaryZeroCncvty_SUMOB_EF_Map[*,*] +
iMaxSgmntAfterEqvlnCRemoval_EF_OB_CumulativeNonZeroCncvtySgmntnMap) *
(iTarget_3level_SUMOB_EF_Map[*,*] EQ cEvenSymtricFilterZeroValueLabel)
```

```
iOutput_EF_OB_CumulativeNonZeroCncvtySgmntnMap =
TwoPassConnectdComponentMultiLvlImgLablngAlgrthm(iInputFinalSegmentationOneOrntnMap)
```

-----

Program stage 12 of 12. Provide quantitative quality indicators of the iOutput\_EF\_OB\_CumulativeNonZeroCncvtySgmntnMap product, specifically: (i) a piecewise-constant approximation of each band of the input image. (ii) an 8-adjacency cross-aura measure of the iInputFinalSegmentationOneOrntnMap.

-----

(End pseudocode)

Results of an automated ZX segment detection starting from ZX pixels based on a multi-orientation single-scale MS color-sensitive tripole (consisting of only three pixels) are shown in Fig. 3.20-1, Fig. 3.20-2, Fig. 3.20-3, Fig. 3.20-4 and Fig. 3.20-5.

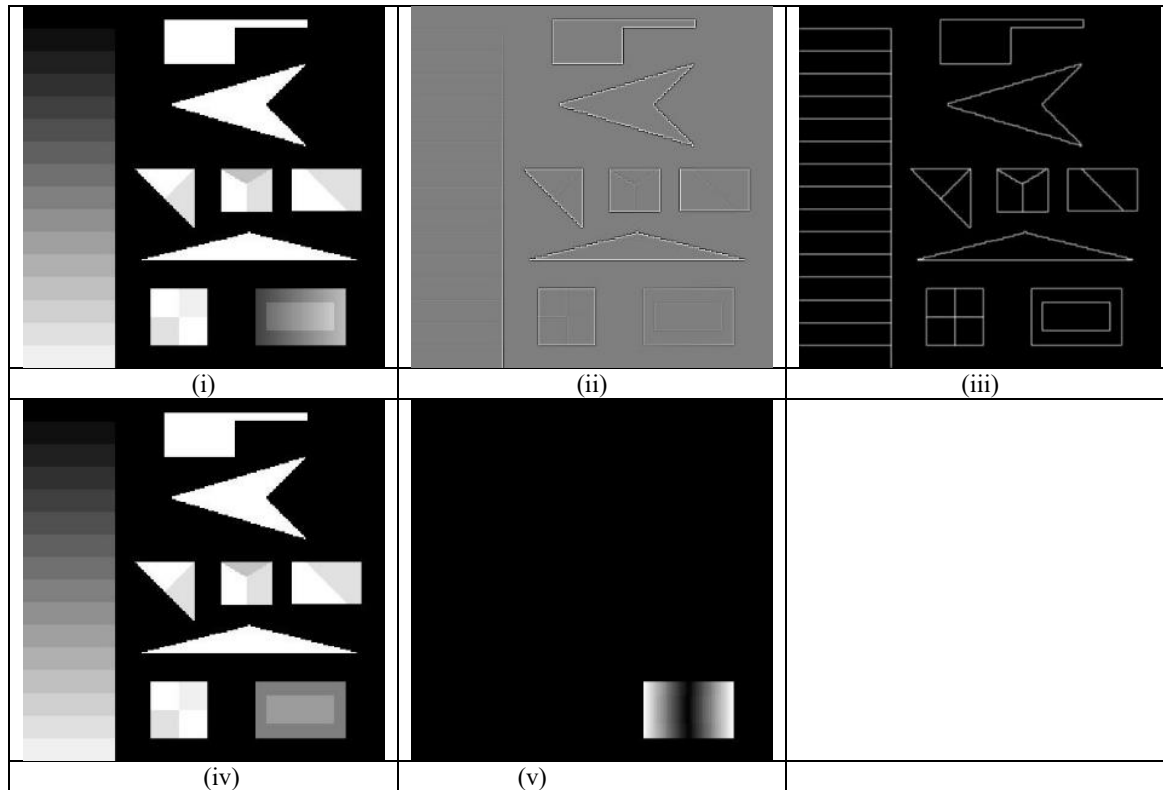


Fig. 3.20-1. (i) SUSAN synthetic test panchromatic image [90], consisting of 31 segments according to human perception. (ii) One-scale multi-orientation even- and odd-symmetric filter combination. Normalized perceptual contrast, in range  $\{-1, 1\}$ , with sign provided by zero-crossing pixels detected by even-symmetric filter combinations, exclusively. (iii) Automated image segmentation into ZX segments. Exactly 31 segments are detected. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$ . (iv) Object mean view =

object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction, per-pixel root mean square error.

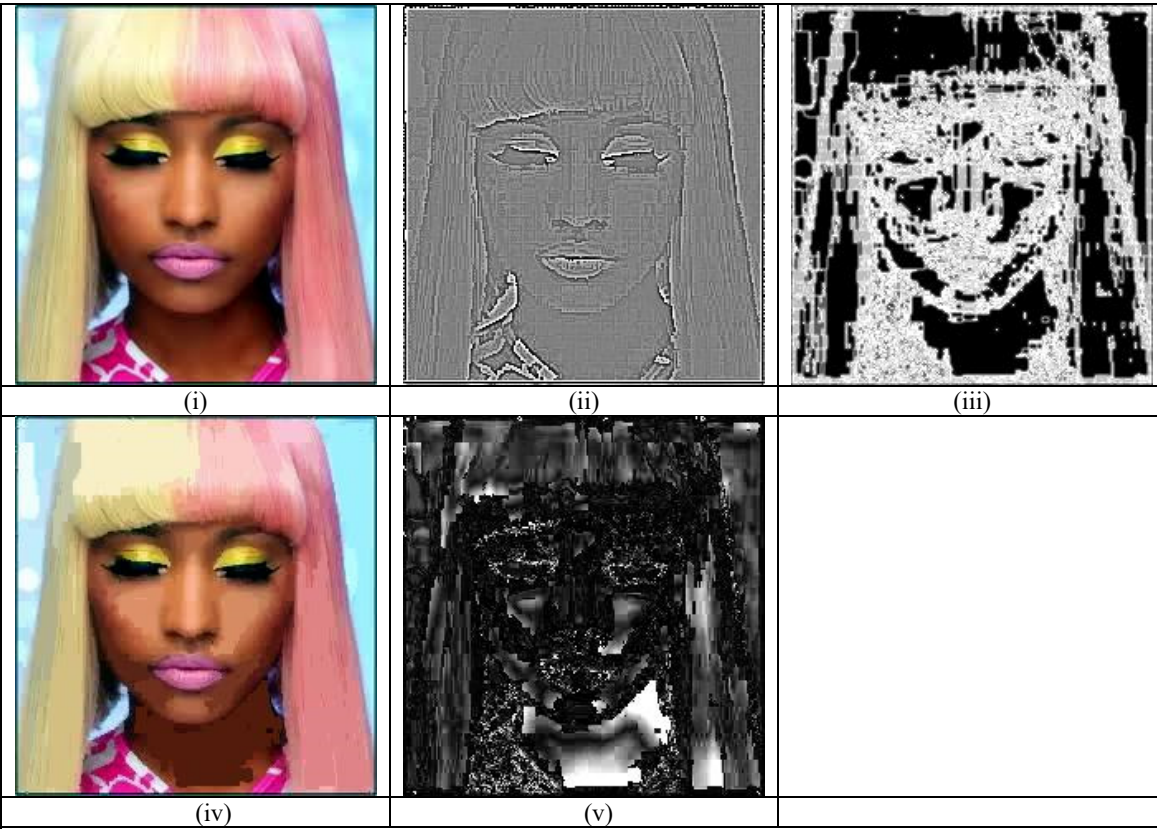
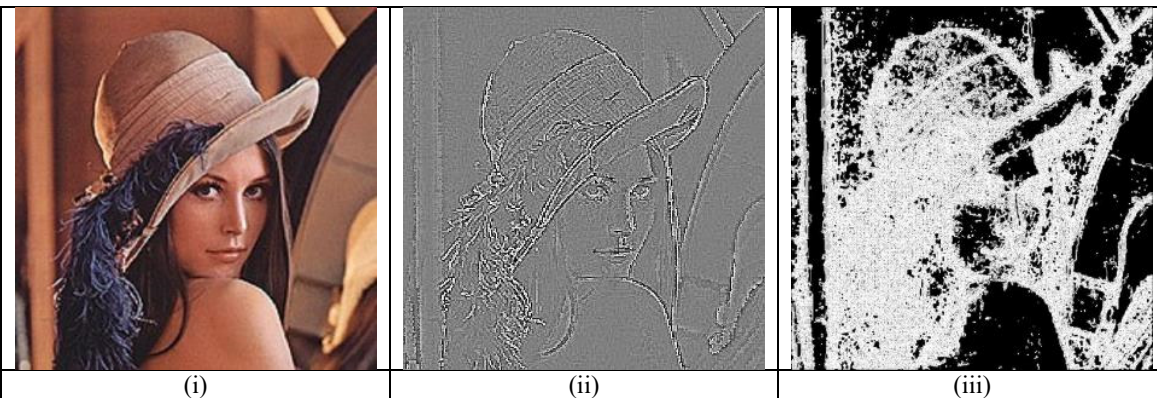


Fig. 3.20-2. (i) RGB image of a human face downloaded from Google Image. (ii) One-scale multi-orientation even- and odd-symmetric filter combination. Normalized perceptual contrast, in range  $\{-1, 1\}$ , with sign provided by zero-crossing pixels detected by even-symmetric filter combinations, exclusively. (iii) Automated image segmentation into ZX segments, exactly 31 segments are detected. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$ . (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction, per-pixel root mean square error.





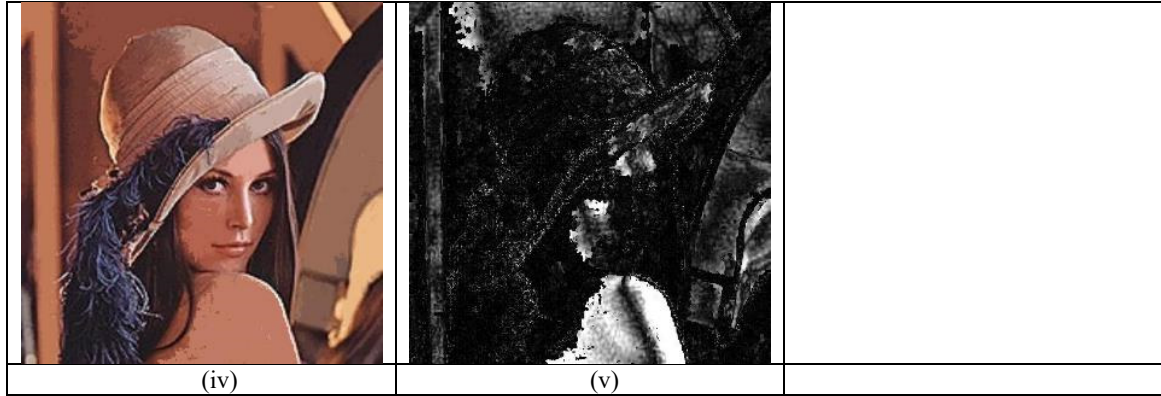


Fig. 3.20-3. (i) RGB image of Lenna (without any enhancement). (ii) One-scale multi-orientation even- and odd-symmetric filter combination. Normalized perceptual contrast, in range  $\{-1, 1\}$ , with sign provided by zero-crossing pixels detected by even-symmetric filter combinations, exclusively. (iii) Automated image segmentation into ZX segments. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$ . (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction, per-pixel root mean square error.

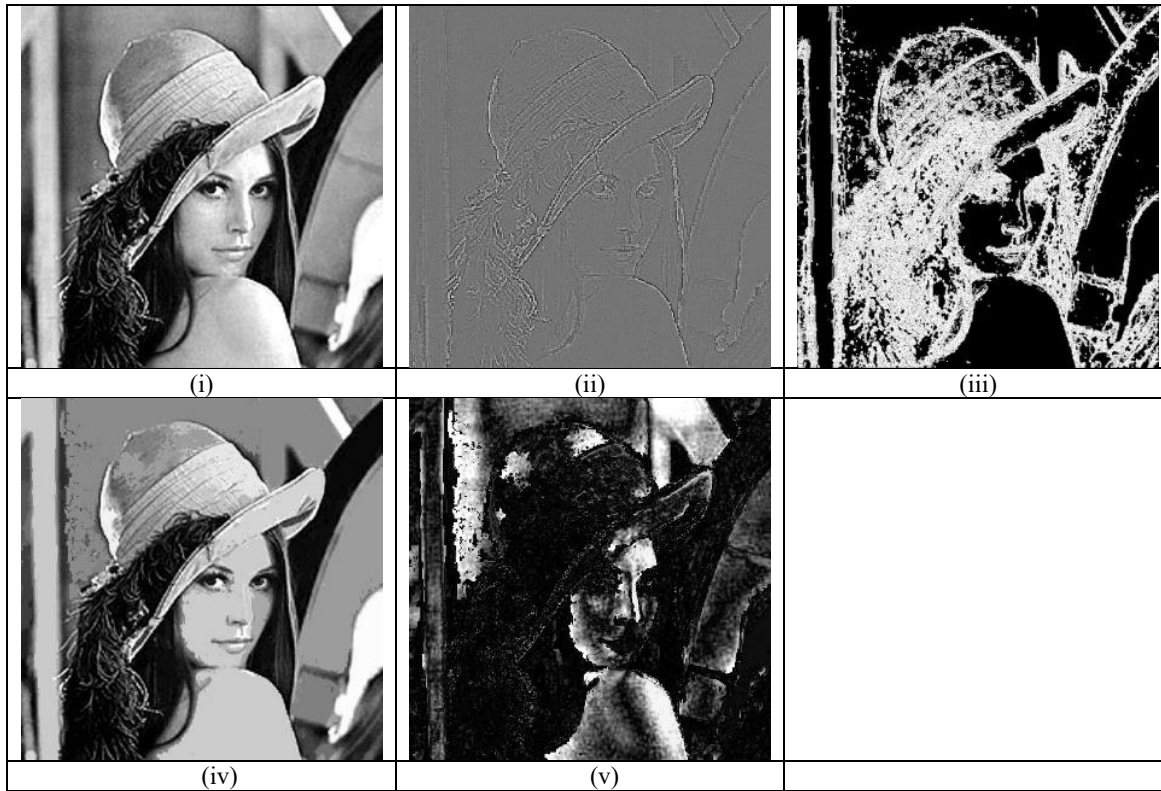
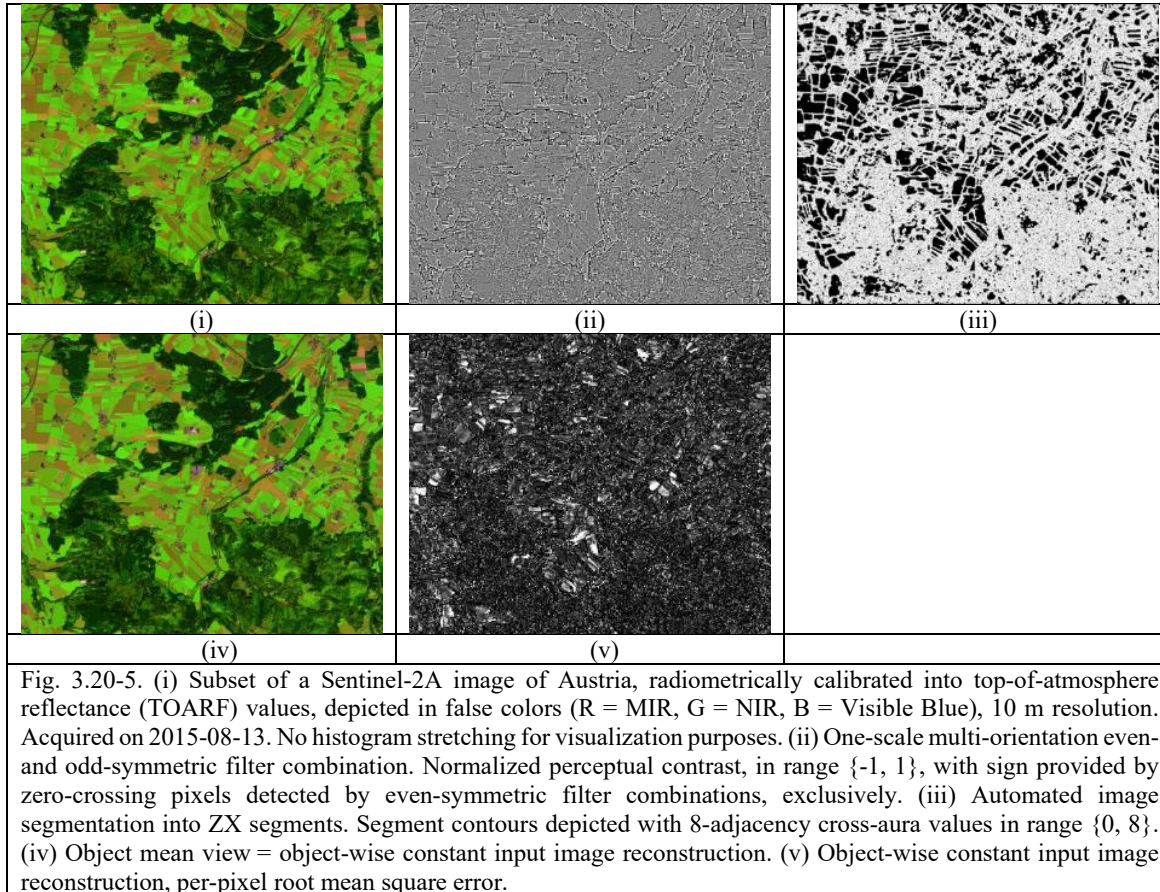


Fig. 3.20-4. (i) Panchromatic image of Lenna (without any enhancement). (ii) One-scale multi-orientation even- and odd-symmetric filter combination. Normalized perceptual contrast, in range  $\{-1, 1\}$ , with sign provided by zero-crossing pixels detected by even-symmetric filter combinations, exclusively. (iii) Automated image segmentation into ZX segments. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$ . (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction, per-pixel root mean square error.



### 3.20.3 The Marr's full primal sketch or perceptual grouping of tokens into larger-scale tokens, such as texture contours

Perceptual organization (PO) lies at the center of vision. It refers to the human visual ability to extract significant image relations from low-level (pre-attentive vision) image plane entities (tokens) without any knowledge of the image content and group them to obtain meaningful intermediate-level data representations that make high-level (attentive vision) object-recognition possible.

Gestalt psychologists undertook the first detailed study of the grouping phenomenon in human vision in the first part of this century [7]. The Gestalt psychologists at the time identified certain rules or principles to explain the particular way the human perceptual system groups tokens together. They suggested that grouping among tokens takes place based on the following Gestalt criteria of perceptual grouping:

- I. Symmetry.
- II. Similarity.
- III. Proximity.
- IV. Closure (closed contour).
- V. Smoothness (curved completion, good continuation).
- VI. Collinearity, (7)
- VII. Parallelism.

In his seminal work [5], Marr states explicitly that, to investigate the *spatial organization* of tokens in an image, elsewhere called *spatial distribution* of tokens (p. 79), *perceptual grouping* (p. 91) or *texture discrimination* (p. 96), *discontinuities* (to be intended as synonyms of “*abrupt changes*” or “*singularities*”, refer to Chapter 3.7) in six image parameters



(quantitative properties, numerical features) must be investigated. Three of them are intrinsic to a token (token-specific) and three pertain to the spatial arrangement of tokens, refer to Chapter 3.5.3 and Chapter 3.7

1. Token-specific metrological attributes, affecting perceptual grouping of tokens.
  - (i) Average achromatic intensity (brightness, where brightness is defined as perceived luminance [1]) or (chromatic) color.
  - (ii) Geometric properties, i.e., size and shape properties (e.g., length, width, compactness, rectangularity, roughness/straightness of boundaries, simple connectivity, etc.). For example, refer to [29].
  - (iii) Orientation.
2. Spatial arrangement of tokens. If there is a discontinuity in any of the following attributes, then there is a change in perceived texture, i.e., there is a texture boundary.
  - (i) Local density of tokens.
  - (ii) Distance apart of tokens. Estimated by the so-called Steven's algorithm for recovering the local orientation of tokens [30], based on an information primitive called, by Marr, inter-token virtual line (p. 82).
  - (iii) Local orientation of tokens, also estimated by the Steven's inter-token virtual line detection algorithm [30].

According to Vecera and Farah: "we have demonstrated that image segmentation can be influenced by the familiarity of the shape being segmented", "these results are consistent with the hypothesis that image segmentation is an interactive (hybrid inference) process" "in which top-down knowledge partly guides lower level processing". "If an unambiguous, yet unfamiliar, shape is presented, top-down influences are unable to overcome powerful bottom-up cues. Some degree of ambiguity is required to overcome bottom-up cues in such situations. The main conclusion from these simulation studies is that while bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization " [8] (p. 1294).

In [6] (pp. 105-117), the primary visual cortex (V1) hypothesis is formulated for creating a bottom-up saliency map for pre-attentive selection and perceptual segmentation without classification, where the saliency of a visual location is defined as the degree to which the spatial location attracts selection by bottom-up mechanisms only, also refer to [71], [142].

### 3.20.4 Original automated implementation of a full primal sketch for texture segmentation: multi-scale texture binary profile

Unfortunately, in his seminal work Marr proposed no algorithm to extract large-scale tokens from smaller-scale tokens according to a constructive (iterative and hierarchical) grouping process ([5], p. 52, p. 91).

To date, no perceptual grouping mechanism has been implemented at the full primal sketch, yet. However, an oversimplistic solution to texture segmentation has been implemented as a multi-scale texture binary profile, inspired to a single-scale binary texture detector proposed by Nagao and Matsuyama in their seminal work on aerial image classification [140]. In their original single-scale binary texture detection criterion, Nagao and Matsuyama start from a binary contour map. The criterion is [140] (p. 116):

Move an  $N \times N$  window over the picture of binary region boundaries. If the window contains more than  $2 \times N$  boundary pixels, mark the central point of the window as high texture.

Noteworthy, if the local window is  $N \times N$  pixels in size, then the window diagonal is  $\sqrt{2} \cdot N$  pixels in length. If an 8-adjacency cross-aura contour map is available, a one-pixel width diagonal is at least 3-pixel wide. Hence,  $3 \cdot 1.41 \cdot N = 4.23 \cdot N$ . In this hypothesis, the aforementioned criterion becomes the following.

Move an  $N \times N$  window over the picture of binary region boundaries extracted from an 8-adjacency cross-aura contour map. If the window contains more than  $4.5 \times N$  8-adjacency binary contour-pixels, mark the central point of the window as high texture.

A multi-scale texture binary profile is generated at three dyadic scales.

- Local window, Scale 0, 3 pixels =  $\pm 3\sigma_0 = 6\sigma_0, 2^0 \cdot \sigma_0 = 0.5$  pixels  $\Rightarrow 3 \times 3$  pixels in size.





- Local window, Scale 1,  $2^1 \cdot \sigma_0 = 1.0$  pixels,  $6\sigma_1 = 6$  pixels  $\Rightarrow 7 \times 7$  pixels in size.
- Local window, Scale 2,  $2^2 \cdot \sigma_0 = 2$  pixels,  $6\sigma_2 = 12$  pixels  $\Rightarrow 13 \times 13$  pixels in size.

A fast implementation of the moving window statistics is accomplished in agreement with [141].

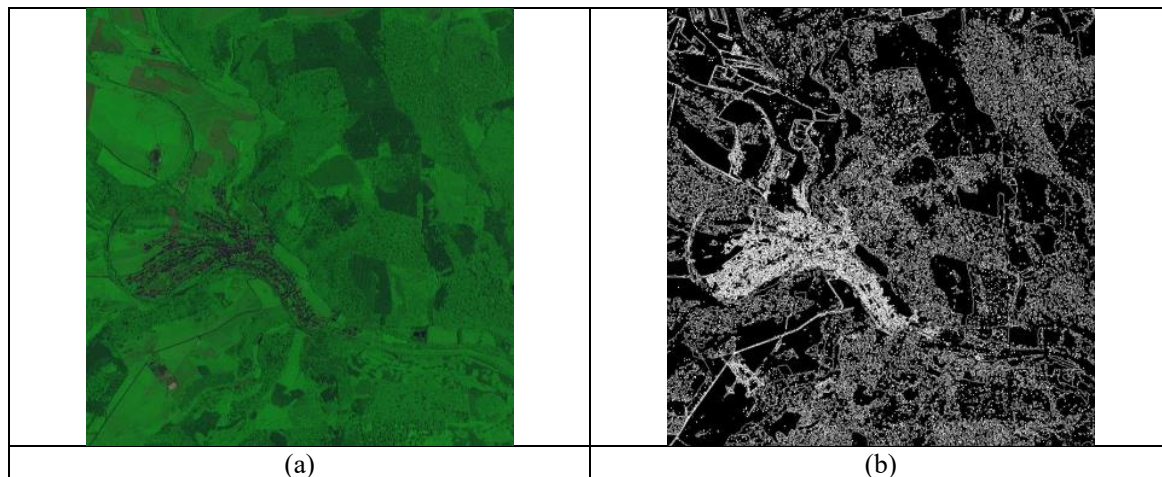
The multi-scale texture binary profile is implemented as follows.

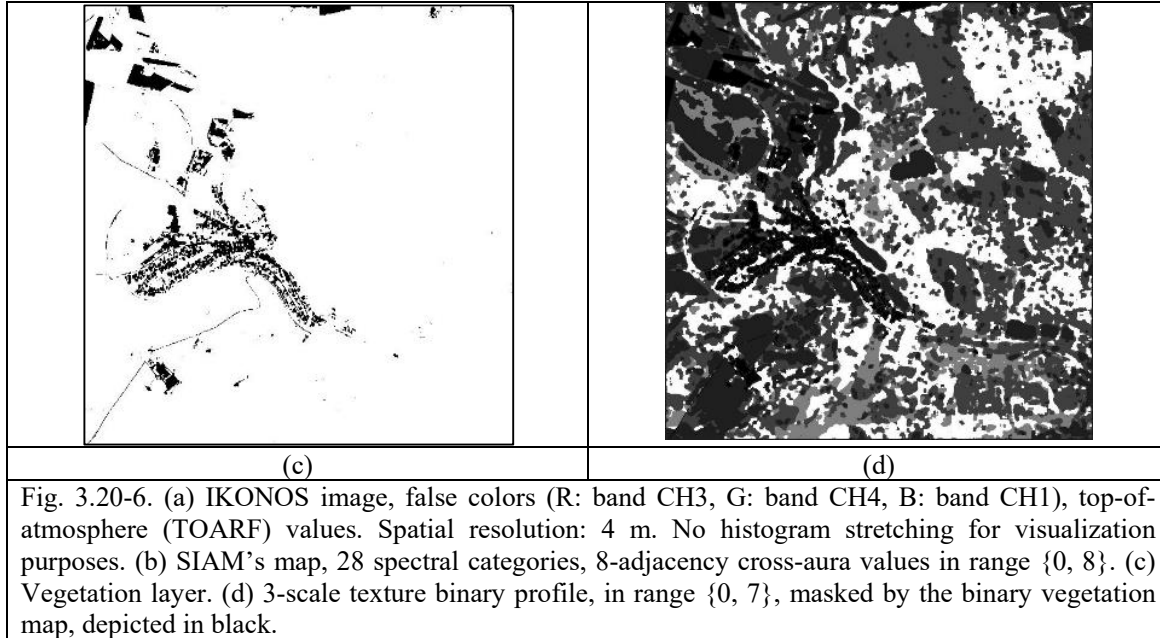
Case ID	Binary High = 1 / Low = 0 texture Scale 2 (coarsest), weight $2^0$	Binary High = 1 / Low = 0 texture Scale 1, weight $2^1$	Binary High = 1 / Low = 0 texture Scale 0 (finest), weight $2^2$
0	0	0	0
1	1	0	0
2	0	1	0
3	1	1	0
4	0	0	1
5	1	0	1
6	0	1	1
7	1	1	1

For example, the multi-scale texture binary profile is mapped onto two fuzzy sets, High and Low, if the texture ID is  $\geq 3$  and  $< 3$ , respectively.

An instantiation of a stratified (masked) multi-scale texture binary profile is shown in Fig. 3.20-6.

In general, an automated texture segmentation algorithm capable of a full primal sketch in the Marr sense remains an open problem to date.





### 3.21 Original conceptual unifying framework for spatial variance, spatial autocorrelation and the proposed 2D wavelet filter bank

In statistics it is common to assume that the variable  $Z(x)$  is stationary, i.e., its distribution is invariant under translation, therefore, for any increment  $h$ , the distribution of  $Z(x_1), Z(x_2), \dots, Z(x_k)$  is the same as that of  $Z(x_1+h), Z(x_2+h), \dots, Z(x_k+h)$ , where  $h$  is the scalar lag (displacement). In its strictest sense stationarity requires all the moments to be invariant under translation, but under the hypothesis of "weak" or second order stationarity only the first two moments (the mean and the covariance) are required to be constant. That is, it is required that, firstly:

$$E[Z(x)] = m(x) = m = const, \quad (21-1)$$

where  $m$  is the mean, and, secondly,

$$C(h) = E[(Z(x+h) - m)(Z(x) - m)] = \tau(h) - m^2, \quad (21-2)$$

where  $C(h)$  is the spatial covariance, such that  $C(0) = \sigma^2$  (variance) and  $\tau(h)$  is the spatial autocorrelation

$$\tau(h) = E[Z(x+h)Z(x)], \quad (21-3)$$

where  $\tau(h) \geq 0$ , in particular  $\tau(h)$  belongs to range { lower bound =  $E[Z(x)]^2$  when  $h \rightarrow \infty$ , such that  $Z(x+h)$  and  $Z(x)$  tend to be independent, therefore  $E[Z(x+h)Z(x)] = E[Z(x+h)]E[Z(x)] = E[Z(x)]^2$ , upper bound =  $E[Z(x)^2]$  when  $h \rightarrow 0$  }, i.e.,  $\tau(h) \in \{ E[Z(x)]^2, E[Z(x)^2] \}$ .

Hence, the spatial covariance  $C(h)$  increases (respectively, decreases) monotonically with spatial autocorrelation  $\tau(h)$ . The spatial autocorrelation  $\tau(h)$  is such that:

- $\tau(h) = \tau(-h)$ , i.e., it is even symmetric.
- $\tau(0) = E[Z(x)^2] \geq \tau(h), \forall h$ , where

$$E[Z(x)^2] = \sigma^2 + mean^2 \quad (21-4)$$





- When  $h$  increases to the point that  $Z(x)$  and  $Z(x+h)$  are statistically independent, then

$$\tau(h \rightarrow \infty) = E[Z(x+h)Z(x)] = E[Z(x+h)]E[Z(x)] = E[Z(x)]^2 = mean^2,$$

i.e.,  $\tau(h)$  decreases from  $\tau(0) = E[Z(x)^2]$  to  $\tau(h \rightarrow \infty) = E[Z(x)]^2$ .

In geostatistics [9], the semivariogram (spacial semivariance, where “semi” means: divided by a factor of 2) measures the inverse degree of spatial correlation between different pixels in an image according to the expression:

$$\begin{aligned} \gamma(h) &= \frac{1}{2} Var[Z(x+h) - Z(h)] = \\ &= \frac{1}{2} E\{[(Z(x+h) - Z(h)) - E[Z(x+h) - Z(h)]]^2\} = \\ &= \frac{1}{2} E\{[Z(x+h) - Z(h)]^2\} \geq 0. \end{aligned} \tag{21-5}$$

Hence,

$$\begin{aligned} \gamma(h) &= \frac{1}{2} E\{[Z(x+h) - Z(h)]^2\} = 2C(0) - 2C(h) = 2\sigma^2 - 2C(h) = \\ &= 2\sigma^2 - 2\tau(h) + 2m^2 = \\ &= 2E[Z(x)^2] - 2\tau(h) \geq 0, \end{aligned} \tag{21-6}$$

where image-wide  $E[Z(x)^2] = \text{image-wide var} + m^2 = \sigma^2 + m^2$ , such that  $\gamma(h) \geq 0$  and  $\gamma(0) = 2E[Z(x)^2] - 2\tau(0) = 2E[Z(x)^2] - 2E[Z(x)^2] = 0$ .

To recapitulate,

$$\gamma(h) = \frac{1}{2} E\{[Contrast(h)]^2\} = 2\sigma^2 - 2C(h) = 2E[Z(x)^2] - 2\tau(h), \tag{21-7}$$

where  $\tau(h)$  is the spatial autocorrelation. Based on Eq. (21-7), it is straightforward to conclude that, if the spatial semivariance  $\gamma(h)$  increases (respectively, decreases), then spatial covariance  $C(h)$  and spatial autocorrelation  $\tau(h)$  decrease (respectively, increase), in particular spatial autocorrelation  $\tau(h)$  decreases from  $E[Z(x)^2]$  to  $E[Z(x)]^2$ , see Fig. 3.21-1.

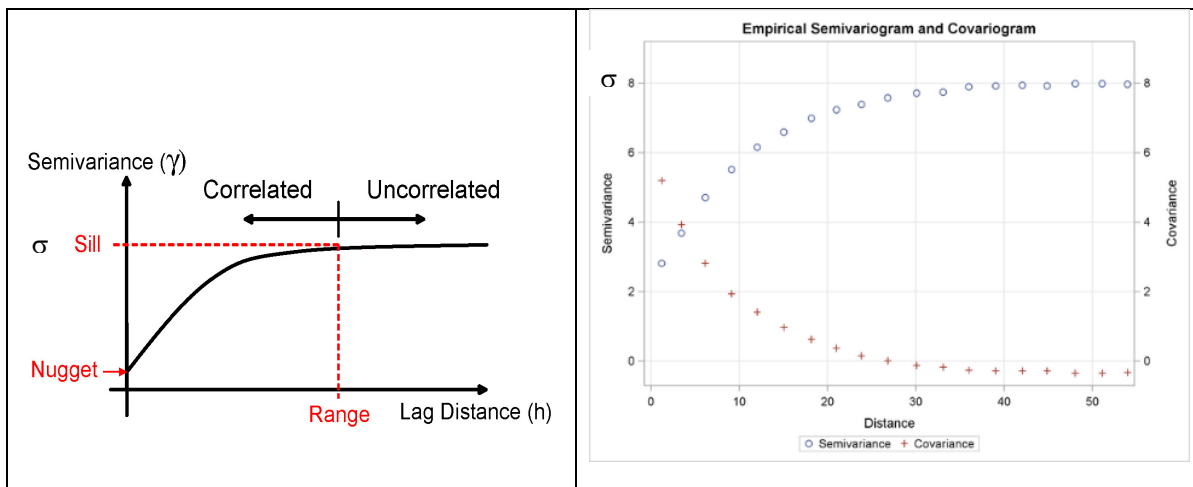


Fig. 3.21-1. In the "weak" stationarity hypothesis of a random variable  $Z(x)$ , whereas the variogram starts from zero and rises up to a limit  $\leq \sigma^2$ , the spatial covariance (or the spatial autocorrelation) starts out from the variance (or the mean squared) and decreases.

In common practice, where a finite set of samples is available, then Eq. (21-7) becomes:

$$\gamma(h) = \sum_{i=1}^N \frac{[Z(x_i + h) - Z(x_i)]^2}{2N} \geq 0. \quad (21-8)$$

It is possible to identify a one-to-one relationship between the semivariogram  $\gamma(h)$ , see Eq. (21-8), and the Geary's C spatial autocorrelation function, defined as follows.

$$\text{Geary's } C = \frac{(N-1) \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} (X_i - X_j)^2}{2 \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} \sum_{i=1}^N (X_i - \bar{X})^2} = \frac{(N-1) \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} (X_i - X_j)^2}{2W \sum_{i=1}^N (X_i - \bar{X})^2} \in [0, 2], \quad (21-9)$$

where  $N$  is the number of *spatial units* indexed by  $i$  and  $j$ ,  $X$  is the variable of interest,  $\bar{X}$  is the mean of  $X$ ,  $w_{i,j}$  is a matrix of spatial weights and  $W$  is the sum of all spatial weights  $w_{i,j}$ . The value of Geary's  $C$  lies between 0 and 2, such that 1 means no spatial autocorrelation. Values lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values higher than 1 illustrate increasing negative spatial autocorrelation. The one-to-one relationship between the semivariogram  $\gamma(h)$ , Eq. (21-8), and the Geary's  $C$  spatial autocorrelation, Eq. (21-9), reveals that the latter should be rather called Geary's  $C$  semi-variogram.

Noteworthy, Geary's  $C$  is inversely related to the Moran's  $I$  spatial autocorrelation, but it is not identical. Moran's  $I$  is a measure of global spatial autocorrelation, while Geary's  $C$  is more sensitive to local spatial autocorrelation. A Reversed Geary's  $C$  is:

$$\text{Reversed Geary's } C \text{ spatial autocorrelation} = \text{RvrsdGearyC} = 1 - C \in [-1, 1], \quad (21-10)$$

such that  $\text{RvrsdGearyC}$  is equal to 1 if there is maximum positive spatial autocorrelation and -1 if there is maximum negative spatial autocorrelation, in agreement with Moran's  $I$ , whose equation is defined as follows.

$$\text{Moran's } I = \frac{N \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} \sum_{i=1}^N (X_i - \bar{X})^2} = \frac{N \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{2W \sum_{i=1}^N (X_i - \bar{X})^2} \in [-1, 1], \quad (21-11)$$

Noteworthy, Moran's  $I$  is one-to-one related to the spatial co-variance  $C(h)$ , see Eq. (21-2), and, as a consequence, to the spatial autocorrelation  $\tau(h)$ , see Eq. (21-3).

In addition to their global (image-wide) expressions, Moran's  $I$  and Geary's  $C$  indexes can be formulated as local (spatial unit-specific) indicators of spatial association (LISA). There is one LISA estimate for each  $i$ -th spatial unit in the dataset and each spatial lag. For example:

$$\text{Local Geary's } C_i = \frac{1}{2 \sum_{j=1, j \neq i}^N w_{i,j} \sum_{i=1}^N (X_i - \bar{X})^2} \sum_{j=1, j \neq i}^N w_{i,j} (X_i - X_j)^2, i = 1, \dots, N. \quad (21-12)$$



$$\text{Local Moran's } I_i = \frac{1}{\sum_{j=1, j \neq i}^N w_{i,j}} \frac{\sum_{j=1, j \neq i}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}, i = 1, \dots, N. \quad (21-13)$$

Therefore, the sum of LISAs for all spatial units in a given dataset area is proportional to a corresponding global indicator of spatial association for that dataset. By “decomposing” a global autocorrelation result into its local parts, LISAs are very useful to uncover hidden, local patterns in data that the global statistics average over. For example, LISAs can detect when: (i) a significant global autocorrelation statistic at a given spatial lag may hide large spatial patches of no autocorrelation, (ii) an insignificant global autocorrelation statistic may hide patches of autocorrelation.

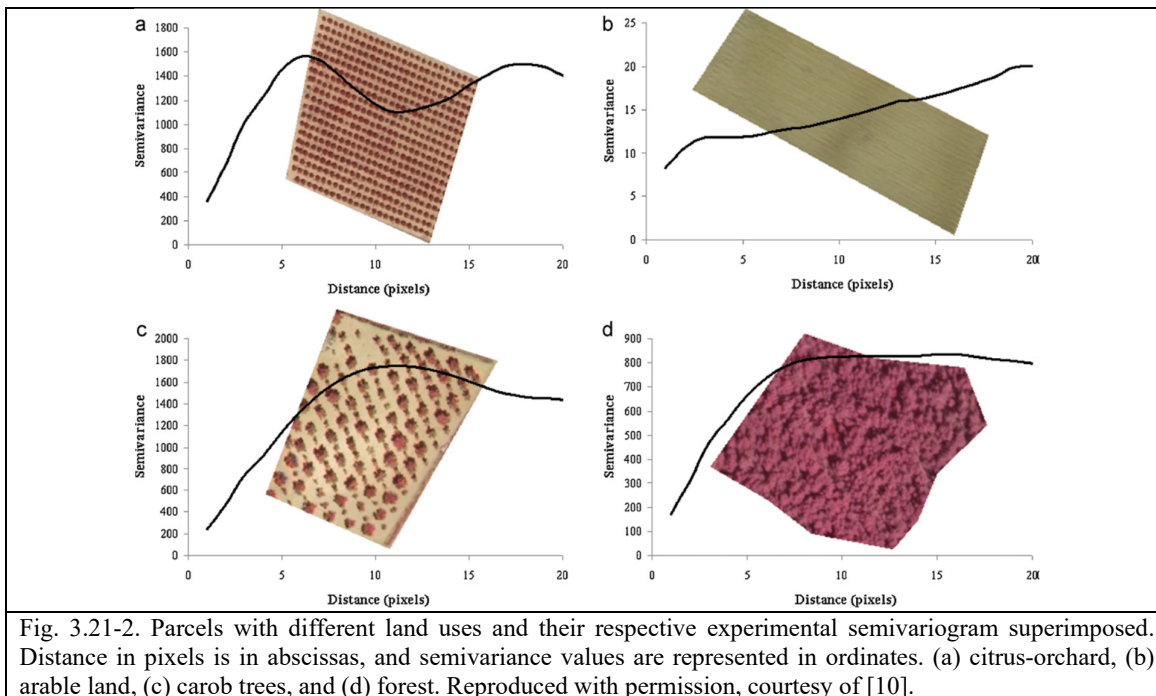
To recapitulate, if the semivariogram  $\gamma(h)$  increases, see Eq. (21-8), , then the Geary’s C, Eq. (21-9), and the Local Geary’s C, Eq. (21-12), increase while spatial covariance  $C(h)$ , Eq. (21-2), spatial autocorrelation  $\tau(h)$ , Eq. (21-3), the Reversed Geary’s C, Eq. (a01a), the Moran’s I, Eq. (21-11), and the Local Moran’s I, Eq. (21-13), decrease, see Fig. 3.21-1, and vice versa.

The first conclusion is that the statistical concept of spatial autocorrelation, obscure to many geographers, is the quantitative counterpart of the qualitative Tobler’s First Law (TFL) of geography, familiar to any geographer. The TFL of geography states that “all things are related, but nearby things are more related than distant things” [50], although certain phenomena clearly constitute exceptions [51].

The second conclusion is that multi-scale multi-orientation filter banks, implemented as the 2D three-pole even-symmetric Gabor filter and the 2D dipole odd-symmetric Gabor filter shown in Fig. 3.5-2, capable of estimating local contrast (first-order derivative), local concavity (second-order derivative) and change in sign of the local concavity (ZX pixel detection) for image contour detection and scale-invariant keypoint extraction, are not only related to the (E)TAU theorem proposed by Yellott, stating that two images are perceptually identical when they share the same image-wide combinations of local statistics, up to third-order statistics (refer to Chapter 3.2.4), but they are also linked by:

- an analytic monotonically increasing relationship to the semivariogram  $\gamma(h)$ , Eq. (21-8), the Geary’s C, Eq. (21-9), and the Local Geary’s C, Eq. (21-12), where local contrast is estimated.
- An analytic monotonically decreasing relationship to the spatial covariance  $C(h)$ , Eq. (21-2), spatial autocorrelation  $\tau(h)$ , Eq. (21-3), the Reversed Geary’s C, Eq. (21-10), the Moran’s I, Eq. (21-11), and the Local Moran’s I, Eq. (21-13).

An *omnidirectional semivariogram* can be obtained by averaging the semivariograms of all possible directions of a given lag distance (inter-point distance)  $h$ . The semivariogram has been widely employed in digital image processing. Its usefulness in remote sensing has been demonstrated, complementing the spectral variables with information related to the spatial structure of the image. In homogeneous objects, semivariance values tend to be higher as the lag increases. However, when the elements inside an image are spatially arranged following a regular pattern, the semivariogram has a cyclic behaviour, and it is known as *hole-effect semivariogram*, otherwise, when an image does not present regular patterns or spatial cyclicity, then the semivariogram curve follows a monotonous rising trend, see Fig. 3.21-2.



Omnidirectional semivariogram	Wavelet filter bank
Contrast(h)	2ndOrdrDerivEvenSymWaveletOutput value = convolution of the filter with the image, where the filter is even-symmetric and zero-dc-component, like $\nabla^2 G$ in [5]. This second-order derivative filter's output is $\neq 0$ if there is a change in the first-order derivative of the image.
Contrast(h) <sup>2</sup>	Pixel-based wavelet energy signature(x,s) = $\sum_{orn=1,Orn} [2ndOrdrDerivEvenSymWaveletOutput\ value\ _s, orn(x)]^2$ , scale s = 1, ..., S.
Lag distance h	Size (scale) of the filter, s = 1, ..., S.
Orientation of the lag distance	Orientation of the filter, orn = 1, ..., Orn.
Omnidirectional semivariogram $\gamma(h) = \frac{1}{2} E[\{Contrast(h)\}^2]$	Image-wide wavelet energy signature(s) = $\sum_{x=1,X} Wavelet\ energy\ signature(x,s)$ , scale s = 1, ..., S.

Table 3.21-1. Relations between an omnidirectional semivariogram and a wavelet filter bank.

It is possible to highlight the following relations between an omnidirectional semivariogram, defined as the mean contrast,  $E[Contrast(h)]$ , computed by averaging the semivariograms of all possible directions of a given lag distance h, and a multi-scale multi-orientation even-symmetric zero-dc-component filter bank where each filter computes the change of the intensity change (second-order derivative of the intensity) across the filter domain of activation, see Table 3.21-1. To recapitulate, by analogy with the omnidirectional semivariogram, it is possible to generate a pixel-based omnidirectional Wavelet energy signature(x,s) =  $\sum_{orn=1,Orn} [2ndOrdrDerivEvenSymWaveletOutput\ _s, orn(x)]^2$ , scale s = 1, ..., S.

In the following, we propose a novel local spatial autocorrelation descriptor (LSAD) computed by means of the even-symmetric local filter shown in Fig. 3.5-2. It means that a multi-scale multi-orientation bank of even-symmetric local filters, such as that shown in Fig. 3.5-2, is:

- Necessary and sufficient to detect image contours, refer to Chapter 3.11.
- Necessary to provide an image reconstruction,  $Rcnstrct(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G / \partial x^2] / 2$ , refer to Chapter 3.6.



- Necessary to estimate a “perceptual contrast” (PrcptlCntrst2),  $\text{PrcptlCntrst2}(x) = \text{abs}[f(x) \circ \partial^2 G/\partial x^2] + \text{abs}[f(x) \circ \partial G/\partial x]/2 = \text{abs}(\text{EvnSym}) + [\text{abs}(\text{OddSym})/2]$ , hence  $\text{PrcptlCntrst2}(x) \geq 0$ , related to the detection of scale-invariant keypoints, refer to Chapter 3.6.
- Necessary and sufficient to work as an LSAD.

The LSAD design problem is constrained by the following project requirements, provided with a strong statistical ground and a physical meaning. Therefore, these project’s constraints are easy (intuitive) to understand, i.e., they are perceptually significant.

1. It employs the same machinery consisting of the even-symmetric local filter shown in Fig. 3.5-2.
2. In a b/w chessboard, each spatial unit (e.g., a black parcel) with respect to its 4-adjacency neighbors (e.g., four white parcels) features:
  - (i) Spatial autocorrelation = -1.
  - (ii) Contrast = Maximum.
3. In a ramp edge, related to the Mach bands illusion:
  - (i) Spatial autocorrelation = 1.
  - (ii) Contrast > 0.
4. In a flat area, also related to the Mach bands illusion:
  - (i) Spatial autocorrelation = 1.
  - (ii) Contrast = 0.

### 3.22 Conclusions and open challenges

To our best knowledge, if we require that an EO-IUS or CV system should score “high” in matching the original set of requirements specified for an innovative computational model of human vision (Chapter 3.6), the number of published models becomes surprisingly small, eventually empty, in both the CV and the RS literature.

The original set of requirements specified for an innovative computational model of human vision (Chapter 3.6) can be considered fulfilled to a large degree by the proposed EO-IUS / CV system design and implementation as reported below (see Table 3.23-1).

1. **Mandatory image enhancement (pre-processing) for data harmonization across time, space and sensors.** Fulfilled. Radiometric calibration of digital numbers into TOARF or SURF values, with  $\text{TOARF} \supseteq \text{SURF}$ , is considered mandatory when a radiometric calibration metadata file is available (Chapter 3.10.1). Otherwise, apply an original self-organizing statistical model-based algorithm for color constancy (Chapter 3.9, Chapter 3.10.2).
2. **Complex system = distributed processing system = artificial neural network (ANN) paradigm.** Fulfilled by a bank of several families of local spatial filters, specifically, even-symmetric filters, odd-symmetric filters and end-stopped cells (for keypoint detection, as a non-linear spatial combination of even- and odd-symmetric filter outputs). Neighboring cells interact through lateral connections, e.g., to detect zero-crossing (ZX) pixels as positions in the (2D) image-domain where a change in sign of the local concavity of the image surface occurs (Chapter 3.11, Chapter 3.12, Chapter 3.15).
3. **ANN capable of topology-preserving feature mapping.** Accomplished in an innovative topology-preserving spatial filter bank, capable of automated low-level vision tasks such as ZX image-contour detection (Chapter 3.12, Chapter 3.15), ZX image-segment detection (Chapter 3.20.2) and keypoint detection (Chapter 3.12, Chapter 3.15) in both panchromatic and color images.
4. **Hybrid (combined deductive/top-down/physical model-based and inductive/bottom-up/statistical model-based) inference.** Accomplished in a novel “complete” 6-stage hybrid feedback EO-IUS, employing a deductive pre-classification first stage for color naming and a convergence-of-evidence approach, which is adopted in a novel EO-IU4SQ system capable of SCBIR in multi-source large-scale EO image databases, see Fig. 3.6.2.
5. **Feedback loops.** Accomplished in a novel “complete” 6-stage hybrid feedback EO-IUS, see Fig. 3.6.2.
6. **Hierarchical submodular approach to form neural modularity systems, equivalent to a network of sub-networks according to a divide-and-conquer problem solving approach.** Accomplished in a novel “complete” 6-



stage hybrid feedback EO-IUS, see Fig. 3.6.2. For example, refer to automated ZX image-contour detection (raw primal sketch) in series with texture segmentation (full primal sketch) in low-level vision (Chapter 3.20.4).

7. **Capability to fully exploit spatial topological and spatial non-topological information components in the (2D) image-domain and in the 4D spatiotemporal scene-domain.** A necessary and sufficient condition for a CV system to fully exploit spatial topological and spatial non-topological information components in addition to color is to perform nearly as well when input with panchromatic or color imagery. Accomplished by the proposed spatial filter bank, able to scale seamlessly from color to panchromatic images and vice versa (Chapter 3.20).
8. **Foveated vision.** Promoted at the level of understanding of an innovative CV system design, see Fig. 3.6-8 and Fig. 3.6-9.

Legend of fuzzy sets of a quantitative variable.	
	<div style="display: inline-block; width: 15px; height: 10px; background-color: red; margin-right: 5px;"></div> LOW <div style="display: inline-block; width: 15px; height: 10px; background-color: blue; margin-right: 5px;"></div> MEDIUM <div style="display: inline-block; width: 15px; height: 10px; background-color: green; margin-right: 5px;"></div> HIGH
<b>Computer vision (CV) Process (Pracs) and Outcome (Otcn) Q<sup>2</sup>Is ± δ ⊆ QA4EO Val</b>	<b>Low-level vision: raw/full primal sketch</b>
<b>Degree of automation (Pracs):</b> (a) inversely related to the number, physical meaning and range of variation of user-defined parameters, (b) inversely related to the collection of the required training data set, if any.	<b>HIGH (unsurpassed, no free-paramtr)</b>
<b>Effectiveness (Otcn), in agreement with human visual perception, i.e., CV ⊃ human vision, where human visual perception is a lower bound of CV.</b>	
a) Color constancy or radiometric calibration (when radiometric calibration metadata are available)	a) HIGH
b) 2D image analysis/ Retinotopic visual information representation/ Topology-preserving visual feature mapping / Spatial topological information extraction. ➤ Necessary not sufficient condition: panchromatic vision performs nearly as well as chromatic vision.	b) YES  ➤ HIGH
c) Pre-attentive image contour detection/ image segmentation quality, consistent with the Mach bands illusion in ramp-edge detection: spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	c) HIGH
d) High-level vision (classification). (a) thematic Qis (TQIs) and (b) spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	d) None
<b>Semantic information level (Pracs)</b>	<b>LOW (sub-symbolic)</b>
<b>Efficiency (Pracs):</b> (a) computation complexity in image size: Polynomial (P), P and linear (L), non-P (NP), and (b) run-time memory occupation.	(a) HIGH (L), (b) HIGH
<b>Robustness to changes in input image (Pracs)</b> , e.g., large spatial extent data mapping (no toy problems).	HIGH
<b>Robustness to changes in input parameters (Pracs)</b> , e.g., sensitivity analysis.	HIGH (unsurpassed, no free-paramtr)
<b>Scalability to changes in the sensor's specifications or user's needs (Pracs)</b> , e.g., (a) pncchrmtc, (b) RGB, true- or false-color, (c) multi-spectral (MS), (d) super-spectral (SS), (e) hyper-spectral (HS).	HIGH, (a) YES, (b) YES, (c) YES, (d) YES, (e) YES.
<b>(Inverse of) Timeliness (Otcn)</b> , from data acquisition to high-level product generation, increases with manpower and computing power.	HIGH
<b>(Inverse of) Costs (Otcn)</b> , increasing with (a) manpower and (b) computing power.	(a) HIGH, (b) HIGH

Table 3.23-1. Outcome and process (OP) quantitative quality indicators (OP-Q<sup>2</sup>Is) of the proposed low-level CV system design and implementation.

9. **Consistency with visual illusions, starting from the Mach bands illusion, where a bright and a dark band are seen at ramp edges.** Accomplished by the proposed spatial filter bank (Chapter 3.12, Chapter 3.20).
10. **Multi-source image scalability to varying imaging sensor specifications**, i.e., capability to process multi-source imagery, including: (i) Uncalibrated panchromatic and RGB images. (ii) Radiometrically calibrated multi-spectral (MS), Super-spectral (SS) and Hyper-spectral (HS) images, whose number of spectral channels N is {2, 9}, {10, 20} and > 20 respectively. (iii) Bi-temporal Red-Green-Blue (RGB) synthetic aperture radar (SAR) images [168]. Accomplished by the novel “complete” 6-stage hybrid feedback EO-IUS, see Fig. 3.6.2, where zero-stage image harmonization across time, space and sensors is followed by a first-stage application- and user-independent original expert system for automated color naming in a calibrated MS reflectance space or in an uncalibrated RGB color space, either true- or false-color (Chapter 3.10).
11. **Convergence-of-evidence approach.** Accomplished by the novel “complete” 6-stage hybrid feedback EO-IUS (Chapter 3.6).
12. **Operating mode.** To be considered in operating mode, a CV system is required to score “high” in a minimally dependent maximally informative (mDMI) set [123], [124] of outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is), to be community-agreed upon in advance in agreement with the QA4EO Validation (Val) guidelines [111]. The proposed mDMI set of OP-Q<sup>2</sup>Is includes the following [56], [57], refer to Chapter 3.3. (i) Degree of automation,



inversely related to user-machine interaction. (ii) Accuracy, e.g., mapping accuracy. (iii) Efficiency, in computation time and memory occupation. (iv) Robustness to changes in input data. (v) Robustness to changes in input parameters, if any. (vi) Scalability to changes in sensor and user specifications. (vii) Timeliness, defined as the time interval between data acquisition and product generation. (ix) Costs in manpower and computer power. Accomplished by the novel “complete” 6-stage hybrid feedback EO-IUS (Chapter 3.6) which is, up to the full primal sketch in low-level vision (Chapter 3.20.4), multi-source, accurate, automated, robust and efficient with computational complexity increasing linearly with the image size.

A detailed description of the original contributions of the present R&D study is provided below.

- Chapter 3.5.4. Original relationship between the Yellot’s Enhanced Triple Autocorrelation Uniqueness (ETAU) principle and the proposed 2D even- and odd-symmetric spatial filter bank, shown in Fig. 3.5-2.
- Chapter 3.6. Original CV system’s design and implementation requirements specification, to pursue a computational model of human vision, including a novel foveated imaging system design.
- Chapter 3.9. Original automated statistical model-based color constancy algorithm for non-calibrated color/panchromatic image harmonization.
- Chapter 3.10. Original expert systems for automated color naming in a calibrated MS reflectance space or in an uncalibrated RGB color space, either true- or false-color: the Satellite Image Automatic Mapper™ (SIAM™) and RGB Image Automatic Mapper (RGBIAM™) lightweight computer programs for static color naming, superpixel detection and vector quantization (VQ) quality assessment.
- Chapter 3.11. Original 1D simulations for image analysis and synthesis, including the zero-frequency signal component, image-contour detection and keypoint detection consistent with the Mach bands illusion, see Fig. 3.11-1.
- Chapter 3.12. Original operational definition of zero-crossing (ZX) pixels for automated (parameter-free) image-contour detection and 2D scale-invariant keypoint (corners, endpoints, T-junctions and X-junctions) detection consistent with the Mach bands illusion.
- Chapter 3.13. Original perceptual image-pair dissimilarity metric.
- Chapter 3.15 to Chapter 3.19. Original design and implementation of a 2D stratified (masked) 4-scale 2-orientation even- and odd-symmetric Gabor filter bank, shown in Fig. 3.5-2, considered necessary and sufficient to accomplish an automated (parameter-free) near-orthogonal image analysis/synthesis, ZX pixel detection and keypoint (corners, endpoints, T-junctions and X-junctions) detection in compliance with the Mach bands illusion. This is a novel automated image processing system framework capable of unifying the ZX pixel detection accomplished by isotropic multi-scale Laplacian-of-a-Gaussian operators as proposed by Marr [5] with the scale-invariant keypoint detection, proposed by David Lowe [11] as extrema of the multi-scale difference of Gaussian (DOG), [12], known as Scale Invariant Feature Transform (SIFT).
- Chapter 3.20.2. Original automated detection of ZX segments from ZX pixels at the raw primal sketch, These ZX segments agree with the Marr quote [5] (p.67): "ZXs provide a natural way of moving from an analogue or continuous representation like the two-dimensional image intensity values to a discrete representation (into discrete tokens, namely, blobs, edges, bars and termination points through so-called ZX segments [5], p. 60). A fascinating thing about this transformation is that it probably incurs no loss of information... (in practice), a one-octave band-pass signal can be completely reconstructed (up to an overall multiplicative constant) from its ZXs". The proposed ZX segment detector successfully finalizes a raw primal sketch in full compliance with the project requirements specification proposed in Chapter 3.4.
  - It is automated, i.e., it requires no user-machine interaction to run. Since it is physical model-based, it requires no system’s free-parameter to be user-defined based on heuristics.
  - It complies with the Mach bands illusion. If we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published models is surprisingly small, refer to [25].
  - It down-scales seamlessly from color to panchromatic images, which means it thoroughly exploits spatial topological and spatial non-topological information which typically dominate color information in both the image-domain and the scene-domain.



- To the best of this author's knowledge, no automated computation of ZX segments defined by Marr has ever been presented in the existing literature to date.
- Chapter 3.20.4. Original multi-scale multi-scale texture binary profile.
- Chapter 3.21. Novel conceptual unifying framework between spatial variance, spatial autocorrelation and the proposed 2D wavelet filter bank.

The following R&D open problems are intended to be further investigated.

- ✓ What is the rationale capable of combining with profit, e.g., for keypoint detection purposes, the odd-symmetric imaginary part of a Gabor filter, equivalent to a first-order derivative of a Gaussian filter, with the even-symmetric real part of a Gabor filter, equivalent to a second-order derivative of a Gaussian filter? An experimental comparison with existing computational models of end-stopped cells responding to singularities (line/edge crossings, vertices, end points), such as those proposed in [28], [43], [66], is highly recommended.
- ✓ Augment the computational efficiency of Gabor filters, to be accomplished by separable filter design and implementation principles proposed in [20], [33], [35], [92].
- ✓ No perceptual grouping of texels (texture segmentation) in compliance with the Marr's full primal sketch has been implemented, yet, refer to Chapter 3.16.2. To date, no automated texture segmentation algorithm has ever been proposed by the CV or remote sensing community.
- ✓ Implementation and testing of the novel foveated imaging system sketched in Fig. 3.6-8 and Fig. 3.6-9.



### References in Chapter 3

- [1] A. Baraldi and F. Parmiggiani, "Combined detection of intensity and chromatic contours in color images," *Optical Engineering*, vol. 35, no. 5, pp. 1413-1439, May 1996.
- [2] J. I. Yellott, "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," *Optical Society of America*, vol. 10, no. 5, pp. 777-793, May 1993.
- [3] J. Victor, "Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: Comment," *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1680-1684, May 1994.
- [4] G. Van de Wouwer, P. Scheunders, and D. Van Dyck, "Statistical texture characterization from discrete wavelet representation," *IEEE Trans. Image Processing*, vol. 8, no. 4, pp. 592-598, April 1999.
- [5] D. Marr, *Vision*. New York: Freeman and C., 1982.
- [6] Li Zhaoping, *Understanding Vision: Theory, Models, and Data*. University College London, UK, 2012.
- [7] Perceptual Grouping, Purdue University, [Online]. Available: <http://cs.iupui.edu/~tuceryan/research/ComputerVision/perceptual-grouping.html>
- [8] S. P. Vecera and M. J. Farah, "Is visual image segmentation a bottom-up or an interactive process?," *Perception & Psychophysics*, vol. 59, pp. 1280-1296, 1997.
- [9] M. Armstrong, *Basic Linear Geostatistics*, Springer: Berlin, 1998.
- [10] L.A. Ruiz, J.A. Recio, A. Fernández-Sarría, and T. Hermosilla, "A feature extraction software tool for agricultural object-based image analysis," *Computers and Electronics in Agriculture*, vol. 76, pp. 284-296, 2011.
- [11] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer: Dordrecht, The Netherlands, 1994.
- [12] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [13] C. Mason and E. R. Kandel, "Central Visual Pathways," in *Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Norwalk, CT, USA: Appleton and Lange, pp. 420-439, 1991.
- [14] P. Gouras, "Color Vision," in *Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Norwalk, CT, USA: Appleton and Lange, pp. 467-479, 1991.
- [15] E. R. Kandel, "Perception of Motion, Depth and Form," in *Principles of Neural Science*; Kandel, E., Schwartz, J., Eds.; Appleton and Lange: Norwalk, CT, USA; pp. 441-466, 1991.
- [16] H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vision Res.*, vol. 19, pp. 19-32, 1979.
- [17] D. Hubel and T. Wiesel, "Receptive fields of single neurons in the cat's striate cortex," *Journal of Physiology*, vol. 148, pp. 574-591, 1959.
- [18] J. Mutch, and D. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *Int J. Comput. Vis.*, vol. 80, pp. 45-57, 2008.
- [19] T. N. Wiesel and D. H. Hubel, "Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey," *J. Neurophys.*, vol. 29, pp. 1115-1156, 1966.
- [20] A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Trans. Image Proc.*, vol. 7, no. 1, pp. 124-128, 1998.
- [21] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [22] T. Gevers, A. Gijzenij, J. van de Weijer, and J-M. Geusebroek, *Color in Computer Vision*, Hoboken, New Jersey, USA: Wiley, 2012.
- [23] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. of Computer Vision*, vol. 30, number 2, 1998.
- [24] H. du Buf and J. Rodrigues, *Image morphology: from perception to rendering*, in *IMAGE - Computational Visualistics and Picture Morphology*, 2007.
- [25] L. Pessoa, "Mach Bands: How Many Models are Possible? Recent Experimental Findings and Modeling Attempts", *Vision Res.*, Vol. 36, No. 19, pp. 3205-3227, 1996.
- [26] D. C. Burr and M. C. Morrone, "A nonlinear model of feature detection," in *Nonlinear Vision: Determination of Neural Receptive Fields, Functions, and Networks*, R. B. Pinter and N. Bahram, Eds., pp. 309-327, CRC Press, Boca Raton, FL, 1992.
- [27] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. and Mach. Intell.* vol. 8, no. 6, pp. 679-698, 1986.



- [28] J. Rodrigues and J.M. Hans du Buf, “Multi-scale lines and edges in V1 and beyond: Brightness, object categorization and recognition, and consciousness,” *BioSystems*, vol. , pp. 1-21, 2008.
- [29] Andrea Baraldi, João V. B. Soares, “Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis,” Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01941. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01941.pdf>
- [30] K. A. Stevens, Computation of Locally Parallel Structure, *Biol. Cybernetics*, vol. 29, pp. 19--28, 1978.
- [31] R. Capurro and B. Hjørland, The concept of information, *Annual Review of Information Science and Technology*, B. Cronin, Ed., Medford, NJ, USA: Information Today Inc., 2003, vol. 37, ch. 8, pp. 343–411. [Online]. Available: <http://www.capurro.de/infoconcept.html>
- [32] R. Capurro, Hermeneutics and the Phenomenon of Information, in *Metaphysics, Epistemology, and Technology. Research in Philosophy and Technology*, vol. 19. Amsterdam, The Netherlands: Elsevier, 2000, pp. 79–85.
- [33] The Gaussian kernel. [Online]. Available: <http://www.stat.wisc.edu/~mchung/teaching/MIA/reading/diffusion.gaussian.kernel.pdf.pdf> (accessed on 15 Sept. 2014).
- [34] C. Axel, “Fusion of Terrestrial LiDAR Point Clouds with Color Imagery”, Senior Project, Rochester Institute of Technology, 16 May 2013. [Online]. Available: <http://www.cis.rit.edu/DocumentLibrary/admin/uploads/CIS000202.PDF> (accessed on 15 Sept. 2014).
- [35] J. Kranauskas, “Accelerated Calculation of Gabor Features in Spatial Domain”, *ELECTRONICS AND ELECTRICAL ENGINEERING*, No. 1(97), 2010.
- [36] Benavente, R., Vanrell, M. & Baldrich, R. (2008). Parametric fuzzy sets for automatic color naming. *Journal of the Optical Society of America A*, 25, 2582-2593.
- [37] J. van de Weijer, C. Schmid, J. Verbeek, D. Larlus, Learning color names for real-world applications, *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1512 – 1523, 2009.
- [38] A. Baraldi, SIAM Report, 2014. [Online]. Available: [http://siam.andreabaraldi.com/content/Documentation/SIAM\\_Report\\_BACRES\\_v1.17.pdf](http://siam.andreabaraldi.com/content/Documentation/SIAM_Report_BACRES_v1.17.pdf) (accessed on 15 Sept. 2014).
- [39] Moran's I Spatial Autocorrelation Measure. [Online]. Available: [en.wikipedia.org/wiki/Moran's\\_I](http://en.wikipedia.org/wiki/Moran's_I) (accessed on 15 Sept. 2014).
- [40] Geary's C Spatial Autocorrelation Measure. [Online]. Available: [en.wikipedia.org/wiki/Geary%27s\\_C](http://en.wikipedia.org/wiki/Geary%27s_C) (accessed on 15 Sept. 2014).
- [41] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, “Simulation of neural contour mechanisms: from simple to end-stopped cells,” *Vision Res.*, vol. 32, no. 5, pp. 963–981, 1992.
- [42] Adelson, E. H. & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284-299.
- [43] J. Rodrigues and J.M.H. du Buf, “Multi-scale keypoints in V1 and beyond: Object segregation, scale selection, saliency maps and face detection”, *BioSystems*, vol. 86, pp. 75–90, 2006.
- [44] M. Bertero, T. Poggio, and V. Torre, “Ill-posed problems in early vision,” *Proc. IEEE*, vol. 76, pp. 869–889, 1988.
- [45] Matsuyama, T. & Hwang, V.S. (1990). *SIGMA – A Knowledge-based Aerial Image Understanding System*. New York, NY: Plenum Press.
- [46] Baatz, M., Hoffmann, C., & Willhauck, G. (2008). Progressing from object-based to object-oriented image analysis. In Blaschke, T., Lang, S., Hay, G.J. (Eds.), *Object-Based Image Analysis–Spatial Concepts for Knowledge-driven Remote Sensing Applications* (pp. 29-42). New York, NY: Springer-Verlag.
- [47] L. Delves, R. Wilkinson, C. Oliver, and R. White, “Comparing the performance of SAR image segmentation algorithms,” *Int. J. Remote Sens.*, vol. 13, no. 11, pp. 2121–2149, 1992.
- [48] G. J. Hay and G. Castilla, “Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT),” in *Proc. 1st Int. Conf. OBIA*, S. Lang, T. Blaschke, and E. Schöpfer, Eds., 2006. [Online]. Available: [www.commission4.isprs.org/obia06/Papers/01\\_Opening%20Session/OBIA2006\\_Hay\\_Castilla.pdf](http://www.commission4.isprs.org/obia06/Papers/01_Opening%20Session/OBIA2006_Hay_Castilla.pdf)
- [49] Q. Iqbal and J. K. Aggarwal, “Image retrieval via isotropic and anisotropic mappings,” in *Proc. IAPR Workshop Pattern Recognit. Inf. Syst.*, Setubal, Portugal, Jul. 2001, pp. 34–49.
- [50] Tobler, W.R., 1970, A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 46, pp. 234–240.





- [51] P. A. Longley, M. F. Goodchild, D. J. Maguire D. W. Rhind, *Geographic Information Systems and Science*, Second Edition. New York: Wiley, 2005.
- [52] C. Chubb and J. I. Yellott, "Every discrete, finite image is uniquely determined by its (two-dimensional) dipole histogram," *Vision Research*, vol. 40, pp. 485–492, 2000.
- [53] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Proc. Letters*, vol. 9, no. 3, pp. 81-84, March 2002.
- [54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, p. 1-14, 2004.
- [55] Open source computer vision library (OpenCV), [Online]. Available: <http://opencv.org/>. Accessed on March 20, 2015.
- [56] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction," *Remote Sens.*, vol. 4, pp. 2694-2735, 2012.
- [57] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation." *Remote Sens.*, vol. 4, pp. 2768-2817, 2012.
- [58] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall, 1994.
- [59] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception*, vol. 2, pp. 391-405, 1973.
- [60] B. Julesz, "Texton gradients: The texton theory revisited," in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, vol. 54, no. 4-5, Aug. 1986.
- [61] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 1, pp. 1-8, vol. 6, no. 1, 2011.
- [62] A. Baraldi and F. Parmiggiani, "An investigation of textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 293-304, March 1995.
- [63] H. Anys and D. C. He, "Evaluation of textural and multipolarization radar features for crop classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1170-1181, 1995.
- [64] C. Chubb and J. I. Yellott, "Every discrete, finite image is uniquely determined by its dipole histogram," *Vision Research*, vol. 40, pp. 485–492, 2000.
- [65] R. M. Boynton, "Human color perception," in *Science of Vision*, K. N. Leibovic, Ed., pp. 211–253, Springer-Verlag, New York, 1990.
- [66] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, "Simulation of neural contour mechanisms: from simple to end-stopped cells," *Vision Res.*, vol. 32, no. 5, pp. 963–981, 1992.
- [67] The Weber–Fechner sensitivity law, Wikipedia, [https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner\\_law](https://en.wikipedia.org/wiki/Weber%E2%80%93Fechner_law)
- [68] J. K. Tsotsos, "Analyzing vision at the complexity level," *Behavioral and Brain Sciences*, vol. 13, pp. 423-469, 1990.
- [69] T. Martinetz, G. Berkovich, and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
- [70] D. Van Essen and J. Maunsell, "Hierarchical organization and functional streams in the visual cortex," *Trends in Neuroscience*, vol. 6, pp. 370-75, 1983.
- [71] S. Frintrop, "Computational visual attention," in *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds., Springer, 2011.
- [72] P. J. Erichsen and J.M. Woodhouse. *Human and animal vision. Machine Vision Handbook*, 2013.
- [73] Available: [https://en.wikipedia.org/wiki/Foveated\\_imaging](https://en.wikipedia.org/wiki/Foveated_imaging)
- [74] M. Ranzato, On Learning Where To Look, arXiv 2014. [Online]. Available: [http://www.cs.toronto.edu/~ranzato/publications/ranzato\\_arxiv14.pdf](http://www.cs.toronto.edu/~ranzato/publications/ranzato_arxiv14.pdf)
- [75] J. Hadamard, "Sur les problemes aux derivees partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.
- [76] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley, 1998.
- [77] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons, 2004.
- [78] L. D. Griffin, "Optimality of the basic color categories for classification", *J. R. Soc. Interface*, vol. 3, pp. 71–85, 2006.



- [79] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- [80] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
- [81] D. Parisi, "La scienza cognitive tra intelligenza artificiale e vita artificiale," in *Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza*, Bologna, Italy: Patron Editore, 1991.
- [82] S. Kosslyn, *Image and Brain*. MIT Press, Cambridge, MA, 1994.
- [83] S. D. Slotnick, W. L. Thompson and S. M. Kosslyn, *Visual Mental Imagery Induces Retinotopically Organized Activation of Early Visual Areas, Cerebral Cortex* October, vol. 15, pp. 1570-1583, 2005.
- [84] G. A. Miller, "The cognitive revolution: a historical perspective", in *Trends in Cognitive Sciences*, vol. 7, pp. 141-144, 2003.
- [85] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.
- [86] F. Capra and P. L. Luisi, *The Systems View of Life: A Unifying Vision*. Cambridge, UK: Cambridge University Press, 2014.
- [87] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, "Geographic object-based image analysis - Towards a new paradigm," *ISPRS J. Photogram. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.
- [88] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. and Mach. Intell.* vol. 8, no. 2, pp. 147–163, 1986.
- [89] A. L. Yuille and T. Poggio, "Fingerprints theorems for zero-crossings," *IEEE Trans. Pattern Anal. and Mach. Intell.* vol. 8, no. 1, pp. 15–25, 1986.
- [90] S.M. Smith and J.M. Brady. *SUSAN - a new approach to low level image processing*. *Int. Journal of Computer Vision*, 23(1):45--78, May 1997.
- [91] [https://en.wikipedia.org/wiki/Action\\_potential](https://en.wikipedia.org/wiki/Action_potential)
- [92] O. Nestares, R. Navarro, J. Portilla and A. Taberero, "Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions," *Journal of Electronic Imaging*, vol. 7, no. 1, pp. 166–173, January 1998.
- [93] J. Sylvester and J. Reggia, "Engineering neural systems for high-level problem solving," *Neural Networks*, vol. 79, pp. 37–52, 2016.
- [94] J. Feldman, "The neural binding problem(s)," *Cogn. Neurodyn.*, vol. 7, pp. 1-11, 2016.
- [95] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, abs/1411.4038, 2014.
- [96] C. Gatta, A. Romero, and J. van de Weijer, "Unrolling loopy top-down semantic feedback in convolutional deep networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2014, pp. 498-505.
- [97] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–353, 1965.
- [98] D. Tiede, A. Baraldi, M. Sudmanns, M. Belgiu, and S. Lang, "ImageQuerying (IQ) – Earth Observation Image Content Extraction & Querying across Time and Space," submitted (Oral presentation and poster session), *ESA 2016 Conf. on Big Data From Space, BIDS '16*, Santa Cruz de Tenerife, Spain, 15-17 March, 2016.
- [99] Weisi Lin, C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image R.*, vol. 22, pp. 297–312, 2011.
- [100] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, p. 1-14, 2004.
- [101] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, and F. Nencini, "A new method for MS+Pan image fusion assessment without reference," *IEEE*, 2006.
- [102] M. C. El-Mezouar, N. Taleb, K. Kpalma, and J. Ronsin, "A new evaluation protocol for image pan-sharpening methods," *ICCIT 2012*, pp. 144-148.
- [103] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193-200, 2008.
- [104] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Proc. Letters*, vol. 9, no. 3, pp. 81-84, March 2002.
- [105] R. Dosselmann and Xue Dong Yang, "A comprehensive assessment of the structural similarity index," *SIViP 2011*, vol. 5, pp. 81–91, 2011.



- [106] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, et al., "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data Fusion Contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012-3021, 2007.
- [107] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, Perceptual image quality assessment using a normalized Laplacian pyramid, *Proc. IS&T Int'l Symposium on Electronic Imaging, Conf. on Human Vision and Electronic Imaging*, vol. 2016(16), Feb 2016.
- [108] B. Veeramallu, C. LavanyaSusanna, and S. Sahitya, "Survey on an image quality assessment metric based on early vision features," *Int. J. Soft Computing and Engineering (IJSCE)*, vol. 2, no. 6, pp. 447-449, Jan. 2013.
- [109] T. Poggio, J. Mutch, and L. Isik, "Computational role of eccentricity dependent cortical magnification," *CBMM Memo*, No. 017, June 6, 2014.
- [110] MIT, Centers for Brains, Minds + Machines, "The computational role of eccentricity dependent resolution in the retina: consequences for hierarchical models of object recognition." [Online]. Available: <https://cbmm.mit.edu/research/projects/thrust/theories/intelligence/computationalrole/eccentricitydependentresolution>
- [111] A Quality Assurance Framework for Earth Observation, version 4.0, Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS), 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- [112] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments", *IEEE Trans. Image Proc.*, vol. 20, no. 9, pp. 2475-2489, 2010.
- [113] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1209-1221, 2001.
- [114] B. Zhou, A. Lapedriza, Jianxiong Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using Places database," *NIPS*, pp. 1-9, 2014.
- [115] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems 25 (NIPS 2012)*.
- [116] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 3, pp. 1349-1362, Oct. 2015.
- [117] M. Långkvist, A. Kiselev, M. Alirezaie and A. Loutfi, "Classification and segmentation of satellite orthoimagery using convolutional neural networks", *Remote Sens.*, vol. 8, no. 329, pp. 1-21, 2016.
- [118] S. Mallat, "Understanding Deep Convolutional Networks", *Phil. Trans. R. Soc. A*, vol. 374: 20150203, pp. 1-16, 2016.
- [119] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, May 2014, DOI: 10.5244/C.28.6.
- [120] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi, "Deep filter banks for texture recognition, description, and segmentation," *CoRR*, abs/1411.6836, 2014.
- [121] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pp. 1398-1402, 2003.
- [122] N. Hunt and S. Tyrrell, *Stratified Sampling*. Coventry University, 2012. [Online] Available: <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>
- [123] Si Liu, Hairong Liu, L. J. Latecki, Shuicheng Yan, Changsheng Xu, Hanqing Lu, "Size adaptive selection of most informative features," *Assoc. Advanc. Artificial Intel.*, 2011.
- [124] H. Peng, H. F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, pp. 1226-1238, 2005.
- [125] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, 1948.
- [126] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 3, pp. 411-426, 2007.
- [127] A. Vedaldi and A. Zisserman, *VGG Convolutional Neural Networks Practical*, Oxford Visual Geometry Group (VGG), computer vision practical, 2015.
- [128] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. XX, no. X, pp. 1-16, 2015.
- [129] J. Vazquez-Corral, M. Vanrell, R. Baldrich, F. Tous, "Color constancy by category correlation". *IEEE Trans. Image Proc.*, vol. 21, no. 4, pp. 1997-2007, 2012.



- [130] E. Land, "Recent advances in retinex theory," *Vision Research*, vol. 26, pp. 7–21, 1986.
- [131] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Proc.*, vol. 16, no. 9, pp. 2207–2214, 2007.
- [132] B. Berlin and P. Kay, *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.
- [133] *ENVI EX User Guide 5.0*, ITT Visual Information Solutions, Dec. 2009. [Online]. Available: [http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI\\_EX\\_User\\_Guide.pdf](http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf)
- [134] *Exelis VIS Technical Support*, Personal communication, March 13, 2013.
- [135] Big data, 2016. [Online]. Available: [en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)
- [136] European Commission, "Big Data Public Private Forum", *Cordis.europa.eu*. 2012-09-01. [Online]. Available: [http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ\\_RCN=13267529](http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=13267529). Retrieved 2013-03-05.
- [137] A. Baraldi, L. Boschetti, and M. Humber, "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, Part: 2, pp. 701-760, Jan. 2014.
- [138] Cognitive Science. [Online]. Available: [https://en.wikipedia.org/wiki/Cognitive\\_science](https://en.wikipedia.org/wiki/Cognitive_science). Retrieved 2016-09-05.
- [139] A. C. Watts, V. G. Ambrosia and E. A. Hinkle, "Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use," *Remote Sens.*, vol. 4, pp. 1671-1692, 2012.
- [140] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, 1980.
- [141] D. Clausi and Yongping Zhao, "Gray level co-occurrence integrated algorithm (GLCIA): a superior computational method to rapidly determine co-occurrence probability texture features," *Computers & Geosciences*, vol. 29, pp. 837-850, 2003.
- [142] K. Fukushima, "A neural network model for selective attention in visual pattern recognition," *Biological Cybernetics*, vol. 55, no. 1, pp. 5-15, 1986.
- [143] J. Shotton, J.M. Winn, C. Rother, and A. Criminisi, "Texton-Boost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context," *Int. J. Comp. Vision*, vol. 81, no. 1, pp. 2–23, 2009.
- [144] R. Khan, J. van de Weijer, F. Shahbaz Khan, D. Muselet, C. Ducottet, C. Barat, "Discriminative color descriptors", *CVPR 2013*.
- [145] B. Fritzsche, *Some Competitive Learning Methods*, 1997. Draft document. [Online]. Available: <http://www.demogng.de/JavaPaper/t.html>
- [146] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–94, Jan. 1980.
- [147] N. Memarsadeghi, D. Mount, N. Netanyahu, and J. Le Moigne, "A fast implementation of the ISODATA clustering algorithm," *Int. J. Comp. Geometry & Applications*, vol. 17, no. 1, pp. 71-103, 2007.
- [148] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 2563-2586, 2006.
- [149] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1299 - 1325, March 2010.
- [150] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1326 - 1354, March 2010.
- [151] A. Baraldi, "Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 6, pp. 2113 - 2134, June 2011.
- [152] Y. Li et al., Use of second derivatives of canopy reflectance for monitoring prairie vegetation over different soil backgrounds, *Remote Sens. Environment*, vol. 44, no. 1, pp. 81-87, 1993.
- [153] Bayesian inference, Wikipedia. [Online]. Available: [https://en.wikipedia.org/wiki/Bayesian\\_inference](https://en.wikipedia.org/wiki/Bayesian_inference). Retrieved 2016-09-05.



- [154] A. Baraldi, M. L. Humber, D. Tiede and S. Lang, "Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation," submitted for consideration for publication in *Int. J. Remote Sens.*, 2016.
- [155] A.-V. Vo, L. Truong-Hong, D.F. Lafer, D. Tiede, S. d'Oleire-Oltmanns, A. Baraldi, M. Shimoni, G. Moser, D. Tuia "Processing of Extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS Data Fusion Contest. Part-B: 3D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5560-5575, Dec. 2016.
- [156] Dillencourt, M. B., Samet, H., & Tamminen, M. (1992). A general approach to connected component labeling for arbitrary image representations. *J. Assoc. Computing Machinery*, 39, 253-280.
- [157] Spreadsheet, Wikipedia. [Online]. Available: <https://en.wikipedia.org/wiki/Spreadsheet>
- [158] P. Chavez, An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data, *Remote Sens. Environ.*, vol. 24, pp. 459-479, 1988.
- [159] J. Tsien, "Principles of Intelligence: On Evolutionary Logic of the Brain," *Frontiers in Systems Neuroscience*, 3 Feb. 2016, pp. 1-7.
- [160] K. Kuzera and R. G. Pontius Jr., "Importance of matrix construction for multiple-resolution categorical map comparison," *GIScience and Remote Sens.* vol. 45, pp. 249-274, 2008.
- [161] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, Boca Raton, FL, USA: Lewis Publishers, 1999.
- [162] R. Lunetta and D. Elvidge, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, London, UK: Taylor & Francis, 1999.
- [163] M. Beauchemin and K. Thomson, "The evaluation of segmentation results and the overlapping area matrix," *Int. J. Remote Sens.*, vol. 18, pp. 3895-3899, 1997.
- [164] A. Ortiz and G. Oliver, "On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures," *Pattern Recognition Letters*, vol. 27, pp. 1916-1926, 2006.
- [165] S. V. Stehman and R. L. Czaplewski, "Design and analysis for thematic map accuracy assessment: Fundamental principles," *Remote Sens. Environ.*, vol. 64, pp. 331-344, 1998.
- [166] Wenwen Li, M. F. Goodchild and R. L. Church, "An efficient measure of compactness for 2D shapes and its application in regionalization problems," *Int. J. of Geographical Info. Science*, pp. 1-24, 2013.
- [167] S. Grove, "Knowledge-based interpretation of multisensory and multitemporal remote sensing images," *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June, 1999.
- [168] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A New Framework for SAR multitemporal data RGB representation: Rationale and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117-133, 2015.
- [169] R. Serra and G. Zanarini, *Complex Systems and Cognitive Processes*, Berlin: Springer-Verlag, 1990.
- [170] *Imaging Spectrometry*, Eds. F. Van der Meer and S. De Jong, Dordrecht, The Netherlands: Springer 2011.
- [171] Fred Hatt, *The full gamut - Drawing life*, 2016. [Online] Available: <http://fredhatt.com/blog/2011/10/23/the-full-gamut/>
- [172] IBM, *The Four V's of Big Data*, IBM Big Data & Analytics Hub. [Online] Available: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>. Accessed on 31 Dec. 2016.
- [173] M. Baatz and A. Schäpe, "Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation," In *Angewandte Geographische Informationsverarbeitung XII*; Strobl, J., Ed.; Herbert Wichmann Verlag: Berlin, Germany, 58: 12-23, 2000.
- [174] G. M. Espindola, G. Camara, I. A. Reis, L. S. Bins, and A. M. Monteiro, "Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation," *Int. J. Remote Sens.* 27(14): 3035-3040, 2006.
- [175] G. Camara, R. Souza, U. Freitas, and J. Garrido, "SPRING: Integrating remote sensing and GIS by object-oriented data modelling," *Computers & Graphics*, 20(3): 395-403, 1996.
- [176] *eCognition® Developer 9.0 Reference Book*, Trimble, 2015.
- [177] A. Baraldi, D. Tiede, M. Sudmanns, M. Belgiu, and S. Lang, "Automated near real-time Earth observation Level 2 product generation for semantic querying," *GEOBIA 2016*, 14-16 Sept. 2016, University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands.
- [178] T. Villmann, R. Der, M. Herrmann, and T. M. Martinetz, "Topology preservation in self-organizing feature maps: Exact definition and measurement," *IEEE Trans. Neural Networks*, vol. 8, no. 2, pp. 256-266, 1997.
- [179] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part I," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 645-661, 2002.
- [180] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part II," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 662-677, 2002.





- [181] National Aeronautics and Space Administration (NASA). Getting Petabytes to People: How the EOSDIS Facilitates Earth Observing Data Discovery and Use. [Online]. Available: <https://earthdata.nasa.gov/getting-petabytes-to-people-how-the-eosdis-facilitates-earth-observing-data-discovery-and-use>. Accessed on December 29, 2016.
- [182] G. M. Pinna and F. Ferrante, “The ESA Earth Observation Payload Data Long Term Storage Activities,” European Space Agency, 2009. Retrieved December 26, 2016, from [http://www.cosmos.esa.int/documents/946106/991257/13\\_Pinna-Ferrante\\_ESALongTermStorageActivities.pdf/813babe0-58db-4e23-b710-3bd9d6b58b12](http://www.cosmos.esa.int/documents/946106/991257/13_Pinna-Ferrante_ESALongTermStorageActivities.pdf/813babe0-58db-4e23-b710-3bd9d6b58b12)
- [183] V. Manilici, S. Kiemle, C. Reck, and M. Winkler, “EOLib Architecture Concept for an Information Mining System for Earth Observation Data,” PV2013, ESRIN, Frascati, 2013-11-04.
- [184] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, “Selective search for object recognition,” Technical Report 2012, submitted to IJCV 2013. Available: <http://www.huppelen.nl/publications/selectiveSearchDraft.pdf>. Accessed on Feb. 15, 2017.
- [185] Andrea Baraldi, Michael Laurence Humber, Dirk Tiede, Stefan Lang, “Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation, Part 1 Theory,” Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01930. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01930.pdf>
- [186] Andrea Baraldi, Michael Laurence Humber, Dirk Tiede, Stefan Lang, “Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation, Part 2 Validation,” Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01932. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01932.pdf>
- [187] H. Öğmen and M. Herzog, “The Geometry of Visual Perception: Retinotopic and Nonretinotopic Representations in the Human Visual System,” Proc IEEE Inst Electr Electron Eng., 98(3): 479–492, 2010.
- [188] Andrea Baraldi, Dirk Tiede, Stefan Lang, “Automated Linear-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images,” Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01940. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01940.pdf>
- [189] T. Poggio, “The Levels of Understanding framework, revised,” Computer Science and Artificial Intelligence Laboratory, Technical Report, MIT-CSAIL-TR-2012-014, CBCL-308, May 31, 2012.
- [190] P. Quinlan, “Marr’s Vision 30 years on: From a personal point of view”, Perception, vol. 41, pp. 1009 – 1012, 2012.



## **4 Manuscript 1 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal *Remote Sensing of Environment*): Systematic Earth Observation Level 2 product generation for semantic querying**

### **Motivation and Contributions to the Dissertation**

No EO image understanding system (EO-IUS) exists to date capable of transforming large-scale multi-source EO image databases into timely, comprehensive and operational EO value-adding information products and services, in compliance with the visionary goals of the intergovernmental Group on Earth Observations (GEO)'s Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015, submitted to (constrained by) the Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements. In addition to the lack of EO-IUSs in operating mode, no semantic content-based image retrieval (SCBIR) system has ever been implemented in operating mode by the RS community, although many content-based image retrieval (CBIR) systems are available to query image databases based on metadata text information (e.g., EO image acquisition time, geographic area of interest, etc.), including image-wide summary statistics (e.g., per-image cloud cover quality index), and/or by either image, image-object or multi-object examples. SCBIR is synonym of semantics-enabled information/knowledge discovery in image databases where, in user-speak, semantic querying capability is required to process semantic queries such as “retrieve all EO images not necessarily cloud-free acquired by imaging sensor X where wetland areas are visible and located adjacent to a highway near a coast in the eastern part of country Y”.

To fill the analytic and pragmatic information gap from CBIR to SCBIR in large-scale multi-source EO image databases, systematic semantic enrichment of EO sensory data was considered a necessary not sufficient pre-condition of SCBIR, i.e.,  $SCBIR \supset CV \supset EO-IU$  in operating mode  $\supset$  human vision, where human visual perception is considered a lower bound of CV, refer to Chapter 1 (Doctoral Research Objectives).

To prove this working hypothesis, the original CV algorithms proposed in Chapter 3 (Technical report 1) were employed by an innovative closed-loop Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system, proposed as a proof-of-concept of a GEOSS in support of SCBIR. The original EO-IU4SQ system comprises, first, an automated near real-time multi-source EO-IU subsystem for single-date and multi-temporal EO big image understanding. The primary (dominant) EO-IU subsystem, capable of filling the semantic gap from sensory data to thematic information products starting from systematic generation of European Space Agency (ESA) EO Level 2 product, is considered a necessary not sufficient pre-condition of SCBIR. Second, connected in closed-loop with the primary EO-IU subsystem, a secondary (dependent) EO-SQ subsystem prototype is proposed. It comprises a graphic user interface (GUI) to streamline high-level user- and application-specific EO image interpretation and SCBIR operations and a raster database management system.

In Chapter 1 (Doctoral Research Objectives), the closed-loop EO-IU4SQ system architecture was sketched in Fig. 1-9, which is reported hereafter for the sake of clarity. The modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 4 (Manuscript 1) are color filled.

It is worth mentioning that the EO-IU4SQ system prototype research and development (R&D) was funded in part by:

- 2015-2016. Austrian Research Promotion Agency (FFG), project call ASAP-11, AutoSentinel-2/3 project (Knowledge-based pre-classification of Sentinel-2/3 images for operational product generation and content-based image retrieval).
- 2016-2017. Austrian Research Promotion Agency (FFG), project call Proposals to ICT of the Future, SemEO project (Semantic enrichment of optical EO data to enhance spatio-temporal querying capabilities).

The EO-IU4SQ system prototype was awarded 1st place in the T-Systems Big Data Challenge of the Copernicus Masters 2015 (Awards Ceremony at the Satellite Masters Conference, 20-22 Oct. 2015, German Federal Ministry of Transport and Digital Infrastructure, Berlin, Germany).

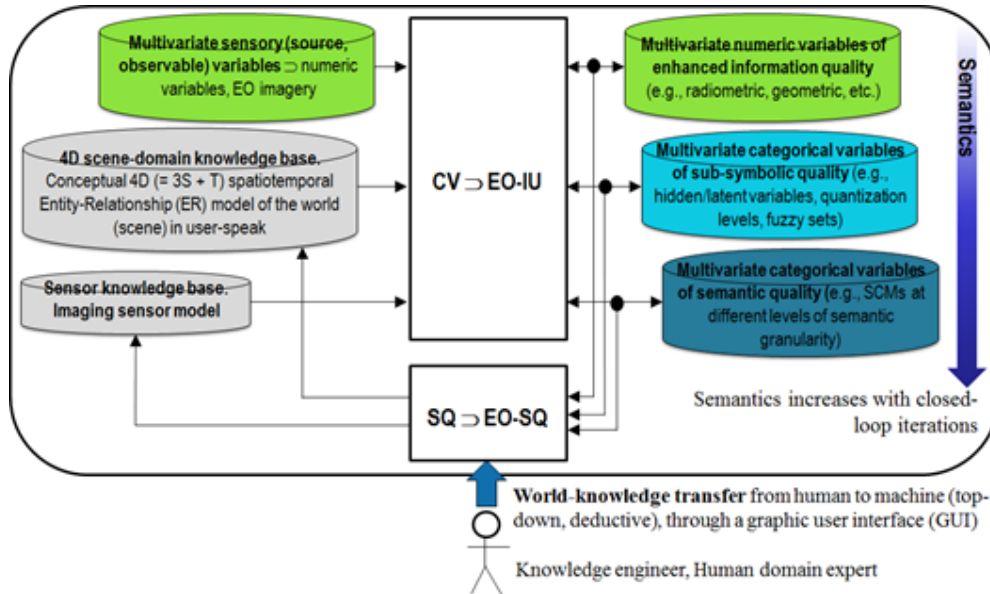


Fig. 1-9, reposed for the sake of clarity. Top-level modular design of a closed-loop EO image understanding (EO-IU) for semantic querying (EO-IU4SQ) system architecture, suitable for incremental learning. It comprises a primary (dominant, necessary not sufficient) hybrid (combined deductive and inductive) EO-IU subsystem in closed-loop with a secondary (dominated) hybrid EO-SQ subsystem. The EO-IU subsystem must be automatic (requiring no human-machine interaction) and near real-time to provide the EO-SQ subsystem with useful information products, including Scene Classification Maps (SCMs) of symbolic quality, as initial necessary not sufficient pre-condition for semantic querying and semantics-enabled information/knowledge discovery. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline high-level user- and application-specific semantic querying and semantics-enabled information/knowledge discovery. Output products generated by the closed-loop EO-IU4SQ system are expected to monotonically increase their value-added with closed-loop iterations.

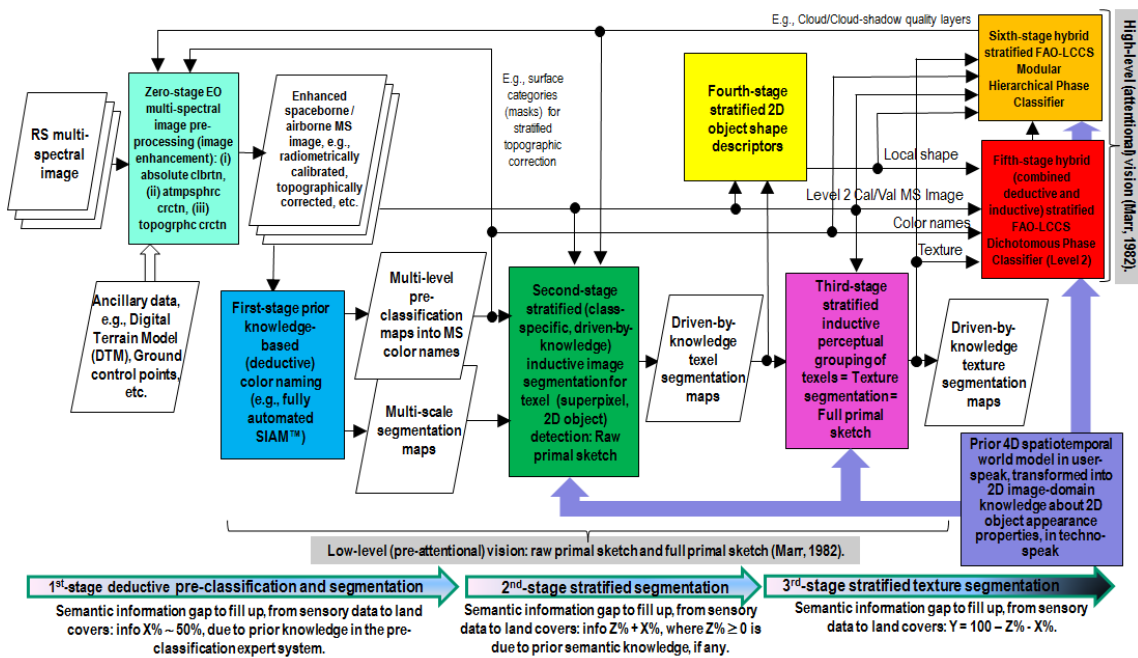


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by



the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 4 (Manuscript 1).



# Systematic Earth Observation Level 2 product generation for semantic querying

A. Baraldi<sup>a,\*</sup>, D. Tiede<sup>b</sup>, M. Sudmanns<sup>b</sup>, M. Belgiu<sup>b</sup>, and S. Lang<sup>b</sup>

<sup>a</sup> Department of Agricultural Sciences, University of Naples Federico II, 80055 Portici (NA), Italy

<sup>b</sup> Paris-Lodron Universität Salzburg, IFFB Geoinformatik, Schillerstraße 30, A-5020 Salzburg, Austria

**Keywords:** Algebra for spatiotemporal objects and events, cognitive science, Earth observation Level 2 product, semantic content-based image retrieval, semantic network, vision.

\* Corresponding author: andrea6311@gmail.com

## Abstract

An innovative Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system prototype is proposed as a proof-of-concept where automated multi-source EO big data spatiotemporal analytics is accomplished as a pre-condition for semantic content-based image retrieval (SCBIR). The EO-IU4SQ prototype consists of two hybrid feedback inference subsystems, where deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based inference are combined with feedback loops. The first EO-IU subsystem automatically transforms in near real-time multi-source multi-spectral EO images into general-purpose user- and application-independent European Space Agency (ESA) EO Level 2 products in compliance with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines. An ESA EO Level 2 product includes an EO image corrected for atmospheric and topographic effects stacked with its general-purpose scene classification map (SCM), whose legend includes cloud and cloud-shadow quality layers. Built upon the first EO-IU subsystem, a second EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline human-machine interaction in support of spatiotemporal EO big data analytics and SCBIR operations. In this paper the EO-IU4SQ system architecture is presented and discussed. Degrees of novelty of the hybrid feedback EO-IU subsystem are highlighted at the levels of abstraction of system design, knowledge/information representation, algorithms and implementation.

## 4.1. Introduction

Traditional Earth observation (EO) content-based image retrieval (CBIR) systems support human-machine interaction through queries by metadata text information (e.g., EO image acquisition time, geographic area of interest, etc.), image-wide summary statistics (e.g., a cloud cover quality index), or by either image, object or multi-object examples (Shyu et al., 2007; Smeulders et al., 2000) (Fig. 4-1). In spite of a wide availability of EO CBIR systems, inherently related to the quantitative unequivocal (“easy”) concept of *information-as-thing* (Capurro and Hjørland, 2003), no EO semantic CBIR (SCBIR) system in operating mode has ever been delivered by the remote sensing (RS) community (Shyu et al., 2007; Dumitru et al., 2015). Related to the qualitative equivocal (“difficult”) concept of *information-as-data-interpretation* (Capurro and Hjørland, 2003), an EO SCBIR system is expected to accomplish systematic processing of spatiotemporal semantic queries, such as “retrieve all EO images not necessarily cloud-free acquired by imaging sensor X showing wetland areas located adjacent to a highway near a coast in the eastern part of country Y”. An EO SCBIR system can be considered in operating mode if it scores “high” in a set of outcome and process (OP) quantitative quality indicators (Q<sup>2</sup>Is), to be community-agreed upon in agreement with the Quality Assurance Framework for Earth Observation (QA4EO) Calibration/Validation (*Cal/Val*) requirements (Group on Earth Observation, 2010). Proposed OP-Q<sup>2</sup>Is include: (i) degree of automation, (ii) accuracy, (iii) efficiency in computation time and run-time memory occupation, (iv) scalability to cope with changes in sensor specifications and user requirements, (v) robustness to changes in input data, (vi) robustness to changes in input parameters, (vii) timeliness from data acquisition to product generation, and (ix) costs in manpower and computer power (Baraldi and Boschetti, 2012a and 2012b). For example, based on these definitions the EO Image Librarian (EOLib), recently presented in the remote sensing (RS) literature as an SCBIR prototype (Dumitru et al., 2015), cannot be considered in operating mode. EOLib is built upon a support vector machine (SVM) for 1D image analysis, whose input vector sequence consists of image convolutional values generated by 2D spatial filters. The EOLib’s 1D image analysis approach scores “low” in several OP-Q<sup>2</sup>Is, including degree of automation, timeliness and costs. First, any inductive



learning-from-data algorithm, such as an SVM, is inherently ill-posed and requires *a priori* knowledge in addition to data to become better conditioned for numerical solution (Cherkassky and Mulier, 1998).

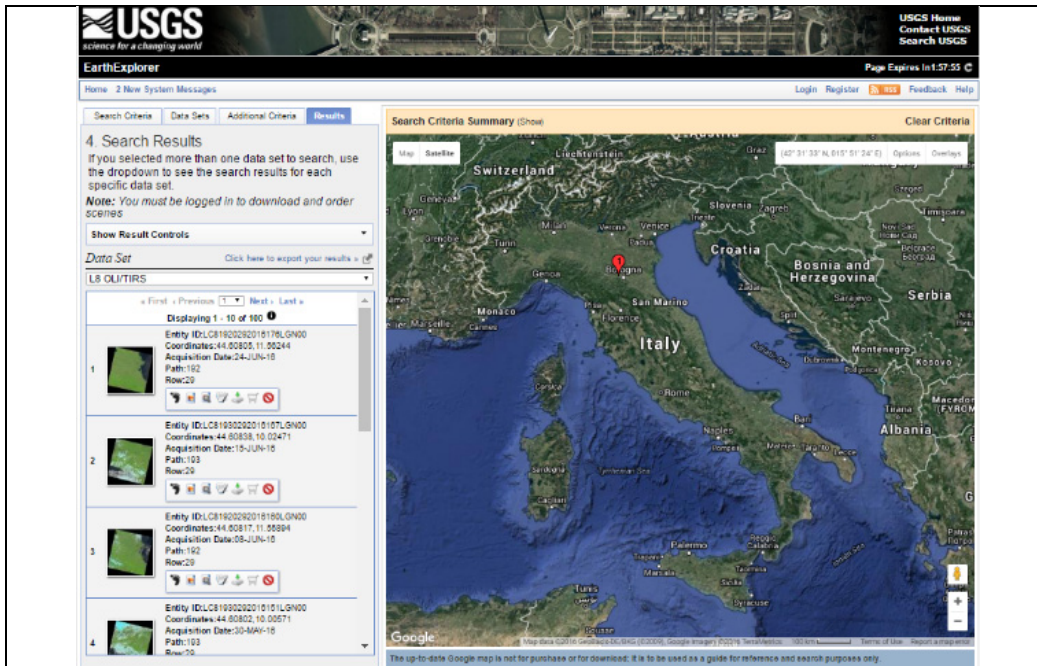


Fig. 4-1. Typical EO image querying systems in operating mode support image retrieval by metadata text information, e.g., acquisition time, target geographic area, etc., or summary statistics, e.g., image-wide cloud cover percentage value. Although they comply with a traditional content-based image retrieval (CBIR) paradigm, they support no semantic CBIR (SCBIR) operation because they lack EO image understanding capabilities.

As a consequence EOLib is semi-automatic; it requires user interaction, first, to collect supervised data samples for training and, second, to define the SVM's free-parameters based on heuristics. In addition, any 1D image classifier is an orderless encoder invariant to permutations in the input vector sequence, where spatial topological information in the (2D) image domain is lost. This is in contrast with an undisputable cognitive fact: spatial information typically dominates color information in both the 4D spatiotemporal real-world domain and the (2D) image domain involved with both chromatic and achromatic visions, which are nearly as effective in scene-from-image representation (Matsuyama and Hwang, 1990). This evidence forms the very foundation of the object-based image analysis (OBIA) paradigm, proposed as a viable alternative to pixel-based image analysis traditionally adopted by the RS community (Blaschke et al., 2014). As a special case of 1D image analysis, pixel-based image classification ignores both spatial non-topological and topological information, because color/gray-value information is the sole visual property available at pixel resolution. In contrast with traditional 1D image interpretation approaches widely adopted by the RS community where non-topological and/or topological information are ignored, combined exploitation of local spatial filters with image topology preserving mapping functions explains the increasing popularity of deep convolutional neural networks in the computer vision (CV) community (Cimpoi et al, 2014).

Our working hypothesis was that existing EO CBIR systems support no semantic querying because they lack EO image understanding capabilities, i.e., their implemented EO image interpretation algorithms fall short in operating mode. This conjecture has two corollaries. First, to solve the SCBIR problem a necessary not sufficient pre-condition is the computational solution of the cognitive (*information-as-data-interpretation*) problem of vision, understood as synonym of scene-from-image reconstruction and understanding, see Fig. 4-2. In other words, the complexity of SCBIR is not inferior to the complexity of vision, acknowledged to be a very difficult problem to solve based on the following three observations.

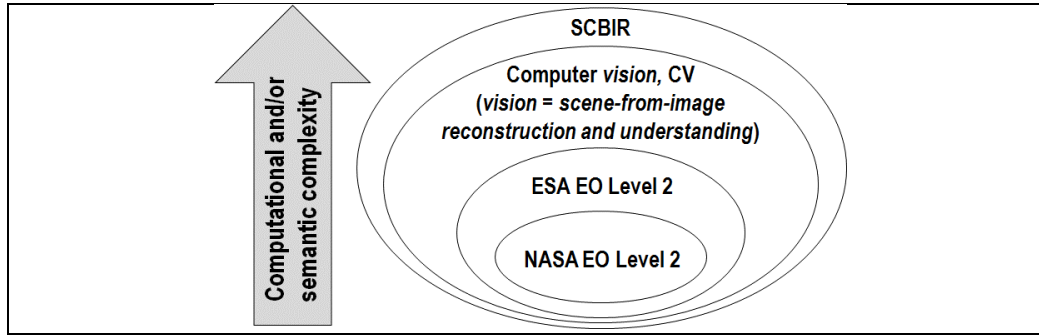


Fig. 4-2. Parent-child inheritance relationship between computer vision (CV) and semantic content-based image retrieval (SCBIR), such that  $SCBIR \supset CV \supset$  Earth observation (EO) image understanding in operating mode  $\supset$  European Space Agency (ESA) EO Level 2 product generation  $\supset$  human vision, also refer to Fig. 4-3. To solve the SCBIR problem, a necessary not sufficient pre-condition is the computational solution of the cognitive problem of vision, intended as 4D spatiotemporal scene from (2D) image reconstruction and understanding. If this working hypothesis holds true, then the complexity of SCBIR is not inferior to the complexity of vision, acknowledged to be an inherently ill-posed cognitive problem in the Hadamard sense, very difficult to solve because: (i) non-polynomial (NP)-hard in computational complexity (Frintrop, 2011; Tsotsos, 1990), (ii) affected by a 4D-to-2D data dimensionality reduction, e.g., responsible of occlusion phenomena, and (iii) affected by a semantic information gap from ever-varying sub-symbolic sensory data (sensations) in the physical world to stable symbolic percepts in the modeled world. Within the CV domain, a National Aeronautics and Space Administration (NASA) EO Level 2 product, defined as “a data-derived geophysical variable at the same resolution and location as Level 1 source data” (NASA, 2016), is a special case of the ESA EO Level 2 product defined as (ESA, 2015; DLR and VEGA, 2011; Telespazio VEGA, 2016): (i) a single-date multi-spectral (MS) image whose digital numbers (DNs) are radiometrically corrected into surface reflectance (SURF) values for atmospheric, adjacency and topographic effects, stacked with (ii) its data-derived general-purpose, user- and application-independent scene classification map (SCM), whose thematic map legend includes quality layers such as cloud and cloud-shadow.

Vision is a cognitive (inference) process inherently ill-posed in the Hadamard sense (Bertero et al, 1988; Matsuyama and Hwang, 1990), i.e., it is non-polynomial (NP)-hard (Frintrop, 2011; Tsotsos, 1990), see Fig. 4-3. Since it admits no solution or multiple plausible solutions (Hadamard, 1902), vision is a very difficult inference task. Its goal is to back-project the numeric/quantitative information in the (2D) image domain onto qualitative/categorical/nominal and semantic information in the 4D spatiotemporal scene domain (Matsuyama and Hwang, 1990). In particular, human vision is an inference process responsible of scene-from-image interpretation in our brain on the basis of a brief glance at a 4D spatiotemporal scene in the real-world, projected onto a 2D retinotopic array (image plane). Scene-from-image representation ranges from pre-attentive global gist to a spatial layout of individual objects in the scene, from attentive local syntax of individual objects to multiple plausible semantic interpretations of the observed 4D spatiotemporal scene encompassing even emotions (du Buf and Rodrigues, 2007; Marr, 1982). Vision is inherently ill-posed because, first, it is affected by a 4D to 2D data dimensionality reduction responsible of visual occlusion phenomena. Second, it is affected by a semantic information gap, from quantitative sub-symbolic ever-varying 2D sensory data (sensations) in the image domain to stable symbolic percepts (concepts) in the 4D spatiotemporal scene domain (Matsuyama and Hwang, 1990). Since it is ill-posed, vision requires *a priori* knowledge in addition to sensory data to make the inference problem better conditioned for solution (Cherkassky and Mulier, 1998). Hence, vision is a *hybrid* inference process where deductive/ top-down/ physical model-based knowledge must be combined with inductive/ bottom-up/ statistical model-based learning-from-examples mechanisms to take advantage of the unique features of each and overcome their shortcomings (Liang, 2004). This thesis complies with the observation that, in nature, no cognitive system starts from scratch (tabula rasa). In particular, any biological inductive learning-from-examples phenotype explores the neighborhood of its initial conditions in a solution space, where initial conditions are set *a priori* by genotype available in addition to sensory data. For example, Vecera and Farah proved that “image segmentation can be influenced by the familiarity of the shape being segmented... Results are consistent with the hypothesis that image segmentation is an interactive (hybrid) process, in which top-down knowledge partly guides lower level processing... If an unambiguous, yet unfamiliar, shape is presented, top-down influences are unable to overcome powerful bottom-up cues... While bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization” (Vecera & Farah, 1997). In contrast with the undisputable fact that vision is a *hybrid* inference process, typical CV systems, including EOLib, consist of inductive algorithms capable of learning-from-examples exclusively.

A second observation useful to assess the complexity of chromatic and achromatic scene-from-image representations, which are nearly equally effective in human common practice, is that color information is typically dominated by spatial information in both the 4D spatiotemporal real-world domain and the (2D) image domain (Matsuyama and Hwang, 1990). As a consequence, while imaging sensors are ever-improving in quality and quantity, CV systems should increasingly focus on modeling and processing spatial rather than color information (Marr, 1982). Unfortunately, existing CV systems, including EOLib, tend to ignore spatial non-topological and/or spatial topological visual information difficult to cope with, in favor of color information easier to deal with because it is the sole visual information available at pixel resolution.

A third observation about the complexity of human vision is that it is a feedback process. For example, visual mental imagery is known to induce retinotopically organized activation of early visual areas via feedback connections, which is tantamount to saying that “mental images in the mind's eye can alter the way we see things in the retina” (Slotnik et al. 2005; Kosslyn 1994). In contrast with this evidence existing CV systems, including EOLib, are typically feedforward, i.e., they are not provided with any feedback loop.

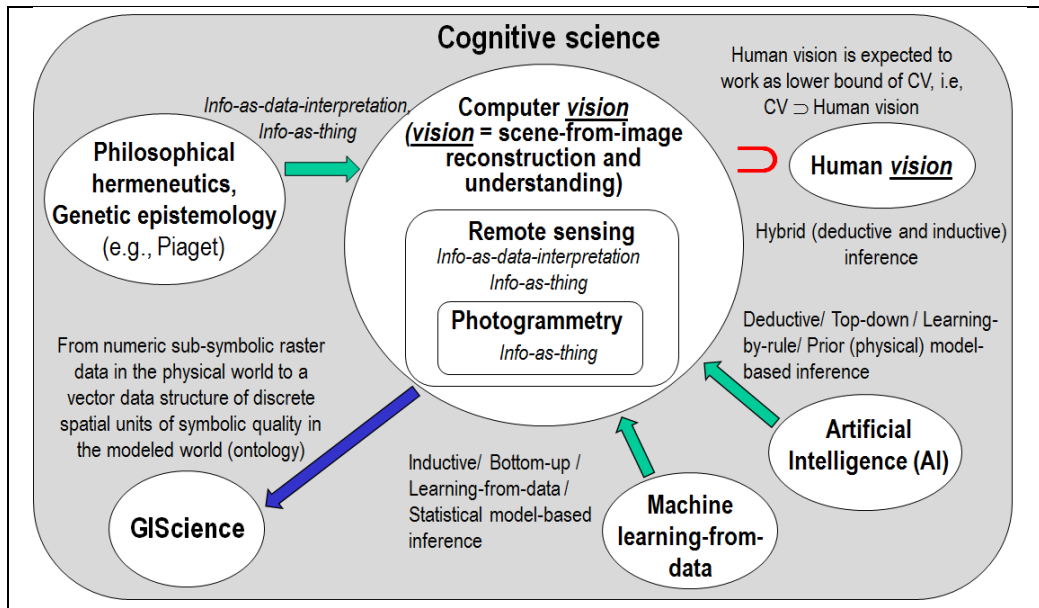


Fig. 4-3. Multi-disciplinary cognitive science. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning, adaptation, self-organization) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred either in the neuro-cerebral apparatus of living organisms or in machines, e.g., computers. Neuroscience, in particular neurophysiology, studies the neuro-cerebral apparatus of living organisms. Neural network (NN) is synonym of distributed processing system consisting of neurons as elementary processing elements and synapses as lateral connections. Is it convenient and even possible to mimic biological mental functions, e.g., human reasoning, by an artificial mind whose physical support is not an electronic brain implemented as an artificial NN (ANN)? The answer is no according to the “connectionists approach” promoted by traditional cybernetics, where a complex system always comprises an “artificial mind-electronic brain” combination. This is alternative to traditional artificial intelligence (AI) whose symbolic approach investigates an artificial mind independently of its physical support (Serra and Zanarini, 1990).

The aforementioned three converging facts provide a realistic estimate of the high complexity of the vision cognitive task, typically underestimated in the RS common practice. They allow to conclude that any realistic and systematic solution to the EO SCBIR problem stems from the development of a novel family of hybrid feedback EO image understanding systems (EO-IUSs) in operating mode in compliance with the QA4EO guidelines (Group on Earth Observation, 2010), alternative to traditional inductive feedforward EO-IUSs widely adopted by the RS community.

The second corollary of our conjecture about an ongoing lack of EO-SCBIR solutions is that existing EO-IUSs to choose from score “low” in operating mode, i.e., they fall short in transforming multi-source EO big data into comprehensive, timely and operational information products. If this conjecture holds true, then existing EO-IUSs do not comply with the



visionary goal of the Group on Earth Observation (GEO), formulated in the Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015 (Group on Earth Observation, 2005) in accordance with the QA4EO *Cal/Val* requirements (Group on Earth Observation, 2010). The conjecture that existing EO-IUSs score “low” in operating mode is supported by several facts. First, the percentage of EO data ever downloaded from the European Space Agency (ESA) databases is estimated at about 10% or less (European Space Agency, 2002). Second, no EO data-derived Level 2 prototype product has ever been generated systematically at the ground segment (European Space Agency, 2015). By definition an EO Level 2 product comprises an EO image corrected for atmospheric, adjacency and topographic effects, and a general-purpose user- and application-independent scene classification map (SCM), including cloud and cloud-shadow quality layers (European Space Agency, 2015).

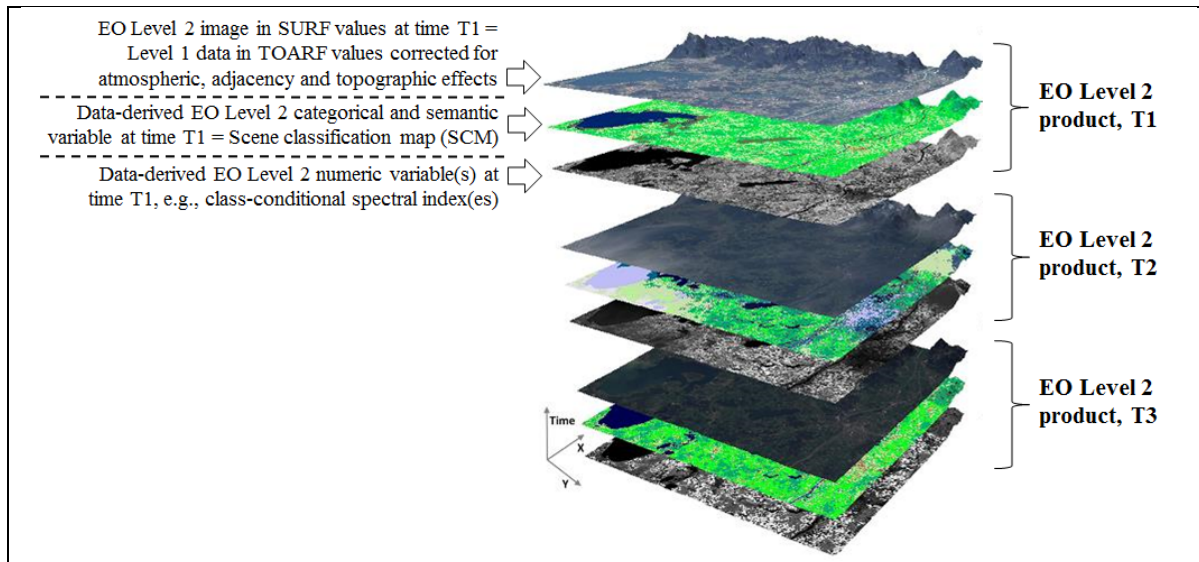


Fig. 4-4. The EO-IU4SQ system for spatiotemporal analytics of multi-source EO big image databases. Each single-date EO Level 1 source image, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and stored in the database, is automatically transformed into an ESA EO Level 2 product comprising: (i) a single-date multi-spectral (MS) image radiometrically calibrated from TOARF into surface reflectance (SURF) values, corrected for atmospheric, adjacency and topographic effects, stacked with (ii) its EO data-derived value-adding scene classification map (SCM), equivalent to a sensory data-derived categorical/qualitative variable of semantic quality, where the thematic map legend is general-purpose, user- and application-independent and comprises quality layers such as cloud and cloud-shadow, eventually stacked with (iii) its EO data-derived value-adding numeric/quantitative variables, e.g., biophysical variables (e.g., leaf area index), class-conditional spectral indexes (e.g., vegetation class-conditional spectral index), etc., or categorical variables of sub-symbolic quality (geographic field-objects), e.g., fuzzy sets/discretization levels low/medium/high of a numeric variable (Goodchild et al., 2007).

To contribute toward filling an ongoing information gap from EO big data to EO value-adding products and services, this paper presents an innovative EO Image Understanding for Semantic Querying (EO-IU4SQ) system prototype as a proof-of-concept where multi-source EO big data spatiotemporal analytics is a pre-condition for SCBIR (Fig. 4-2 and Fig. 4-4). The proposed EO-IU4SQ prototype consists of two hybrid feedback inference subsystems. The first EO-IU subsystem transforms automatically (i.e., without user interaction) and in near real-time any multi-source EO multi-spectral (MS) image, featuring a radiometric calibration metadata file in agreement with the QA4EO *Cal/Val* requirements, into an ESA EO Level 2 product consisting of: (a) an enhanced EO image corrected for atmospheric and topographic effects, and (b) an SCM whose legend includes quality layers, such as cloud and cloud-shadow image masks, in addition to the 8-class land cover (LC) taxonomy adopted by the initial Dichotomous Phase (DP) of the two-phase Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) (Di Gregorio and Jansen, 2000). The general-purpose user- and application-independent 8-class LCCS-DP taxonomy consists of three “nested” dichotomous layers: (i) vegetation/non-vegetation, (ii) terrestrial/aquatic and (iii) managed/natural. They deliver as output the following eight LCCS-DP classes. (A11) Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas. (A12) Natural and Semi-Natural Terrestrial Vegetation. (A23) Cultivated Aquatic or Regularly Flooded Areas. (A24) Natural and



Semi-Natural Aquatic or Regularly Flooded Vegetation. (B35) Artificial Surfaces and Associated Areas. (B36) Bare Areas. (B47) Artificial Waterbodies, Snow and Ice. (B48) Natural Waterbodies, Snow and Ice (Fig. 4-5). The general-purpose 8-class LCCS-DP legend is preliminary to the LCCS Modular Hierarchical Phase (MHP) taxonomy, consisting of a hierarchical battery of application- and user-specific one-class LC classifiers. Unlike alternative hierarchical LC class taxonomies whose Level 1 is multi-class, such as the CORINE Land Cover (CLC) (Bossard et al., 2000) and the EOLib's taxonomy (Dumitru et al., 2015), the two-phase LCCS taxonomy is fully “nested”, starting from the first-level DP layer vegetation/non-vegetation, whose relevance becomes paramount according to a well-known garbage in, garbage out (GIGO) principle of error propagation.

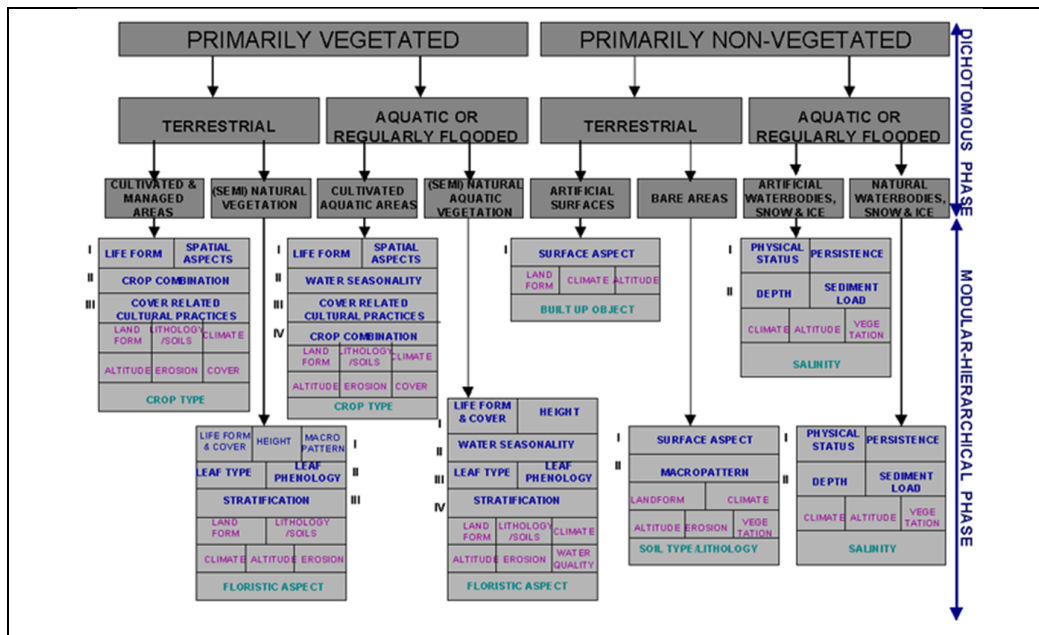


Fig. 4-5. The FAO LCCS taxonomy comprises a general-purpose, user- and application-independent 3-level 8-class Dichotomous Phase (DP) legend, followed by a user- and application-specific Modular Hierarchical Phase (MHP) taxonomy. The FAO LCCS 3-level hierarchical DP mapping criteria are: (i) vegetation/non-vegetation, (ii) terrestrial/aquatic, (iii) managed/natural or semi-natural.

In the rest of this paper the proposed EO-IU4SQ system architecture is described and discussed. Next, degrees of novelty of the hybrid feedback EO-IU subsystem are highlighted at the four levels of abstraction of system design, knowledge/information representation, algorithms and implementation. According to Marr, the linchpin of success of any information processing system is system design and knowledge/information representation, rather than algorithms and implementation (Marr, 1982).

## 4.2. Materials and Methods

In this section the EO-IU4SQ system architecture is proposed and its prototypical materials and methods are discussed, with special emphasis on the hybrid feedback EO-IU subsystem.

### 4.2.1. System design

The EO-IU4SQ system architecture (Fig. 4-6 and Fig. 4-7) combines two hybrid feedback inference subsystems overlapping in part: an EO-IU subsystem, capable of automated general-purpose EO Level 2 product generation, is preliminary to a second EO-SQ subsystem, provided with a GUI for high-level user- and application-specific EO image interpretation and SCBIR operations. These two subsystems share (Laurini and Thompson, 1992): (i) a fact base of EO data and derived products, at either Level 2 or higher. (ii) A knowledge base available in addition to facts. This knowledge base comprises a “default” baseline of physical laws, first-principle models, if-then decision rules, methods, processes, etc., to be streamlined by an inference engine to extract new information from the fact base. (iii) An inference engine, which applies the knowledge base to the fact base to infer new information, including EO Level 2 products.



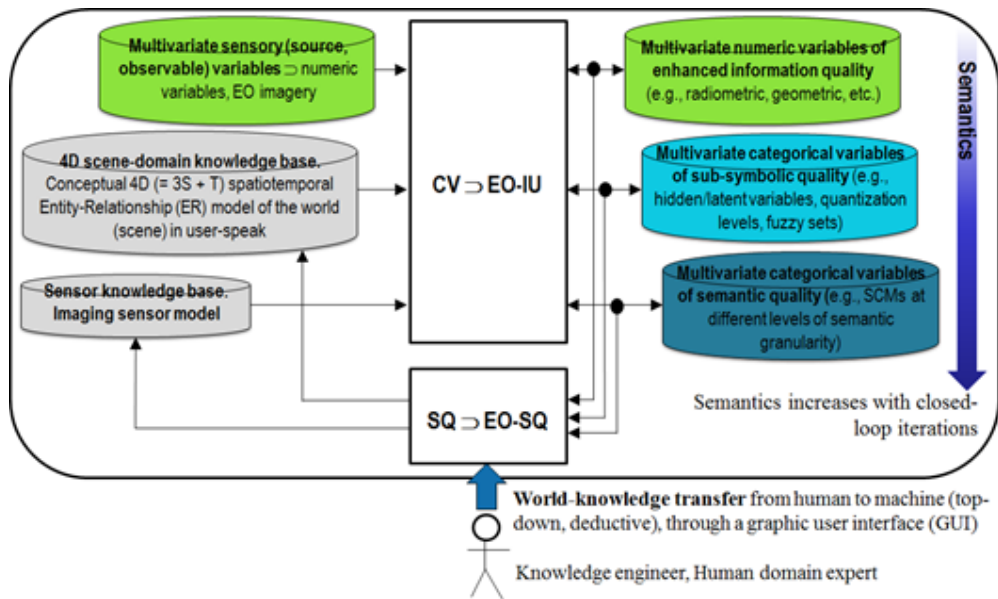


Fig. 4-6. Top-level modular design of a closed-loop EO image understanding (EO-IU) for semantic querying (EO-IU4SQ) system architecture, suitable for incremental learning. It comprises a primary (dominant, necessary not sufficient) hybrid (combined deductive and inductive) EO-IU subsystem in closed-loop with a secondary (dominated) hybrid EO-SQ subsystem. The EO-IU subsystem must be automatic (requiring no human-machine interaction) and near real-time to provide the EO-SQ subsystem with useful information products, including Scene Classification Maps (SCMs) of symbolic quality, as initial necessary not sufficient pre-condition for semantic querying and semantics-enabled information/knowledge discovery. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline high-level user- and application-specific semantic querying and semantics-enabled information/knowledge discovery. Output products generated by the closed-loop EO-IU4SQ system are expected to monotonically increase their value-added with closed-loop iterations.

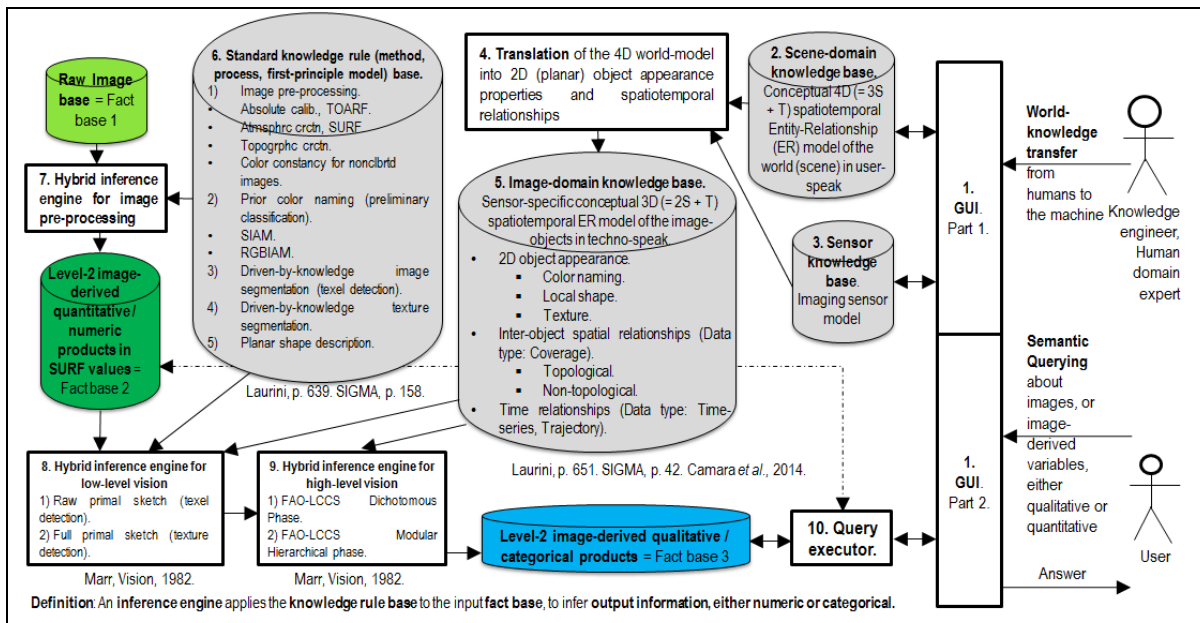


Fig. 4-7. EO-IU4SQ system modular design proposed at a finer level of detail than the top-level modular design shown in Fig. 4-6. Processing modules are shown as rectangles and databases as cylinders. Acronyms adopted in the rest of this paper: graphic user- interface (GUI), surface reflectance (SURF), top-of-atmosphere reflectance (TOARF), Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color

names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain (Baraldi et al., 2006), RGB Image Automatic Mapper (RGBIAM) lightweight computer program for RGB space polyhedralization into color names, superpixel detection and vector quantization (VQ) quality assessment in the image-domain (Baraldi et al., 2017), FAO Land Cover Classification System (LCCS) taxonomy (Di Gregorio and L. Jansen, 2000).

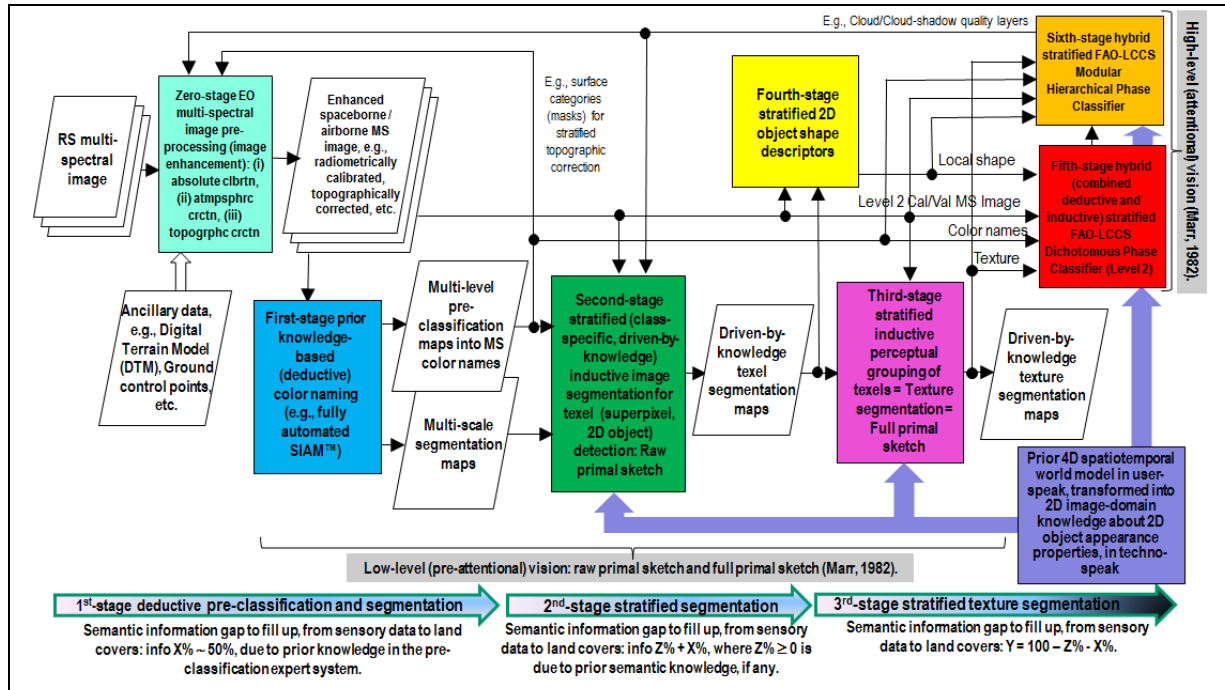


Fig. 4-8. Six-stage *hybrid* feedback EO image understanding (EO-IU) subsystem modular design, based on a convergence-of-evidence approach to low-level (pre-attentive) and high-level (attentive) vision tasks, in agreement with the Marr terminology (Marr, 1982). For the sake of visualization each of the six data processing stages plus stage-zero for EO data pre-processing is depicted as a rectangle with a different color fill. Visual evidence stems from color, local shape, texture and inter-object spatial relationships. Low-level EO image classification is general-purpose, user- and application independent, such as an ESA EO Level 2 Scene Classification Map (SCM) product provided with an 8-class FAO Land Cover Classification System (LCCS) Dichotomous Phase (DP) map legend in addition to quality layers, such as cloud and cloud-shadow. High-level EO image classification is user- and application-specific, such as an EO SCM product provided with an LCCS Modular Hierarchical Phase (MHP) legend.

#### 4.2.2. Knowledge/information representation

The EO-IU4SQ knowledge base consists of four types of knowledge (Fig. 4-7).

(1) 4D spatiotemporal scene-domain knowledge in user-speak, including physical units of measure. Known as world model (Matsuyama and Hwang, 1990), it can be thought of as a spatiotemporal entity-relationship (ER) conceptual model of semantic entities, such as classes of 4D real-world objects, provided with an identification field, numeric/quantitative or nominal/categorical attributes, e.g., appearance properties, and inter-class relationships, either spatiotemporal or not, e.g., part-of, subset-of, etc. (Sonka et al., 1994; Laurini and Thompson, 1992). In EO data applications the world model consist of LC classes belonging to a discrete and finite taxonomy (dictionary) of LC classes to be community-agreed upon in advance, such as the two-phase LCCS taxonomy (Di Gregorio and L. Jansen, 2000). To date grammars (syntactic models) can rarely be inferred from supervised (labeled) or unsupervised (unlabeled) data by statistical approaches. For example, non-spatial inter-class relationships, such as part-of and subset-of, are nearly impossible to be learned from data (Sonka et al., 1994). In common practice, grammars require significant top-down human-machine interaction to be acquired by a machine. In addition, “no amount of syntax will ever produce semantics” (Mayo, 2003). Equivalent to a grammar provided with semantics, a world model can be transferred top-down from human domain-experts to the EO-IU4SQ machine through a GUI, where the world model can be graphically represented as a semantic network (Grove, 1999). This deductive/top-down knowledge transfer from human-to-machine, typically investigated by artificial intelligence (Laurini and Thompson, 1992), is the dual problem of (complementary not alternative to) inductive/bottom-up learning-from-data, typically

investigated by machine learning (Fig. 4-3). In the EO-IU4SQ strategy, the prior knowledge-based world model provides initial conditions to the EO-IU and EO-SQ hybrid inference subsystems capable of learning from data.

(2) Image-domain knowledge in techno-speak, e.g., in pixel units, accounting for quantitative visual features, such as color, local shape (Baraldi and Soares, 2017), texture and inter-object spatial relationships (Matsuyama and Hwang, 1990; Smeulders et al., 2000; Sonka et al., 1994).

(3) Knowledge about the sensor transfer functions (translation rules), capable of mapping the scene-domain knowledge in physical units onto the image-domain knowledge in pixel units.

(4) “Default” baseline of general-purpose, user- and application-independent physical laws, first-principle models, if-then decision rules, methods, processes, etc., capable of automated near real-time EO image pre-processing and low-level (pre-attentive) vision (Mason and E. R. Kandel, 1991). These are software modules in operating mode selected and streamlined by the EO-IU4SQ inference engine to generate new information layers in the 4D scene domain, including EO Level 2 products, from the existing fact base of EO images and derived products.

#### 4.2.3. “Default” rule base for EO image pre-processing and low-level vision

An original battery of automated near real-time EO image pre-processing and low-level vision algorithms in operating mode was developed in compliance with the QA4EO *Cal/Val* requirements and the EO-IU4SQ system’s target OP-Q<sup>2</sup>I values discussed in Chapter 4.1. In the terminology of neurophysiology (Mason & Kandel, 1991), low-level (pre-attentive) vision is preliminary to high-level (attentive) vision. The proposed hybrid feedback EO-IU subsystem design encompasses both low- and high-level vision modules (Fig. 4-8). According to Marr, low-level vision comprises two phases (Marr, 1982): (i) a raw primal sketch, responsible of image contour detection followed by image segmentation into connected image-objects (segments); the raw primal sketch is input to (ii) a full primal sketch, where texture segmentation occurs as perceptual grouping of texture elements, traditionally known as texels (Julesz et al., 1973). In low-level vision, numeric variables to be investigated in the image domain are color, which is the sole visual property available at pixel resolution, local shape and size of image-objects (segments) (Baraldi and Soares, 2017), texture and inter-object spatial relationships, either topological, such as adjacency, inclusion, etc., or non-topological, such as distance and angle measures (Marr, 1982; Matsuyama and Hwang, 1990; Smeulders et al., 2000; Sonka et al., 1994).

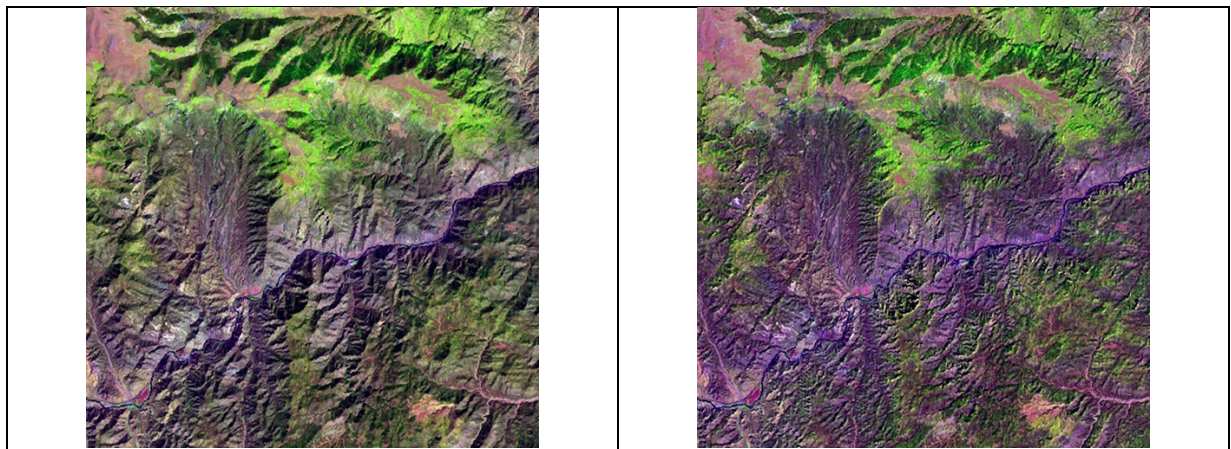


Fig. 4-9. Left: Zoomed area of a Landsat 7 ETM+ image of Colorado, USA (path: 128, row: 021, acquisition date: 2000-08-09), depicted in false colors (R: band ETM5, G: band ETM4, B: band ETM1), 30 m resolution, radiometrically calibrated into TOARF values. Right: Automated stratified topographic correction (TOC), based on a SIAM’s spectral map at coarse spectral granularity, consisting of 16 spectral categories, and a Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM), 30 m resolution. No systematic EO Level 2 product generation at the ground segment featuring topographic correction has ever been accomplished by the RS community.

Visual attributes in the (2D) image domain should never be confused with materials (surface types) in the 4D scene domain (Cimpoi et al, 2014). For example, numeric color values mapped onto categorical color names, e.g., colors red (R), green (G) and blue (B), “cannot always be inverted to unique LC class names”; the same consideration holds for discrete spectral endmembers typically adopted in hyper-spectral image interpretation (Adams et al, 1995). According to a convergence-of-evidence approach typical of human symbolic reasoning (Matsuyama and Hwang, 1990), traditionally





mimicked by fuzzy logic (Zadeh, 1965), it is the combination of visual attributes, with special emphasis on spatial information, capable of identifying a target material in the scene domain (Fig. 4-8).

The several degrees of novelty of the implemented battery of general-purpose EO image pre-processing and low-level vision algorithms are highlighted in the rest of this section.

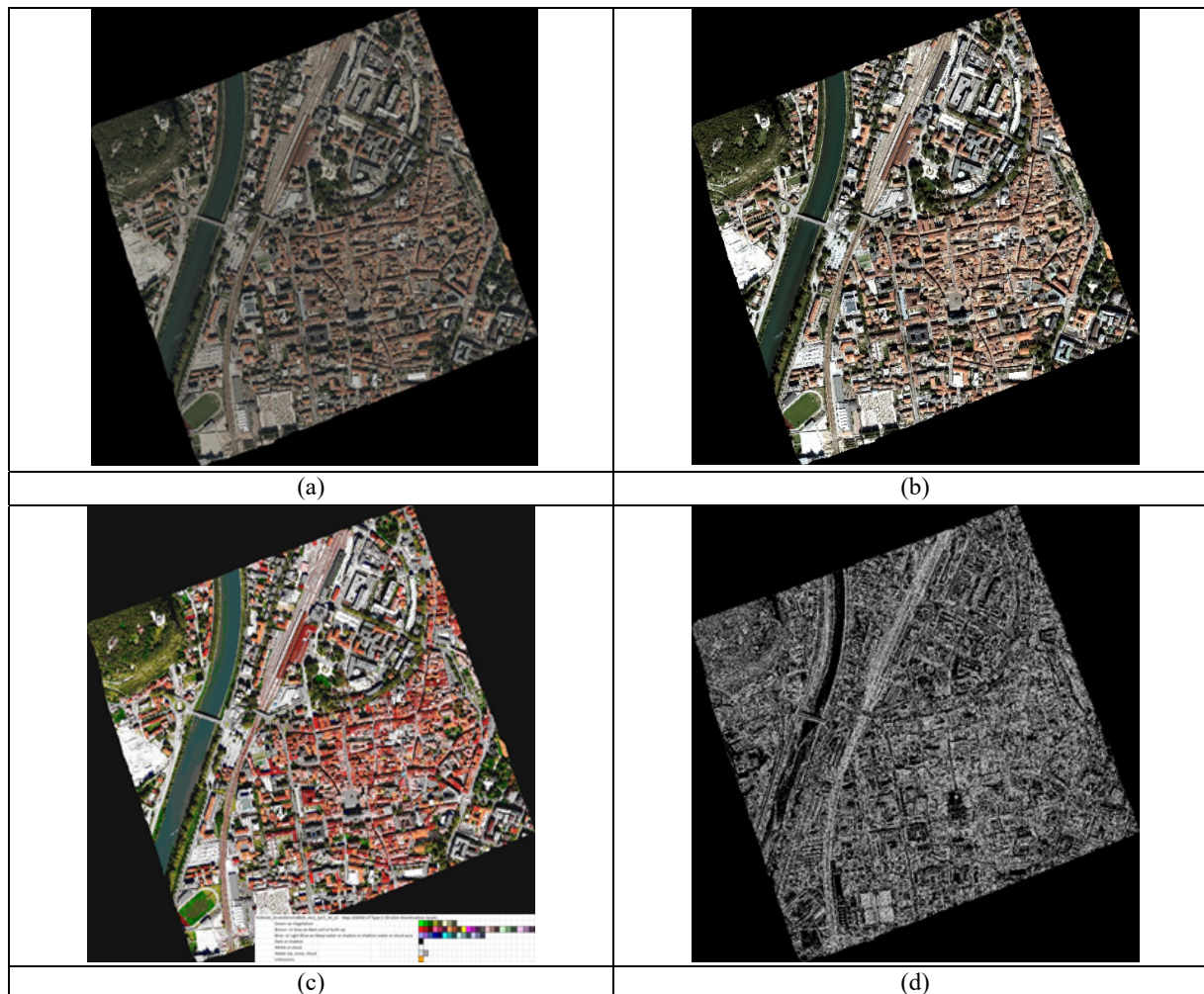


Fig. 4-10. (a) Airborne 10 cm resolution true-color RGB orthophoto of Trento, Italy, 4017 x 4096 x 3 pixels in size, acquired in 2014 and provided with no radiometric calibration metadata file. No histogram stretching for visualization purposes. Courtesy of Bruno Kessler Foundation (FBK), Trento, Italy. (b) Same RGB orthophoto subject to self-organizing statistical color constancy. (c) RGBIAM polyhedralization of the RGB color space and prior color map of the RGB image subject to color constancy. The RGBIAM map legend, consisting of 50 spectral categories, is depicted in pseudocolors, see Fig. 4-11. Input parameters: none. Processing time (one-pass, IDL implementation) = 2 min. (d) To visualize contours of image-segments in the multi-level RGBIAM color map-domain, an automatic 4- or 8-adjacency cross-aura measure is estimated in linear time (refer to Fig. 4-18).

RGBIAM_StndrdStrtchdBGR_r6v2_SpCt_50_12 - Map LEGEND of Type 1: 50 color discretization levels	
Green-as-Vegetation	
Brown- or Gray-as-Bare soil or built-up	
Blue- or Light Blue-as-Deep water or shadow or shallow water or cloud aura	
Dark or shadow	
White or cloud	
Water ice, snow, cloud	
Unknowns	

(a)

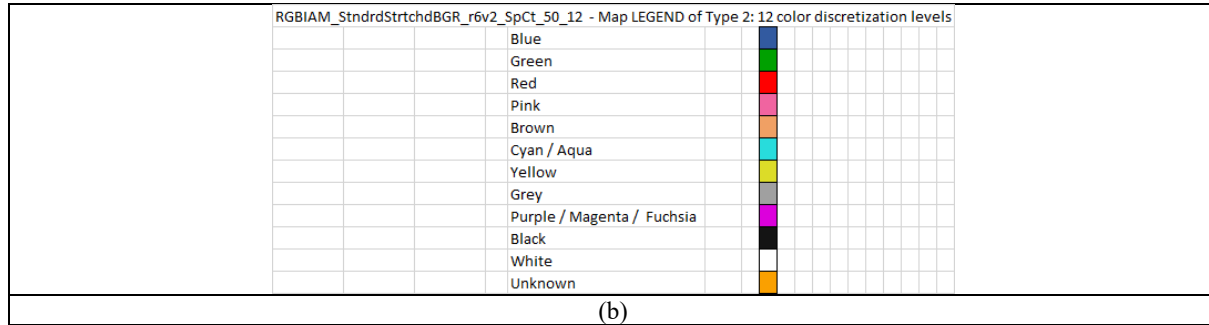


Fig. 4-11. Two-level RGBIAM's quantization of the RGB cube. (a) RGB color map's legend at fine granularity: 49 color names + 1 class unknown. (b) RGB color map's legend at coarse granularity: 11 color names + 1 class unknown, where the coarse vector quantization (VQ) is a mutually exclusive and totally exhaustive combination of the fine VQ.

#### 4.2.3.1. EO image pre-processing for radiometric enhancement

(I) Absolute radiometric *Cal*, in compliance with the QA4EO *Cal* requirements. A battery of sensor-specific EO image radiometric calibrators, e.g., Landsat-4 to Landsat-8, AVHRR, VIIRS, MeteoSat, AVNIR, SPOT-1 to SPOT-7, Pleiades, IRS, DMC, RapidEye, Ikonos, QuickBird, WorldView, etc., is required to transform dimensionless digital numbers into a radiometric unit of measure, specifically, top-of-atmosphere reflectance (TOARF)  $\in [0, 1]$ . Sensory data *Cal* is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” (Schaeppman-Strub et al., 2006). In other words, sensory data *Cal* is a necessary not sufficient condition for both physical model-based and hybrid inference-based data understanding, including the hybrid EO-IU subsystem discussed herein. Whereas physical variables can be investigated by physical, statistical and hybrid inference systems, uncalibrated sensory data can be investigated by statistical data models exclusively. Although they do not require physical variables as input, statistical models can benefit from physical variables, since physical units of measure inherently harmonize data acquired across time, space and sensors. Irrespective of this unquestionable benefit, in the RS literature, where statistical models dominate physical ones, EO data *Cal* is largely ignored, in disagreement with the QA4EO *Cal* requirements (Baraldi & Boschetti, 2012a and 2012b).

(II) Automated “stratified” (“layered”, driven-by-knowledge) approach (Mather, 1994) to atmospheric correction of TOARF into surface reflectance (SURF) values. Atmospheric correction is a typical physical model-based inversion problem (Castelletti et al. 2016) inherently ill-posed in the Hadamard sense (Hadamard, 1902). To become better conditioned for numerical treatment, an atmospheric correction algorithm can be run on class-conditional EO data layers (Vermote and Saleous, 2007; Richter and D. Schlöpfer, 2016), such as MS color names detected by a static decision tree for MS reflectance space polyhedralization (Baraldi et al., 2006). In practice, an unconditional quantitative data distribution, such as an image-wide TOARF distribution, is transformed into class-conditional distributions through prior knowledge-based stratification. A “stratified” or “layered” multivariate data analysis approach (Mather, 1994) complies with the popular divide-and-conquer problem solving criterion and with the principle of statistical stratification (Hunt & Tyrrell, 2012). Well known in statistics, it states that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” (Hunt & Tyrrell, 2012).

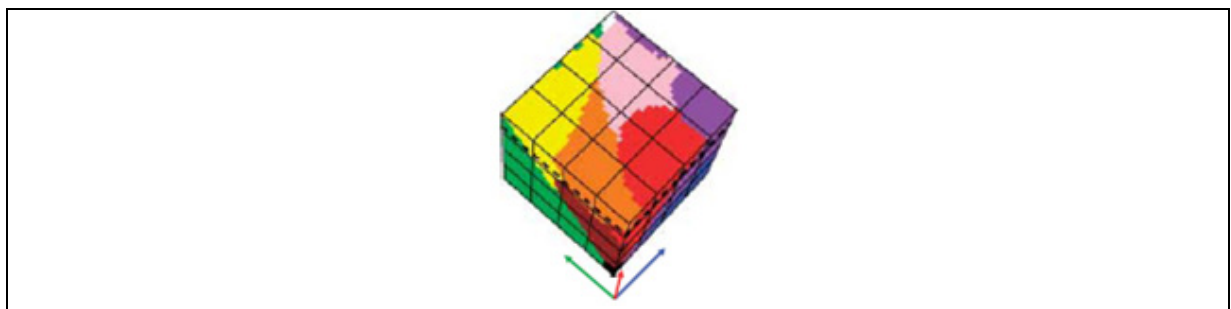


Fig. 4-12. Courtesy of (Griffin, 2006). Unlike a MS space polyhedralization impossible to visualize when the number of channels is superior to three, an RGB data cube polyhedralization is intuitive to display. For example, based on



psychophysical evidence, human basic color (BC) names can be mapped onto the RGB cube. Central to this consideration is Berlin and Kay’s landmark study of a “universal” inventory of eleven BC words in twenty human languages: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown (Berlin and Kay, 1969).

(III) Automated “stratified” topographic correction (TOC). TOC is a well-known chicken-and-egg dilemma. To become better posed for automated numerical solution, an inherently ill-conditioned TOC algorithm is run on informative EO image strata. In line with (Baraldi et al., 2010b), information layers for TOC stratification comprise MS color names, detected by a static decision tree for MS reflectance space polyhedralization (Baraldi et al., 2006), overlapped with image masks inferred from ancillary data. Specifically, a digital surface model (DSM) data set and the EO image metadata parameters sun position and sensor position are required to partition an EO image into three masks: horizontal areas where no TOC is required, slopes-facing-the-sun and slopes-facing-away-from-the-sun. Within each MS color layer, e.g., vegetation, bare soil, etc., stratified TOC occurs between the two image masks slopes-facing-the-sun and slopes-facing-away-from-the-sun (Fig. 4-9).

(IV) Color constancy for non-calibrated MS image inter-channel harmonization. In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions, so that they appear identical to a “canonical” (reference) image subject to a “canonical” (known) light source (of controlled quality), e.g., under a white light source (Gijssen et al., 2010). In short, solution of the color constancy problem is the recovery “of an illuminant-independent representation of the reflectance values in a scene” (Finlayson et al., 2001). In common practice color constancy supports image harmonization and interoperability when no radiometric *Cal* parameter is available (Fig. 4-10). Its goal is analogous to that of absolute radiometric *Cal* of an image, considered mandatory by the QA4EO guidelines when radiometric *Cal* parameters are available (Group on Earth Observation, 2010). Inspired by human vision, a novel self-organizing (autonomous) statistical algorithm for multi-band image color constancy was implemented (patent pending). It is eligible for use with any MS image provided with no radiometric calibration metadata file, such as images typically acquired by consumer-level RGB color cameras, either true- or false-color, mounted onboard Unmanned Aerial Vehicles (UAVs) (Vo et al., 2016).

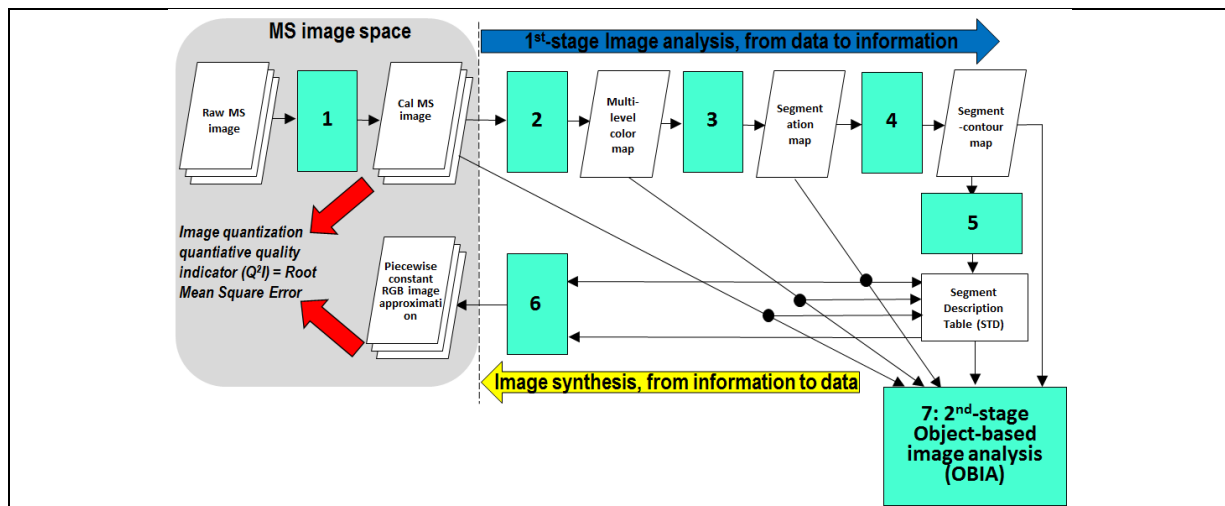


Fig. 4-13. The SIAM lightweight computer program for automated prior knowledge-based MS reflectance space hyperpolyhedralization, superpixel detection and vector quantization (VQ) quality assessment, consisting of boxes 1 to 6. Phase 1-of-2 = Encoding phase/Image analysis - Stage 1: sensor-specific MS data calibration into TOARF or SURF values. Stage 2: Prior knowledge-based SIAM reflectance space partitioning. Stage 3: Well-posed two-pass connected-component multi-level image labeling (Sonka et al., 1994; Dillencourt et al., 1992). Connected-components in the image-domain are connected sets of pixels featuring the same color label. They are also called image-objects, segments or superpixels. Stage 4: Well-posed segment-contour extraction. Stage 5: Well-posed Superpixel Description Table (STD) allocation and initialization. Phase 2-of-2 = Decoding phase/Image synthesis - Stage 6: Superpixelwise-constant input image approximation (‘Object-mean view’) and per-pixel VQ error estimation. (Stage 7: Object-based image analysis (OBIA), in cascade to the SIAM color naming).

SIAM™, r88v6	Input bands	Preliminary classification map output products: Number of output spectral categories			
		Fine discretization levels	Intermediate discretization levels	Coarse discretization levels	Inter-sensor discretization levels (*)
L-SIAM™	7 – B, G, R, NIR, MIR1, MIR2, TIR	96	48	18	33 * employed for inter-sensor post-classification change/no-change detection
S-SIAM™	4 – G, R, NIR, MIR1	68	40	15	
AV-SIAM™	4 – R, NIR, MIR1, TIR	83	43	17	
AA-SIAM™	5 – G, R, NIR, MIR1, TIR	83	43	17	
Q-SIAM™	4 – B, G, R, NIR	61	28	12	
D-SIAM™	3 – G, R, NIR	61	28	12	

Table 4-1. SIAM is an EO system of systems, in compliance with the GEOSS guidelines (Group on Earth Observation, 2005). SIAM consists of six subsystems, to be input with MS images of different spectral resolution acquired by any past, present or future MS imaging sensor radiometrically calibrated into TOARF or SURF values (Fig. 4-14). The 7-band Landsat-like SIAM (L-SIAM) is the “master” SIAM decision tree. “Slave” SIAM decision trees are derived from the “master” L-SIAM when the MS sensor’s spectral resolution overlaps with Landsat’s, but is inferior to Landsat’s. “Slave” SIAM implementations are the 4-band SPOT-like SIAM (S-SIAM), 4-band AVHRR-like SIAM (AV-SIAM), 5-band AATSR-like SIAM (AA-SIAM), 4-band QuickBird-like SIAM, and 3-band DMC-like SIAM. Depending on the informativeness of the MS sensor’s spectral resolution, three different levels of polyhedralization of the MS reflectance space are implemented at, respectively, fine, intermediate and coarse granularity.

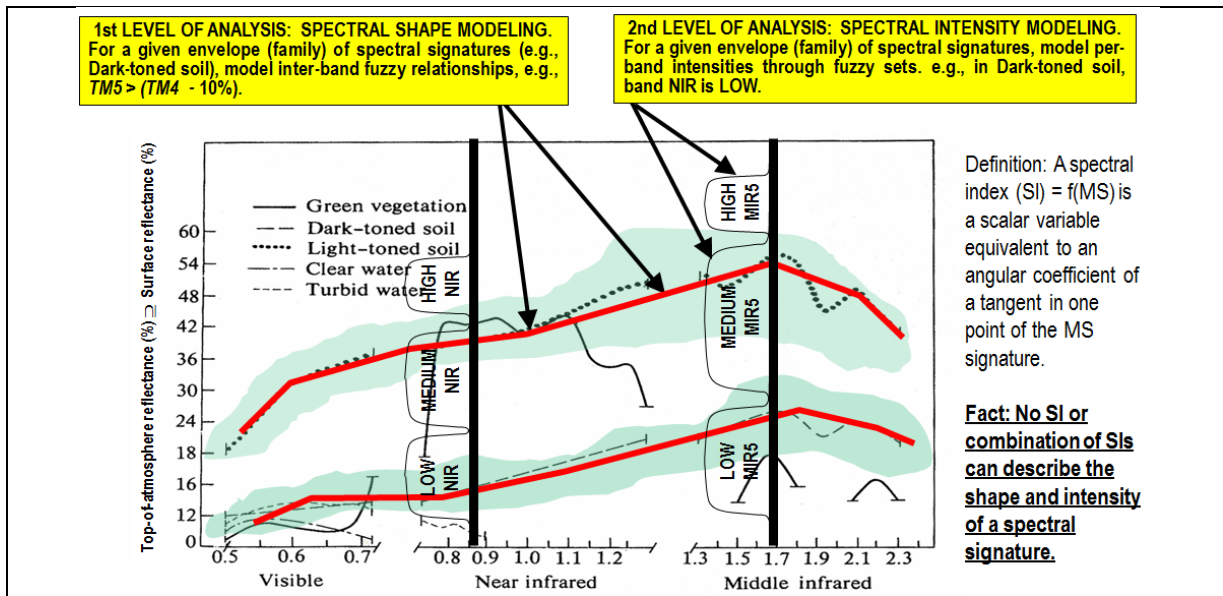


Fig. 4-14. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF values forms a buffer zone (hyperpolyhedron, envelope, manifold) which includes surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions. Like a vector quantity has two characteristics, a magnitude and a direction, any LC class-specific MS manifold is characterized by a multivariate shape and a multivariate intensity information component. A typical spectral index (SI) is a scalar band ratio or inter-band difference conceptually equivalent to an angular coefficient of a tangent in one point of the spectral signature. Infinite functions can feature the same tangent value in one point. In practice, no univariate spectral index or multivariate combination of spectral indexes can reconstruct the multivariate shape and multivariate intensity of a spectral signature.

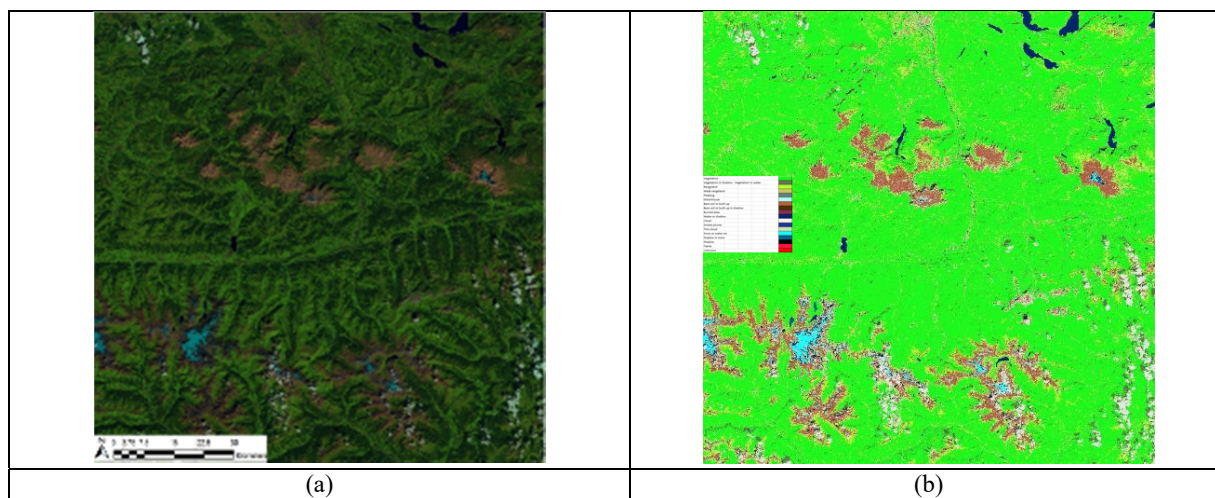
#### 4.2.3.2. Pre-attentive vision: Raw and full primal sketch

(I) Raw primal sketch, pixel-based MS color naming in a measurement space of either TOARF or SURF values. Color is the sole visual property available at pixel resolution. To mimic fuzzy human reasoning based on fuzzy sets including



color names, pixel-based values in a numeric color space can be partitioned into hyperpolyhedra corresponding to a finite and discrete dictionary of basic color (BC) names to be community-agreed upon in advance (Griffin, 2006), see Fig. 4-12. Presented in the RS literature in recent years (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi and Boschetti, 2012a and 2012b), the Satellite Image Automatic Mapper (SIAM) software product is an expert system (prior knowledge-based decision tree) for physical model-based/deductive/top-down vector quantization (VQ) and VQ quality assessment in a MS reflectance space (Fig. 4-13). Since it is a physical data model, SIAM requires as input EO data *Cal* into either TOARF or SURF values where, for every LC class in the real-world domain, SURF values are a special case of TOARF values in clear sky and flat terrain conditions (Fig. 4-14). By definition an expert system, including SIAM, relies exclusively on *a priori* knowledge available in addition to data; hence, any expert system is fully automated, i.e., it requires neither user-defined parameters nor training data to run. In the VQ encoding phase, SIAM partitions the MS reflectance space into static (non-adaptive to data) hyperpolyhedra, not necessarily convex or connected, equivalent to a dictionary (codebook) of MS color names (codewords), see Table 4-1. Each MS pixel is mapped onto one MS hyperpolyhedron associated with a BC name. Unfortunately, when the MS space dimensionality is greater than three, a prior dictionary of mutually exclusive and totally exhaustive hyperpolyhedra is difficult to think of and impossible to visualize (Fig. 4-12). Next, for image analysis purposes a 2D multi-level VQ map is automatically generated (Fig. 4-15), together with its segmentation map (Fig. 4-17). To visualize contours of image segments, an automatic 4- or 8-adjacency cross-aura measure is estimated in linear time (Fig. 4-18). In the VQ decoding phase, equivalent to image synthesis generated via segmentwise-constant MS image reconstruction (Fig. 4-13), also known as “image-object mean view” in OBIA (ITT, 2009), a VQ quality assessment is accomplished in compliance with the QA4EO *Val* requirements. In VQ problems, a typical community-agreed  $Q^2I$  is the per-vector encoding-decoding Euclidean distance (Cherkassky and Mulier, 1998). In practice, the SIAM expert system automatically detects texels as connected sets of pixels featuring the same color name (Julesz et al., 1973). In CV applications, traditional texels have been recently renamed superpixels, to be detected by semi-automatic inductive data learning algorithms where at least two free-parameters are user-defined based on heuristics (Achanta et al., 2011). Texels automatically detected by SIAM can be input to the full primal sketch for texture detection.

(II) Raw primal sketch, pixel-based RGB color naming in an uncalibrated true- or false-color RGB cube. The RGB Image Automatic Mapper (RGBIAM) is a novel expert system (patent pending) capable of partitioning a three-band RGB data cube, either true- or false-color, into a pre-defined dictionary of RGB color names (Baraldi et al., 2017). The RGBIAM software design is the same as SIAM’s (Fig. 4-13). When a true- or false-color space dimensionality is equal to three, a prior dictionary of mutually exclusive and totally exhaustive polyhedra, not necessarily convex or connected, is intuitive to think of and easy to visualize (Fig. 4-12). The novel RGBIAM’s static decision tree requires as input a true- or false-color RGB image subject to color constancy for inter-channel harmonization (Fig. 4-10), in place of the mandatory QA4EO *Cal* requirements (Fig. 4-19). Texels automatically detected by RGBIAM can be input to the full primal sketch for texture detection.





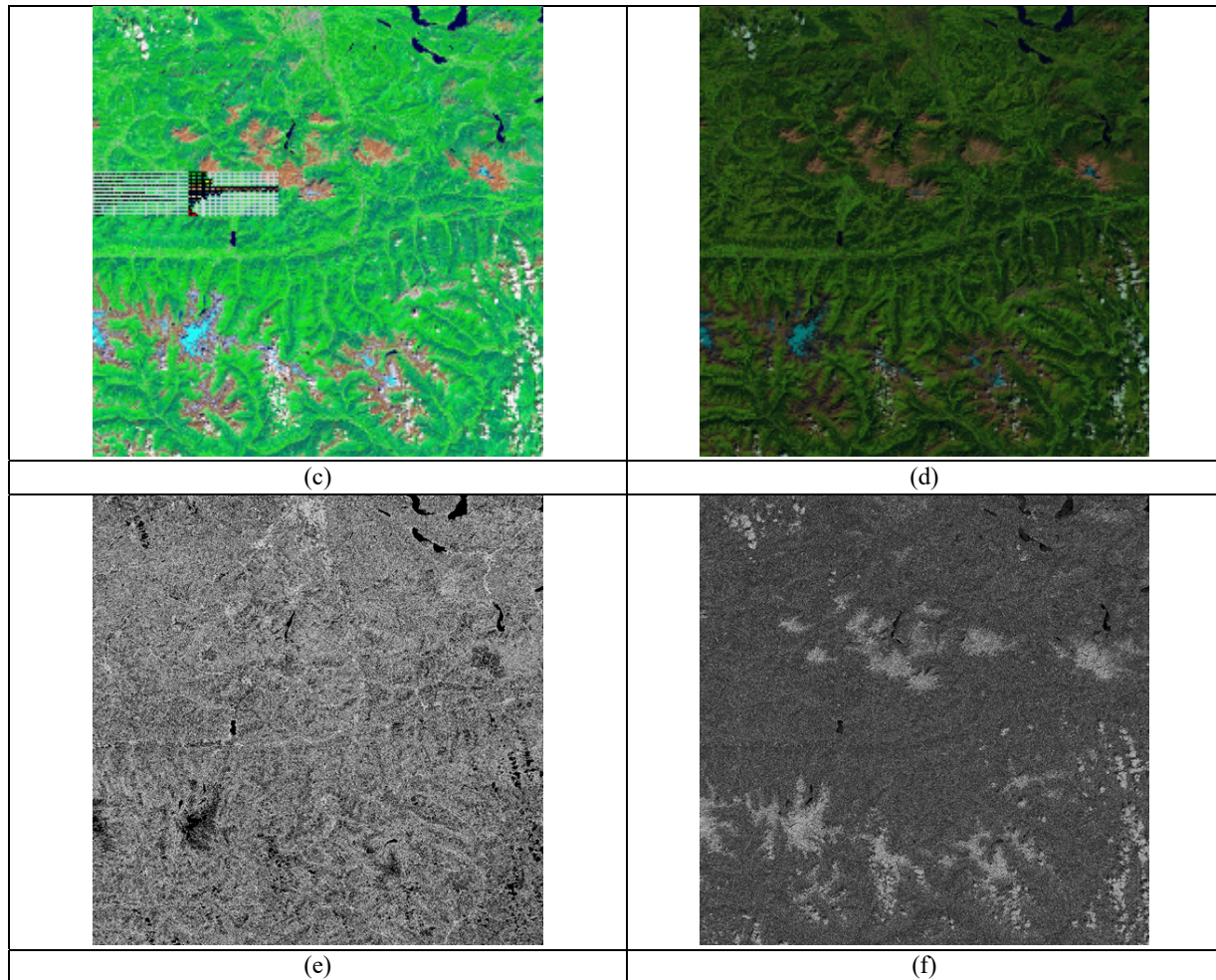


Fig. 4-15. (a) Sentinel-2A MSI Level-1C image of the Earth surface, located south of the city of Salzburg, Austria. The city area is visible around the middle of the image upper boundary (Lat-long coordinates: 47°48'25.0"N 13°02'43.6"E). Acquired on 2015-09-11. Spatial resolution: 10 m. Image size: 110×110 km. Radiometrically calibrated into TOARF values in range  $\{0, 255\}$ , it is depicted as a false color RGB image, where: R = Medium InfraRed (MIR) = Band 11, G = Near IR (NIR) = Band 8, B = Blue = Band 2. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at coarse color granularity, consisting of 18 spectral categories depicted in pseudo colors, refer to the map legend shown in Fig. 4-16. Coarse-granularity color categories are generated by merging color hyperpolyhedra at fine color granularity, according to pre-defined parent-child relationships. (c) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, refer to the map legend shown in Fig. 4-16. (d) Superpixelwise-constant approximation of the input image (“image-object mean view”) generated from the L-SIAM’s 96 color map at fine granularity. Depicted in false colors: R = MIR = Band 11, G = NIR = Band 8, B = Blue = Band 2. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (e) 8-adjacency cross-aura contour map in range  $\{0, 8\}$  automatically generated from the L-SIAM’s 96 color map at fine granularity. It shows contours of connected sets of pixels featuring the same color label. These connected-components are also called image-objects, segments or superpixels. (f) Per-pixel scalar difference between the input MS image shown in (a) and the superpixelwise-constant MS image reconstruction shown in (d). This scalar difference is computed as the per-pixel Root Mean Square Error (RMSE) in range  $\{0, 255\}$ . The RMSE is a well-known vector quantization (VQ) error. Image-wide basic statistics: Min = 0, Max = 130, Mean = 2.60, Stdev = 3.45. Histogram stretching is applied for visualization purposes.

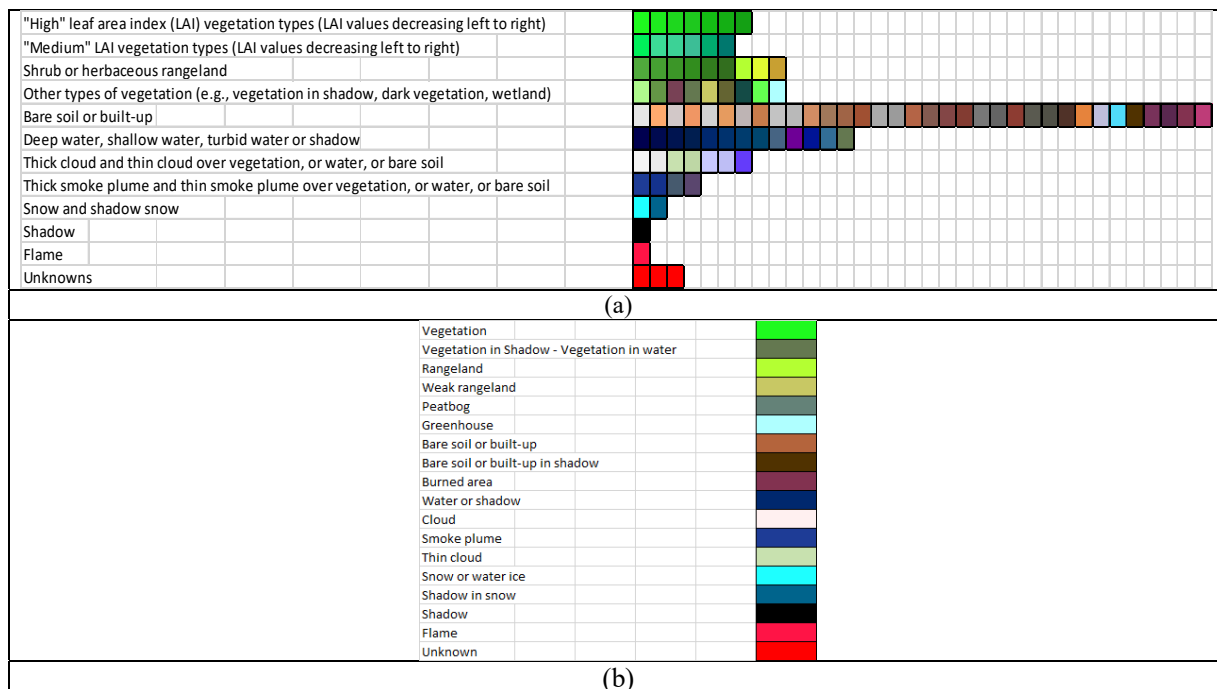


Fig. 4-16. (a) L-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at fine granularity. Pseudocolors of 96 spectral categories are gathered based on their spectral end member (e.g., bare soil or built-up) or parent spectral category (e.g., "high" leaf area index vegetation). The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that MS hyperpolyhedron. (b) L-SIAM™ lightweight computer program, r88v6. Multi-spectral (MS) reflectance space hyperpolyhedralization into MS color names at coarse granularity. The coarse vector quantization (VQ) is a mutually exclusive and totally exhaustive combination of the fine VQ. Pseudocolors of the 18 spectral categories, equivalent to parent spectral categories, are chosen as to mimic natural colors of pixels belonging to each MS hyperpolyhedron.

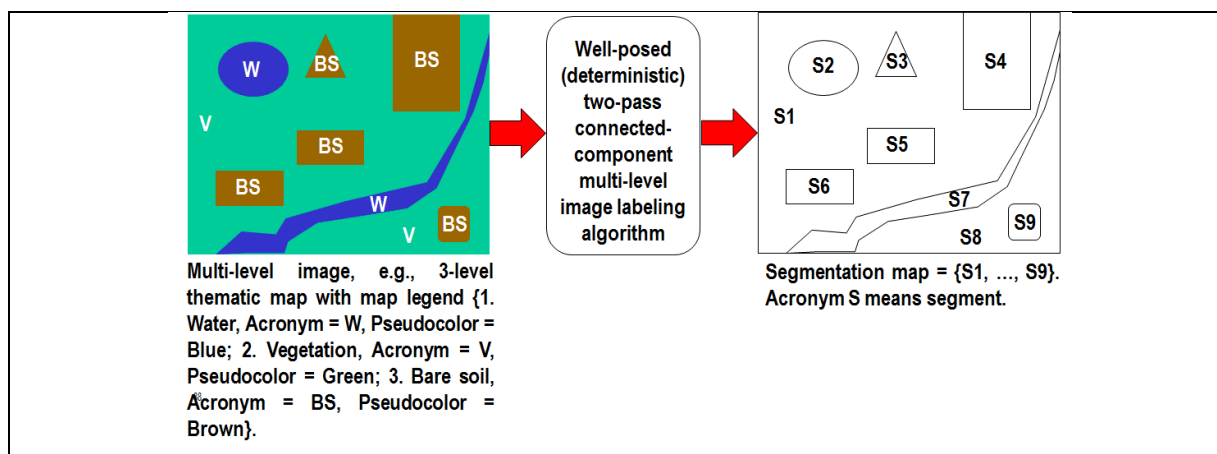


Fig. 4-17. One segmentation map is deterministically generated from one multi-level image, such as a thematic map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. In this example, nine image-objects/segments S1 to S9 can be detected in the 3-level thematic map shown at left. Each segment consists of a connected set of pixels sharing the same thematic map label (Sonka et al., 1994; Dillencourt et al., 1992). Each stratum/layer/level consists of one or more segments, e.g., stratum Vegetation (V) consists of the two disjoint segments S1 and S8. In any multi-level (categorical, nominal) image domain, three spatial primitives co-exist and are provided with parent-child relationships: pixel (row-column coordinate pair) with a parent segment identifier (ID) and a super-parent level ID, segment (polygon) with a segment ID and a parent level ID, and a level/stratum (multi-part polygon) with a level ID.



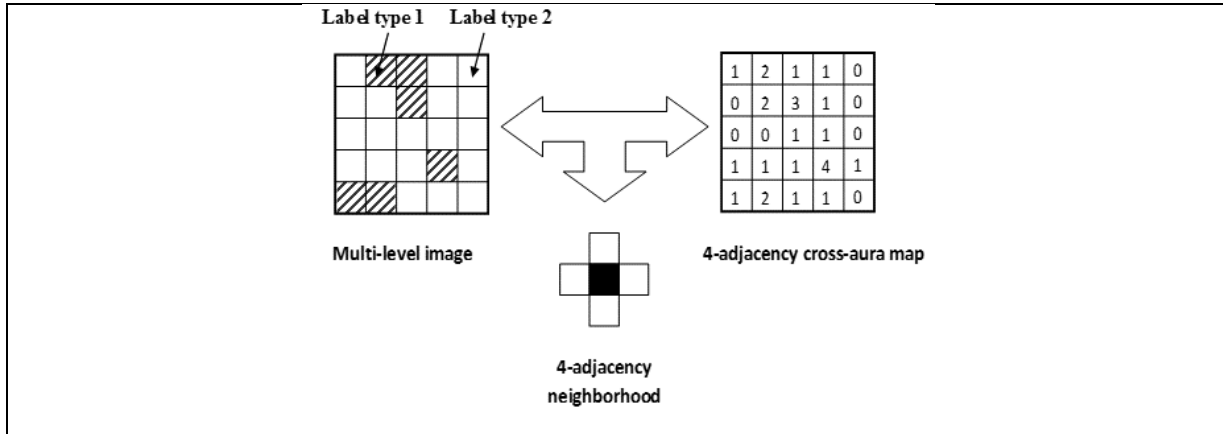


Fig. 4-18. Example of a 4-adjacency cross-aura map, shown at right, generated from a two-level image, shown at left.

(III) Raw primal sketch, pixel-based MS color naming through time in a measurement space of either TOARF or SURF values. When two single-date image-derived SCMs of the same Earth surface area are available and co-registered at the same spatial resolution, post-classification through time is straightforward according to a two-way contingency table, where post-classification LC change/no-change (LCC) overall accuracy,  $OA-LCC_{1,2} \in [0, 1]$ , is such that

$$OA-LCC_{1,2} \leq (OA-LC_1 \times OA-LC_2), \text{ where } OA-LC_1 \text{ and } OA-LC_2 \in [0, 1].$$

For example, if  $OA-LC_1 = 0.90 = OA-LC_2$ , then  $OA-LCC_{1,2} \leq 0.81$ . It means that post-classification change/no-change detection-through-time is recommended if and only if single-date map accuracies score “high”. According to literature, this requirement is accomplished by the SIAM maps. To extend the SIAM color naming to the time domain, specifically, to develop a bi-temporal SIAM-based post-classification LC change/no-change detection (Fig. 4-20), a two-way square LCC contingency table (transition matrix) was defined, where the two input categorical variables at time T1 (corresponding to table rows) and time T2 > T1 (corresponding to table columns) are the same multi-source SIAM map’s legend of 33 color names (refer to Table 4-1). The implemented bi-temporal SIAM-based post-classification LC change/no-change detection map’s legend features 29 LC change/no-change classes (Fig. 4-21).

(IV) Raw primal sketch, deterministic/well-posed (hence, automated) two-pass connected-component multi-level image labeling. Any SCM whose legend is a discrete and finite dictionary of LC class labels is a multi-level image (Sonka et al., 1994; Dillencourt et al., 1992). As such, it can be partitioned (segmented) into planar spatial units (image-objects), either (0D) pixel, (1D) line or (2D) polygon (Open Geospatial Consortium, 2015), where each image-object is a connected set of pixels featuring the same LC class label, by a well-posed (deterministic) two-pass connected component multi-level image labeling algorithm (Sonka et al., 1994; Dillencourt et al., 1992) (Fig. 4-17). Each SCM-derived planar object in the image domain is provided with both a segment identifier and an LC class label in the SCM’s taxonomy. For example, semantic change/no-change through time of spatial units, either 0D pixel, 1D line or 2D polygon, can be tracked by the EO-IU4SQ inference engine, in compliance with an OBIA-through-time paradigm (Blaschke et al., 2014) as a viable alternative to context-insensitive (pixel-based) image analysis-through-time traditionally adopted by EO-IUSs (Mather, 1994).

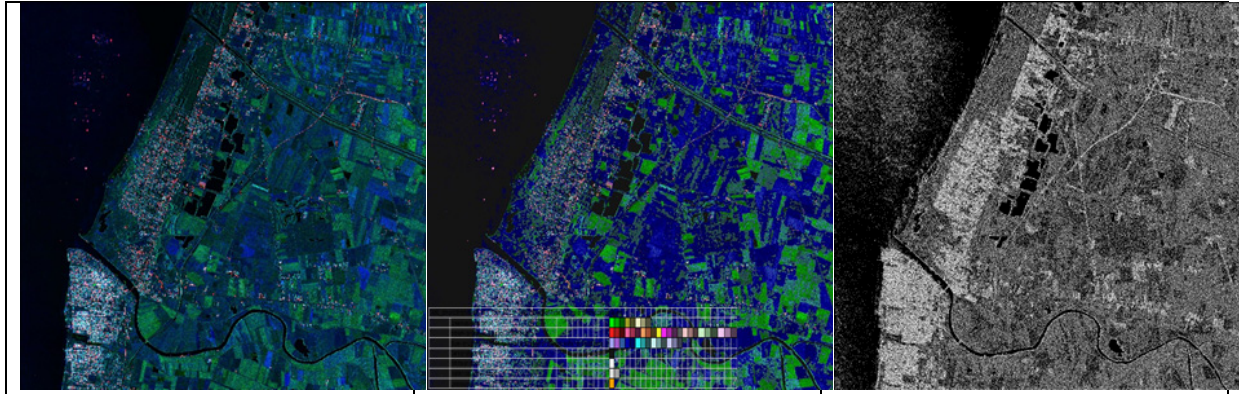


Fig. 4-19. Left: False-color RGB image subject to preliminary color constancy. It shows three heterogeneous channels generated from a bi-temporal pair of SAR COSMO-SkyMed images of Caserta (Italy). Band R = Coherence magnitude (in range  $[0, 1]$ ), band G = Intensity of a test image, acquired in high-leaf season, B = Intensity (energy ratio, in range  $[0, 1]$ ) of a reference image, acquired in low-leaf season. Courtesy of Università degli Studi di Napoli Federico II. No histogram stretching is applied for visualization purposes. Center: RGBIAM color space discretization into 50 polyhedra associated with color names depicted in pseudocolors (Baraldi et al., 2017). The map legend, overlapped to the map product, is shown in Fig. 4-11. Right: 8-adjacency cross-aura contour map, with contour values in range  $\{0, 8\}$  (refer to Fig. 4-21). These contours visualize boundaries of segments, equivalent to texture elements (texels, superpixels), automatically detected in the RGBIAM's color map by a well-posed two-pass connected-component multi-level image labeling algorithm (Sonka et al., 1994; Dillencourt et al., 1992).

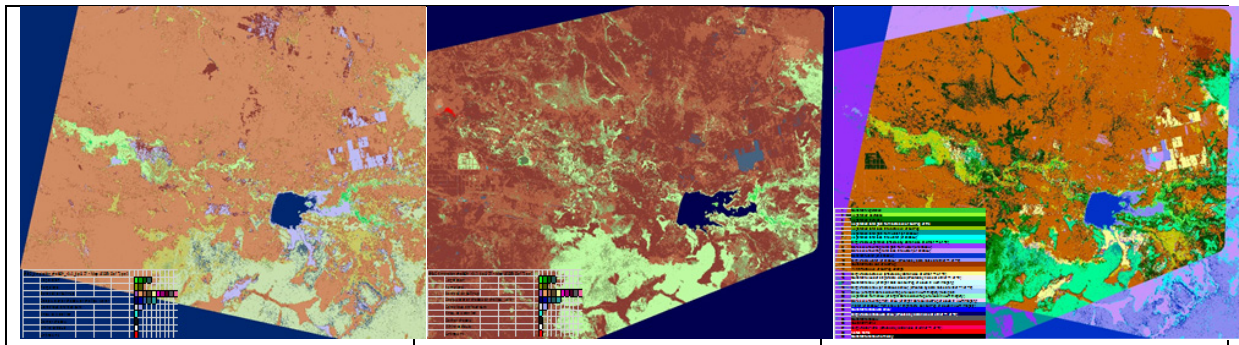


Fig. 4-20. Automated SIAM-based post-classification change/no-change detection. Left: SIAM map depicted in 33 pseudo colors, generated from a SPOT-5 image in TOARF values, upscaled to 5 m resolution. Center: SIAM map depicted in 33 pseudo colors, generated from a RapidEye image in TOARF values, 5 m resolution. Left: Multi-source SIAM-based post-classification change/no-change detection map, depicted in pseudocolors, according to a two-way  $33 \times 33$  contingency table. The bi-temporal SIAM-based post-classification map legend is shown in Fig. 4-21.



1	Constant vegetation
2	Vegetation decrease
3	Vegetation increase
4	Vegetation total gain from bare soil or built-up or fire
5	Vegetation total loss into bare soil or built-up
6	Vegetation total gain from water (or shadow)
7	Vegetation total loss into water (or shadow)
8	Single-date vegetation (affected by data noise at either T1 or T2)
9	Bare soil or built-up total gain from water (or shadow)
10	Bare soil or built-up total loss into water (or shadow)
11	Constant water (or shadow)
12	Single-date water (or shadow) (affected by data noise at either T1 or T2)
13	Constant bare soil or built-up
14	Within-bare soil or built-up change
15	Single-date bare soil (affected by data noise at either T1 or T2)
16	Constant cloud or single-date cloud (affected by noise at either T1 or T2)
17	Constant snow (or bright bare soil/built-up or cloud in VHR imagery)
18	Single-date snow (or shadowed snow) (affected by data noise at either T1 or T2)
19	Snow (or bright bare soil/built-up or cloud in VHR imagery) total gain
20	Vegetation from snow (or bright bare soil/built-up or cloud in VHR imagery)
21	Bare soil or built-up from snow (or bright bare soil/built-up or cloud in VHR imagery)
22	Water (or shadow) from snow (or bright bare soil/built-up or cloud in VHR imagery)
23	Constant shadowed snow
24	Single-date shadowed snow (affected by data noise at either T1 or T2)
25	Constant shadow
26	Constant flame
27	Single-date flame (affected by data noise at either T1 or T2)
28	Active flame
29	Constant unknown or noisy

Fig. 4-21. Bi-temporal SIAM-based post-classification map legend, consisting of 29 change/no-change LC classes-through-time, depicted with pseudocolors.

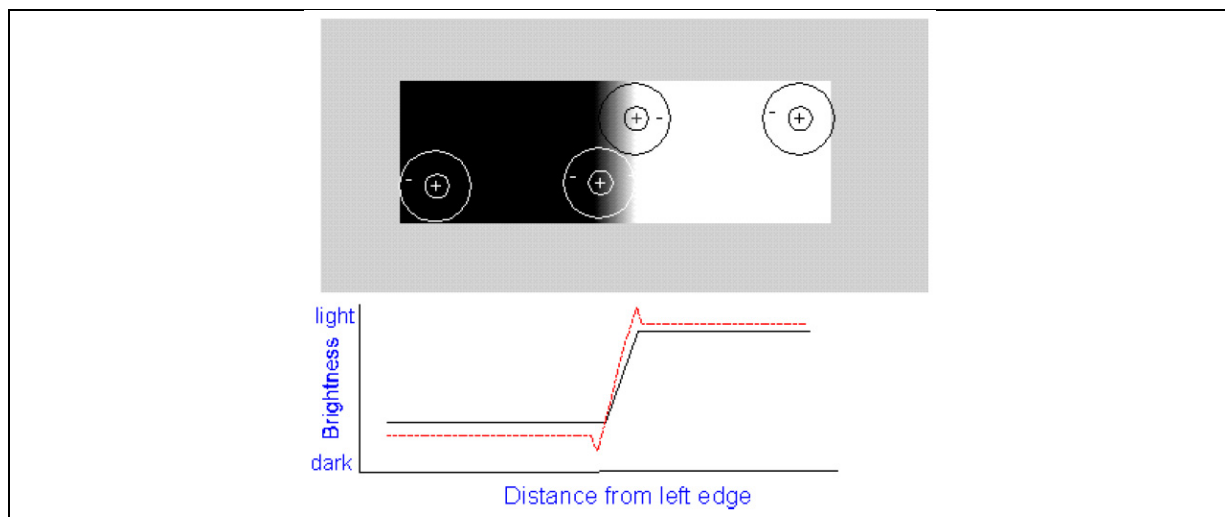


Fig. 4-22. Mach bands illusion. In black: Ramp in luminance units across space. In red: Brightness (perceived luminance) across space. One of the best-known brightness illusions (where brightness is defined as a subjective aspect of vision, i.e., brightness is the perceived luminance of a surface) is the psychophysical phenomenon of the Mach bands: where a luminance (radiance, intensity) ramp meets a plateau, there are spikes of brightness, although there is no discontinuity in the luminance profile. Hence, human vision detects two boundaries, one at the beginning and one at the end of the ramp in luminance. Since there is no discontinuity in luminance where brightness is spiking, the Mach bands effect is called a visual “illusion”. Along a ramp, no image-contour is perceived by human vision, irrespective of the ramp’s local contrast (gradient) in range  $(0, +\infty)$ . In the words of Pessoa, “if we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands which are seen at ramp edges, the number of published models is surprisingly small” (Pessoa, 1996). In statistical 2D signal processing the lesson to be learned from the Mach bands illusion is that local variance, contrast and first-order derivative (gradient) are statistical features (data-derived numeric variables) computed locally in the (2D) image-domain NOT suitable to detect image-objects (segments, closed contours) required to be perceptually “uniform” (“homogeneous”). In other words, these popular local statistics are not suitable visual features if detected image-segments/image-contours are required to be consistent with human visual perception, including ramp-edge detection. This straightforward (obvious), but not trivial observation is in contrast with a large portion of existing literature, where many semi-automatic image segmentation/image-contour detection algorithms are based on thresholding the local



variance, contrast or gradient (Baatz et al., 2000; Espindola et al., 2006; Trimble, 2015; Camara et al., 1996; Canny, 1986), where a system-free threshold parameter  $\in (0, +\infty)$  must be user-defined based on heuristics.

(V) Raw primal sketch, innovative automated wavelet-based zero-crossing (ZX) contour-pixel detection followed by an automated ZX segment detection, consistent with the Mach bands illusion (Pessoa, 1996). In the words of Pessoa, “if we require that a CV system should be able to predict perceptual effects, such as the well-known Mach bands illusion where bright and dark bands are seen at ramp edges (Fig. 4-22), then the number of published vision models becomes surprisingly small” (Pessoa, 1996). In statistical 2D signal processing the unquestionable fact to be learned from the Mach bands illusion is that local variance, local contrast and local first-order derivative (gradient) are statistical features (data-derived numeric variables) computed locally in the (2D) image-domain, e.g., within a moving window or within an image-object or by means of a 2D spatial filter with a finite support, NOT suitable to detect image-objects (segments, closed contours) required to be perceptually “uniform” (“homogeneous”). In other words, these popular local statistics are not suitable visual features if detected image-segments/image-contours are required to be consistent with human visual perception, including ramp-edge detection. This observation (true-fact) can be considered straightforward (obvious), but not trivial. It is in contrast with a large portion of existing CV and RS literature, where many heuristic semi-automatic image segmentation/image-contour detection algorithms inconsistent with the Mach bands illusion do detect image-segments/image-contours by thresholding either a local estimate of the gray-level first-order derivative (gradient, angular coefficient of the tangent to the 2D surface in one point), for example a local gradient estimated by means of an odd-symmetric spatial filter such as the Canny edge detector (Canny, 1986), or by thresholding the local variance or local contrast estimated within a moving window or within an image-region, such as in image-region growing algorithms proposed by Baatz et al. (2000) and by Espindola et al. (2006), adopted respectively by the well-known eCognition commercial software product (Trimble, 2015) and the freeware SPRING software for RS image processing (Camara et al., 1996). Inconsistent with the Mach bands illusion, these traditional edge detectors and image-region growing algorithms are semi-automatic because they depend on a system-free local variance, contrast or gradient threshold  $\in (0, +\infty)$  to be user-defined based on heuristics. For example, in the eCognition commercial software (Trimble, 2015), the infamous spatial “scale parameter” to be user-defined is nothing else than a system-free within-segment variance threshold  $\in (0, +\infty)$ . If it is relaxed, then image-regions grow larger (at a coarser spatial scale). Their perceptual inconsistency with the Mach bands illusion explains why image-segments detected by semi-automatic image-region growing algorithms based on heuristics (Baatz et al., 2000; Espindola et al., 2006), as well as image-contours detected by semi-automatic edge detectors based on heuristics (Canny, 1986), often appear counter-intuitive, i.e., inconsistent with human vision to a varying degree depending on the combination of two random variables, the ever-varying complexity of an image as a peculiar combination of four spatial primitives, specifically flat area, step-edge, line and ramp-edge, with a user-defined local variance, contrast or gradient threshold value  $\in (0, +\infty)$ . Alternative to the aforementioned semi-automatic image segmentation/image-contour detection algorithms based on heuristics inconsistent with the Mach bands illusion, an original multi-scale multi-orientation quasicomplete spatial wavelet transform (Nestares et al., 1998), eligible for image analysis, coding and synthesis (Baraldi and Parmiggiani, 1996), consistent with the Mach bands illusion, is proposed for automated ZX contour-pixel detection, followed by an original automated physical model-based image segmentation algorithm for ZX segment detection. To our best knowledge, ZX segments mentioned by Marr in his seminal work (Marr, 1982) have never been automatically generated from ZX contours (Bertero et al., 1988; Torre and Poggio, 1986; Yuille and Poggio, 1986). In (Baraldi and Parmiggiani, 1996), to provide a best-fit of a simple cell of the cat visual cortex, a physical model-based relationship  $F \times \sigma = 0.22$  was proposed between the Gabor elementary function parameters  $\sigma$  and  $F$ , where  $\sigma$  is the spread of a Gaussian function modulated by a complex sinusoid with central frequency  $F$  (Fig. 4-23). Intuitively, a zero DC-component even-symmetric real part of a 1D Gabor mother-wavelet with  $F \times \sigma = 0.22$  can be considered equivalent to a 2nd-order derivative of a 1D Gaussian ( $G$ ) function,  $\partial^2 G / \partial x^2$ . A zero DC-component odd-symmetric imaginary part of a 1D Gabor mother-wavelet with  $F \times \sigma = 0.22$  can be considered equivalent to a 1st-order derivative of a 1D Gaussian function,  $\partial G / \partial x$ . It is important to remind that the 2nd-order derivative  $\partial^2 / \partial x^2$  is a nonlinear operator which neither commutes nor associates with the convolution (Torre and Poggio, 1986). Therefore,

$$(\partial^2 G / \partial x^2 \bullet f(x)) \neq \partial^2 / \partial x^2 (G(x) \bullet f(x)).$$

To retain consistency with the Mach bands illusion starting from 1D signals, Fig. 4-24 shows a 1D function  $f(x)$  comprising flat areas, ramps, step edges and lines, where the following local spatial filters are applied.

- 3 pixel-wide zero DC-component 1D odd-symmetric filter:  $(+1, 0, -1)$ .



- 3 pixel-wide zero DC-component 1D even-symmetric filter: (-0.5, 1, -0.5).
- 3 pixel-wide 1D Gabor filter: (+0.25, 0.5, +0.25).

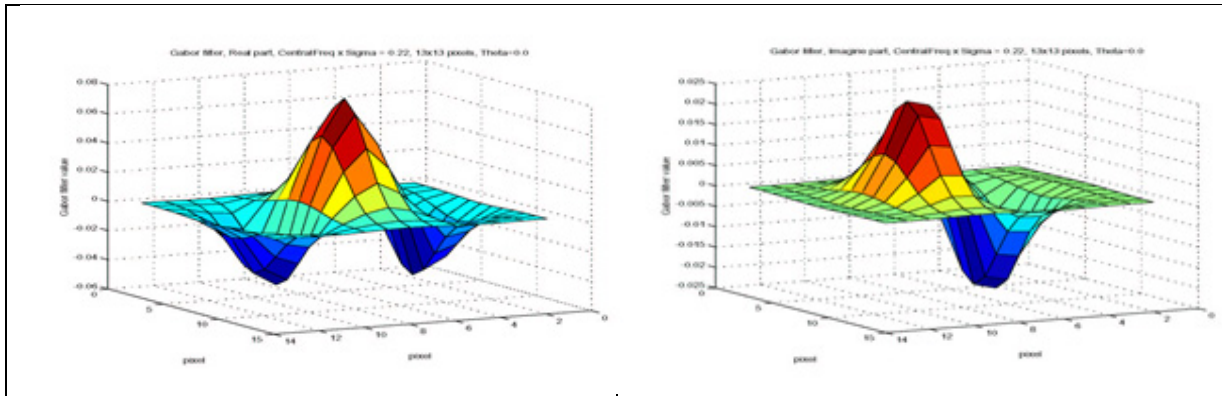
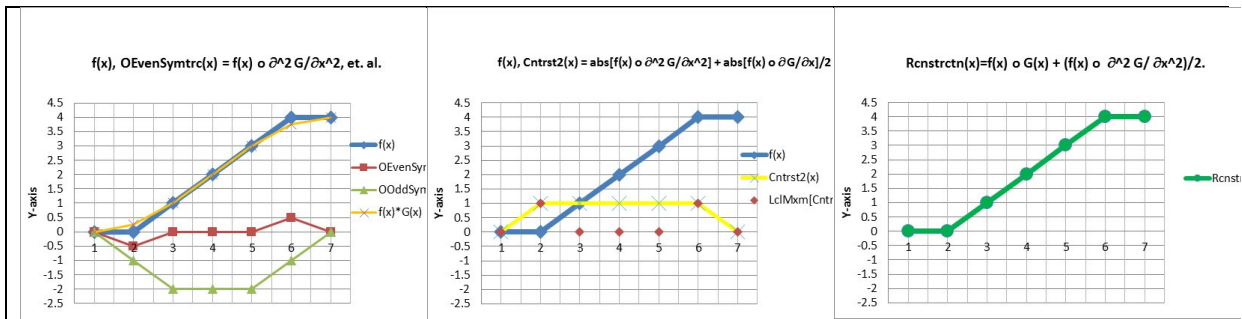
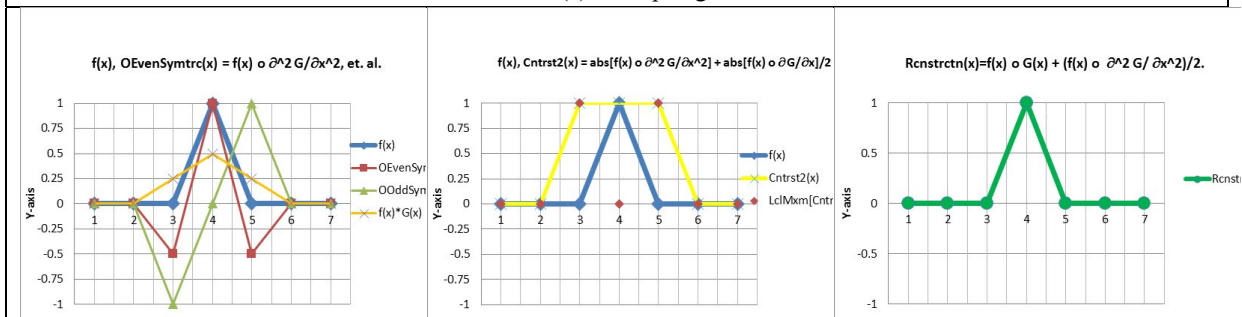


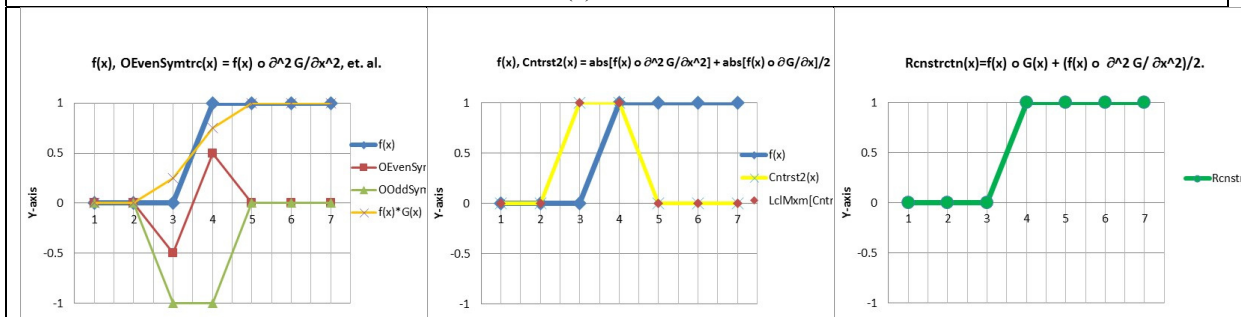
Fig. 4-23. Best-fitting 2-D Gabor elementary function for a simple cell of the cat visual cortex, where the Gaussian spread  $\sigma$  and the central frequency  $F$  of the complex sinusoid are related by the physical model-based relationship  $F \times \sigma = 0.22$ . The zero DC-component filter's even-symmetric real part can be considered equivalent to a 2nd-order derivative of a Gaussian function,  $\partial^2 G/\partial x^2$ . The zero DC-component filter's odd-symmetric imaginary part can be considered equivalent to a 1st-order derivative of a Gaussian function,  $\partial G/\partial x$ .



(a) Ramp edge



(b) White line







(c) Step edge

Fig. 4-24. The proposed multi-scale wavelet-based image decomposition/reconstruction algorithm encompasses an original unified approach to: (i) image-contour detection and (ii) keypoint (endpoint, corner, junction) detection. A 1D synthetic signal decomposition/reconstruction is considered a necessary not sufficient pre-condition for (2D) image analysis. Any synthetic 1D or 2D signal is a combination of four spatial primitives: flat area, step-edge, line and ramp-edge. In 1D signal processing, the proposed 3 pixel-wide 1D zero DC-component odd-symmetric filter profile is  $\partial G/\partial x = (+1, 0, -1)$ ; a 3 pixel-wide 1D zero DC-component even-symmetric filter is  $\partial^2 G/\partial x^2 = (-0.5, 1, -0.5)$ ; a 3 pixel-wide Gaussian filter profile is  $G(x) = (0.25, 0.50, 0.25)$ . A perfect (lossless) 1D signal reconstruction is provided by equation  $\text{Rcnstrct}(x) = f(x) \bullet G(x) + [f(x) \bullet \partial^2 G/\partial x^2]/2$ , where a Gaussian filter output,  $f(x) \bullet G(x)$ , is combined with an even-symmetric filter output,  $f(x) \bullet \partial^2 G/\partial x^2$ . The even-symmetric filter  $\partial^2 G/\partial x^2$  computes a local concavity value,  $f(x) \bullet \partial^2 G/\partial x^2$ , alternative to a traditional local gradient estimate,  $f(x) \bullet \partial G/\partial x$ . The even-symmetric filter  $\partial^2 G/\partial x^2$  is a necessary and sufficient local spatial operator to detect all kinds of 1D contour pixels, where the signal boundary position is localized by a zero-crossing (ZX) pixel defined according to Marr (1980).

The following original considerations stem from Fig. 4-24.

- ✓ Function reconstruction  $\text{Rcnstrct}(x) = f(x) \bullet G(x) + [f(x) \bullet \partial^2 G/\partial x^2]/2$ , where a Gaussian filter is combined with an even-symmetric filter, provides a perfect (lossless) signal reconstruction.
- ✓ An even-symmetric filter  $\partial^2 G/\partial x^2$  is necessary and sufficient to detect all kinds of 1D contour pixels shown in Fig. 4-24, where the signal boundary position is localized by a ZX pixel defined as follows, in compliance with the Mach bands illusion.
  - A ZX pixel is located where the output convolutional value  $[f(x) \bullet \partial^2 G/\partial x^2]$  passes from positive to non-positive, i.e., from positive to either zero or negative values, or vice versa, where it passes from negative to non-negative, i.e., from negative to either zero or positive values.
  - The convolutional value  $[f(x) \bullet \partial^2 G/\partial x^2]$  is an estimate of the local curvature, also called local concavity, of the function  $f(x)$ . The term "local curvature" = "local concavity" is synonym of "second-order derivative", "change in first-order derivative", "change in local slope" or "change in gradient". For example, in a ramp, the gradient is constant, but the local curvature  $[f(x) \bullet \partial^2 G/\partial x^2]$  is zero. In a flat function area, the gradient is zero and the local curvature  $[f(x) \bullet \partial^2 G/\partial x^2]$  is zero. In particular:
    - If  $[f(x) \bullet \partial^2 G/\partial x^2] > 0$  (positive value), then the estimated local concavity is down.
    - If  $[f(x) \bullet \partial^2 G/\partial x^2] < 0$  (negative value), then the estimated local concavity is up.
    - In a horizontal area there is no local curvature, then  $[f(x) \bullet \partial^2 G/\partial x^2] = 0$ .
    - In a ramp there is no local curvature, then  $[f(x) \bullet \partial^2 G/\partial x^2] = 0$ . This observation complies with the Mach bands illusion.

These observations lead to an unequivocal operational definition of image contours, to be considered original, neither obvious nor trivial because different from alternative definitions found in literature, unless proved otherwise (Adelson and Bergen, 1985; Bertero et al., 1988; Burr and Morrone, 1992; Canny, 1986; Heitger et al., 1992; Pessoa, 1996; Rodrigues and du Buf, 2008; Smith and Brady, 1997; Torre and Poggio, 1986; Yuille and Poggio, 1986).

*A pixel  $I(n)$  with pixel coordinates  $n = (x, y)$  in a 2D array is an image-contour pixel if it is a ZX pixel, where the image local concavity, equal to  $[I(n) \bullet \partial^2 G/\partial n^2]$ , changes in sign, either from positive to non-positive, i.e., from positive to either zero or negative, or from negative to non-negative, i.e., from negative to either zero or positive, in comparison with the local concavity of any of its 8-adjacency neighboring pixel.*

Noteworthy, because the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution. Therefore, the filtered image  $(\partial^2 G/\partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2/\partial n^2$  applied to the low-pass image adopted by both Canny (1986) and Bertero et al. (1988).

$$(\partial^2 G/\partial n^2 * I) \neq \partial^2/\partial n^2 (G * I).$$

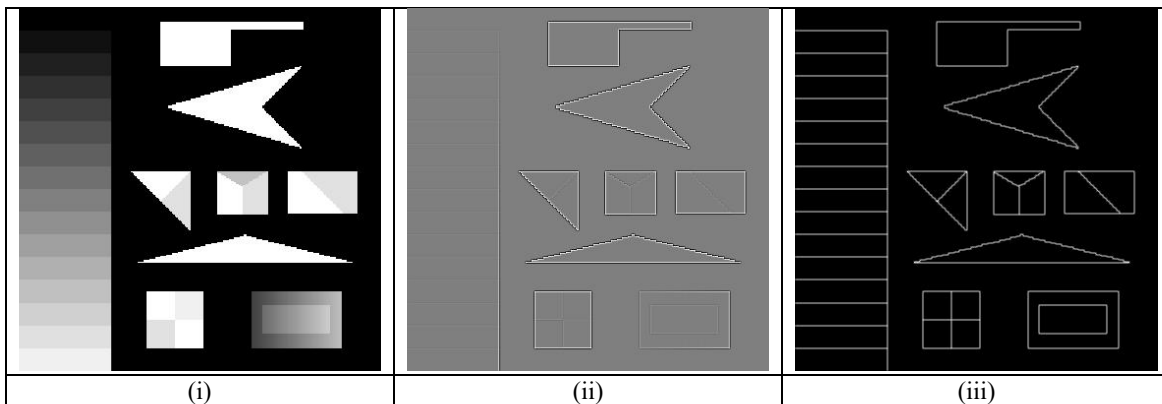
- ✓ Function perceptual contrast,  $\text{PrcptlCntrst}(x) = \text{abs}[f(x) \bullet \partial^2 G/\partial x^2] + \text{abs}[f(x) \bullet \partial G/\partial x]/2$ , hence  $\text{PrcptlCntrst}(x) \geq 0$ , is an original non-negative combination of even- and odd-symmetric simple cells alternative to complex cells



featuring a second-degree (squaring) nonlinearity proposed by (Adelson and Bergen, 1985; Burr and Morrone, 1992; Canny, 1986; Heitger et al., 1992; Pessoa, 1996; Rodrigues and du Buf, 2008).

- ✓  $\text{PrcptlCntrst}(x)$  allows to partition (discriminate) zero-concavity (ZC) segments, where  $[f(x) \bullet \partial^2 G/\partial x^2] = 0$ , into either ramps or flat areas: in any ramp  $[f(x) \bullet \partial^2 G/\partial x^2] = 0$  AND  $\text{PrcptlCntrst}(x) > 0$ , in any flat area  $[f(x) \bullet \partial^2 G/\partial x^2] = 0$  AND  $\text{PrcptlCntrst}(x) = 0$ .
- ✓ The local extrema (local maxima and local minima) of  $\text{PrcptlCntrst}(x) \geq 0$  feature the following properties.
  - They represent a small subset of  $\text{PrcptlCntrst}(x) \geq 0$ , hence their scrutiny should be easier to accomplish.
  - According to Fig. 4-24, the local extrema of  $\text{PrcptlCntrst}(x) \geq 0$  can either coincide or not with ZX pixels.
  - They appear of particular interest for two low-level vision problems.
    - Perceptual contour detection, in combination with ZX pixels detected by the even-symmetric filtering operator  $\partial^2 G/\partial x^2$ .
    - Image saliency perception, including keypoint detection, such as end-point, T-junction, X-junction and corner detection, called terminations by Marr (Marr, 1982),

According to the existing literature (Heitger et al., 1992; Rodrigues and du Buf, 2008), the information represented at the keypoints complements the edge representation. The edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of this map difficult. One can see that the representation of keypoints indicates precisely these critical locations, like terminations, corners and junctions. Many of the keypoints are located on occluding contours. In Fig. 4-24, the local extrema of  $\text{PrcptlCntrst}(x) \geq 0$  appear related to keypoints by Lowe (Lowe, 2004) and/or end-stopped cell's outputs (Heitger et al., 1992; Rodrigues and du Buf, 2008). For example, in (Lowe, 2004), Lowe searches for scale-invariant keypoints (scale invariant feature transform, SIFT) as local extrema in the difference of Gaussian (DOG), equivalent to an isotropic Laplacian of a Gaussian ( $\nabla^2 G$ )-filtered image, that are not, simultaneously, contour pixels, i.e., pixels whose local concavity, estimated by means of a second-order derivative Hessian matrix is strong because they lie along edges. In (Heitger et al., 1992; Rodrigues & du Buf, 2008), the computational model of so-called end-stopped cells allows to detect “keypoints” as the peaks (local maxima in a  $3 \times 3$  neighbourhood) in the summed end-stopped representation. These keypoints are conceptually one-to-one related to, i.e., they are the biological counterpart of, the Lowe SIFT operators, although keypoint estimation through simple- and double-stopped operators adopted in (Heitger et al., 1992; Rodrigues & du Buf, 2008) is not computationally equivalent to Lowe’s SIFT (Lowe, 2004).



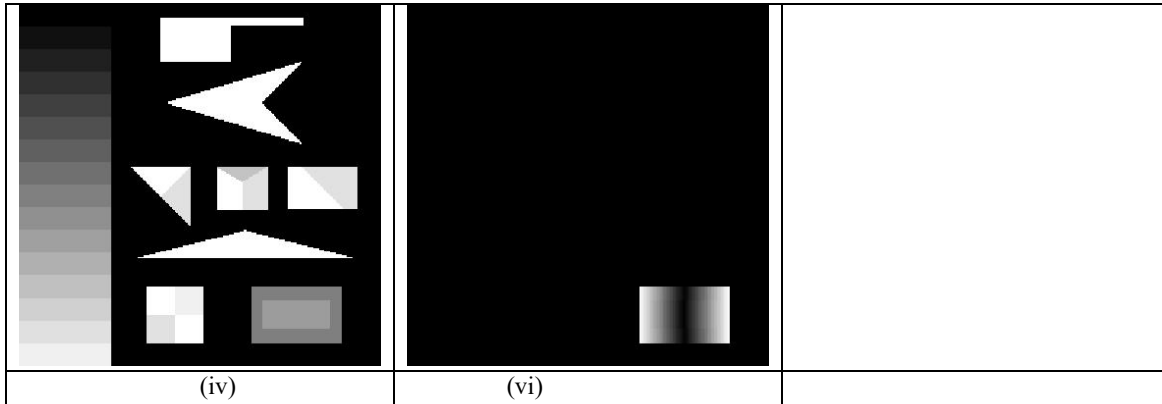


Fig. 4-25. (i) SUSAN synthetic test panchromatic image (Smith and Brady, 1997), consisting of 31 segments according to human perception, including a ramp edge (in the bottom right). (ii) Multi-scale multi-orientation even-symmetric filter-based image reconstruction in range  $\{-1, 1\}$ , where ZX pixels can be detected. (iii) Automated image segmentation into ZX segments. Exactly 31 segments are detected. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$  (refer to Fig. 4-18). (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction in comparison with the input image, per-pixel root mean square error. It shows that ramp edges, irrespective of their slope, are detected as on one single segment, in agreement with the Mach bands illusion. Where a ramp-edge object is replaced by its object mean value in the piecewise-constant image reconstruction phase (object mean view), within that object the per-pixel root mean square error can be high.

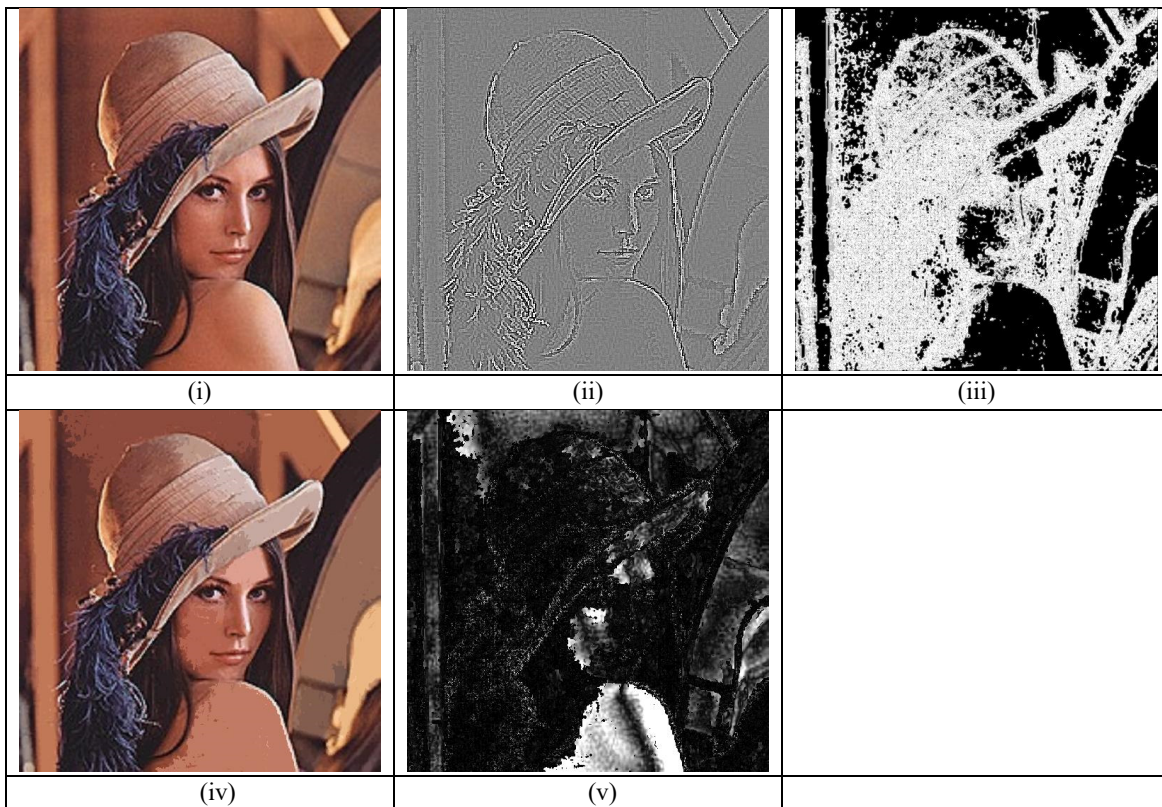


Fig. 4-26. (i) Natural RGB image of Lenna, intuitive to understand by a human observer. (ii) Multi-scale multi-orientation even-symmetric filter-based image reconstruction in range  $\{-1, 1\}$ , where ZX pixels can be detected. (iii) Automated image segmentation into ZX segments. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$  (refer to Fig. 4-18). (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction in comparison with the input image, per-pixel root mean square error. It shows that ramp edges, irrespective of their slope, are detected as on one single segment, in agreement with the Mach bands illusion. Where a ramp-edge object is replaced by its object mean value in the piecewise-constant image reconstruction phase (object mean view), within that



object the per-pixel root mean square error can be high.

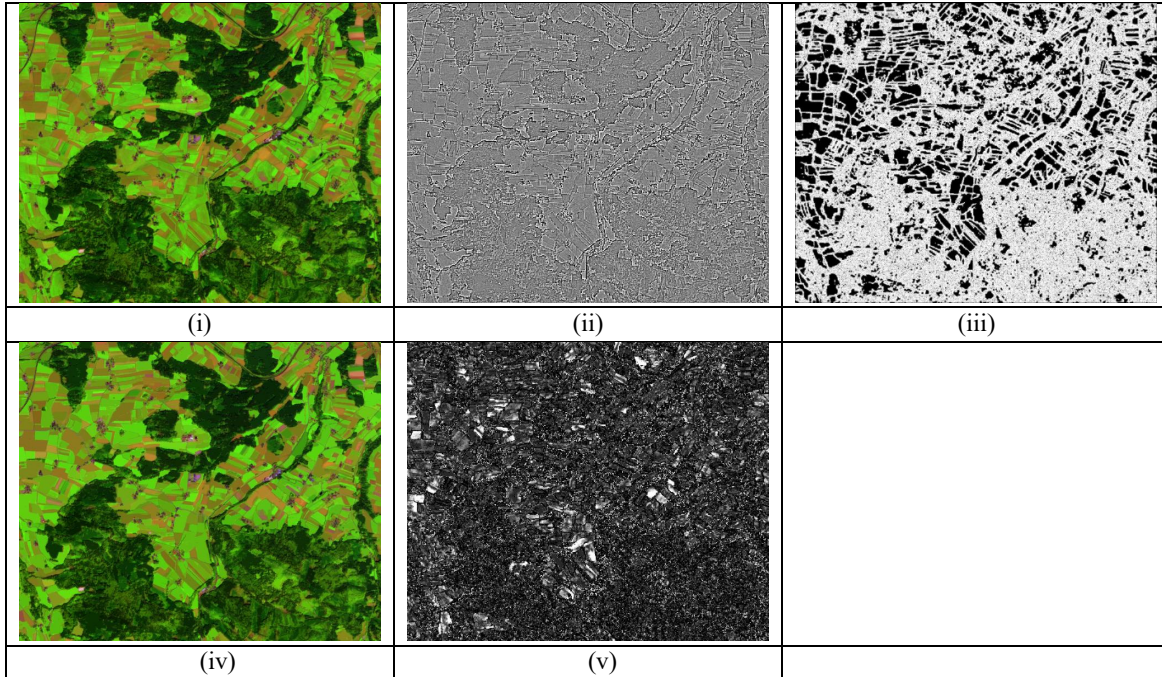


Fig. 4-27. (i) Subset of a Sentinel-2A image of Austria, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values, depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 10 m resolution. Acquired on 2015-08-13. (ii) Multi-scale multi-orientation even-symmetric filter-based image reconstruction in range  $\{-1, 1\}$ , where ZX pixels can be detected. (iii) Automated image segmentation into ZX segments. Segment contours depicted with 8-adjacency cross-aura values in range  $\{0, 8\}$  (refer to Fig. 4-18). (iv) Object mean view = object-wise constant input image reconstruction. (v) Object-wise constant input image reconstruction in comparison with the input image, per-pixel root mean square error. It shows that ramp edges, irrespective of their slope, are detected as on one single segment, in agreement with the Mach bands illusion. Where a ramp-edge object is replaced by its object mean value in the piecewise-constant image reconstruction phase (object mean view), within that object the per-pixel root mean square error can be high.

As a non-trivial extension of the 1D spatial filters shown in Fig. 4-24 to a (2D) image domain, an even- and odd-symmetric multi-scale multi-orientation wavelet-based 2D spatial filter bank was implemented for near-orthogonal image analysis (decomposition), synthesis (reconstruction, by summing the near-orthogonal signal components), and ZX contour-pixel detection in compliance with the Mach bands illusion. Four spatial filter scales one-octave apart and two filter orientations suffice to reduce the image reconstruction error below human perception. ZX pixels automatically detected in synthetic, natural and satellite images are shown in Fig. 4-25, Fig. 4-26 and Fig. 4-27 respectively.

In Marr's words an intermediate information primitive in the image domain, called ZX segment, was defined as "a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform" (Marr, 1982, p. 60). Hence, it is at the level of detection of ZX segments that 0D ZX pixels and 1D image-contours turn into sub-symbolic discrete 2D image-objects (polygons). To automatically extract 2D ZX segments from 0D ZX pixels, an original algorithm was implemented in operating mode to account for real-world image background "noise" below human visual perception. It is described hereafter in pseudo-code.

- Let us consider a (2D) image function  $I(n)$  where the 2D pixel coordinate in the image domain is  $n = (x, y)$ . A 2D even-symmetric spatial filtered-image  $[I(n) \circ \partial^2 G/\partial n^2] = \text{EFI}(n) \in [-1, 1]$ , can be automatically partitioned into a 3-level image consisting of "white", "gray" and "dark" connected segments, whose EFI values are respectively  $> 0$ , equal to zero (defined as zero-concavity segments, ZC) and  $< 0$ . According to this terminology, a ZC segment can be either a flat area or a ramp.
- The 3-level EFI partition can be transformed into a so-called preliminary ZX segment map by a well-posed two-pass connected-component image labeling algorithm (Sonka et al., 1994; Dillencourt et al., 1992).

• Based on Fig. 4-24, it is intuitive to understand that, starting from the aforementioned preliminary ZX segment map, adjacent pixel pairs can be merged into a new or pre-existing ZC segment according to the following three criteria described in pseudo-code.

- Adjacent pixel-pair merging into a new or pre-existing ZC segment, Rule 1. Any ZX pixel should be merged with the 8-adjacency neighboring pixel whose  $PreptlCntrst(x) = 0$  or “low”, if any. This is equivalent to requiring the 8-adjacency neighboring pixel to belong to a flat area. If this condition occurs, these two neighboring pixels are merged into the same ZC segment, either new or pre-existing.
- Adjacent pixel-pair merging into a new or pre-existing ZC segment, Rule 2. Any pair of 8-adjacency neighboring pixels, either ZX or not, should be merged with the neighboring pixel whose  $InterPixelDeltaGrayValue = 0$  or “low”, if any. If this condition occurs, these two neighboring pixels are merged into the same ZC segment, either new or pre-existing.
- Any pair of 8-adjacency neighboring pixels, either ZX or not, both featuring  $PreptlCntrst(x) = 0$  or “low”, should be merged into the same ZC segment, either new or pre-existing.

This ZX segment detection algorithm depends on two hidden parameters, to be defined based on general-purpose user- and application-independent psychophysical evidence. The Weber Sensitivity Fraction in range [0, 1] was set equal to  $0.010 = 1\%$  in range [0, 1] to decide whether an  $InterPixelDeltaGrayValue$  is fuzzy “low”. The Weber–Fechner sensitivity law states that a physical entity, e.g., a weight, seems to have to increase by an X% for someone to be able to reliably detect (sense) the increase. This minimum required fractional increase of a physical entity, e.g., 5/100 of the original weight, is referred to as the “Weber (sensitivity) fraction” for detecting changes in a physical entity. The Normalized Perceptual Contrast Action Potential in range [0, 1] was set equal to  $0.012 = 1.2\%$  in range [0, 1] to decide whether a  $PreptlCntrst$  value is fuzzy “low”. In physiology, synaptic inputs to a neuron trigger an action potential when inputs are superior to a threshold potential. When these two hidden parameters are set as described above, ZX segments automatically detected in synthetic, natural and satellite images are shown in Fig. 4-25, Fig. 4-26 and Fig. 4-27 respectively.

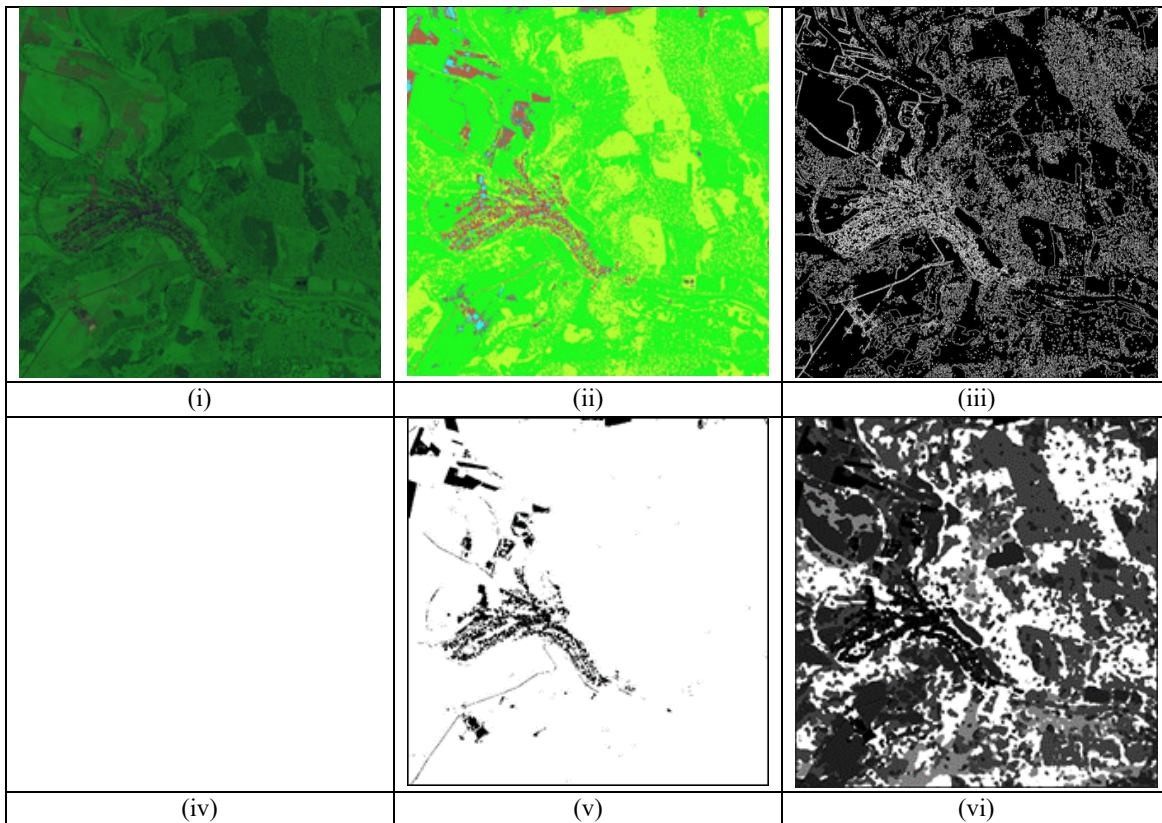


Fig. 4-28. (i) IKONOS image, depicted in false colors (R: band CH3, G: band CH4, B: band CH1), calibrated into TOARF values. Spatial resolution: 4 m. No histogram stretching applied for visualization purposes. (ii) SIAM map at coarse color granularity, 12 spectral categories, depicted in pseudo colors. (iii) 8-adjacency cross-aura measure showing the boundaries





of segments automatically detected in the SIAM map domain. (v) SIAM binary vegetation map. Vegetated areas, detected by SIAM according to their spectral signature (color) properties exclusively, are shown in white. They can be split into natural forested areas, featuring high-texture values, and non-forested areas, either managed (agricultural fields) or natural (pasture), featuring low-texture values. (vi) 3-bit 3-scale texture binary profile, in range  $\{0, 7\}$ , depicted in gray shade from black to white, stratified by the binary vegetation map. Non-vegetated mask-out areas, excluded from stratified texture analysis of vegetated areas, are also shown in black.

(VI) Full primal sketch, automated “stratified” texture segmentation as a multi-scale analysis of the spatial distribution of image contours. This is a multi-scale extension of the binary single-scale texture detection algorithm proposed in (Nagao and Matsuyama, 1980): for a central pixel centered on a moving window of size  $N \times N$  in pixel units, the central pixel is high-texture if there are more than  $N \times \sqrt{2} = 1.4 \times N$  contour pixels within the local window. In the proposed  $M$ -scale binary texture analysis, with scales  $M = 3$ , where dyadic window sizes in pixel units are equal to  $N_1 \times N_1 = 3 \times 3$  up to  $7 \times 7$  and  $13 \times 13$ , where moving windows are estimated via the fast implementation proposed in (Clausi and Zhao, 2003), then the single-scale binary texture values are combined through scales into an  $M = 3$ -bit 3-scale texture binary profile, whose overall value ranges in  $\{0, 7 = 2^2 + 2^1 + 2^0\}$ , where the largest texture weight  $2^{M-1} = 2^2 = 4$  is applied to the finest resolution window  $N_1 = 3$  and the lowest texture weight  $2^0 = 1$  is applied to the coarsest resolution window  $N_3 = 13$ . In practice, a numeric texture variable is transformed into a nominal texture variable suitable for fuzzy reasoning (Fig. 4-28).

(VII) Full primal sketch, planar shape and size description of image-objects. An image segmentation map (Fig. 4-17) consists of planar objects belonging to the spatial unit types (0D) point, (1D) line or (2D) polygon (Open Geospatial Consortium, 2015). Any planar object can be described in geometric (shape) and size terms by a minimally dependent and maximally informative (mDMI) set of: (i) planar shape indexes, specifically, scale-invariant roundness, elongatedness, straightness of boundaries, simple connectivity, rectangularity and convexity (Fig. 4-29), and (ii) size indexes in pixel units, specifically, area and characteristic scale, equivalent to a within-segment average spatial autocorrelation estimate (Baraldi and Soares, 2017; Soares et al., 2014). Noteworthy, excluding size-related parameters area and characteristic scale, all shape (geometric) descriptors belong to range  $[0, 1]$ ; hence, they are intuitive to partition into fuzzy sets, e.g., “high”, “medium” or “low”, employed by fuzzy decision rules typical of human fuzzy reasoning on categorical variables (Zadeh, 1965). The implemented mDMI set of planar shape indexes is alternative to existing libraries of planar shape descriptors available in commercial software toolboxes (OpenCV, 2015), provided with no quality assurance in terms of shape feature independence and informativeness, which is in contrast with the QA4EO *Val* guidelines. Last but not least, shape and size descriptor values computed for each image-object (segment) automatically detected in a multi-level image, such as an SCM, are encoded as raster information layers in the EO-IU4SQ system’s array database to allow shape-related spatiotemporal queries, e.g., change in size of a water body through time, where co-existing spatial units of information are (0D) pixel, (1D) line and (2D) polygon. This allows to overcome the traditional ill-fated dichotomy between (0D) pixel-based image analysis and the 2D polygon-specific OBIA paradigm (Blaschke et al., 2014).

#### 4.2.4. GUI

The EO-IU4SQ system’s GUI supports the following human-machine interactions (Tiede et al., 2016).

(1) Scene-domain knowledge transfer from human-to-machine. In the 4D real world, observations (true-facts) are discrete  $n$ -tuples (space  $x$ ,  $y$  and  $z$ , time  $t$ ; “theme”; plus other numeric or categorical attributes, e.g., weight, size, etc.), where the 4-tuple  $(x, y, z, t)$  is the location in space and time of the observation, while attribute “theme” identifies the real-world phenomenon or object being observed (Ferreira et al., 2014). Hence, “theme” may account for semantics involved with the observed object or phenomenon. All possible combinations of attributes (space, time, theme), can be modeled as three data types, called time series, coverage, and trajectory, where one attribute is measured, the second is fixed and the third is controlled. A time series represents the measured variations of a theme over a controlled time in a fixed location. A trajectory measures locations of a fixed theme over a controlled time. A coverage measures attribute theme within a controlled spatial extent at a fixed time. A world model is an ontology of real-world geospatiotemporal objects/continuants and events/occurrences derived from the three spatiotemporal data types time series, coverage and trajectory. A real-world object/continuant (e.g., a car) is an identifiable (discrete) entity, i.e., it is provided with a unique “identity”, which remains constant during its lifetime, while its attributes, whether spatiotemporal or not, including semantics, can change during its lifetime. An object is present as a whole unit at each moment of its existence. In EO applications, a real-world object can be: (i) a periodic object whose identity is fixed while its attributes change with a cyclic behavior, i.e., the object’s identity comprises a given sequence of different states in a fixed time periodicity, e.g., a corn cropland whose growth cycle is



decided by agricultural practices specific to a geographic region, or (ii) persistent/non-periodic objects, e.g. forested areas or lakes. An event is an individual episode with a definite beginning and an end. It only exists as a whole across the interval over which it occurs, either instant or durative. An event does not change over time. While an event can involve one or more objects, the same object can be involved in any number of events. Events can be: (i) instant, (ii) durative, including short-term transition events and slowly transient events. For example, in EO applications vegetated areas can change into different LC types, such as bare soil, building or water, according to slowly transient events, such as urban sprawl due to policies enforced by responsible authorities, or due to some short-term transition events, e.g., natural or artificial fire and flooding events. The world model can be graphically represented as a semantic network with LC classes as nodes and spatiotemporal relationships, including events, as arcs between nodes (Grove, 1999). It is constrained by an algebra which describes spatiotemporal data types and operations in a language-independent and formal way (Ferreira et al., 2014).

Segment Number	Chromatic	Panchromatic	Segment	Convexity and No Hole	Elongatedness	Polygon-Based Approximate Rectangularity	Roundness and No Hole	Simple-Connectivity	Straightness of Boundary	Angle of MER (in degrees)	Area (in pixels)	Average Contrast Along Boundary	Morphological multiscale characteristic	Mean Panchromatic Intensity
1				0.96	1.10	1.00	0.90	1.00	0.68	90.00	81	30.53	5.59	94.95
2				0.85	2.92	0.95	0.66	1.00	0.63	75.07	666	1.91	15.14	57.62
3				0.86	4.68	1.00	0.62	1.00	0.77	-16.50	237	36.07	5.12	113.29
4				0.87	8.72	1.00	0.53	0.72	0.83	74.05	1406	27.24	7.27	89.84
5				0.78	4.86	1.00	0.58	0.89	0.89	73.30	1812	27.00	15.62	76.79
6				0.48	9.24	0.78	0.44	1.00	0.79	155.85	461	7.83	16.31	59.71
7				0.35	50.42	0.72	0.22	1.00	0.89	-105.95	727	17.72	9.64	54.21
8				0.67	22.35	0.05	0.33	1.00	0.85	-5.57	340	20.51	8.87	55.27
9				0.84	9.61	1.00	0.54	0.93	0.85	167.83	555	20.22	8.67	49.82

Fig. 4-29. Screenshot of the GUI specifically developed to show a human expert values of the proposed set of geometric attributes (Baraldi and Soares, 2017; Soares et al., 2014). In this GUI, darker cells correspond to: (i) higher values of geometric attributes and (ii) lower values of photometric attributes, like the panchromatic mean intensity shown at the rightmost column. In this figure, for reasons of readability only nine segments are shown simultaneously for comparison. Detected by the SIAM expert system in a spaceborne very high resolution (VHR) QuickBird image of an urban area, segments 1 through 6 correspond to buildings or parts of buildings while segments 7 through 9 belong to roads. These two families of segments appear easy to discriminate based on different combinations of ranges of change (fuzzy sets) of their geometric attributes.

(2) Graphic selection of existing semantic queries/decision rules or generation of new semantic queries/decision rules, constrained by the world model. In general, each query instantiation is associated with an information pair, specifically, one spatiotemporal scene-domain knowledge (e.g., target LC classes) and one set of sensor-specific transfer functions required to map scene-domain knowledge in user-speak into image-domain knowledge in techno-speak. A query pipeline is a combination of spatial and temporal operators and/or “default” algorithms whose inputs are qualitative/categorical information layers (e.g., SCMs) or quantitative/numeric variables (e.g., spectral indexes) available in the fact base. There are two types of queries. (i) To accomplish SCBIR operations, where the fact base is investigated for EO image retrieval purposes. For example, retrieve EO images that are cloud-free across the selected geographic area of interest (AOI), or those where a specific vegetated LC is found in the AOI, etc. (ii) To infer new information layers from the fact base. For example, detect-through-time flooded areas as a post-classification combination through time of available single-date



SCMs, etc.

#### 4.2.5. Array fact base

To accomplish efficient geospatial data querying and analysis through space and time within a user-defined AOI and a target time interval, the latest EO-IU4SQ system implementation stores its fact base, consisting of multi-sensor multi-temporal EO images, e.g., Landsat-4/5/7/8 images, Sentinel-2A images, etc., each one provided with information products, either numeric or categorical, in an array database implemented as the Rasdaman (Baumann et. al, 2015). Considered a viable alternative to traditional flat files adopted in relational databases (Tiede et. al, 2016), an array database instantiates multiple spatiotemporal data cubes in compliance with the Open Geospatial Consortium (OGC) standards (OGC, 2015), to guarantee inter-system harmonization and compatibility. In a spatiotemporal data cube the third dimension is time, defined in eXtensible Markup Language (XML) as a 1D temporal coordinate system accessible Unified Resource Identifiers (URIs). Time overlays the 2D spatial coordinate system, specified by European Petroleum Survey Group (EPSG) codes defined in XML using URIs. In the implemented Rasdaman array database, any EO image, its derived categorical variables (e.g., an SCM) and derived numeric variables (e.g., spectral indexes) are stored in three different data cubes (also refer to Fig. 4-4).

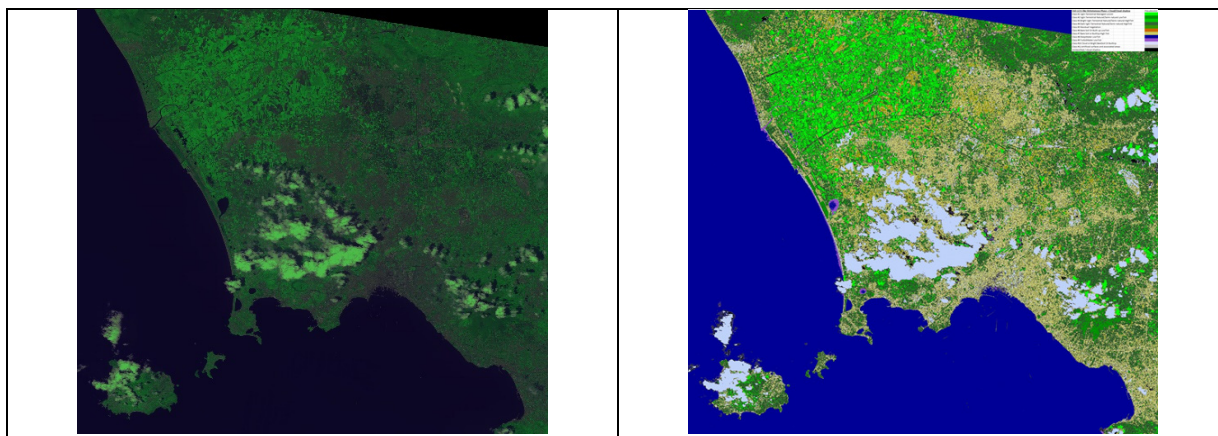


Fig. 4-30. Left: 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into TOARF values and depicted in false colors (R = Visible Red, G = NIR, B = Visible Blue), 10 m resolution. Acquired on 2004-13-06. No histogram stretching for visualization purposes. Right: Automatically generated low-level 12-class EO image classification map, approximating an “ideal” 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow. Visual features input to the implemented ESA EO Level 2 SCM classifier are: MS color names detected by the Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names and superpixel detection in the image-domain, and texture segmentation automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range  $\{0, 7\}$  (refer to Section 4.2.3.2). Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size. Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier are: local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships (refer to Section 4.2.3.2). No cloud/cloud-shadow detection and masking is employed either. Map legend: refer to Fig. 4-32.

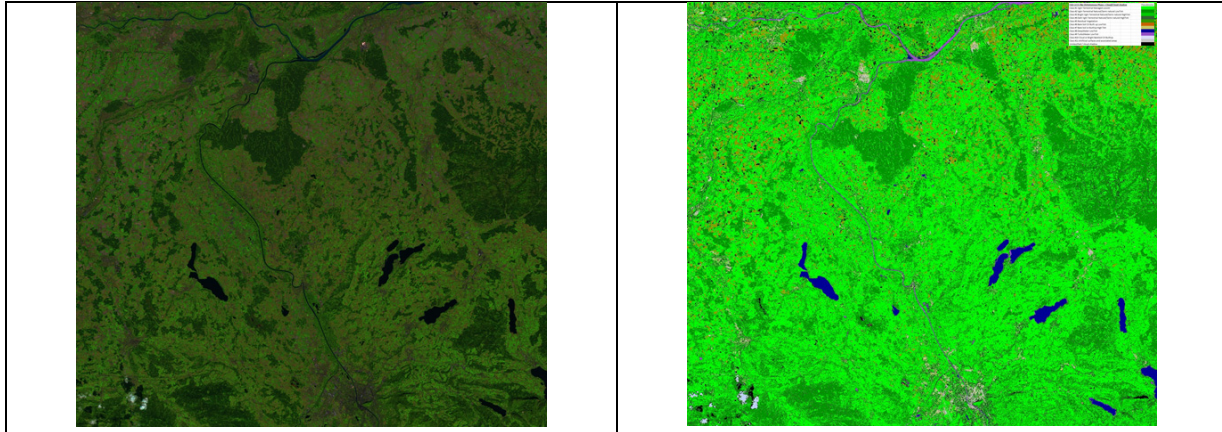


Fig. 4-31. Left: Subset of a 6-band (B, G, R, NIR, MIR1, MIR2) Sentinel-2A (S2A) image of Austria, reduced to 6 bands Landsat-like (B, G, R, NIR, MIR1, MIR2), radiometrically calibrated into TOARF values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 10 m resolution. Acquired on 2015-08-13. No histogram stretching for visualization purposes. Right: Automatically generated low-level 12-class EO image classification map, approximating an “ideal” 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow. Visual features input to the implemented ESA EO Level 2 SCM classifier are: MS color names detected by the Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names and superpixel detection in the image-domain, and texture segmentation automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range {0, 7} (refer to Section 4.2.3.2). Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size. Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier are: local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships (refer to Section 4.2.3.2). No cloud/cloud-shadow detection and masking is employed either. Map legend: refer to Fig. 4-32.

	Pseudocolor
<b>Class #1 Vgtn Terrestrial Managed LowTxtr</b>	
<b>Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxtr</b>	
<b>Class #3 Bright Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #5 Residual Vegetation</b>	
<b>Class #6 Bare Soil Or Built-up LowTxtr</b>	
<b>Class #7 Bare Soil or BuiltUp High Txtr</b>	
<b>Class #8 DeepWater LowTxtr</b>	
<b>Class #9 TurbidWater LowTxtr</b>	
<b>Class #10 Cloud or Bright BareSoil Or BuiltUp</b>	
<b>Class #11 Artificial surfaces and associated areas</b>	
<b>Unclassified / cloud-shadow</b>	

Fig. 4-32. Implemented ESA EO Level 2 scene classification map (SCM) legend, consisting of 12 classes, approximately equivalent to a FAO LCCS Dichotomous Phase (DP)-like 1st (veg/non-veg) and 2nd level (water/terrestrial) + quality layers cloud and cloud-shadow. This is a mere approximation of an “ideal” 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow. Visual features input to the implemented ESA EO Level 2 SCM classifier are: MS color names detected by the Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names and superpixel detection in the image-domain, and texture segmentation automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range {0, 7} (refer to Chapter 4.2.3.2). Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size. Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier are: local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships (refer to Chapter 4.2.3.2).

Similar to standard relational databases investigated by the Structured Query Language (SQL), an array database



consisting of data cubes can be queried by a declarative query language. In an array database storage-related characteristics, such as indexing, tiling and horizontal scaling, can be investigated and optimized independently of application logic. In addition, the data cube model has been proven to be scalable and reliable in operational applications (Baumann et. al, 2015). Based on these considerations it was selected as storage backend in the EO-IU4SQ system prototypical implementation.

Once users successfully create and store semantic queries in the web-based query interface, these queries are published as OGC compliant Web Processing Services (WPSs) and become available within the client-server architecture, to be executed by any WPS client remotely on a single server or server cluster. This client-server architecture, where EO data processing capabilities are available together with “ready-to-analyze” data and products, guarantees fast response to queries (Tiede et. al, 2016).

#### 4.2.6. Hybrid inference engine

To be shared between the EO-IU and EO-SQ subsystems, a hybrid inference engine combines rule base and fact base to infer new information products, including EO Level 2 products generated by “default” by the EO-IU subsystem. This hybrid inference engine consists of four inference subsystems.

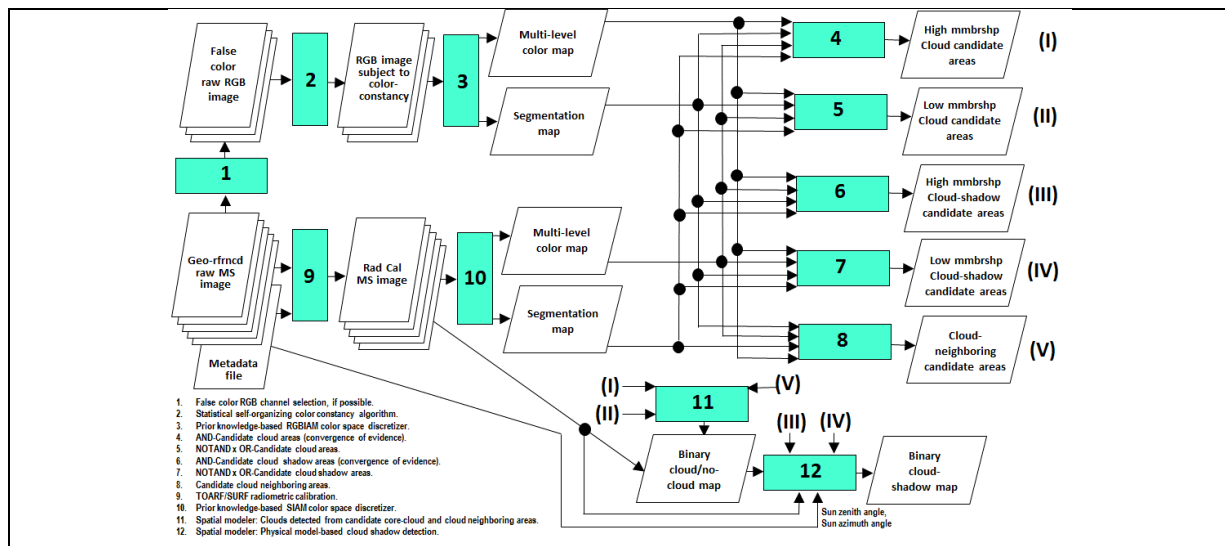


Fig. 4-33. Hybrid (combined physical model-based and statistical model-based) context-sensitive cloud/cloud-shadow detection system architecture (Baraldi, 2015). (1) False color RGB channel selection, if possible. (2) Statistical self-organizing color constancy algorithm. (3) Prior knowledge-based RGBIAM color space discretizer. (4) AND-Candidate cloud areas (convergence of evidence). (5) NOT(AND-Candidate cloud areas). (6) AND-Candidate cloud-shadow areas (convergence of evidence). (7) NOT(AND-Candidate cloud shadow areas). (8) Candidate cloud neighboring areas. (9) TOARF/SURF radiometric calibration. (10) Prior knowledge-based SIAM color space discretizer. (11) Spatial modeler: Clouds detected from candidate core-cloud and cloud neighboring areas. (12) Spatial modeler: Physical model-based cloud shadow detection.

(i) Transformation of space domain knowledge into image domain knowledge through the sensor transfer functions. It provides: (1) an estimate of the required EO imaging sensor’s physical properties, including spatial, spectral and temporal resolutions, considered necessary to intercept a target LC class phenomenon. (2) An estimate of the visual features (photometric attributes) in the image domain, such as color, planar shape, texture and spatial relationships, mapped from each LC class modeled by a human domain-expert in the 4D spatiotemporal scene domain (Tiede et. al, 2016)..

(ii) Automated EO image pre-processing (enhancement), e.g., absolute *Cal*, “stratified” atmospheric correction, “stratified” TOC, etc., refer to Chapter 4.2.3.1.

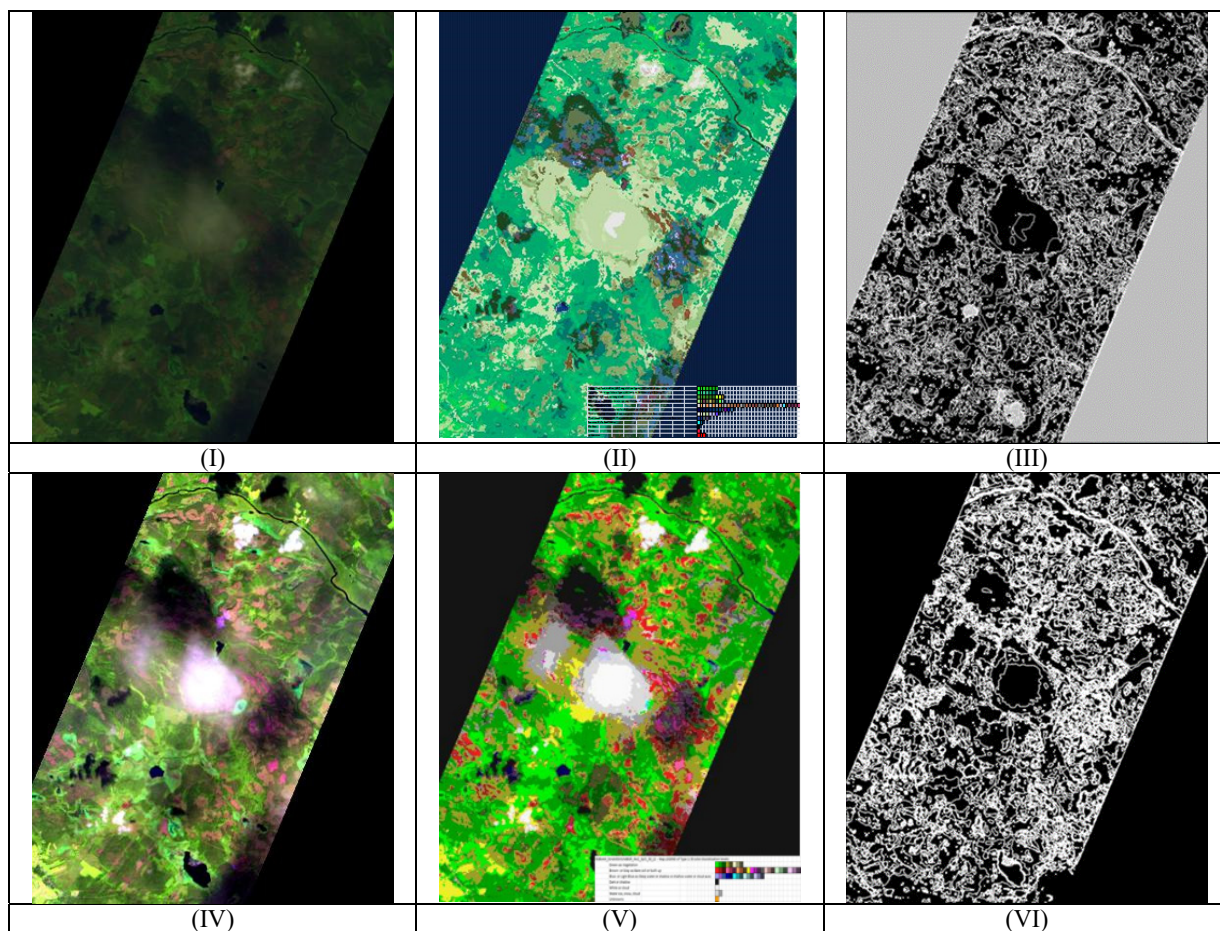
(iii) Automated low-level (pre-attentional) vision, either raw primal sketch or full primal sketch (Marr, 1982; Sonka et al., 1994), refer to Chapter 4.2.3.2.

(iv) High-level (attentional) vision, specifically, EO image classification based on a hierarchical convergence-of-visual-evidence approach (Fig. 4-8). High-level EO image understanding is split into two stages (Fig. 4-7): a preliminary ESA EO Level 2 cloud/cloud-shadow detection (Baraldi, 2015; Baraldi and Tiede, 2015) in addition to an “ideal” general-





purpose user- and application-independent 8-class LCCS-DP classification (Fig. 4-5) is followed by a LCCS-MHP classification phase, consisting of a hierarchical battery of application- and user-specific hybrid feedback one-class LC classifiers. In the implemented ESA EO Level 2 SCM product generator available to date (Fig. 4-30 and Fig. 4-31), whose map legend is shown in Fig. 4-32, input visual features are: MS color names detected by the Satellite Image Automatic Mapper (SIAM) lightweight computer program for MS reflectance space hyperpolyhedralization into MS color names and superpixel detection in the image-domain, and texture segmentation automatically estimated from spatial densities of image-contours by a 3-scale texture binary profile in range  $\{0, 7\}$  (refer to Section 4.2.3.2). Computational complexity of the implemented ESA EO Level 2 SCM classifier is linear in image size. Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier are: local shape and size of image-objects, inter-object spatial topological and spatial non-topological relationships (refer to Section 4.2.3.2). No cloud/cloud-shadow detection and masking is employed either. In an “ideal” ESA EO Level 2 product generated by “default” by the EO-IU subsystem from every EO image stored in the fact base, cloud/cloud-shadow quality masks are detected by a novel hybrid OBIA approach (Fig. 4-33) (Baraldi, 2015), based on converging spatial and color evidence detected by both physical- and statistical model-based data models (Fig. 4-34), alternative to the state-of-the-art Fmask algorithm, which is a purely deductive/top-down static decision tree (Zhe Zhu and. Woodcock, 2012). Collected by D. Tiede (Z-GIS, Univ. of Salzburg) by the end of March 2017, recent results provided by the first prototypical implementation of the proposed hybrid EO-IU system architecture for automatic spatial context-sensitive cloud/cloud-shadow detection in multi-source MS imagery, where input information sources include the SIAM and RGBIAM color maps automatically generated in linear time from a single-date MS image according to a convergence-of-evidence approach (Fig. 4-34), appear extremely encouraging as shown in Fig. 4-35.



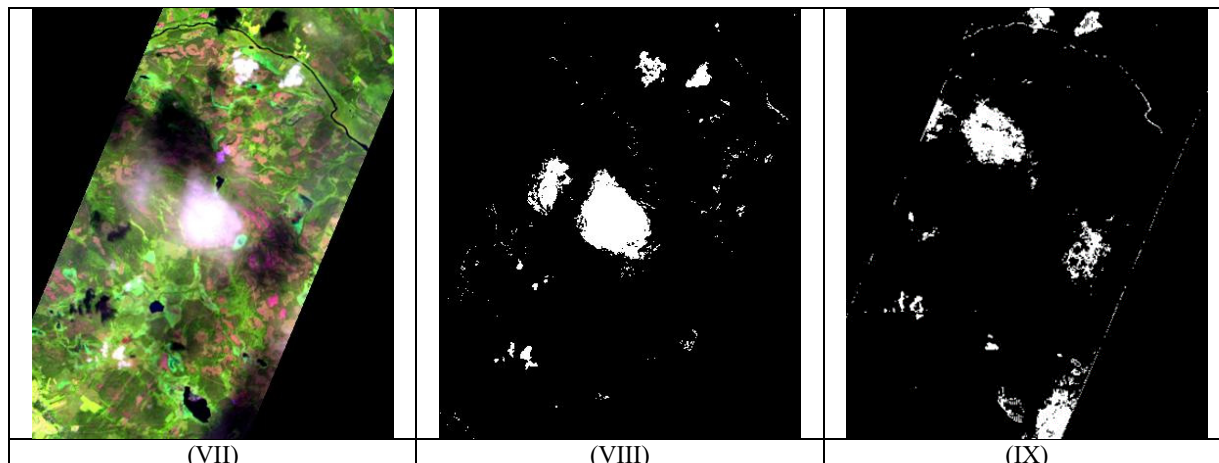
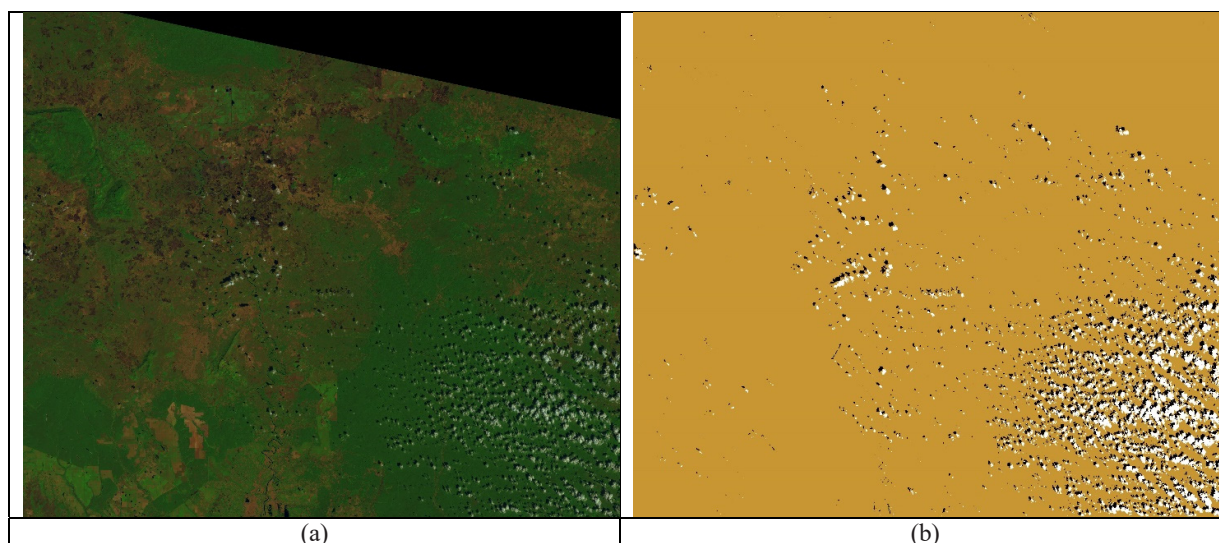


Fig. 4-34. Hybrid (combined physical model-based and statistical model-based) cloud/cloud-shadow detection (Baraldi, 2015). (I) Hyperspectral EO-1 Hyperion image, spatial resolution: 30 m., 198 bands, radiometrically calibrated into TOARF values. Transformed into a 7-band Landsat-like image, with band TIR == dumb. Depicted in false colors: R = MIR, G = NIR, B = Visible Red. No histogram stretching is employed for visualization purposes. (II) L-SIAM map of the 7-band Landsat-like image. Its map legend consists of 96 spectral categories (refer to Table 4-1), depicted in pseudocolors. (III) 8-adjacency cross-aura contours in range  $\{0, 8\}$  of the image-objects automatically detected in the multi-level SIAM map (refer to Fig. 4-18). (IV) Hyperspectral EO-1 Hyperion image, spatial resolution: 30 m, 198 bands, radiometrically calibrated into TOARF values. Transformed into a 7-band Landsat-like image, reduced to a false-color RGB image with R = MIR, G = NIR, B = Visible Red. Subject to statistical histogram stretching for color constancy. No histogram stretching is employed for visualization purposes. (V) RGBIAM map of the false-color RGB image. Its map legend consists of 50 spectral categories, depicted in pseudocolors. (VI) 8-adjacency cross-aura contours in range  $\{0, 8\}$  of the image-objects automatically detected in the multi-level RGBIAM map (refer to Fig. 4-18). (VII) Same as Figure (IV), for comparison purposes with Figure (VIII) and Figure (IX). (VIII) Binary cloud candidate mask = SIAM-based Cloud candidate 1 AND RGBIAM-based Cloud candidate 2. These cloud candidate areas are subject to further spatial analysis to enhance true positive and remove false positive cloud-candidate areas. (IX) Binary cloud-shadow candidate mask, combining evidence from SIAM and RGBIAM color naming. These cloud-shadow candidate areas are subject to spatial analysis to enhance true positive and remove false positive cloud-shadow candidate areas.





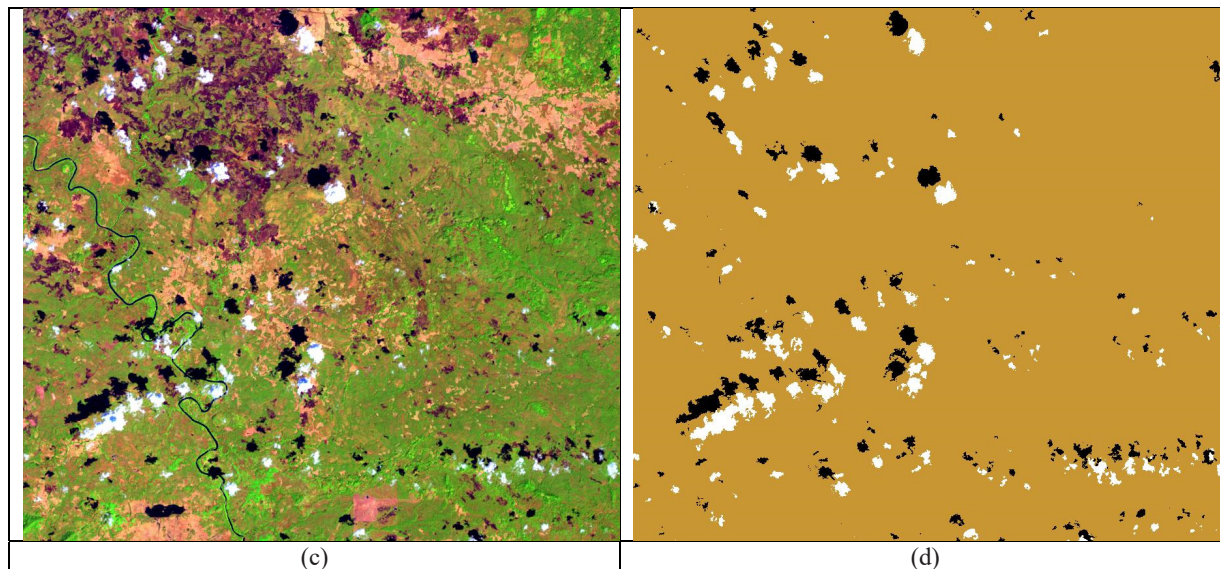
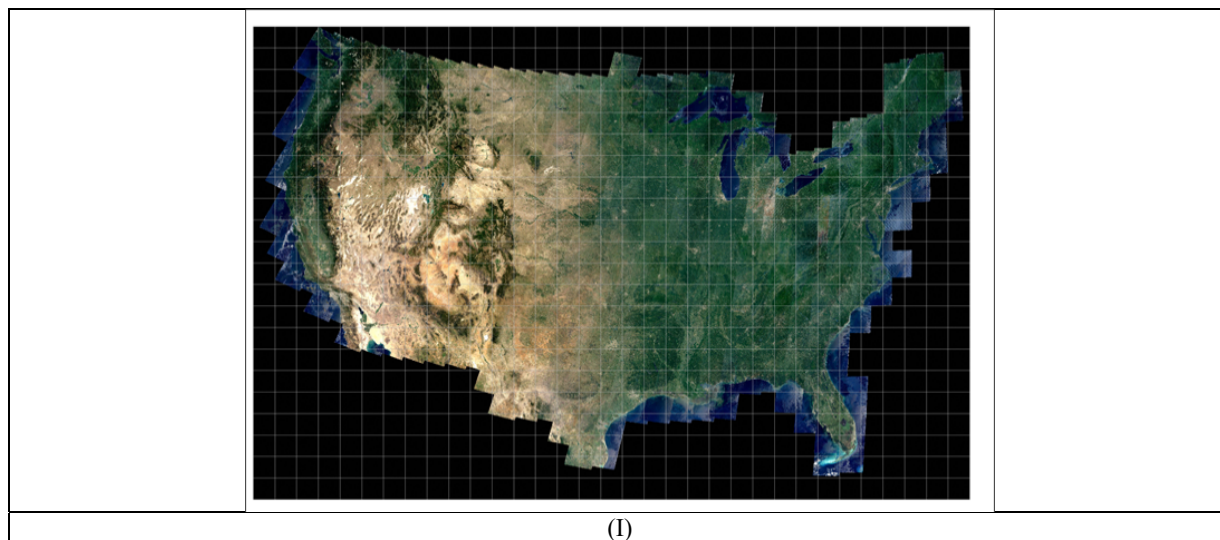


Fig. 4-35. Courtesy of D. Tiede, Z-GIS, Univ. of Salzburg. Automatic hybrid (combined physical model-based and statistical model-based) spatial context-sensitive cloud/cloud-shadow detection (Baraldi, 2015). (a) Subset of a Landsat-8 OLI image of Cambodia (LC81260512017036LGN00, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values, depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 30 m resolution, acquisition date: 03-13-2017. No histogram stretching is applied for visualization purposes. (b) Cloud/cloud-shadow thematic map. Legend in pseudolocors: White = cloud, Black = cloud-shadow, Brown = Otherwise. (c) Zoom-in of image (a), with ENVI standard histogram stretching applied for visualization purposes. (d) Zoom-in of thematic map (b).

### 4.3. Results

To comply with the QA4EO *Val* requirements each step in the EO-IU4SQ system prototype (Fig. 4-6 and Fig. 4-7) should be provided with an mDMI set of community-agreed OP-Q<sup>2</sup>Is, each index featuring a degree of uncertainty in measurement  $\pm\delta \in [0\%, 100\%]$ . These OP-Q<sup>2</sup>I  $\pm\delta$  values should allow a human supervisor to assess the propagation of errors in comparison with reference standards (Group on Earth Observation, 2010). This outcome and process *Val* requirements specification coincides with an intuitive GIGO principle of error propagation and monitoring.



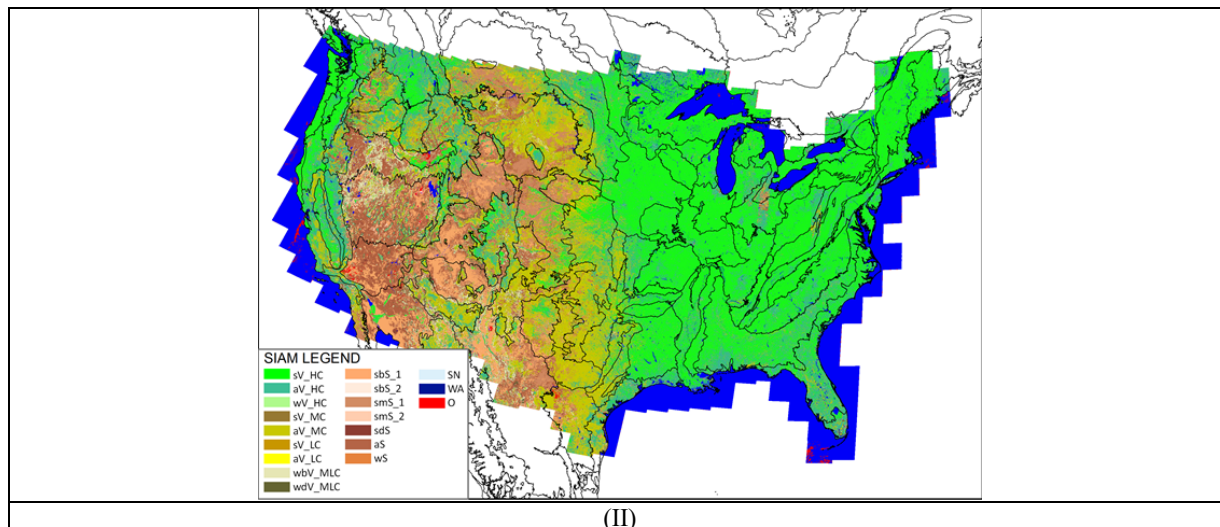
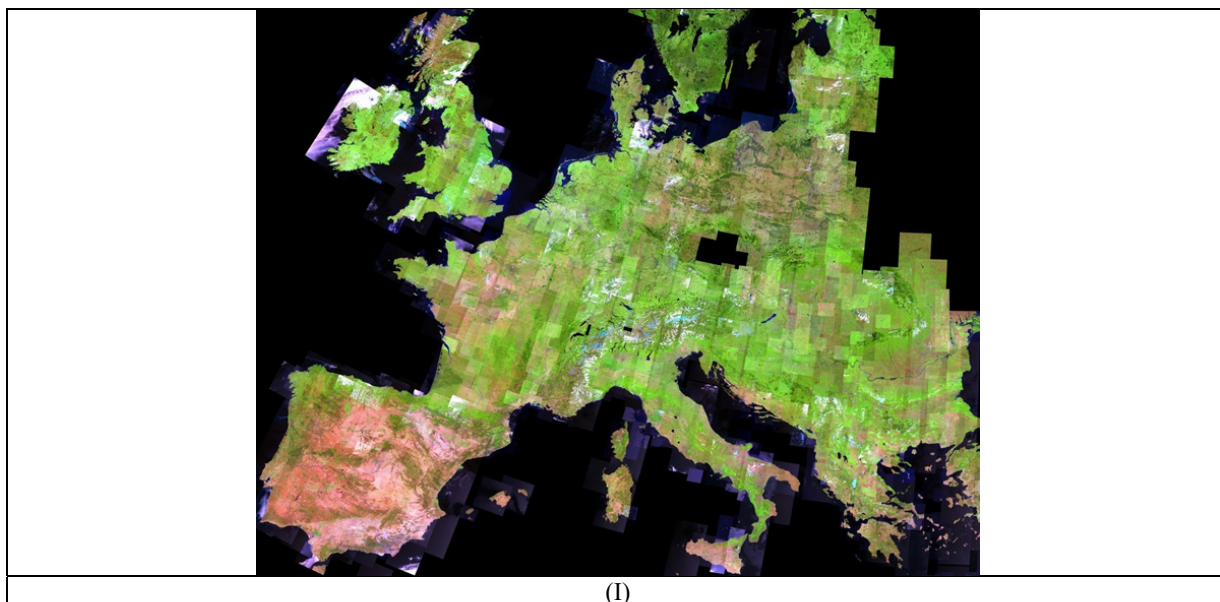


Fig. 4-36. (I) Web Enabled Landsat Data (WELD) annual composite for the year 2006 of the conterminous U.S. (CONUS), 30 m resolution, radiometrically calibrated into TOARF values. Depicted in true colors (red: Band 3, 0.63-0.69  $\mu\text{m}$ ; green: Band 2, 0.53-0.61  $\mu\text{m}$ , and blue: Band 1, 0.45-0.52  $\mu\text{m}$ ). The white grid shows locations of the 501 WELD tiles of the CONUS. Each tile is 5000 $\times$ 5000 pixels in size, covering a surface area of 150 $\times$ 150 km. Pixels are geographically projected in the Albers Equal Area projection. (II) L-SIAM map of the 7-band WELD 2006 annual composite, where 96 spectral categories (refer to Table 4-1) were reassembled into 19 spectral macro-categories, easier to deal with for map quality *Val* purposes. Black lines across the SIAM-WELD 2006 map represent the boundaries of the 86 Environmental Protection Agency (EPA) Level III ecoregions of the CONUS. The SIAM-WELD map legend is shown on the left bottom corner (Baraldi et al., 2016).





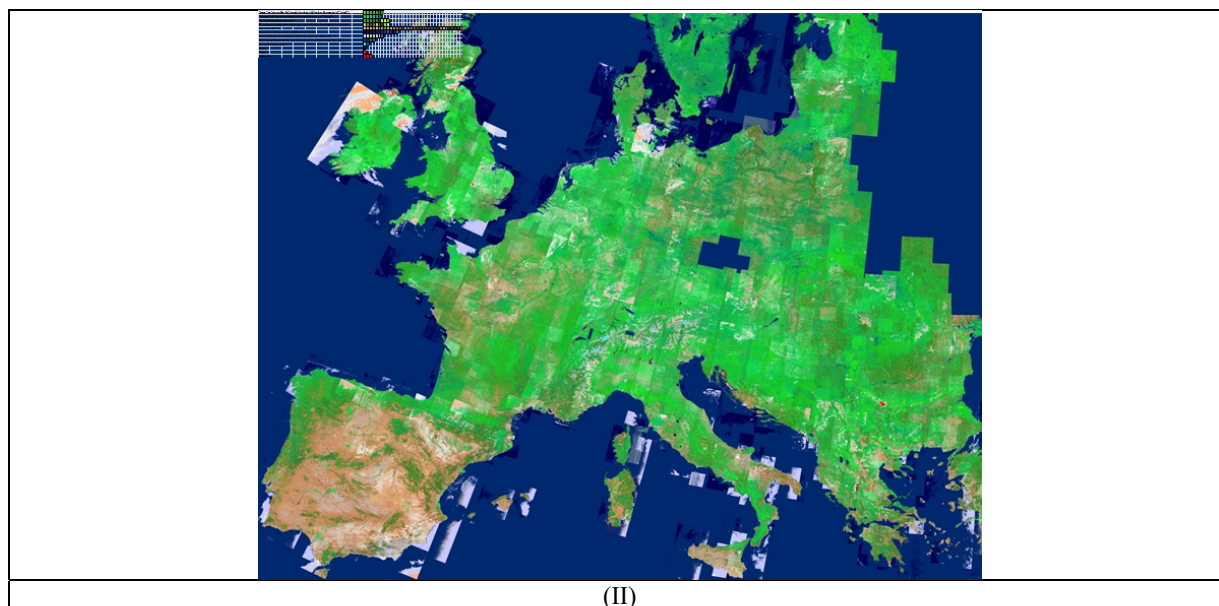


Fig. 4-37. (I) Image 2006 Coverage 1 mosaic, 10 m resolution, consisting of two thousands 4-band IRS-P6 LISSIII, SPOT-4, and SPOT-5 images, mostly acquired during 2006, radiometrically calibrated into TOARF values, geometrically orthorectified and depicted in false colors: Red – Band 4 (Short Wave InfraRed, SWIR), Green – Band 3 (Near IR, NIR), Blue – Band 1 (Visible Green). (II) S-SIAM map of the 4-band Image 2006 Coverage 1 mosaic. The S-SIAM map legend, consisting of 68 spectral categories (refer to Table 4-1) depicted in pseudocolors, is shown in the left top corner (Baraldi et al., 2010a).

For each information processing block in the EO-IU4SQ pipeline, proposed OP-Q<sup>2</sup>Is to be estimated were (refer to Chapter 4.1): (i) degree of automation, inversely related to human-machine interaction. (ii) Accuracy, e.g., thematic map accuracy with degree of uncertainty in measurement and inter-vocabulary similarity index computed according to (Baraldi, Humber & Boschetti, 2014). (iii) Efficiency, e.g., computation time and dynamic memory occupation. (iv) Scalability to cope with changes in sensor specifications and user requirements. (v) Robustness to changes in input data. (vi) Robustness to changes in input parameters, if any. (vii) Timeliness from data acquisition to product generation, which includes the time required to collect supervised data for training, if any. (viii) Costs in manpower and computer power (Baraldi and Boschetti, 2012a and 2012b).

For the EO-IU subsystem's algorithms whose research and development predates the EO-IU4SQ system prototyping, estimated OP-Q<sup>2</sup>I values can be found in the existing literature. About the automated SIAM-based stratified TOC, refer to (Baraldi et al., 2010b). About the automated near real-time SIAM expert system, refer to (Baraldi et al., 2006; Baraldi et al., 2010a; Baraldi et al., 2010b; Baraldi & Boschetti, 2012a and 2012b; Baraldi et al., 2014; Baraldi et al., 2016). In these works, the scalability of SIAM to different MS imaging sensors and its robustness to changes in the input data set were tested in multi-source EO "big data" mapping projects at continental scale, see Fig. 4-36 and Fig. 4-37. The SIAM software product adopts a tile streaming implementation, where the SIAM dynamic memory maximum size is fixed, irrespective of the image size. In these experiments the SIAM's dynamic memory (random access memory) maximum size parameter was set equal to 800 MB, which can be considered a "small" value for dynamic memory occupation. When it was run on a Dell Power Edge 710 server with dual Intel Xeon @ 2.70 GHz processor with 64 GB of RAM and a 64-bit Linux operating system, SIAM required less than 45 seconds to generate its complete set of per-image output products from a 30 m resolution 7-band Web Enabled Landsat Data (WELD) tile of 5000 × 5000 pixels (Roy et al., 2010), which means about 8 hours to map an annual WELD composite of the conterminous U.S. (CONUS), see Fig. 4-37. In our data mapping workflow, such an output rate was considered not inferior to the input rate of an annual WELD composite being implemented or delivered to end-users. Hence, the SIAM computation time was considered equivalent to near real-time.



		Reference samples											Total samples	UsrAcc	± δ
		Class #1 Vgtn Terrestrial Managed LowTxr	Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxr	Class #3 Vgtn Terrestrial Natural/Semi-natural HighTxr	Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxr	Class #5 Residual Vegetation	Class #6 Bare Soil Or Built-up LowTxr	Class #7 Bare Soil Or Built-up High Txr	Class #8 DeepWater LowTxr	Class #9 TurbidWater LowTxr	Class #10 Cloud or Bright BareSoil Or BuiltUp	Class #11 Artificial surfaces and associated areas			
T	Unclassified	0											8	0	0
e	Class #1 Vgtn Terrestrial Managed LowTxr	62	10				1	1				5	1	72	0.25 0.131399
s	Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxr	18	70											88	0.795455 0.110718
t	Class #3 Bright Vgtn Terrestrial Natural/Semi-natural HighTxr			68										68	1 0
s	Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxr				48									48	1 0
s	Class #5 Residual Vegetation			2	2	0								4	0 0
a	Class #6 Bare Soil Or Built-up LowTxr						74							74	1 0
m	Class #7 Bare Soil Or Built-up High Txr							76						76	1 0
p	Class #8 DeepWater LowTxr								80					80	1 0
l	Class #9 TurbidWater LowTxr									50				50	1 0
e	Class #10 Cloud or Bright BareSoil Or BuiltUp										44	4		48	0.916667 0.102719
s	Class #11 Artificial surfaces and associated areas						5	3				1	35	44	0.795455 0.156579
	Total samples	0	80	80	70	50	0	80	80	80	50	50	40	660	
	PrdrAcc	0	0.225	0.875	0.971429	0.96	0	0.925	0.95	1	1	0.88	0.875	OA=	0.919697
	± δ	0	0.120214	0.095207	0.051272	0.071357	0	0.075825	0.062742	0	0	0.118332	0.134644	± δ <sub>OA</sub> =	0.020733

Table 4-2. Two-way contingency table (bivariate table or frequency table, BIVRTAB) of a 12-class EO image classification map automatically generated from a Sentinel-2A image of Austria, shown in Figure 4-31, in comparison with ground truth samples selected by non-independent means (potentially biased) by the same authors of the ESA EO Level 2 SCM product under testing. Since the two input categorical variables estimated from the same geospatial population coincide, then the BIVRTAB instance is a (square and sorted) confusion matrix, CMTRX, whose main diagonal guides an intuitive interpretation process. The implemented ESA EO Level 2 SCM classifier is approximately equivalent to a standard 2-level 4-class LCCS Dichotomous Phase (DP)-like 1st (veg/non-veg) and 2nd level (water/terrestrial) taxonomy. Multiple sources of visual evidence are: SIAM's color names and a 3-scale texture binary profile in range {0, 7} automatically estimated from image-contours. Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier are per-segment geometric (shape) and size properties, and inter-segment spatial relationships, either topological or non-topological. No cloud and cloud-shadow masking is applied either. Stratified random samples = 660, spatial unit = (0D) pixel. Overall accuracy  $OA \pm \delta_{OA} = 0.92 \pm 0.02$ , with level of significance  $\alpha = 0.12$ .

With regard to the novel CV algorithms implemented in the EO-IU subsystem as part of the EO-IU4SQ system prototype and never published before, they included the following: (i) EO image pre-processing: self-organizing color constancy (refer to Chapter 4.2.3.1), (ii) pre-attentive vision: color naming, ZX contour detection, keypoint detection, ZX segment detection, and multi-scale texture segmentation (refer to Chapter 4.2.3.2), and (iii) attentive vision: ESA EO Level 2 SCM generation by a world model-driven decision tree pursuing a convergence-of-visual-evidence approach (refer to Chapter 4.2.6). In our tests, these original algorithms shared the following OP-Q<sup>2</sup>I values: full degree of automation, i.e., they require neither training samples nor system's free-parameters to be user-defined based on heuristics; computational complexity increasing linearly with the image size, which means high efficiency; independence from the input image, which means high robustness to changes in the input data set and high scalability to changes in imaging sensor specifications.

For the implemented ESA EO Level 2 product generator, collection of Level 2 SCM's Q<sup>2</sup>Is of accuracy is still ongoing. If reference samples must be identified in the test image-domain to avoid space-time differences between reference and test samples and if these reference samples must be labeled with the same thematic map legend of the implemented ESA EO Level 2 SCM product to avoid inter-dictionary translation problems, any expert photointerpreter, including the present authors, is required to accomplish an inherently equivocal (subjective) *information-as-data-interpretation* process where additional difficulties in data interpretation stem from the poor (vague, ambiguous) semantics featured by the thematic legend of interest, see Fig. 4-32. Table 4-2 reports on a preliminary test of accuracy of the ESA EO Level 2 SCM instance automatically generated from a Sentinel-2A image of Austria shown in Fig. 4-31.

With regard to Table 4-2, the selected thematic map accuracy Q<sup>2</sup>Is were the overall accuracy (OA), users' and producers' accuracies in agreement with recommendations found in (Baraldi, Boschetti & Humber, 2014). A stratified random sampling strategy was applied, with an overall number of samples equal to 340 and a sample spatial unit equal to (0D) pixel. In agreement with the QA4EO guidelines, any classification OA probability estimate,  $p_{OA} \in [0, 1]$ , is a random variable (sample statistic) with a confidence interval (error tolerance) associated with it, identified as  $\pm \delta$ , where  $\delta$  represents the half-width of the error tolerance at a specified *confidence level*  $(1 - \alpha)$  such that  $0 < \delta < p_{OA} \leq 1$ , with  $\alpha \in [0, 1]$ , known as the desired *level of significance* (e.g.,  $\alpha = 0.05$ ), which is the risk that the actual error is larger than  $\pm \delta$ . Hence, the specified confidence level  $(1 - \alpha)$  (e.g.,  $1 - \alpha = 1 - 0.05 = 95\%$ ) is the required probability that the actual error falls within the confidence interval  $\pm \delta$ . In practice  $p_{OA} \pm \delta$  is a function of the specific test data set used for its estimation, and vice

versa. For example, for a given reference sample set size ( $SSS$ ) comprising independent and identically distributed (i.i.d.) reference samples and an estimated classification accuracy probability  $p_{OA}$ , it is possible to prove that the half width  $\delta$  of the error tolerance  $\pm\delta$  at a desired confidence level (e.g., if confidence level  $(1 - \alpha) = 95\%$  then the critical value is 1.96) can be computed as follows (Lunetta & Elvidge, 1998).

$$\delta = \sqrt{\frac{(1.96)^2 \cdot p_{OA} \cdot (1 - p_{OA})}{SSS}} \quad (1)$$

Vice versa, minimum  $SSS = f(\text{target } p_{OA}, \text{target } \delta)$  can be computed as follows.

$$SSS = \frac{(1.96)^2 \cdot p_{OA} \cdot (1 - p_{OA})}{\delta^2} \quad (2)$$

For each  $c$ -th class simultaneously involved in the classification process, with  $c=1, \dots, C$ , where  $C$  is the total number of classes, with  $C \geq 2$ , it is possible to prove that (Lunetta & Elvidge, 1998):

$$\delta_c = \sqrt{\frac{\chi^2_{(1,1-\alpha/C)} \cdot p_{OA,c} \cdot (1 - p_{OA,c})}{SSS_c}}, c=1, \dots, C, \quad (3)$$

where  $\alpha$  is the desired level of significance, i.e., the risk that the actual error is larger than  $\pm\delta_c$  (e.g.,  $\alpha = 0.05$ ),  $1 - \alpha/C$  is the level of confidence. For example, if  $\alpha = 0.05$  and  $C = 5$ , then  $1 - 0.05/5 = 0.99$ , and  $\chi^2_{(1,1-\alpha/C)}$  is the upper  $(1 - \alpha/C)$  \* 100<sup>th</sup> percentile of the chi-square distribution with one degree of freedom. If the level of confidence is  $(1 - 0.05/5) = 0.99$ , then  $\chi^2_{(1,0.99)} = 6.63$ . In the contingency table shown in Table 4-2, where  $C = 12$ ,  $\alpha$  was chosen equal to 0.12, so that the level of confidence  $1 - \alpha/C = 0.99$ , hence  $\chi^2_{(1,0.99)} = 6.63$ .

Built upon an EO-IU subsystem whose “default” algorithms are automated, an EO-SQ subsystem prototype was implemented as a proof-of-concept. One example of a semantic spatiotemporal query formulated by a user through the EO-SQ subsystem’s GUI to the fact base of EO images and products is shown (Fig. 4-38). In this GUI the user selects the AOI (1), the target time period (2) and creates or selects via graphical elements the decision-rule pipeline for semantic querying the fact base (3). Once it is executed, the query results are shown in the image domain (4) and as summary statistics. Geospatial results can be downloaded as a GeoTiff file and/or input to further queries. The query itself can be saved and re-used on different AOIs and target time periods.

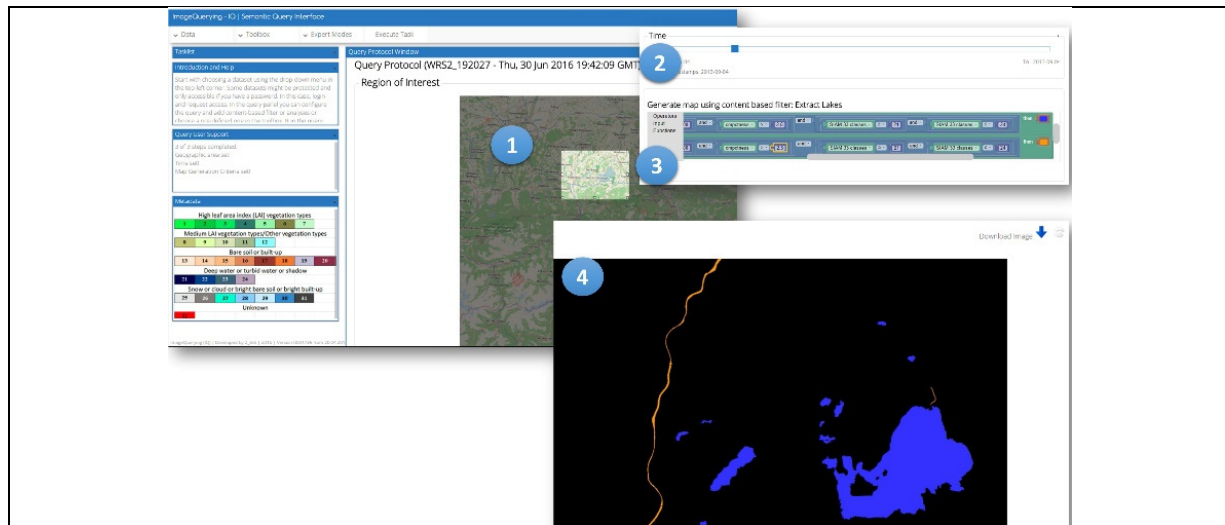




Fig. 4-38. Example of spatiotemporal semantic querying to infer new information layers from the fact base: splitting of the LCCS-DP class B48, Natural Waterbodies, into the LCCS-MHP classes Lake and River based on planar shape and size properties. In the EO-IU4SQ's GUI, the user selects (1) the area of interest (AOI), (2) the target time period and (3) creates or selects via graphical elements the decision-rule pipeline for semantic querying the fact base. Once it is executed, the query results are shown in the image domain and/or as summary statistics (4).

#### 4.4. Discussion

According to the Group on Earth Observation (GEO), in Stage 3 *Val* “spatial and temporal consistency of the product with similar products are evaluated by independent means over multiple locations and time periods representing global conditions. In Stage 4 validation, results for Stage 3 are systematically updated when new product versions are released and as the time-series expands” (Group on Earth Observation, 2015).

In agreement with the U.S. Geological Survey (USGS) classification system guidelines (Lillesand & Kiefer, 1979), a target  $OA \in [0, 1] \pm \delta_{OA}$  should be fixed within range  $[0.80, 0.85]$  with a reasonable degree of uncertainty in measurement equal to  $\pm 2\%$ . A class-specific classification accuracy,  $CA_c \in [0, 1] \pm \delta_c$ ,  $c = 1, \dots, C$ , should be about equal across classes and never below 70%, whereas a reasonable reference standard for  $\delta_c$  is about 5% (Baraldi, Humber, & Boschetti, 2014).

To comply with the GEO Stage 4 *Val* requirements, the EO-IU4SQ system's OP-Q<sup>2</sup>Is should be validated by independent means on large-scale EO image time-series. This would be a huge outcome and process *Val* task, to be mainly focused on accuracy Q<sup>2</sup>Is to be validated in comparison with those of state-of-the-art alternative approaches, if any (Baraldi et al., 2016). With regard to the implemented ESA EO Level 2 SCM generator, an accuracy test was conducted by dependent means (eventually biased), specifically, by the same authors of the outcome and process under testing. The CMTRX shown in Table 4-2 features an  $OA \pm \delta_{OA} = 0.92 \pm 0.02$ , with level of significance  $\alpha = 0.12$ , where most of the inter-class confusion is between Class #1 Vegetation Terrestrial Managed Low Texture and Class #2 Vegetation Terrestrial Natural/Semi-natural Low Texture. These mapping accuracy results appear reasonable, which means in line with theoretical expectations if we consider that, in this CMTRX, the reference sample consists of 660 pixels manually selected and labeled in the Sentinel-2A image of Austria, shown in Fig. 4-31(a), by photointerpreters who were the same (potentially biased) authors of the ESA EO Level 2 SCM product under testing and who had to provide sub-symbolic pixels with semantics according to the thematic map legend shown in Fig. 4-32. This inherently ill-posed (equivocal, subjective) *information-as-data-interpretation* process was made even more difficult by the low-level (vague, ambiguous) semantics featured by the target dictionary of LC class names shown in Fig. 4-32.

In Chapter 4.3, the mDMI set of OP-Q<sup>2</sup>I values reported for the implemented low-level CV system for automated near real-time raw and full primal sketch generation can be summarized as shown in Table 4-3. These OP-Q<sup>2</sup>Is comply with the primary EO-IU subsystem requirements specification proposed in Chapter 4.1.

In agreement with the QA4EO *Val* requirements, the mDMI set of OP-Q<sup>2</sup>I values of the automated near real-time ESA EO Level 2 SCM generator can be summarized as shown in Table 4-4. Noteworthy, the prototypical implementation of the tested ESA EO Level 2 SCM generator did not employ as input information components either local shape, spatial topological or spatial non-topological information, and did not employ as quality layers any cloud and cloud-shadow detector either. Nevertheless, OP-Q<sup>2</sup>Is shown in Table 4-4 comply with the primary EO-IU subsystem requirements specification proposed in Chapter 4.1, with the sole exception of the semantic information level, inferior to the target 3-level 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow. This semantic information gap is expected to be filled in as soon as the missing visual information primitives, already made available as described in Chapter 4.2.3.2, will be provided as input to the EO-IU subsystem prototype.





Legend of fuzzy sets of a  
quantitative variable.

LOW
MEDIUM
HIGH

Computer vision (CV) Process (Pracs) and Outcome (Otcn) Q <sup>2</sup> Is ± δ ⊆ QA4EO Val	Low-level vision: raw/full primal sketch
<b>Degree of automation (Pracs):</b> (a) inversely related to the number, physical meaning and range of variation of user-defined parameters, (b) inversely related to the collection of the required training data set, if any.	HIGH (unsurpassed, no free-paramtr)
<b>Effectiveness (Otcn), in agreement with human visual perception, i.e., CV ⊃ human vision, where human visual perception is a lower bound of CV.</b>	
a) Color constancy or radiometric calibration (when radiometric calibration metadata are available)	a) HIGH
b) 2D image analysis/ Retinotopic visual information representation/ Topology-preserving visual feature mapping / Spatial topological information extraction. > Necessary not sufficient condition: panchromatic vision performs nearly as well as chromatic vision.	b) YES > HIGH
c) Pre-attentive image contour detection/ image segmentation quality, consistent with the Mach bands illusion in ramp-edge detection: spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	c) HIGH
d) High-level vision (classification). (a) thematic Qis (TQIs) and (b) spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	d) None
<b>Semantic information level (Pracs)</b>	LOW (sub-symbolic)
<b>Efficiency (Pracs):</b> (a) computation complexity in image size: Polynomial (P), P and linear (L), non-P (NP), and (b) run-time memory occupation.	(a) HIGH (L), (b) HIGH
<b>Robustness to changes in input image (Pracs)</b> , e.g., large spatial extent data mapping (no toy problems).	HIGH
<b>Robustness to changes in input parameters (Pracs)</b> , e.g., sensitivity analysis.	HIGH (unsurpassed, no free-paramtr)
<b>Scalability to changes in the sensor's specifications or user's needs (Pracs)</b> , e.g., (a) pncchrmtc, (b) RGB, true- or false-color, (c) multi-spectral (MS), (d) super-spectral (SS), (e) hyper-spectral (HS).	HIGH, (a) YES, (b) YES, (c) YES, (d) YES, (e) YES.
<b>(Inverse of) Timeliness (Otcn)</b> , from data acquisition to high-level product generation, increases with manpower and computing power.	HIGH
<b>(Inverse of) Costs (Otcn)</b> , increasing with (a) manpower and (b) computing power.	(a) HIGH, (b) HIGH

Table 4-3. Outcome and process (OP) quantitative quality indicators (OP-Q<sup>2</sup>Is) of the proposed low-level CV system design and implementation for automated near real-time raw and full primal sketch generation.

#### 4.5. Conclusions

Although it has not been submitted to a Stage 4 *Val* process yet, the proposed closed-loop EO-IU4SQ system prototypical implementation works as a proof-of-concept of a GEOSS in support of SCBIR, where SCBIR ⊃ CV ⊃ EO-IU in operating mode, synonym of GEOSS ⊃ systematic ESA EO Level 2 product generation ⊃ human vision, i.e., human visual perception is adopted as a lower bound of CV in operating mode. The closed-loop EO-IU4SQ system modular design comprises two information processing subsystems, a primary (dominant, necessary not sufficient) EO-IU subsystem in operating mode and a secondary (dependent) EO-SQ subsystem. The former consists of a multi-source hybrid feedback inference system featuring linear complexity in image size and requiring no human-machine interaction to systematically generate ESA EO Level 2 product from EO big data as necessary not sufficient initial condition for EO-SQ operations. Collected OP-Q<sup>2</sup>I values agree with the project objectives, working hypotheses and with the GEOSS visionary goals (refer to Chapter 4.1). In terms of degree of automation, efficiency, robustness to changes in input parameters, robustness to changes in the input data set, scalability to cope with changes in imaging sensor specifications, timeliness and costs in manpower and computer power, collected OP-Q<sup>2</sup>I values appear superior to those of traditional EO-IUSs and SCBIR system prototypes, such as EOLib, which are built upon an inductive feedforward inference system for 1D image classification, e.g., an SVM, known to be inherently semi-automatic, training data-dependent and insensitive to spatial topological information which typically dominates color information in vision (refer to Chapter 4.1).

In future works, first, within the proposed EO-IU4SQ system architecture (Fig. 4-6 and Fig. 4-7), any information processing block will be considered eligible for permanent optimization, modification or replacement to improve the mDMI set of OP-Q<sup>2</sup>Is of the EO-IU and EO-SQ subsystem implementations to be considered in operating mode (refer to Chapter 4.1). Second, in the EO-SQ subsystem the semantic network formalism required to graphically represent the world model will be augmented and integrated with an algebra capable of describing spatiotemporal data types and operations in a language-independent and formal way (Ferreira et al., 2014). Third, an EO-IU4SQ system Stage 4 *Val* plan by independent means over multiple locations and time periods representing global conditions will be scheduled and pursued in agreement with the QA4EO guidelines.



Legend of fuzzy sets of a  
quantitative variable.

LOW
MEDIUM
HIGH

Computer vision (CV) Process (Prcs) and Outcome (Otcn) Q <sup>2</sup> Is ± δ ⊆ QA4EO Va/	ESA EO Level 2 product generator
<b>Degree of automation (Prcs):</b> (a) inversely related to the number, physical meaning and range of variation of user-defined parameters, (b) inversely related to the collection of the required training data set, if any.	HIGH
<b>Effectiveness (Otcn), in agreement with human visual perception, i.e., CV ⊃ human vision, where human visual perception is a lower bound of CV.</b>	
a) Color constancy or radiometric calibration (when radiometric calibration metadata are available)	a) HIGH
b) 2D image analysis/ Retinotopic visual information representation/ Topology-preserving visual feature mapping / Spatial topological information extraction. > Necessary not sufficient condition: panchromatic vision performs nearly as well as chromatic vision.	b) YES > HIGH
c) Pre-attentive image contour detection/ image segmentation quality, consistent with the Mach bands illusion in ramp-edge detection: spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	c) HIGH
d) High-level vision (classification). (a) thematic Qis (TQIs) and (b) spatial Qis (SQIs), provided with a degree of uncertainty in measurement ±δ.	d) HIGH
<b>Semantic information level (Prcs)</b>	MEDIUM/LOW (LEVEL 2 SCM legend)
<b>Efficiency (Prcs):</b> (a) computation complexity in image size: Polynomial (P), P and linear (L), non-P (NP), and (b) run-time memory occupation.	(a) HIGH (L), (b) HIGH
<b>Robustness to changes in input image (Prcs)</b> , e.g., large spatial extent data mapping (no toy problems).	HIGH
<b>Robustness to changes in input parameters (Prcs)</b> , e.g., sensitivity analysis.	HIGH (unsurpassed, no free-paramtr)
<b>Scalability to changes in the sensor's specifications or user's needs (Prcs)</b> , e.g., (a) pncchrmtc, (b) RGB, true- or false-color, (c) multi-spectral (MS), (d) super-spectral (SS), (e) hyper-spectral (HS).	HIGH, (a) YES, (b) YES, (c) YES, (d) YES, (e) YES.
<b>(Inverse of) Timeliness (Otcn)</b> , from data acquisition to high-level product generation, increases with manpower and computing power.	HIGH
<b>(Inverse of) Costs (Otcn)</b> , increasing with (a) manpower and (b) computing power.	(a) HIGH, (b) HIGH

Table 4-4. Outcome and process (OP) quantitative quality indicators (OP-Q<sup>2</sup>Is) of the proposed EO-IU subsystem design and implementation for systematic ESA EO Level 2 product generation.

### Acknowledgements

The test implementation of the EO-IU4SQ system, called Image Querying system, was awarded 1st place in the Copernicus Masters 2015 - T-Systems Big Data Challenge. The study was supported by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel2/3 (Knowledge-based pre-classification of Sentinel-2/3 images for operational product generation and content-based image retrieval), ID 848009, and project SemEO (Semantic enrichment of optical EO data to enhance spatio-temporal querying capabilities). Andrea Baraldi thanks Prof. Raphael Capurro for his hospitality, patience, politeness and open-mindedness. He also thanks Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, for his friendship and support. The authors wish to thank the Editor-in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.

### References in Chapter 4

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Susstrunk, S. (2011). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(1), 1-8.
- Adams, J. B., Donald, E. S., Kapos, V., Almeida Filho, R., Roberts, D. A., Smith, M. O., & Gillespie, A. R. (1995). Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote Sens. Environ.*, 52, 137-154.
- Baatz, M. & Schäpe, A. (2000). Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. In *Angewandte Geographische Informationsverarbeitung XII*; Strobl, J., Ed.; Herbert Wichmann Verlag: Berlin, Germany, 58: 12–23.
- Baraldi, A. (2015). "Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+," Invitation to tender ESA/AO/1-8373/15/I-NB – "VAE: Next Generation EO-based Information Services", DOI: 10.13140/RG.2.2.34162.71363. arXiv: 1701.04256. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.04256.pdf>
- Baraldi, A., & Boschetti, L. (2012a). Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction. *Remote Sens.*, 4, 2694-2735.



- Baraldi, A., & Boschetti, L. (2012b). Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation. *Remote Sens.*, 4, 2768-2817.
- Baraldi, A., Durieux, L., Simonetti, D., Conchedda, G., Holecz, F., & Blonda, P. (2010a). Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation,” *IEEE Trans. Geosci. Remote Sensing*, 8(3), 1299 - 1325.
- Baraldi, A., Humber, M., & Boschetti, L. (2014). Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images. *IEEE Trans. Geosci. Remote Sensing*, 52(1), Part: 2, 701-760.
- Baraldi, A., & Soares, J. (2017). Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis. Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01941. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01941.pdf>
- Baraldi, A., Gironde, M., & Simonetti, D. (2010b). Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier. *IEEE Trans. Geosci. Remote Sensing*, 48(1), 112-146.
- Baraldi, A., Humber, M. L., Tiede, D., & Lang, S. (2016). Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation – Part 1: Theory, Part 2 - Validation. arXiv:1701.01932. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01932.pdf>
- Baraldi, A., & Parmiggiani, F. (1996). Combined detection of intensity and chromatic contours in color images. *Optical Eng.*, 35(5), 1413-1439.
- Baraldi, A., Puzzolo, P., Blonda, P., Bruzzone, L., & Tarantino, C. (2006). Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. *IEEE Trans. Geosci. Remote Sens.*, 44, 2563-2586.
- Baraldi, A., Tiede, D., & Lang, S. (2017). Automated Linear-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images. arXiv:1701.01940. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01940.pdf>
- Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., Bigagli L, et al. (2015). Big data analytics for earth sciences: the EarthServer approach. *International Journal of Digital Earth*, 1–27.
- Berlin B., & Kay, P. (1969). Basic color terms: their universality and evolution. *Berkeley: University of California*.
- Bertero, M., Poggio, T., & Torre, V. (1988). Ill-posed problems in early vision. *Proc. IEEE*, 76, 869–889.
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., & Tiede, D. (2014). Geographic object-based image analysis - towards a new paradigm. *ISPRS J. Photogram. Remote Sens.*, 87, 180–191.
- Bossard M., Feranec J., & Othel J. (2000). *CORINE land cover technical guide—addendum 2000*. Technical Report, 40, EEA
- Burr, D. C., & Morrone, M. C. (1992). A nonlinear model of feature detection, in *Nonlinear Vision: Determination of Neural Receptive Fields, Functions, and Networks*, R. B. Pinter and N. Bahram, Eds., 309–327, CRC Press, Boca Raton, FL.
- Camara, G., Souza, R., Freitas, U., & Garrido, J. (1996). SPRING: Integrating remote sensing and GIS by object-oriented data modelling, *Computers & Graphics*, 20(3): 395-403.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6), 679–698.
- Capurro, R. & Hjørland, B. (2003). The concept of information. *Annual Review of Information Science and Technology*, 37, 343-411.
- Castelletti, D., Pasolli, L., Bruzzone, L., Notarnicola, C., & Demir, B. (2016). A novel hybrid method for the correction of the theoretical model inversion in bio/geophysical parameter estimation. *IEEE Trans. Geosci. Remote Sens.*, XX(Y), 1–11.
- Cherkassky, V., & Mulier, F. (1998). *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley.
- Cimpoi M., Maji S., Kokkinos I., & Vedaldi A. (2014). Deep filter banks for texture recognition, description, and segmentation. CoRR, abs/1411.6836, 2014.
- Clausi, D., & Yongping Zhao (2003). Gray level co-occurrence integrated algorithm (GLCIA): a superior computational method to rapidly determine co-occurrence probability texture features. *Computers & Geosciences*, 29, 837-850.



- Di Gregorio, A., & Jansen, L. (2000). *Land Cover Classification System (LCCS): Classification Concepts and User Manual*. FAO: Rome, Italy, FAO Corporate Document Repository. [Online]. Available: <http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm>
- Dillencourt, M. B., Samet, H., & Tamminen, M. (1992). A general approach to connected component labeling for arbitrary image representations. *J. Assoc. Computing Machinery*, 39, 253-280.
- Dumitru, C. O., Cui, S., Schwarz, G., & Datcu, M. (2015). Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies. *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.*, 8(4), 1635 – 1650.
- du Buf, H., & Rodrigues, J. (2007). Image morphology: from perception to rendering. In *IMAGE - Computational Visualistics and Picture Morphology*.
- Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., & Monteiro, A. M. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation, *Int. J. Remote Sens.* 27(14): 3035–3040.
- European Space Agency (2015). *Sentinel-2 User Handbook, Standard Document*. Issue 1 Rev 2.
- European Space Agency (2002). D'Elia, S. Personal communication.
- Ferreira, K. R., Camara, G., & Monteiro, A. M. V. (2014). An algebra for spatiotemporal data: From observations to events. *Trans. in GIS*, 1(2), 253–269.
- Finlayson, G. D., Hordley, S. D., & Hubel, P. M. (2001). Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(11), 1209-1221.
- Frintrop, S. (2011). Computational visual attention. In *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds., Springer.
- Gijssen, A., Gevers, T., & van de Weijer, J. (2010). Computational color constancy: Survey and experiments. *IEEE Trans. Image Proc.*, 20(9), 2475-2489.
- Goodchild, M. F., Yuan M., & Cova, T. J. (2007). Towards a general theory of geographic representation in GIS, *International Journal of Geographical Information Science*, 21:3, 239-260.
- Griffin, L. D. (2006). Optimality of the basic color categories for classification. *J. R. Soc. Interface*, 3, 71–85.
- Group on Earth Observation (GEO). (2005). *The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan*. Retrieved January 10, 2012, from <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>
- Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS) (2010). *A Quality Assurance Framework for Earth Observation, version 4.0*. Retrieved November 15, 2012, from [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- Group on Earth Observation (2015). *Land Product Validation (LPV), Committee on Earth Observation Satellites (CEOS) - Working Group on Calibration and Validation (WGCV)*. [Online]. Available: <http://lpvs.gsfc.nasa.gov/>. Accessed on March 20, 2015.
- Grove, S. (1999). Knowledge-based interpretation of multisensor and multitemporal remote sensing images. *Int. Archives of Photogrammetry and Remote Sensing*, 32, Part 7–4–3 W6, Valladolid, Spain, 3–4 June, 1999.
- Hadamard, J. (1902). Sur les problemes aux derivees partielles et leur signification physique. *Princet. Univ. Bull.*, 13, 49–52.
- Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., & Kubler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res.*, 32(5), 963–981.
- Hunt, N., & Tyrrell, S. (2012). *Stratified Sampling*. Coventry University. [Online] Available: <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>.
- ITT Visual Information Solutions (2009). *ENVI EX User Guide 5.0*. [Online]. Available: [http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI\\_EX\\_User\\_Guide.pdf](http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf)
- Lillesand, T., & Kiefer, R. (1979). *Remote Sensing and Image Interpretation*, New York: John Wiley & Sons.
- Lunetta, R.S., and Elvidge, C.D. (1998). *Remote sensing and Change Detection: Environmental Monitoring Methods and Applications*. Chelsea, Michigan: Ann Arbor Press.
- Julesz, B., Gilbert, E. N., Shepp, L. A., & Frisch, H. L. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited. *Perception*, 2, 391-405.
- Kosslyn, S. M. (1994). *Image and Brain*. MIT Press, Cambridge, MA.
- Laurini, R. & Thompson, D. (1992). *Fundamentals of Spatial Information Systems*. London, UK: Academic Press.
- Liang, S. (2004). *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons.





- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2), 91–110.
- Marr, D. (1980). *Vision*. New York, NY: Freeman and C.
- Mason, C. & Kandel, E. R. (1991). Central visual pathways. In *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 420–439.
- Mather, P. (1994). *Computer Processing of Remotely-Sensed Images—An Introduction*. Hoboken, NJ, USA: Wiley.
- Matsuyama, T. & Hwang, V.S. (1990). *SIGMA – A Knowledge-based Aerial Image Understanding System*. New York, NY: Plenum Press.
- Mayo, M. (2003). Symbol grounding and its implications for artificial intelligence, *26th Australian Computer Science Conference (ACSC2003)*, Adelaide, Australia, 16, 2003.
- Nagao, M., & Matsuyama, T. (1980). *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York.
- National Aeronautics and Space Administration (NASA) (2016). Data Processing Levels. [Online]. Available: <https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products>. Accessed on December 20, 2016.
- Nestares, O., Navarro, R., Portilla, J., & Taberner, A. (1998). Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *J. of Electronic Imaging*, 7(1), 166–173.
- Open Source Computer Vision Library (OpenCV)* (2015). [Online]. Available: <http://opencv.org/>. Accessed on March 20, 2015.
- Open Geospatial Consortium (OGC) Inc. (2015). *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. [Online]. Available: <http://www.opengeospatial.org/standards/is>
- Pessoa, L. (1996). Mach Bands: How Many Models are Possible? Recent Experimental Findings and Modeling Attempts. *Vision Res.*, 36(19), 3205–3227.
- Richter, R., & Schlöpfer, D. (2012). *Atmospheric / Topographic Correction for Satellite Imagery – ATCOR-2/3 User Guide, Version 8.2 BETA*. [Online] Available: [http://www.dlr.de/eoc/Portaldat/60/Resources/dokumente/5\\_tech\\_mod/atcor3\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldat/60/Resources/dokumente/5_tech_mod/atcor3_manual_2012.pdf)
- Rodrigues, J., & Hans du Buf, J. M. (2008). Multi-scale lines and edges in V1 and beyond: Brightness, object categorization and recognition, and consciousness. *BioSystems*, xxx, 1-21.
- Roy, D., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., Loveland, T. R., Vermote, E. & Zhang, C. S. (2010). Web-enabled Landsat Data (WELD): Landsat ETM plus composited mosaics of the conterminous United States, *Remote Sens. Environ.*, 114, 35-49.
- Schaepman-Strub G., Schaepman M.E., Painter T.H., Dangel S., & Martonchik J.V. (2006). Reflectance quantities in optical remote sensing—Definitions and case studies. *Remote Sens. Environ.*, 103, 27–42.
- Serra, R. & Zanarini, G. (1990). *Complex Systems and Cognitive Processes*, Berlin: Springer-Verlag.
- Shyu, C.-R., Klaric, M., Scott, G. J., Barb, A. S., Davis, C. H., & Palaniappan, K. (2007). GeoIRIS: Geospatial Information Retrieval and Indexing System—Content mining, semantics modeling, and complex queries. *IEEE Trans. Geosci. Remote Sens.*, 45(4), 839–852.
- Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas, *Cerebral Cortex*, 15, 1570-1583.
- Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(12), 1349-1380.
- Smith, S. M., & Brady, J. M. (1997). SUSAN - a new approach to low level image processing. *Int. J. of Computer Vision*, 23(1), 45--78.
- Soares, J. V. B., Baraldi, A., & Jacobs, D. W. (2014). Segment-based simple-connectivity measure design and implementation. *Tech. Rep., University of Maryland, College Park*. [Online]. Available: <http://hdl.handle.net/1903/15430>.
- Sonka, M., Hlavac, V., & Boyle, R. (1994). *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall.
- Tiede, D., Baraldi, A., Sudmanns, M., Belgiu, M., & Lang, S. (2016). Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases, *European J. Remote Sens.*, submitted for consideration for publication.
- Torre, V., & Poggio, T. (1986). On edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(2), 147–163.
- Trimble (2015). *eCognition® Developer 9.0 Reference Book*.
- Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13, 423-469.



- Vecera, S., & Farah, M. (1997). Is visual image segmentation a bottom-up or an interactive process?. *Percept. Psychophys.*, 59, 1280–1296.
- Vermote, E., & Saleous, N. (2007). *LEDAPS surface reflectance product description - Version 2.0*, University of Maryland at College Park /Dept Geography and NASA/GSFC Code 614.5.
- Vo, A.-V., Truong-Hong, L., Laefer, D. F., Tiede, D., d'Oleire-Oltmanns, S., Baraldi, A., Shimoni, M. (2016). Processing of extremely high resolution LiDAR and optical data: Outcome of the 2015 IEEE GRSS Data Fusion Contest. Part-B: 3D contest. *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.*, vol. 9, no. 12, pp. 5560-5575, Dec. 2016.
- Yuille, A. L., & Poggio, T. (1986). Fingerprints theorems for zero-crossings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(1), 15–25.
- Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control*, 8, 338–353.
- Zhe Zhu & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83–94.

## 5 Manuscript 2 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal *European Journal of Remote Sensing*): Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases

### Motivation and Contributions to the Dissertation

An innovative closed-loop Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system prototype was proposed as a proof-of-concept of a Global Earth Observation System of Systems (GEOSS) in support of semantic content-based image retrieval (SCBIR). It comprises, first, an EO-IU subsystem mainly described in Chapter 4 (Manuscript 1), and, second, an EO-SQ subsystem prototype for semantic content-based image retrieval (SCBIR), which is the focus of attention of the present Chapter 5 (Manuscript 2).

In Chapter 1 (Doctoral Research Objectives), the closed-loop EO-IU4SQ system architecture was sketched in Fig. 1-9, reported hereafter for the sake of clarity.

It is worth mentioning that the EO-IU4SQ system prototype research and development (R&D) was funded in part by:

- 2015-2016. Austrian Research Promotion Agency (FFG), project call ASAP-11, AutoSentinel-2/3 project (Knowledge-based pre-classification of Sentinel-2/3 images for operational product generation and content-based image retrieval).
- 2016-2017. Austrian Research Promotion Agency (FFG), project call Proposals to ICT of the Future, SemEO project (Semantic enrichment of optical EO data to enhance spatio-temporal querying capabilities).

The EO-IU4SQ system prototype was awarded 1st place in the T-Systems Big Data Challenge of the Copernicus Masters 2015 (Awards Ceremony at the Satellite Masters Conference, 20-22 Oct. 2015, German Federal Ministry of Transport and Digital Infrastructure, Berlin, Germany).

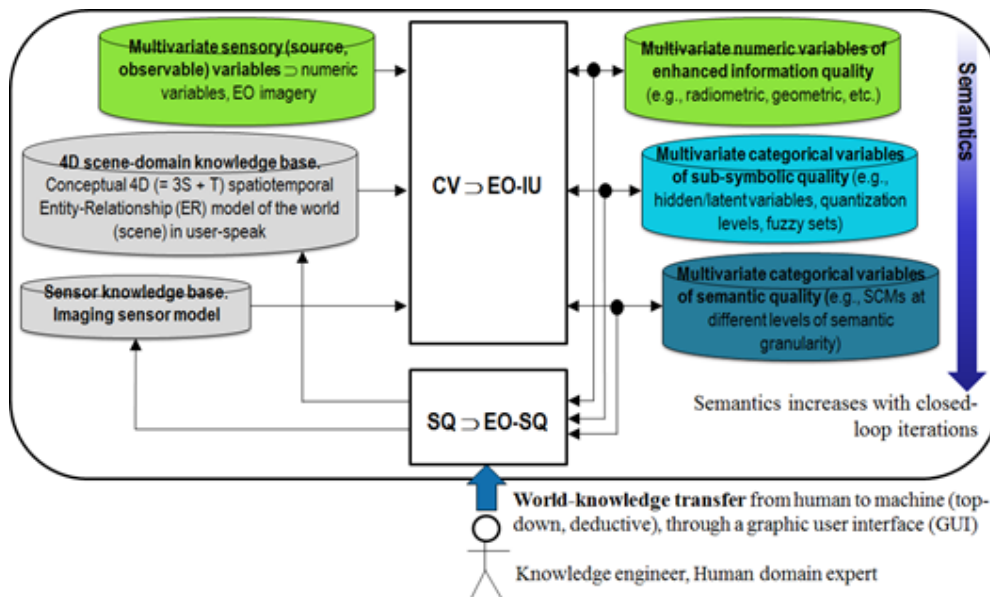


Fig. 1-9. Top-level modular design of a closed-loop EO image understanding (EO-IU) for semantic querying (EO-IU4SQ) system architecture, suitable for incremental learning. It comprises a primary (dominant, necessary not sufficient) hybrid (combined deductive and inductive) EO-IU subsystem in closed-loop with a secondary (dominated) hybrid EO-SQ subsystem. The EO-IU subsystem must be automatic (requiring no human-machine interaction) and near real-time to provide the EO-SQ subsystem with useful information products, including Scene Classification Maps (SCMs) of symbolic quality, as initial necessary not sufficient pre-condition for semantic querying and semantics-enabled information/knowledge discovery. The EO-SQ subsystem is provided with a graphic user interface (GUI) to streamline high-level user- and application-specific semantic querying and semantics-enabled information/knowledge discovery. Output products generated by the closed-loop EO-IU4SQ system are expected to monotonically increase their value-added with closed-loop iterations.



# Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases

Dirk Tiede<sup>1\*</sup>, Andrea Baraldi<sup>2</sup>, Martin Sudmanns<sup>1</sup>, Mariana Belgiu<sup>1</sup> and Stefan Lang<sup>1</sup>

<sup>1</sup>Department of Geoinformatics – Z\_GIS, University of Salzburg, 5020 Salzburg, Austria

<sup>2</sup>Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

\*Corresponding author, e-mail address: dirk.tiede@sbg.ac.at

## Abstract

Spatiotemporal analytics of multi-source EO big data is a pre-condition for semantic content-based image retrieval (SCBIR). As a proof-of-concept, an innovative Earth observation (EO) semantic querying (EO-SQ) subsystem was designed and prototypically implemented in series with an EO image understanding (EO-IU) subsystem. The EO-SQ subsystem comprises a graphical user interface (GUI) and an array database embedded in a client server model. In the array database, every EO image is stored as a space-time data cube together with its Level 2 products automatically generated by the EO-IU subsystem. The GUI allows users to: (a) develop a conceptual world model, graphically represented as a semantic network with land cover classes as nodes and inter-class spatiotemporal relationships as arcs between nodes, and (b) create, save and share within the client-server architecture complex semantic queries/decision rules, suitable for SCBIR and/or spatiotemporal EO image analytics, consistent with the world model.

**Keywords:** Big data, Earth observation, ESA EO Level 2 product, spatiotemporal objects and events, array database, semantic content-based image retrieval.

## 5.1 Introduction

### 5.1.1 Semantic content-based image querying

Vision (image understanding) is a cognitive process responsible of scene-from-image representation, ranging from local syntax of individual objects to global gist and spatial layout of objects in the 4D spatiotemporal scene domain, including multiple plausible semantic scene interpretations and even emotions [du Buf and Rodrigues, 2007]. Vision is inherently ill-posed, affected by a 4D to 2D data dimensionality reduction problem, responsible of occlusion phenomena, and by a semantic information gap, from quantitative sub-symbolic ever-varying sensations, specifically, 2D sensory data in the image domain, to stable symbolic percepts in the 4D spatiotemporal scene domain [Matsuyama and Hwang, 1990].

Traditional content-based image retrieval (CBIR) systems, including photographic image search engines such as the Google Image Search, support human-machine interaction through queries by metadata text information, image-wide summary statistics or by either image, image-object or multi-object examples [Gudivada and Raghavan, 1995, Seidel et al., 1998; Shyu et al., 2007; Smeulders et al., 2000]. Queries by image examples typically compute statistical similarities between numeric sub-symbolic low-level vision variables (e.g., pixel values, texture parameters, etc.) extracted from a reference and a test image pair without extracting high-level vision semantics for scene-from-image reconstruction.

In the remote sensing (RS) domain, Earth observation (EO) CBIR systems support human-machine interaction through queries by either metadata text information or image-wide summary statistics [Shyu et al., 2007]. For example, the popular U.S. Geological Survey (USGS) Earth Explorer and the European Space Agency (ESA) Sentinels Scientific Data Hub allow a user to choose a geographic area of interest (AOI), a target timespan and some textual metadata, such as the name of the mission, image path/row, data category, etc. The sole EO image filtering criterion they support is input with a maximum cloud-cover percentage value to be user-defined. This scalar threshold is compared with a cloud-cover quality index estimated off-line for every EO image stored in the database. Any image-wide quality index (summary statistic) neither refers to the user-selected AOI nor provides information about its geospatial distribution across the AOI.

Unlike a traditional CBIR system, a semantic CBIR (SCBIR) system is expected to cope with spatiotemporal semantic queries such as “retrieve all images in the database where a lake is not covered by clouds and larger than a certain area”. Such an SCBIR system must rely on image understanding (vision) as a pre-condition for SCBIR. This makes the SCBIR problem at least as difficult (or ill-posed) as vision. Since computer vision is still an open problem, this may explain why





very few SCBIR system prototypes have been presented in the remote sensing (RS) and computer vision literature [Li and Bretschneider, 2004; Dumitru et al., 2015]. To our best knowledge no SCBIR system in operating mode is available to date. We define an information processing system in operating mode if it scores “high” in a minimally dependent and maximally informative (mDMI) set of outcome and process quality indicators (QIs), encompassing accuracy, efficiency, degree of automation, scalability, robustness to changes in input data as well as to changes in input parameters, timeliness from data acquisition to product generation and costs in manpower and computer power [Baraldi and Boschetti 2012].

In the RS domain, the EO SCBIR system prototype proposed by Li and Bretschneider (2004) adopts, first, a relational database to store planar information layers (symbolic strata) in a scene classification map (SCM) whose legend (codebook) is a dictionary of target land cover (LC) classes (codewords). Second, it employs a semi-automatic pixel-based statistical classifier. Because colour information is the sole visual feature available at pixel resolution, any pixel-based classifier ignores spatial information in the image domain. This is in contrast with the fact that, since chromatic and achromatic biological vision systems are nearly as effective in scene-from-image representation, spatial information dominates colour information in both the (2D) image domain and the 4D spatiotemporal scene domain [Matsuyama and Hwang, 1990]. An alternative EO SCBIR prototype, called Earth Observation Image Librarian (EOLib), was recently proposed by Dumitru et al. [2015]. EOLib is built upon a support vector machine (SVM) for 1D image classification, where the 1D vector data sequence consists of image convolutional values generated by 2D spatial filters. Any inductive learning-from-data algorithm, such as an SVM, requires *a priori* knowledge in addition to data to become better conditioned for numerical solution [Cherkassky and Mulier, 1998]. As a consequence, EOLib can be considered as semi-automatic only. In addition, any 1D image classifier is an orderless encoder invariant to permutations, where spatial topological information in the (2D) image domain is lost.

Our conjecture is that existing EO CBIR systems support no semantic querying because they lack EO image understanding capabilities, where spatial information, either topological (e.g., adjacency, inclusion, etc.) or non-topological (e.g., metric distance, angle measure, etc.), dominates colour information [Matsuyama and Hwang, 1990], in agreement with the object-based image analysis (OBIA) paradigm [Blaschke et al., 2014]. This is tantamount to saying that a necessary not sufficient pre-condition for SCBIR system development in operating mode is the computational solution of the inherently ill-posed vision problem [Matsuyama and Hwang, 1990]. Existing EO image understanding systems (EO-IUSs) fall short in transforming multi-source EO big data into comprehensive, timely and operational information products. For example, no European Space Agency (ESA) EO data-derived Level 2 prototype product has ever been generated systematically at the ground segment [European Space Agency, 2015]. By definition an ESA EO Level 2 product consists of: (a) an enhanced EO image corrected for atmospheric and topographic effects, and (b) a general-purpose, user- and application-independent SCM, whose legend includes quality layers, such as cloud and cloud-shadow [European Space Agency, 2015].

### 5.1.2 Scalable processing

Existing EO CBIR systems handle large-volume EO imagery as flat files in relational databases. Array databases provide a valuable alternative to classical file handling. First, an array database can be queried by means of a declarative query language analogous to the Standard Query Language (SQL). Thus, the array database is able to perform internal optimizations, such as identifying the best access patterns using query plans, which can lead to a significant improvement in efficiency [Baumann and Holstein, 2010]. Second, input/output (IO) activities can be reduced by using a data model optimized regarding the data source, expected queries and the indexing of the database contents. Query examples are time series analyses where potentially large amount of EO images have to be interpreted in a specified AOI smaller than one image and/or across a time interval spanning multiple images. In these query examples the handling of files using file-system and operating-system capabilities (such as locking) is inferior to the use of array databases, since it does not easily support typical concurrent large-scale processing properties such as query planning, load balancing, transactions or database indexing. Additional properties of databases can be exploited in an image querying application. For example, the technical and logical separation from the application logic, and, relevant for multi-user access, a better security through a finer granulated user rights management, a transaction manager which allows concurrent queries and database-inherent backup capabilities [Brinkhoff and Kresse, 2011]. Although applications of array databases to EO (S)CBIR systems are rare, some promising works have been carried out in recent years [Planthaber et al., 2012]. Examples are the EarthServer [Baumann et al., 2015] and the Australian Geoscience Data Cube [Purss et al., 2015]. In the EarthServer, the array database implementation known as Rasdaman [Baumann et al., 2015] proved its capability to handle large-volume EO image bases. In the remainder of this paper an innovative EO-SQ system prototype, called *ImageQuerying* (IQ), is proposed as a proof-of-concept to work in series with an EO-IU subsystem, capable of multi-source EO big data spatiotemporal analytics as a pre-condition for SCBIR. The combined EO-SQ and EO-IU subsystems form an integrated EO Image Understanding and



Semantic Querying (EO-IU&SQ) system. Within the EO-IU&SQ system design, the EO-SQ/IQ subsystem prototypical implementation in a distributed client-server architecture will be discussed in details.

## 5.2 EO-IU&SQ system design

An integrated EO-IU&SQ system architecture [Baraldi et al., 2016] consists of two subsystems: the EO-IU subsystem (section (1) in Fig. 5-1) and the EO-SQ subsystem (sections (2) and (3) in Fig. 5-1). These two subsystems share: (a) a fact base, where each EO image is stored together with its information products; (b) a knowledge base, encompassing physical laws, first-principle models, if-then decision rules, methods, processes, etc., eligible for generating new information from the fact base, and (c) an inference engine, which links the knowledge base to the fact base to infer new information. To be considered in operating mode the EO-IU subsystem requires EO data-derived information layers, to be generated automatically (without user interaction) and in near real-time. Next, EO images are associated/linked with their information products (Fig. 5-2), either nominal/categorical/qualitative, such as SCMs, or numeric/quantitative, such as spectral indexes, stored in a data cube model known as a space-time data cube within the array database (see also Fig. 5-6). Finally, EO images and associated information products are input to the EO-SQ subsystem for spatiotemporal semantic querying.

### 5.2.1 Generic rule base for low-level information layer generation

The generic rule base adopted by the EO-IU subsystem comprises a battery of automated EO data processing algorithms, requiring no user interaction to run in pipeline. They cope with EO data calibration, stratified (driven-by-knowledge) atmospheric and topographic effect removal [Baraldi et al., 2010], categorization of colour values into colour names for texel (superpixel) detection [Baraldi et al., 2006], wavelet-based image-contour detection, image segmentation (raw primal sketch) [Marr 1980], texture segmentation (full primal sketch) [Marr, 1982], local shape analysis [Soares et al., 2014], stratified cloud/cloud-shadow detection [Baraldi and Tiede, 2015]. Some of these baseline algorithms, which were used in the prototypical implementation, are commented below.

#### 5.2.1.2 EO image calibration, physical model-based colour naming and texel detection

Sensory data calibration (*Cal*) is a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” [Schaeppman-Strub et al., 2006]. In practice, EO data *Cal* is mandatory to employ physical model-based and hybrid (combined deductive and inductive) EO-IUSs. To comply with the EO data *Cal* requirement, the proposed hybrid EO-IU subsystem employs a battery of sensor-specific EO image radiometric calibrators, e.g., Landsat-5 to Landsat-8, MeteoSat, RapidEye, WorldView, etc., to transform digital numbers into physical units of measure, e.g., top-of-atmosphere reflectance (TOARF) values.

Presented in the RS literature in recent years [Baraldi et al., 2006; Baraldi et al., 2010; Baraldi and Boschetti, 2012], the Satellite Image Automatic Mapper (SIAM) software product is a physical model-based decision tree (expert system) for deductive/top-down vector quantization (VQ) and VQ quality assessment in a multispectral (MS) reflectance space. VQ is synonym of vector space polyhedralization [Cherkassky and Mulier, 1998]. Since it is based on *a priori* knowledge available in addition to data, the SIAM expert system is fully automated, i.e., it requires neither user-defined parameters nor training data to run. It maps each MS pixel onto one MS polyhedron, either convex or not, either connected or not, associated with a MS colour name in a predefined legend of MS colour labels. Per-pixel colour labels form a 2D multi-level VQ map in the (2D) image domain (Fig. 5-3). This multi-level image is input to a well-posed (deterministic) two-pass connected component multi-level image labelling algorithm to extract a multi-level image segmentation map [Dillencourt et al., 1992; Sonka et al., 1994]. These image-segments are connected sets of pixels featuring the same colour label. Traditionally known as texels (texture elements), whose detection occurs at the raw primal sketch of low-level vision [Marr, 1982], texels have been recently renamed superpixels by the computer vision community [Achanta et al., 2011].

#### 5.2.1.2 Planar shape description

Given an image segmentation map consisting of planar objects, each planar object can be described in geometric (shape and size) terms by an mDMI set of planar shape indexes, such as area, characteristic scale, scale-invariant roundness, elongatedness, straightness of boundaries, simple connectivity, rectangularity and convexity [Soares et al., 2014]. Noteworthy, excluding size-related parameters area and characteristic scale, all shape (geometric) descriptors belong to range [0, 1]; hence, they are intuitive to partition into fuzzy sets, e.g., “high”, “medium” or “low”, employed by fuzzy decision rules typical of human reasoning. Shape descriptor values computed for each image-object (segment) detected in an EO image are encoded as raster information layers in the array database to allow shape-related spatiotemporal queries, e.g., change in size of a water body through time, where co-existing spatial units of information are pixel and image-object. This allows to overcome the traditional dualism between pixel-based image analysis and the OBIA paradigm [Blaschke et al., 2014].



### 5.2.2 Spatiotemporal conceptual modelling of real-world objects in the scene domain

In the 4D real world, observations (true-facts) can be represented by  $n$ -tuples (space  $x$ ,  $y$  and  $z$ , time  $t$ ; ‘theme’; plus other numeric or categorical attributes, e.g., weight, size, etc.), where the 4-tuple  $(x, y, z, t)$  is the location in space and time of the observation, and the attribute ‘theme’ identifies the real-world phenomenon or object being observed [Reis Ferreira et al., 2014]. ‘Theme’ may account for semantics involved with the observed object or phenomenon. All possible combinations of attributes (space, time, theme), can be modelled according to the conceptualization of Reis Ferreira et al. [2014] as three data types, called *time series*, *coverage*, and *trajectory*, where one attribute is measured, the second is fixed and the third is controlled. A *time series* represents the measured variations of a theme over a controlled time in a fixed location. A *trajectory* measures locations of a fixed theme over a controlled time. A *coverage* measures attribute theme within a controlled spatial extent at a fixed time.

A world model is an ontology of real-world geospatiotemporal *objects/continuants* and *events/occurrence* derived from the three spatiotemporal data types time series, coverage and trajectory. In EO applications, a real-world object can be: (a) a periodic object whose identity is fixed while its attributes change with a cyclic behaviour, i.e., the object’s identity comprises a given sequence of different states in a fixed time periodicity, e.g., a corn cropland whose growth cycle is decided by agricultural practices specific to a geographic region (Fig. 5-4), or (b) persistent/non-periodic objects, e.g. forested areas or lakes (Fig. 5-5). An event is an individual episode with a definite beginning and an end [Belgiu et al., 2016]. It exists as a whole across the interval over which it occurs, either instant or durative. An event does not change over time. While an event can involve one or more objects, the same object can be involved in any number of events. Events can be: (a) instant, or (b) durative, including short-term transition events and slowly transient events. For example, in EO applications vegetated areas can change into different LC types, such as bare soil, building or water, according to slowly transient events, such as urban sprawl due to policies enforced by responsible authorities, or due to some short-term transition events, e.g., natural or artificial fire and flooding events. The world model can be graphically represented as a semantic network with LC classes as nodes and spatiotemporal relationships, including events, as links between nodes [Grove, 1999].

A graphical user interface (GUI) allows users to easily select existing semantic queries/decision rules or to intuitively generate new ones. Each query instantiation is associated with an information pair, specifically, one spatiotemporal scene-domain knowledge (e.g., target LC classes) and one set of sensor-specific transfer functions required to map scene-domain knowledge in user-speak into image-domain knowledge in techno-speak. A query pipeline is a combination of spatial and temporal operators and/or standard algorithms whose inputs are qualitative/categorical information layers (e.g., SCMs) or quantitative/numeric variables (e.g., spectral indexes) available in the fact base. There are two types of semantic queries: (a) to accomplish SCBIR operations, where the fact base is investigated for EO image retrieval purposes. For example, retrieve EO images that are cloud-free across the selected geographic area of interest (AOI); (b) to infer new information layers from the fact base. For example, detect flooded areas as a post-classification combination through time of available single-date SCMs.

### 5.2.3 Array fact base

To accomplish efficient geospatial data querying and analysis through space and time within a user-defined AOI and a target time interval, a fact base stores multi-sensor multi-temporal EO images, e.g., Landsat-5/7/8 and Sentinel-2A images, together with their information products, either numeric, such as spectral indexes, or categorical, such as SCMs. This is realised within an array database (Rasdaman), where multiple spatiotemporal data cubes are instantiated in compliance with the Open Geospatial Consortium (OGC) coordinate reference systems (CRS) to ensure inter-system harmonization and compatibility (Fig. 5-6). In a spatiotemporal data cube the third dimension is time, defined in the eXtensible Markup Language (XML) as a 1D temporal coordinate system using Unified Resource Identifiers (URIs). Time overlays the 2D spatial coordinate system, specified by European Petroleum Survey Group (EPSG) codes defined in XML using URIs. Similar to standard relational databases with its SQL, an array database consisting of data cubes can be queried by a declarative query language. In the selected array database approach, storage-related characteristics, such as indexing, tiling and horizontal scaling, can be investigated and optimized independently of data models. In addition, the data cube model has been proven to be scalable and reliable in operational applications [Evans et al., 2015; Baumann et al., 2015]. These considerations make it best suited for the proposed EO-SQ system as storage backend. Semantic queries stored by the web-based query interface are accessible as OGC compliant web processing service (WPS) within the client-server architecture (Fig. 5-7) and can be executed by any WPS client remotely on a single server or server cluster. By providing EO data processing capabilities together with “ready-to-analyse” data, this client-server architecture guarantees a fast-time response to queries.



### 5.3 EO-SQ subsystem prototypical implementation

The proposed prototypical implementation of the EO-SQ subsystem, hereafter identified as IQ, is built upon the Rasdaman array database implementation. It stores every multi-source radiometrically calibrated EO image together with its information layers as one dense temporal stack, known as space-time data cube, whose third dimension is time. In greater detail, one EO image, its categorical and continuous information variables are stored in three different data cubes whose atomic element is a voxel. To improve IO performance, the data cube is automatically divided into smaller partitions (tiles) with the same dimensionality. Each partition is then stored as separate file on the hard disk. Additionally, large data cubes can be distributed on multiple servers. This distributed system can be scaled horizontally using the capability of the Rasdaman database to deploy multiple worker units coordinated by a management unit. Associated metadata, e.g. for the spatial reference, are stored in an object-relational PostgreSQL database.

The main application tier is accessible over the internet using an Apache httpd webserver. It was written in python and developed in-house (Fig. 5-8). Its purpose is twofold. First, it translates queries formulated according to human reasoning and encoded as an XML by the IQ frontend into a valid database query to be executed against the database. Additionally, if an allowed user decides to store the query, the query is automatically provided other users in compliance with the OGC WPS. In this case the query can be executed with different AOIs and time spans by alternate clients, e.g., ArcGIS, QGIS, etc., as long as they provide an OGC compliant WPS client. This system is architecture- and software-agnostic and it can be fully integrated into an existing workflow. Second, to provide API endpoints for system- and user-management. The user management encompasses the user login and restrictions to accessible datasets as well as permissions to create and store queries. Registered users are the first of three user groups. Every query in the knowledge base is accessible by any registered user. However, non-registered users are limited by a maximum AOI. Restriction to access is applied to datasets only and not to queries in order to foster the idea of the community- and knowledge-sharing-based approach. The second user group are expert users; they are a subset of the registered users having the permission to create and share queries. The third user group consists of administrators with permissions to manage the system as well as users. Administrators are allowed to create a new dataset using an importer wizard within the IQ's GUI. It allows to select the data source (e.g. Sentinel 2, Landsat), an AOI, a time span and available information layers. Next, system parameters are automatically fed into the IQ image loader which creates a single or routinely recurring instantiation of a process to download images from the official archive, transform them into products and store products into the fact base. Additionally, each dataset can have a moderator, who is allowed to select single users or groups who have access to that dataset.

EO big data require scalable and efficient processes to store, process and visualize EO images and products, but also affect quality assurance (QA). Besides automatic QA of sensory data, QA has to guarantee consistency of a user's semantic query/decision rule, whose input variables are images and products available in the fact base and whose functions are operators and processes available in the rule base. To enforce QA of decision rules, the EO-SQ subsystem delivers quality-ensuring metadata for each query, in agreement with the available 4D world model. Query-specific metadata are the possible target AOI, time interval and spatial, temporal, radiometric and spectral resolution of the imaging sensor, the name of the user who created the query and the time when the query was created. Query performance metrics are collected during each execution and updated regularly.

### 5.4 Use case examples

The IQ's GUI allows a user (1) to select the AOI and (2) the target time period and (3) to create or select via graphical elements the decision-rule pipeline for semantic querying the fact base. (4) Once a decision rule is executed, the query results are shown in the image domain and/or as summary statistics. Three examples of semantic spatiotemporal queries instantiated by users through the IQ's GUI are discussed hereafter.

#### 5.4.1 User case I – LC change detection through time

A snow cover analysis through time is conducted across an alpine area, specifically, the Hohe Tauern National Park in Austria (Fig. 5-9, point 1). The spatiotemporal semantic query (Fig. 5-9, points 2 and 3) shows the selected time span as well as the user-defined LC change (LCC) classes to be extracted from the fact base. Such a query can be easily saved, modified and shared with other users. The GUI shows the query output product, which can be downloaded as a geoTiff or re-used for further queries. The output map areas depicted in blue/white identify loss/persistent snow cover detected from January to April 2015 (Fig. 5-9, point 4).

#### 5.4.2 User case II – Planar shape descriptors to infer high-level LC classes from EO Level 2 products

In the hierarchical two-phase Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) [Di Gregorio and Jansen, 2000], an application-independent general-purpose 8-class LCCS-Dichotomous





Phase (DP) taxonomy is preliminary to the LCCS Modular Hierarchical Phase (MHP), consisting of a hierarchical battery of application- and user-specific one-class LC classifiers. In agreement with a hierarchical LCCS taxonomy, high-level LC classes Lake and River were extracted from an EO Level 2 SCM, by incorporating shape descriptors in a user's query. More specifically, LC lake and river candidate areas were selected as water objects in the image domain whose area in pixel units was within a given physical model-based range of values and whose shape compactness in the user query (Fig. 5-10 point 1) is fuzzy "high" (lake, Fig. 5-10 point 3) or "low" (river, Fig. 5-10 point 2), elongatedness is "low" and "high" respectively, etc. The output lake/river-from-water LC binary mask is a geoTiff file, to be stored or reused in further queries (Fig. 5-10).

### 5.4.3 User case III – Cloud-free SCBIR

Within a user-defined AOI and time interval, retrieve all multi-source EO images available in the fact base where no cloud is found across the AOI, based on the available data-derived EO Level 2 SCMs. In the implemented prototype, image retrieval queries turn back the matching image IDs and associated download links. Other examples for such an AOI based SCBIR are e.g. searching for images showing flooded areas/burnt areas/deforested areas or similar in the selected AOI.

## 5.5 Conclusions

This work started from the conjecture that existing EO CBIR systems support no semantic querying because they lack EO image understanding capabilities, where spatial information, either topological or non-topological dominates colour information [Matsuyama and Hwang, 1990], in agreement with the OBIA paradigm [Blaschke et al., 2014]. To our best knowledge, no EO SCBIR system in operating mode has ever been developed by the RS community. To be considered in operating mode any information processing system must score "high" in an mDMI set of output and process QIs [Baraldi and Boschetti, 2012].

A prototypical implementation of an EO-SQ subsystem (identified as IQ) plugged-in an innovative EO-IU&SQ system was proposed as a proof-of-concept. Capable of providing every EO image stored in the database with EO Level 2 products generated automatically (without user interaction) and in near real-time, the EO-IU subsystem is preliminary to IQ. Within the proposed EO-IU&SQ system architecture, the EO-IU and EO-SQ subsystems, their algorithms and their implementations can be modified or replaced to accomplish SCBIR capabilities in operating mode.

Further research and development will be focused on six areas. (I) Increase the 4D spatiotemporal world model. (II) Develop and implement an algebra to describe spatiotemporal data types and operations in a language-independent and formal way inspired to Reis Ferreira et al. [2014]. (III) Augment the GUI for semantic query selection and writing; (IV) Improve efficiency of on-the-fly EO image processing capabilities within array databases, e.g., to compute shape descriptors of image-objects selected by spatiotemporal queries. (V) Validation of outcome and process QIs in a multi-scale EO big data scenario. (VI) Improve outcome and process QIs in the knowledge base, with special emphasis on: (a) cloud/cloud-shadow quality layer detection [Baraldi and Tiede, 2015], (b) EO Level 2 product generation, where the general-purpose application- and user-independent Level 2 SCM's legend coincides with the 8-class LCCS-DP taxonomy [Di Gregorio and Jansen, 2000].

### Acknowledgements

The test implementation of the EO-SQ system, called ImageQuerying system, was awarded 1st place in the Copernicus Masters 2015 - T-Systems Big Data Challenge. The study was supported by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel-2/3, ID 848009.

### References in Chapter 5

- Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Susstrunk S. (2011) - *SLIC superpixels compared to state-of-the-art superpixel methods*, IEEE Trans. Pattern Anal. Machine Intell., 6(1): 1-8.
- Baraldi A., Boschetti L. (2012) - *Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation*. Remote Sens., 4: 2768-2817.
- Baraldi A., Girona M., Simonetti D. (2010) - *Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier*. IEEE Trans. Geosci. Remote Sensing, 48(1): 112-146.
- Baraldi A., Puzzolo P., Blonda P., Bruzzone L., Tarantino C. (2006) - *Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images*. IEEE Trans. Geosci. Remote Sens., 44: 2563-2586.



- Baraldi A., Tiede D. (2015). *AutoCloud+*, ESA Invitation to tender - Next Generation EO-based Information Services, ESA/AO/1-8373/15/I-NB.
- Baraldi A., Tiede D., Sudmanns M., Belgiu M., Lang, S. (2016) – *Automated near real-time Earth observation Level 2 product generation and semantic querying*. Int. Conf. GEOBIA 2016, in press.
- Baumann, P., Holsten, S. (2011) - *A comparative analysis of array models for databases*. In Database Theory and Application, Bio-Science and Bio-Technology. Springer, Berlin Heidelberg, pp. 80-89.
- Baumann P., Mazzetti P., Ungar J., Barbera R., Barboni D., Beccati A., Bigagli L. (2015) - *Big data analytics for earth sciences: the EarthServer approach*. International Journal of Digital Earth, 1–27.
- Belgiu M., Sudmanns M., Tiede D., Baraldi A., Lang, S. (2016) - Spatiotemporal enabled content-based image retrieval, Int. Conf. GIScience Montreal, Canada, (submitted).
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., and Tiede, D. (2014) - *Geographic object-based image analysis - towards a new paradigm,* ISPRS J. Photogram. Remote Sens., 87: 180–191.
- Brinkhoff T., Kresse, W. (2011). - *Databases. Springer Handbook of Geographic Information*. Springer, Berlin Heidelberg, pp. 11-34.
- Cherkassky V., Mulier F. (1998) - *Learning from Data: Concepts, Theory, and Methods*. Wiley, New York, NY.
- Di Gregorio A., Jansen L. (2000) - *Land Cover Classification System (LCCS): Classification Concepts and User Manual*, FAO: Rome, Italy, FAO Corporate Document Repository. Available online at: <http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm>
- Dillencourt, M. B., Samet H., Tamminen M. (1992) - *A general approach to connected component labeling for arbitrary image representations*, J. Assoc. Computing Machinery, 39: 253-280.
- Dumitru C. O., Cui S., Schwarz G., Datcu M. (2015) - *Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies*, IEEE J. Selected Topics Applied Earth Obs. Remote Sens., 8(4): 1635 – 1650.
- du Buf H., Rodrigues, J. (2007) - *Image morphology: from perception to rendering*. In IMAGE - Computational Visualistics and Picture Morphology.
- European Space Agency (2015) - *Sentinel-2 User Handbook*, Standard Document, Issue 1 Rev 2.
- Evans B., Wyborn L., Pugh T., Allen C., Antony J., Gohar K., Porter D. (2015) - *The NCI high performance computing and high performance data platform to support the analysis of petascale environmental data collections*. In Environmental software systems. Infrastructures, services and applications. Springer, pp. 569–77.
- Reis Ferreira K., Camara G., Monteiro A.M.V. (2014) - *An Algebra for Spatiotemporal Data: From Observations to Events*. Transactions in GIS, 18: 253-269.
- Grove S. (1999) - *Knowledge-based interpretation of multisensor and multitemporal remote sensing images*, Int. Archives of Photogrammetry and Remote Sensing, 32, Part 7–4–3 W6, Valladolid, Spain, pp. 3–4.
- Gudivada V. N., Raghavan, V.V. (1995) - *Content based image retrieval systems*. Computer, 28(9): 18-22.
- Li Y., Bretschneider T. (2004) - *Semantics-based satellite image retrieval using low-level features*. In Proceedings of the Geoscience and Remote Sensing Symposium, 2004 (IGARSS04), pp. 4406–4409.
- Matsuyama T., Hwang V.S. (1990) - *SIGMA – A Knowledge-based Aerial Image Understanding System*. Plenum Press, New York, NY.
- Marr D. (1982) - *Vision*. Freeman and C, New York, NY.
- Open Geospatial Consortium (OGC) Inc.,(2015) *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Available online at: <http://www.opengeospatial.org/standards/is>
- Planthaber G., Stonebraker M., Frew J. (2012) - *EarthDB: scalable analysis of MODIS data using SciDB*. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data. ACM, pp 11-19.
- Purss M.B.J., Levis A., Simon O., Ip A., Sixsmith J., Evans B., Edberg R., Frankish G., Hurst L., Chan T. (2015) - *Unlocking the Australian Landsat Archive – From dark data to High Performance Data infrastructures*. GeoResJ, 6: 135-140.
- Reis Ferreira K., Camara G., Vieira Monteiro A.M. (2014) *An Algebra for Spatiotemporal Data: From Observations to Events*, Trans. in GIS, 1(2): 253–269.
- Schaepman-Strub G., Schaepman M.E., Painter T.H., Dangel S., Martonchik J.V. (2006) - *Reflectance quantities in optical remote sensing—Definitions and case studies*. Remote Sens. Environ. 103: 27–42.
- Seidel K., Schroder M., Rehrauer H., Schwarz G., Datcu M. (1998) - *Query by image content from remote sensing archives*. Proc. Of the Geoscience and Remote Sensing Symposium 1998 (IGARSS98), pp. 393-396.

- Shyu C.-R., Klaric M., Scott G.J., Barb, A.S., Davis C.H., Palaniappan, K. (2007) - *GeoIRIS: Geospatial Information Retrieval and Indexing System—Content mining, semantics modeling, and complex queries*, IEEE Trans. Geosci. Remote Sens., 45(4): 839–852.
- Smeynders A., Worring M., Santini S., Gupta A., Jain R. (2000) - *Content-based image retrieval at the end of the early years*, IEEE Trans. Pattern Anal. Machine Intell., 22(12): 1349-1380.
- Soares J.V.B., Baraldi A., Jacobs D.W. (2014) - *Segment-based simple-connectivity measure design and implementation*, Tech. Rep., University of Maryland, College Park, Available online at: <http://hdl.handle.net/1903/15430>.
- Sonka M., Hlavac V., Boyle R. (1994) *Image Processing, Analysis and Machine Vision*. Chapman and Hall, London, U.K..

**Figures in Chapter 5**

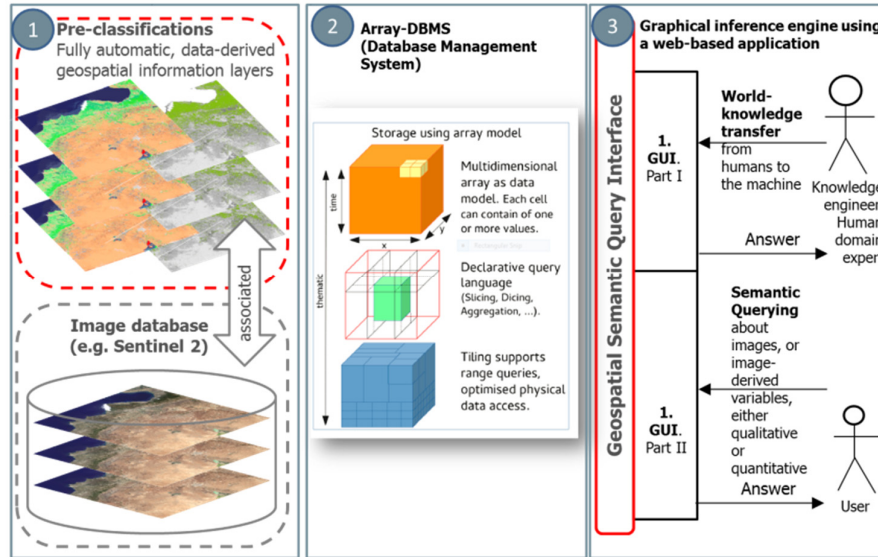


Fig. 5-1: EO-IU&SQ system architecture. The EO-SQ subsystem is identified as sections (2) and (3).

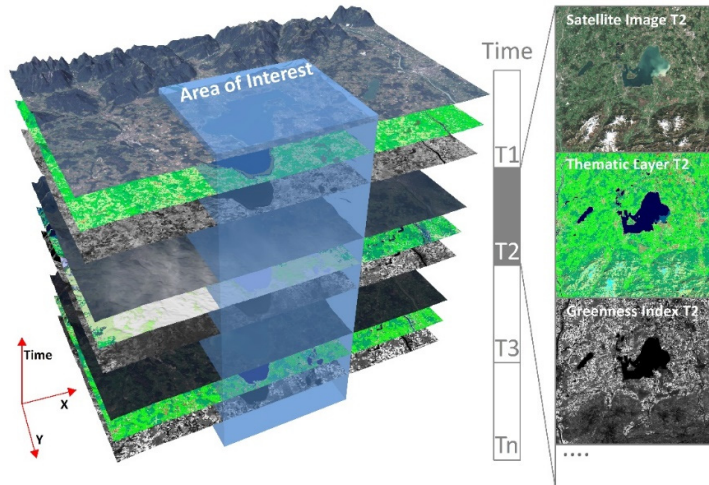


Fig. 5-2: EO Level 2 information layers, either numeric/quantitative or categorical/qualitative, are automatically generated by the EO-IU subsystem and linked with the EO data to be employed as input by the EO-SQ subsystem for spatiotemporal semantic querying.



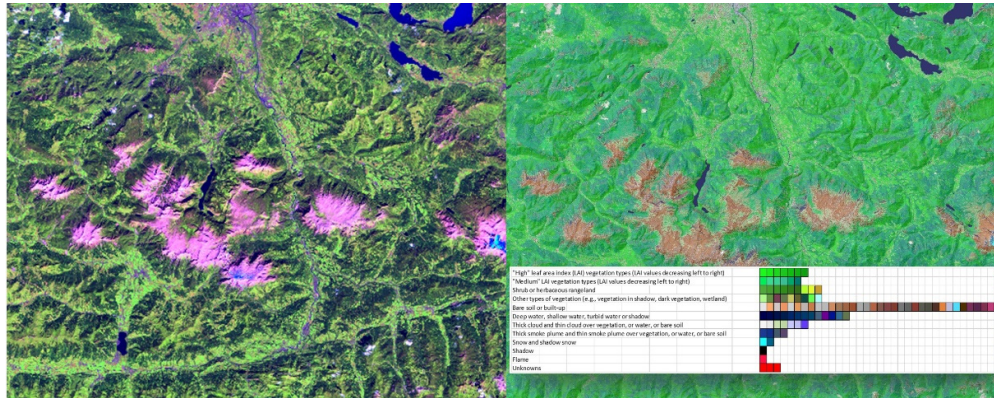


Fig. 5-3: Left: Sentinel-2A (S2A) image of Salzburg, Austria, acquired on 2015-08-13, depicted in false colours: R = short wave infrared (SWIR), G = Near IR (NIR), B = Visible blue. No histogram stretching for visualization purposes. Right: Automatic SIAM mapping of the S2A image onto a legend of 96 MS colour names (spectral categories), depicted as pseudo colours (green as vegetation, blue as water or shadow, etc.).

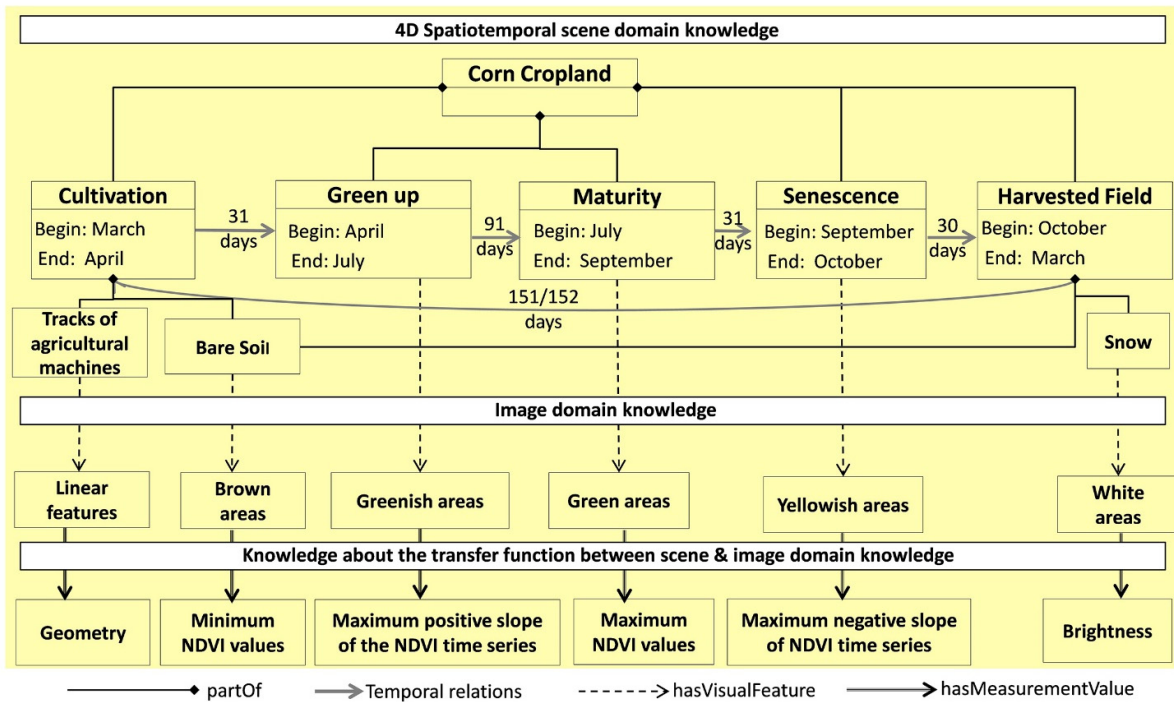


Fig. 5-4: Semantic network of a real-world object with cyclic behaviour, specifically, a corn cropland in northern hemisphere.



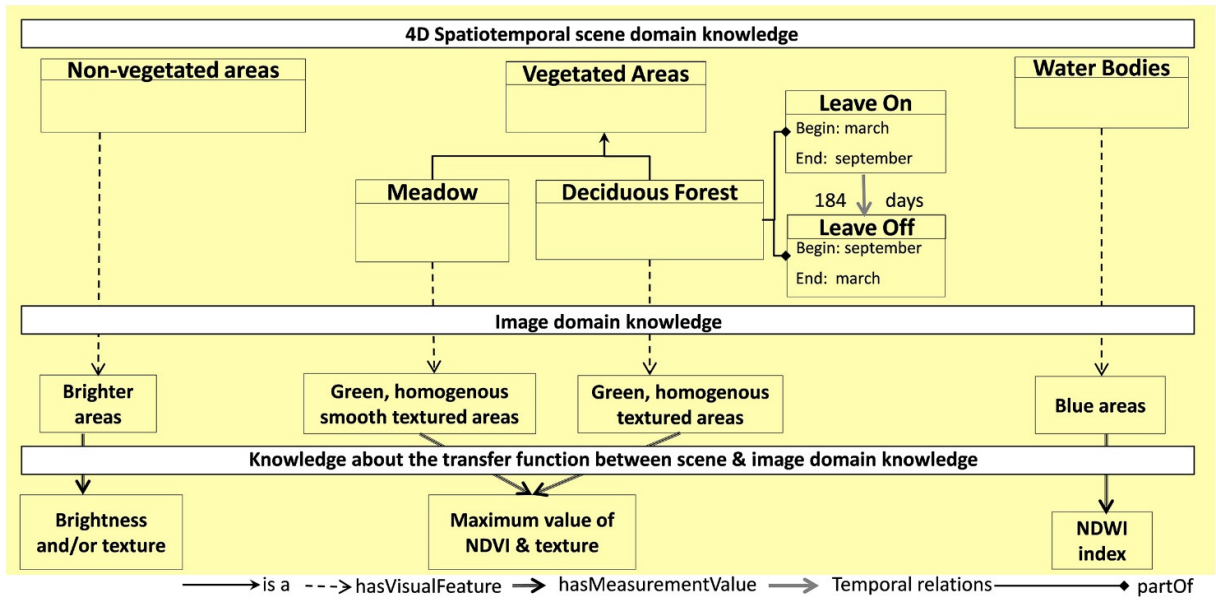


Fig. 5-5: Semantic networks of two real-world persistent objects, specifically, non-vegetated areas and water bodies, and one periodic object, such as deciduous forest.

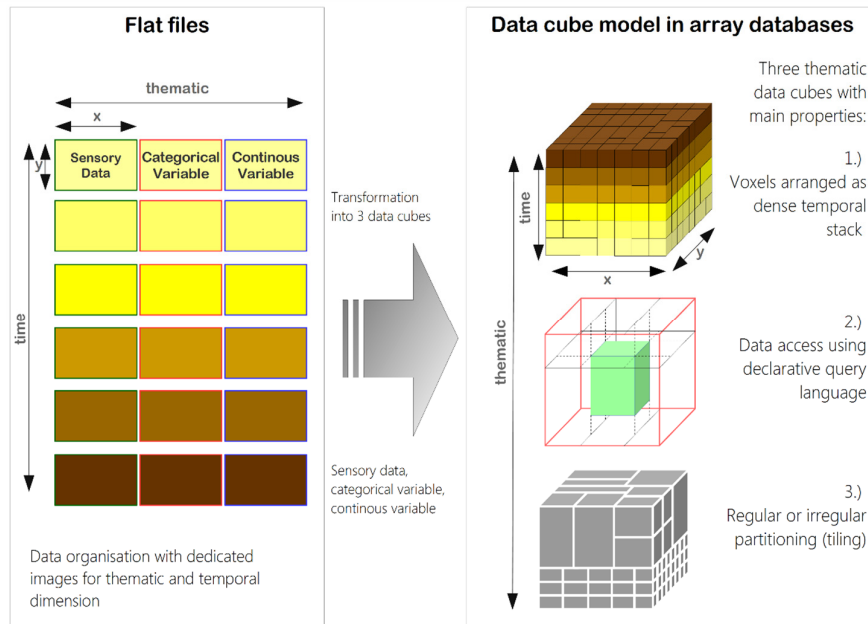


Fig. 5-6. Storage using flat files versus storage of images in an array database.

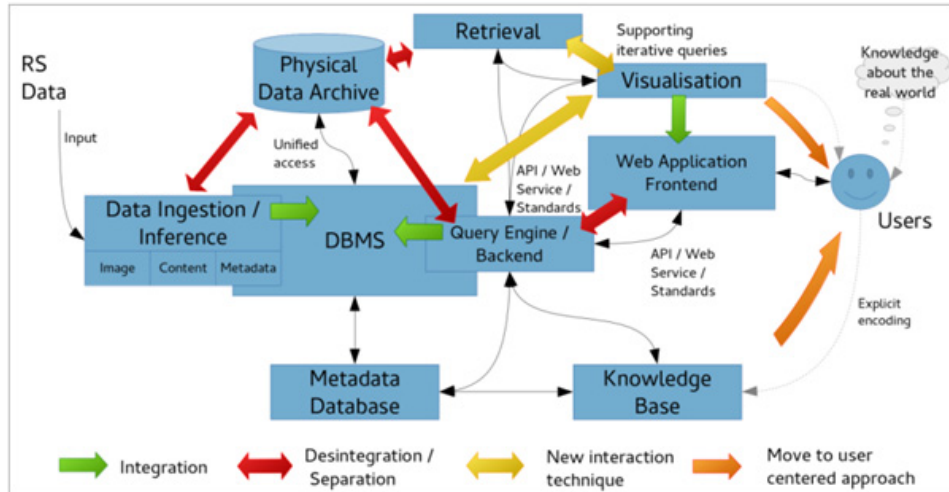


Fig. 5-7. Client-server array database architecture

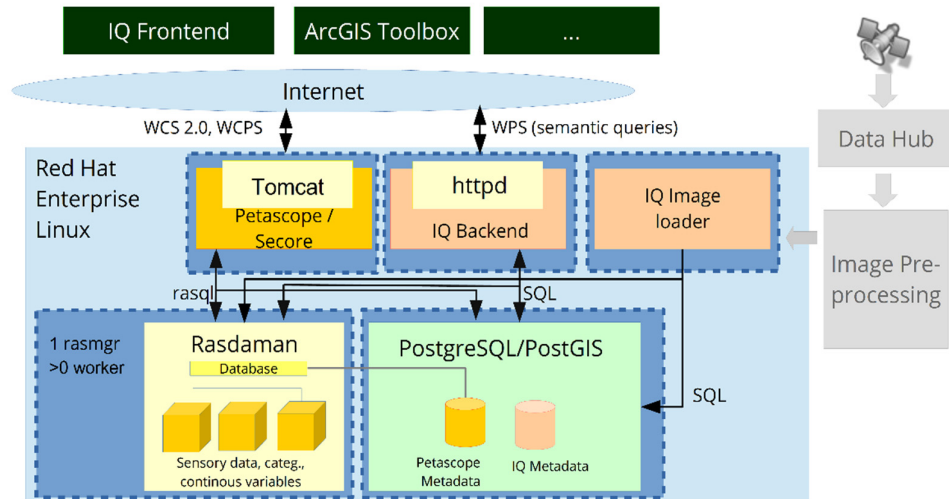


Fig. 5-8. Web-based Image Querying (IQ) prototype. The term IQ is used for the prototypical implementation of the EO-SQ subsystem.

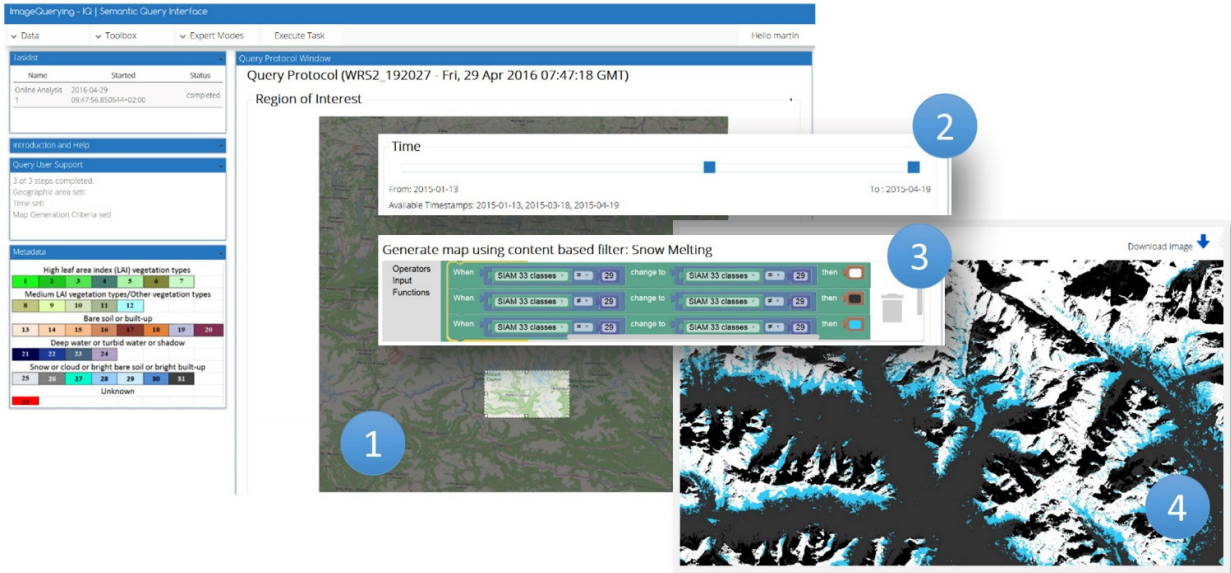


Fig. 5-9. Semantic querying to infer new information layers from the fact base, here: snow cover analysis. For a detailed description see main text.

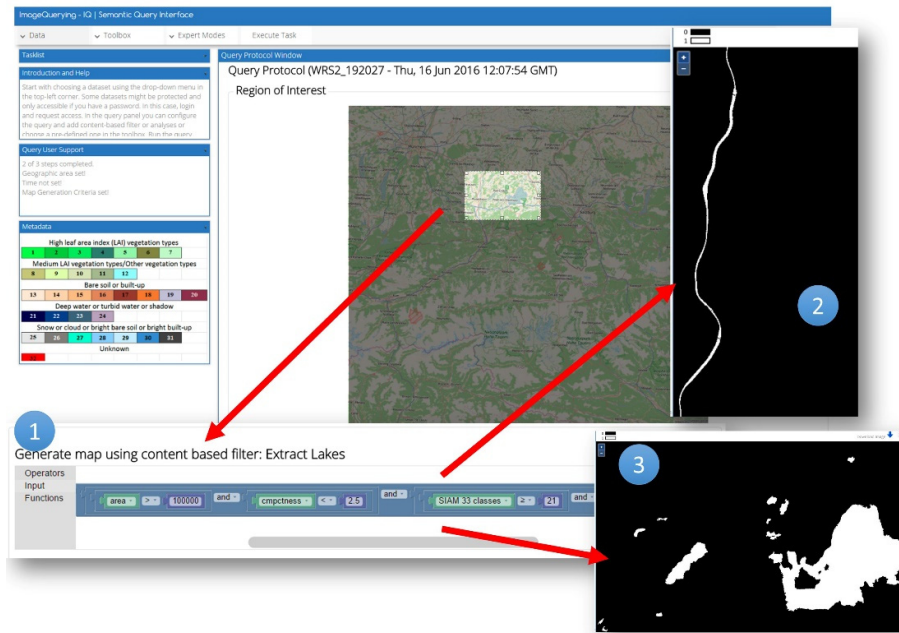


Fig. 5-10. Example of spatiotemporal semantic querying to infer new information layers from the fact base: splitting of the LCCS-DP class B48, Natural Waterbodies, into the LCCS-MHP classes Lake and River based on planar shape and size properties.

## 6 Manuscript 3 (published, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5560-5575, Dec. 2016): Automated Hierarchical 2D and 3D Object-Based Recognition and Reconstruction of ISO Containers in a Harbor Scene

### Motivation and Contributions to the Dissertation

Original CV algorithms presented in Chapter 3 (Technical report 1, Computational models of human vision) and adopted by an innovative Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system, proposed as a proof-of-concept of a Global Earth Observation System of Systems (GEOSS) in support of semantic content-based image retrieval (SCBIR) in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), are applied to the experimental framework of the IEEE GRSS Data Fusion Contest 2015. The present Chapter 6 (Manuscript 3) was derived from a conference paper, titled “Geospatial 2D AND 3D object-based classification and 3D reconstruction of ISO-containers depicted in a LiDAR dataset and aerial imagery of a harbor”, presented in the IGARSS 2015 conference, Milan, Italy, 27-31 July 2015, which ranked 2nd in the IEEE GRSS Data Fusion Contest 2015.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 6 (Manuscript 3) are color filled.

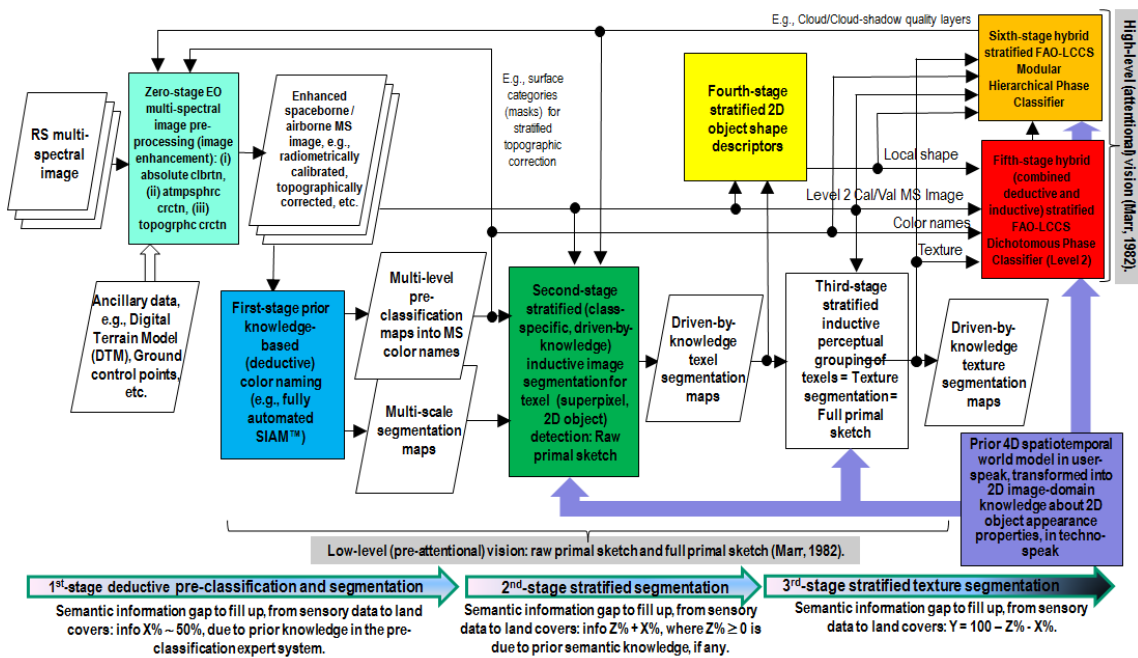


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype [42]. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 6 (Manuscript 3).





# Processing of Extremely high resolution LiDAR and optical data: Outcome of the 2015

## IEEE GRSS Data Fusion Contest.

### Part-B: 3D contest

A.-V. Vo, L. Truong-Hong, D.F. Laefer, D. Tiede, S. d'Oleire-Oltmanns, A. Baraldi, M. Shimoni, *Member, IEEE*, G. Moser, *Senior Member, IEEE*, D. Tuia, *Senior Member, IEEE*.

AVV, LTH and DL are with the University College Dublin, Ireland. E-mail: anh-vu.vo@ucdconnect.ie, linh.truonghong@ucd.ie, debra.laefer@ucd.ie

DT and SOO are with the Dept. of Geoinformatics - Z GIS, University of Salzburg, Austria. E-mail: dirk.tiede@sbg.ac.at, Sebastian.dOleire-Oltmanns@sbg.ac.at

AB is with the Dept. of Agricultural and Food Sciences, University of Naples Federico II, Italy, and with the Dept. of Geoinformatics - Z GIS, University of Salzburg, Austria. E-mail: andrea6311@gmail.com

MS is with the Signal and Image Centre, Dept. of Electrical Engineering, Royal Military Academy (SIC-RMA), Brussels, Belgium, E-mail: mshimoni@elec.rma.ac.be.

GM is with the University of Genoa, Dept. of Electrical, Electronic, Telecommunications Eng. and Naval Architecture (DITEN), Italy, E-mail: gabriele.moser@unige.it.

DTuia is with the Department of Geography, University of Zurich, Switzerland. E-mail: devis.tuia@geo.uzh.ch.

#### Abstract

In this paper, we report the outcomes of the 2015 data fusion contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (IEEE GRSS). As for previous years, the IADF TC organized a data fusion contest aiming at fostering new ideas and solution for multisource studies. The 2015 edition of the contest proposed a multiresolution and multisensorial challenge involving extremely high resolution color images and a 3D LiDAR point cloud. The competition was framed in two parallel tracks, considering 2D and 3D products, respectively. In this Part-B, we report the results obtained by the winners of the 3D contest, which explored the synergistic use of the LiDAR point cloud and color data for 3D analysis at extremely high spatial resolution. The 2D part of the contest, as well as the dataset, are discussed in [1].

#### Index Terms

Image analysis and data fusion, IADF, Multiresolution-, Multisource-, Multimodal-data fusion, land cover classification, LiDAR, color VHR data.

#### 6.1 Introduction to Part-B

Three-dimensional (3D) high-spatial-resolution data becomes a fundamental part in a growing number of applications ranging from urban planning, cartographic mapping, environmental impact assessment, cultural heritage protection, transportation management and civilian and military emergency responses [2]. Among the data sources available, a light detection and ranging (LiDAR) sensor offers a fast and effective way to acquire 3D data [3]. It consists of a laser scanner which transmits and detects signals to measure range, a GPS receiver, which provides a sensor position, and an inertial navigation system which provides orientation information. The acquired data represent intricate height surfaces including artificial objects, such as buildings, and natural objects.

In this framework, the present paper is the second of a two-part manuscript presenting and critically discussing the scientific outcomes of the 2015 edition of the Data Fusion Contest organized by the IADF TC of the IEEE-GRSS. The 2015 Contest released to the international community of remote sensing a topical and complex image data set involving 3D information, multiresolution / multisensor imagery, and extremely high spatial resolutions. The data set was composed of a color orthophoto and of a LiDAR point cloud acquired over a urban and harbor area in Zeebrugge, Belgium (see Section II of [1]).



Given the relevance of this data set for the modelling and extraction of both 2D and 3D thematic results, the Contest was framed as two independent and parallel competitions. The 2D Contest was focused on multisource fusion for the generation of 2D processing results at extremely high spatial resolution: the interested reader can find the presentation and discussion of the results in [1]. The 3D Contest explored the synergistic use of 3D point cloud and 2D color data for 3D analysis at extremely high spatial resolution. Its results are discussed in detail in this paper. In either case, participating teams submitted original open-topic manuscripts proposing scientifically relevant contributions to the fields of 2D/3D extremely high resolution image analysis. Even though LiDAR [4] and VHR color [5] data were considered in the past contests, for the first time a 3D competition considering their joint use is proposed to the community.

Indeed, advanced airborne LiDAR systems provide high-density points in a square meter (more than 100) [6] and spatial resolution that is finer than 10 cm. However, due to the characteristics of point cloud data (discrete distribution and loss of spectral information), the laser measurements do not always allow a precise reconstruction of the 3-D shape of the target (e.g., building corner) [7], [8]. To overcome these shortcomings, integrated approaches fusing LiDAR data and aerial images are increasingly used for the extraction and the quantification of trees [7], [9], [10], buildings [8], [11]–[13], roads [14]–[18], vehicles [19], [20], and other small objects in the urban scenes [20], [21]. These multisource fusion methods can reduce the difficulties of processing, whereas they have strict requirements for data acquisition and registration [8]. For the latter, several approaches provide frameworks for automated registration of 2D images onto 3D range scans. Most of the available methods are based on extracting and matching features (e.g., points, lines, edges, rectangles or rectangular parallelepipeds) [22]–[24]. Others describe solutions that combine 2D-to-3D registration with multi-view geometry algorithms obtained from the parameters of the camera [25], [26].

Another major topic that appeared in the LiDAR-color fusion literature is roads and object detection (i.e., buildings, containers, marine vessels and vehicles). The integration of 2D optical and 3D LiDAR data sets provides photorealistic textured models that facilitate the detection and the extraction of large-scale objects from the scene. The literature mainly covers the topic of feature based fusion for building extraction [11], building surface description [18], [27], detection of roof planes and boundaries [12], structure monitoring [13], and urban building modelling [28], but it also addresses the extraction and the identification of small to medium-scale objects [19]– [21], [29]. Automatic extraction of roads in complex urban scenes from remotely sensed data has also been an open problem that is unsolvable using a single remote sensing source [14]. In passive imagery, the occlusion of the road surface by vertical objects creates artefacts such as shadows, radiometric inhomogeneity, and mix spectra that complicate the road detection. The property of airborne LiDAR imagery makes it a better data source for road extraction in urban scenes. Free of shadow effects, relatively narrow scanning angle (typically 20°-40° [15]), laser reflectance and elevation information allow good separation of road from other urban objects [16]. However, LiDAR data lack spectral information, which creates difficulties for reliable object recognition. Moreover, due to the irregular distribution of LiDAR points, more effort is needed to extract accurate break lines or features, such as the edges of roads [14]. Given the pros and the cons of LiDAR and aerial imagery, it has been suggested that these data be fused to improve the degree of automation and the robustness of automatic road detection [14]–[17].

In this paper, we present the works of the winning teams of the 3D contest and provide a general discussion of the results: first for the 3D contest and then overall for the 2015 IEEE GRSS Data Fusion Contest. We invite the readers interested in the 2D contest, as well as in the detailed presentation of the datasets to refer to the sister publication, the Part-A manuscript [1]. For the track 3D, the papers awarded were:

- 1<sup>st</sup> place. Aerial laser scanning and imagery data fusion for road detection in city scale by *Anh-Vu Vo, Linh Truong-Hong, and Debra F. Laefer* from University College Dublin (Ireland).
- 2<sup>nd</sup> place. Automated hierarchical 2D and 3D object-based recognition and reconstruction of ISO-containers in a harbor scene by *Dirk Tiede, Sebastian d'Oleire-Oltmanns and Andrea Baraldi* from the University of Salzburg (Austria) and the University of Naples Federico II (Italy).

In the present Part-B paper, the approaches proposed by the winning teams of the 3D contest are presented in Sections II (omitted) and III (now Chapter 6.2), respectively; then, a critical discussion of these approaches and of the overall set of the manuscripts submitted to the 3D contest is presented in Section IV. A general discussion on the Data Fusion Contest 2015 concludes the paper in Section V.

## 6.2 Automated hierarchical 2D and 3D object-based recognition and reconstruction of ISO containers in a harbor scene

### 6.2.1 Introduction

The inventory of rapidly changing logistics infrastructures, such as depots and harbors, is crucial to their efficiency. For example, an optimized exploitation of an intermodal storage volume for shipping containers requires an inventory where positional, geometric and identification attributes of individual containers are known real-time. In the 2015 GRSS 3D Data Fusion Contest this work tackled the problem of freight container localization and classification at a harbor extent by an automated near real-time computer vision (CV) system, in agreement with the International Standards Organization (ISO) 668 - Series 1 freight containers documentation adopted as a source of *a priori* (3D) scene-domain knowledge [42].

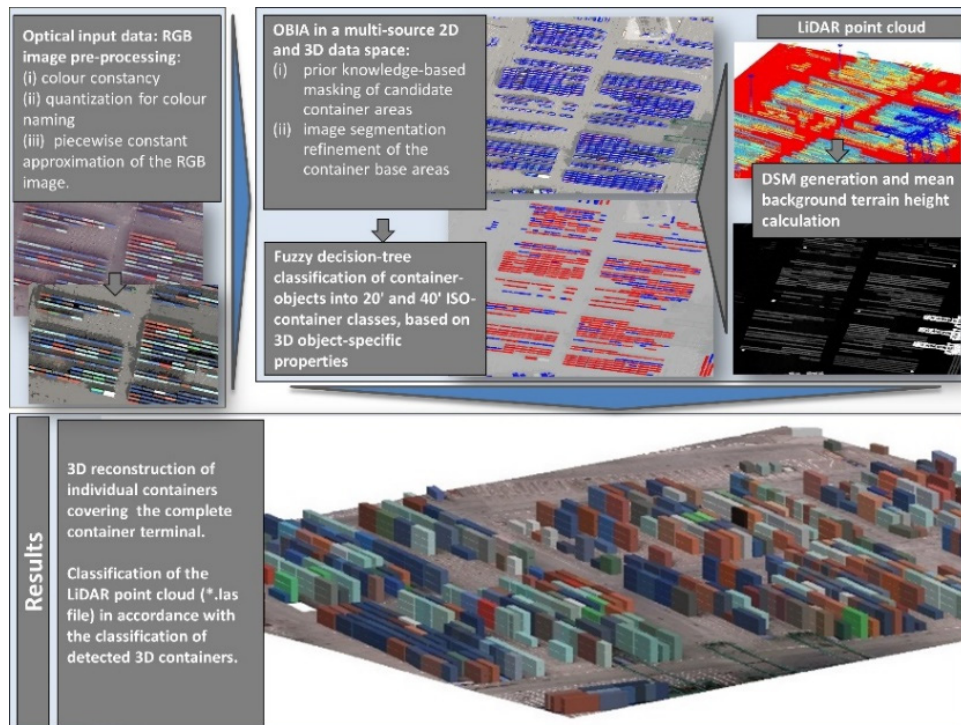


Fig. 6-8. Adopted workflow. (1) RGB image pre-processing (upper left) and LiDAR data pre-processing (top right). (2) Integration of two object-based image and point cloud analyses (center top). (3) Reconstruction (synthesis) of tangible 3D container-objects (bottom).

### 6.2.2 Methods

1) *Selected 2D and 3D sensory data sets.* Focusing on the harbor area visible in Tiles number 1-2-3-4, refer to Fig. 1 in [1], the input data sets selected for use were the uncalibrated three-band true-color RGB aerial (2D) orthophoto featuring very high spatial resolution (VHR), below 10 cm, and the dense 3D LiDAR point cloud described in Section II of [1]. The available DSM was not used because it appeared to lack non-stationary surface-elevated objects, such as cranes and freight containers, in disagreement with the LiDAR point cloud. In addition, a slight tilting effect was observed to affect the RGB orthophoto in comparison with the LiDAR point cloud, across image locations where above-ground objects, such as container stacks, were depicted. Such a tilting effect could be caused by an image orthorectification process employing as input the aforementioned DSM, where non-stationary above-ground objects were absent. In practice, the target CV system was required to cope with the observed tilting effect when 2D and 3D data sets were spatially overlapped.

2) *In-house DSM generation from the LiDAR point cloud.* To reveal above-ground scene elements, such as containers and cranes, the 3D LiDAR point cloud was integrated as a raster (2D gridded) point cloud, where the DSM pixel size was the same of the input orthophoto and the DSM pixel value was the LiDAR highest elevation value  $z$  occurring per pixel.

3) *Main workflow.* Human panchromatic vision is nearly as effective as color vision in the provision of a complete scene-from-image representation, from local syntax of individual objects to global gist and layout of objects in space, including semantic interpretations and even emotions [44], [45]. This fact means that spatial information dominates color information in the spatiotemporal 4D real world-through-time domain, described by humans in user-speak [42], as well as in a (2D) VHR image domain, to be described in techno-speak [44], irrespective of data dimensionality reduction from 4D to 2D [45], [46]. It agrees with the increasing popularity of the object-based image analysis (OBIA) paradigm [47], [48], [49], proposed as a viable alternative to traditional 1D image analysis, where inter-vector topological (neighborhood) relationships are lost when a 2D gridded vector set is mapped onto a 1D vector stream [50]. To develop a CV system capable of 2D spatial reasoning in a VHR image-domain for 3D scene reconstruction, a *hybrid* inference system architecture was selected. According to Marr, the linchpin of success of any data interpretation system is architecture and knowledge/information representation, rather than algorithms and implementation [43]. In hybrid inference, deductive and inductive inference are combined to take advantage of each and overcome their shortcomings [45], [48], [49], [52], [53], [52], [55]. On the one hand, inductive (bottom-up, learning-from-data, statistical model-based) algorithms, capable of learning from either unsupervised or supervised data, are inherently ill-posed and require *a priori* knowledge in addition to data to become better posed for numerical solution ([51], p. 39). In the remote sensing (RS) common practice they are semi-automatic and site-specific [52], [53], [54]. On the other hand, expert systems (deductive, top-down, physical model-based, prior knowledge-based inference systems) are automated, since they rely on *a priori* knowledge available in addition to data, but lack flexibility and scalability [51], [52], [53], [54]. An original *hybrid* CV system architecture was selected to: (I) comply with the OBIA paradigm [47], [48], [49], alternative to 1D image analysis [50], (II) start from an automated deductive inference first stage, to provide second-stage inductive learning-from-data algorithms with initial conditions without user interaction, and (III) employ feedback loops, to enforce a “stratified” (driven-by-knowledge, class-conditional) approach to unconditional sensory data interpretation [45], [46], [54], equivalent to a focus of visual attention (FoA) mechanism [45], [55a] and to the popular divide-and-conquer problem solving approach [51]. Sketched in Fig. 6-8, the implemented CV system consisted of three main modules. (i) An application-independent automated RGB image pre-processing first stage for uncalibrated RGB image harmonization, enhancement and preliminary classification. (ii) High-level second-stage classification with spatial reasoning in a heterogeneous 2D and 3D data space. (iii) 3D reconstruction (synthesis) of individual ISO containers.

### 6.2.3 Automated RGB Image Pre-processing First Stage

Calibration/Validation (*Cal/Val*) of EO data and data-derived information products are considered mandatory by the Quality Assurance Framework for Earth Observation (QA4EO) guidelines [56]. The QA4EO’s *Val* principle requires each stage of an EO data processing pipeline to be provided with community-agreed quality indicators (Q<sup>2</sup>Is), to compare error propagation with quality standards [56]. The QA4EO’s *Cal* principle requires dimensionless digital numbers to be transformed into a physical variable, provided with a radiometric unit of measure, by means of radiometric *Cal* metadata parameters [56]. Physical variables can be analyzed by both physical and statistical models, therefore *hybrid* models too [52]. On the contrary, quantitative variables provided with no physical unit of measure can be investigated by statistical models exclusively [52]. Irrespective of this common knowledge, EO data *Cal* is largely overlooked in the RS common practice, dominated by statistical model-based data analysis [46], [56]. To comply with the QA4EO *Cal/Val* requirements when no radiometric calibration metadata are available, the proposed *hybrid* CV system included an original uncalibrated RGB image pre-processing first stage.

1) *RGB color constancy for uncalibrated RGB image harmonization.* Color constancy is a perceptual property of the human vision system ensuring that the perceived colors of objects in a (3D) scene remain relatively constant under varying illumination conditions [57]. By augmenting image harmonization and interoperability without relying on radiometric calibration parameters, it was considered a viable alternative to the mandatory QA4EO’s data *Cal* [56]. An original automated (self-organizing) statistical model-based algorithm for MS color constancy, including RGB color constancy as a special case, was designed and implemented in-house (unpublished, patent pending).



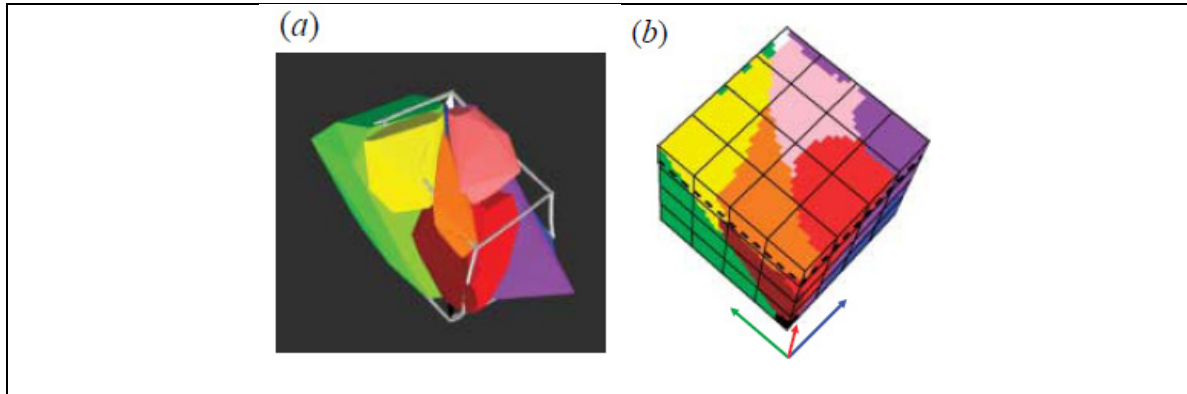


Fig. 6-9. Stages in mapping human basic colors (BCs) into the RGB cube. (a) CIE-Lab space with regions of eleven basic color labels as colored polyhedra and the edges of the monitor-typical RGB cube (in grey). (b) 323 quantization of the RGB space with the basic color extents from (a) mapped into it. The uniform 43 quantization of the RGB cube shown in (b) was adopted for representing color category systems whose classification performance was assessed. Images reproduced courtesy of [58].

2) *Forward RGB image analysis by prior knowledge-based vector quantization (VQ) and inverse RGB image synthesis for VQ quality assessment.* Widely investigated by the CV community [57], a finite and discrete dictionary of prior RGB color names is equivalent to a static (non-adaptive to data) RGB cube polyhedralization, where polyhedra can be any, either convex or not, either connected or not. In his seminal work [57], Griffin proved the hypothesis that the best partition of a monitor-typical RGB data cube into color categories for pragmatic purposes coincides with human basic colors (BCs), see Fig. 6-9. Central to this consideration is Berlin and Kay's landmark study of color words in 20 human languages, where they claimed that the "basic color terms of any given language are always drawn" from a universal inventory of eleven color names: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown [59]. These perceptual BC categories are expected to be "universal", i.e., users can apply the same universal color representation independently of the image-understanding problem at hand [57]. Equivalent to color naming in natural languages [59], prior knowledge-based color space discretization is the deductive automatic counterpart of inductive learning-from-data VQ algorithms (not to be confused with unsupervised data clustering algorithms [51]). In machine learning, the class of predictive VQ optimization problems requires to minimize a known VQ error function, typically a root mean square error (RMSE), where the number and location of VQ bins are the system's free-parameters [51]. For example, in the popular k-means VQ algorithm, the number of VQ levels,  $k$ , must be user-defined based on heuristics [60]. When they adopt a Euclidean metric distance in their minimization criterion and they reach convergence, inductive VQ algorithms accomplish a Voronoi tessellation of the input vector space, which is a special case of convex polyhedralization [78]. In contrast with inductive VQ algorithms capable of convex hyper-polyhedralizations, prior spectral knowledge-based decision trees can be designed to partition an input data hyper-space into hyper-polyhedra of any possible shape and size, either convex or concave, either connected or not. Unfortunately, when a data space dimensionality is superior to three, such as in Fig. 6-9, a prior partition of hyper-polyhedra is difficult to think of and impossible to visualize. This is the case of the Satellite Image Automatic Mapper™ (SIAM), an expert software system for MS color naming presented to the RS community in recent years [46]. By definition, expert systems require neither training data sets nor user-defined parameters to run, i.e., they are fully automated. Inspired by the SIAM expert system [46], a novel RGB Image Automatic Mapper™ (RGBIAM) was designed and implemented (unpublished, patent pending). RGBIAM is an expert software system for RGB cube partitioning into an *a priori* dictionary of RGB color names. By analogy with SIAM, since no total number  $k$  of VQ bins can be considered "best" (universal) in general, the implemented RGBIAM supports two co-existing VQ levels, fine and coarse, corresponding to 50 and 12 color names respectively, provided with inter-dictionary parent-child relationships, see Fig. 6-10. Whereas the physical model-based SIAM requires as input a radiometrically calibrated MS image [56], the first-principle model-based RGBIAM requires as input an uncalibrated RGB image, in either true- or false-colors, pre-processed by a color constancy algorithm, to guarantee data harmonization and interoperability across images and sensors.

Green-as-Vegetation	
Brown- or Gray-as-Bare soil or built-up	
Blue- or Light Blue-as-Deep water or shadow or shallow water or cloud aura	
Dark or shadow	
White or cloud	
Water ice, snow, cloud	
Unknowns	

(a) RGBIAM's fine color map legend, consisting of 50 color names and their pseudocolors.

Blue	
Green	
Red	
Pink	
Brown	
Cyan / Aqua	
Yellow	
Grey	
Purple / Magenta / Fuchsia	
Black	
White	
Unknown	

(b) RGBIAM's coarse color map legend, consisting of 12 color names and their pseudocolors.

Fig. 6-10. RGBIAM is an expert system for a monitor-typical RGB color-space discretization into two prior quantization levels, consisting of (a) 50 color bins and (b) 12 color bins, linked by inter-legend child-parent relationships. They are fixed *a priori*, to be community-agreed upon.

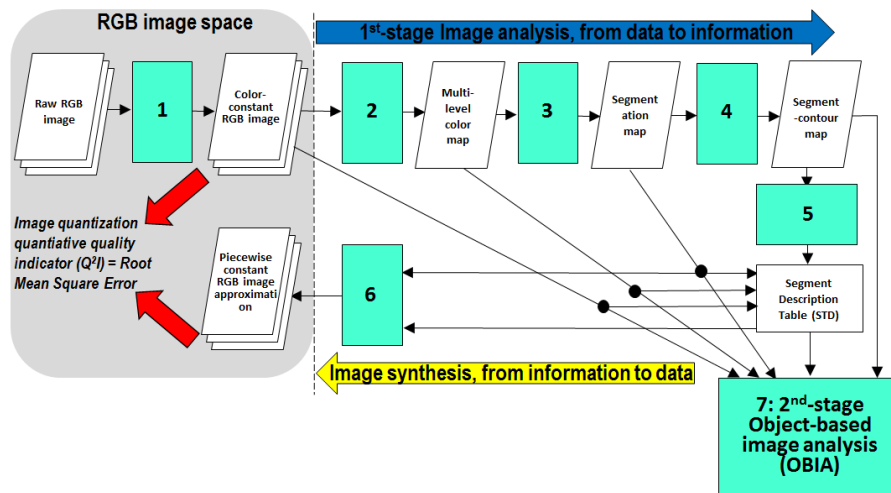


Fig. 6-11. First-stage RGBIAM's QNQ transform, consisting of six information processing blocks identified as 1 to 6, followed by a high-level object-based image analysis (OBIA) second stage, shown as block 7. Blocks 1 to 5 cope with direct image analysis. 1: Self-organizing statistical algorithm for color constancy. 2: Deductive Vector Quantization (VQ) stage for prior knowledge-based RGB cube polyhedralization. 3: Well-posed two-pass connected-component multi-level image labeling. 4: Well-posed extraction of image-object contours. 5: Well-posed Superpixel Description Table (STD) allocation and initialization. Block 6: inverse image synthesis, specifically, superpixelwise (piecewise)-constant RGB image approximation.

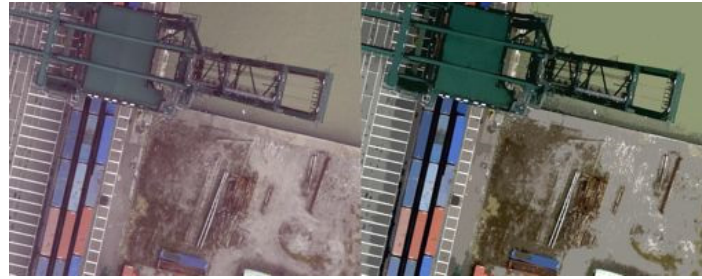


Fig. 6-12. Left: Subset of the original uncalibrated RGB image, before color constancy. Right: same subset, after automatic RGB image pre-processing, consisting of image enhancement by statistical color constancy, RGBIAM's image mapping into 50 color names and RGBIAM's segment-constant edge-preserving image reconstruction. No image histogram stretching is applied for visualization purposes.

The automated RGBIAM pipeline for a quantitative-to-nominal-to-quantitative (QnQ) transform of a monitor-typical true- or false-color RGB image is shown in Fig. 6-11. It consists of: (1) a forward RGBIAM's Q-to-N variable transform. It maps an RGB image onto two multi-level color maps, whose legends consist of 50 and 12 color names, see Fig. 6-10. (2) An inverse RGBIAM's N-to-Q variable transform. It provides an RGB color VQ error estimation, specifically, an RMSE image estimation, in compliance with the QA4EO's *Val* requirements [56]. To this end, each of the two RGBIAM's multi-level color maps was deterministically partitioned into an image segmentation map by a well-known two-pass connected-component multi-level image-labeling algorithm [53]. The RGBIAM's planar segments identified in the 2D color map domain, consisting of connected pixels featuring the same color name, are traditionally known as texels (texture elements), textons [61], tokens [43], or superpixels in the recent CV literature [62]. In other words, RGBIAM works as a texel detector at the Marr's raw primal sketch in low-level (pre-attentional) vision. Next, for each RGB color quantization level, fine or coarse, a segment description table (SDT) was generated as a tabular representation of the texel information [45]. In an SDT estimated in one image pass, each texel was described by its positional (e.g., minimum enclosing rectangle), photometric (e.g., mean MS value) and geometric attributes (e.g., area). Finally, based on each pair of one SDT and one segmentation map, a texel-constant edge-preserving approximation of the input RGB image (mean value of the RGB image per texel object) was automatically generated in linear time (see Fig. 6-12). The comparison of the input RGB image with the output reconstructed RGB image allowed estimation of an RMSE image as a community-agreed Q<sup>2</sup>I in VQ problems [51], [56].

#### 6.2.4 Second-Stage Classification with Spatial Reasoning in an Heterogeneous 2D and 3D Data Space

Traditionally mimicked by fuzzy logic [63], symbolic human reasoning is grounded on the transformation of a quantitative (numeric) variable, such as ever-varying sensations, into a qualitative (categorical, nominal) variable consisting of fuzzy sets, such as discrete and stable percepts [45]. Our hybrid CV system classification second stage was input with five geospatial variables, either quantitative (*information-as-thing* [64]) or categorical (*information-as-data-interpretation* [64]): (I) A quantitative 3D LiDAR point cloud. (II) A quantitative LiDAR data-derived DSM. (III) A quantitative piecewise-constant edge-preserving simplification of the original RGB image, see Fig. 6-12. (IV) Two categorical and semi-symbolic RGBIAM's pre-classification maps into color names of the input RGB image, see Fig. 6-10. (V) Two categorical sub-symbolic RGB image segmentation maps, consisting of planar objects automatically extracted from the two multi-level pre-classification maps, provided with inter-map parent-child relationships. To infer output variables of higher information quality from the combination of numeric and categorical geospatial variables, our *hybrid* CV system's classification second stage adopted two strategies: (1) A "stratified" approach to quantitative variable analysis, where geospatial numeric variables are conditioned by geospatial categorical variables in compliance with the principle of statistical stratification [45], [46], [54]. (2) An OBIA approach to spatial symbolic reasoning on geospatial categorical variables, consisting of discrete geometric objects, either planar (2D) or 3D, by means of physical model-based syntactic decision trees [53]. Unlike traditional data fusion techniques, where multiple sources of quantitative (sensory) data are merged, the proposed system fuses data sources only after they were individually transformed into higher-level qualitative variables capable of carrying some sort of semantics. The implemented *hybrid* CV system's classification second stage consisted of five subsystems, coded in the Cognition Network Language (CNL), within the eCognition Developer



software environment (Trimble Geospatial).

1) *Convergence-of-evidence criterion for automated background terrain extraction from the DSM and the RGB image.* An automated (well-conditioned) eCognition multiresolution image segmentation algorithm [55b] was run to extract planar objects in the DSM image-domain featuring within-object nearly constant DSM values. Merging adjacent 2D objects whose height differences was below 1 m resulted in, amongst others, one very large planar object, corresponding to the dominating background terrain across the depicted surface area. Next, color names of background surface types, such as asphalt, bare soil or water, were visually selected, combined by a logical OR operator and overlapped with the DSM-derived background mask. Finally, a foreground binary mask was generated as the inverse of the background binary mask. Foreground planar objects were candidates for 3D ISO container detection.

2) *Candidate 3D object selection based on converging 2D and 3D data-derived information.* Masked by the foreground binary mask detected at step 1, the RGBIAM's planar objects (texels) detected in the RGB image-domain were considered as input geospatial information primitives. In the orthophoto domain, the top view of a 3D ISO container looked like a single foreground image-object provided with its RGBIAM's color name. To assign an object-specific height value  $z$  to each foreground planar object, the tilting effect observed in the orthorectified RGB image (refer to Chapter 9.3.B) had to be coped with. To this end, a 2D object-specific height value was estimated as the 90% quantile of the LiDAR point cloud's  $z$  elevation values whose  $(x, y)$  coordinates fell on the target planar object. Foreground image-objects with an estimated height higher than a physical model-based maximum height of 26 m, corresponding to ten stacked ISO containers (considered as the possible maximum stacking height [42a]) were removed from the set of 3D container-candidate objects, such as image-objects related to cranes in the scene-domain. Finally, a spatial decision rule exploiting inter-object spatial relationships was applied to mask out small-size non-elevated 3D objects, whose planar projection was below the minimum ISO container area and that were isolated, i.e., surrounded by background areas exclusively. The result was a binary mask of container-candidate planar areas, where containers could be stacked up to ten layers.

3) *Driven-by-knowledge refined segmentation of the RGB image.* Masked by the binary candidate-container image areas detected in step 2, the edge-preserving smoothed RGB image (see Fig. 6-10) was input to a well-posed multiresolution eCognition segmentation algorithm [55b], whose free-parameter "planar shape compactness" was selected in accordance with prior physical knowledge of the ISO container's length and width [42]. Unlike the first RGBIAM's image partition, based on a non-adaptive-to-data spectral knowledge base and applied image-wide, this second adaptive-to-data image segmentation algorithm was provided with physical constrains and run on a masked image subset (container area only), to make it less prone to inherent segmentation errors and faster to compute. This stage accounted for the LiDAR data-derived DSM image indirectly through the input binary mask, rather than directly by stacking it with the input RGB image, to avoid the aforementioned tilting effect.

4) *3D ISO container recognition and classification.* Classes of ISO containers in the scene-domain were described in user-speak by the following shape and size properties [42].

- ISO class 1 of 20' container: rectangular area of 15 m<sup>2</sup>, rectangular shape, height 2.6 m.
- ISO class 2 of 40' container: rectangular area of 30 m<sup>2</sup>, rectangular shape, height 2.6 m.

These size values were projected onto the image domain in techno-speak, based on a sensor-specific transformation function [45]. The planar projection of a 3D rectangular object belonging to the ISO container classes 1 and 2 was found to match an eCognition's image object-specific rectangular fit index of 0.75 in range [0, 1]. Input image-objects, detected at the previous step 3, were selected based on their fuzzy rectangular shape membership value. Surviving image-objects whose height was divided by the standard height of an ISO container, equal to 2.6 meters, provided an estimate of the number of stacked ISO containers per image-object. Finally, vector container-objects were assigned to ISO container classes 1 or 2 based on their length/width relation. In addition, for visualization and 3D reconstruction purposes, an eCognition function was run per container-object to simplify and orthogonalize vector object boundaries in agreement with the main direction identified as the angle featuring the largest sum of object-edges per container-object. Classified and 3D container-objects were exported as polygon vectors in a standard GIS-ready file format (e.g., .shp).

5) *Semantic labelling of the 3D LiDAR point cloud.* Geospatial locations of the container-objects classified in step 4 were spatially intersected with the 3D LiDAR point cloud, to provide semantic labels to the LiDAR's  $z$  values whose  $(x, y)$  spatial coordinates fell on the planar projection of a 3D container. The semantic labels were written into the point class field of the \*.las file, as shown in Fig. 6-13.





### 6.2.5 3D Reconstruction of ISO Containers

The synthesis of tangible 3D container-objects took place in a GIS commercial software product (ArcScene, ESRI). The RGB orthophoto was draped over the DSM. Vector 3D container-objects were extruded according to their relative height. Stacked containers were extruded several times, based on the estimated number of stacked containers. Stacked containers were visualized in the same RGB color value estimated for the container on top, see Fig. 6-14.

### 6.2.6 Results and Discussion

The implemented *hybrid CV* system ran automatically, because prior knowledge initialized inductive 2D and 3D data analysis without user interaction, and near real-time, because of its linear complexity with the data size. In a standard laptop computer, computation time was 1 min for the RGBIAM pre-classification and less than 5 min for the second-stage OBIA classification. It detected 1659 containers distributed in 690 container stacks, including 118 single containers, 201 stacks of two, 361 stacks of three, 2 stacks of four and 8 stacks of six containers, see Fig. 6-13 and Fig. 6-14. In comparison with a reference “container truth”, acquired by an independent photointerpreter from the RGB image, detected ISO containers revealed a high overall accuracy, see Table 6-1, together with producer's and user's accuracies superior to 96%, although the user's accuracy for the 20' container class, affected by 39 false-positive occurrences, scored 81%. Further investigation revealed that the majority of these false positives were due to either true containers (e.g., few 40' container-objects were recognized as pairs of 20' container-objects due to spectral disturbances) or real-world objects similar to ISO containers in shape and size (e.g., trucks). A qualitative comparison of the 3D container reconstruction with the original LiDAR point cloud revealed a qualitatively “high” accuracy of height estimation per container stack, see Fig. 6-15.

Table 6-1

Accuracy assessment of the extracted container types

Container type	Automatic Assessment (2D)	Visual Assessment (2D)	Matches	Producer's accuracy (%)	User's accuracy (%)
40'	425	426	415	97.42	97.65
20'	265	226	217	96.02	81.89
Aggregated	690	652	632	96.93	91.59

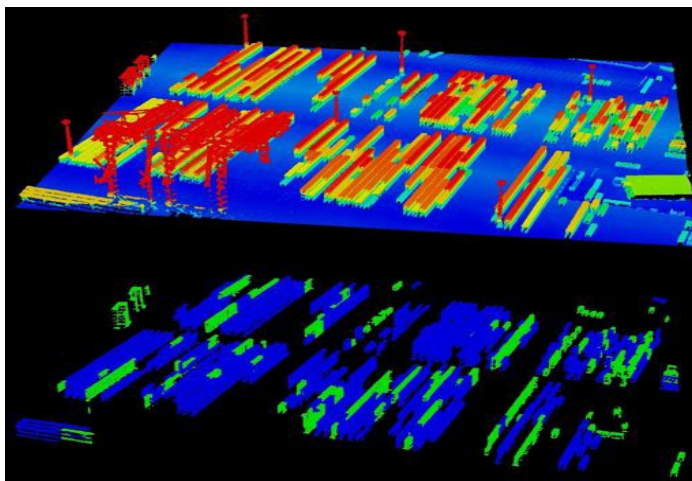




Fig. 6-13. Above: original LiDAR point cloud, colored from blue to red according to increasing point elevation values. Below: detected and classified LiDAR "container-points", where green = 20' container class 1, blue = 40' container class 2 (scene extent: (275x325 m)).



Fig. 6-14. Subset of the 3D reconstructed container terminal. Each container is a tangible 3D object featuring positional, colorimetric, geometric and identification attributes. Stacked containers were visualized in the same color of the container on top, according to the per-object mean RGB value extracted from the RGB orthophoto.

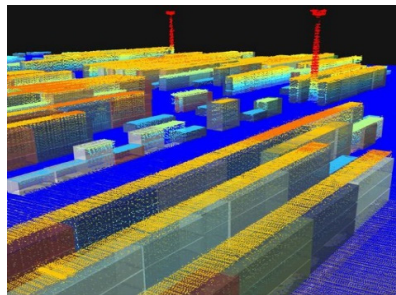


Fig. 6-15. Overlay of the 3D container reconstruction with the original LiDAR point cloud, colored from blue over orange to red according to increasing point elevation values. It qualitatively shows the estimated height per container stack looked accurate.

In line with theoretical expectations about hybrid inference, these experimental results reveal that the implemented *hybrid CV* system can be considered in operating mode by scoring "high" in a set of minimally redundant and maximally informative  $Q^2I$ s, to be community-agreed upon in accordance with the QA4EO's *Val* requirements [56]. Estimated  $Q^2I$  values encompassed [46]: accuracy ("high", see Table 6-1 and Fig. 6-15), efficiency ("high", linear time), degree of automation ("high", no user interaction), timeliness from data acquisition to product generation including training data collection, to be kept low ("low", no training-from-data, physical models were intuitive to tune, etc.) and scalability to different CV problems ("high", the CV system pipeline is data- and application-independent up to the CV system's target-specific classification second stage). The conclusion is, that *hybrid feedback CV* system design and implementation strategies can contribute to tackle the increasing demand for off-the-shelf software products capable of filling the semantic information gap from 2D and 3D big sensory data to high-level geospatial information products, where 2D data are typically uncalibrated, such as images acquired by consumer-level color cameras mounted on mobile devices, including unmanned aircraft systems (UASs).

### 6.3 Discussion of the 3D Contest

#### 6.3.1 Submissions

A total of ten submissions were received for the 3D contest. The participants mainly addressed object detection as the topic of their research. In total 80% of the submissions (Fig. 6-14) used the data sets to extract and to improve the detection of different objects in the urban and the harbor environments. Regardless of the homogeneity in the selections of topics, and as may be observed in Fig. 6-16, the processing methods were highly heterogeneous. In general, the proposed processing schemes were complex and time consuming, and integrated various learning and



optimisation procedures. The complex but nevertheless novel approaches proposed by the two winning teams proposed image modelling and processing tools for overcoming the different geometrical scanning modes and the limitations of each data set.

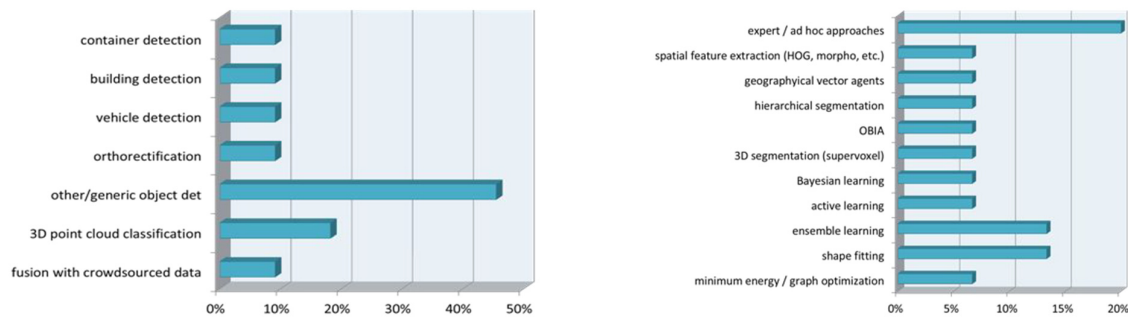


Fig. 6-16. Summary of the 10 submissions to the 3D contest by topics (left) and approaches considered (right).

### 6.3.2 The winners

- The winning team considered the problem of road extraction (Section II). As discussed above, automatic extraction of roads in complex urban scenes using remote sensing sources is an open and challenging research topic, less active in comparison to the detection of buildings or trees. The innovative processing solution proposed by the winning team was found remarkable for two reasons: (1) it takes into account the characteristics of the modern urban landscape that includes many different road materials and types but also road obstacles such as speed bumps and road curbs; (2) it applies and combines methods on different levels of processing including the laser's point clouds, grids and regions. This approach supports the importance of LiDAR as a highly accurate remote sensing source for road and object detection in urban areas. The integration of optical data is mainly used to exclude vegetal features such as grassy areas as well as to reduce the ambiguity in areas near the buildings.
- Runner up team addressed the detection of containers in the harbor of Zeebrugge (Section III). Typically, spatial information dominates color information in a spatiotemporal 4D real world-through-time domain and in an image domain, irrespective of data dimensionality reduction from 4D to 2D [45], [46]. In agreement with the increasingly popular OBIA paradigm [47] and with the *Rad/Cal* requirements of the QA4EO guidelines [56], a *hybrid feedback* CV system was selected for spatial reasoning upon a VHR uncalibrated 2D data set and a dense 3D data set, where deductive inference provided initial conditions to inherently ill-posed inductive learning-from-data algorithms without user interaction. An automated *hybrid* CV system first stage comprised a mandatory RGB color constancy algorithm, required for uncalibrated image harmonization and interoperability, followed by an *a priori* knowledge-based decision tree for static RGB cube partitioning (color naming). It differs from the inductive CV system solution proposed in Part B - Section II, where inductive learning-from-data occurred without requiring any uncalibrated RGB image harmonization policy. The *hybrid* CV system's classification second stage fused evidence stemming from the 2D and 3D data sets in two forms. (i) Fuzzy physical model-based decision-tree combinations of geospatial categorical variables. (ii) "Stratified" statistical analysis of geospatial numeric variables conditioned by categorical variables. Fusion of categorical variables (*information-as-data-interpretation* [70]) is alternative to traditional fusion of numeric variables (*information-as-thing*), where 2D and 3D vector data were stacked before interpretation. In addition, the *hybrid* CV system's classification second stage applied spatial reasoning on geometric objects in the (2D) image-domain, according to an OBIA paradigm, avoiding the loss of topological information in the (2D) image-domain compared to traditional image analysis.

### 6.3 Conclusion on the Data Fusion Contest 2015

In this double paper, we presented and discussed the outcomes of the IEEE GRSS Data Fusion Contest 2015. In compliance with the 2-track structure of the contest, we discussed the results in two parts: the 2D contest in Part A [1] and the results of the 3D contest in Part B (this manuscript). The winners of both tracks showed innovative ways of dealing with the very timely sources of data proposed: extremely high resolution color data and high density LiDAR point clouds.

In the 2D contest, the emerging technology of Convolutional Neural Networks, which is becoming a very



prominent standard in computer vision, has emerged as a powerful and effective way of extracting knowledge from these complex data. The need of understanding the information learned by the network is highlighted by the winning team, which provided an in depth analysis of the properties of the deep network filters. The effectiveness in classifying land cover types was highlighted by the runner-up team, which provided a comprehensive and thorough benchmark and disclosed their evaluation ground truth to the community.

In the 3D contest, the need of working with the point cloud directly emerged as a clear need, since the precision of the DSM used for calculation proved to be fundamental for the detection tasks addressed by the participating teams. Solutions to computational problems were also deeply considered by the winning team, since high density LiDAR point cloud calls for new standards of storage and access to data. The runner-up team showed a *hybrid* CV approach to multi-source EO data interpretation where a deductive inference first stage, capable of modeling physical knowledge of the 3D scene and psychophysical evidence about human vision, provided second-stage inductive learning-from-examples algorithm with initial conditions without requiring user interaction. The implemented hybrid CV system showed that automated interpretation of a VHR uncalibrated RGB image was possible in combination with a dense 3D LiDAR data set.

Summing up, the organizers were extremely pleased by the quality of the solutions proposed and by the variety of fusion problems addressed and processing approaches adopted by the participants to the Contest. They ranged from cutting-edge machine learning methodologies to case-specific processing pipelines and to prior knowledge-based systems, with the goal of capturing the information conveyed by extremely high resolution 2D and 3D remote sensing data. The organizers do hope that the concepts emerging from the 2015 Data Fusion Contest will inspire new researches at the interface of computer vision and remote sensing and foster new joint uses of laser and optical data.

### Acknowledgements

The authors wish to express their greatest appreciation to the department of Communication, Information, Systems & Sensors (CISS) of Belgian Royal Military Academy (RMA), for acquiring and providing the data used in the competition and for indispensable contribution to the organization of the Contest, and the IEEE GRSS for continuously supporting the annual Data Fusion Contest through funding and resources. D. Tuia acknowledges the Swiss National Science Foundation for financial support, through the grant PP00P2-150593. The work presented in Section II was sponsored by funding from the European Commission in the form of European Research Council grant ERC-2012-StG 20111012 "RETURN - Rethinking Tunnelling in Urban Neighbourhoods" Project 30786.

### References in Chapter 6

- [1] M. Camps-Taberner, A. Romero-Soriano, C. Gatta, G. Camps-Valls, A. Lagrange, B. L. Saux, A. Beaupère, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, M. Ferecatu, M. Shimoni, G. Moser, and D. Tuia, "Processing of extremely high resolution LiDAR and optical data: Outcome of the 2015 IEEE GRSS Data Fusion Contest. Part A: 2D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, submitted.
- [2] T. Lakes, P. Hostert, B. Kleinschmit, S. Lauf, and J. Tigges, "Remote sensing and spatial modelling of the urban environment," in *Perspectives in urban ecology*, W. Endlicher, Ed. Springer, 2011.
- [3] J. Jung, E. Pasolli, S. Prasad, J. Tilton, and M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 491–502, 2014.
- [4] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, 2014.
- [5] W. Liao, X. Huang, F. Van Collie, A. Gautama, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, and D. Tuia, "Processing of thermal hyperspectral and digital color cameras: outcome of the 2014 data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2015.
- [6] J. Yoon, J. Shin, and K. Lee, "Land cover characteristics of airborne LiDAR intensity data: A case study," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 4, pp. 801–805, 2008.
- [7] C. Paris and L. Bruzzone, "A three-dimensional model-based approach to the estimation of the tree top height by fusing low-density LiDAR data and very high resolution optical images," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 1, pp. 467–480, 2015.





- [8] Y. Chen, L. Cheng, M. Li, J. Wang, L. Tong, and K. Yang, "Multiscale grid method for detection and reconstruction of building roofs from airborne LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4081–4094, 2014.
- [9] J. Secord and A. Zakhor, "Tree detection in urban regions using aerial LiDAR and image data," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 196–200, 2007.
- [10] M. Bandyopadhyay, J. A. N. van Aardt, and K. Cawse-Nicholson, "Classification and extraction of trees and buildings from urban scenes using discrete return LiDAR and aerial color imagery," in *Proc. SPIE Laser Radar Technology and Applications XVIII*, M. D. Turner and G. W. Kamerman, Eds., vol. 8731, 2013.
- [11] G. Zhou and X. Zhou, "Seamless fusion of LiDAR and aerial imagery for building extraction," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 11, pp. 7393–7407, 2014.
- [12] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 135–149, 2007.
- [13] A. Brook, M. Vandewal, and E. Ben-Dor, "Fusion of optical and thermal imagery and LiDAR data for application to 3-D urban environment and structure monitoring," in *Remote Sensing - Advanced Techniques and Platforms*, B. Escalante, Ed. InTech, 2012, pp. 29–50.
- [14] X. Hu, Y. Li, J. Shan, J. Zhang, and Y. Zhang, "Road centerline extraction in complex urban scenes from lidar data based on multiple features," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 11, pp. 7448–7456, 2014.
- [15] F. Rottensteiner and S. Clode, "Building and road extraction by LiDAR and imagery," in *Topographic Laser Ranging and Scanning: Principles and Processing*, 1st ed. Boca Raton, FL: CRC press, 2008, pp. 445–478.
- [16] Y.-W. Choi, Y.-W. Jang, H.-J. Lee, and G.-S. Cho, "Three-dimensional LiDAR data classifying to extract road point in urban area," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 4, pp. 725–729, 2008.
- [17] P. Zhu, Z. Lu, X. Chen, K. Honda, and A. Elumnoch, "Extraction of city roads through shadow path reconstruction using laser data," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 12, pp. 1433–1440, 2004.
- [18] A. Habib, M. Ghanma, M. Morgan, and R. Al-Ruzouq, "Photogrammetric and lidar data registration using linear features," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 6, pp. 699–707, 2005.
- [19] L. Zhou, "Fusing laser point cloud and visual image at data level using a new reconstruction algorithm," in *Proc. IEEE Intelligent Vehicles Symposium*, 2013, pp. 1356–1361.
- [20] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained partbased models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [21] R. Tylecek and R. Šára, "Spatial pattern templates for recognition of objects with regular structure," *Pattern Recognition*, vol. 8142, pp. 364–374, 2013.
- [22] I. Stamos and P. K. Allen, "Geometry and texture recovery of scenes of large scale," *Comput. Vis. Image Underst.*, vol. 88, no. 2, pp. 94–118, 2002.
- [23] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, and S. Zokai, "Integrating automated range registration with multi-view geometry for the photorealistic modeling of large-scale scenes," *Int. J. Comp. Vision*, vol. 78, pp. 237–260, 2008.
- [24] H. Kim, C. D. Correa, and N. Max, "Automatic registration of LiDAR and optical imagery using depth map stereo," in *Proc. IEEE International Conference on Computational Photography*, Santa Clara, CA, 2014.
- [25] C. Frueh, R. Sammon, and A. Zakho, "Automated texture mapping of 3d city models with oblique aerial imagery," in *Proc. 3DPVT*, Tokyo, Japan, 2004, pp. 396–403.
- [26] L. Liu, I. Stamos, G. Yu, G. Wolberg, and S. Zokai, "Multiview geometry for texture mapping 2D images onto 3D range data," in *Proc. Comp. Vision Pattern Rec.*, New York, NY, 2006, pp. 2293–2300.
- [27] T. Schenk and B. Csatho, "Fusion of lidar data and aerial imagery for a more complete surface description," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 3A, pp. 301–317, 2002.
- [28] K. Fujii and T. Arikawa, "Urban object reconstruction using airborne laser elevation image and aerial image," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2234–2240, 2002.
- [29] J. Xu, K. Kim, Z. Zhang, H.-W. Chen, and O. Yu, "2D/3D sensor exploitation and fusion for enhanced object detection," in *Proc. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2014, pp. 778–784.
- [30] A. Alharthy and J. Bethel, "Automated road extraction from lidar data," in *Proceedings of the ASPRS Annual Conference*, 2003, pp. 05–09.



- [31] S. Clode, P. Kootsookos, and F. Rottensteiner, "The automatic extraction of roads from LiDAR data," in *The International Society for Photogrammetry and Remote Sensing's Twentieth Annual Congress*, vol. 35, 2004, pp. 231–237.
- [32] S. Clode, F. Rottensteiner, P. Kootsookos, and E. Zelniker, "Detection and vectorization of roads from lidar data," *Photogrammetric Engineering & Remote Sensing*, vol. 73, no. 5, pp. 517–535, 2007.
- [33] X. Hu, C. V. Tao, and Y. Hu, "Automatic road extraction from dense urban area by integrated processing of high resolution imagery and lidar data," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Istanbul, Turkey*, vol. 35, p. B3, 2004.
- [34] Y.-W. Choi, Y. W. Jang, H. J. Lee, and G.-S. Cho, "Heuristic road extraction," in *Information Technology Convergence, 2007. ISITC 2007. International Symposium on*. IEEE, 2007, pp. 338–342.
- [35] F. Samadzadegan, M. Hahn, and B. Bigdeli, "Automatic road extraction from lidar data based on classifier fusion," in *Urban Remote Sensing Event, 2009 Joint*. IEEE, 2009, pp. 1–6.
- [36] C. Ünsalan and K. L. Boyer, "Review on building and road detection," in *Multispectral Satellite Image Understanding*, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2011.
- [37] H. Kaartinen, J. Hyypä, X. Yu, M. Vastaranta, H. Hyypä, A. Kukko, M. Holopainen, C. Heipke, M. Hirschmugl, F. Morsdorf, E. Naesset, J. Pitkänen, S. Popescu, S. Solberg, B. M. Wolf, and J. C. Wu, "An international comparison of individual tree detection and extraction using airborne laser scanning," *Remote Sens.*, vol. 4, pp. 950–974, 2012.
- [38] J. Skilling, "Programming the hilbert curve," in *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 707, no. 1. AIP Publishing, 2004, pp. 381–387.
- [39] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [40] G. Sithole, "Detection of bricks in a masonry wall," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 1–6, 2008.
- [41] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 88–100, 2015.
- [42a] Y. Wild, R. Scharnow, and M. Ruhmann. "Container handbook." GDV German Insurance Association, Berlin, 2005.
- [42] "Iso 668:2013 series 1 freight containers - classification, dimensions and ratings." 2015. [Online]. Available: <https://www.document-center.com/standards/show/ISO-668>
- [43] D. Marr, *Vision*. New York, NY: Freeman and C., 1982.
- [44] H. du Buf and J. Rodrigues, *Image morphology: from perception to rendering*, in *IMAGE - Computational Visualistics and Picture Morphology*, 2007.
- [45] T. Matsuyama and V. S.-S. Hwang, *SIGMA: A knowledge-based aerial image understanding system*. Springer Science & Business Media, 2013.
- [46] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery - part i: System design and implementation," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1299–1325, 2010.
- [47] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, "Geographic object-based image analysis - towards a new paradigm." *ISPRS J. Photo. Remote Sens.*, vol. 87, no. 100, pp. 180–191, 2014.
- [48] D. Tiede, S. Lang, F. Albrecht, and D. Hölbling, "Object-based class modeling for cadastre-constrained delineation of geo-objects," *Photo. Eng. Remote Sens.*, vol. 76, no. 2, pp. 193–202, 2010.
- [49] D. Tiede, "A new geospatial overlay method for the analysis and visualization of spatial change patterns using object-oriented data modeling concepts," *Cartography and Geographic Information Science*, vol. 41, no. 3, pp. 227–234, May 2014.
- [50] T. Martinetz, G. Berkovich, and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
- [51] V. S. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1998.
- [52] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons, 2004.



- [53] M. Sonka, V. Hlavax, and R. Boyle, *Image Processing, Analysis and Machine Vision*, 1st ed. UK: Chapman & Hall, 1993.
- [54] P. Mather, *Computer Processing of Remotely-Sensed Images—An Introduction*. Hoboken, NJ, USA: Wiley, 1994.
- [55a] S. Frintrop, “Computational visual attention,” in *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds., Springer, 2011.
- [55b] M. Baatz and A. Schäpe, “Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation.” in *Angewandte Geographische Informationsverarbeitung*, J. Strobl, T. Blaschke, and G. Griesebner, Eds. Heidelberg: Wichmann-Verlag, 2000, pp. 12–23.
- [56] GEO/CEOS, “A Quality Assurance Framework for Earth Observation,” Group on Earth Observation / Committee on Earth Observation Satellites, Tech. Rep. 4, 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- [57] T. Gevers, A. Gijsenij, J. van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision : Fundamentals and Applications*. Wiley, 2012.
- [58] L. D. Griffin, “Optimality of the basic color categories for classification”, *J. R. Soc. Interface*, vol. 3, pp. 71–85, 2006.
- [59] B. Berlin and P. Kay, *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.
- [60] Y. Linde, A. Buzo, and R. M. Gray, “An algorithm for vector quantizer design,” *IEEE Trans. Commun.*, vol. 28, pp. 89–94, 1980.
- [61] B. Julesz, “Texton gradients: The texton theory revisited,” in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, vol. 54, no. 4-5, Aug. 1986.
- [62] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [63] L. A. Zadeh, “Fuzzy sets,” *Inform. Control*, vol. 8, pp. 338–353, 1965.
- [64] R. Capurro and B. Hjørland, “The concept of information,” *Annual Review of Information Science and Technology*, vol. 37, pp. 343-411, 2003.

## 7 Manuscript 4 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01930): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 1: Theory

### Motivation and Contributions to the Dissertation

Among the original CV algorithms proposed in Chapter 3 (Technical report 1) and adopted by an Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), there is an original pair of expert systems (prior knowledge-based decision trees) for color naming in a calibrated multi-spectral (MS) reflectance space or in an uncalibrated RGB color space, either true- or false-color. Color naming transforms a numeric variable (color value, colorimetric sensation) into a categorical variable, specifically, into color names belonging to a pre-defined dictionary of color names, equivalent to a latent/hidden variable and eligible for use in symbolic human reasoning. The present Chapter 7 (Manuscript 4) presents and discusses the non-trivial multi-disciplinary background of color naming, ranging from cognitive science to artificial intelligence (AI) and machine learning-from-data.

Provided with a relevant survey value, Chapter 7 (Manuscript 4) features several degrees of novelty. First, to cope with dictionaries of MS color names and land cover class names that do not coincide and must be harmonized, an original hybrid (combined deductive and inductive) guideline is proposed to identify a categorical variable-pair (binary) relationship. Second, an original quantitative measure of categorical variable-pair association is proposed, given a categorical variable-pair relationship, independent of frequency counts of the two univariate categorical variables generated from a single population.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 7 (Manuscript 4) are color filled.

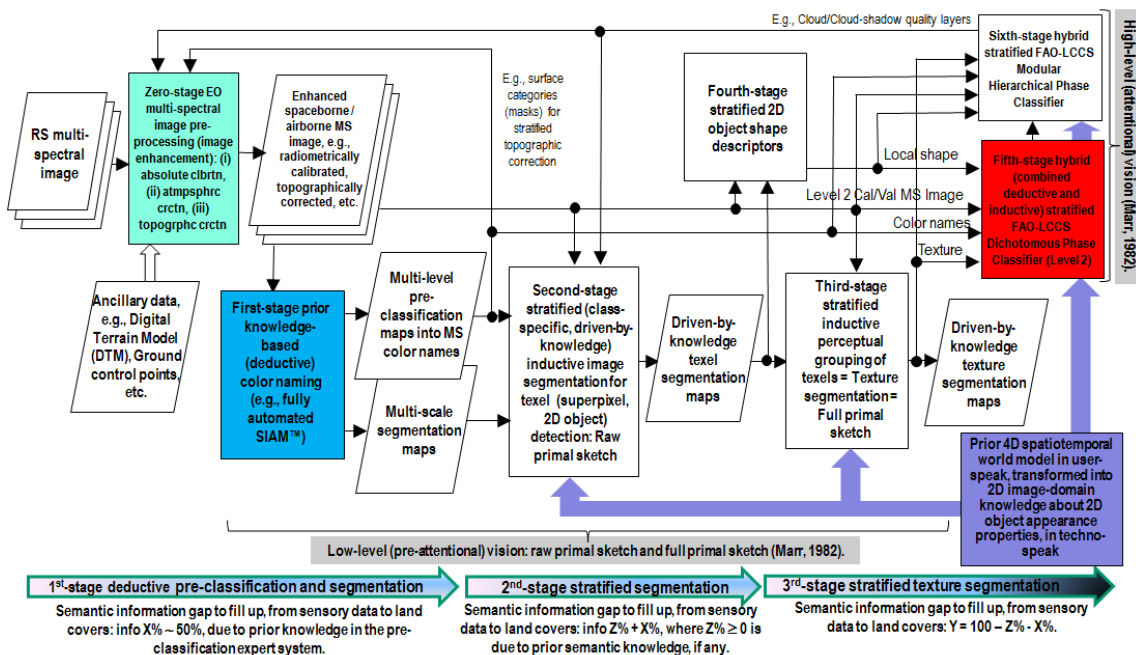


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid





feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 7 (Manuscript 4).



## Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation – Part 1: Theory

Andrea Baraldi<sup>a,c,\*</sup>, Michael Laurence Humber<sup>b</sup>, Dirk Tiede<sup>c</sup> and Stefan Lang<sup>c</sup>

<sup>a</sup> Department of Agricultural Sciences, University of Naples Federico II, Portici (NA), Italy.

<sup>b</sup> Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA.

<sup>c</sup> Department of Geoinformatics – Z\_GIS, University of Salzburg, Salzburg 5020, Austria.

\*Corresponding author. Email: andrea6311@gmail.com

### Abstract

The European Space Agency (ESA) defines an Earth Observation (EO) Level 2 product as a multi-spectral (MS) image corrected for geometric, atmospheric, adjacency and topographic effects, stacked with its data-derived scene classification map (SCM), whose legend includes quality layers such as cloud and cloud-shadow. No ESA EO Level 2 product has ever been systematically generated at the ground segment. To contribute toward filling an analytic and pragmatic information gap from EO big data to the ESA EO Level 2 product, an original Stage 4 validation (*Val*) of the Satellite Image Automatic Mapper (SIAM) lightweight computer program was conducted by independent means on an annual Web-Enabled Landsat Data (WELD) image composite time-series of the conterminous U.S. The non-iterative SIAM application was designed to run automatically in near real-time on the web and on mobile devices. Its core is a one-pass prior knowledge-based decision tree for MS reflectance space hyperpolyhedralization into static (non-adaptive-to-data) color names presented in literature in recent years. For the sake of readability this paper is split into two. The present Part 1 – Theory provides the multidisciplinary background of *a priori* color naming in cognitive science, from linguistics to computer vision, and surveys related works on static MS color naming. To cope with dictionaries of MS color names and land cover class names that do not coincide and must be harmonized, an original hybrid (combined deductive and inductive) guideline is proposed to identify a categorical variable-pair relationship. An original quantitative measure of categorical variable-pair association is also proposed. The subsequent Part 2 – Validation presents and discusses Stage 4 *Val* results collected by an original protocol for wall-to-wall thematic map quality assessment without sampling where the test and reference map legends can differ. Conclusions are that the SIAM-WELD maps instantiate a Level 2 SCM product whose legend is the 4-class taxonomy of the Food and Agriculture Organization of the United Nations – Land Cover Classification System (LCCS) at the Dichotomous Phase Level 1 (vegetation/non-vegetation), Level 2 (terrestrial/aquatic) or superior LCCS level.

### Keywords

Artificial intelligence; binary relationship; Cartesian product; color naming; connected-component multi-level image labeling; deductive inference; Earth observation; land cover class taxonomy; high- (attentive) and low-level (pre-attentive) vision; hybrid inference; image segmentation; inductive inference; machine learning-from-data; outcome and process quality indicators; radiometric calibration; remote sensing; thematic map comparison; two-way contingency table; unsupervised data discretization/vector quantization.

### 7.1 Introduction

Proposed by the intergovernmental Group on Earth Observations (GEO), the visionary goal of the Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015 (GEO 2005) is the systematic transformation of multi-source EO “big data” into timely, comprehensive and operational EO value-adding products and services, submitted to the Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements (GEO-CEOS 2010). To date the GEOSS mission cannot be considered fulfilled by the remote sensing (RS) community. Existing Earth observation (EO) image understanding systems (EO-IUSs) tend to score low in productivity because outpaced by the rate of collection of EO sensory data, whose quality and quantity are ever-increasing. To be considered affected by low levels of productivity, which means to be considered in non-operating mode, an EO-IUS suffices to fall short in one of its outcome and process (OP) quantitative quality indicators (Q<sup>2</sup>Is), to be community-agreed



upon in compliance with the QA4EO guidelines. A proposed minimally dependent and maximally informative (mDMI) set of EO OP-Q<sup>2</sup>Is includes degree of automation, effectiveness, e.g., thematic mapping accuracy, efficiency in computation time and in memory occupation, robustness (vice versa, sensitivity) to changes in input data, robustness to changes in input parameters to be user-defined, scalability to changes in user requirements and in sensor specifications, timeliness from data acquisition to information product generation, and costs in manpower and computer power (Baraldi and Boschetti 2012a, 2012b; Duke 2016). According to the Pareto formal analysis of multi-objective optimization problems, optimization of an mDMI set of OP-Q<sup>2</sup>Is is an inherently-ill posed problem in the Hadamard sense (Hadamard 1902), where many Pareto optimal solutions lying on the Pareto efficient frontier can be considered equally good (Boschetti et al. 2004).

The conjecture that existing EO-IUSs are Pareto sub-optimal and tend to score low in operating mode is supported by several facts. First, the percentage of EO data ever downloaded from the European Space Agency (ESA) databases is estimated at about 10% or less (D'Elia 2002). Second, EO-IUSs presented in the RS literature are typically assessed and compared based on the sole mapping accuracy, which means their mDMI set of OP-Q<sup>2</sup>Is remains largely unknown to date. For example, when a large-scale EO data-derived thematic map product was generated by a supervised data learning EO-IUS at “high” accuracy, the most limiting factors turned out to be the cost, timeliness, quality and availability of adequate supervised (labeled) data samples, collected from field sites, existing maps or geospatial data archives in tabular form (Gutman et al. 2004). Third, no ESA EO data-derived Level 2 prototype product has ever been systematically generated at the ground segment (ESA 2015; CNES 2016). An EO Level 2 product is defined by ESA as a multi-spectral (MS) image corrected for geometric, atmospheric, adjacency and topographic effects, equivalent to a multivariate numeric variable of enhanced information quality (related to the concept of quantitative unequivocal *information-as-thing* in the terminology of Capurro and Hjørland [2003]), see Fig. 7-1, stacked with its data-derived scene classification map (SCM) (ESA 2015; CNES 2016), equivalent to a categorical variable of semantic quality (related to the concept of qualitative equivocal *information-as-data-interpretation* in the terminology of Capurro and Hjørland [2003]). An ESA EO Level 2 SCM legend is expected to consist of general-purpose, user- and application-independent land cover (LC) classes, in addition to quality layers such cloud and cloud-shadow (ESA 2015; CNES 2016).

Noteworthy, ESA EO Level 2 product is superset of National Aeronautics and Space Administration (NASA) EO Level 2 product, defined as “a data-derived geophysical variable at the same resolution and location as Level 1 source data” (NASA 2016), see Fig. 7-2.

A possible example of ESA EO Level 2 SCM legend is the 3-level 8-class Dichotomous Phase (DP) taxonomy of the Food and Agriculture Organization of the United Nations (FAO) – Land Cover Classification System (LCCS) (Di Gregorio and Jansen 2000). The LCCS-DP hierarchy comprises three “nested” dichotomous LC class layers: Level 1 – Vegetation versus non-vegetation, Level 2 – Terrestrial versus aquatic and Level 3 – Managed versus natural or semi-natural. The 3-level 8-class LCCS-DP taxonomy is listed in Fig. 7-3. For the sake of generality, a 3-level 8-class LCCS-DP legend added with LC class “other” or “rest of the world”, which includes information layers such as cloud and cloud-shadow, is identified hereafter as “augmented” 9-class LCCS-DP taxonomy. In the complete two-phase LCCS hierarchy a low-level general-purpose LCCS-DP legend is preliminary to a high-level application- and user-specific LCCS Modular Hierarchical Phase (MHP) taxonomy, consisting of a hierarchical battery of one-class LC class-specific classifiers (Di Gregorio and Jansen 2000). In recent years the two-phase LCCS taxonomy has become increasingly popular (Ahlqvist 2008). One reason of its popularity is that the LCCS hierarchy is “fully nested” while alternative LC class hierarchies, such as the CORINE Land Cover (CLC) taxonomy (Bossard et al. 2000) and the EO Image Librarian LC taxonomy (Dumitru et al. 2015), start from a Level 1 which is multi-class. In a hierarchical EO-IUS architecture submitted to a garbage in, garbage out information principle, the fully-nested LCCS hierarchy highlights the dependence of OP-Q<sup>2</sup>Is featured by any high-level LCCS-MHP data processing module on OP-Q<sup>2</sup>Is featured by low-level LCCS-DP data processing units, starting from the LCCS-DP Level 1 vegetation/non-vegetation information layer whose relevance becomes paramount for all subsequent LCCS layers. This multi-level dependence is neither trivial nor obvious to underline. For example, vegetation/non-vegetation discrimination is acknowledged to be very challenging when pursued in EO image composites at continental or global scale by means of traditional supervised data learning EO-IUSs (Gutman et al. 2004), which are inherently semi-automatic and training data-specific (Liang 2004).

Our working thesis was that a necessary not sufficient pre-condition for a yet-unfulfilled GEOSS development (GEO 2005) is the systematic generation at the ground segment of an ESA EO Level 2 product, whose general-purpose SCM is constrained as follows. First, the SCM legend agrees with the 3-level 9-class “augmented” LCCS-DP taxonomy (see Fig. 7-3). Second, to comply with the QA4EO *Cal/Val* requirements the SCM product must be submitted to a GEO



Stage 4 *Val*, where an mDMI set of OP-Q<sup>2</sup>Is is evaluated by independent means (GEO-CEOS 2010). By definition a GEO Stage 3 *Val* requires that “spatial and temporal consistency of the product with similar products are evaluated by independent means over multiple locations and time periods representing global conditions. In Stage 4 *Val*, results for Stage 3 are systematically updated when new product versions are released and as the time-series expands” (GEO-CEOS WGCV 2015).

To contribute toward filling an analytic and pragmatic information gap from multi-source EO big data to the ESA EO Level 2 product, the primary goal of this interdisciplinary study was to undertake an original (to the best of these authors’ knowledge, the first) outcome and process Stage 4 *Val* of an off-the-shelf Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program. Implemented in operating mode in the C/C++ programming language, the SIAM software executable runs: (i) automatically, i.e., it requires no human-machine interaction, (iii) in near real-time, specifically, it is non-iterative (one-pass with a single subsystem that is two-pass, refer to the text below), with a computational complexity increasing linearly with the image size, and (iii) in tile streaming mode, i.e., it requires a fixed runtime memory occupation (Baraldi et al. 2006, 2010a, 2010b, 2010c, 2012a, 2012b, 2013, 2015, 2016; Baraldi and Humber 2015). In addition to running on laptop and desktop computers the SIAM lightweight computer program is eligible for use in a mobile software application. Eventually provided with a mobile user interface, a mobile software application is a lightweight computer program specifically designed to run on web browsers and mobile devices, such as tablet computers and smartphones. The SIAM software pipeline comprises six non-iterative subsystems for MS image analysis (decomposition) and synthesis (reconstruction). Its core is a one-pass prior knowledge-based decision tree for MS reflectance space hyperpolyhedralization into static (non-adaptive-to-data) color names, presented in the RS literature where enough information was provided for the implementation to be reproduced (Baraldi et al. 2006). Sketched in Fig. 7-4, the SIAM application workflow is summarized hereafter.

(1) MS data radiometric calibration, in agreement with the QA4EO *Cal* requirements (GEO-CEOS 2010). The SIAM expert system instantiates a physical data model; hence, it requires as input sensory data provided with a physical meaning. Specifically, digital numbers must be radiometrically calibrated into a physical unit of radiometric measure to be community-agreed upon, such as top-of-atmosphere reflectance (TOARF), surface reflectance (SURF) or Kelvin degrees for thermal channels. Relationship  $TOARF \supseteq SURF$  holds because SURF is a special case of TOARF in clear sky and flat terrain conditions (Chavez 1988), i.e.,  $TOARF \approx SURF + \text{atmospheric noise} + \text{topographic effects} + \text{surface adjacency effects}$ . In a spectral decision tree this relationship means that if decision boundaries are able to cope with MS hyperpolyhedra of “noisy” TOARF values then they can also deal with “noiseless” SURF values as special cases of the former, while the vice versa does not necessarily hold, see Fig. 7-5.

(2) One-pass prior knowledge-based SIAM decision tree for MS reflectance space hyperpolyhedralization into three static codebooks (dictionaries) of sub-symbolic color names as codewords, see Fig. 7-5. Provided with inter-level parent-child relationships, the SIAM’s three-level dictionary of static color names features a ColorDictionaryCardinality value which decreases from fine to intermediate to coarse, refer to Table 7-1 and Fig. 7-6. MS reflectance space hyperpolyhedra for color naming are difficult to think of and impossible to visualize when the MS data space dimensionality is superior to three. This is not the case of basic color names adopted in human languages (Berlin and Kay 1969), whose mutually exclusive and totally exhaustive perceptual 280labelled280d, neither necessarily convex nor connected, are intuitive to think of and easy to visualize in a 3D monitor-typical red-green-blue (RGB) cube, see Fig. 7-7 (Griffin 2006; Benavente et al. 2008). When each pixel of a MS image is mapped onto a color space partitioned into a set of mutually exclusive and totally exhaustive hyperpolyhedra equivalent to a dictionary of color names, then a 2D multi-level color map is generated automatically (without human-machine interaction) in near real-time (with a computational complexity increasing linearly with the image size), where the number  $k$  of 2D map levels (color strata, color names) belongs to range  $\{1, \text{ColorDictionaryCardinality}\}$ . Popular synonyms of measurement space hyperpolyhedralization (discretization, partition) are vector quantization (VQ) in inductive machine learning-from-data (Cherkassky and Mulier 1998) (Fritzke 1997a, 1997b; Patanè and Russo 2001, 2002; Linde et al. 1980; Lee et al. 1997; Lloyd 1982; Elkan 2003), and deductive fuzzification of a numeric variable into fuzzy sets in fully logic (Zadeh 1965). Typical inductive learning-from-data VQ algorithms aim at minimizing a known VQ error function, e.g., a root mean square error (RMSE), given a number of  $k$  discretization levels selected by a user based on *a priori* knowledge and/or heuristic criteria. One of the most widely used VQ heuristics in RS and computer vision (CV) applications is the  $k$ -means VQ algorithm (Linde et al. 1980; Lee et al. 1997; Lloyd 1982; Elkan 2003), capable of convex Voronoi tessellation of an 280labelled280d data space (Fritzke 1997a; Cherkassky and Mulier 1998). For example, in a bag-of-words model applied to CV tasks, a numeric color space is typically





discretized into a categorical color variable by an inductive VQ algorithm, such as  $k$ -means; next, the categorical color variable is simplified by a 1<sup>st</sup>-order histogram representation, which disregards word grammar, semantics and even word-order, but keeps multiplicity; finally, the frequency of each color codeword is used as a feature for training a supervised data learning classifier (Cimpoi et al. 2014). Unlike the  $k$ -means VQ algorithm where  $k$  is user-defined and the VQ error is estimated from 281labelled281d data, a user can fix the target VQ error value, so that it is the free-parameter  $k$  to be dynamically learned from 281labelled281d data by an inductive VQ algorithm (Patanè and Russo 2001, 2002), such as ISODATA (Memarsadeghi et al. 2007). It means there is no universal number  $k$  of static hyperpolyhedra in a vector data space suitable for satisfying any VQ error specification. As a viable strategy to cope with the inherent ill-posedness of VQ problems (Cherkassky and Mulier 1998), the SIAM expert system provides its three pre-defined VQ levels with a per-pixel RMSE estimation required for VQ quality assurance, in compliance with the QA4EO guidelines, refer to point (6) below.

(3) Well-posed (deterministic) two-pass detection of connected-components in the multi-level color map-domain (Dillencourt et al. 1992; Sonka et al. 1994), where the number  $k$  of map levels is  $\leq$  ColorDictionaryCardinality, see Fig. 7-8. These connected-components consist of connected sets of pixels featuring the same color label. They are typically known as superpixels in the CV literature (Achanta et al. 2011), homogeneous segments or image-objects in the object-based image analysis (OBIA) literature (Blaschke et al. 2014; Nagao and Matsuyama 1980; Matsuyama and Hwang 1990; Shackelford and Davis 2003a, 2003b), and texture elements (texels) in human vision (Julesz 1986; Julesz et al. 1973). Whereas the physical model-based SIAM expert system requires no human-machine interaction to detect top-down superpixels whose shape and size can be any, superpixels detected bottom-up in CV applications typically require a pair of statistical model's free-parameters to be user-defined based on heuristics. This pair of user-defined parameters typically thresholds the superpixel maximum area and forces a superpixel to stay compact in shape (Achanta et al. 2011). In a multi-level image domain where  $k$  is the number of levels (image-wide strata), superpixels, pixels and strata co-exist as labelled spatial units provided with a parent-child relationship, where each superpixel is a 2-tuple [superpixel ID, level 1-of- $k$ ] and each pixel is a 2-tuple [row-column coordinate pair, superpixel ID].

(4) Well-posed 4- or 8-adjacency cross-aura representation in linear time of superpixel-contours, see Fig. 7-9. These cross-aura contour values allow estimation of a scale-invariant planar shape index of compactness (Soares et al. 2014), eligible for use by a high-level OBIA approach (Blaschke et al. 2014), see Fig. 7-4.

(5) Superpixel description table allocation and initialization, to describe superpixels in a 1D tabular form (list) in combination with their 2D raster representation, to take advantage of each data structure and overcome their shortcomings (Nagao and Matsuyama 1980; Matsuyama and Hwang 1990; Marr 1980). Typically local spatial searches are computational more efficient in the raster domain than in the list representation (Marr 1980).

(6) Superpixelwise-constant input image approximation (reconstruction), also known as “image-object mean view” in OBIA applications (Trimble 2015), followed by a per-pixel RMSE estimation between the original MS image and the reconstructed piecewise-constant MS image. This VQ error estimation strategy enforces a product quality assurance policy considered mandatory by the QA4EO guidelines. For example, VQ quality assurance supported by SIAM allows a user to adopt quantitative (objective) criteria in the selection of pre-defined VQ levels to fit user- and application-specific VQ error requirement specifications.

An example of the SIAM output products automatically generated in linear time from a radiometrically calibrated 13-band 10 m-resolution Sentinel-2A image is shown in Fig. 7-10.

The potential impact on the RS community of a Stage 4 *Val* of an off-the-shelf SIAM lightweight computer program for prior knowledge-based MS reflectance space hyperpolyhedralization, superpixel detection and per-pixel VQ quality assessment is expected to be relevant, with special emphasis on existing or future *hybrid* (combined deductive and inductive) EO-IUSs. In the RS discipline there is a long history of prior knowledge-based MS reflectance space partitioners for static color naming, alternative to SIAM's, developed but never validated by space agencies, public organizations and private companies for use in hybrid EO-IUSs in operating mode, see Fig. 7-11. Examples of hybrid EO data pre-processing applications (*information-as-thing*) conditioned by static color naming are large-scale MS image compositing (Ackerman et al. 1998; Luo et al. 2008; Lück and van Niekerk 2016), MS image atmospheric correction (Richter and D. Schläpfer 2012a; Richter and D. Schläpfer 2012b; Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Dorigo et al. 2009; Vermote and Saleous 2007; DLR and VEGA 2011; Lück and van Niekerk 2016), MS image topographic correction (Richter and D. Schläpfer 2012a; Richter and D. Schläpfer 2012b; Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Dorigo et al. 2009; Baraldi et al. 2010c; DLR and VEGA 2011; Lück and van Niekerk 2016), see Fig. 7-12, MS image adjacency effect correction (DLR and VEGA 2011) and radiometric quality assurance of pan-sharpened MS



imagery (Despini et al. 2014). Examples of hybrid EO image classification applications (*information-as-data-interpretation*) conditioned by static color naming are cloud and cloud-shadow quality layer detection (Baraldi et al. 2015; DLR and VEGA 2011; Lück and van Niekerk 2016), single-date LC classification (Muirhead and Malkawi 1989; Simonetti et al. 2015a; GeoTerraImage 2015; DLR and VEGA 2011; Lück and van Niekerk 2016), multi-temporal post-classification LC change (LCC)/no-change detection (Baraldi et al. 2016; Tiede et al. 2016; Simonetti et al. 2015a), multi-temporal vegetation gradient detection and quantization in fuzzy sets (Arvor et al. 2016), multi-temporal burned area detection (Boschetti et al. 2015), and prior knowledge-based LC mask refinement (cleaning) of supervised data samples employed as input to supervised data learning EO-IUSs (Baraldi et al. 2010a, 2010b). Due to their large application domain ranging from low- (pre-attentional) to high-level (attentional) vision tasks, existing hybrid EO-IUSs in operating mode conditioned by static color naming are natural candidates for the systematic transformation of multi-source single-date MS imagery into ESA EO Level 2 product at the ground segment.

The terminology adopted in the rest of this paper is mainly driven from the multidisciplinary domain of cognitive science, see Fig. 7-13. Popular synonyms of deductive inference are top-down, prior knowledge-based, learning-from-rule and physical model-based inference. Synonyms of inductive inference are bottom-up, learning-from-data, learning-from-examples and statistical model-based inference (Baraldi and Boschetti 2012a, 2012b; Liang 2004). Hybrid inference systems combine statistical and physical models to take advantage of the unique features of each and overcome their shortcomings (Baraldi and Boschetti 2012a, 2012b; Cherkassky and Mulier 1998; Liang 2004). For example, in biological cognitive systems “there is never an absolute beginning” (Piaget 1970), where an *a priori* genotype provides initial conditions to an inductive learning-from-examples phenotype (Parisi 1991). Biological cognitive systems are hybrid inference systems where inductive/phenotypic learning-from-examples mechanisms explore the neighbourhood of deductive/genotypic initial conditions in a solution space (Parisi 1991). In line with biological cognitive systems an artificial hybrid inference system can alternate deductive and inductive inference units, starting from a deductive first stage for initialization purposes, see Fig. 7-11. It means that no deductive inference subsystem, such as SIAM, should be considered stand-alone, but eligible for use in a hybrid inference system architecture to initialize (pre-condition, stratify) inductive learning-from-data algorithms, which are inherently ill-posed (difficult-to-solve) and require *a priori* knowledge in addition to data to become better posed for numerical solution (Cherkassky and Mulier 1998).

To comply with the GEO Stage 4 *Cal/Val* requirements, the selected ready-for-use SIAM application had to be validated by independent means on a radiometrically calibrated EO image time-series at large spatial extent. This input data set was identified in the open-access U.S. Geological Survey (USGS) 30 m resolution Web Enabled Landsat Data (WELD) annual composites of the conterminous U.S. (CONUS) for the years 2006 to 2009, radiometrically *Cal* into TOARF values (Roy et al. 2010; Homer et al. 2004; WELD 2015). The 30 m resolution 16-class U.S. National Land Cover Data (NLCD) 2006 map, delivered in 2011 by the U.S. Geological Survey (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC) (Vogelmann et al. 1998, 2001; Wickham et al. 2010; Wickham et al. 2013; Xian and Homer 2010; EPA 2007), was selected as the reference thematic map at continental spatial extent. The 16-class NLCD map legend is summarized in Table 7-2. To account for typical non-stationary geospatial statistics, the NLCD 2006 thematic map was partitioned into 86 Level III ecoregions of North America collected from the Environmental Protection Agency (EPA) (EPA 2013; Griffith and Omernik 2009).

In this experimental framework the test SIAM-WELD annual color map time-series and the reference NLCD 2006 map share the same spatial extent and spatial resolution, but their map legends are not the same. These working hypotheses are neither trivial nor conventional in the RS literature where thematic map quality assessment strategies typically adopt an either random or non-random sampling strategy and assume that the test and reference thematic map dictionaries coincide (Stehman and Czaplewski 1998). Starting from a stratified random sampling protocol presented in (Baraldi et al. 2014), the secondary contribution of the present study was to develop a novel protocol for wall-to-wall comparison without sampling of two thematic maps featuring the same spatial extent and spatial resolution, but whose legends can differ.

For the sake of readability this paper is split into two, the present Part 1 – Theory and the subsequent Part 2 – Validation. An expert reader familiar with static color naming in cognitive science, spanning from linguistics to human vision and computer vision, can skip the present Part 1. To make this paper self-contained and provided with a relevant survey value, the Part 1 is organized as follows. The multidisciplinary background of color naming is discussed in Chapter 7.2. Chapter 7.3 reviews prior knowledge-based decision trees for MS color naming presented in the RS literature. To cope with thematic map legends that do not coincide and must be harmonized (reconciled, associated, translated) (Ahlqvist 2005), such as dictionaries of MS color names and LC class names, Chapter 7.3 proposes an original hybrid guideline to identify a categorical variable-pair relationship, where prior beliefs are combined with additional evidence inferred from



new data. An original measure of categorical variable-pair association is proposed in Chapter 7.4. In the subsequent Part 2 Stage 4 *Val* results are collected by an original protocol for wall-to-wall thematic map quality assessment without sampling where the test SIAM-WELD map legend and the reference NLCD 2006 map legend are harmonized. Conclusions are that the SIAM-WELD maps instantiate an ESA EO Level 2 SCM product whose legend is the FAO LCCS taxonomy at the DP Level 1 (vegetation/non-vegetation), Level 2 (terrestrial/aquatic) or superior LCCS level.

## 7.2 Color naming problem background in cognitive science

Vision is an inherently ill-posed cognitive problem where scene-from-image representation is affected, first, by data dimensionality reduction from the 4D spatiotemporal scene-domain to the (2D) image-domain and, second, by a semantic information gap from ever-varying sub-symbolic numeric sensations to stable categorical and semantic (symbolic) percepts (Matsuyama and Hwang 1990).

Ever-varying sensations are observable numeric/quantitative variables of “sub-symbolic” quality, i.e., sensations are equivalent to sensory data, directly measured in the real world and provided with no semantic content. Stable percepts are nominal/categorical/qualitative variables of “symbolic” quality, i.e., they are categorical variables provided with a semantic content in a modeled world, also known as world ontology or “world model” (Matsuyama and Hwang 1990). In statistics, latent/hidden variables are not directly measured, but rather inferred from observable numeric variables to link sensory data in the real world to categorical variables of semantic quality in the modeled world. The terms hypothetical variable or hypothetical construct may be used when latent variables correspond to abstract concepts, like perceptual categories or mental states. Hence, to fill the semantic gap from low-level numeric variables of sub-symbolic quality to high-level categorical variables of semantic quality, hypothetical variables are expected to be mid-level categorical variables of “semi-symbolic” quality, i.e., qualitative variables provided with some degree of semantic content.

In vision, spatial topological and spatial non-topological information typically dominate color information. This thesis is proved by the undisputable fact that achromatic (panchromatic) human vision, familiar to everybody when wearing sunglasses, is nearly as effective as chromatic vision in scene-from-image representation. It means that a necessary not sufficient condition for a CV system to fully exploit spatial topological and non-topological information components in addition to color is to perform nearly as well when input with panchromatic or color imagery.

Deeply investigated in CV (Sonka et al. 1994; Frintrop 2011), content-based image retrieval (Smeulders et al. 2000) and RS applications (Baraldi and Boschetti 2012a, 2012b; Nagao and Matsuyama 1980; Matsuyama and Hwang 1990; Shackelford and Davis 2003a, 2003b), popular visual features are: (i) color (Griffin 2006; Gevers et al. 2012; ii) local shape (Wenwen Li et al. 2013; iii) texture, defined as the perceptual spatial grouping of texture elements known as texels (Julesz 1986; Julesz et al. 1973) or tokens (Marr 1982; iv) inter-object spatial topological relationships, e.g., adjacency, inclusion, etc., and (v) inter-object spatial non-topological relationships, e.g., spatial distance, angle measure, etc. Color is the sole visual property available at the imaging sensor resolution. In other words, pixel-based information is spatial context-independent and purely color-specific. Among the aforementioned visual variables, per-pixel color values are the sole non-spatial numeric variable.

Neglecting the fact that spatial topological and non-topological information typically dominate color information in both the (2D) image-domain and the 4D spatiotemporal scene-domain involved with vision (Matsuyama and Hwang 1990), traditional EO-IUSs adopt a 1D image analysis approach, see Fig. 7-14. In 1D image analysis, a 1D streamline of vector data, either spatial context-sensitive (e.g., window-based or image object-based) or context-insensitive (pixel-based), is processed irrespective of the order of presentation of the input sequence. In practice 1D image analysis is invariant to permutations, such as in orderless encoders (Cimpoi et al. 2014). When vector data are spatial context-sensitive then 1D image analysis ignores spatial topological information. When vector data are pixel-based then 1D image analysis ignores both spatial topological and non-topological information. Prior knowledge-based color naming of a spatial unit  $x$  in the image-domain, where  $x$  is either (0D) point, (1D) line or (2D) polygon defined according to the Open Geospatial Consortium (OGC) nomenclature (OGC 2015), is a special case of 1D image analysis, either pixel-based or image object-based, where spatial topological and/or non-topological information are ignored.

Alternative to 1D image analysis, 2D image analysis/retinotopic/topology-preserving visual feature representation relies on a sparse (distributed) 2D array (2D regular grid) of local spatial filters (Tsotsos 1990), suitable for topology-preserving feature mapping (Martinetz et al. 1994; Fritzke 1997a), see Fig. 7-15. The human brain’s organizing principle is topology-preserving feature mapping (Feldman 2016). In the biological visual system, topology-preserving feature maps are primarily spatial, where activation domains of physically adjacent processing units in the 2D array of convolutional



filters are spatially adjacent regions in the 2D visual field. Provided with a superior degree of biological plausibility in modelling 2D spatial topological and non-topological information, distributed processing systems capable of 2D image analysis/retinotopic/topology-preserving visual feature representation, such as deep convolutional neural networks (DCNNs), outperform 1D image analysis approaches (Cimpoi et al. 2014). This apparently trivial consideration is at odd with a relevant portion of the RS literature, where pixel-based 1D image analysis is mainstream followed by context-sensitive 1D image analysis implemented within the OBIA paradigm (Blaschke et al. 2014).

Since traditional EO-IUSs adopt a 1D image analysis approach where dominant spatial information is neglected in favour of secondary color information, it is useful to turn attention to the multidisciplinary framework of cognitive science to shed light on how humans deal with color information. According to cognitive science, which includes linguistics, the study of languages, humans discretize (fuzzify) ever-varying quantitative (numeric) photometric and spatiotemporal sensations into stable qualitative (categorical, nominal) percepts, eligible for use in symbolic human reasoning based on a convergence-of-evidence approach (Matsuyama and Hwang 1990). In their seminal work, Berlin and Kay proved that 20 human languages, spoken across space and time in the real-world, partition quantitative color sensations collected in the visible portion of the electromagnetic spectrum (see Fig. 7-1) onto the same “universal” dictionary of eleven basic color (BC) names (Berlin and Kay 1969): black, white, gray, red, orange, yellow, green, blue, purple, pink and brown. In a 3D monitor-typical red-green-blue (RGB) cube, BC names are intuitive to think of and easy to visualize. They provide a mutually exclusive and totally exhaustive partition of the RGB cube into RGB polyhedra neither necessarily connected nor convex, see Fig. 7-7 (Griffin 2006; Benavente et al. 2008). Since they are community-agreed upon to be used by members of the community, RGB BC polyhedra are prior knowledge-based, i.e., stereotyped, non-adaptive-to-data (static), general-purpose, application- and data-independent. Multivariate measurement space hyperpolyhedralization is the transformation of a numeric variable into a categorical variable. This is a typical problem in many scientific disciplines, such as inductive VQ in machine learning-from-data (Cherkassky and Mulier 1998) and deductive numeric variable fuzzification into discrete fuzzy sets in fuzzy logic (Zadeh 1965), refer to Chapter 7.1.

To summarize, in perceptual human vision for scene-from-image understanding, color names belonging to a stable (non-adaptive to data, prior knowledge-based) dictionary of color names are, first, physically equivalent to color hyperpolyhedra in a numeric MS color space  $\mathfrak{R}^{\text{MS}}$  and, second, conceptually equivalent to a latent/hypothetical categorical variable of semi-symbolic quality, capable of linking sub-symbolic sensory data in the real world, specifically color values in color space  $\mathfrak{R}^{\text{MS}}$ , to categorical variables of semantic (symbolic) quality in the world model.

In an analytic model of vision based on a convergence-of-evidence approach, the first original contribution of the present Part 1 is to intuitively show how prior knowledge-based color value discretization into a static dictionary of color names affects image classification. Irrespective of their Pearson’s cross-correlation, if any, it is easy to prove that individual sources of visual evidence, such as color, local shape, texture and inter-object spatial relationships are statistically independent (because cross-correlation does not mean causation) (Baraldi and Soares 2017). According to a “naive” hypothesis of conditional independence of features, which holds true for individual sources of visual evidence, such as color, local shape, texture and inter-object spatial relationships, when target classes of observed objects in the real-world scene are  $c = 1, \dots, \text{ObjectClassLegendCardinality}$ , for a given discrete spatial unit  $x$  in the image-domain, either point, line or polygon (OGC 2015), then the well-known “naive” Bayes classification formulation becomes

$$\begin{aligned}
 p(c| \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) &= p(c|F_1, \dots, F_I) = \\
 p(c) \prod_{i=1}^I p(F_i|c) &= \\
 p(c) \bullet p(\text{ColorValue}(x)|c) \bullet p(\text{ShapeValue}(x)|c) \bullet p(\text{TextureValue}(x)|c) \bullet p(\text{SpatialRelationships}(x, \text{Neigh}(x))|c) &\leq \\
 \min\{p(c| \text{ColorValue}(x)), p(c| \text{ShapeValue}(x)), p(c| \text{TextureValue}(x)), p(c| \text{SpatialRelationships}(x, \text{Neigh}(x)))\}, & \\
 c = 1, \dots, \text{ObjectClassLegendCardinality}, & \quad (1-1)
 \end{aligned}$$

where  $\text{ColorValue}(x)$  belongs to a MS measurement space  $\mathfrak{R}^{\text{MS}}$ , i.e.,  $\text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , and  $\text{Neigh}(x)$  is a generic 2D spatial neighborhood of spatial unit  $x$  in the image-domain. Equation (1-1) shows that any convergence-of-evidence approach is more selective than each individual source of evidence, in line with a focus-of-visual attention mechanism (Frintrop 2011). For the sake of simplicity, if priors are ignored because considered equiprobable in a maximum class-conditional likelihood inference approach alternative to a maximum *a posteriori* optimization criterion, then Equation (1-1) becomes



$$\begin{aligned}
 & p(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\
 & p(\text{ColorValue}(x) | c) \cdot p(\text{ShapeValue}(x) | c) \cdot p(\text{TextureValue}(x) | c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c) = \\
 & \left[ \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} p(\text{ColorValue}(x) | \text{ColorName}) p(\text{ColorName} | c) \right] \cdot p(\text{ShapeValue}(x) | c) \cdot \\
 & p(\text{TextureValue}(x) | c) \cdot p(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c), \\
 & c = 1, \dots, \text{ObjectClassLegendCardinality}, \tag{1-2}
 \end{aligned}$$

where color space  $\mathfrak{R}^{\text{MS}}$  is partitioned into hyperpolyhedra, equivalent to a discrete and finite dictionary of static color names, with  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ . To further simplify Equation (1-2), its canonical interpretation based on frequentist statistics can be relaxed by fuzzy logic (Zadeh 1965), so that the logical-AND operator is replaced by a fuzzy-AND (min) operator, inductive class-conditional probability  $p(x | c) \in [0, 1]$ , where  $\sum_{c=1}^{\text{ObjectClassLegendCardinality}} p(x | c) \geq 0$ , is replaced by a deductive membership (compatibility) function  $m(x | c) \in [0, 1]$ , where  $\sum_{c=1}^{\text{ObjectClassLegendCardinality}} m(x | c) \geq 0$ , and color space hyperpolyhedra are considered mutually exclusive and totally exhaustive. If these simplifications are adopted, then Equation (1-2) becomes

$$\begin{aligned}
 & m(c | \text{ColorValue}(x), \text{ShapeValue}(x), \text{TextureValue}(x), \text{SpatialRelationships}(x, \text{Neigh}(x))) \propto \\
 & \min \left\{ \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x) | \text{ColorName}) m(\text{ColorName} | c), m(\text{ShapeValue}(x) | c), \right. \\
 & \left. m(\text{TextureValue}(x) | c), m(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c) \right\} = \\
 & \min \{ m(\text{ColorName}^* | c), m(\text{ShapeValue}(x) | c), m(\text{TextureValue}(x) | c), m(\text{SpatialRelationships}(x, \text{Neigh}(x)) | c) \}, \\
 & c = 1, \dots, \text{ObjectClassLegendCardinality}, \text{ where } \text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}, \text{ such that} \\
 & m(\text{ColorValue}(x) | \text{ColorName}^*) = 1 \text{ and } m(\text{ColorName}^* | c) \in \{0, 1\}. \tag{1-3}
 \end{aligned}$$

In Equation (1-3), the following considerations hold.

- Each numeric  $\text{ColorValue}(x)$  in color space  $\mathfrak{R}^{\text{MS}}$  belongs to a single color name (hyperpolyhedron)  $\text{ColorName}^*$  in the static color dictionary, i.e.,  $\forall \text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}, \sum_{\text{ColorName}=1}^{\text{ColorDictionaryCardinality}} m(\text{ColorValue}(x) | \text{ColorName}) = m(\text{ColorValue}(x) | \text{ColorName}^*) = 1$  holds, where  $m(\text{ColorValue}(x) | \text{ColorName}) \in \{0, 1\}$ ,  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ .
- The set  $A = \text{DictionaryOfColorNames}$ , with cardinality  $|A| = a = \text{ColorDictionaryCardinality}$ , and the set  $B = \text{LegendOfObjectClassNames}$ , with cardinality  $|B| = b = \text{ObjectClassLegendCardinality}$ , can be considered a bivariate categorical random variable where two univariate categorical variables  $A$  and  $B$  are generated from a single population. A binary relationship (product set) from set  $A$  to set  $B$ ,  $R: A \Rightarrow B$ , is a subset of the 2-fold Cartesian product  $A \times B$ , whose size is rows  $\times$  columns =  $a \times b$ . The Cartesian product of two sets  $A \times B$  is a set whose elements are ordered pairs. Hence, the Cartesian product is non-commutative,  $A \times B \neq B \times A$ . In agreement with common sense, see Table 7-3,  $R: \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames}$  is a set of ordered pairs where each  $\text{ColorName}$  can be assigned to none, one or several classes  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  of observed scene-objects, whereas each class of observed objects can be assigned with none, one or several color names as the class-specific color attribute. Binary membership values  $m(\text{ColorName} | c) \in \{0, 1\}$  and  $m(c | \text{ColorName}) \in \{0, 1\}$ , with  $c = 1, \dots, \text{ObjectClassLegendCardinality}$  and  $\text{ColorName} = 1, \dots, \text{ColorDictionaryCardinality}$ , can be community-agreed upon based on various kinds of evidence, whether viewed all at once or over time, such as a combination of prior beliefs with additional evidence inferred from new data in agreement with a Bayesian updating rule (Bayesian inference), largely applied in artificial intelligence and expert systems. A binary relationship  $R: A \Rightarrow B \subseteq A \times B$  where sets  $A$  and  $B$  are categorical variables generated from a single population guides the interpretation process of a two-way *contingency table* (also known as association matrix, cross tabulation, bivariate table or frequency table) (Kuzera and Pontius 2008; Pontius and Connors 2006),  $\text{BIVRTAB} = \text{FrequencyCount}(A \times B)$ . In the conventional domain of frequentist inference with no reference to prior beliefs, a BIVRTAB is the 2-fold Cartesian product  $A \times B$  instantiated by the bivariate frequency counts of the two univariate categorical variables  $A$  and  $B$  generated from a single population. For any BIVRTAB instance, either square or non-square, there is a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  that guides the interpretation process, where "correct" entry-pair cells of the 2-fold Cartesian product  $A \times B$  can be either off-diagonal (scattered) or on-diagonal, if a main diagonal exists. When a BIVRTAB is estimated from a



geospatial population without sampling, it is called *overlapping area matrix* (OAMTRX) (Baraldi et al. 2014; Baraldi et al. 2006; Lunetta and Elvidge 1999; Beauchemin and Thomson 1997; Ortiz and Oliver 2006; Baraldi et al. 2005; Pontius and Connors 2006). When the binary relationship  $R: A \Rightarrow B$  is a bijective function (both 1-1 and onto), i.e., when the two categorical variables  $A$  and  $B$  estimated from a single population coincide, then the BIVRTAB is square and sorted and typically called confusion matrix (CMTRX) or error matrix (Stehman and Czaplewski 1998; Congalton and Green 1999; Lunetta and Elvidge 1999). In a CMTRX the main diagonal guides the interpretation process. For example, a square OAMTRX = FrequencyCount( $A \times B$ ), where  $A$  = test thematic map legend,  $B$  = reference thematic map legend and cardinality  $a = b$ , is a CMTRX if and only if  $A = B$ , i.e., if the test and reference codebooks are the same sorted set of semantic concepts. In general the class of (square and sorted) CMTRX instances is a special case of the class of OAMTRX instances, either square or non-square, i.e.,  $OAMTRX \supset CMTRX$ . A similar consideration holds about summary  $Q^2$ s generated from an OAMTRX or a CMTRX, i.e.,  $Q^2I(OAMTRX) \supset Q^2I(CMTRX)$ .

Equation (1-3) shows that for any spatial unit  $x$  in the image-domain, when a hierarchical CV classification approach estimates posterior  $m(c | ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x, Neigh(x)))$  starting from a near real-time context-insensitive color naming first stage where condition  $m(ColorValue(x) | ColorName^*) = 1$  holds, if condition  $m(ColorName^* | c) = 0$  is true according to a static community-agreed binary relationship  $R: DictionaryOfColorNames \Rightarrow LegendOfObjectClassNames$  (and vice versa) known *a priori*, see Table 7-3, then  $m(c | ColorValue(x), ShapeValue(x), TextureValue(x), SpatialRelationships(x, Neigh(x))) = 0$  irrespective of any second-stage assessment of spatial terms  $ShapeValue(x)$ ,  $TextureValue(x)$  and  $SpatialRelationships(x, Neigh(x))$ , whose computational model is typically difficult to find and computationally expensive. Intuitively Equation (1-3) shows that static color naming allows the stratification of unconditional multivariate spatial variables into color class-conditional data distributions, in agreement with the statistic stratification principle (Hunt and Tyrrell 2012) and the divide-and-conquer problem solving approach (Bishop 1995; Cherkassky and Mulier 1998). Well known in statistics, the principle of statistic stratification guarantees that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” (Hunt and Tyrrell 2012).

Whereas color polyhedra are easy to visualize and intuitive to think of in a true- or false-color RGB cube, see Fig. 7-7, hyperpolyhedra are difficult to think of and impossible to visualize in a MS reflectance space whose dimensionality is superior to three, with spectral channels ranging from visible to thermal portions of the electromagnetic spectrum, see Fig. 7-1. Since it is non-adaptive-to-data, any static hyperpolyhedralization of a MS measurement space must be based on *a priori* physical knowledge available in addition to sensory data. Since it relies on a physical data model, static hyperpolyhedralization of a MS data space requires all spectral channels to be provided with a physical unit of radiometric measure, i.e., MS data must be radiometrically calibrated in compliance with the QA4EO *Cal* requirements, refer to Chapter 7.1.

Noteworthy, sensory data provided with a physical unit of measure can be input to statistical and physical models, including hybrid inference systems, refer to Chapter 7.1. On the contrary, uncalibrated dimensionless sensory data can be input to statistical data models exclusively.

Although considered mandatory by the QA4EO guidelines and regarded as a well-known “prerequisite for physical model-based analysis of airborne and satellite sensor measurements in the optical domain” (Schaeppman-Strub et al. 2006, 29), EO data *Cal* is ignored by a relevant portion of the existing RS literature (Baraldi 2009). One consequence is that, to date, statistical model-based EO-IUSs dominate the RS literature and commercial EO image processing software toolboxes, which typically consist of overly complicated collections of inherently ill-posed inductive machine learning-from-data algorithms (Cherkassky and Mulier 1994) to choose from based on heuristics (Baraldi and Boschetti 2012a, 2012b).

### 7.3 Related works in static MS reflectance space hyperpolyhedralization

In the RS discipline there is a long history of hybrid EO-IUSs in operating mode, suitable for either low-level EO image enhancement or high-level EO image classification, where an *a priori* knowledge-based decision tree for static MS reflectance space hyperpolyhedralization is plugged-in without validation, refer to Chapter 7.1. For example, the SIAM stratification of single-date MS imagery contributed to make MS image topographic correction, which is a traditional chicken-and-egg dilemma, better conditioned for automated solution (Baraldi et al. 2010c), in compliance with the ESA EO Level 2 product requirements (ESA 2015; CNES 2016), see Fig. 7-12. In the Atmospheric/Topographic Correction for



Satellite Imagery (ATCOR) commercial software product, several deductive decision trees are implemented for use in different stages of an EO data enhancement pipeline (Richter and D. Schlöpfer 2012a; Richter and D. Schlöpfer 2012b; Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Dorigo et al. 2009; Richter et al. 2009), see Fig. 7-11. One of the ATCOR's prior knowledge-based per-pixel decision trees delivers as output a haze/ cloud/ water (and snow) classification mask file ("image\_out\_hcw.bsq"). In addition ATCOR includes a so-called prior knowledge-based Spectral Classification of surface reflectance signatures (SPECL) (Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Dorigo et al. 2009), see Table 7-4. Unfortunately, SPECL has never been tested by its authors in the RS literature, although it has been validated by independent means (Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015). Supported by the National Aeronautics and Space Administration (NASA), atmospheric effect removal by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) project relies on exclusion masks for water, cloud, shadow and snow surface types detected by a simple set of prior knowledge-based spectral decision rules. Unfortunately, quantitative analyses of LEDAPS products revealed that these exclusion masks are prone to errors, to be corrected in future LEDAPS releases (Vermote and Saleous 2007). In the 1980s, to provide an automatic alternative to a visual and subjective assessment of the cloud cover on Advanced Very High Resolution Radiometer (AVHRR) quicklook images in the European Space Agency (ESA) Earthnet archive, Muirhead and Malkawi developed a simple algorithm to classify daylight AVHRR images on a pixel-by-pixel basis into land, cloud, sea, snow or ice and sunglint, such that the classified quicklook image was presented in appropriate pseudocolors, e.g., green: land, blue: sea, white: cloud, etc. (Muirhead and Malkawi 1989). Developed independently by NASA (Ackerman et al. 1998) and the Canadian Center for Remote Sensing (Luo et al. 2008), pixel-based static decision trees contribute, to date, to the generation of clear-sky Moderate Resolution Imaging Spectroradiometer (MODIS) composites, see Fig. 7-16. To pursue high-level LC/LCC detection through time, extensions to the time domain of a single-date *a priori* spectral rule base for MS reflectance space quantization have become available to the general public in 2015 through the Google Earth Engine (GEE) platform (Simonetti et al. 2015b) or in the form of a commercial LC/LCC map product at national scale (GeoTerraImage 2015). These two multi-temporal spectral decision trees for LC/LCC detection share the same operational limitations, specifically, they are Landsat sensor series-specific and pixel-based. They were both inspired by a year 2006 SIAM instantiation of a static decision tree for Landsat reflectance space hyperpolyhedralization, presented in pseudo-code in the RS literature (Baraldi et al. 2006) and further developed into the SIAM application software available to date (Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Baraldi and Boschetti 2012a, 2012b; Baraldi et al. 2014; Baraldi et al. 2010a, 2010b; Baraldi 2011; Baraldi et al. 2010d). In Boschetti et al. (2015), a year 2013 SIAM application was successfully employed to accomplish burned area detection in MS image time-series. Among the aforementioned static decision trees for MS color naming, only SIAM claims scalability to several families of EO imaging sensors featuring different spectral resolutions, see Table 7-1.

It is obvious but not trivial to emphasize to the RS community that, in human vision and CV, an *a priori* dictionary of general-purpose data- and application-independent color names is equivalent to a static sub-symbolic categorical variable non-coincident with a symbolic categorical variable whose levels are user- and application-specific classes of objects observed in the 4D spatiotemporal scene-domain, refer to Table 7-3 and Equation (1-3). The very same consideration holds for any discrete and finite set of spectral endmembers in mixed pixel analysis, which "cannot always be inverted to unique LC class names" (Adams et al. 1995, 147).

Quite surprisingly, the non-coincident assumption between an *a priori* dictionary of sub-symbolic color names and a scene- and application-specific legend of symbolic classes of real-world objects appears somehow difficult to acknowledge by relevant portions of the RS community. For example, in the DigitalGlobe Geospatial Big Data platform (GBDX), a patented prior knowledge-based decision tree for pixel-based very high resolution WorldView-2 and WorldView-3 image mapping onto static MS reflectance space hyperpolyhedra (GBDX Registered Name: protogenV2LULC, Provider: GBDX) was proposed to RS end-users as an "Automated Land Cover Classification" (DigitalGlobe 2016). This program name can be considered somehow misleading because assigned to a static sub-symbolic color space partitioner, see Equation (1-3). In fact, so-called "known issues" linked to the "Automated Land Cover Classification" computer program included: "Vegetation: Thin cloud (cloud edges) might be misinterpreted as vegetation; Water: False positives maybe present due to certain types of concrete roofs or shadows; Soils: Ceramic roofing material and some types of asphalt may be misinterpreted as soil," etc. (DigitalGlobe 2016). In Salmon et al. (2013), a year 2006 SIAM's *a priori* dictionary of static sub-symbolic MS color names was downscaled in cardinality and sorted in the order of presentation to form a bijective function with a legend of symbolic classes of target objects in the scene-domain. In practice these authors forced a non-square BIVRTAB to become a CMTRX, where the main diagonal guides the interpretation process (Congalton and Green 1999), to make it more intuitive and familiar to RS practitioners. In general, no binary relationship  $R: A \Rightarrow B$  between an *a priori* dictionary



A of static sub-symbolic color names and a user- and application-dependent dictionary B of symbolic classes of observed objects in the scene-domain is a bijective function, refer to Table 7-3 and Equation (1-3). As a consequence of its unrealistic hypothesis in color information/knowledge representation, the 1D image classification approach proposed in (Salmon et al. 2013) scored low in accuracy. Unfortunately, to explain their poor MS image classification outcome these authors concluded that, in their experiments, a year 2006 SIAM's static dictionary of color names was useless to identify target LC classes. The lesson to be gained by these authors' experience is that well-established RS practices, such as 1D image analysis based on supervised data learning algorithms and thematic map quality assessment by means of a CMTRX where test and reference thematic legends are the same, can become malpractices when an *a priori* dictionary of static color names is employed for MS image classification purposes in agreement with Equation (1-3) and common sense, see Table 7-3. This lesson learned is supported by the fact that one of the same co-authors of paper (Salmon et al. 2013) reached opposite conclusions when a year 2013 SIAM application software, the same investigated by the present paper, was employed successfully in detecting burned areas from MS image time-series according to a convergence of color names with spatiotemporal visual properties in agreement with Equation (1-3) (Boschetti et al. 2015).

#### 7.4 Original hybrid eight-step guideline for identification of a categorical variable-pair relationship

Our experimental project required to compare an annual time-series of test SIAM-WELD maps of sub-symbolic color names, see Fig. 7-6, with a reference NLCD 2006 map whose legend of symbolic LC classes is summarized in Table 7-2. Since these test and reference map legends do not coincide, they must be reconciled through a binary relationship R: DictionaryOfColorNames  $\Rightarrow$  LegendOfObjectClassNames (and vice versa), refer to Equation (1-3).

The harmonization of ontologies and the comparison of thematic maps with different legends are the subject of research of a minor body of literature, e.g., refer to works in ontology-driven geographic information systems (ODGIS) (Fonseca et al. 2002; Guarino 1995; Sowa 2000). Ahlqvist writes that “to negotiate and compare information stemming from different classification systems (Bishr 1998; Mizen et al. 2005)... a translation can be achieved by *matching the concepts in one system with concepts in another*, either directly or through an intermediate classification (Feng and Flewelling 2004; Kavouras and Kokla 2002)” (Ahlqvist 2005). Stehman describes four common types of map-pair comparisons (Stehman 1999). In the first type, different thematic maps, either crisp or fuzzy, of the same region of interest and employing the same sorted set (legend) of LC classes are compared (Kuzera and Pontius 2008). In the second type, which includes the first type as a special case, thematic maps, either crisp or fuzzy, of the same region of interest, but featuring map legends that differ in their basic terms with regard to semantics and/or cardinality and/or order of presentation are compared. The third and fourth types of thematic maps comparison regard maps of different surface areas featuring, respectively, the same dictionary or different dictionaries of basic terms. Whereas a large portion of the RS community appears concerned with the aforementioned first type of map comparisons exclusively, the protocol proposed in (Baraldi et al. 2014) is focused on the second type, which includes the first type as a special case. In (Couclelis 2010) Couclelis observed that inter-dictionary concept matching (“conceptual matching” (Ahlqvist 2005)) is an inherently equivocal *information-as-data-interpretation* process (Capurro and Hjørland 2003), see Table 7-3. In common practice two independent human domain-experts (cognitive agents, knowledge engineers [Laurini and Thompson 1992]) are likely to identify different binary associations between two codebooks of codewords. The conclusion is that no “universal best match” of two different codebooks can exist, but identification of the most appropriate binary relationship between two different nomenclatures becomes a subjective matter of negotiation to become community-agreed upon (Couclelis 2010; Capurro and Hjørland 2003).

To streamline the inherently subjective selection of “correct” entry-pairs in a binary relationship R:  $A \Rightarrow B \subseteq A \times B$  between two univariate categorical variables A and B of a single population, an original hybrid 8-step guideline was designed for best practice, where deductive/top-down prior beliefs and inductive/bottom-up learning-from-data inference are combined. This hybrid protocol is sketched hereafter as the second original and pragmatic contribution of the present Part 1 to fill the gap from EO sensory data to ESA EO Level 2 product. As an example, let us consider a binary relationship R:  $A \Rightarrow B = \text{DictionaryOfColorNames} \Rightarrow \text{LegendOfObjectClassNames} \subseteq A \times B$  where rows are a test set of three semi-symbolic color names, say,  $A = \{\text{MS green-as-“Vegetation”}, \text{MS white-as-“Cloud”}, \text{“Unknowns”}\}$ , where  $|A| = a = \text{ColorDictionaryCardinality} = TC = 3$  is the row (test) cardinality, and where columns are a reference set of three symbolic LC classes, say,  $B = \{\text{“Evergreen Forest”}, \text{“Deciduous Forest”}, \text{“Others”}\}$ , where  $|B| = b = \text{ObjectClassLegendCardinality} = RC = 3$  is the column (reference) cardinality.





1. Display multivariate frequency distributions of the two univariate categorical variables estimated from a single population in the  $BIVRTAB = \text{FrequencyCount}(A \times B)$  whose size is  $TC \times RC$ .
2. Estimate probabilities in the  $BIVRTAB$  cells.
3. Compute class-conditional probability  $p(r | t)$  of reference class  $r = 1, \dots, RC$ , given test class  $t = 1, \dots, TC$ .
4. Reset to zero all  $p(r | t)$  below  $TH1 \in [0, 1]$  (e.g.,  $TH1 = 9\%$ ), otherwise set that cell to 1. Let us identify this contingency table instantiation as  $DataDrivenConditionalProb(r|t)(x, y)$ ,  $x = 1, \dots, RC$ ,  $y = 1, \dots, TC$ .
5. Compute class-conditional probability  $p(t | r)$  of test class  $t = 1, \dots, TC$ , given reference class  $r = 1, \dots, RC$ .
6. Reset to zero all  $p(t | r)$  below  $TH2 \in [0, 1]$  (e.g.,  $TH2 = 6\% \leq TH1$ ), otherwise set that cell to 1. Let us identify this contingency table instantiation as  $DataDrivenConditionalProb(t|r)(x, y)$ ,  $x = 1, \dots, RC$ ,  $y = 1, \dots, TC$ .
7. Compute  $DataDrivenTemporaryCells(x, y) = \max\{DataDrivenConditionalProb(t|r)(x, y), DataDrivenConditionalProb(r|t)(x, y)\}$ ,  $x = 1, \dots, RC$ ,  $y = 1, \dots, TC$ . At this point, based exclusively on bottom-up evidence stemming from frequency data, in the 2-fold Cartesian product  $A \times B$  each cell is equal to 0 or 1. Then that cell is termed either “temporary non-correct” or “temporary correct”.
8. Top-down scrutiny by a human domain-expert of each cell in the  $BIVRTAB$ , which is either “temporary correct” or “temporary non-correct” at this point, to select those cells to be finally considered as “correct entry-pairs”. Actions undertaken by this top-down scrutiny are twofold.
  - Switch any data-derived “temporary correct” cell to a “final non-correct” cell if it is provided with a strong prior belief of conceptual mismatch. For example, based on experimental evidence a test spectral category MS white-as-“Cloud” can match a reference LC class “Evergreen Forest”: this data-derived entry-pair match must be considered non-correct in the final  $R: A \Rightarrow B$  following semantic scrutiny by a human expert.
  - Switch any data-derived “temporary non-correct” cell to a “final correct” cell if it is provided with a strong prior belief of conceptual match. For example, the test spectral category MS green-as-“Vegetation” is considered a superset of the reference LC class “Deciduous Forest” irrespective of whether there are frequency data in support of this conceptual relationship.

Table 7-5 shows an example of how this 8-step protocol can be employed in practice. In Table 7-5 the last step 8 identifies an inherently equivocal *information-as-data-interpretation* process, where a human decision maker has a pro-active role in providing frequency data with semantics (meanings) (Capurro and Hjørland 2003). It is highly recommended that any inherently subjective *information-as-data-interpretation* activity occurs as late as possible in the information processing workflow, to avoid propagation of “errors” due to personal preferences not yet community-agreed upon. Noteworthy, in the proposed eight-step guideline there are two “hidden” free-parameters to be user-defined based on heuristics (trial-and-error strategy), the binary thresholds  $TH1$  and  $TH2$ , whose normalized range of change and intuitive meaning in terms of probability should make their selection easy and, to a certain extent, application- and user-independent.

### 7.5 Original measure of association in a categorical variable-pair relationship

Traditional scalar indicators of bivariate categorical variable association estimated from a  $BIVRTAB = \text{FrequencyCount}(A \times B)$ , either square or non-square, include the Pearson’s chi-square index of statistical independence and the normalized Pearson’s chi-square index, also known as Cramer’s coefficient V (Sheskin 2000). These frequentist statistics do not apply to a binary relationship  $R: A \Rightarrow B$  such as that shown in Table 7-3. Hereafter, a scalar indicator of association estimated from a binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , where A and B are two univariate categorical variables of a single population, is called Categorical Variable Pair Association Index (CVPAI) in range  $[0, 1]$ .

Proposed in (Baraldi et al. 2014), a CVPAI version 1 (CVPAI1)  $\in [0, 1]$  is maximized (tends to 1) if the binary relationship  $R: A \Rightarrow B$  from set  $A =$  test categorical variable, e.g., dictionary of color names, to set  $B =$  reference categorical variable, e.g., dictionary of land cover (LC) class names, is a bijective function, both injective (one-to-one) and surjective (onto). Original CVPAI version 2 (CVPAI2) and version 3 (CVPAI3) formulations are proposed hereafter as the third original and analytic contribution of the present Part 1. Unlike the CVPAI1, a novel CVPAI2 formulation was constrained as follows, see Fig. 7-17. (i) The “most discriminative” test-to-reference class relation  $R: A \Rightarrow B$  is a function, i.e., each test color name matches with only one reference LC class name. (ii) The “most discriminative” reference-to-test class relation is either a surjective function, i.e., each reference LC class matches with at least one test color name, or a bijective function, both surjective and injective as a special case of the former, i.e., each reference LC class matches with only one test color name, see Fig. 7-17. In short, CVPAI2 is maximum when the binary relationship  $R: A \Rightarrow B$  is either a surjective



function or a bijective function. Although it is maximized by the same type of distribution of “correct” entry-pair cells in a binary relationship  $R: A \Rightarrow B$ ,  $CVPAI3$  is a more severe formulation of  $CVPAI2$ , i.e.,  $CVPAI2 \geq CVPAI3 \in [0, 1]$ . Since the  $CVPAI2$  design relaxes the  $CVPAI1$  formulation, it is always true that  $CVPAI2 \geq CVPAI1 \in [0, 1]$ . Whereas the  $CVPAI1$  and  $CVPAI2/CVPAI3$  are maximized by different distributions of “correct” entry-pair cells in a binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , they are all independent of frequency counts generated by a bivariate categorical variable distribution to be displayed in a  $BIVRTAB = \text{FrequencyCount}(A \times B)$ .

To appreciate the conceptual difference between the  $CVPAI1$  and  $CVPAI2$  designs, see Fig. 7-17, let us compare a test dictionary  $A$  of color names, such as SIAM’s, see Fig. 7-6, with a reference dictionary  $B$  of LC class names, such as NLCD’s, see Table 7-2. In terms of capability of color names to discriminate LC class names, the ideal test-to-reference class relation is 1-1, where one color name matches with only one reference LC class. On the other hand, the color attribute of a real-world LC class can be typically linked to one or more discrete color names, see Table 7-3. In this realistic example the expected  $CVPAI2$  value would belong to range  $(0, 1]$ , while the  $CVPAI1$  formulation proposed in (Baraldi et al. 2014) scores below its maximum, i.e.,  $CVPAI1 \in (0, 1)$ .

Another example where the difference between the  $CVPAI1$  and  $CVPAI2$  formulations is highlighted is when the test dictionary  $A$  is a specialized version of the reference dictionary  $B$ . For example, a test taxonomy of LC classes is  $A = \text{LegendOfObjectClassNamesA} = \{\text{LC class “Dark-tone bare soil”}, \text{LC class “Light-tone bare soil”}, \text{LC class “Deciduous Forest”}, \text{LC class “Evergreen Forest”}\}$  and a reference LC class taxonomy is  $B = \text{LegendOfObjectClassNamesB} = \{\text{LC class “Bare soil”}, \text{LC class “Forest”}\}$ . Based on our prior knowledge-based understanding of these two semantic dictionaries  $A$  and  $B$ , a reasonable binary relationship can be considered  $R: A \Rightarrow B = \{(\text{LC class “Dark-tone bare soil”}, \text{LC class “Bare soil”}); (\text{LC class “Light-tone bare soil”}, \text{LC class “Bare soil”}); (\text{LC class “Deciduous Forest”}, \text{LC class “Forest”}); (\text{LC class “Evergreen Forest”}, \text{LC class “Forest”})\}$ . In this case, the  $CVPAI1$  formulation scores below its maximum, i.e.,  $CVPAI1 \in (0, 1)$ , while the expected  $CVPAI2$  value would score maximum, i.e.,  $CVPAI2 = 1$ .

These two examples illustrate the intuitive meaning and practical use of the normalized quantitative indicator  $CVPAI2 \in [0, 1]$  in a hierarchical EO-IUS based on a convergence-of-evidence approach in agreement with Equation (1-3). When the semantic information gap from sub-symbolic sensory data to a symbolic set  $B = \text{LegendOfObjectClassNames}$  is filled by an EO-IUS starting from a static color naming first stage, provided with a semi-symbolic set  $A = \text{DictionaryOfColorNames}$ , if the binary relationship  $R: A \Rightarrow B$  features a degree of association  $CVPAI2 \in [0, 1]$ , then  $(1 - CVPAI2) \in [0, 1]$  is the semantic information gap from sub-symbolic sensory data to the symbolic  $\text{LegendOfObjectClassNames}$  left to be filled by further stages in the EO-IUS pipeline. If  $CVPAI2 = 1$ , then secondary color information discretized by set  $A = \text{DictionaryOfColorNames}$  suffices to detect the target set  $B = \text{LegendOfObjectClassNames}$  with no further need to investigate primary spatial information in a convergence-of-evidence image classification approach.

The analytic formulation of the  $CVPAI2$  is proposed as follows. In a binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , set  $A$  is a test codebook of cardinality  $|A| = TC$  as rows and set  $B$  is a reference codebook of cardinality  $|B| = RC$  as columns, so that the size of the 2-fold Cartesian product  $A \times B$  is  $TC \times RC$ . The total number of “correct” entry-pair cells in  $R: A \Rightarrow B$  is identified as  $CE$ , where  $0 \leq CE \leq TC \times RC$ . In addition, symbol ‘ $\equiv$ ’ is adopted to mean ‘equal to’. The  $CVPAI2$  formulation is constrained as follows.

- $$CE = \sum_{t=1}^{TC} \sum_{r=1}^{RC} CE_{t,r}, \text{ with } CE_{t,r} \in \{0,1\} =$$
- (a)  $\{ \text{"non - correct" entry - pair } (t, r) = 0, \text{ "correct" entry - pair } (t, r) = 1 \},$   
 $CE \in \{0, RC \times TC\}.$
  - (b) If  $(CE \equiv 0)$  then  $CVPAI2 = 0$  must hold. It means that, when no “correct” cell exists, then the degree of conceptual match between the two categorical variables is zero.
  - (c) If  $(CE \equiv TC \times RC)$  then  $CVPAI2 \rightarrow 0$  must hold. It means that when all table cells are considered “correct”, then no entry-pair is discriminative (informative), i.e., nothing makes the difference between the two categorical variables.
  - (d) If

$$\left\{ \left[ \left[ \begin{array}{l} TC \\ \sum_{t=1}^{TC} CE_{t,r} = CE_{+,r} \end{array} \right] > 0, r = 1, \dots, RC \right] \text{AND} \left[ \begin{array}{l} RC \\ \sum_{r=1}^{RC} CE_{t,r} = CE_{t,+} \end{array} \right] = 1, t = 1, \dots, TC \right\},$$

i.e., for each reference class  $r = 1, \dots, RC$  there is at least one match while each test class  $t = 1, \dots, TC$  features one single match, then  $CVPAI2$  must be maximum, i.e.,  $CVPAI2 = 1$ .

(e) If [not condition(b) AND not condition(c) AND not condition(d)] then  $CVPAI2 \in (0,1)$ .

In a square binary relationship  $R: A \Rightarrow B$  where  $TC = RC$ , to maximize the  $CVPAI2$  (to become equal to 1), submitted to condition (d), the binary relationship must be a 1-1 function (an injective function, 1-1 forward and 1-1 backward). To satisfy the set of aforementioned constraints (a) to (e), the following set of original equations is proposed.

$$CVPAI2 \in [0,1], CVPAI2 = \frac{1}{RC + TC} \left( \sum_{r=1}^{RC} f_{RC}(CE_{+,r}) + \sum_{t=1}^{TC} f_{TC}(CE_{t,+}) \right), \quad (1-4)$$

with

$$f_{RC}(i) = \begin{cases} 0 & \text{if } i=0, \\ 1 & \text{if } i>0, \end{cases} \quad i \in \{0, TC\} \subset I_0^+, \text{ where } i = CE_{+,r}, r \in \{1, RC\}, \quad (1-5)$$

$$f_{TC}(j) = \begin{cases} 0 & \text{if } j = 0, \\ \text{GaussianMembership}(j, \text{Center} = 1, \text{StnDev} = RC/3) = \\ -\frac{1(j-1)^2}{2\left(\frac{RC}{3}\right)^2} & \\ e & \in [0, 1], \quad \text{if } j > 0, \text{ with } j \in \{0, RC\} \subset I_0^+, \text{ where } j = CE_{t,+}, t \in \{1, TC\}. \end{cases} \quad (1-6)$$

A more severe formulation of  $CVPAI2$  is the following  $CVPAI3$ , such that  $0 \leq CVPAI3 \leq CVPAI2 \leq 1$ .

$$CVPAI3 \in [0,1], CVPAI3 = \min \left\{ \frac{\sum_{r=1}^{RC} f_{RC}(CE_{+,r})}{RC}, \frac{\sum_{t=1}^{TC} f_{TC}(CE_{t,+})}{TC} \right\}. \quad (1-7)$$

In Equation (1-6)  $\text{GaussianMembership}(j, \text{Center} = 1, \text{StnDev} = RC/3) \rightarrow 0$  when it covers approximately 99.73% of the area underneath the curve at distance  $j \approx \pm 3 \cdot \text{StnDev} = \pm 3 \cdot RC / 3 = \pm RC$  from its Center = 1. If  $j = 1$ , then  $\text{GaussianMembership}(j, \text{Center} = 1, \text{StnDev} = RC/3) = 1$ . It is trivial to prove that Equation (1-4) to Equation (1-6) satisfy the aforementioned requirements (a) to (d). By means of a numeric example, it can be shown that requirement € is satisfied too. For example, estimated from the binary relationship instantiated at step 8 of Table 7-5,  $CVPAI2 = (1/6) * (1 + 1 + 1 + 1 + 1 + \exp(-0.5 * (3-1)^2 / (3/3)^2)) = (1/6) * (5 + 0.1353) = 0.8558$ . Intuitively closer to 1 than 0 this  $CVPAI2$  value shows that the harmonization between the two test and reference nominal variables is (fuzzy) “high” ( $> 0.8$ ) in Table 7-5.



## 7.6 Conclusions

To pursue a GEOSS mission not-yet accomplished by the RS community, this interdisciplinary work aimed at filling an analytic and pragmatic information gap from EO big sensory data to ESA EO Level 2 product. For the sake of readability this paper is split into two, the present Part 1 – Theory and the following Part - Validation.

The original contribution of the present Part 1 is fourfold. A first lesson was learned from published works on prior knowledge-based MS reflectance space hyperpolyhedralization into static (non-adaptive-to-data) color names. It was observed that well-established RS practices, such as 1D image analysis based on supervised data learning algorithms, where dominant spatial information is neglected in favor of secondary color information, and thematic map quality assessment where test and reference map legends are required to coincide, can become malpractices when an *a priori* dictionary of static color names is employed for MS image classification purposes in agreement with Equation (1-3). When test and reference thematic map legends A and B are the same, the binary relationship  $R: A \Rightarrow B \subseteq A \times B$  becomes a bijective function (both 1-1 and onto) and the main diagonal of the 2-fold Cartesian product  $A \times B$  guides the interpretation process of a BIVRTAB = FrequencyCount( $A \times B$ ) = CMTRX. This constraint makes a CMTRX intuitive to understand and more familiar to RS practitioners. Quite surprisingly, the non-coincident assumption between an *a priori* dictionary A of static sub-symbolic color names and a scene- and application-specific legend B of symbolic classes of real-world objects appears somehow difficult to acknowledge by relevant portions of the RS community, in contrast with common sense, see Table 7-3.

Second, Equation (1-3) was proposed as an analytic expression of a biologically plausible hybrid (combine deductive and inductive) CV system suitable for convergence of color and spatial evidence, in agreement with the statistic stratification principle and the divide-and-conquer problem solving approach. In compliance with common sense (see Table 7-3), Equation (1-3) shows that a static color naming first stage can be employed for stratification purposes of further spatial-context sensitive image classification stages. In the static color naming first stage, a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from a set A of general-purpose static color names to a set B of user- and application-specific LC class names can be established by human experts based on top-down prior beliefs, if any, in combination with bottom-up evidence inferred from new data.

Third, for best practice an eight-step protocol was streamlined to infer a categorical variable-pair relationship  $R: A \Rightarrow B$  from categorical variable A to categorical variable B as a hybrid combination of deductive prior beliefs with inductive evidence from data.

Fourth, an original CVPAl2 formulation was proposed as a categorical variable-pair degree of association in a binary relationship  $R: A \Rightarrow B$ .

To comply with the QA4EO *Call/Val* requirements, the subsequent Part 2 of this paper presents and discusses a Stage 4 *Val* of the SIAM-WELD annual map time-series in comparison with the reference NLCD 2006 map, based on an original protocol for wall-to-wall inter-map comparison without sampling where the test and reference maps feature the same spatial resolution and spatial extent, but whose legends are not the same and must be harmonized.

## Acknowledgments

To accomplish this work Andrea Baraldi was supported in part by the National Aeronautics and Space Administration (NASA) under Grant No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. Dirk Tiede was supported in part by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel2/3, ID 848009. Prof. Ralph Maughan, Idaho State University, is kindly acknowledged for his contribution as active conservationist and for his willingness to share his invaluable photo archive with the scientific community as well as the general public. Andrea Baraldi thanks Prof. Raphael Capurro, Hochschule der Medien, Germany, and Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, for their support. Above all, the authors acknowledge the fundamental contribution of Prof. Luigi Boschetti, currently at the Department of Forest, Rangeland and Fire Sciences, University of Idaho, Moscow, Idaho, who conducted by independent means all experiments whose results are proposed in this validation paper. The authors also wish to thank the Editor-in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.

## Disclosure statement

In accordance with XXX policy and his ethical obligation as a researcher, Andrea Baraldi reports he is the sole developer





and IPR owner of the Satellite Image Automatic Mapper™ (non-registered trademark) computer program licensed to academia, public institutions and private companies, eventually free-of-charge, by the one-man-company Baraldi Consultancy in Remote Sensing that may be affected by the research reported in the enclosed paper. Andrea Baraldi has disclosed those interests fully to XXX, and he has in place an approved plan for managing any potential conflicts arising from that involvement.

## References in Chapter 7

- Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. 2011. "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Machine Intell.* 6(1): 1-8.
- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley. 1998. "Discriminating clear sky from clouds with MODIS", *J. Geophys. Res.* 103(32): 141-157.
- Adams, J. B., E. S. Donald, V. Kapos, R. Almeida Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie. 1995. "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon." *Remote Sens. Environ.* 52: 137-154.
- Ahlqvist, O. 2005. "Using uncertain conceptual spaces to translate between land cover categories." *Int. J. Geographic. Info. Science* 19: 831–857.
- Ahlqvist, O. 2008. "In search of classification that supports the dynamics of science: the FAO Land Cover Classification System and proposed modifications." *Environment and Planning B: Planning and Design* 35: 169-186.
- Arvor, D., B. D. Madiela, and T. Corpetti. 2016. "Semantic pre-classification of vegetation gradient based on linearly unmixed Landsat time series." In *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International* 4422-4425.
- Baraldi, A. 2009. "Impact of radiometric calibration and specifications of spaceborne optical imaging sensors on the development of operational automatic remote sensing image understanding systems." *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 2(2): 104-134.
- Baraldi, A. 2011. "Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images." *IEEE Trans. Geosci. Remote Sens.* 49: 2113-2134.
- Baraldi, A., and L. Boschetti. 2012a. "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction," *Remote Sens.* 4: 2694-2735.
- Baraldi, A., and L. Boschetti. 2012b. "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation." *Remote Sens.* 4: 2768-2817.
- Baraldi, A., L. Boschetti, and M. Humber. 2014. "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images." *IEEE Trans. Geosci. Remote Sens.* 52(1): 701-760.
- Baraldi, A., L. Bruzzone, and P. Blonda. 2005. "Quality assessment of classification and cluster maps without ground truth knowledge." *IEEE Trans. Geosci. Remote Sens.* 43: 857-873.
- Baraldi, A., M. Girona, and D. Simonetti. 2010c. "Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sensing* 48(1): 112-146.
- Baraldi, A., and M. Humber. 2015. "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 1 – Theory," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 8(3): 1307-1329.
- Baraldi, A., M. Humber, and L. Boschetti. 2013. "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 2 – Experimental results," *Remote Sens.* 5: 5209-5264.
- Baraldi, A., and J. V. B. Soares. 2017. "Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis," Subjects: Computer Vision and Pattern Recognition (cs.CV), arXiv:1701.01941. Date: 8 Jan. 2017. [Online] Available: <https://arxiv.org/ftp/arxiv/papers/1701/1701.01941.pdf>



- Baraldi, A., V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino. 2006. "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images." *IEEE Trans. Geosci. Remote Sens.* 44:2563-2586.
- Baraldi, A., D. Tiede and S. Lang. 2015. "Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+." Project proposal (1-year duration): ESA Invitation to tender ESA/AO/1-8373/15/I-NB – "Value Adding Element (VAE): Next Generation EO-based Information Services", University of Salzburg, Department of Geoinformatics - Z\_GIS.
- Baraldi, A., D. Tiede, M. Sudmanns, M. Belgiu, and S. Lang. 2016. "Automated near real-time Earth observation Level 2 product generation for semantic querying," *GEOBIA 2016*, 14-16 Sept., University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands.
- Baraldi, A., T. Wassenaar, and S. Kay. 2010d. "Operational performance of an automatic preliminary spectral rule-based decision-tree classifier of spaceborne very high resolution optical images." *IEEE Trans. Geosci. Remote Sens.* 48: 3482-3502.
- Benavente, R., M. Vanrell, and R. Baldrich. 2008. "Parametric fuzzy sets for automatic color naming." *J. Optical Society of America A* 25:2582-2593.
- Berlin, B., and P. Kay. 1969. "Basic color terms: their universality and evolution." Berkeley: University of California.
- Beauchemin, M., and K. Thomson. 1997. "The evaluation of segmentation results and the overlapping area matrix." *Int. J. Remote Sens.* 18: 3895–3899.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon.
- Bishr, Y. 1998. "Overcoming the semantic and other barriers to GIS interoperability." *Int. J. Geographic. Info. Science* 12: 299-310.
- Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede. 2014. "Geographic object-based image analysis - towards a new paradigm." *ISPRS J. Photogram. Remote Sens.* 87: 180–191.
- Boschetti, L., S. P. Flasse, and P. A. Brivio. 2004. "Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto boundary." *Remote Sens. Environ.* 91: 280–292.
- Boschetti, L., D. P. Roy, C. O. Justice, and M. L. Humber. 2015. "MODIS–Landsat fusion for large area 30 m burned area mapping." *Remote Sens. Environ.* 161: 27-42.
- Bossard, M., J. Feranec and J. Otahel. 2000. "CORINE land cover technical guide – Addendum 2000, Technical report No 40." European Environment Agency.
- Capurro, R., and B. Hjørland. 2003. "The concept of information." *Annual Review of Information Science and Technology* 37: 343-411.
- Chavez, P. 1988. "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data." *Remote Sens. Environ.* 24:459–479.
- Cherkassky, V., and F. Mulier. 1998. *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley.
- Cimpoi M., S. Maji, I. Kokkinos I., and A. Vedaldi. 2014. "Deep filter banks for texture recognition, description, and segmentation," CoRR, abs/1411.6836.
- CNES. 2015. Venus Satellite Sensor Level 2 Product. Accessed 5 January 2016. [https://venus.cnes.fr/en/VENUS/prod\\_12.htm](https://venus.cnes.fr/en/VENUS/prod_12.htm)
- Congalton, R. G., and K. Green. 1999. *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton, FL, USA: Lewis Publishers.
- Couclelis, H. 2010. "Ontologies of geographic information." *Int. J. Geographic. Info. Science*, 24(12): 1785-1809.
- Despini, F., S. Teggi and A. Baraldi. 2014. "Methods and metrics for the assessment of pan-sharpening algorithms", in *SPIE Proceedings*, Vol. 9244: Image and Signal Processing for Remote Sensing XX, L. Bruzzone, J. A. Benediktsson, and F. Bovolo, Eds., Amsterdam, Netherlands, Sept. 22.
- Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) and VEGA Technologies. 2011. "Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document." Document S2PAD-ATBD-0001, European Space Agency.
- D'Elia, S. 2002. Personal communication, European Space Agency.
- Di Gregorio, A., and L. Jansen. 2000. Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO: Rome, Italy, FAO Corporate Document Repository. Accessed 10 February 2015. <http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm>
- Dillencourt, M. B., H. Samet, and M. Tamminen. 1992. "A general approach to connected component labeling for arbitrary image representations," *J. Association for Computing Machinery* 39: 253-280.



- Dorigo, W., R. Richter, F. Baret, R. Bamler, and W. Wagner. 2009. "Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach," *Remote Sens.* 1: 1139-1170.
- Duke Center for Instructional Technology, 2016. *Measurement: Process and Outcome Indicators*. Accessed 20 June 2016. [http://patientsafetyed.duhs.duke.edu/module\\_a/measurement/measurement.html](http://patientsafetyed.duhs.duke.edu/module_a/measurement/measurement.html)
- Dumitru C. O., S. Cui, G. Schwarz, and M. Datcu. 2015. "Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies." *IEEE J. Selected Topics Applied Earth Obs. Remote Sens.* 8(4): 1635 – 1650.
- Environmental Protection Agency (EPA). 2007. "Definitions" in *Multi-Resolution Land Characteristics Consortium (MRLC)*. Accessed 13 November 2013. <http://www.epa.gov/mrlc/definitions.html#2001>
- Environmental Protection Agency (EPA). 2013. *Western Ecology Division*. Accessed 13 November 2013. <http://www.epa.gov/wed/pages/ecoregions.htm>
- Elkan, C. 2003. "Using the triangle inequality to accelerate k-means," *Int. Conf. Machine Learning*.
- European Space Agency (ESA). 2015. Sentinel-2 User Handbook, , Standard Document, Issue 1 Rev 2.
- Feldman, J. 2016. "The neural binding problem(s)." *Cogn. Neurodyn.* 7: 1-11.
- Feng, C. C., and D. M. Flewelling. 2004. "Assessment of semantic similarity between land use/land cover classification systems." *Computers, Environ. Urban Systems* 28: 229–246.
- Fonseca, F., M. Egenhofer, P. Agouris, and G. Camara. 2002. "Using ontologies for integrated geographic information systems," *Transactions in GIS*, 6: 231–257.
- Frintrop, S. 2011. "Computational visual attention," in *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds., Springer.
- Fritzke, B. 1997a. "Some Competitive Learning Methods." Accessed 17 March 2015: <http://www.demogng.de/JavaPaper/t.html>
- Fritzke, B. 1997b. "The LBG-U method for vector quantization - An improvement over LBG inspired from neural networks," *Neural Processing Lett.* 5(1): xx-yyy.
- GeoTerraImage. 2015. Provincial and national land cover 30m. Accessed 22 September 2015. <http://www.geoterraimage.com/productslandcover.php>
- Gevers, T., A. Gijzenij, J. van de Weijer, and J. M. Geusebroek. 2012. *Color in Computer Vision*. Hoboken, NJ, USA: Wiley.
- Griffin, L. D. 2006. "Optimality of the basic color categories for classification." *J. R. Soc. Interface* 3: 71–85.
- Griffith, G. E., and J. M. Omernik. 2009. "Ecoregions of the United States-Level III (EPA)," in C.J. Cleveland (Ed.), *Encyclopedia of Earth*. Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment.
- Group on Earth Observation (GEO). 2005. "The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, adopted 16 February 2005." Accessed 10 January 2012. <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>
- Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS). 2010. "A Quality Assurance Framework for Earth Observation, version 4.0." Accessed 15 November 2012. [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS) - Working Group on Calibration and Validation (WGCV). 2015. *Land Product Validation (LPV)*. Accessed March 20, 2015. <http://lpvs.gsfc.nasa.gov/>
- Guarino, N. 1995. "Formal ontology, conceptual analysis and knowledge representation." *Int. J. Human Computer Studies* 43: 625–640.
- Gutman, G., A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss, D. Skole, B. L. Turner, M. A. Cochrane, Eds. 2004. *Land Change Science*. Dordrecht, The Netherlands: Kluwer.
- Hadamard, J. 1902. "Sur les problemes aux derivees partielles et leur signification physique," *Princet. Univ. Bull.* 13: 49–52.
- Homer, C., C. Q. Huang, L. M. Yang, B. Wylie, and M. Coan. 2004. "Development of a 2001 National Land-Cover Database for the United States." *Photo. Engin. Remote Sens.* 70:829-840.
- Hunt, N., and S. Tyrrell. 2012. *Stratified Sampling*. Coventry University. Accessed 7 February 2012. <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>
- Julesz, B. 1986. "Texton gradients: The texton theory revisited," in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, 54(4-5).



- Julesz, B., E. N. Gilbert, L. A. Shepp, and H. L. Frisch. 1973. "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception* 2: 391-405.
- Kavouras, M., and M. Kokla. 2002. "A method for the formalization and integration of geographical categorizations." *Int. J. Geographic. Info. Science* 16: 439-453.
- Kuzera, K., and R. G. Pontius jr. 2008. "Importance of matrix construction for multiple-resolution categorical map comparison." *GIScience and Remote Sens.* 45: 249–274.
- Laurini, R., and D. Thompson. 1992. *Fundamentals of Spatial Information Systems*. London, UK: Academic Press.
- Lee, D., S. Baek, and K. Sung. 1997. "Modified k-means algorithm for vector quantizer design," *IEEE Signal Processing Lett.*, 4: 2–4.
- Liang, S. 2004. *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons.
- Linde, Y., A. Buzo, and R. M. Gray. 1980. "An algorithm for vector quantizer design," *IEEE Trans. Commun.* 28: 84–94.
- Lloyd, S. P. 1982. "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, 28(2): 129–137.
- Lück W., and A. van Niekerk. 2016. "Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images," *Int. J. of Applied Earth Observation and Geoinformation* 47: 1–14.
- Lunetta, R., and D. Elvidge. 1999. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London, UK: Taylor & Francis.
- Luo, Y., A.P. Trishchenko and K.V. Khlopenkov. 2008. "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sensing of Environment* 112: 4167–4185.
- Nagao, M., and T. Matsuyama. 1980. *A Structural Analysis of Complex Aerial Photographs*. New York, NY, USA: Plenum, 1980.
- National Aeronautics and Space Administration (NASA). 2016. Data Processing Levels. [Online]. Available: <https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products>. Accessed on December 20, 2016.
- Marr, D. 1982. *Vision*. New York, NY: Freeman and C.
- Martinetz, T., G. Berkovich, and K. Schulten. 1994. "Topology representing networks." *Neural Networks* 7(3): 507–522.
- Matsuyama, T., and V. S. Hwang. 1990. *SIGMA – A Knowledge-based Aerial Image Understanding System*. New York, NY: Plenum Press.
- Memarsadeghi, N., D. Mount, N. Netanyahu, and J. Le Moigne. 2007. "A fast implementation of the ISODATA clustering algorithm." *Int. J. Comp. Geometry & Applications*. 17(1): 71-103.
- Mizen, H., C. Dolbear, and G. Hart. 2005. "Ontology ontogeny: Understanding how an ontology is created and developed," in Rodriguez, M., Cruz, I., Levashkin, S. and Egenhofer, M.J. (Eds.) *GeoSpatial Semantics: First International Conference, GeoS 2005, Mexico City, Mexico, November 29-30, 2005*. Proceedings, Springer Berlin Heidelberg, 15-29.
- Muirhead, K., and O. Malkawi. 1989. "Automatic classification of AVHRR images," in Proc. Fourth AVHRR Data Users Meeting, Rottenburg, Germany, 5-8 September 1989, 31-34.
- Open Geospatial Consortium (OGC) Inc. 2015. *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Accessed 8 March 2015. <http://www.opengeospatial.org/standards/iso>
- Ortiz, A., and G. Oliver. 2006. "On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures." *Pattern Recognition Letters* 27: 1916-1926.
- Parisi, D. 1991. "La scienza cognitive tra intelligenza artificiale e vita artificiale," in *Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza*. Bologna, Italy: Patron Editore.
- Patanè, G., and M. Russo. 2001. "The enhanced-LBG algorithm," *Neural Networks* 14(9): 1219–1237.
- Patanè, G., and M. Russo. 2002. "Fully automatic clustering system," *IEEE Trans. Neural Networks* 13(6): 1285-1298.
- Piaget, J. 1970. *Genetic Epistemology*. New York: Columbia University Press.
- Pontius Jr., R.G., and J. Connors. 2006. "Expanding the conceptual, mathematical and practical methods for map comparison," in Caetano, M. and Painho, M. (Eds.) Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon. Instituto Geográfico Português, 5 – 7 July 2006, 64-79.
- Richter, R., and D. Schlöpfer. 2012a. "Atmospheric / Topographic Correction for Satellite Imagery – ATCOR-2/3 User Guide, Version 8.2 BETA." Accessed 12 April 2013. [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor3\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor3_manual_2012.pdf)





- Richter, R., and D. Schlöpfer. 2012b. "Atmospheric / Topographic correction for airborne imagery – ATCOR-4 User Guide, Version 6.2 BETA, 2012." Accessed 12 April 2013. [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor4\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf)
- Roy, D. J., Ju, K. Kline, P. L. Scaramuzza, V. Kovalsky, M. Hansen, T. R. Loveland, E. Vermote, and C. S. Zhang. 2010. "Web-enabled Landsat Data (WELD): Landsat ETM plus composited mosaics of the conterminous United States." *Remote Sens. Environ.* 114:35-49.
- Salmon, B., K. Wessels, F. van den Bergh, K. Steenkamp, W. Kleyhans, D. Swanepoel, D. Roy, and V. Kovalsky. 2013. "Evaluation of rule-based classifier for Landsat-based automated land cover mapping in South Africa," *IEEE Int. Geosci. Remote Sensing Symposium (IGARSS)*, 21-26 July 2013, 4301-4304.
- Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik. 2006. "Reflectance quantities in optical remote sensing - Definitions and case studies," *Remote Sens. Environ.* 103: 27-42.
- Shackelford, K., and C. H. Davis. 2003a. "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.* 41(10): 2354 – 2363.
- Shackelford, K., and C. H. Davis. 2003b. "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.* 41(10): 1920-1932.
- Sheskin, D. 2000. *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL: Chapman & Hall/CRC.
- Simonetti, D., E. Simonetti, Z. Szantoi, A. Lupi, and H. D. Eva. 2015b. "First results from the phenology based synthesis classifier using Landsat-8 imagery," *IEEE Geosci. Remote Sens. Lett.* 12(7): 1496-1500.
- Simonetti, D., A. Marelli, and H. Eva. 2015b. *Impact Toolbox*, JRC Technical EUR 27358 EN.
- Smeulders, A., M. Worring, S. Santini, A. Gupta, and R. Jain. 2000. "Content-based image retrieval at the end of the early years." *IEEE Trans. Pattern Anal. Machine Intell.* 22(12): 1349-1380.
- Soares, J., A. Baraldi, and D. Jacobs. 2014. "Segment-based simple-connectivity measure design and implementation." Tech. Rep., University of Maryland, College Park, MD.
- Sonka, M., V. Hlavac, and R. Boyle. 1994. *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall.
- Sowa, J. F. 2000. *Knowledge representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA, USA: Brooks/Cole.
- Stehman, S. V. 1999. "Comparing thematic maps based on map value," *Int. J. Remote Sens.*, 20: 2347-2366.
- Stehman, S. V., and R. L. Czaplewski. 1998. "Design and analysis for thematic map accuracy assessment: Fundamental principles." *Remote Sens. Environ.* 64: 331-344.
- Tiede, D., A. Baraldi, M. Sudmanns, M. Belgiu, and S. Lang. 2016. "ImageQuerying (IQ) – Earth Observation Image Content Extraction & Querying across Time and Space," submitted (Oral presentation and poster session), ESA 2016 Conf. on Big Data From Space, BIDS '16, Santa Cruz de Tenerife, Spain, 15-17 March.
- Trimble. 2015. eCognition® Developer 9.0 Reference Book.
- Tsotsos, J. K. 1990. "Analyzing vision at the complexity level." *Behavioral and Brain Sciences* 13: 423-469.
- Vermote, E., and N. Saleous. 2007. "LEDAPS surface reflectance product description - Version 2.0." University of Maryland at College Park /Dept Geography and NASA/GSFC Code 614.5
- Vogelmann, J. E., T. L. Sohl, P. V. Campbell, and D. M. Shaw. 1998. "Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources." *Environ. Monitoring Assess.* 51: 415-428.
- Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. "Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources." *Photo. Eng. Remote Sens.* 67: 650-662.
- Web-Enabled Landsat Data (WELD) Tile FTP. Accessed 12 December 2016. <https://weld.cr.usgs.gov/>
- Wenwen Li, M. F. Goodchild and R. L. Church. 2013. "An efficient measure of compactness for 2D shapes and its application in regionalization problems." *Int. J. of Geographical Info. Science* 1-24.
- Wickham, J. D., S. V. Stehman, J. A. Fry, J. H. Smith, and C. G. Homer. 2010. "Thematic accuracy of the NLCD 2001 land cover for the conterminous United States." *Remote Sens. Environ.* 114: 1286-1296.
- Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J.A. Fry, and T.G. Wade. 2013. "Accuracy assessment of NLCD 2006 land cover and impervious surface." *Remote Sens. Environ.* 130: 294-304.
- Xian, G., and C. Homer. 2010. "Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods." *Remote Sens. Environ.* 114: 1676-1686.



Zadeh, L.A. 1965. "Fuzzy sets," *Inform. Control* 8: 338–353.

Figures and figure captions in Chapter 7

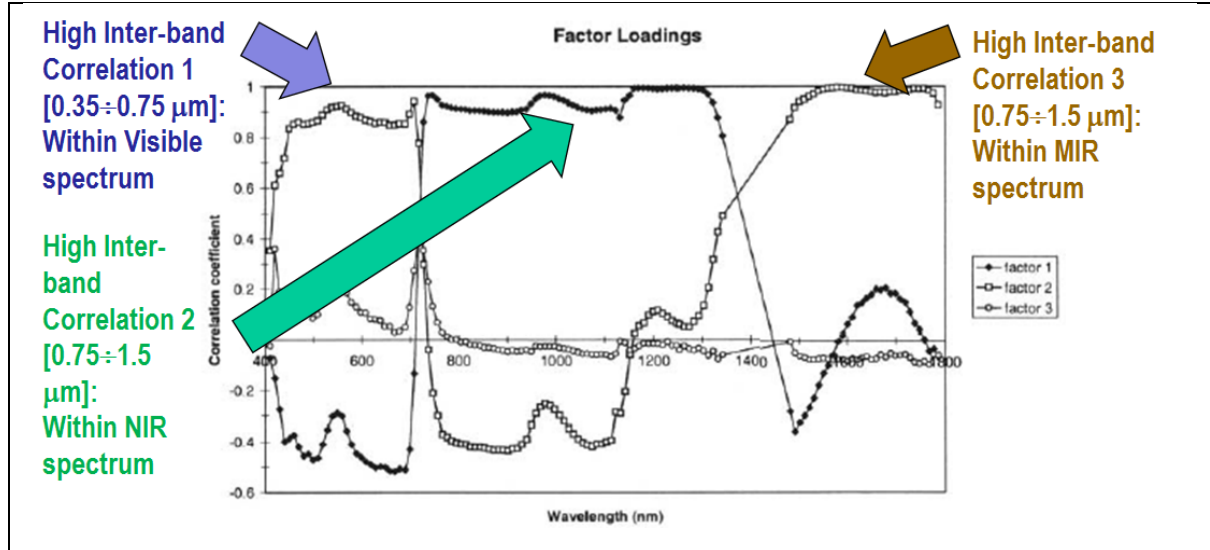


Fig. 7-1. Courtesy of van der Meer and De John (2000). Pearson's cross-correlation (CC) coefficients for the main factors resulting from a principal component analysis and factor rotation for an agricultural data set based on spectral bands of the AVIRIS hyper-spectral (HS) spectrometers 1, 2 and 3. Flevoland test site, July 5th 1991. Inter-band CC values are "high" ( $> 0.8$ ) within the visible spectral range, the Near Infra-Red (NIR) wavelengths and the Medium IR (MIR) wavelengths. The general conclusion is that, irrespective of non-stationary local information, the global (image-wide) information content of a multi-spectral (MS) image whose number  $N$  of spectral channels  $\in \{2, 9\}$ , a super-spectral (SS) image with  $N \in \{10, 20\}$ , or an hyperspectral (HS) image with  $N > 20$ , can be preserved by selecting one visible, one NIR, one MIR and one thermal IR (TIR) band, such as in the spectral resolution of the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) imaging sensor series in operating mode from 1978 to date.

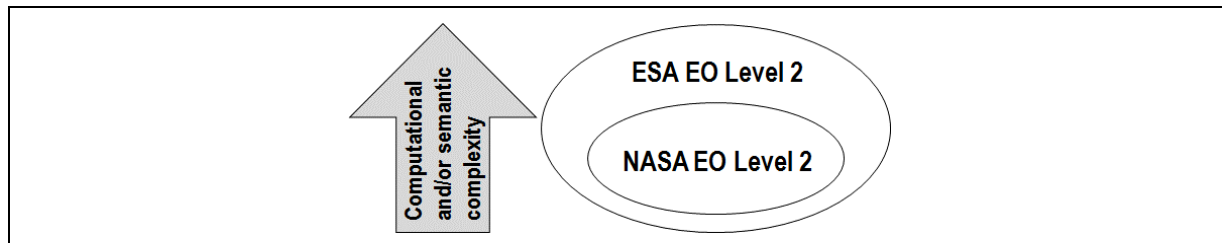


Fig. 7-2. A NASA EO Level 2 product is defined as "a data-derived geophysical variable at the same resolution and location as Level 1 source data" (NASA 2016). This definition is a special case of the ESA EO Level 2 product definition, encompassing a numeric variable provided with a physical unit of measure, e.g., surface reflectance (SURF) values corrected for atmospheric, topographic and adjacency effects, stacked with its data-derived scene classification map (SCM), equivalent to a categorical variable provided with semantics (ESA 2015).

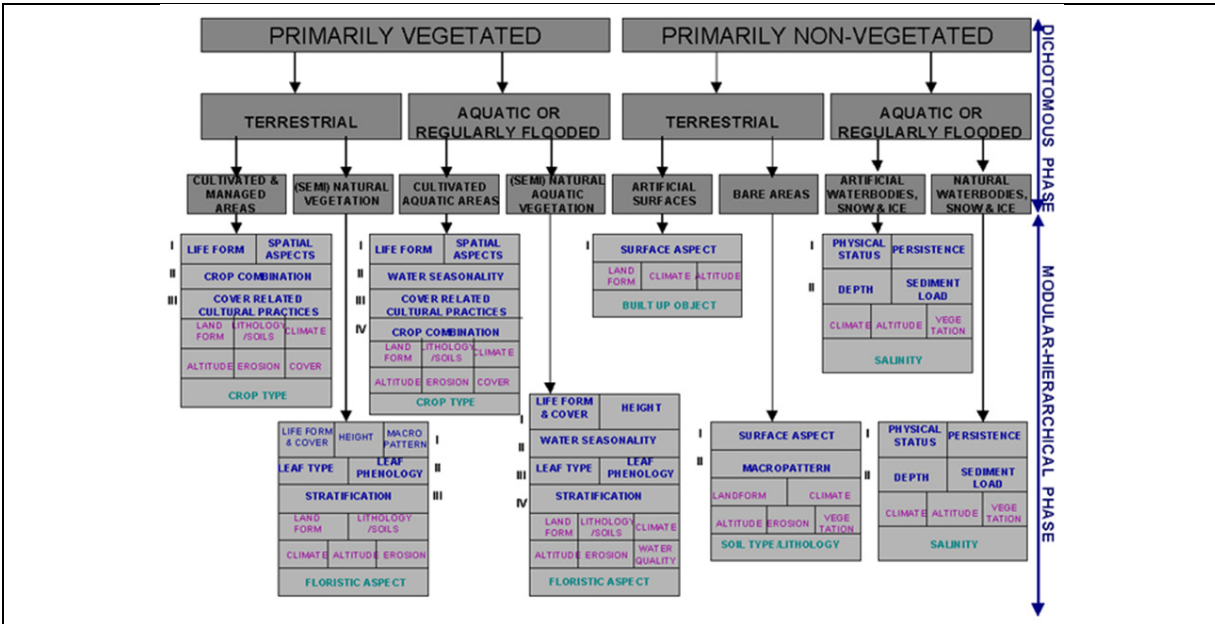


Fig. 7-3. The nested 3-level LCCS-DP layers are: (i) vegetation versus non-vegetation, (ii) terrestrial versus aquatic, and (iii) managed versus natural or semi-natural. They deliver as output the following 8-class LCCS-DP taxonomy. (A11) Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas. (A12) Natural and Semi-Natural Terrestrial Vegetation. (A23) Cultivated Aquatic or Regularly Flooded Vegetated Areas. (A24) Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation. (B35) Artificial Surfaces and Associated Areas. (B36) Bare Areas. (B47) Artificial Waterbodies, Snow and Ice. (B48) Natural Waterbodies, Snow and Ice. The general-purpose user- and application-independent 8-class LCCS-DP taxonomy is preliminary to a user- and application-specific LCCS Modular Hierarchical Phase (MHP) taxonomy.

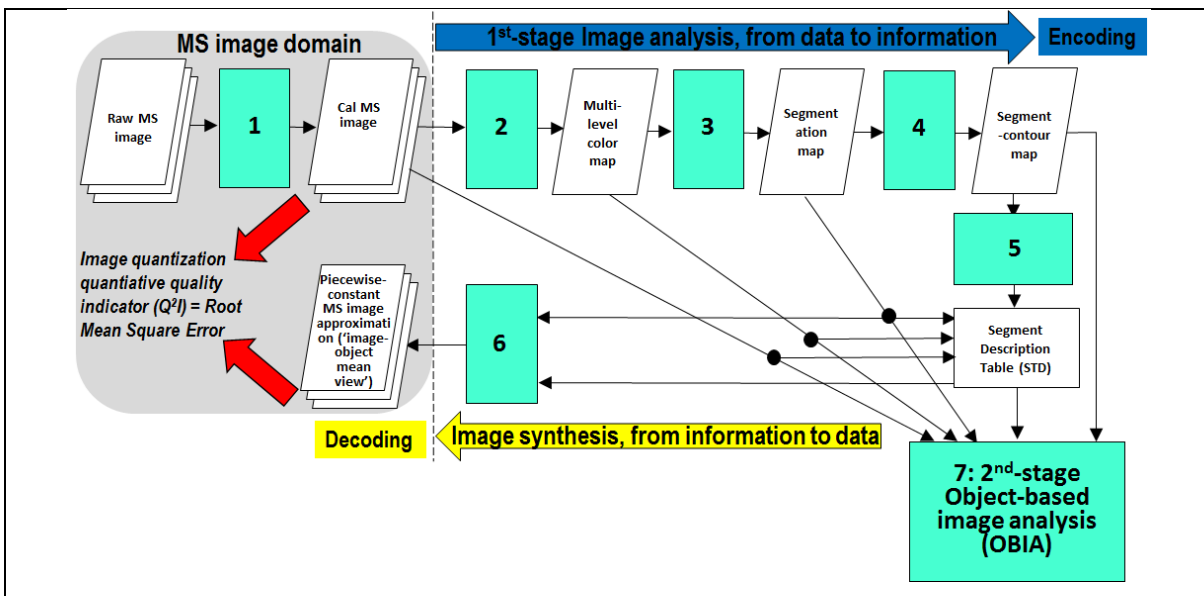


Fig. 7-4. The SIAM application software for prior knowledge-based vector quantization (VQ) in a radiometrically calibrated MS data space. It consists of six subsystems, identified as 1 to 6. Phase 1-of-2 = Encoding phase/Image analysis - Stage 1: MS data calibration into top-of-atmosphere reflectance (TOARF) or surface reflectance (SURF) values. Stage 2: Prior knowledge-based SIAM decision tree for MS reflectance space partitioning (quantization, hyperpolyhedralization). Stage 3: Well-posed (deterministic) two-pass connected-component detection in the multi-level color map-domain. Connected-components in the color map-domain are connected sets of pixels featuring the same color label. These connected-components are also called image-objects, segments or superpixels. Stage 4: Well-





posed superpixel-contour extraction. Stage 5: Superpixel description table allocation and initialization. Phase 2-of-2 = Decoding phase/Image synthesis - Stage 6: Superpixelwise-constant input image approximation (“image-object mean view”) and per-pixel VQ error estimation. (Stage 7: in cascade to the SIAM superpixel detection, a high-level object-based image analysis (OBIA) approach can be adopted).

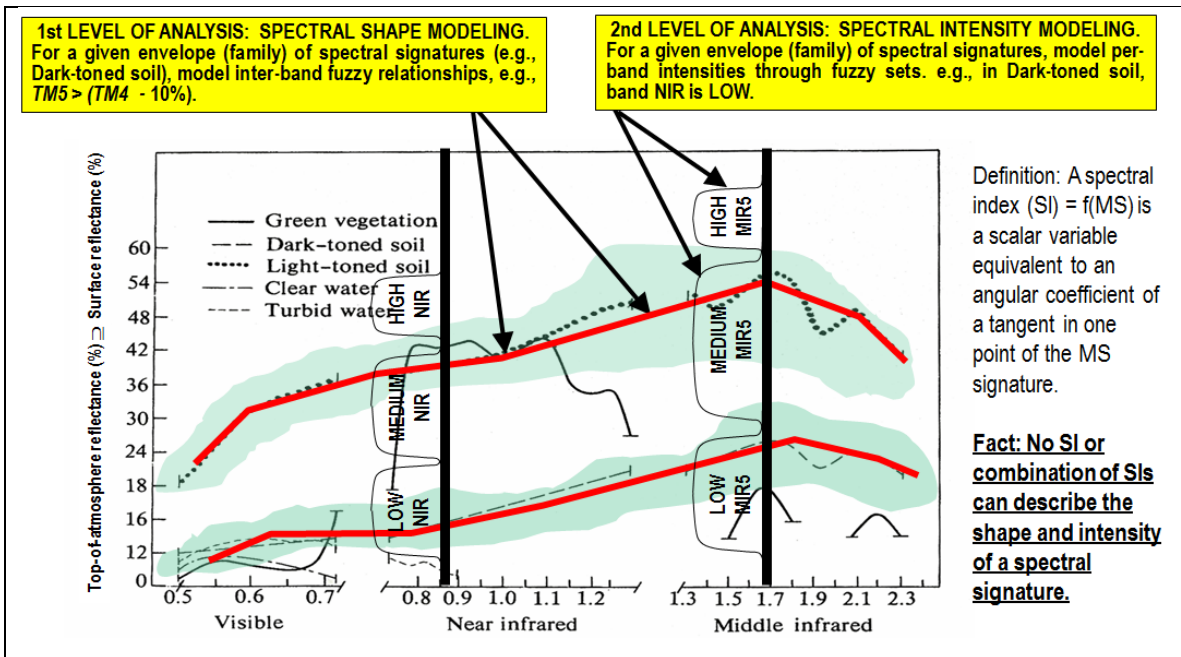


Fig. 7-5. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values which include surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF values forms a buffer zone (hyperpolyhedron, envelope, manifold). The SIAM decision tree models each target family of spectral signatures in terms of multivariate shape and multivariate intensity as a viable alternative to multivariate analysis of spectral indexes. A typical spectral index is a scalar band ratio equivalent to an angular coefficient of a tangent in one point of the spectral signature. Infinite functions can feature the same tangent value in one point. In practice, no spectral index or combination of spectral indexes can reconstruct the multivariate shape and multivariate intensity of a spectral signature.

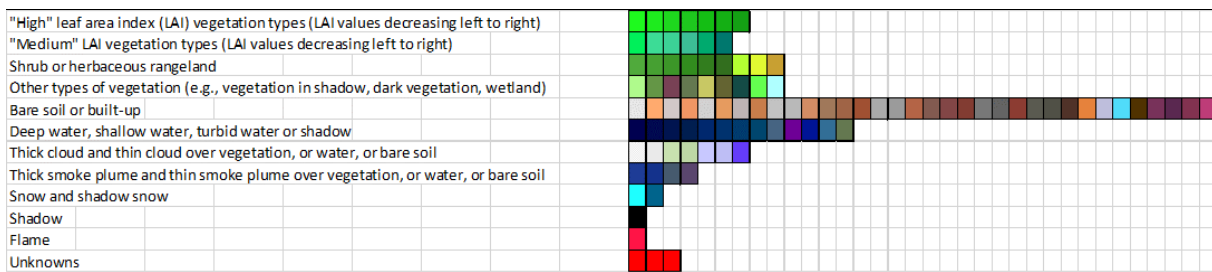


Fig. 7-6. Prior knowledge-based color map legend adopted by the Landsat-like SIAM (L-SIAM™, release 88 version 6) implementation. For the sake of representation compactness pseudocolors of the 96 spectral categories are gathered along the same row if they share the same parent spectral category in the decision tree, e.g., "strong" vegetation, equivalent to a spectral end-member. The pseudocolor of a spectral category is chosen as to mimic natural colors of pixels belonging to that spectral category. These 96 color names at fine color granularity are aggregated into 48 and 18 color names at intermediate and coarse color granularity respectively, according to parent-child relationships defined *a priori*, also refer to Table 7-1.

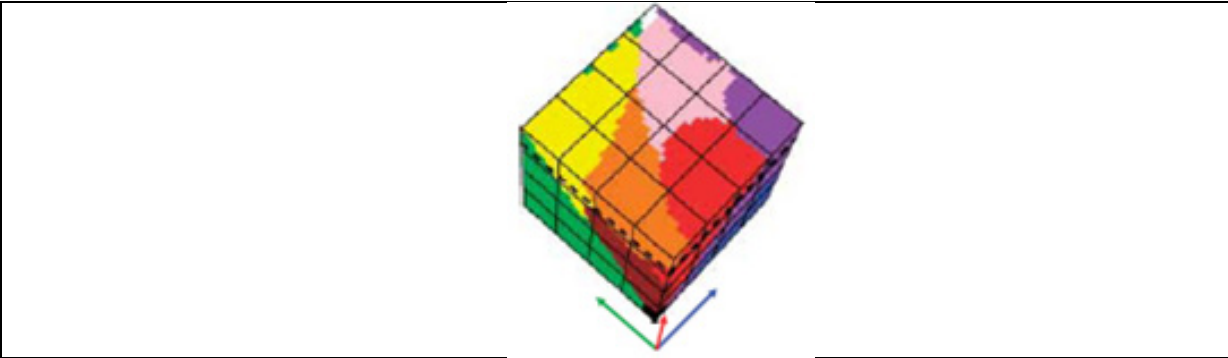


Fig. 7-7. Courtesy of Griffin (2006). Monitor-typical RGB cube partitioned into perceptual polyhedra corresponding to a discrete and finite dictionary of basic color (BC) names, to be community-agreed upon in advance to be employed by members of the community. The mutually exclusive and totally exhaustive polyhedra are neither necessarily convex nor connected. In practice BC names belonging to a finite and discrete color dictionary are equivalent to Vector Quantization (VQ) levels belonging to a VQ codebook (Cherkassky and Mulier 1998).

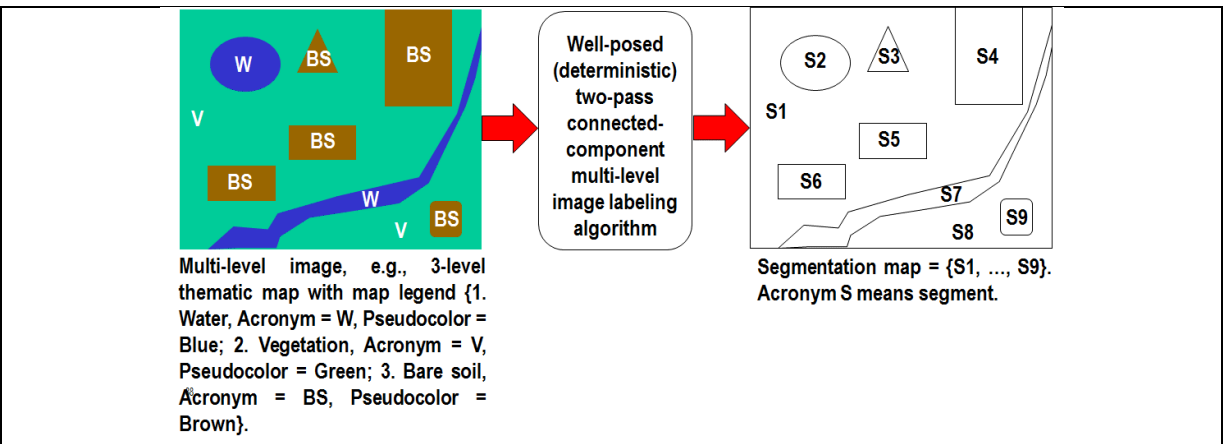


Fig. 7-8. One segmentation map is deterministically generated from one multi-level image, such as a thematic map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. In this example, nine image-objects/segments S1 to S9 can be detected in the 3-level thematic map shown at left. Each segment consists of a connected set of pixels sharing the same multi-level map label. Each stratum/layer/level consists of one or more segments, e.g., stratum Vegetation (V) consists of two disjoint segments, S1 and S8. In any multi-level (categorical, nominal, qualitative) image domain, three labeled spatial primitives co-exist and are provided with parent-child relationships: pixel with a level-label and a pixel identifier (ID, e.g., the row-column coordinate pair), segment (polygon) with a level-label and a segment ID, and stratum (multi-part polygon) with a level-label equivalent to a stratum ID. This overcomes the ill-fated dichotomy between traditional unlabeled sub-symbolic pixels versus labeled sub-symbolic segments in the numeric (quantitative) image domain traditionally coped with by the object-based image analysis (OBIA) paradigm (Blaschke et al. 2014).

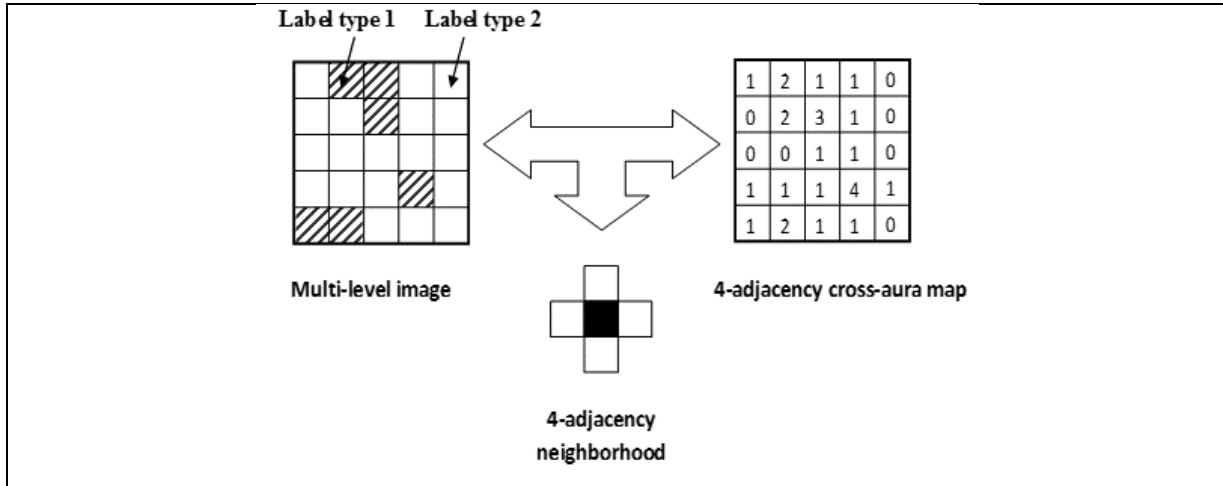
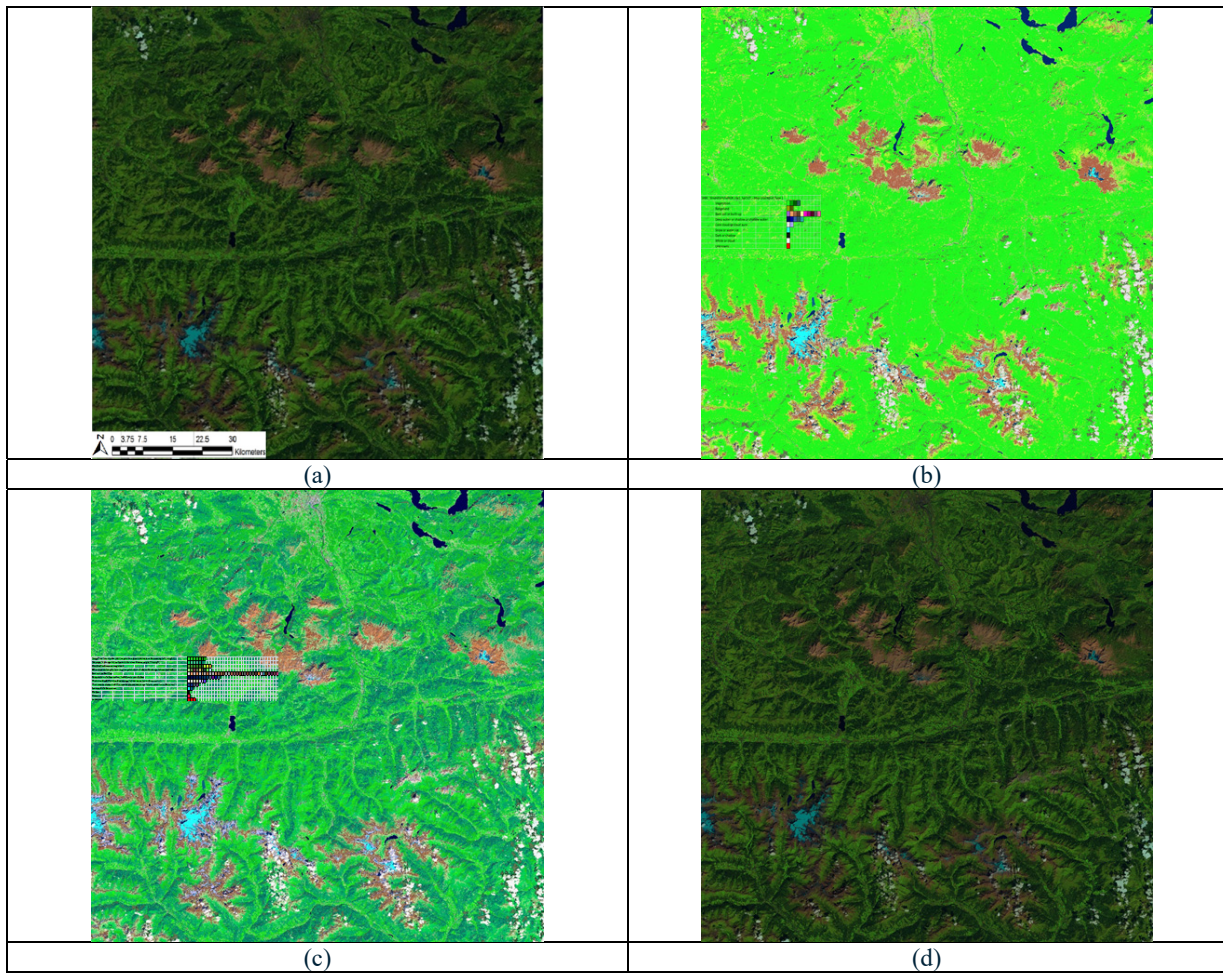


Fig. 7-9. Example of a 4-adjacency cross-aura map, shown at right, generated in linear time from a two-level image shown at left.





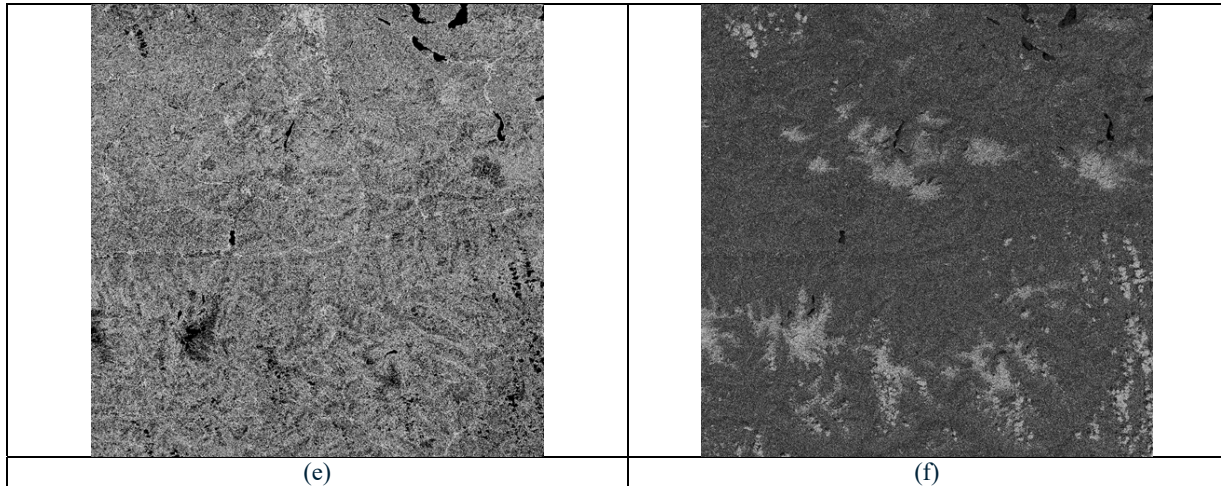


Fig. 7-10. (a) Sentinel-2A MSI Level-1C image of the Earth surface south of the city of Salzburg, Austria. The city area is visible around the middle of the image upper boundary (Lat-long coordinates: 47°48'25.0"N 13°02'43.6"E). Acquired on 2015-09-11. Spatial resolution: 10 m. Image size: 110×110 km. Radiometrically calibrated into TOARF values in range  $\{0, 255\}$ , it is depicted as a false color RGB image, where: R = Medium InfraRed (MIR) = Band 11, G = Near IR (NIR) = Band 8, B = Blue = Band 2. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at coarse color granularity, consisting of 18 spectral categories depicted in pseudo colors shown in the map legend. Coarse-granularity color categories are generated by merging color hyperpolyhedra at fine color granularity, according to pre-defined parent-child relationships, refer to Table 7-1. (c) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors shown in the map legend. (d) Superpixelwise-constant approximation of the input image (“image-object mean view”) generated from the L-SIAM’s 96 color map at fine granularity. Depicted in false colors: R = MIR = Band 11, G = NIR = Band 8, B = Blue = Band 2. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (e) 8-adjacency cross-aura contour map in range  $\{0, 8\}$  automatically generated from the L-SIAM’s 96 color map at fine granularity. It shows contours of connected sets of pixels featuring the same color label. These connected-components are also called image-objects, segments or superpixels. (f) Per-pixel scalar difference between the input MS image shown in (a) and the superpixelwise-constant MS image reconstruction shown in (d). This scalar difference is computed as the per-pixel Root Mean Square Error (RMSE) in range  $\{0, 255\}$ . The RMSE is a well-known vector quantization (VQ) error. Image-wide basic statistics: Min = 0, Max = 130, Mean = 2.60, Stdev = 3.45. Histogram stretching is applied for visualization purposes.



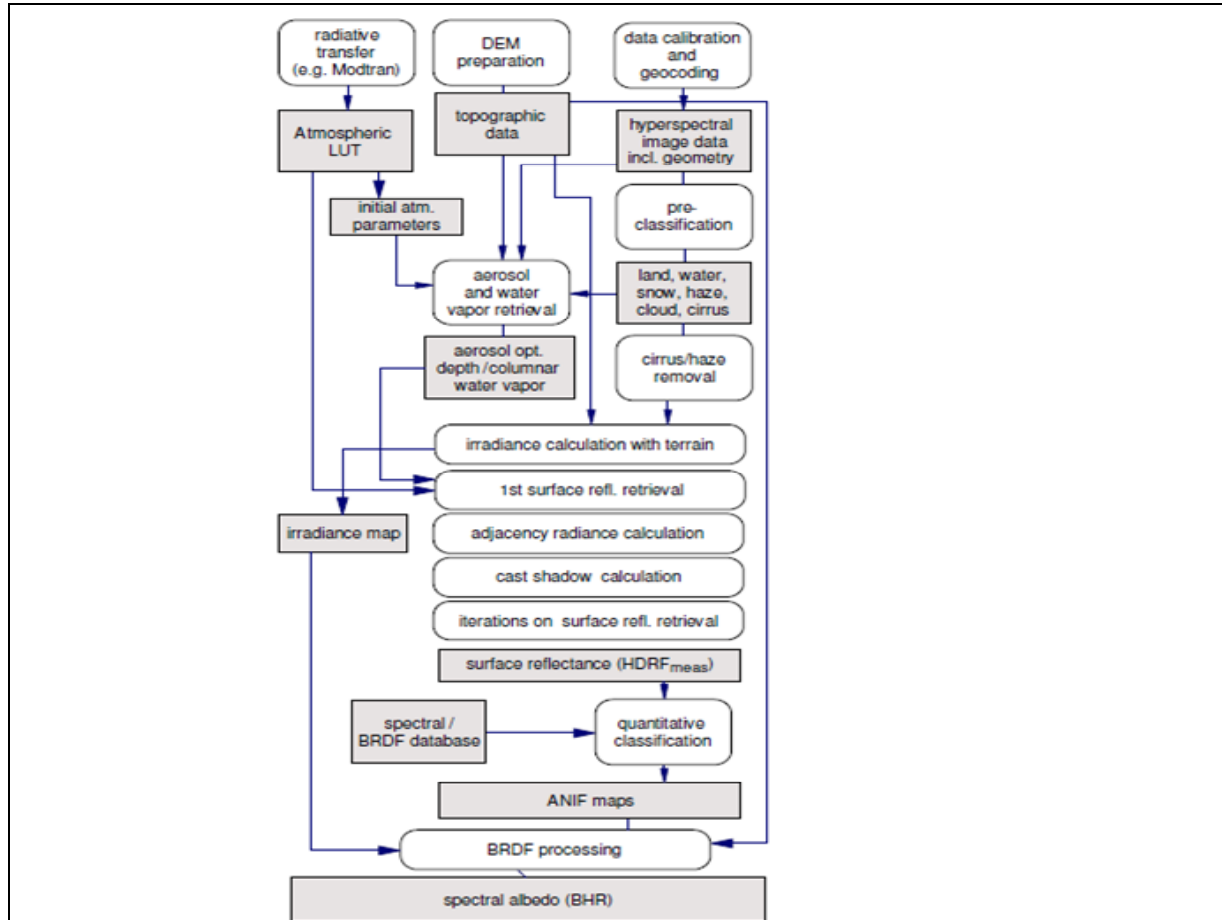


Fig. 7-11. Same as in Richter et al. (2009), courtesy of Daniel Schläpfer, ReSe Applications Schläpfer. A complete (“augmented”) hybrid inference workflow for MS image correction from atmospheric, adjacency and topographic effects. It combines a standard Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software workflow (Richter and D. Schläpfer 2012a; Richter and D. Schläpfer 2012b), with a bidirectional reflectance distribution function (BRDF) effect correction. Processing blocks are represented as circles and output products as rectangles. This hybrid (combined deductive and inductive) workflow alternates deductive/prior knowledge-based and inductive/learning-from-data inference units, starting from initial conditions provided by a first-stage deductive Spectral Classification of surface reflectance signatures (SPECL) decision tree for color naming (pre-classification), implemented within the ATCOR commercial software toolbox (Richter and D. Schläpfer 2012a; Richter and D. Schläpfer 2012b). Categorical variables generated by the pre-classification and classification blocks are employed to stratify (mask) unconditional numeric variable distributions, in line with the statistic stratification principle (Hunt and Tyrrell 2012). Through statistic stratification, inherently ill-posed inductive learning-from-data algorithms are provided with prior knowledge required in addition to data to become better posed for numerical solution, in agreement with the machine learning-from-data literature (Cherkassky and Mulier 1998).

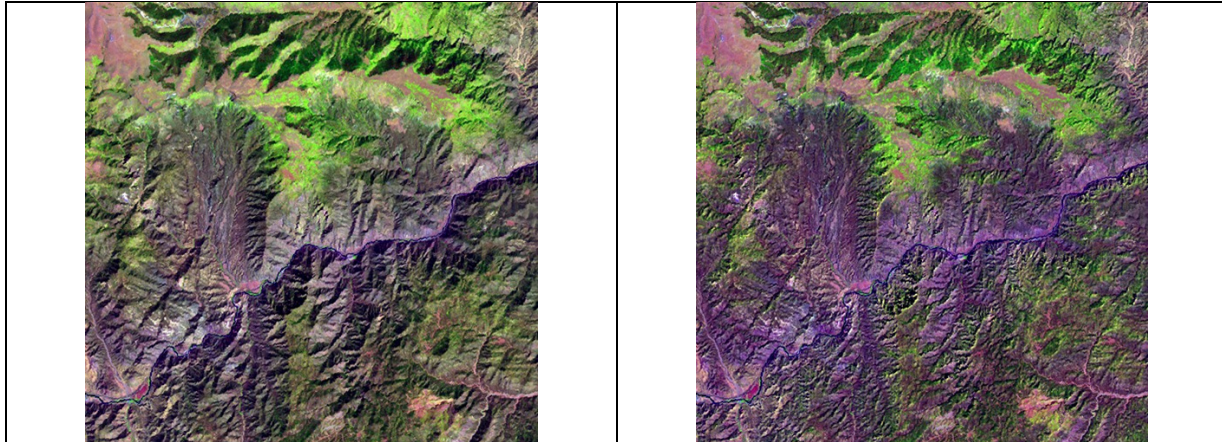


Fig. 7-12. Left: Zoomed area of a Landsat 7 ETM+ image of Colorado, USA (path: 128, row: 021, acquisition date: 2000-08-09), depicted in false colors (R: band ETM5, G: band ETM4, B: band ETM1), 30 m resolution, radiometrically calibrated into TOARF values. Right: Output product automatically generated without human-machine interaction by the stratified topographic correction (STOC) algorithm proposed in Baraldi et al. (2010), whose input datasets are one Landsat image, its data-derived L-SIAM color map at coarse color granularity, consisting of 18 spectral categories for stratification purposes (see Table 7-1), and a standard 30 m resolution Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM).

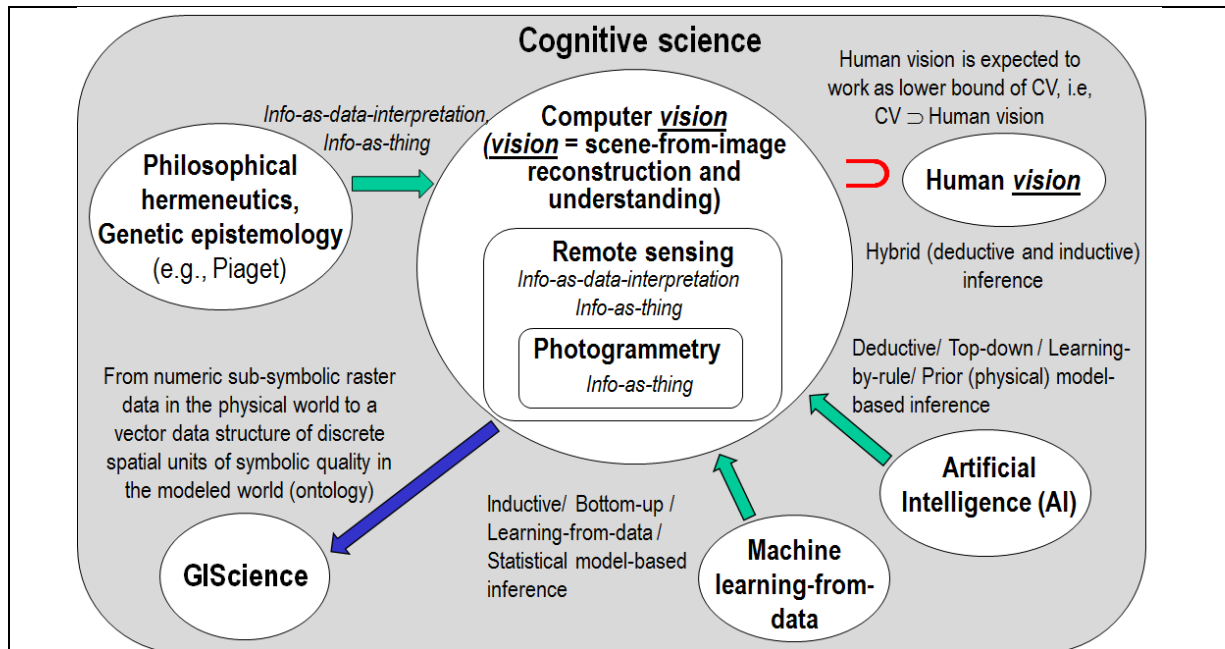


Fig. 7-13. Like engineering, RS is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred within nervous systems (distributed processing systems in humans, such as the human brain, or other animals) and machines (e.g., computers). Neurophysiology studies nervous systems, including the brain.

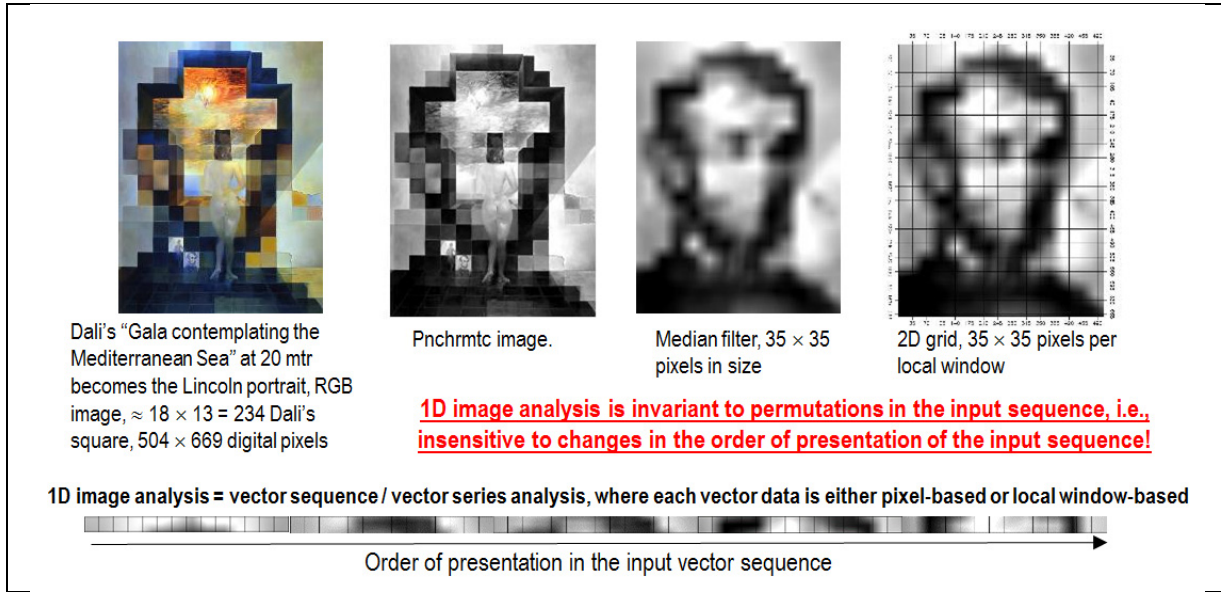


Fig. 7-14. Example of 1D image analysis. The (2D) image at left is transformed into the 1D vector data stream shown at bottom, where vector data are either pixel-based or spatial context-sensitive, e.g., local window-based. This 1D vector data stream means nothing to a human photointerpreter. When it is input to a traditional inductive data learning classifier, it is what the inductive classifier actually sees when watching the (2D) image at left. Undoubtedly, computers are more successful than humans in 1D image analysis. Nonetheless, humans are still far more successful than computers in (2D) image analysis.

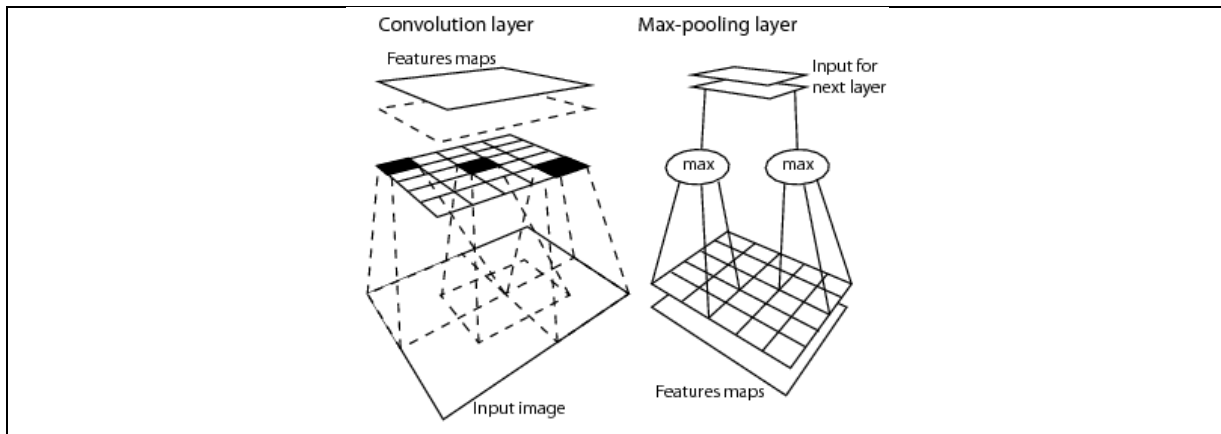


Fig. 7-15. 2D image analysis as synonym of retinotopic/topology-preserving feature mapping in a (2D) image-domain. Activation domains of physically adjacent processing units in the 2D array of convolutional filters are spatially adjacent regions in the 2D visual field. Provided with a superior degree of biological plausibility in modelling 2D spatial topological and non-topological information, distributed processing systems capable of 2D image analysis, such as deep convolutional neural networks (DCNNs), typically outperform traditional 1D image analysis approaches. Will computers become as good as humans in 2D image analysis?

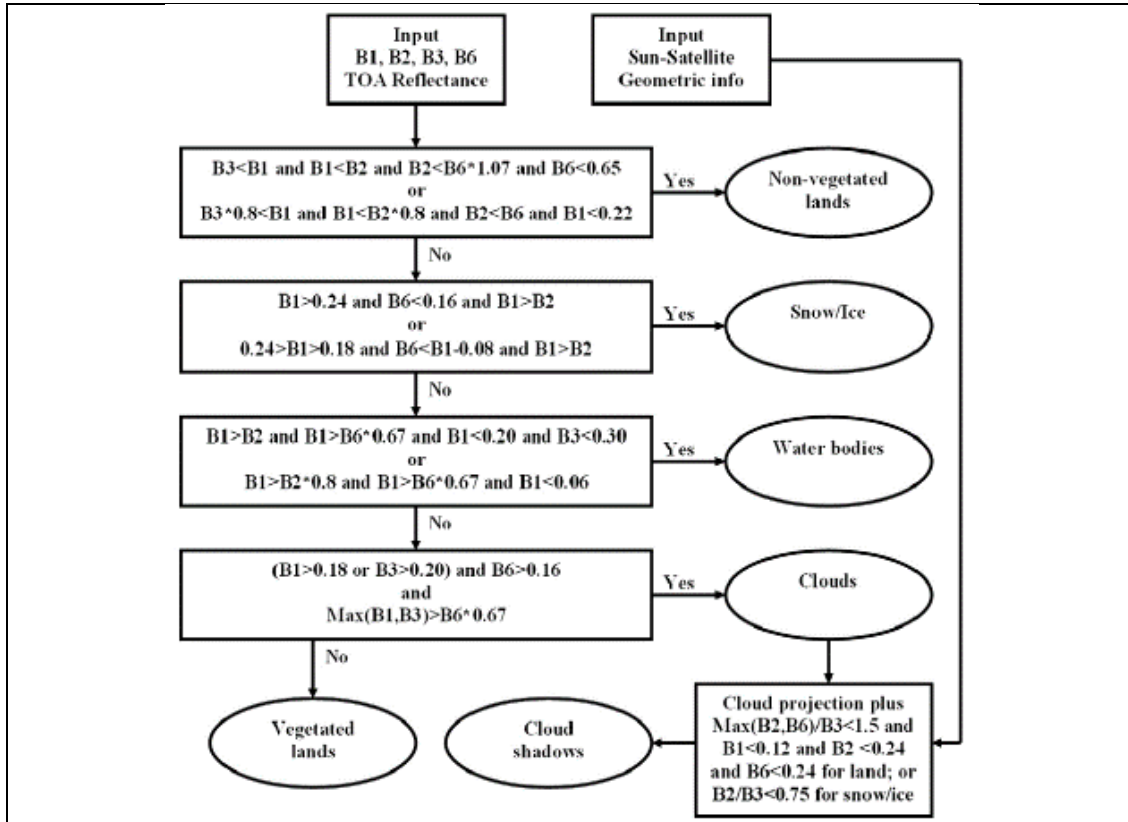


Fig. 7-16. Courtesy of Luo et al. (2008). Canada Centre for Remote Sensing (CCRS)'s flow chart of a physical model-based per-pixel MODIS image classifier integrated into a clear-sky multi-temporal MODIS image compositing system in operating mode. Acronym B stands for MODIS Band.

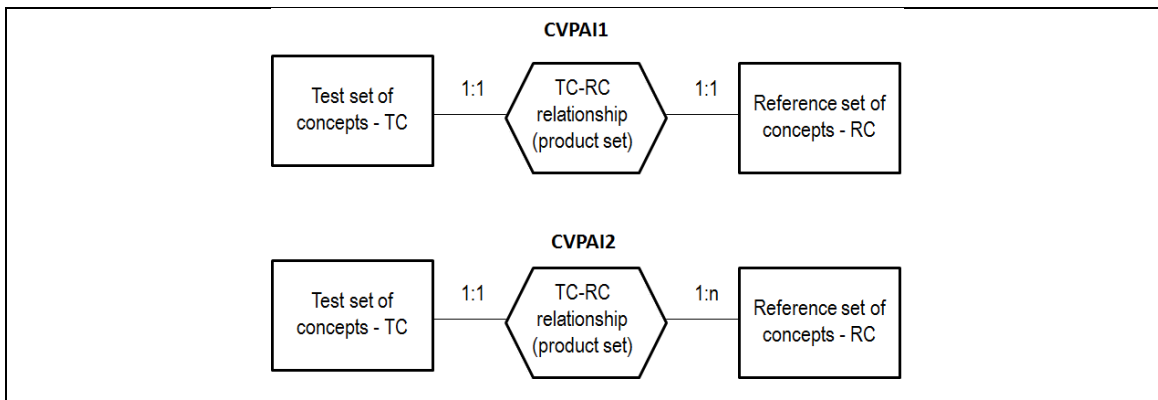


Fig. 7-17. Entity-relationship conceptual model representation of the binary relationship  $R: A \Rightarrow B$  from set  $A =$  test categorical variable to set  $B =$  reference categorical variable, provided with the min:max cardinality required by the Categorical Variable-Pair Association Index (CVPAI) formulation 1 (CVPAI1) and formulation 2 (CVPAI2) to score maximum in range  $[0, 1]$ . Inequality  $CVPAI1 \leq CVPAI2$  holds, i.e., the latter is a relaxed version of the former. In particular, CVPAI1 is maximum (equal 1) when the binary relationship  $R: A \Rightarrow B$  from set  $A =$  test categorical variable to set  $B =$  reference categorical variable is a bijective function, both injective (one-to-one) and surjective (onto). CVPAI2 is maximum when the binary relationship  $R: A \Rightarrow B$  is either a surjective function or a bijective function.





Tables and table captions in Chapter 7

SIAM, r88v6	Input bands	Prior knowledge-based color map legends: Number of output spectral categories			
		Fine discretization levels	Intermediate discretization levels	Coarse discretization levels	Inter-sensor discretization levels (*)
L-SIAM	7 – B, G, R, NIR, MIR1, MIR2, TIR	96	48	18	33 * employed for inter-sensor post-classification change/no-change detection
S-SIAM	4 – G, R, NIR, MIR1	68	40	15	
AV-SIAM	4 – R, NIR, MIR1, TIR	83	43	17	
Q-SIAM	4 – B, G, R, NIR	61	28	12	








Table 7-1. The SIAM computer program is an EO system of systems scalable to any existing or future MS imaging sensors provided with radiometric calibration metadata parameters. It encompasses the following subsystems. (i) 7-band Landsat-like SIAM™ (L-SIAM™), with input channels Blue (B), Green (G), Red (R), Near Infra-Red (NIR), Medium IR1 (MIR1), Medium IR2 (MIR2), and Thermal IR (TIR). (ii) 4-band (channels G, R, NIR, MIR1) SPOT-like SIAM™ (S-SIAM™). (iii) 4-band (channels R, NIR, MIR1, and TIR) Advanced Very High Resolution Radiometer (AVHRR)-like SIAM™ (AV-SIAM™). (iv) 4-band (channels B, G, R, and NIR) QuickBird-like SIAM™ (Q-SIAM™).

NLCD 2001/2006/2011 Classification Scheme (Legend), Level II				LCCS-DP, level 1: A = Veg, B = Non-Veg, and level 2: 1 = Terrestrial, 2 = Aquatic
Code	ID	Name	Land cover (LC) Class Definition	ID
11	OW	Open water	OW: Areas of open water, generally with less than 25% cover of vegetation or soil	B4 - Non-vegetated aquatic
12	PIS	Perennial Ice/Snow	PIS: Areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.	B4
21 22 23 24	DOS DLI DMI DHI	Developed, Open Space Developed, Low Intensity Developed, Medium Intensity Developed, High Intensity	DOS: Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes. DLI, DMI, DHI: refer to the “National Land Cover Database 2006 (NLCD2006),” Multi-Resolution Land Characteristics Consortium (MRLC), 2013.	B3 - Non-vegetated terrestrial / A1 - Vegetated terrestrial
31	BL	Barren Land (Rock/Sand/Clay)	BL: Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover. As a consequence of this constraint, class BL covers only 1.21% of the CONUS total surface.	B3
41 42 43	DF EF MF	Deciduous Forest Evergreen Forest Mixed Forest	DF: Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change. EF: Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.	A1



			MF: Mixed Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.	
51 52	- SS	Dwarf Scrub <sup>2</sup> Scrub/Shrub	SS: Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions. The aforementioned definition of class BL means that class SS may feature a vegetated cover which accounts for 15% of total cover or more.	A1/ B3
71 72 73 74	GH - - -	Grassland/Herbaceous Sedge Herbaceous <sup>2</sup> Lichens <sup>2</sup> Moss <sup>2</sup>	GH: Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing. The aforementioned definition of class BL means that class GH may feature a vegetated cover which accounts for 15% of total cover or more.	A1/B3
81 82	PH CC	Pasture/Hay Cultivated Crops	PH: Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation. CC: Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.	A1
90 95	WW EHW	Woody Wetlands Emergent Herbaceous Wetland	WW: Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water. EHW: Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.	A2 – Vegetated aquatic

Table 7-2. Definition of the NLCD 2001/2006/2011 classification taxonomy, Level II. <sup>2</sup>Alaska only. For further details, refer to the “National Land Cover Database 2006 (NLCD2006),” Multi-Resolution Land Characteristics Consortium (MRLC), 2013.. The right column instantiates a possible binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{NLCD legend}$  to set  $B = 2\text{-level } 4\text{-class Dichotomous Phase (DP) taxonomy of the Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) (Di Gregorio and Jansen 2000), refer to Fig. 7-3.$

		Target classes of individuals (entities in a conceptual model for knowledge representation built upon an ontology language)		
		Class 1, Water body	Class 2, Tulip flower	Class 3, Italian tile roof
Color names	black			√
	blue		√	√
	brown		√	√
	grey			
	green		√	√
	orange			√
	pink			√



	purple			√	
	red			√	√
	white			√	
	yellow			√	

Table 7-3. Example of a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{DictionaryOfColorNames}$ , with cardinality  $|A| = a = \text{ColorDictionaryCardinality} = 11$ , and the set  $B = \text{LegendOfObjectClassNames}$ , with cardinality  $|B| = b = \text{ObjectClassLegendCardinality} = 3$ . The latter dictionary is a superset of the typical taxonomy of land cover (LC) classes adopted by the RS community. “Correct” entry-pairs (marked with  $\checkmark$ ) must be: (i) selected by domain experts based on a hybrid combination of deductive prior beliefs with inductive evidence from data (refer to Table 7-5) and (ii) community-agreed upon.

Index	Spectral Categories	Spectral Rule (based on reflectance measured at Landsat TM central wave bands: b1 is located at 0.48 $\mu\text{m}$ , b2 at 0.56 $\mu\text{m}$ , b3 at 0.66 $\mu\text{m}$ , b4 at 0.83 $\mu\text{m}$ , b5 at 1.6 $\mu\text{m}$ , b7 at 2.2 $\mu\text{m}$ )	Pseudocolor
1	Snow/ice	$b4/b3 \leq 1.3$ AND $b3 \geq 0.2$ AND $b5 \leq 0.12$	
2	Cloud	$b4 \geq 0.25$ AND $0.85 \leq b1/b4 \leq 1.15$ AND $b4/b5 \geq 0.9$ AND $b5 \geq 0.2$	
3	Bright bare soil / sand / cloud	$b4 \geq 0.15$ AND $1.3 \leq b4/b3 \leq 3.0$	
4	Dark bare soil	$b4 \geq 0.15$ AND $1.3 \leq b4/b3 \leq 3.0$ AND $b2 \leq 0.10$	
5	Average vegetation	$b4/b3 \geq 3.0$ AND $(b2/b3 \geq 0.8$ OR $b3 \leq 0.15)$ AND $0.28 \leq b4 \leq 0.45$	
6	Bright vegetation	$b4/b3 \geq 3.0$ AND $(b2/b3 \geq 0.8$ OR $b3 \leq 0.15)$ AND $b4 \geq 0.45$	
7	Dark vegetation	$b4/b3 \geq 3.0$ AND $(b2/b3 \geq 0.8$ OR $b3 \leq 0.15)$ AND $b3 \leq 0.08$ AND $b4 \leq 0.28$	
8	Yellow vegetation	$b4/b3 \geq 2.0$ AND $b2 \geq b3$ AND $b3 \geq 8.0$ AND $b4/b5 \geq 1.5^a$	
9	Mix of vegetation / soil	$2.0 \leq b4/b3 \leq 3.0$ AND $0.05 \leq b3 \leq 0.15$ AND $b4 \geq 0.15$	
10	Asphalt / dark sand	$b4/b3 \leq 1.6$ AND $0.05 \leq b3 \leq 0.20$ AND $0.05 \leq b4 \leq 0.20^a$ AND $0.05 \leq b5 \leq 0.25$ AND $b5/b4 \geq 0.7^a$	
11	Sand / bare soil / cloud	$b4/b3 \leq 2.0$ AND $b4 \geq 0.15$ AND $b5 \geq 0.15^a$	
12	Bright sand / bare soil / cloud	$b4/b3 \leq 2.0$ AND $b4 \geq 0.15$ AND $(b4 \geq 0.25$ OR $b5 \geq 0.30^b)$	
13	Dry vegetation / soil	$(1.7 \leq b4/b3 \leq 2.0$ AND $b4 \geq 0.25^c)$ OR $(1.4 \leq b4/b3 \leq 2.0$ AND $b7/b5 \leq 0.83^c)$	
14	Sparse veg. / soil	$(1.4 \leq b4/b3 \leq 1.7$ AND $b4 \geq 0.25^c)$ OR $(1.4 \leq b4/b3 \leq 2.0$ AND $b7/b5 \leq 0.83$ AND $b5/b4 \geq 1.2^c)$	
15	Turbid water	$b4 \leq 0.11$ AND $b5 \leq 0.05^a$	
16	Clear water	$b4 \leq 0.02$ AND $b5 \leq 0.02^a$	
17	Clear water over sand	$b3 \geq 0.02$ AND $b3 \geq b4 + 0.005$ AND $b5 \leq 0.02^a$	
18	Shadow		
19	Not classified (outliers)		

<sup>a</sup> These expressions are optional and only used if b5 is present. <sup>b</sup> Decision rule depends on presence of b5. <sup>c</sup> Decision rule depends on presence of b7.

Table 7-4. Rule set (structural knowledge) and order of presentation of the rule set (procedural knowledge) adopted by the prior knowledge-based MS reflectance space quantizer called Spectral Classification of surface reflectance signatures (SPECL), implemented within the ATCOR commercial software toolbox (Dorigo et al. 2009; Richter and D. Schläpfer 2012a, 2012b).



<b>STEP 1. Dictionary-pair relationship, multivariate occurrence distributions</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	10	30	60	100	
	Cloud	2	0	10	12	
	Unknowns	0	5	100	105	
		12	35	170	217	
<b>STEP 2. Dictionary-pair relationship, multivariate probability distributions</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	0.046082949	0.138248848	0.276498	0.460829	
	Cloud	0.00921659	0	0.046083	0.0553	
	Unknowns	0	0.023041475	0.460829	0.483871	
		0.055299539	0.161290323	0.78341	1	
<b>STEP 3. Dictionary-pair relationship, cond. prob. (RC TC)</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	0.1	0.3	0.6	1	
	Cloud	0.166666667	0	0.833333	1	
	Unknowns	0	0.047619048	0.952381	1	
<b>STEP 4. Crisp membership function(RC TC) &gt; TH1 = 0.09.</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	1	1	1		
	Cloud	1	0	1		
	Unknowns	0	0	1		
<b>STEP 5. Dictionary-pair relationship, cond. prob. (TC RC)</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	0.833333333	0.857142857	0.352941		
	Cloud	0.166666667	0	0.058824		
	Unknowns	0	0.142857143	0.588235		
		1	1	1		
<b>STEP 6. Crisp membership function(TC RC) &gt; TH2 = 0.06 &lt;= TH1 = 0.09.</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	1	1	1		
	Cloud	1	0	0		
	Unknowns	0	1	1		
<b>STEP 7. OR{Crisp membership function(TC RC), Crisp membership function(RC TC)}</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	1	1	1		
	Cloud	1	0	1		
	Unknowns	0	1	1		
<b>STEP 8. Top-down (driven-by-prior knowledge) scrutiny of bottom-up (data-driven) "temporary correct" or "temporary non-correct" cells</b>						
		<i>Reference Classification (RC)</i>				
		EvergreenF	DeciduousF	Others		
<b>Test Classification (TC)</b>	Vegetation	1	1	1		
	Cloud	0	0	1		
	Unknowns	0	0	1		

Table 7-5. 8-step guideline for best practice in the identification of a dictionary-pair relationship based on a hybrid combination of prior beliefs, if any, with frequentist inference.





## **8 Manuscript 5 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01932): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 2: Validation**

### **Motivation and Contributions to the Dissertation**

Among the original expert systems (prior knowledge-based decision trees) for color naming presented in Chapter 3 (Technical report 1) and adopted by an Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program was designed and implemented to provide multi-spectral (MS) reflectance space hyperpolyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time (with a computational complexity increasing linearly with the image size). The multidisciplinary background of color naming was proposed and discussed in the Part 1 of this paper (Chapter 7, Manuscript 4) where, first, an original hybrid (combined deductive and inductive) guideline was proposed to identify a categorical variable-pair relationship and, second, an original quantitative measure of categorical variable-pair association was presented. In the present Part 2 (Chapter 8, Manuscript 5) an off-the-shelf SIAM lightweight computer program was submitted to an intergovernmental Group on Earth Observations (GEO)'s Stage 4 validation, by independent means on a large-scale EO image time-series, for systematic ESA EO Level 2 product generation.

Chapter 8 (Manuscript 5) features several degrees of novelty. First, it presents a novel protocol suitable for wall-to-wall thematic map quality assessment without sampling, where the test and reference thematic maps share the same spatial extent and spatial resolution, but whose map legends can differ in agreement with Chapter 8 (Manuscript 5). Second, it provides several instantiations of the categorical variable-pair relationship and quantitative measure of categorical variable-pair association proposed in Chapter 7 (Manuscript 4). Last but not least, it provides a non-trivial solution to the important question of general interest: if the “ultimate” accuracy of reference map A is validated in comparison with “absolute” ground-truth while the accuracy of test map B is assessed in relative terms of agreement in comparison with reference map A, what is the inferred “ultimate” accuracy of test map B in comparison with “absolute” ground-truth?

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 8 (Manuscript 5) are color filled.

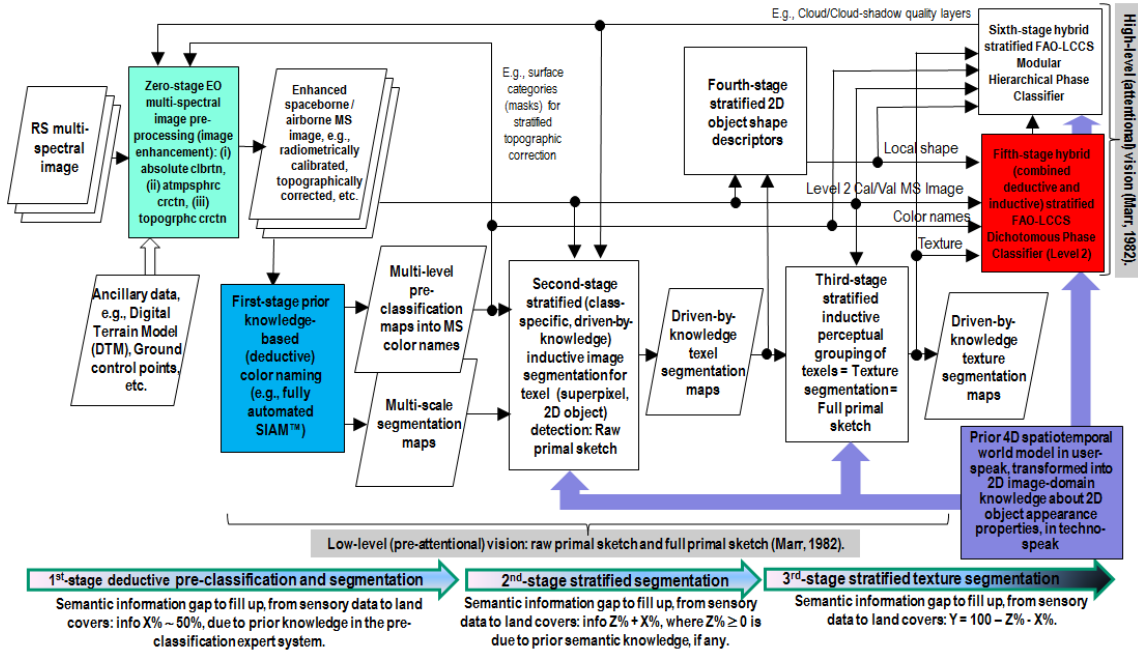


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 8 (Manuscript 5).



## Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation – Part 2: Validation

Andrea Baraldi<sup>a,c,\*</sup>, Michael Laurence Humber<sup>b</sup>, Dirk Tiede<sup>c</sup> and Stefan Lang<sup>c</sup>

<sup>a</sup> Department of Agricultural Sciences, University of Naples Federico II, Portici (NA), Italy.

<sup>b</sup> Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA.

<sup>c</sup> Department of Geoinformatics – Z\_GIS, University of Salzburg, Salzburg 5020, Austria.

\*Corresponding author. Email: andrea6311@gmail.com

### Abstract

The European Space Agency (ESA) defines an Earth Observation (EO) Level 2 product as a multi-spectral (MS) image corrected for geometric, atmospheric, adjacency and topographic effects, stacked with its data-derived scene classification map (SCM) whose legend includes quality layers such as cloud and cloud-shadow. No ESA EO Level 2 product has ever been systematically generated at the ground segment. To contribute toward filling an analytic and pragmatic information gap from EO big sensory data to the ESA EO Level 2 product, a Stage 4 validation (*Val*) of an off-the-shelf Satellite Image Automatic Mapper (SIAM) lightweight computer program for prior knowledge-based MS color naming was conducted by independent means. A time-series of annual Web-Enabled Landsat Data (WELD) image composites of the conterminous U.S. (CONUS) was selected as input dataset. The annual SIAM-WELD maps of the CONUS were validated in comparison with the U.S. National Land Cover Data (NLCD) 2006 map. These test and reference maps share the same spatial resolution and spatial extent, but their map legends are not the same and must be harmonized. For the sake of readability this paper is split into two. The previous Part 1 – Theory provided the multidisciplinary background of *a priori* color naming and proposed, first, an original guideline to identify a categorical variable-pair relationship based on a hybrid (combined deductive and inductive) inference approach; second, an original measure of categorical variable-pair association. The present Part 2 – Validation presents and discusses Stage 4 *Val* results collected from the test SIAM-WELD map time-series and the reference NLCD map by an original protocol for wall-to-wall thematic map quality assessment without sampling, where the test and reference maps share the same spatial extent and spatial resolution, but whose legends can differ in agreement with the Part 1. Conclusions are that the SIAM-WELD maps instantiate a Level 2 SCM product whose legend is the Food and Agriculture Organization of the United Nations – Land Cover Classification System (LCCS) taxonomy at the Dichotomous Phase (DP) Level 1 (vegetation/non-vegetation), Level 2 (terrestrial/aquatic) or superior LCCS level.

### Keywords

Artificial intelligence; binary relationship; Cartesian product; color naming; connected-component multi-level image labeling; deductive inference; Earth observation; land cover class taxonomy; high- (attentive) and low-level (pre-attentive) vision; hybrid inference; image segmentation; inductive inference; machine learning-from-data; outcome and process quality indicators; radiometric calibration; remote sensing; thematic map comparison; two-way contingency table; unsupervised data discretization/vector quantization.

### 8.1 Introduction

For the sake of readability this paper is split into two. The preliminary Part 1 – Theory advanced the thesis that a necessary not sufficient pre-condition for a yet-unfulfilled Global Earth Observation System of Systems (GEOSS) development, promoted by the intergovernmental Group on Earth Observations (GEO) (GEO 2005), is the systematic transformation at the ground segment of multi-source single-date multi-spectral (MS) imagery into a European Space Agency (ESA) EO Level 2 product (ESA 2015), instantiated as follows. (i) The MS image is corrected for geometric, atmospheric, adjacency and topographic effects (ESA 2015; CNES 2016). (ii) The enhanced MS image is stacked with its data-derived Scene Classification Map (SCM) (ESA 2015; CNES 2016). (iii) A general-purpose, user- and application-independent SCM's legend agrees with the 3-level 9-class “augmented” Dichotomous Phase (DP) taxonomy of the Food and Agriculture Organization of the United Nations (FAO) – Land Cover Classification System (LCCS) (Di Gregorio and Jansen 2000).



Such an “augmented” land cover (LC) class taxonomy encompasses the canonical 3-level 8-class LCCS-DP legend in addition to a thematic layer “other” or “rest of the world” which includes, for example, quality layers such as cloud and cloud-shadow, see Fig. 8-1. (iv) A GEO Stage 4 validation of the ESA EO Level 2 outcome and process is considered mandatory to comply with the GEO Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements (GEO-CEOS 2010). By definition a GEO Stage 3 *Val* requires that “spatial and temporal consistency of the product with similar products are evaluated by independent means over multiple locations and time periods representing global conditions. In Stage 4 *Val*, results for Stage 3 are systematically updated when new product versions are released and as the time-series expands” (GEO-CEOS WGCV 2015).

No ESA EO Level 2 product has ever been systematically generated at the ground segment (ESA 2015; CNES 2016). To contribute toward filling an analytic and pragmatic information gap from multi-source EO big data to ESA EO Level 2 product, the primary goal of this interdisciplinary study was to undertake an original (to the best of these authors’ knowledge, the first) outcome and process Stage 4 *Val* of an expert system for top-down (deductive) MS reflectance space hyperpolyhedralization, specifically, an off-the-shelf Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program. Implemented in operating mode in the C/C++ programming language, the SIAM application software is “lightweight” because it runs automatically (without human-machine interaction), in near real-time (it is non-iterative, either one-pass or two-pass, with a computational complexity increasing linearly with the image size) and in tile streaming mode (it requires a fixed runtime memory occupation) (Baraldi et al. 2006, 2010a, 2010b, 2010c, 2011, 2012a, 2012b, 2012c, 2012d, 2013, 2015, 2016; Baraldi and Humber 2015). In addition to running on laptop and desktop computers the SIAM lightweight computer program is eligible for use in a mobile software application. In addition to running on laptop and desktop computers the SIAM lightweight computer program is eligible for use in a mobile software application. Eventually provided with a mobile user interface, a mobile software application is a lightweight computer program specifically designed to run on web browsers and mobile devices, such as tablet computers and smartphones. The core of the non-iterative SIAM software pipeline is a one-pass prior knowledge-based decision tree (expert system) for MS reflectance space hyperpolyhedralization (quantization, partitioning) into static (non-adaptive-to-data) color names, see Fig. 8-2. Presented in the remote sensing (RS) literature where enough information was provided for the implementation to be reproduced (Baraldi et al. 2006), the SIAM expert system for color naming is followed by a well-posed two-pass superpixel detector in the color map-domain (Dillencourt et al. 1992; Sonka et al. 1994) and a per-pixel vector quantization (VQ) error assessment for VQ quality assurance, in agreement with the QA4EO *Val* guidelines, refer to Fig. 7-4 in the Part 1.

There is a long history of prior knowledge-based MS reflectance space partitioners for static color naming developed but never validated by space agencies, public organizations and private companies for use in hybrid EO-IUSs in operating mode. EO value-adding products and services targeted by existing hybrid EO-IUSs conditioned by static color naming encompass a large variety of low-level EO image enhancement tasks (Ackerman et al. 1998; Luo et al. 2008; Lück and van Niekerk 2016; Richter and D. Schläpfer 2012a; Richter and D. Schläpfer 2012b; Baraldi, Humber and Boschetti 2013; Baraldi and Humber 2015; Dorigo et al. 2009; Vermote and Saleous 2007; DLR and VEGA 2011; Lück and van Niekerk 2016; Baraldi et al. 2010c; Despini et al. 2014) and high-level EO image understanding applications (Baraldi et al. 2015; DLR and VEGA 2011; Lück and van Niekerk 2016; Muirhead and Malkawi 1989; Simonetti et al. 2015; GeoTerraImage 2015; Arvor et al. 2016; Boschetti et al. 2015; Baraldi et al. 2010a, 2010b). The potential impact on existing or future hybrid EO-IUSs in operating mode of an original (to the best of these authors’ knowledge, the first) outcome and process Stage 4 *Val* of an off-the-shelf SIAM lightweight computer program for prior knowledge-based MS reflectance space hyperpolyhedralization, superpixel detection and per-pixel VQ quality assessment is expected to be relevant.

To comply with the GEO Stage 4 *Val* requirements an off-the-shelf SIAM application had to be validated by independent means on a radiometrically calibrated EO image time-series at large spatial extent. The open-access U.S. Geological Survey (USGS) 30 m resolution Web Enabled Landsat Data (WELD) annual composites of the conterminous U.S. (CONUS) for the years 2006 to 2009, radiometrically *Cal* into top-of-atmosphere reflectance (TOARF) values (Roy et al. 2010; Homer et al. 2004; WELD 2016) was identified as a viable input dataset. The 30 m resolution 16-class U.S. National Land Cover Data (NLCD) 2006 map, delivered in 2011 by the U.S. Geological Survey (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC) (Vogelmann et al. 1998, 2001; Wickham et al. 2010; Wickham et al. 2013; Xian and Homer 2010; EPA 2007), was selected as the reference thematic map at the CONUS spatial extent. The 16-class NLCD map legend is described in Table 8-1. To account for typical non-stationary geospatial statistics, the NLCD 2006 thematic map was partitioned into 86 Level III ecoregions of North America collected from the Environmental Protection Agency (EPA) (EPA 2013; Griffith and Omernik 2009).





In the proposed experimental framework the test SIAM-WELD color map time-series and the reference NLCD 2006 map share the same spatial extent and spatial resolution, but their map legends are not the same. These working hypotheses are neither trivial nor conventional in the RS literature, where thematic map quality assessments typically adopt a sampling strategy, either random or non-random, and assume that the test and reference thematic map dictionaries coincide (Stehman and Czaplewski 1998). Starting from a stratified random sampling protocol presented in literature (Baraldi et al. 2013), the present Part 2 – Validation proposes an original protocol for wall-to-wall comparison without sampling of two thematic maps featuring the same spatial extent and spatial resolution, but whose legends can differ. This novel protocol incorporates two original contributions of the Part 1 where, first, a hybrid (combined deductive and inductive) guideline was proposed to streamline a human decision maker in the identification of a binary relationship  $R: A \Rightarrow B$  between two univariate categorical variables  $A$  and  $B$  of the same population. This is an inherently ill-posed (equivocal, subjective) *information-as-data-interpretation* process (Capurro and Hjørland 2003) belonging to the multi-disciplinary domain of cognitive science, refer to Fig. 7-13 in the Part 1. Second, version 2 of a categorical variable-pair association index (CVPAI2)  $\in [0, 1]$  was proposed to cope with the entity-relationship conceptual model shown in Fig. 7-17 of the Part 1.

The rest of the present Part 2 is organized as follows. Chapter 8.2 describes materials including the SIAM computer program, the time-series of annual WELD composites, the reference NLCD 2006 map and the EPA Level III ecoregion map of North America. Methods, specifically, an original protocol to compare without sampling the test SIAM-WELD and reference NLCD 2006 maps of the CONUS, whose map legends do not coincide and must be harmonized (reconciled, associated, translated (Ahlqvist 2005)), is proposed in Chapter 8.3. Experimental results are presented in Chapter 8.4 and discussed in Chapter 8.5. Conclusions are reported in Chapter 8.6.

## 8.2 Materials

Presented in the RS literature, four alternative implementations of a prior knowledge-based decision tree for static MS reflectance space hyperpolyhedralization onto static color names were compared for model selection. (i) The year 2006 SIAM decision tree presented in Baraldi et al. (2006). (ii) The static decision tree for Spectral Classification of surface reflectance signatures (SPECL), proposed by Dorigo et al. (2009), see Table 7-4 in the Part 1, and implemented by the Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software product (Richter and D. Schläpfer 2012a, 2012b). (iii) The static decision tree for Single-Date Classification (SDC), proposed by Simonetti et al. (2015). (iv) The Canada Centre for Remote Sensing (CCRS) spectral decision tree shown in Fig. 7-16 of the Part 1 (Luo et al. 2008). Whereas the SDC, SPECL and CCRS decision trees declare their applicability to Landsat images exclusively, SIAM claims its scalability to MS imaging sensors featuring different spectral resolution specifications, see Table 7-1 in the Part 1. Moreover, the SIAM decision tree outperforms its counterparts in terms of spectral quantization capability, parameterized by the total number of detected color names, equal to 96 for the 7-band Landsat-like SIAM (L-SIAM) subsystem, see Table 7-1 in the Part 1, versus 13, 19 and 7 color names detected in Landsat images by the SDC, SPECL (see Table 7-4 in the Part 1) and CCRS (see Fig. 7-16 in the Part 1) decision trees respectively. To explain their broad differences in terms of number of detected color names and scalability to MS imaging sensors whose spectral and spatial resolution specifications can vary, the four static spectral decision trees of interest were compared at the level of understanding of spectral information/knowledge representation (Marr 1982), irrespective of the implementation of the decision rule set (structural knowledge in the decision tree) and of the order of presentation of decision rules (procedural knowledge in the decision tree).

To investigate the scalability of an *a priori* knowledge-based spectral decision tree to varying MS imaging sensor specifications we started observing that, given a partition of a MS color space  $\mathfrak{R}^{\text{MS}}$  into a discrete and finite dictionary of hyperpolyhedra equivalent to color names  $\{1, \text{ColorDictionaryCardinality}\}$ , for any spatial unit  $x$ , either (0D) pixel, (1D) line or (2D) polygon defined according to the Open Geospatial Consortium (OGC) nomenclature (OGC 2015) and featuring a numeric  $\text{ColorValue}(x) \in \mathfrak{R}^{\text{MS}}$ , the photometric attribute of spatial unit  $x$  can be assigned with a categorical  $\text{ColorName}^* \in \{1, \text{ColorDictionaryCardinality}\}$ , such that membership  $m(\text{ColorValue}(x) | \text{ColorName}^*) = 1$ , see Equation (1-3) in the Part 1. In practice any prior knowledge-based spectral decision tree for color naming can work at the sensor spatial resolution whatever it is, i.e., it can work pixel-based irrespective of the spatial resolution of the imaging sensor.

Since they are independent of the spatial resolution of the imaging sensor, static decision trees for color naming depend on spectral resolution specifications exclusively. Inter-sensor differences in spectral resolution can vary from minor differences in a band-specific sensitivity curve to the major lack of a whole spectral channel. To gain robustness to changes



in spectral resolution specifications, the necessary not sufficient pre-condition for spectral rules is to infer “strong” (robust and reliable) conjectures based on the redundant convergence of multiple independent sources of spectral evidence, each of which is individually “weak”. This rationale is alternative to, for example, pruning of redundant processing elements in distributed processing systems such as multi-layer perceptrons (Bishop 1995; Cherkassky and Mulier 1994). If this diagnosis holds true, i.e., redundancy of spectral evidence is a value-added of spectral rules to scale to varying spectral resolutions, then information redundancy of a spectral rule is expected to increase monotonically with the number of independent spectral terms.

In a MS reflectance space, any target family of LC class-specific spectral signatures is a multivariate data distribution (envelope, hyperpolyhedron, manifold). Like a vector quantity has two characteristics, a magnitude and a direction, any LC class-specific MS manifold is characterized by a multivariate shape and a multivariate intensity, see Fig. 8-2. Hence, spectral information redundancy required to gain robustness to changes in spectral resolution specifications can regard the modelling of both the MS shape and MS intensity information components of a target MS manifold. Among the spectral decision trees being compared, only the SIAM decision tree adopts two different sets of spectral rules to model the MS shape and the MS intensity as two independent spectral information components of a target manifold of MS signatures. On the contrary, in the SDC, SPECL and CCRS decision trees MS shape and MS intensity properties are modeled simultaneously. For example, a typical SDC spectral rule applied to a Landsat pixel vector, radiometrically calibrated into a TOARF value in range [0, 1] in each Landsat band 1 to 6, is

If  $NDVI < 0.5$  and  $NIR (= \text{Landsat band } 4) \geq 0.15$  then do something else do otherwise.

In this spectral decision rule, the normalized difference vegetation index,  $NDVI = (NIR - Red) / (NIR + Red)$ , where  $NIR = \text{Landsat band } 4$  and  $Red = \text{Landsat band } 3$ , is a well-known spectral index, whose unbounded version is the band ratio  $NIR/Red$ . For example, band ratios are scalar spectral indexes widely employed in the SPECL decision tree, see Table 7-4 in the Part 1, and in the CCRS decision tree shown in Fig. 7-16 of the Part 1 (Luo et al. 2008). Any scalar spectral index, either normalized band difference or band ratio, is conceptually equivalent to the slope of a tangent to the spectral signature in one point. This spectral slope is a MS shape descriptor independent of the MS intensity, i.e., infinite functions with different intensity values can feature the same tangent value in one point. Although appealing due to its conceptual and numerical simplicity (Liang 2004), any scalar spectral index is unable *per se* to represent either the multivariate shape information or the multivariate intensity information component of a MS signature. Intuitively a scalar spectral index causes a dramatic N-to-1 loss in spectral resolution by reducing an N-channel MS image to a panchromatic image. No photointerpreter whose objective is a single LC class detection, e.g., vegetation detection, would typically consider a panchromatic image as informative as a MS image or simple enough to be considered a binary image, e.g., vegetation/non-vegetation, where a simple crisp intensity thresholding criterion can be applied for target detection. In practice no univariate or multivariate spectral index is representative of the multivariate shape and multivariate intensity information components of a MS manifold, see Fig. 8-2. This obvious but not trivial observation explains why, in spectral pattern recognition applications, lossy scalar spectral indexes are ever-increasing in number and variety in the endless search for yet-another scalar spectral index, supposedly more informative (Liang 2004; Baraldi et al. 2010a, 2010b). In the SDC rule reported above, the first spectral term,  $NDVI < 0.5$ , constrains a MS shape attribute; it is employed in logical combination with a second spectral term, where a MS intensity value is constrained as  $NIR \geq 0.15$ . The conclusion is that, unlike SIAM's, neither the SDC nor the SPECL nor the CCRS decision tree decompose a target MS manifold into its multivariate shape and multivariate intensity information components to make each information component easier to be investigated by multivariate data analysis. In each of its two independent sets of spectral rules for MS shape and MS intensity modelling SIAM pursues redundancy of spectral terms as a value-added for spectral scalability. Possible combinations of these two independent sets of spectral rules make the SIAM decision tree implementations, starting from that proposed in pseudo-code in (Baraldi et al. 2006), capable of representing the multivariate shape and multivariate intensity information components of a target MS hyperpolyhedron, neither necessarily convex nor connected, as a converging combination of many independent  $j$ th-variable functions, with  $j \in \{1, \text{total number } N \text{ of spectral channels}\}$ . Multivariate data statistics are known to be more informative than a sequence of univariate data statistics. For example, maximum likelihood data classification, accounting for multivariate data correlation and variance (covariance), is typically more accurate than parallelepiped data classification whose rectangular decision regions, equivalent to a concatenation of univariate data constraints, poorly fit multivariate data in the presence of bivariate cross-correlation (Lillesand and Kiefer 1979). In the RS common practice, thanks to its spectral redundancy of multivariate data statistics the “master” 7-band Landsat-like SIAM (L-SIAM) decision tree can be down-scaled to cope with “slave” MS imaging sensors whose spectral resolution is



inferior to, but overlaps with, Landsat's, see Table 7-1 in the Part 1 (Baraldi et al. 2010a, 2010b).

Based on this decision tree model comparison these authors concluded that the SIAM's peculiar design for multivariate spectral information representation and the SIAM implementation complexity, superior in terms of number of rules and number of terms per rule, appeared sufficient to justify the SIAM claims of a finer spectral quantization capability and a superior spectral scalability. Based on these conclusions an off-the-shelf SIAM application software was selected and considered worth of a Stage 4 *Val* to comply with the QA4EO *Call/Val* requirements, refer to Chapter 8.1.

To pursue a Stage 4 *Val* of the SIAM application the 30 m resolution U.S. NLCD 2006 map was selected as reference LC map. When this experimental work was conducted the NLCD 2006 map was the most recent release of the U.S. NLCD map series developed by the USGS EDC (Vogelmann et al. 1998, 2001; Wickham et al. 2010; Wickham et al. 2013; Xian and Homer 2010; EPA 2007; Homer et al. 2015), see Fig. 8-3. By now the U.S. NLCD map series comprises the NLCD 1992, 2001, 2001 Version 2.0, 2006 (released in 2011) and 2011 (released in 2015) editions. The timeliness from image collection to NLCD product delivery, which includes information layers such as tree cover fraction and impervious fraction, has steadily decreased from the about 5 years of the initial NLCD product. Made available for public access in a provisional version in Feb. 2011, the NLCD 2006 map was "based primarily on the unsupervised classification" of Landsat-5 Thematic Mapper <sup>TM</sup> and "Landsat-7 Enhanced Thematic Mapper (ETM)+ images acquired in circa 2006" (Xian and Homer 2010). It is a 30 m resolution raster product in the Albers Equal Area projection, which is the reference standard for continental scale cartography produced by U.S. agencies. Its legend consists of 16 LC classes defined according to the Level II LC classification system, refer to Table 8-1 (EPA 2007). Validation of the NLCD 2006 map provided an overall accuracy (OA) of 78%, which increased to 84% when the 16 LC classes were aggregated into 9 LC classes (Wickham et al. 2010; Wickham et al. 2013; Stehman et al. 2008). Noteworthy, these 9 LC classes are conceptually equivalent to an "augmented" 9-class LCCS-DP taxonomy, refer to Chapter 8.1. The validated NLCD 2006 map's OA values of 78% and 84% with, respectively, a 16 and a 9 LC class legend can be considered state-of-the-art. For example, they are superior to Oas featured by national-scale maps recently generated by pixel-based random forest classifiers from monthly WELD composites, whose OA is 65%–67% using 22 detailed classes and 72%–74% using 12 aggregated national classes (Wessels et al. 2016). In general renowned experts in Geographical Information Science (GIScience) suggest that "the widely used target accuracy of 85% may often be inappropriate and that the approach to accuracy assessment adopted commonly in RS can be pessimistically biased" (Foody 2006, 2016).

Based on these observations we considered the NLCD 2006 map's official OA estimate of 84% realistic and state-of-the-art at the U.S. national scale when the 3-level 9-class "augmented" LCCS-DP legend is adopted. As a consequence the reference NLCD 2006 map was considered suitable for a Stage 4 *Val* of the SIAM application software whose Level 2 SCM product's legend had to comply with the "augmented" 9-class LCCS-DP taxonomy. When a test SIAM map and a reference NLCD map share the same 30 m spatial resolution and spatial extent, then they can be compared wall-to-wall without sampling. Since no conventional sampling-theory procedure is employed (Lunetta and Elvidge 1999), a wall-to-wall OA(Test SIAM  $\Rightarrow$  Reference NLCD) estimate in range  $[0, 100\%]$  is provided with a confidence interval (degree of uncertainty in measurement),  $\pm \delta \in [0, 100\%]$ , whose estimation is considered mandatory by the QA4EO *Val* guidelines, equal to  $\pm \delta = 0\%$ .

From a statistic standpoint the aforementioned experimental work specifications imply the following. Let's identify with OA(Test SIAM  $\Rightarrow$  "Ultimate" GroundTruth)  $\in [0, 100\%] = 100\% - \text{Mismatch}(\text{Test SIAM} \Rightarrow \text{"Ultimate" GroundTruth})$  the OA of an EO data-derived SIAM test map with respect to an "ultimate" (ideal) ground truth and with OA(Test SIAM  $\Rightarrow$  Reference NLCD)  $\pm 0\% = 100\% - \text{Mismatch}(\text{Test SIAM} \Rightarrow \text{Reference NLCD}) \pm 0\%$  the overall degree of agreement provided with its confidence interval of a test SIAM map compared wall-to-wall without sampling with a reference NLCD map at the same spatial resolution and spatial extent. It is known that

$$\text{OA}(\text{Reference NLCD 2006, "augmented" LCCS-DP 9 classes} \Rightarrow \text{"Ultimate" GroundTruth 2006, "augmented" LCCS-DP 9 classes}) = 84\% = 100\% - \text{Mismatch}(\text{Reference NLCD 2006, "augmented" LCCS-DP 9 classes} \Rightarrow \text{"Ultimate" GroundTruth 2006, "augmented" LCCS-DP 9 classes}) = 100\% - 16\%.$$

Similarly,

$$\text{OA}(\text{Reference NLCD 2006, NLCD 16 classes} \Rightarrow \text{"Ultimate" GroundTruth 2006, NLCD 16 classes}) = 78\% = 100\% - \text{Mismatch}(\text{Reference NLCD 2006, NLCD 16 classes} \Rightarrow \text{"Ultimate" GroundTruth 2006, NLCD 16 classes}) = 100\% - 22\%.$$

Based on the superposition principle it is possible to write



$OA(\text{Test SIAM} \Rightarrow \text{“Ultimate” GroundTruth}) \in [0, 100\%] = OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD} \Rightarrow \text{“Ultimate” GroundTruth}) = \{OA(\text{Reference NLCD} \Rightarrow \text{“Ultimate” GroundTruth}) \pm \text{Mismatch}(\text{Test SIAM} \Rightarrow \text{Reference NLCD}) \pm 0\%\} \in [\text{Worst Case, Best Case}]$ , where Worst Case =  $\max\{0\%, \text{Lower Bound}\}$  and Best Case =  $\min\{100\%, \text{Upper Bound}\}$ , with Lower Bound  $\leq$  Upper Bound  $\in [0\%, 100\%]$ ,

where

Lower Bound =  $[OA(\text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes}) - \text{Mismatch}(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%] =$

$[100\% - \text{Mismatch}(\text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes}) - (100\% - OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%)] =$

$[OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\% - \text{Mismatch}(\text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes})] =$

$[OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\% - 16\%]$ ,

and

Upper Bound =  $[OA(\text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes}) + \text{Mismatch}(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%] =$

$[84\% + (100\% - OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%)] = [184\% - OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%]$ .

To recapitulate, when the “Ultimate” GroundTruth adopts an “augmented” 9-class LCCS-DP legend, it is expected that

$OA(\text{Test SIAM} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$

$[\max\{0\%, OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\% - 16\%\}, \min\{100\%, 184\% - OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes}) \pm 0\%\}]$ . (2-1)

Similarly, when the “Ultimate” GroundTruth adopts a 16-class NLCD legend, then it is expected that

$OA(\text{Test SIAM} \Rightarrow \text{“Ultimate” GroundTruth 2006, NLCD 16 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$

$[\max\{0\%, OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\% - 22\%\}, \min\{100\%, 178\% - OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\%\}]$ . (2-2)

Equation (2-1) and Equation (2-2) are useful because, first, they highlight the undisputable fact that *per se* the NLCD reference map is not a “ground truth” for the SIAM test map, but only a reference baseline for comparison purposes. Second, they support the validity of this experimental project by showing that a summary statistic  $OA(\text{Test SIAM} \Rightarrow \text{“Ultimate” GroundTruth}) = OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD} \Rightarrow \text{“Ultimate” GroundTruth})$  can be inferred from an estimated  $OA(\text{Test SIAM} \Rightarrow \text{Reference NLCD 2006}) \pm 0\%$  known that  $OA(\text{Reference NLCD 2006} \Rightarrow \text{“Ultimate” GroundTruth 2006})$  is equal to 84% or 78% when the “Ultimate” GroundTruth adopts an “augmented” 9-class LCCS-DP legend or the 16-class NLCD legend respectively.

Supported by NASA and distributed by the USGS EDC (WELD 2016), the annual WELD composites for years 2006, 2007, 2008 and 2009 were selected as a large-scale radiometrically calibrated EO image time-series required by a Stage 4 *Val* of the SIAM application software in comparison with the reference NLCD 2006 map, see Fig. 8-3. Each annual WELD composite consists of approximately 8,000 Landsat-5/7 image acquisitions per year over the CONUS, starting from year 2003 to 2012. The current WELD processing workflow requires as input Landsat sensor series L1T images with cloud cover  $\leq 20\%$ . The WELD mosaic of the CONUS encompasses 501 fixed location tiles defined in the Albers Equal Area projection. Each tile is  $5000 \times 5000$  pixels in size, equal to  $150 \times 150$  km (Homer et al. 2004). The Landsat sensor series L1T image geolocation error in the CONUS, including areas with substantial terrain relief, is less than 30 m ( $< 1$  pixel)





(Lee et al. 2004). The most recent Landsat data radiometric *Cal* expertise is employed in the WELD workflow to ensure harmonization and interoperability of multi-sensor Landsat image time-series, with a 5% absolute reflective band *Cal* uncertainty (Markham and Helder 2012), in agreement with the QA4EO's *Cal/Val* requirements (GEO-CEOS 2010). Fig. 8-4 shows the WELD 2006 composite over the CONUS, where TOARF values are depicted in true colors, with the WELD tiling scheme overlaid in white.

To account for typical non-stationary geospatial statistics, an inter-map statistical comparison on a stratified (masked) basis should be accomplished at a local spatial extent, where strata convey some geospatial criteria of land surface information invariance. The NLCD 2006 reference map was partitioned into Level III ecoregions of North America collected from the EPA (EPA 2013). There are 86 ecoregions across the CONUS, each ecoregion featuring similar ecological and climatic characteristics (Griffith and Omernik 2009). Distributed as vector data, the EPA Level III ecoregions were rasterized to 30 m resolution in the Albers Equal Area projection. Fig. 8-3 shows the NLCD 2006 map with boundaries of ecoregions overlaid in black.

### 8.3 Methods

A wall-to-wall comparison without sampling between the test SIAM-WELD map time-series and the reference NLCD 2006 map, sharing the same 30 m spatial resolution at the CONUS spatial extent, but whose legends A = DictionaryOfColorNames (see Table 8-2) and B = LegendOfObjectClassNames (see Table 8-1) do not coincide and must be harmonized, was designed and implemented for Stage 4 *Val* purposes. These working hypotheses differ from thematic map accuracy assessment protocols adopted by the large majority of the RS community, typically based on an either random or non-random sampling and/or a confusion matrix (CMTRX). A CMTRX is defined as a special case of a two-way contingency table (bivariate table),  $BIVRTAB = FrequencyCount(A \times B)$  where  $A \times B$  is a 2-fold Cartesian product between two univariate categorical variables A and B estimated from the same population (Kuzera and Pontius 2008; Pontius and Connors 2006). In particular a CMTRX is square and sorted because the test and reference categorical variables A and B of the same population are required to be the same, to let the main diagonal guide the interpretation process, refer to the Part 1, Chapter 7.2.

In (Baraldi et al. 2014), a crisp thematic map assessment protocol was proposed based on probability sampling, a pair of test and reference thematic legends A and B that may differ, an overlapping area matrix (OAMTRX) (Beauchemin and Thomson 1997; Ortiz and Oliver 2006), whose spatial unit  $x$  is (0D) pixel, a set of thematic quantitative quality indicators ( $Q^2$ Is),  $TQ^2$ Is, extracted from the OAMTRX and a set of spatial  $Q^2$ Is ( $SQ^2$ Is) extracted from sub-symbolic image-objects in the multi-level map domain, where image-objects are either (0D) pixels, (1D) lines or (2D) polygons according to the OGC nomenclature (OGC 2015). Equivalent to an either square or non-square  $BIVRTAB = FrequencyCount(A \times B)$ , an OAMTRX is generated from the cross-tabulation of any possible pair of univariate categorical variables A and B, either coincident or not, estimated from the same geospatial population, refer to Part 1, Chapter 7.2 (Beauchemin and Thomson 1997; Ortiz and Oliver 2006). Whereas the construction of an OAMTRX is straightforward and non-controversial when the semantic labels of sampling units are crisp (hard), the method to construct an OAMTRX when semantic labels are soft (fuzzy) is not obvious at all, e.g., refer to (Kuzera and Pontius 2008). Hence, we focused on crisp OAMTRX instances, exclusively. To accomplish our work hypotheses, the crisp thematic map probability sampling protocol proposed in (Baraldi et al. 2014) was modified as follows.

- The original hybrid eight-step guideline proposed in the Part 1, Chapter 7.4 was adopted to streamline the inherently subjective selection by human experts of a binary relationship  $R: A = DictionaryOfColorNames \Rightarrow B = LegendOfObjectClassNames \subseteq A \times B$  that guides the interpretation process of a crisp  $OAMTRX = FrequencyCount(A \times B)$ , see Table 7-3 in the Part 1.
- Given a binary relationship  $R: A = DictionaryOfColorNames \Rightarrow B = LegendOfObjectClassNames$  that guides the interpretation process of a crisp  $OAMTRX = FrequencyCount(A \times B)$ , a novel  $CVPAl2(R: A \Rightarrow B)$  formulation was adopted as a relaxed version of the  $CVPAl1$  formulation proposed in (Baraldi et al. 2014), refer to the Part 1, Chapter 7.5.
- Traditional 30 m resolution Landsat image classifiers are pixel-based, due to the lack of contextual information in 30 m resolution imagery. Hence, in the 30 m resolution WELD composites, the most informative planar entity is (0D) pixel, rather than image-object, either (1D) line or (2D) polygon (OGC 2015). As a consequence, for the sake of simplicity, in the present thematic map comparison image-object-based  $SQ^2$ Is were omitted. Rather, the following pixel-based  $TQ^2$ Is



were estimated from the crisp OAMTRX = FrequencyCount(A × B) estimated wall-to-wall with spatial unit x equal to pixel.

- An OA(OAMTRX) ± 0% was computed in line with (Baraldi et al. 2014). This OA estimate is guided by the binary relationship  $R: A = \text{DictionaryOfColorNames} \Rightarrow B = \text{LegendOfObjectClassNames}$  identified and community-agreed upon in advance, refer to this text above. In an OAMTRX estimated from a wall-to-wall inter-map comparison, where no sample data is investigated, any adopted TQ<sup>2</sup>I features a degree of uncertainty in measurement equal to ± 0%, e.g., see Equation (2-1).
- User's and producer's accuracies, computed in (Baraldi et al. 2014), were replaced by class-conditional probabilities,  $p(r | t)$  of reference class  $r$  given test class  $t$  and, vice versa,  $p(t | r)$  of test class  $t$  given reference class  $r$ , with  $r = 1, \dots, RC$ , and  $t = 1, \dots, TC$ , where  $RC = |B| = b = \text{ObjectClassLegendCardinality}$  and  $TC = |A| = a = \text{ColorDictionaryCardinality}$  are the total numbers of reference and test classes respectively,

The proposed ensemble of summary statistics, specifically, CVPAl2(R:  $A \Rightarrow B \subseteq A \times B$ ), OA(OAMTRX = FrequencyCount(A × B)) and class-conditional probabilities(OAMTRX), is an original minimally dependent and maximally informative (mDMI) set (Si Liu et al. 2011; Peng et al. 2005) of TQ<sup>2</sup>Is, to be jointly maximized according to the Pareto formal analysis of multi-objective optimization problems (Boschetti et al. 2004), refer to the Part 1, Chapter 7.1.

#### 8.4 Validation session

*Val* is the process of assessing (GEO-CEOS WGCV 2015), by independent means, the quality of an information processing system by means of an mDMI set (Si Liu et al. 2011; Peng et al. 2005) of community-agreed outcome and process (OP) Q<sup>2</sup>Is provided with a degree of uncertainty in measurement, ±δ, in compliance with the QA4EO *Call/Val* guidelines (GEO-CEOS 2010). The SIAM-WELD data mapping process and outcome were validated by a human expert independent of the present authors (refer to Acknowledgments). This independent human expert accomplished the following tasks. (I) Run without user interaction an off-the-shelf SIAM application upon the 30 m resolution annual WELD 2006 to 2009 composites of the CONUS. (II) Overlap the test SIAM-WELD annual map time-series with the reference NLCD 2006 map to generate instances of an OAMTRX = FrequencyCount(A × B) (Baraldi et al. 2014). (III) Estimate an mDMI set of OP-Q<sup>2</sup>Is, encompassing the following, refer to the Part 1, Chapter 7.1 (Baraldi and Humber 2015; Baraldi et al. 2013; Baraldi and Boschetti 2012a, 2012b). (i) Product effectiveness. Proposed outcome Q<sup>2</sup>Is (O-Q<sup>2</sup>Is) were the TQ<sup>2</sup>Is presented in Chapter 8.3: (a) CVPAl2(R:  $A \Rightarrow B \subseteq A \times B$ ); (b) OA(OAMTRX = FrequencyCount(A × B)), and (c) class-conditional probabilities  $p(r | t)$  and  $p(t | r)$  with reference class  $r = 1, \dots, RC = |B| = \text{ObjectClassLegendCardinality}$  and test class  $t = 1, \dots, TC = |A| = \text{ColorDictionaryCardinality}$ . (ii) Process efficiency. Proposed process Q<sup>2</sup>Is (P-Q<sup>2</sup>Is) were computation time and memory occupation. (iii) Process degree of automation, monotonically decreasing with the number of system's free-parameters to be user-defined. (iv) Process robustness to changes in the input dataset. For post-classification change/no-change detection (Lunetta and Elvidge 1999), the SIAM-WELD 2006 to 2009 maps were compared one another when one year apart. (v) Process robustness to changes in input parameters, if any. (vi) Process scalability, to keep up with changes in users' needs and sensor properties. (vii) Product timeliness, defined as the time between data acquisition and product generation. (viii) Product costs in manpower and computer power. In the present study the following definition is adopted: an information processing system can be considered in operating mode (ready-to-go) if it scores “high” in all of its OP-Q<sup>2</sup>I values, refer to the Part 1, Chapter 7.1.

For the sake of paper simplicity, the following decisions were undertaken.

- The SIAM-WELD 2006 color maps at fine (96) color names and intermediate (48) color names were compared with the NLCD 2006 map, while the SIAM map at coarse (18) color names was ignored, see Table 7-1 in the Part 1. This implied the following.
  - In the case of the SIAM-WELD 2006 map at fine discretization level, an OAMTRX instance consisted of test set A = 96 spectral categories as rows and reference set B = 16 NLCD classes as columns. Due to its excessive size, this OAMTRX instance cannot be shown in a technical paper. However, it is made available on an anonymous ftp site (SIAM-WELD-NLCD FTP 2016) and its TQ<sup>2</sup>I summary statistics are reported in the present paper.
  - In the case of the SIAM-WELD 2006 map at the intermediate color discretization level, an OAMTRX instance consisted of test set A = 48 spectral categories as rows and reference set B = 16 NLCD classes as columns. Due to its excessive size, this OAMTRX instance cannot be shown in a technical paper. Hence, the ensemble of 48 spectral categories were reassembled into 19 spectral macro-categories by the independent human expert, refer to



Table 8-2. This grouping of basic color names into spectral macro-categories of basic colors pertains to the inherently equivocal (subjective) domain of *information-as-data-interpretation*, refer to the Part 1, Chapter 7.4 (Capurro and Hjørland 2003). Among the 19 spectral macro-categories reassembled by the independent human expert, 16 macro-categories coincided exactly with one category in the initial set of 48 spectral categories. The reassembled macro-category "Others" included the SIAM intermediate spectral category "Unknowns" together with 24 of the 48 intermediate spectral categories covering "disturbances", such as cloud, smoke plume, fire front, etc., which are typically minimized or removed in an annual WELD composite. The reduced set of 19 spectral macro-categories is mutually exclusive and totally exhaustive, in line with the Congalton and Green requirements of a classification scheme (Congalton and Green 1999). A simplified OAMTRX instance of reduced size was generated. It consisted of a test set A = 19 spectral macro-categories as rows and a reference set B = 16 NLCD classes as columns. Due to its reduced size, this OAMTRX instance can be shown in the present paper, together with its TQ<sup>2</sup>I estimates.

- In agreement with the previous paragraph, the annual SIAM-WELD 2006 to 2009 maps at the SIAM intermediate discretization level of 48 color names were all reassembled into 19 spectral macro-categories, see Table 8-2.

#### 8.4.1 Verification of the Co-Registration Requirements for Pixel-based Inter-Map Comparison

In the requirements specification of RS projects dealing with per-pixel post-classification change/no-change detection, the required RS image co-registration error is typically < 1 pixel. For example, in (Lunetta and Elvidge 1999), it is recommended that the root-mean-square (RMS) error between any two-date images should not exceed 0.5 pixels.

In (Dai and Khorram 1998), simulated misregistration effects are investigated upon multi-temporal Landsat images of North Carolina across four study areas representative of land cover types: forest land, agricultural land, bare soil and urban/residential area. In these experiments, a registration accuracy < 1/5 of a pixel is considered necessary to achieve a land cover change detection error < 10%. This conclusion is more severe than the one-pixel co-registration constraint typically adopted in most change detection applications.

The annual WELD composites and the NLCD 2006 thematic map were derived from the same sensory dataset of Landsat LIT images acquired by the USGS EDC. It means that the SIAM-WELD 2006 pre-classification maps and the NLCD 2006 reference map were derived from the same sensory dataset. Hence, it is reasonable to assume that the co-registration error between these data-derived maps is negligible.

#### 8.4.2 Inter-Annual SIAM-WELD Map Comparisons for Years 2006 to 2009

The consistency across time and space of the annual SIAM-WELD 2006 to 2009 map time-series with a legend of 19 spectral macro-categories was investigated. Based on *a priori* knowledge of the multi-temporal pixel-based selection criteria adopted by the USGS EDC for the generation of annual WELD composites (refer to Chapter 8.2) and of the LC/LC change (LCC) dynamics in the real-world CONUS, a small percentage of LCC counts was expected to be detected one year apart at the CONUS spatial extent.

Table 8-3 shows percentages of the CONUS assigned to each SIAM-WELD spectral macro-category across the annual time-series. The green-as-*"Vegetation"* spectral categories are predominant (refer to the total vegetation statistic reported in Table 8-3), with an average 79% of the CONUS pixels, followed by color names such as brown-as-*"Bare soils or built-up"* (19% on average), followed by the remaining spectral macro-categories which, altogether, account for about 2%. The standard deviation through time of the occurrence of each SIAM-WELD spectral macro-category at the CONUS spatial extent is lower than 1%, with the exception of two vegetation spectral categories (specifically, aV\_HC and aV\_MC) where a larger variance can be attributed mostly to phenology. If a vegetation-through-time spectral variability due to changes in phenology affects the annual WELD composites then the data-derived SIAM-WELD color quantization maps will be affected by changes in phenology too. This diagnosis was verified as follows. Due to the limited availability of cloud-free Landsat observations at a generic pixel location per year, the Julian day of the year of the observation selected at a given location (pixel) of the WELD composite changes through years (Roy et al. 2010). This is illustrated in Fig. 8-5 where, at any fixed location across a target "ground-truth" area of deciduous forest in a pair of monthly August-November WELD composites, the SIAM spectral labels change significantly, but consistently with the phenological season. The same consideration holds when changes in phenology affect the annual WELD composites. This can explain the "high" intra-vegetation spectral variability observed by the SIAM vegetation macro-categories aV\_HC and aV\_MC in the tested time-series of annual WELD composites.



Non-stationary spatial phenomena occurring at the CONUS spatial extent can be overlooked by global statistics. To be captured spatial non-stationarities require local statistics such as class-conditional statistics described in Table 8-3. According to Chapter 8.3, for every pair of one annual SIAM-WELD test map with legend  $A = 19$  spectral macro-categories for year 2006 to 2009 overlapped at the CONUS spatial extent with a reference NLCD 2006 map with legend  $B = \text{NLCD } 16$  classes, the pair of summary statistics  $\text{CVPAI2}(R: A \Rightarrow B \subseteq A \times B) \in [0, 1.0]$  and  $\text{OA}(\text{OAMTRX} = \text{FrequencyCount}(A \times B)) \in [0, 1.0]$  should be maximized jointly. Shown as gray entry-pair cells in Table 8-4, the binary relationship  $R: A \Rightarrow B$  selected by the independent human expert featured a  $\text{CVPAI2}(R: A \Rightarrow B) = 0.6769$  while the  $\text{OA}(\text{OAMTRX}) = \text{OA}(\text{Test SIAM-WELD } 2006, 19 \text{ spectral macro-categories} \Rightarrow \text{Reference NLCD } 2006, \text{ NLCD } 16 \text{ classes}) = 96.88\% \pm 0\%$ . With a binary relationship  $R: A \Rightarrow B$  kept fixed, where  $\text{CVPAI2}(R: A \Rightarrow B) = 0.6769$ , the  $\text{OA}(\text{OAMTRX})$  estimate became equal to 97.02%, 96.69% and 96.75% for the annual SIAM-WELD map of year 2007 to 2009 compared with the reference NLCD 2006 map.

#### 8.4.3 Comparison of the SIAM-WELD 2006 and NLCD 2006 Thematic Maps

The SIAM-WELD 2006 test maps at intermediate (48) and fine (96) color quantization levels were compared with the NLCD 2006 reference map as described below.

**Test case A.** The SIAM-WELD 2006 map of the CONUS at the intermediate color discretization level reassembled into 19 spectral macro-categories is shown in Fig. 8-6. The OAMTRX instance generated from the overlap between the test SIAM-WELD 2006 map with legend  $A = 19$  spectral macro-categories at the CONUS spatial extent with a reference NLCD 2006 map with legend  $B = \text{NLCD } 16$  classes is shown in Table 8-4, where each cell reports a joint probability value  $p(\text{SIAM-WELD}_t, \text{NLCD}_r)$ ,  $r = 1, \dots, RC = |B| = 16$ , refer to Table 8-2, and  $t = 1, \dots, TC = |A| = 19$ , refer to Table 8-1. Gray entry-pair cells identify the binary relationship  $R: A \Rightarrow B \subseteq A \times B$  chosen by the independent human expert to guide the OAMTRX interpretation process. The distribution of these "correct" entry-pairs shows that every NLCD class overlaps with several discrete color types, with the exceptions of two SIAM-NLCD entry-pairs, specifically, entry-pair [SIAM spectral macro-category, NLCD class] = [MS white-as-"Snow" (SN, see Table 8-2), NLCD class "Perennial ice/snow" (PIS, see Table 8-1)] and entry-pair [SIAM spectral macro-category, NLCD class] = [MS blue-as-"Water or Shadow" (WA, see Table 8-2), NLCD class "Open water" (OW, see Table 8-1)], which are both characterized by a 1-1 matching relation. According to their specific definitions (refer to Table 8-1), anthropic NLCD classes, such as "Developed, Open Space" (DOS), "Developed, Low Intensity" (DLI), "Developed, Medium Intensity" (DMI) and "Developed, High intensity" (DHI), are a mixture of vegetated surfaces, impervious surfaces and bare soil, in agreement with the popular vegetation-impervious surface-soil model for urban ecosystem analysis (Ridd 1995). In agreement with their definitions, these NLCD classes overlap exclusively with the SIAM spectral macro-categories related to vegetation or bare soil. The NLCD class "Barren Land" (BL, see Table 8-1) overlaps with all of the SIAM spectral macro-categories related to bare soil. Noteworthy, according to Table 8-4, the NLCD class BL covers only 1.21% of the CONUS total surface. This is due to the NLCD 2006 definition of class BL (Rock/Sand/Clay), very restrictive with regard to the presence of vegetation, which has to account for less than 15% of total cover. The NLCD definition of class BL means that the NLCD classes "Shrub/Scrub" (SS) and "Grassland/Herbaceous" (GH, refer to Table 8-1) may feature a vegetated cover which accounts for 15% of total cover or more. The NLCD forest classes "Deciduous forest" (DF), "Evergreen Forest" (EF) and "Mixed forest" (MF, refer to Table 8-1) overlap with the SIAM's high and medium canopy-cover spectral macro-categories. The NLCD vegetation classes "Shrub/Scrub" (SS) and "Grassland/Herbaceous" (GH, refer to Table 8-1) overlap with the SIAM-WELD 2006 medium and low canopy-cover spectral macro-categories, but, in case of dry or sparse vegetation, also with some of the SIAM-WELD 2006 spectral macro-categories related to bare soil, namely, sbS\_1, SmS\_1 and aS (refer to Table 8-2). The overlap between the reference NLCD 2006 vegetation classes SS and GH and the test SIAM-WELD 2006 bare soil spectral macro-categories sbS\_1, SmS\_1 and aS is the only case of comprehensive (systematic) "semantic mismatch" recorded across the wall-to-wall SIAM-WELD 2006 and NLCD 2006 thematic map comparison. Hence, it is worth a deeper analysis in comparison with an "ultimate" ground truth. Reported in this section above, the NLCD 2006 definition of class "Barren Land" (BL, see Table 8-1) means that classes SS and GH may feature a vegetated cover which accounts for 15% of total cover or more. Two consequence of these definitions are that, whereas the NLCD class BL covers only 1.21% of the CONUS total surface, the NLCD vegetation classes SS and GH map the near totality of desert areas across the CONUS. Hence, there is a systematic "semantic mismatch" between the NLCD 2006 vegetation classes SS and GH and the SIAM-WELD 2006 bare soil spectral macro-categories across nearly all desert areas of the CONUS. Fig. 8-7 shows real-world examples of geographic locations mapped as vegetation classes "Scrub/Shrub" (SS) or "Grassland/Herbaceous" (GH) in the NLCD 2006 map (refer to Table 8-1), while they are mapped predominantly as the





bare soil spectral categories sbS\_1, SmS\_1 and aS in the SIAM-WELD 2006 map (refer to Table 8-2). For more comments about this systematic case of "conceptual mismatch" between the test SIAM-WELD and reference NLCD 2006 maps, refer to Fig. 8-10.

Additional inter-map overlaps highlighted by Table 8-4 reveal that the NLCD class "Pasture/Hay" (PH, see Table 8-1) occurs together with high and medium canopy-cover color types of the SIAM-WELD map. The NLCD class "Cultivated crops" (CC, see Table 8-1) matches with both spectral macro-categories MS green-as-"Vegetation" and MS brown-as-"Bare soil or built-up". Finally, the NLCD classes of wetland ("Woody Wetlands", WW, and "Emergent Herbaceous Wetland", EHW, see Table 8-1) overlap with the SIAM's vegetated spectral macro-categories or with spectral macro-category MS blue-as-"Water or Shadow" (WA, refer to Table 8-2).

As reported in Chapter 8.4.2, in the OAMTRX instance shown in Table 8-4 summary statistics are  $CVPAI2(R: A \Rightarrow B) = 0.6689$  and  $OA(OAMTRX) = OA(\text{Test SIAM-WELD 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) = 96.88\% \pm 0\%$ . As a consequence, according to Equation (2-2),

$$OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{"Ultimate" GroundTruth 2006, NLCD 16 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$$

$$[\max\{0\%, OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\% - 22\%\}, \min\{100\%, 178\% - OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\%\}] =$$

$$[\max\{0\%, 96.88\% \pm 0\% - 22\%\}, \min\{100\%, 178\% - 96.88\% \pm 0\%\}] =$$

$$[74.88\%, 81.12\%], \text{ with } CVPAI2(R: A = \text{SIAM dictionary of 19 color names} \Rightarrow B = \text{NLCD legend of 16 LC class names}) = 0.6769, \text{ hence the semantic information gap from sub-symbolic sensory data to symbolic NLCD classes left to be filled by further stages in the EO-IUS pipeline} = 1 - CVPAI2 = 0.3231, \text{ refer to the Part 1, Chapter 7.5. (2-3)}$$

When disagreements between the two reference and test maps were back-projected onto the WELD 2006 image domain, these specific WELD sites were photointerpreted by the independent human expert to provide an additional independent source of thematic evidence for Stage 4 *Val* of the SIAM-WELD 2006 test map. The large majority of the CONUS areas where the NLCD vegetation classes overlap with the SIAM spectral macro-category MS blue-as-"Water or Shadow" (WA, refer to Table 8-2) or, vice versa, where the SIAM vegetation spectral macro-categories overlap with the NLCD reference class "Open Water" (OW, refer to Table 8-1) were identified by the independent human photointerpreter as riparian zones. In practice, these riparian zones were labeled by the SIAM-WELD 2006 and NLCD 2006 maps in two different conditions of their annual surface status. Also in this case the SIAM-WELD labeling appears consistent with the human photointerpretation of the WELD composite, irrespective of the semantic disagreement between this SIAM-WELD labeling and the NLCD 2006 reference map.

Based on evidence collected by the independent photointerpreter with regard to the systematic "conceptual mismatches" between the test SIAM-WELD 2006 map and the reference NLCD 2006 map across nearly all desert areas and riparian zones of the CONUS, validated conclusions were twofold. First, according to Equation (2-1) where the NLCD 2006 reference map is not a "ground truth" for the SIAM-WELD 2006 test map, but only a reference baseline for comparison purposes, inter-map "conceptual mismatches" should not be misinterpreted as mapping errors by the test SIAM-WELD 2006 map with respect to an "ultimate" ground truth. Second, an  $OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{"Ultimate" GroundTruth 2006, NLCD 16 classes})$  summary statistic, inferred to belong to range [74.88%, 81.12%] according to Equation (2-3), was considered likely to be lying closer to its upper bound.

**Test case B.** To reveal the inherent ill-posedness of any inter-dictionary "conceptual matching" (refer to the Part 1, Chapter 7.4), one co-author of this paper, different from the independent human expert (refer to Acknowledgments), conducted a second inherently equivocal selection of "correct" entry-pairs in the binary relationship  $R: A \Rightarrow B$  which guides the interpretation process of the OAMTRX instance shown in Table 8-4. This second experiment provided a  $CVPAI2(R: A \Rightarrow B) = 0.6731$  and an  $OA(OAMTRX) = 97.28\% \pm 0\%$ , which are both superior to (better than) the pair of summary statistics  $CVPAI2(R: A \Rightarrow B) = 0.6689$  and  $OA(OAMTRX) = 96.88\% \pm 0\%$  provided by the independent human expert in the test case A. This binary relationship looks sparser and therefore less intuitive to understand than that shown as dark entry-pair cells in Table 8-4. Hence, it is not shown in this paper, although it is made available via anonymous ftp (SIAM-WELD-NLCD FTP 2016). When these summary statistics replace variables in Equation (2-2), we obtain



$OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{“Ultimate” GroundTruth 2006, NLCD 16 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$

$[\max\{0\%, OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\% - 22\%\}, \min\{100\%, 178\% - OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\%\}] =$

$[\max\{0\%, 97.28\% \pm 0\% - 22\%\}, \min\{100\%, 178\% - 97.28\% \pm 0\%\}] =$

$[75.28\%, 80.72\%]$ , with  $CVPAl2(R: A = \text{SIAM dictionary of 19 color names} \Rightarrow B = \text{NLCD legend of 16 LC class names}) = 0.6731$ , hence the semantic information gap from sub-symbolic sensory data to symbolic NLCD classes left to be filled by further stages in the EO-IUS pipeline  $= 1 - CVPAl2 = 0.3196$ , refer to the Part 1, Chapter 7.5. (2-4)

**Test case C.** The wall-to-wall overlap between the test SIAM-WELD 2006 map at fine (96) color discretization (refer to Table 7-1 in the Part 1) and the reference NLCD 2006 map generated another OAMTRX instance featuring a test set  $A = 96$  color names as rows and a reference set  $B = 16$  NLCD classes as columns, available via anonymous ftp (SIAM-WELD-NLCD FTP 2016). Again, the hybrid inference procedure described in the Part 1, Chapter 7.4 was employed by the independent human expert to select “correct” entry-pairs in the binary relationship  $R: A \Rightarrow B$  that guides the interpretation process of this OAMTRX instance, whose estimated  $TQ^2I$  summary statistics became  $CVPAl2(R: A \Rightarrow B) = 0.5809$  and  $OA(OAMTRX) = 95.41\% \pm 0\%$ . When these summary statistics replace variables in Equation (2-2), we obtained

$OA(\text{Test SIAM 2006, 96 spectral macro-categories} \Rightarrow \text{“Ultimate” GroundTruth 2006, NLCD 16 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$

$[\max\{0\%, OA(\text{Test SIAM 2006, 96 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\% - 22\%\}, \min\{100\%, 178\% - OA(\text{Test SIAM 2006, 96 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, NLCD 16 classes}) \pm 0\%\}] =$

$[\max\{0\%, 95.41\% \pm 0\% - 22\%\}, \min\{100\%, 178\% - 95.41\% \pm 0\%\}] =$

$[73.41\%, 82.59\%]$ , with  $CVPAl2(R: A = \text{SIAM dictionary of 96 color names} \Rightarrow B = \text{NLCD legend of 16 LC class names}) = 0.5809$ , hence the semantic information gap from sub-symbolic sensory data to symbolic NLCD classes left to be filled by further stages in the EO-IUS pipeline  $= 1 - CVPAl2 = 0.4191$ , refer to the Part 1, Chapter 7.5. (2-5)

**Test case D.** When the NLCD classification taxonomy becomes less discriminative (coarser) because reassembled from 16 LC class names to 9 and 4 LC class names respectively, it is a fact that the following inequality holds.

$OA(\text{Reference NLCD 2006, NLCD 16 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, NLCD 16 classes}) = 78\% \leq$

$OA(\text{Reference NLCD 2006, “augmented” LCCS-DP 9 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, “augmented” LCCS-DP 9 classes}) = 84\% \leq$

$OA(\text{Reference NLCD 2006, 2-level LCCS-DP 4 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) = XX\%,$  (2-6)

where the 2-level LCCS-DP 4 classes are (see Fig. 8-1):

- A1 = Primarily Vegetated Terrestrial Areas = Cultivated Areas (A11) or (Semi) Natural Vegetation (A12).
- A2 = Primarily Vegetated Aquatic or Regularly Flooded Areas = Cultivated Aquatic Areas (A23) or (Semi) Natural Aquatic Vegetation (A24).
- B3 = Primarily Non-vegetated Terrestrial Areas = Artificial Surfaces (B35) or Bare Areas (B36).
- B4 = Primarily Non-vegetated Aquatic or Regularly Flooded Areas = Artificial (B47) or Natural Waterbodies, Snow and Ice (B48).

In agreement with Equation (2-1), Equation (2-2) and Equation (2-6) we could write

$OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) \in [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] =$



$$[\max\{0\%, \text{OA}(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, 2-level LCCS-DP 4 classes}) \pm 0\% - (100\% - \text{XX}\%), \min\{100\%, 100\% + \text{XX}\% - \text{OA}(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{Reference NLCD 2006, 2-level LCCS-DP 4 classes}) \pm 0\%\}], \text{ with } \text{XX}\% = \text{OA}(\text{Reference NLCD 2006, 2-level LCCS-DP 4 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) \geq 84\%. \quad (2-7)$$

As shown in the test case A, according to the NLCD definitions (refer to Table 8-1), LC classes "Developed, Open Space" (DOS), "Developed, Low Intensity" (DLI), "Developed, Medium Intensity" (DMI) and "Developed, High intensity" (DHI) are a spatial mixture of vegetated surfaces, impervious surfaces and bare soil, in agreement with the popular vegetation-impervious surface-soil model for urban ecosystem analysis (Ridd 1995). It means that a logical OR combination of the NLCD classes DOS or DLI or DMI or DHI mainly matches with the 2-level LCCS-DP classes B3 or A1. In general, it is not possible to backtrack (reconstruct) each of the 2-level LCCS-DP 4 classes starting from the NLCD 16 class taxonomy without ambiguity. To transform a reference NLCD map with an NLCD 16-class legend into a reference NLCD map with a 2-level LCCS-DP 4-class legend we adopted the approximated binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , with set  $A = \text{NLCD 16-class legend}$  and set  $B = \text{2-level LCCS-DP 4-class legend}$ , reported in Table 8-5 and summarized below.

- A1 = Primarily Vegetated Terrestrial Areas = Cultivated Areas or (Semi) Natural Vegetation  $\approx$  NLCD 16-classes DF (41) or EF (42) or MF (43) or SS (52) or GH (71) or PH (81) or CC (82). Actually, this NLCD class OR-combination is an expected mixture of LCCS-DP classes A1 and B3 as first- and second-best match respectively.
- A2 = Primarily Vegetated Aquatic or Regularly Flooded Areas = Cultivated Aquatic Areas or (Semi) Natural Aquatic Vegetation  $\approx$  NLCD 16-classes WV (90) or EHW (95).
- B3 = Primarily Non-vegetated Terrestrial Areas = Artificial Surfaces or Bare Areas  $\approx$  NLCD 16-classes DOS (21) or DLI (22) or DMI (23) or DHI (24) or BL (31). Actually, this NLCD class OR-combination is an expected mixture of LCCS-DP classes B3 and A1 as first- and second-best match respectively.
- B4 = Primarily Non-vegetated Aquatic or Regularly Flooded Areas = Artificial or Natural Waterbodies, Snow and Ice  $\approx$  NLCD 16-classes OW (11) or PIS (12).

Fixed this approximated binary relationship  $R: A \Rightarrow B \subseteq A \times B$  with set  $A = \text{NLCD 16-class legend}$  and set  $B = \text{2-level LCCS-DP 4-class legend}$ , Table 8-5 shows the following.

- A binary relationship  $R: C \Rightarrow B \subseteq C \times B$  with set  $C = \text{SIAM 19-class legend}$  as rows and set  $B = \text{2-level LCCS-DP 4-class legend}$  as columns, where “correct” entry-pairs are shown as grey cells in the 2-fold Cartesian product  $C \times B$ . This binary relationship was (subjectively) selected by the present authors according to the categorical variable-pair relationship identification strategy proposed in the Part 1, Chapter 7.5.
- An OAMTRX = FrequencyCount( $C \times B$ ) generated by the wall-to-wall overlap between the test SIAM map with legend C and the reference NLCD map with legend B.

Table 8-5 provided a  $\text{CVPAI2}(R: C \Rightarrow B) = 0.7486$  and an  $\text{OA}(\text{OAMTRX} = \text{FrequencyCount}(C \times B)) = 93.09\% \pm 0\%$ . When these summary statistics replace variables in Equation (2-7), we obtained

$$\begin{aligned} & \text{OA}(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) \in \\ & [\max\{0\%, \text{Lower Bound}\}, \min\{100\%, \text{Upper Bound}\}] = \\ & [\max\{0\%, 93.09\% \pm 0\% - (100\% - \text{XX}\%), \min\{100\%, 100\% + \text{XX}\% - 93.09\% \pm 0\%\}] = \\ & [\text{XX}\% - 6.91\%, \text{XX}\% + 6.91\%], \text{ where } \text{XX}\% = \text{OA}(\text{Reference NLCD 2006, 2-level LCCS-DP 4 classes} \Rightarrow \text{“Ultimate”} \\ & \text{GroundTruth 2006, 2-level LCCS-DP 4 classes}) \geq 84\%, \text{ with } \text{CVPAI2}(R: C = \text{SIAM dictionary of 19 spectral macro-} \\ & \text{categories} \Rightarrow B = \text{2-level LCCS-DP 4 classes}) = 0.7486, \text{ hence the semantic information gap from sensory data to the} \\ & \text{2-level LCCS-DP 4-class legend left to be filled by further stages in the EO-IUS pipeline} = 1 - \text{CVPAI2} = 0.2514, \\ & \text{refer to the Part 1, Chapter 7.5.} \end{aligned} \quad (2-8)$$

#### 8.4.4 Probabilities of the SIAM-WELD Test Labels Conditioned by the NLCD Reference Labels and Vice Versa

Table 8-4 shows the OAMTRX instance generated from the wall-to-wall overlap of the SIAM-WELD 2006 test map featuring 19 spectral macro-categories, see Table 8-2 and Fig. 8-6, with the NLCD 2006 reference map featuring 16 LC classes, see Table 8-1 and Fig. 8-3. The division of each probability cell of Table 8-4 by its column-sum generates the conditional probability  $p(\text{SIAM-WELD}_t | \text{NLCD}_r)$  of the SIAM-WELD 2006 test spectral category  $t$ , with  $t = 1, \dots, TC = 19$ , given the NLCD 2006 reference class  $r$ , with  $r = 1, \dots, RC = 16$ , refer to Fig. 8-8 and Table 8-6, where Table 8-6 is a summarized text version of Fig. 8-8. To prove their plausibility, conditional probabilities  $p(\text{SIAM-WELD}_t | \text{NLCD}_r)$ ,  $t = 1, \dots, TC$ ,  $r = 1, \dots, RC$ , should agree with theoretical expectations stemming from human experience. For instance, it was



expected that the NLCD 2006 reference classes "*Deciduous Forest*" (DF), "*Evergreen Forest*" (EF) and "*Mixed Forest*" (MF), refer to Table 8-1, overlap with vegetated spectral categories in the test SIAM-WELD 2006 map, while the NLCD reference class "*Developed, High Intensity*" (DHI, see Table 8-1) was expected to be mostly matched by bare soil macro-categories in the test SIAM-WELD 2006 map. Overall, these prior knowledge-based expectations about specific class-conditional probabilities appear satisfied by both Fig. 8-8 and Table 8-6.

In the RS common practice, once a generic user has generated at no cost in manpower and computer power, i.e., in near real-time and without user-machine interaction, a SIAM color map from an unknown EO image, what this user wishes to do is to infer from the EO image a set of LC classes (say, "*Forest*"), conditioned by the detected SIAM spectral categories (say, MS green-as-"*Vegetation*"). To accomplish this spectral category-conditional inference, class-conditional probabilities  $p(\text{SIAM-WELD}_t | \text{NLCD}_r)$ ,  $t = 1, \dots, TC$ ,  $r = 1, \dots, RC$ , shown in Table 8-6, are not useful. Rather, this generic user can find helpful to know the conditional probabilities of an NLCD 2006 reference class  $r$ , with  $r = 1, \dots, RC = 16$ , given the SIAM-WELD 2006 spectral category  $t$ , with  $t = 1, \dots, TC = 19$ . These are the class-conditional probabilities  $p(\text{NLCD}_r | \text{SIAM-WELD}_t)$ ,  $t = 1, \dots, TC$ ,  $r = 1, \dots, RC$ , generated by dividing each probability cell of Table 8-4 by its row-sum. They are shown in Fig. 8-9 and summarized in text form in Table 8-7. Very intuitive to understand, Table 8-7 clearly highlights the two main semantic inconsistencies found between the reference NLCD 2006 and test SIAM-WELD 2006 maps already reported in Chapter 8.4.3. First, the SIAM vegetation spectral macro-category wV\_HC ("*Weak evidence vegetation with high canopy cover*", refer to Table 8-2) is best-matched by the reference NLCD class "*Open Water*" (OW, refer to Table 8-2). Since this semantic mismatch occurs almost exclusively in the CONUS areas recognized by the independent human expert as riparian zones typically depicted as mixed pixels at 30 m resolution, then the 30 m resolution SIAM labeling can be considered reasonable, if we consider that the crisp SIAM implementation is not expected to accomplish pixel unmixing. Second, the NLCD reference class "*Shrub/Scrub*" (SS, refer to Table 8-2) appears to be the best match for several of the SIAM bare soil spectral macro-categories. Fig. 8-7 shows examples of geographic locations where this semantic mismatch occurs. In these locations, 30 m resolution pixels are typically affected by mixed spectral contributions that the crisp SIAM implementation is not expected to unmix.

#### 8.4.5 Stratification by Ecoregions

Due to the central limit theorem, the arithmetic mean of a large number of independent random variables tends to be a Gaussian distribution, where independent "local" data distributions (like basis functions) become indistinguishable from the whole. For example, in human vision, the neural computations are inherently spatially local in the (2D) image-domain; next, a global spatial average is superimposed on the local computational processes. In general, non-stationary local features do not survive the averaging process, i.e., the precise position of each local contribution is no longer perceived after the averaging process (Victor 1994). Since the WELD composite of the CONUS is about ten billion pixels in size, summary statistics of the SIAM mapping quality at the CONUS spatial extent are inadequate to demonstrate the local-scale capability of the SIAM expert system to correctly map EO images, characterized by non-stationary local statistics. To investigate the SIAM mapping capability at local spatial extent, the SIAM-WELD 2006 and NLCD 2006 maps were stratified using the 86 EPA Level III ecoregions of the CONUS (see Fig. 8-3) and an individual OAMTRX was generated per ecoregion. All 86 ecoregion-specific OAMTRX instances are available as supplemental online material (SIAM-WELD-NLCD FTP 2016). As one example of an inter-map comparison at the ecoregion spatial scale of analysis, let us consider the SIAM-WELD 2006 and NLCD 2006 maps of the Wyoming Basin ecoregion, which is predominantly desert, see Fig. 8-10 where the ecoregion boundary is highlighted in red. Table 8-8 reports the corresponding OAMTRX instance. Table 8-8 shows that the predominantly desert Wyoming Basin ecoregion is predominantly classified as the LC classes "*Scrub/Shrub*" (SS) and "*Grassland/Herbaceous*" (GH) in the NLCD 2006 reference map (refer to Table 8-1) and as bare soil spectral categories (sbS\_1, SmS\_1, aS) in the SIAM-WELD 2006 test map (refer to Table 8-2). This semantic disagreement was already observed in Chapter 8.4.3, also refer to Fig. 8-7.

Fig. 8-11 provides a synthetic representation of the full dataset of 86 ecoregion-specific OAMTRX instances (SIAM-WELD-NLCD FTP 2016). It shows for each of the 16 reference NLCD classes, with index  $r = 1, \dots, RC = 16$ , the box-and-whisker diagram of the NLCD-class-conditional probabilities  $p(\text{SIAM-WELD}_{er,t} | \text{NLCD}_{er,r})$ , with  $t = 1, \dots, TC = 19$ , collected across the 86 ecoregions, each ecoregion identified with an index  $er = 1, \dots, ER = 86$ . In each of the  $TC = 19$  boxes of an NLCD class-specific boxplot, the median (shown as a horizontal line within the box) represents the general trend of the distribution and the dispersion around it describes the distribution variability across ecosystems. A small dispersion around the median value indicates a reference-to-test class mapping whose occurrence is nearly constant across





ecosystems, while a large dispersion around the median indicates that occurrences of this inter-map relationship change significantly across ecosystems.

#### 8.4.6 OP-Q<sup>2</sup>I values of the SIAM application and product

OP-Q<sup>2</sup>Is of the SIAM application (refer to the introduction to Chapter 8.4) input with the 30 m resolution annual WELD 2006 to 2009 composites of the CONUS were collected by the independent human expert (refer to Acknowledgments). They are summarized below (Duke 2016).

(i) Process degree of automation. In line with theoretical expectations about expert systems (refer to Chapter 8.1), the SIAM application required neither user-defined parameters nor reference samples to run. Hence, its ease of use cannot be surpassed by any alternative inference approach.

(ii) Outcome effectiveness. An mDMI set of O-Q<sup>2</sup>Is (Si Liu et al. 2011; Peng et al. 2005), comprising a CVPAl2(R:  $A \Rightarrow B$ ), an OA(OAMTRX = FrequencyCount( $A \times B$ )), the class-conditional probabilities  $p(r | t)$  of reference class  $r = I, \dots, RC = |B|$ , given test class  $t = 1, \dots, TC = |A|$ , and class-conditional probabilities  $p(t | r)$ , with  $r = I, \dots, RC = |B|, t = 1, \dots, TC = |A|$ , was estimated in the four test cases described in Chapter 8.4.3.

(iii) Process efficiency: memory occupation and computation time. About memory occupation, the SIAM computer program adopts a tile streaming implementation, where the dynamic memory (random access memory, RAM) maximum occupation is a known function of the tile size to be fixed in advance, irrespective of the image size. In these experiments the RAM maximum occupation was set equal to 800 MB, which can be considered a “small” RAM value. About computation time: Run on a Dell Power Edge 710 server with dual Intel Xeon @ 2.70 GHz processor with 64 GB of RAM and a 64-bit Linux operating system, the SIAM application required less than 45 seconds to generate its complete set of per-image output products from a 7-band Landsat-7 ETM+ WELD tile of  $5000 \times 5000$  pixels, which means about 8 hours to map an annual WELD composite of the CONUS. In our data mapping workflow, such an output rate was not inferior to the input rate of an annual WELD composite being implemented or delivered to end-users. Hence, the SIAM computation time was considered equivalent to near real-time, where the SIAM computational complexity increases linearly with image size.

(iv) Process robustness to changes in the input dataset. The SIAM mapping consistency of the annual WELD composites from year 2006 to 2009 was estimated to be “high” at the CONUS spatial extent, refer to Chapter 8.4.2 to Chapter 8.4.5.

(v) Process robustness to changes in input parameters, if any. Since SIAM requires no user-defined parameter to run, its robustness to changes in input parameters cannot be surpassed by alternative approaches.

(vi) Process maintainability/ scalability/ re-usability, to keep up with changes in users’ needs and sensor properties. The multi-source SIAM physical model can be applied to any existing or future planned spaceborne/airborne MS imaging sensor provided with a radiometric calibration metadata file, refer to the existing literature (Baraldi and Humber 2015; Baraldi et al. 2010c; Baraldi et al. 2010a, 2010b) and to the Part 1, Chapter 7.2.

(vii) Outcome timeliness, defined as the time span between data acquisition and product generation. Since it is prior knowledge-based and near real-time, the SIAM application reduces timeliness from image acquisition to color map generation to almost zero.

(viii) Outcome costs, monotonically increasing with manpower and computer power. Since it is prior knowledge-based and near real-time in a standard laptop computer, the SIAM costs are almost negligible.

### 8.5 Discussion

Table 8-3 shows that the 30 m resolution annual SIAM-WELD map time-series at the CONUS spatial extent with an intermediate color discretization legend of 48 color names reassembled into 19 spectral macro-categories features a standard deviation of the annual frequency counts collected for each spectral macro-category lower than 1%, with the exception of two vegetated spectral macro-categories, specifically, aV\_HC and aV\_MC (see Table 8-2). These two larger spectral category-specific variations in annual frequency counts at the CONUS spatial extent can be attributed mostly to vegetation phenology. This was proved in Chapter 8.4.2: changes in phenology affect the monthly WELD and annual WELD composites and, as a consequence, the data-derived SIAM-WELD maps. These numerical results agree with the *a priori* knowledge of RS experts about the CONUS surface dynamics, whose inter-annual LCC summary statistics are expected to score low. The conclusion is that observations stemming from the annual SIAM-WELD map time-series with



a legend of 19 spectral macro-categories comply with the domain knowledge of RS experts about the LC and LCC dynamics in the physical CONUS.

The interpretation process of the OAMTRX = FrequencyCount( $A \times B$ ) shown in Table 8-4, generated from the wall-to-wall overlap between the test SIAM-WELD 2006 map featuring a set DictionaryOfColorNames =  $A = 19$  spectral macro-categories and the reference NLCD 2006 map with a set LegendOfObjectClassNames =  $B = 16$  LC classes, is guided by the inter-dictionary binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , whose entry-pair cells, shown in gray, were selected by the independent human expert (refer to Acknowledgments). Table 8-4 reveals one single systematic case of “conceptual mismatch” between the NLCD 2006 reference vegetation classes “*Scrub/Shrub*” (SS) or “*Grassland/Herbaceous*” (GH, refer to Table 8-1) and the SIAM-WELD 2006 bare soil spectral categories sbS\_1, SmS\_1 and aS (refer to Table 8-2). These inter-map semantic mismatches occur in geographical locations where the CONUS landscapes look like those shown in Fig. 8-7. When these land surface types are observed from space with a spatial resolution of 30 m, a one-pixel surface area of 900 m<sup>2</sup> becomes a spectral mixture of sparse vegetation, rangeland, cheatgrass, dry long grass and/or short grass as foreground, with a background of sand, clay and/or rocks, especially if the percentage of vegetation cover can be slightly above the 15% of total cover required by the NLCD definitions of classes SS and GH (refer to Table 8-2). In these mixed pixels at 30 m resolution, the spectral detection of the vegetated component is impossible for a hard (crisp) classifier, while it would be more manageable by a fuzzy classifier (Baraldi 2011). In these experiments, since the SIAM expert system is run in crisp mode (refer to Chapter 8.3), then no pixel unmixing strategy can be applied to diminish or avoid the observed case of “semantic mismatch”. The conclusion is that the “conceptual mismatch” between the NLCD 2006 reference vegetation classes SS and GH and the SIAM-WELD 2006 bare soil spectral categories is a possible example of systematic disagreement between the test and reference thematic maps featuring the same spatial resolution whose occurrence should be carefully scrutinized by RS experts in comparison with an “ultimate” ground truth, see Fig. 8-7.

A different strategy to aesthetically (rather than formally) remove the aforementioned inter-dictionary “conceptual mismatch” would be to change color names in the SIAM color map legend, without changing the SIAM decision tree for color space hyperpolyhedralization. In other words, based on thematic evidence collected on an *a posteriori* basis from the NLCD reference map, it would be possible to change color names attached to the SIAM-WELD 2006 map legend and consider that, at the Landsat spectral and spatial resolution of an annual WELD composite of the CONUS, the SIAM spectral categories sbS\_1, SmS\_1 and aS are more likely to map the NLCD reference vegetation classes “*Shrub/Scrub*” (SS) or “*Grassland/herbaceous*” (GH) than bare soil surface types.

Starting from the same OAMTRX = FrequencyCount( $A \times B$ ) shown in Table 8-4, two independent selections by two different RS experts of a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  provided two alternative pairs of independent O-Q<sup>2</sup>I values to be jointly maximized, namely, a CVPAl2( $R: A \Rightarrow B$ ) = 0.6689 with OA(OAMTRX) = 96.88%  $\pm$  0% (test case A) and a CVPAl2( $R: A \Rightarrow B$ ) = 0.6731 with OA = 97.28%  $\pm$  0% with (test case B). These alternative O-Q<sup>2</sup>I pairs highlight the inherent ill-posedness of any inter-dictionary conceptual harmonization, although a specific protocol to reduce heuristic decisions by human experts in the identification of a binary relationship  $R: A \Rightarrow B \subseteq A \times B$  was proposed in the Part 1, Chapter 7.4. According to a Pareto multi-objective optimization principle, the latter O-Q<sup>2</sup>I value pair should be preferred to the former. This choice proves that the OA of the test SIAM-WELD 2006 map compared with the reference NLCD 2006 map scores “very high”, with a semantic information gap from sub-symbolic sensory data to symbolic NLCD classes left to be filled by further stages in the EO-IUS pipeline equal to  $(1 - \text{CVPAl2}) = 0.3196$ .

At the fine discretization level of the SIAM-WELD 2006 test map, featuring a legend  $A = 96$  color names, another inter-map wall-to-wall overlap with the NLCD 2006 reference map, whose legend  $B = 16$  LC classes, provided a pair of O-Q<sup>2</sup>I values equal to CVPAl2( $R: A \Rightarrow B$ ) = 0.5809 and OA(OAMTRX) = 95.41%  $\pm$  0% (test case C). When compared to the two pairs of O-Q<sup>2</sup>I values collected from the test case A and the test case B, this third TQ<sup>2</sup>I value pair proves that a finer reflectance space hyperpolyhedralization for color naming is not necessarily more convenient to cope with by human experts in the stratification of an LC classification problem according to a spectral and spatial convergence-of-evidence approach, refer to Equation (1-3) in the Part 1 (Hunt and Tyrrell 2012).

When an approximated binary relationship  $R: A \Rightarrow B \subseteq A \times B$  was identified from set  $A = \text{NLCD 16-class legend}$  to set  $B = \text{2-level LCCS-DP 4-class legend}$ , a binary relationship  $R: C \Rightarrow B \subseteq C \times B$  was defined from set  $C = \text{SIAM 19-class legend}$  as rows to set  $B = \text{2-level LCCS-DP 4-class legend}$  as columns and an OAMTRX = FrequencyCount( $C \times B$ ) was generated by the wall-to-wall overlap between the test SIAM map with legend  $C$  and the reference NLCD map with legend  $B$  as reported in Table 8-5 (test case D), then O-Q<sup>2</sup>I values were CVPAl2( $R: C \Rightarrow B$ ) = 0.7486 and OA(OAMTRX) = 93.09%  $\pm$  0%. From these results we can infer that



OA(Test SIAM 2006, 19 spectral macro-categories  $\Rightarrow$  “Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes)  $\in$  [XX% - 6.91%, XX% + 6.91%], where XX% = OA(Reference NLCD 2006, 2-level LCCS-DP 4 classes  $\Rightarrow$  “Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes)  $\geq$  84%, with a semantic information gap from sensory data to the 2-level LCCS-DP 4-class legend left to be filled by further stages in the EO-IUS pipeline equal to  $(1 - \text{CVP}AI2) = 0.2514$ .

This inference supports the thesis investigated by the present experimental work, where the off-the-shelf SIAM lightweight computer program for prior knowledge-based MS reflectance space hyperpolyhedralization was considered eligible for systematic Level 2 SCM product generation with an SCM legend consistent with the “augmented” 9-class LCCS-DP taxonomy.

To complete the interpretation of the OAMTRX shown in Table 8-4 two histograms of class-conditional probabilities, shown in Fig. 8-8 and Fig 2-9 respectively, together with their summarized text versions, shown as Table 8-6 and Table 8-7 respectively, were generated from the OAMTRX of interest. Fig. 8-8 and Table 8-6 reveal that any test SIAM-WELD 2006 spectral category conditioned by one NLCD 2006 reference class appears consistent with the NLCD class definition (refer to Table 8-1) and with *a priori* domain knowledge of RS experts about the real-world CONUS spatially sampled at 30 m resolution. Analogously, Fig. 8-9 and Table 8-7 show that any NLCD 2006 reference class conditioned by one SIAM-WELD 2006 spectral category appears consistent with the spectral properties of the SIAM color type and with *a priori* domain knowledge of RS experts about the physical CONUS depicted at 30 m resolution. To conclude, class-conditional probabilities generated from Table 8-4 appear reasonable and confirm the statistical plausibility of the OAMTRX instance shown in Table 8-4 as a whole.

Fig. 8-11 shows that if, for example, the boxplot of the NLCD reference class “*Developed, Open Space*” (DOS) is compared to that of reference class “*Developed, Low Intensity*” (DLI), “*Developed, Medium Intensity*” (DMI) and “*Developed, High Intensity*” (DHI, refer to Table 8-1), then a monotonic decrease of the class-conditional probability of the SIAM-WELD vegetated spectral categories conditioned by the NLCD reference class and collected at local spatial extents for a population of 86 ecoregions is observed in parallel with a monotonic increase of the class-conditional probability of the SIAM-WELD bare soil spectral categories. This is perfectly consistent with the *a priori* domain knowledge of RS experts about the spatial and spectral properties of urban and industrial area in the CONUS, in agreement with the popular vegetation-impervious surface-soil model for urban ecosystem analysis (Ridd 1995). In addition, these boxplots confirm that, at the local spatial extent of individual ecoregions, the NLCD reference classes “*Deciduous Forest*” (DF), “*Evergreen Forest*” (EF) and “*Mixed Forest*” (MF, refer to Table 8-1) are almost entirely (> 90%) covered by the SIAM-WELD vegetation spectral categories, in agreement with theoretical expectations about the SIAM-WELD test map. In line with preliminary outcomes discussed in Chapter 8.4.3 and in Fig. 8-10, boxplots shown in Fig. 8-11 confirm that the NLCD reference classes “*Scrub/Shrub*” (SS) and “*Grassland/Herbaceous*” (GH) have a strong heterogeneity of matches with the SIAM-WELD 2006 spectral categories collected at the ecoregion spatial extent. This is tantamount to saying that spectral signatures of these NLCD classes feature a strong variability when collected at “local” scale, also refer to Fig. 8-7. More properties of the NLCD class-specific box diagrams collected at the local spatial extent of ecoregions appear reasonable, based on *a priori* human knowledge of the physical CONUS at the ecoregion spatial extent. For example, first, the NLCD reference classes “*Pasture/Hay*” (PH) and “*Cultivated Crops*” (CC, refer to Table 8-1) are largely matched across ecoregions by the SIAM-WELD vegetated spectral categories. Second, the NLCD reference class “*Perennial Ice/Snow*” (PIS, refer to Table 8-1) is best-matched across ecoregions by the SIAM-WELD spectral category MS white-as-“*Snow*” (SN, refer to Table 8-2). Third, across ecoregions, the NLCD reference class “*Open water*” (OW, refer to Table 8-1) is best-matched by the SIAM-WELD spectral category MS blue-as-“*Water or Shadow*” (WA, refer to Table 8-2). To summarize, collected at the local extent of ecoregions to account for non-stationary spatial properties, boxplots shown in Fig. 8-11 are considered statistically and semantically consistent with the definitions of the two legends adopted by the test and reference maps, they agree with *a priori* domain knowledge of RS experts about the LC and LCC dynamics in the physical CONUS and appear consistent with global (non-stratified by ecoregions) statistics collected at the CONUS spatial extent reported in Chapter 8.4.2 to Chapter 8.4.4.

## 8.6 Conclusions

To pursue the GEO’s visionary objective of a GEOSS not-yet accomplished by the RS community we advanced the following thesis: a necessary not sufficient pre-condition for a GEOSS development is the systematic generation at the



ground segment of an ESA EO Level 2 product, whose general-purpose SCM legend agrees with the 3-level 9-class “augmented” LCCS-DP taxonomy (see Fig. 8-2). To comply with the GEO QA4EO *Call/Val* requirements an ESA EO Level 2 product must be submitted to a GEO Stage 4 *Val* process by independent means. No ESA EO Level 2 product has ever been systematically generated at the ground segment. This interdisciplinary work aimed at filling an analytic and pragmatic information gap from multi-source EO big sensory data to ESA EO Level 2 product. To fill this gap we focused our attention on a long history of prior knowledge-based MS reflectance space partitioners for static color naming, developed but never validated by space agencies, public organizations and private companies to be plugged into hybrid EO-IUSs for EO image enhancement and classification tasks in operating mode. As a potential candidate for systematic ESA EO Level 2 SCM product generation at the ground segment to be submitted to a GEO Stage 4 *Val* we selected an off-the-shelf SIAM lightweight computer program for prior knowledge-based MS color naming, presented in the RS literature in recent years where enough information was provided for the implementation to be reproduced.

For the sake of readability this paper is split into two, the preliminary Part 1 – Theory and the present Part 2 – Validation. Original contributions of the Part 1 include, first, an eight-step protocol to identify a categorical variable-pair relationship  $R: A \Rightarrow B$  from categorical variable A to categorical variable B as a hybrid combination of deductive prior beliefs with inductive evidence from data. Second, an original CVPAI2 formulation was proposed as a categorical variable-pair degree of association in a binary relationship  $R: A \Rightarrow B$ . The original contribution of the present Part 2 is a novel protocol for wall-to-wall inter-map comparison without sampling, where the test and reference maps feature the same spatial resolution and spatial extent, but whose legends are not the same and must be harmonized.

Conclusions of the present Part 2 are twofold. The off-the-shelf SIAM lightweight computer program can be considered suitable for systematic generation of a Level 2 SCM product in operating mode, where the SCM legend agrees with the “augmented” 9-class LCCS-DP taxonomy. It was inferred that  $OA(\text{Test SIAM 2006, 19 spectral macro-categories} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) \in [XX\% - 6.91\%, XX\% + 6.91\%,]$ , where  $XX\% = OA(\text{Reference NLCD 2006, 2-level LCCS-DP 4 classes} \Rightarrow \text{“Ultimate” GroundTruth 2006, 2-level LCCS-DP 4 classes}) \geq 84\%$ , with a semantic information gap from sensory data to the 2-level LCCS-DP 4-class legend left to be filled by further stages in the EO-IUS pipeline equal to  $(1 - CVPAI2) = 0.2514 \in [0, 1]$ . In agreement with the definition of an information processing system in operating mode proposed in Chapter 8.4, the off-the-shelf SIAM application software submitted to a Stage 4 *Val* can be considered in operating mode because its whole set of OP-Q<sup>2</sup>I values scored “high”.

Future developments will regard the systematic generation of a Level 2 SCM product whose map legend fully addresses the “augmented” 3-level 9-class LCCS-DP taxonomy, including quality layers such cloud and cloud-shadow. Our aim is to develop a hybrid (combined deductive and inductive) EO-IUS based on a convergence-of-evidence approach, where dominant spatial information and secondary color information are combined in line with Equation (1-3) in the Part 1. A prototypical implementation of this hybrid EO-IUS (Baraldi et al. 2016; Tiede et al. 2016) incorporates the SIAM application for color naming and exploits planar shape indexes, such as straightness-of-boundaries (Nagao and Matsuyama 1980; Soares et al. 2014), to discriminate managed (man-made) LC classes from natural surface types, such as the LCCS-DP level 3 class A11 (Cultivated and Managed Terrestrial Vegetated Areas) from class A12 (Natural and Semi-Natural Terrestrial Vegetation) and class B35 (Artificial Surfaces and Associated Areas) from class B36 (Bare Areas), see Fig. 8-1.

### Acknowledgments

To accomplish this work Andrea Baraldi was supported in part by the National Aeronautics and Space Administration (NASA) under Grant No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. Dirk Tiede was supported in part by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel2/3, ID 848009. Prof. Ralph Maughan, Idaho State University, is kindly acknowledged for his contribution as active conservationist and for his willingness to share his invaluable photo archive with the scientific community as well as the general public. Andrea Baraldi thanks Prof. Raphael Capurro, Hochschule der Medien, Germany, and Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, for their support. Above all, the authors acknowledge the fundamental contribution of Prof. Luigi Boschetti, currently at the Department of Forest, Rangeland and Fire Sciences, University of Idaho, Moscow, Idaho, who conducted by independent means all experiments whose results are proposed in this validation paper. The authors also wish to thank the Editor-in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.

### Disclosure statement

In accordance with XXX policy and his ethical obligation as a researcher, Andrea Baraldi reports he is the sole developer





and IPR owner of the Satellite Image Automatic Mapper™ (non-registered trademark) computer program licensed to academia, public institutions and private companies, eventually free-of-charge, by the one-man-company Baraldi Consultancy in Remote Sensing that may be affected by the research reported in the enclosed paper. Andrea Baraldi has disclosed those interests fully to XXX, and he has in place an approved plan for managing any potential conflicts arising from that involvement.

## References in Chapter 8

- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley. 1998. "Discriminating clear sky from clouds with MODIS", *J. Geophys. Res.* 103(32): 141-157.
- Ahlqvist, O. 2005. "Using uncertain conceptual spaces to translate between land cover categories." *Int. J. Geographic. Info. Science* 19: 831–857.
- Arvor, D., B. D. Madiela, and T. Corpetti. 2016. "Semantic pre-classification of vegetation gradient based on linearly unmixed Landsat time series." In *Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE International 4422-4425.
- Baraldi, A. 2011. "Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images." *IEEE Trans. Geosci. Remote Sens.* 49: 2113-2134.
- Baraldi, A., and L. Boschetti. 2012a. "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction," *Remote Sens.* 4: 2694-2735.
- Baraldi, A., and L. Boschetti. 2012b. "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation." *Remote Sens.* 4: 2768-2817.
- Baraldi, A., L. Boschetti, and M. Humber. 2014. "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images." *IEEE Trans. Geosci. Remote Sens.* 52(1): 701-760.
- Baraldi, A., L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda. 2010a. "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery- Part I: System design and implementation." *IEEE Trans. Geosci. Remote Sens.* 48: 1299-1325.
- Baraldi, A., L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda. 2010b. "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery- Part II: Classification accuracy assessment," *IEEE Trans. Geosci. Remote Sens.* 48: 1326-1354.
- Baraldi, A., M. Girona, and D. Simonetti. 2010c. "Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sensing* 48(1): 112-146.
- Baraldi, A., and M. Humber. 2015. "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 1 – Theory," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 8(3): 1307-1329.
- Baraldi, A., M. Humber, and L. Boschetti. 2013. "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 2 – Experimental results," *Remote Sens.* 5: 5209-5264.
- Baraldi, A., V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino. 2006. "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images." *IEEE Trans. Geosci. Remote Sens.* 44:2563-2586.
- Baraldi, A., D. Tiede and S. Lang. 2015. "Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+." Project proposal (1-year duration): ESA Invitation to tender ESA/AO/1-8373/15/I-NB – "Value Adding Element (VAE): Next Generation EO-based Information Services", University of Salzburg, Department of Geoinformatics - Z\_GIS.



- Baraldi, A., D. Tiede, M. Sudmanns, M. Belgiu, and S. Lang. 2016. "Automated near real-time Earth observation Level 2 product generation for semantic querying," GEOBIA 2016, 14-16 Sept., University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, The Netherlands.
- Baraldi, A., T. Wassenaar, and S. Kay. 2010d. "Operational performance of an automatic preliminary spectral rule-based decision-tree classifier of spaceborne very high resolution optical images." *IEEE Trans. Geosci. Remote Sens.* 48: 3482-3502.
- Beauchemin, M. and K. Thomson. 1997. "The evaluation of segmentation results and the overlapping area matrix." *Int. J. Remote Sens.* 18: 3895-3899.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon.
- Boschetti, L., S. P. Flasse, and P. A. Brivio. 2004. "Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto boundary," *Remote Sens. Environ.* 91: 280-292.
- Boschetti, L., D. P. Roy, C. O. Justice, and M. L. Humber. 2015. "MODIS-Landsat fusion for large area 30 m burned area mapping." *Remote Sens. Environ.* 161: 27-42.
- Capurro, R., and B. Hjørland. 2003. "The concept of information." *Annual Review of Information Science and Technology* 37: 343-411.
- CNES. 2015. Venus Satellite Sensor Level 2 Product. Accessed 5 January 2016. [https://venus.cnes.fr/en/VENUS/prod\\_l2.htm](https://venus.cnes.fr/en/VENUS/prod_l2.htm)
- Cherkassky, V., and F. Mulier. 1998. *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley.
- Dai, X., and S. Khorrarn. 1998. "The effects of image misregistration on the accuracy of remotely sensed change detection." *IEEE Trans. Geosci. Remote Sens.* 36: 1566-1577.
- Despini, F., S. Teggi and A. Baraldi. 2014. "Methods and metrics for the assessment of pan-sharpening algorithms", in *SPIE Proceedings*, Vol. 9244: Image and Signal Processing for Remote Sensing XX, L. Bruzzone, J. A. Benediktsson, and F. Bovolo, Eds., Amsterdam, Netherlands, Sept. 22.
- Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) and VEGA Technologies. 2011. "Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document." Document S2PAD-ATBD-0001, European Space Agency.
- Di Gregorio, A., and L. Jansen. 2000. Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO: Rome, Italy, FAO Corporate Document Repository. Accessed 10 February 2015. <http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm>
- Dillencourt, M. B., H. Samet, and M. Tamminen. 1992. "A general approach to connected component labeling for arbitrary image representations," *J. Association for Computing Machinery* 39: 253-280.
- Dorigo, W., R. Richter, F. Baret, R. Bamler, and W. Wagner. 2009. "Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach," *Remote Sens.* 1: 1139-1170.
- Duke Center for Instructional Technology, 2016. *Measurement: Process and Outcome Indicators*. Accessed 20 June 2016. [http://patientsafetyped.uhs.duke.edu/module\\_a/measurement/measurement.html](http://patientsafetyped.uhs.duke.edu/module_a/measurement/measurement.html)
- Environmental Protection Agency (EPA). 2007. "Definitions" in *Multi-Resolution Land Characteristics Consortium (MRLC)*. Accessed 13 November 2013. <http://www.epa.gov/mrlc/definitions.html#2001>
- Environmental Protection Agency (EPA). 2013. *Western Ecology Division*. Accessed 13 November 2013. <http://www.epa.gov/wed/pages/ecoregions.htm>
- European Space Agency (ESA). 2015. Sentinel-2 User Handbook, , Standard Document, Issue 1 Rev 2.
- GeoTerraImage. 2015. Provincial and national land cover 30m. Accessed 22 September 2015. <http://www.geoterraimage.com/productslandcover.php>
- Griffith, G. E. and J. M. Omernik. 2009. "Ecoregions of the United States-Level III (EPA)," in C.J. Cleveland (Ed.), *Encyclopedia of Earth*. Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment.
- Group on Earth Observation (GEO). 2005. "The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, adopted 16 February 2005." Accessed 10 January 2012. <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>
- Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS). 2010. "A Quality Assurance Framework for Earth Observation, version 4.0." Accessed 15 November 2012. [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- Group on Earth Observation / Committee on Earth Observation Satellites (GEO-CEOS) - Working Group on Calibration and Validation (WGCV). 2015. *Land Product Validation (LPV)*. Accessed March 20, 2015. <http://lpvs.gsfc.nasa.gov/>



- Homer, C., C. Q. Huang, L. M. Yang, B. Wylie, and M. Coan. 2004. "Development of a 2001 National Land-Cover Database for the United States." *Photo. Engin. Remote Sens.* 70:829-840.
- Hunt, N. and S. Tyrrell. 2012. *Stratified Sampling*. Coventry University. Accessed 7 February 2012. <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>
- Lee, D. S., J. C. Storey, M. J. Choate, and R. Hayes. 2004. "Four years of Landsat-7 on-orbit geometric calibration and performance." *IEEE Trans. Geosci. Remote Sens.* 42: 2786-2795.
- Lunetta, R., and D. Elvidge. 1999. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London, UK: Taylor & Francis.
- Kuzera, K. and R.G. Pontius. 2008. "Importance of matrix construction for multiple-resolution categorical map comparison." *GIScience and Remote Sens.* 45: 249–274.
- Liang, S. 2004. *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons.
- Lillesand, T., and R. Kiefer. 1979. *Remote Sensing and Image Interpretation*, New York: John Wiley and Sons.
- Lück W., and A. van Niekerk. 2016. "Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images," *Int. J. of Applied Earth Observation and Geoinformation* 47: 1–14.
- Lunetta, R. and D. Elvidge. 1999. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. London, UK: Taylor & Francis.
- Luo, Y., A. P. Trishchenko and K. V. Khlopenkov. 2008. "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sensing of Environment* 112: 4167–4185.
- Markham, B., and D. Helder. 2012. "Forty-year calibrated record of earth-reflected radiance from Landsat: A review." NASA Publications, Paper 70.
- Marr, D. 1982. *Vision*. New York, NY: Freeman and C.
- Nagao, M., and T. Matsuyama. 1980. *A Structural Analysis of Complex Aerial Photographs*. New York, NY, USA: Plenum, 1980.
- Open Geospatial Consortium (OGC) Inc. 2015. *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Accessed 8 March 2015. <http://www.opengeospatial.org/standards/is>
- Ortiz, A., and G. Oliver. 2006. "On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures." *Pattern Recognition Letters* 27: 1916-1926.
- Peng, H., F. Long, and C. Ding. 2005. "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy." *IEEE Trans. Pattern Anal. Machine Intell.* 27: 1226–1238.
- Pontius Jr., R.G., and J. Connors. 2006. "Expanding the conceptual, mathematical and practical methods for map comparison," in Caetano, M. and Painho, M. (Eds). *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon. Instituto Geográfico Português, 5 – 7 July 2006, 64-79.
- Richter, R., and D. Schläpfer. 2012a. "Atmospheric / Topographic Correction for Satellite Imagery – ATCOR-2/3 User Guide, Version 8.2 BETA." Accessed 12 April 2013. [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor3\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor3_manual_2012.pdf)
- Richter, R., and D. Schläpfer. 2012a. "Atmospheric / Topographic correction for airborne imagery – ATCOR-4 User Guide, Version 6.2 BETA, 2012." Accessed 12 April 2013. [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor4\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf)
- Ridd, M. K. 1995. "Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities." *Int. J. Remote Sens.* 16(12): 2165-2185.
- Roy, D., J. Ju, K. Kline, P. L. Scaramuzza, V. Kovalsky, M. Hansen, T. R. Loveland, E. Vermote, and C. S. Zhang. 2010. "Web-enabled Landsat Data (WELD): Landsat ETM plus composited mosaics of the conterminous United States." *Remote Sens. Environ.* 114:35-49.
- Si Liu, Hairong Liu, L. J. Latecki, Shuicheng Yan, Changsheng Xu, Hanqing Lu. 2011. "Size adaptive selection of most informative features." *Assoc. Advanc. Artificial Intel.*
- SIAM-WELD-NLCD FTP, 2016. Accessed 13 December 2016: <http://tinyurl.com/j4nzwzl>
- Simonetti, D., E. Simonetti, Z. Szantoi, A. Lupi, and H. D. Eva. 2015. "First results from the phenology based synthesis classifier using Landsat-8 imagery," *IEEE Geosci. Remote Sens. Lett.* 12(7): 1496-1500.
- Soares, J., A. Baraldi, and D. Jacobs. 2014. "Segment-based simple-connectivity measure design and implementation." Tech. Rep., University of Maryland, College Park, MD.



- Sonka, M., V. Hlavac, and R. Boyle. 1994. *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall.
- Stehman, S. V., and R L. Czaplewski. 1998. "Design and analysis for thematic map accuracy assessment: Fundamental principles." *Remote Sens. Environ.* 64: 331-344.
- Stehman, S. V., J. D. Wickham, T. G. Wade, and J. H. Smith. 2008. "Designing a multi-objective, multi-support accuracy assessment of the 2001 National Land Cover Data (NLCD 2001) of the conterminous United States." *Photo. Engin. Remote Sens.* 74: 1561-1571.
- Tiede, D., A. Baraldi, M. Sudmanns, M. Belgiu, and S. Lang. 2016. "ImageQuerying (IQ) – Earth Observation Image Content Extraction & Querying across Time and Space," submitted (Oral presentation and poster session), ESA 2016 Conf. on Big Data From Space, BIDS '16, Santa Cruz de Tenerife, Spain, 15-17 March.
- Vermote, E., and N. Saleous. 2007. "LEDAPS surface reflectance product description - Version 2.0." University of Maryland at College Park /Dept Geography and NASA/GSFC Code 614.5
- Victor, J. 1994. "Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: Comment." *J. Opt. Soc. Am. A* 11(5): 1680-1684.
- Vogelmann, J. E., T. L. Sohl, P. V. Campbell, and D. M. Shaw. 1998. "Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources." *Environ. Monitoring Assess.* 51: 415-428.
- Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. "Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources." *Photo. Eng. Remote Sens.* 67: 650-662.
- Web-Enabled Landsat Data (WELD) Tile FTP. Accessed 12 December 2016. <https://weld.cr.usgs.gov/>
- Wessels, K., F. van den Bergh, D. Roy, B. Salmon, K. Steenkamp, B. MacAlister, D. Swanepoel and D. Jewitt. 2016. "Rapid Land Cover Map Updates Using Change Detection and Robust Random Forest Classifiers." *Remote Sens.* 8(888): 1-24.
- Wickham, J. D., S. V. Stehman, J. A. Fry, J. H. Smith, and C. G. Homer. 2010. "Thematic accuracy of the NLCD 2001 land cover for the conterminous United States." *Remote Sens. Environ.* 114: 1286-1296.
- Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry, and T. G. Wade. 2013. "Accuracy assessment of NLCD 2006 land cover and impervious surface." *Remote Sens. Environ.* 130: 294-304.
- Xian, G., and C. Homer. 2010. "Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods." *Remote Sens. Environ.* 114: 1676-1686.



Figures and figure captions in Chapter 8

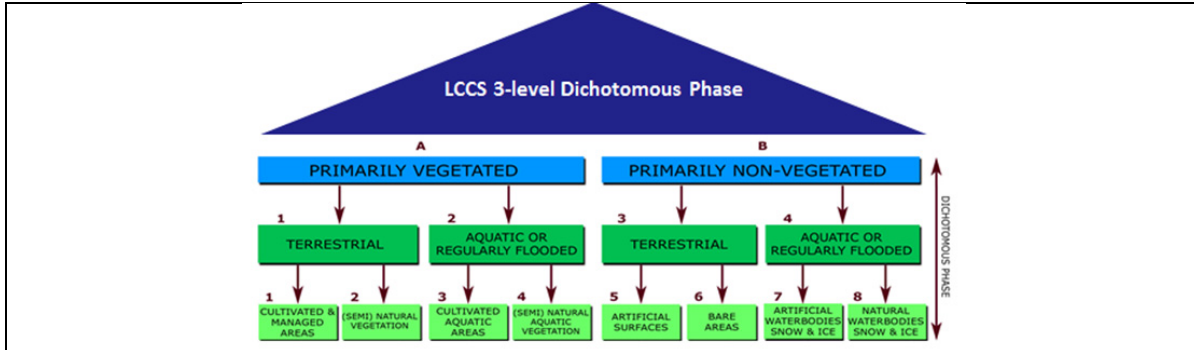


Fig. 8-1. The nested 3-level LCCS-DP layers are: (i) vegetation versus non-vegetation, (ii) terrestrial versus aquatic, and (iii) managed versus natural or semi-natural. They deliver as output the following 8-class LCCS-DP taxonomy. (A11) Cultivated and Managed Terrestrial (non-aquatic) Vegetated Areas. (A12) Natural and Semi-Natural Terrestrial Vegetation. (A23) Cultivated Aquatic or Regularly Flooded Vegetated Areas. (A24) Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation. (B35) Artificial Surfaces and Associated Areas. (B36) Bare Areas. (B47) Artificial Waterbodies, Snow and Ice. (B48) Natural Waterbodies, Snow and Ice. The general-purpose user- and application-independent 8-class LCCS-DP taxonomy is preliminary to a user- and application-specific LCCS Modular Hierarchical Phase (MHP) taxonomy (Di Gregorio and Jansen 2000).

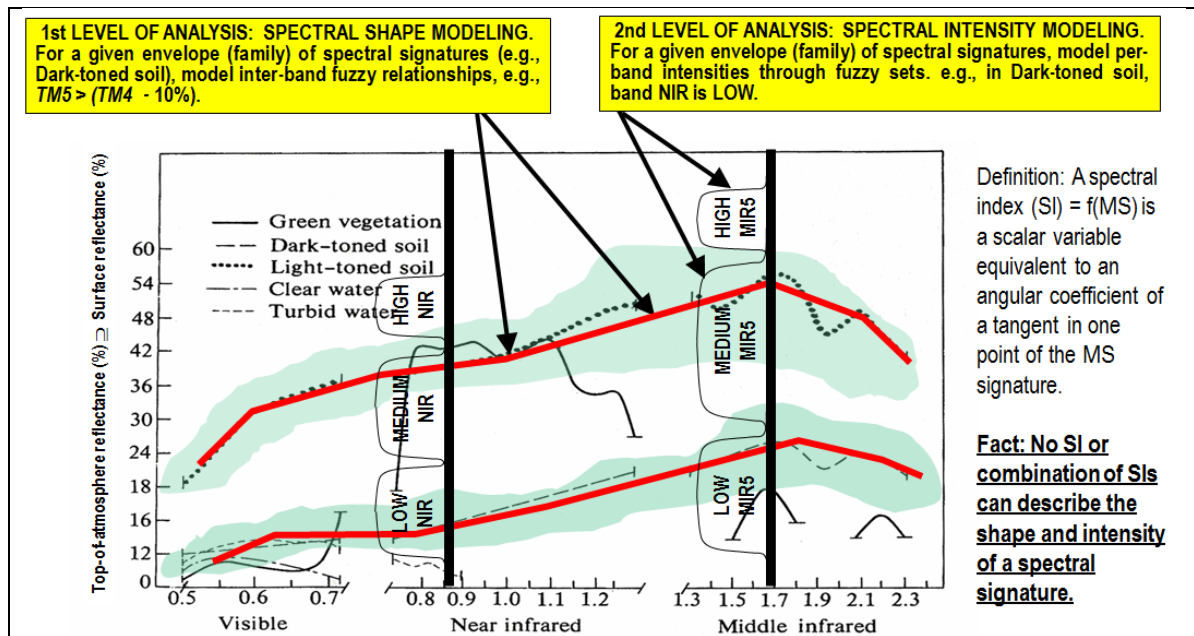


Fig. 8-2. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values which include surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF values forms a buffer zone (support area, envelope). The SIAM decision tree models each target family of spectral signatures in terms of multivariate shape and multivariate intensity as a viable alternative to multivariate analysis of spectral indexes. A typical spectral index is a scalar band ratio equivalent to an angular coefficient of a tangent in one point of the spectral signature. Infinite functions can feature the same tangent value in one point. In practice, no spectral index or combination of spectral indexes can reconstruct the multivariate shape and multivariate intensity of a spectral signature.

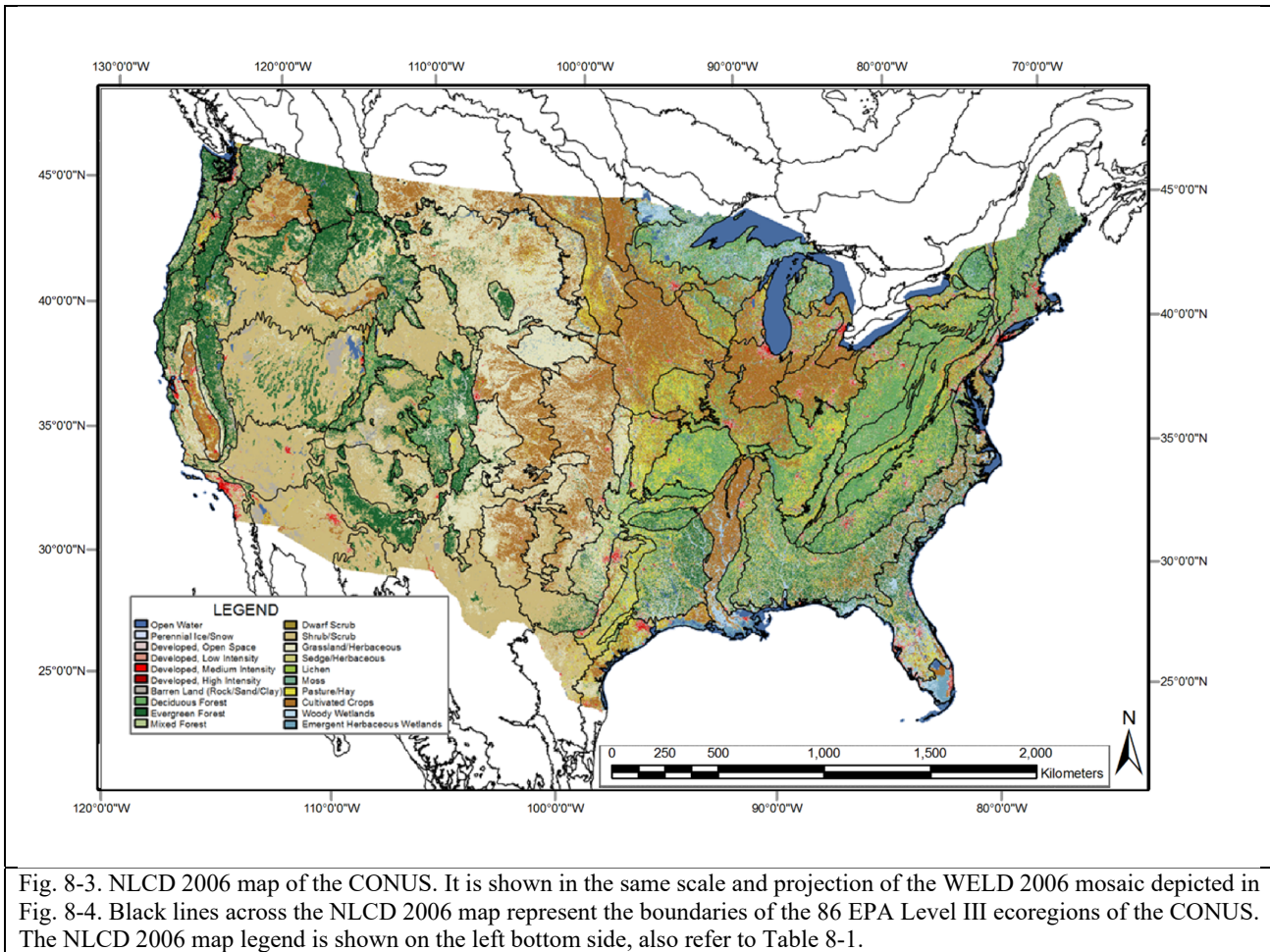


Fig. 8-3. NLCD 2006 map of the CONUS. It is shown in the same scale and projection of the WELD 2006 mosaic depicted in Fig. 8-4. Black lines across the NLCD 2006 map represent the boundaries of the 86 EPA Level III ecoregions of the CONUS. The NLCD 2006 map legend is shown on the left bottom side, also refer to Table 8-1.

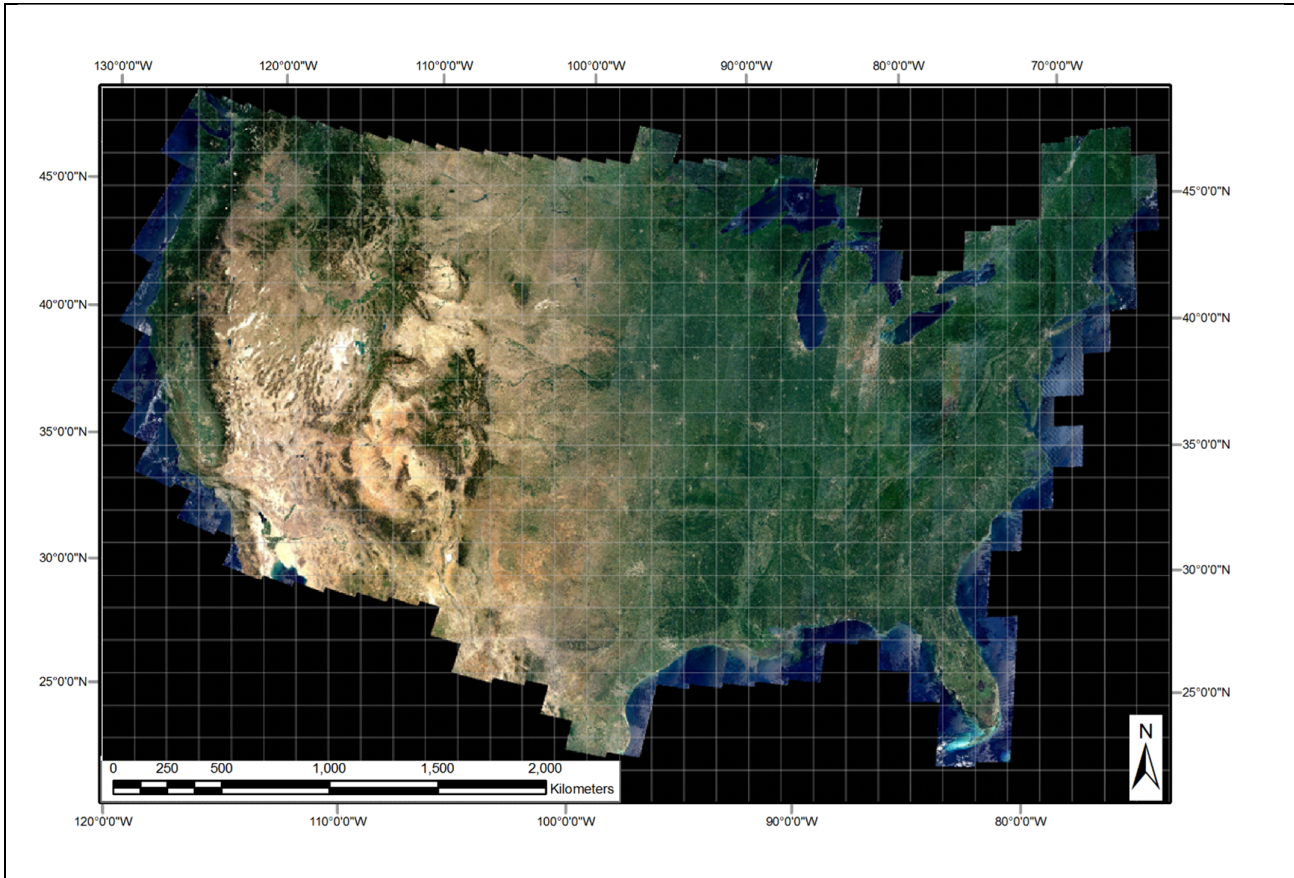


Fig. 8-4. 30 m resolution annual WELD composite for the year 2006 (December 2005 to November 2006) of the conterminous U.S. (CONUS), radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values. Depicted in true colors (red: Band 3, 0.63-0.69  $\mu\text{m}$ ; green: Band 2, 0.53-0.61  $\mu\text{m}$ , and blue: Band 1, 0.45-0.52  $\mu\text{m}$ ). The white grid shows locations of the 501 WELD tiles of the CONUS. Each tile is 5000 $\times$ 5000 pixels in size, covering a surface area of 150 $\times$ 150 km. Pixels are geographically projected in the Albers Equal Area projection.



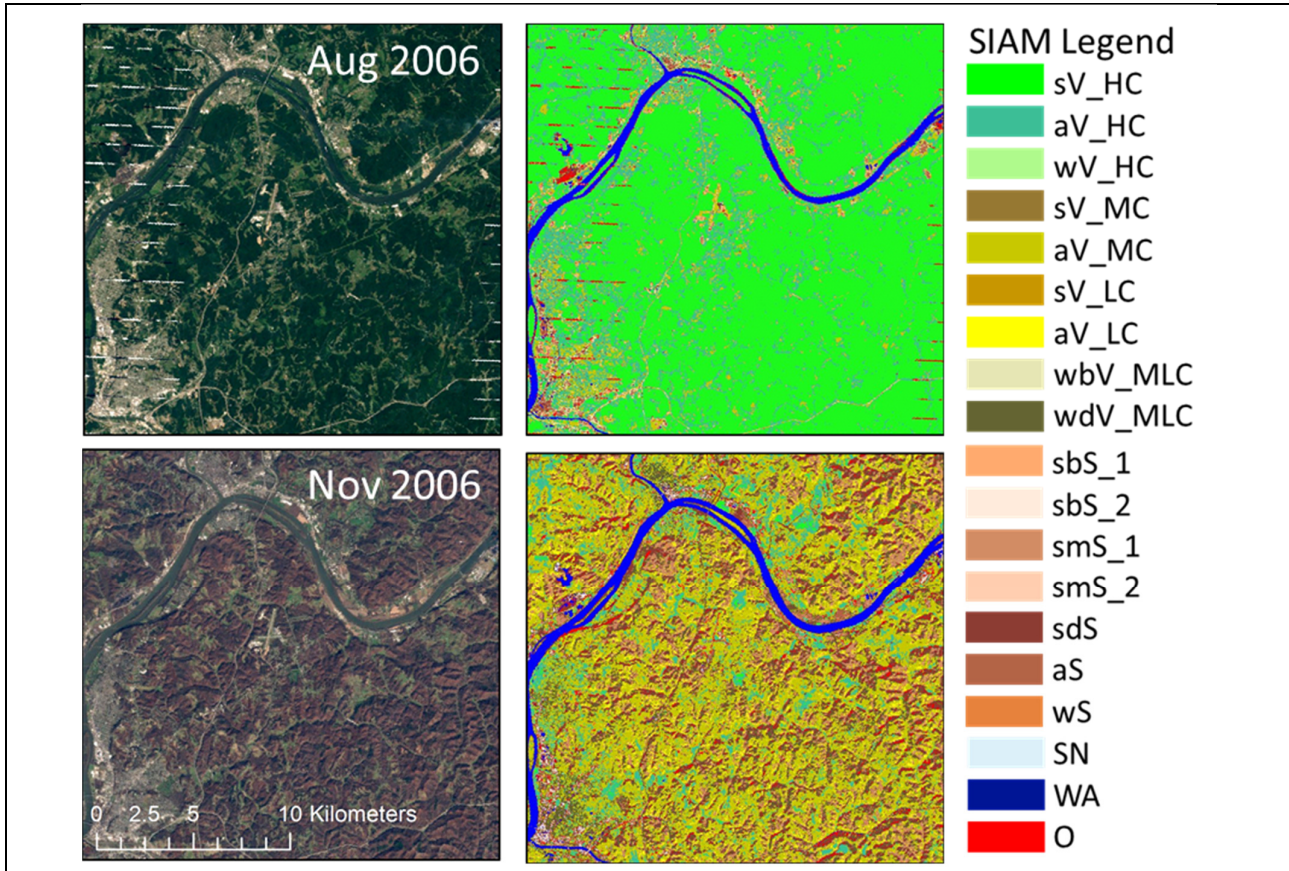


Fig. 8-5. Changes in the SIAM's spectral macro-category labeling due to vegetation phenology affecting the monthly WELD composite. *Left side*: 30 m resolution monthly WELD composites, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values, for August and November 2006, showing an area predominantly covered by broadleaf forest in the Mid-Western United States (Ohio). Depicted in true colors (red: Band 3, 0.63-0.69  $\mu\text{m}$ ; green: Band 2, 0.53-0.61  $\mu\text{m}$ , and blue: Band 1, 0.45-0.52  $\mu\text{m}$ ). To allow inter-image comparison, the two images are displayed with an identical contrast stretch. *Right side*: SIAM-WELD color maps generated from the two WELD images shown on the left side. The SIAM map legend, consisting of 19 spectral macro-categories, is shown on the right side, also refer to Table 8-2.



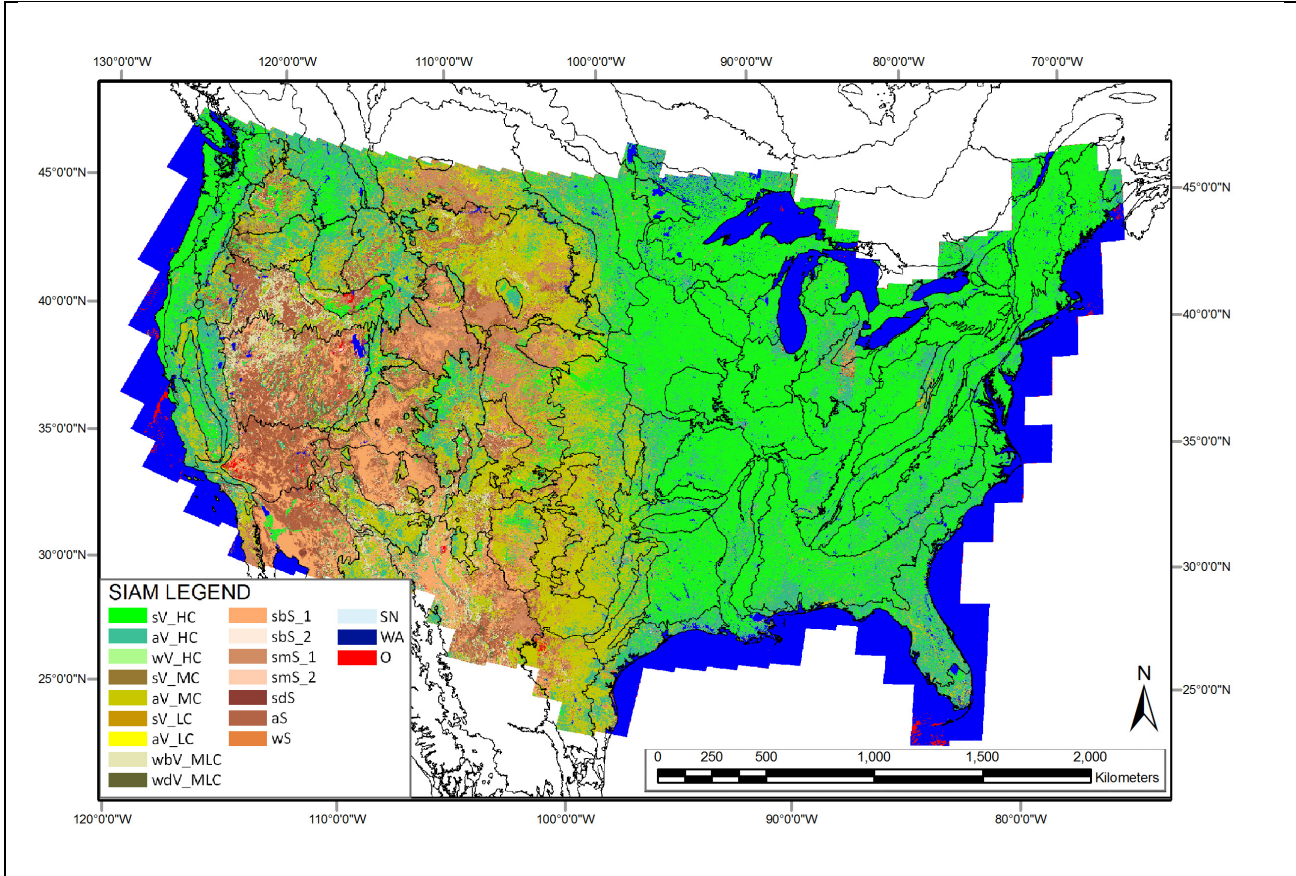
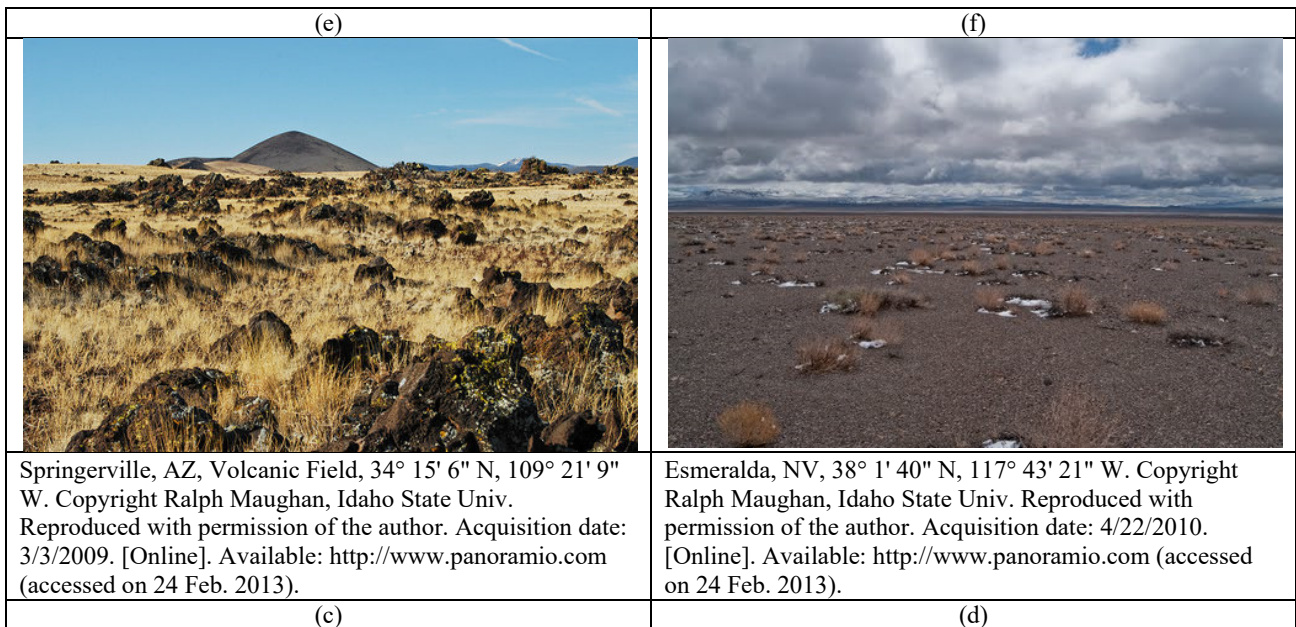


Fig. 8-6. SIAM-WELD 2006 color map at an intermediate discretization level of 48 color names, reassembled into 19 spectral macro-categories. Black lines across the SIAM-WELD 2006 map represent the boundaries of the 86 EPA Level III ecoregions of the CONUS. The SIAM map legend is shown on the left bottom corner, also refer to Table 8-2.







	
<p>Overton, NV, 36° 25' 42" N, 114° 27' 21" W. Copyright Ralph Maughan, Idaho State Univ. Reproduced with permission of the author. Acquisition date: 2/11/2009. [Online]. Available: <a href="http://www.panoramio.com">http://www.panoramio.com</a> (accessed on 24 Feb. 2013).</p>	<p>San Juan, UT, 37° 16' 43" N, 109° 40' 27" W. Copyright Ralph Maughan, Idaho State Univ. Reproduced with permission of the author. Acquisition date: 3/4/2009. [Online]. Available: <a href="http://www.panoramio.com">http://www.panoramio.com</a> (accessed on 24 Feb. 2013).</p>
<p>(a)</p>	<p>(b)</p>
	
<p>Sublette, WY, Rangeland, 42° 51' 37" N, 109° 43' 7" W. Copyright Ralph Maughan, Idaho State Univ. Reproduced with permission of the author. Acquisition date: 6/16/2011. [Online]. Available: <a href="http://www.panoramio.com">http://www.panoramio.com</a> (accessed on 24 Feb. 2013).</p>	<p>Twin Falls, ID, Ripening cheatgrass infestation, 42° 23' 52" N, 114° 21' 9" W. Copyright Ralph Maughan, Idaho State Univ. Reproduced with permission of the author. Acquisition date: April 2010? [Online]. Available: <a href="http://www.panoramio.com">http://www.panoramio.com</a> (accessed on 24 Feb. 2013).</p>
<p>Fig. 8-7. Examples of geographic locations mapped as vegetation classes "<i>Scrub/Shrub</i>" (SS) or "<i>Grassland/Herbaceous</i>" (GH) in the NLCD 2006 reference map (refer to Table 8-1) and predominantly as bare soil spectral categories (sbS_1, SmS_1, aS) in the SIAM-WELD 2006 test map (refer to Table 8-2), as pointed out in Table 8-8. The SIAM's color names sbS_1, SmS_1 and aS mean that, from space, with a pixel size of 30 m × 30 m = 900 m<sup>2</sup>, the contribution of sparse vegetation, rangeland, cheatgrass, dry long grass or short grass as foreground, mixed with a background of sand, clay or rocks, like those shown in these pictures, becomes extremely difficult to detect, especially if a hard (crisp, defuzzified) label rather than a set of fuzzy class labels must be provided as the output product.</p>	

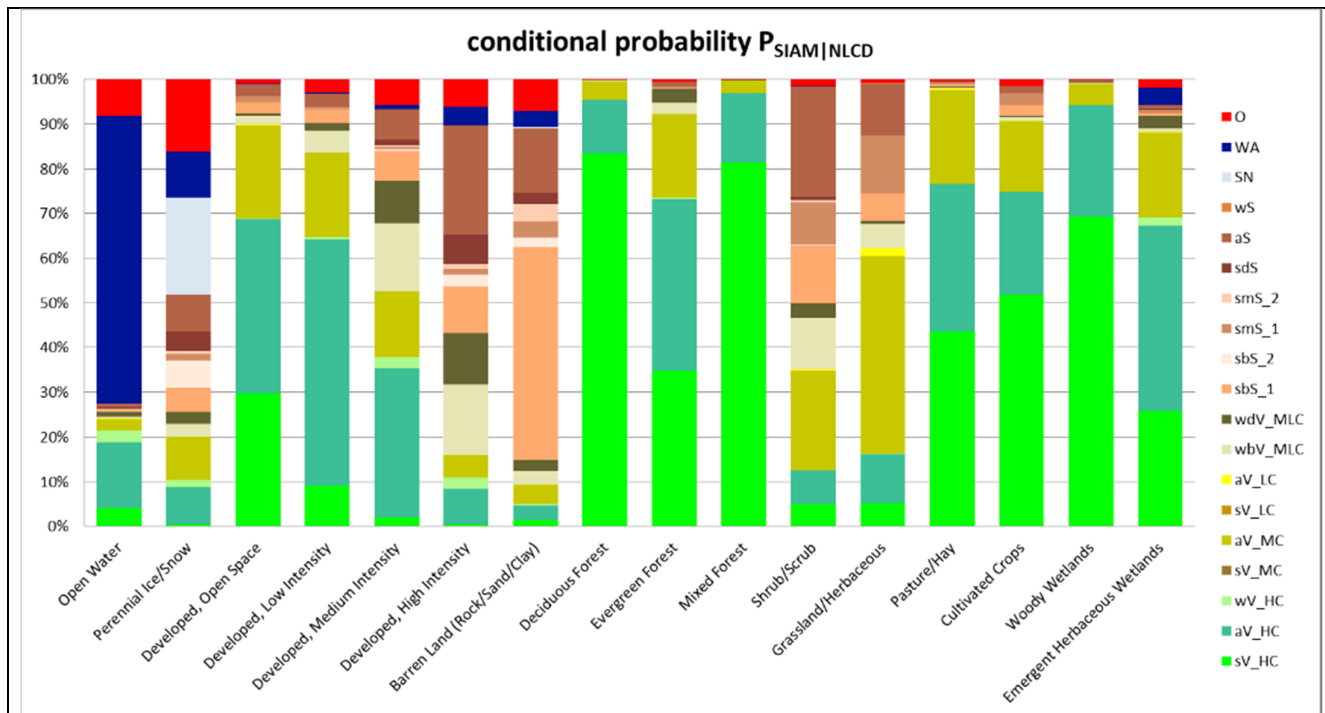


Fig. 8-8. Histogram of the conditional probabilities of the 19 SIAM-WELD 2006 spectral macro-categories (shown as the right column of acronyms, refer to Table 8-2) at the intermediate color discretization level, conditioned by one-of-16 NLCD 2006 classes (listed along the horizontal axis). These conditional probabilities are derived from Table 8-4 by normalizing each cell of Table 8-4 by its column-sum. The same conditional probabilities are summarized in text form in Table 8-6. In this histogram, pseudocolors associated with the SIAM color types make the interpretation of the histogram bins more intuitive. Green pseudocolors are associated with vegetation spectral categories (see labels of type xV\_x on the right column of labels), brown pseudocolors are selected for bare soil spectral categories (see labels of type xS\_x on the right column of labels), the pseudocolor blue is chosen for the “Water or Shadow” (WA) spectral category, the light blue pseudocolor is linked to the snow (SN) spectral category etc. As a consequence, the bin of the NLCD class “Open Water” is expected to look blue, bins of the NLCD vegetation classes are expected to look green, etc.



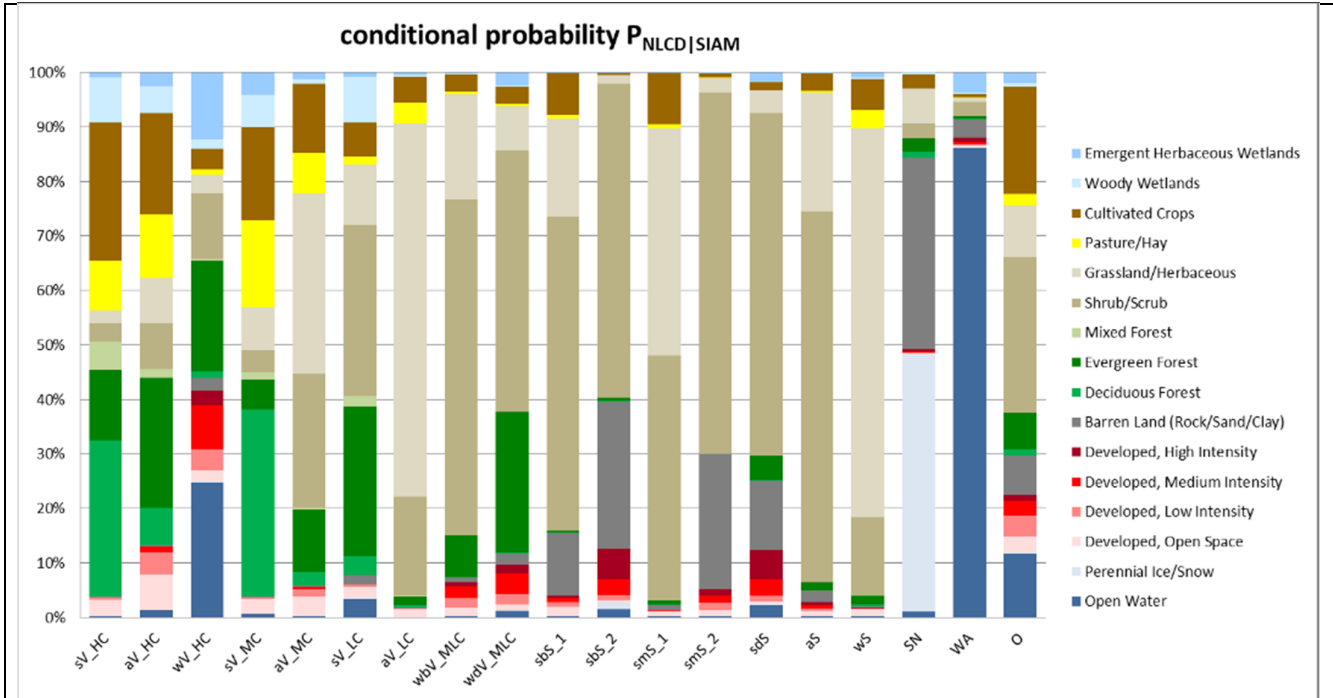


Fig. 8-9. Histogram of the conditional probabilities of the 16 NLCD 2006 classes (shown as the right column of class names) conditioned by one-of-19 SIAM-WELD 2006 spectral macro-categories (listed along the horizontal axis as acronyms, refer to Table 8-2) at the intermediate color discretization level. This histogram is derived from Table 8-4 by normalizing each cell of Table 8-4 by its row-sum. The same conditional probabilities are summarized in text form in Table 8-7. In this histogram, pseudocolors associated with the NLCD classes make the interpretation of the histogram bins more intuitive. Green pseudocolors are associated with vegetation NLCD classes, brown pseudocolors are selected for bare soil NLCD classes, the pseudocolor blue is chosen for the NLCD class “Open Water”, the light blue pseudocolor is linked to the NLCD class “Perennial Ice/Snow”, etc. As a consequence, the bin of the SIAM’s spectral category “Water or Shadow” (WA) is expected to look blue, bins of the SIAM’s vegetation spectral categories (see labels of type xV x along the horizontal axis) are expected to look green, etc.

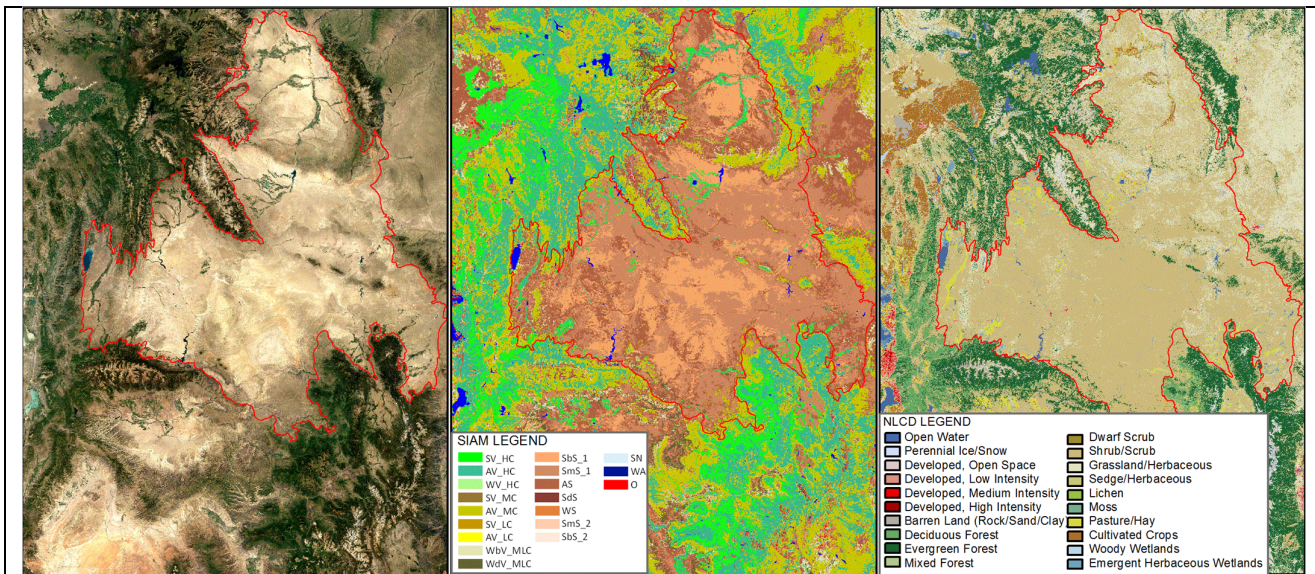
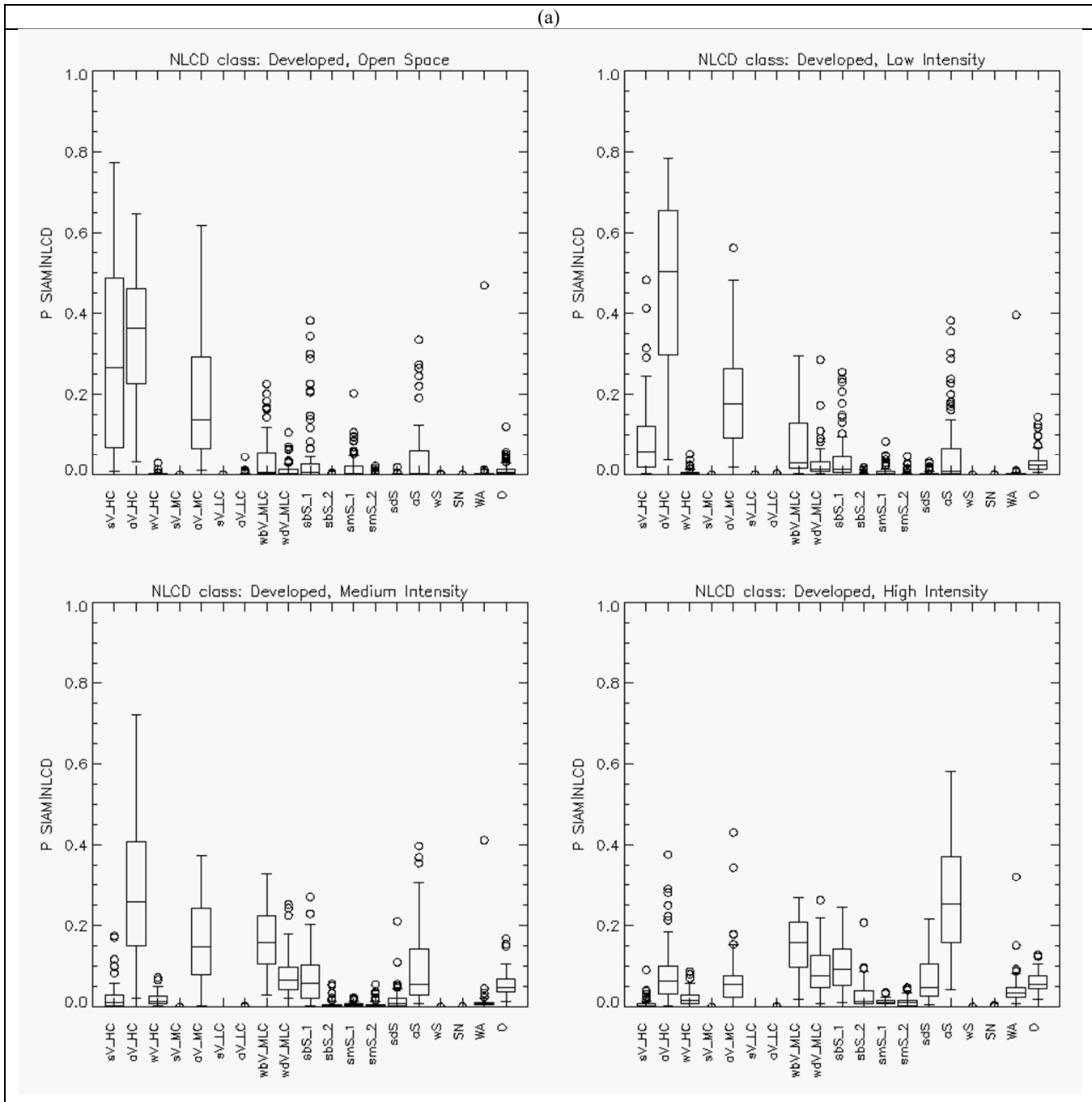


Fig. 8-10. Wyoming Basin ecoregion, as part of the "North American deserts" level 1 ecoregion 10.1.4. *Left*: WELD 2006 tile (true color). *Middle*: SIAM test map of the WELD 2006 tile shown at left, with 19 spectral macro-categories at the intermediate color discretization level. *Right*: NLCD 2006 reference map, featuring 16 LC classes. In these three images, the boundary of the Wyoming Basin ecoregion is overlaid in red. The desertic Wyoming Basin ecoregion is classified as predominantly “Scrub/Shrub”

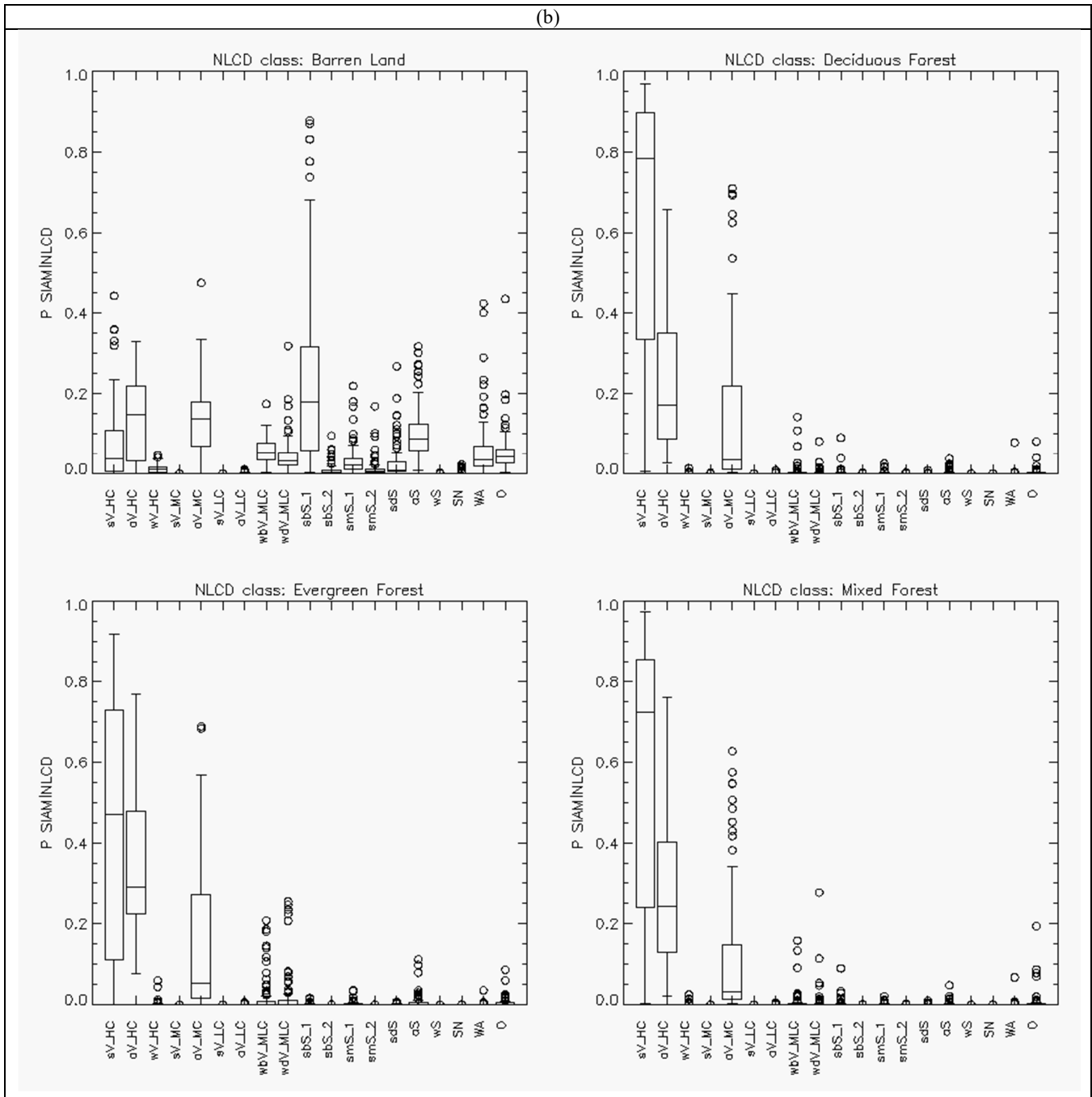


(SS) and "Grassland/Herbaceous" (GH) in the NLCD 2006 reference map (refer to Table 8-1), and predominantly as bare soil (sbS\_1, SmS\_1, aS) in the SIAM-WELD 2006 test map (refer to Table 8-2). This phenomenon of comprehensive "semantic mismatch" between the NLCD 2006 and SIAM-WELD 2006 thematic maps is explained thoroughly in Chapter 8.4.3.



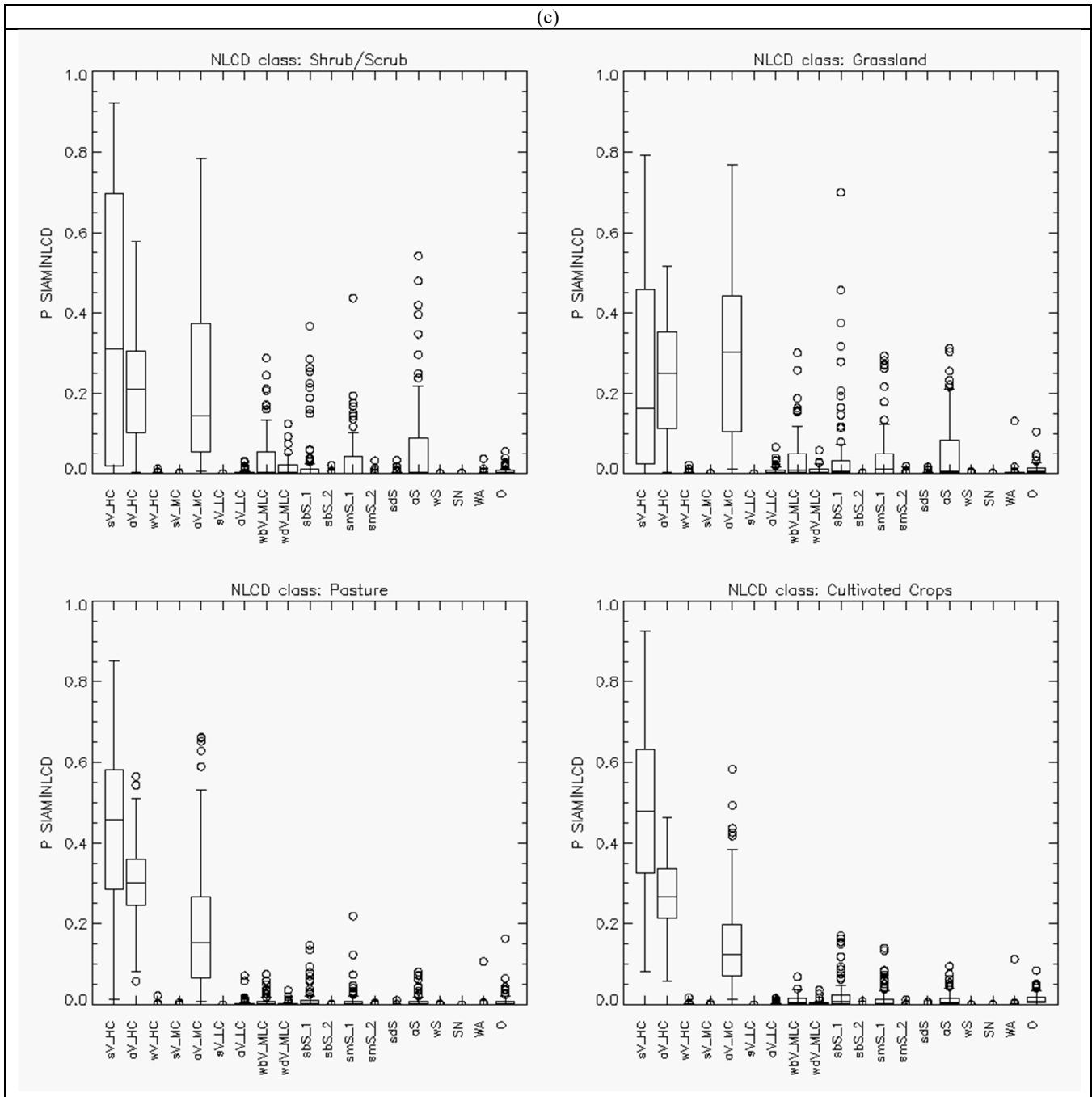


(b)





(c)



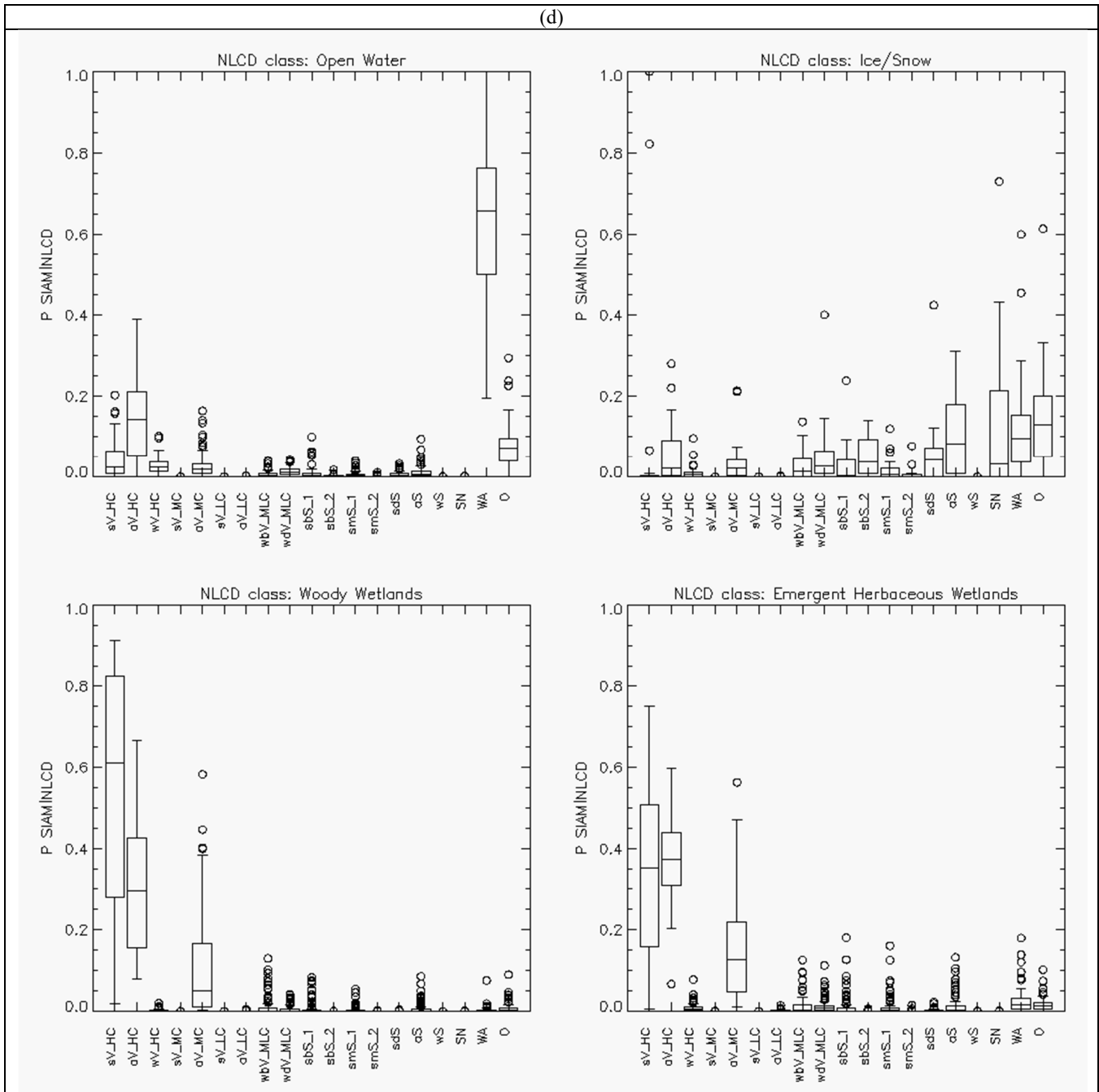


Fig. 8-11(a) - (d). Reference NLCD class-specific box-and-whisker diagrams, identified by index  $r = 1, \dots, RC = 16$ , of the NLCD class-conditional probabilities  $p(\text{SIAM-WELD}_{er,t} | \text{NLCD}_{er,r})$ , with  $t = 1, \dots, TC = 19$ , collected across ecoregions  $er = 1, \dots, ER = 86$ . The 19 spectral categories of the SIAM-WELD test map, identified by their acronyms (refer to Table 8-2), are distributed along the x axis of each NLCD class-specific diagram. Each of the 19 boxes in a box-and-whisker diagram extends from the 25<sup>th</sup> to the 75<sup>th</sup> percentile, with a horizontal line to represent the median (50<sup>th</sup> percentile) of the distribution. The whiskers extend to the maximum or minimum value of the data set, or to 1.5 times the interquartile range, whichever comes first. If there is data beyond this range, it is represented by open circles.





Tables and table captions in Chapter 8

NLCD 2001/2006/2011 Classification Scheme (Legend), Level II				LCCS-DP, level 1: A = Veg, B = Non-Veg, and level 2: 1 = Terrestrial, 2 = Aquatic
Code	ID	Name	Land cover (LC) Class Definition	ID
11	OW	Open water	OW: Areas of open water, generally with less than 25% cover of vegetation or soil	B4 - Non-vegetated aquatic
12	PIS	Perennial Ice/Snow	PIS: Areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.	B4
21 22 23 24	DOS DLI DMI DHI	<ul style="list-style-type: none"> <li>Developed, Open Space</li> <li>Developed, Low Intensity</li> <li>Developed, Medium Intensity</li> <li>Developed, High Intensity</li> </ul>	<p>DOS: Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.</p> <p>DLI, DMI, DHI: refer to the “National Land Cover Database 2006 (NLCD2006),” Multi-Resolution Land Characteristics Consortium (MRLC), 2013.</p>	B3 - Non-vegetated terrestrial / A1 - Vegetated terrestrial
31	BL	Barren Land (Rock/ Sand/ Clay)	BL: Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover. As a consequence of this constraint, class BL covers only 1.21% of the CONUS total surface.	B3
41 42 43	DF EF MF	<ul style="list-style-type: none"> <li>Deciduous Forest</li> <li>Evergreen Forest</li> <li>Mixed Forest</li> </ul>	<p>DF: Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.</p> <p>EF: Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.</p> <p>MF: Mixed Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.</p>	A1
51 52	- SS	<ul style="list-style-type: none"> <li>Dwarf Scrub <sup>2</sup></li> <li>Scrub/Shrub</li> </ul>	SS: Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions. The aforementioned definition of class BL means that class SS may feature a vegetated cover which accounts for 15% of total cover or more.	A1/ B3
71 72 73 74	GH - - -	<ul style="list-style-type: none"> <li>Grassland/Herbaceous</li> <li>Sedge</li> <li>Herbaceous <sup>2</sup></li> <li>Lichens <sup>2</sup></li> <li>Moss <sup>2</sup></li> </ul>	GH: Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing. The aforementioned definition of class BL means that class GH may feature a vegetated cover which accounts for 15% of total cover or more.	A1/B3



81 82	PH CC	<ul style="list-style-type: none"> <li>Pasture/Hay</li> <li>Cultivated Crops</li> </ul>	<p>PH: Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.</p> <p>CC: Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.</p>	A1
90 95	WW EHW	<ul style="list-style-type: none"> <li>Woody Wetlands</li> <li>Emergent Herbaceous</li> <li>Wetland</li> </ul>	<p>WW: Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.</p> <p>EHW: Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.</p>	A2 – Vegetated aquatic

Table 8-1. Definition of the NLCD 2001/2006/2011 classification taxonomy, Level II. <sup>2</sup>Alaska only. For further details, refer to the “National Land Cover Database 2006 (NLCD2006),” Multi-Resolution Land Characteristics Consortium (MRLC), 2013. The right column instantiates a possible binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{NLCD legend}$  to set  $B = \text{2-level 4-class Dichotomous Phase (DP) taxonomy of the Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) (Di Gregorio and Jansen 2000), refer to Fig. 8-1.}$

SIAM, Intermediate discretization level (featuring 48 spectral categories) reassembled into 19 spectral macro-categories				LCCS-DP, level 1: A = Veg, B = Non-Veg, and level 2: 1 = Terrestrial, 2 = Aquatic
ID	Abbreviation	OR-Aggregations	Spectral macro-category name	ID
1	sV_HC	1	Strong evidence vegetation, high canopy cover	A1 - Vegetated terrestrial
2	aV_HC	1	Average evidence vegetation, high canopy cover	A1
3	wV_HC	1	Weak evidence vegetation, high canopy cover	A1
4	sV_MC	1	Strong evidence vegetation, medium canopy cover	A1
5	aV_MC	1	Average evidence vegetation, medium canopy cover	A1
6	sV_LC	1	Strong evidence vegetation, low canopy cover	A1
7	aV_LC	1	Average evidence vegetation, low canopy cover	A1
8	wbV_MLC	1	Weak evidence bright vegetation, medium or low canopy cover	A1
9	wdV_MLC	1	Weak evidence dark vegetation, medium or low canopy cover	A1/A2 - Vegetated aquatic
10	sbS_1	1	Strong evidence bright soil AND NIR <= MIR	B3 - Non-vegetated terrestrial
11	sbS_2	1	Strong evidence bright soil AND NIR > MIR	B3
12	smS_1	1	Strong evidence medium soil AND NIR <= MIR	B3
13	smS_2	1	Strong evidence medium soil AND NIR > MIR	B3
14	sdS	1	Strong evidence dark soil	B3
15	aS	1	Average evidence soil	B3
16	wS	1	Weak evidence soil	B3
17	SN	2	Snow	B4 - Non-vegetated aquatic
18	WA	6	Water or Shadow	B4



19	O	24	Others	B - Non-vegetated
	<b>TOT.</b>	48		

Table 8-2. List of the 19 spectral macro-categories generated from the aggregation of the SIAM's 48 spectral categories originally detected at the intermediate level of color quantization. The "Water or Shadow" (WA) spectral macro-category results from the aggregation of six original SIAM categories, the "Snow" (SN) spectral macro-category from two and the spectral macro-category "Others" (O) from the aggregation of 24 original spectral categories covering disturbances typically minimized or removed in an annual composite (clouds, smoke plumes, fire fronts, etc.) as well as the original spectral category "Unknowns". Hence,  $(19 - 3) + 6 + 2 + 24 = 48$ , which is the SIAM's intermediate discretization level. In the proposed names of spectral macro-categories, acronym Near Infra-Red (NIR) indicates Landsat TM/ETM+ band 4 ( $0.9 \mu\text{m}$ ) and Medium Infra-Red (MIR) indicates Landsat TM/ETM+ band 5 ( $1.6 \mu\text{m}$ ). The right column instantiates a possible binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{SIAM legend}$  to set  $B = 2\text{-level } 4\text{-class Dichotomous Phase (DP) taxonomy of the Food and Agriculture Organization of the United Nations (FAO) - Land Cover Classification System (LCCS) (Di Gregorio and Jansen 2000), refer to Fig. 8-1.$

Spectral Category	2006	2007	2008	2009	Mean	Std Dev
sV HC	33.11%	32.56%	33.79%	34.06%	33.38%	0.68%
aV HC	19.94%	23.31%	20.02%	20.86%	21.03%	1.57%
wV HC	0.18%	0.17%	0.17%	0.19%	0.18%	0.01%
sV MC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
aV MC	20.05%	18.79%	18.07%	17.93%	18.71%	0.97%
sV LC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
aV LC	0.40%	0.18%	0.31%	0.22%	0.28%	0.10%
wbV MLC	4.12%	3.60%	3.73%	3.30%	3.69%	0.34%
wdV MLC	1.48%	1.37%	1.19%	1.53%	1.39%	0.15%
<i>Total vegetation</i>	<i>79.28%</i>	<i>79.98%</i>	<i>77.29%</i>	<i>78.10%</i>	<i>78.66%</i>	<i>1.20%</i>
sbS 1	5.00%	5.44%	6.28%	5.39%	5.53%	0.54%
sbS 2	0.09%	0.13%	0.08%	0.12%	0.11%	0.02%
smS 1	4.65%	3.51%	5.38%	4.78%	4.58%	0.78%
smS 2	0.19%	0.16%	0.24%	0.20%	0.20%	0.03%
sdS	0.25%	0.28%	0.25%	0.29%	0.27%	0.02%
aS	8.04%	8.18%	8.24%	8.70%	8.29%	0.29%
wS	0.02%	0.01%	0.01%	0.01%	0.01%	0.00%
<i>Total soils</i>	<i>18.24%</i>	<i>17.71%</i>	<i>20.48%</i>	<i>19.49%</i>	<i>18.98%</i>	<i>1.25%</i>
SN	0.01%	0.01%	0.02%	0.01%	0.01%	0.01%
WA	1.28%	1.28%	1.25%	1.27%	1.27%	0.02%
O	1.19%	1.02%	0.96%	1.13%	1.07%	0.10%

Table 8-3. Spectral category-specific percentage of occurrences in the SIAM-WELD 2006/ 2007/ 2008/ 2009 test maps at the intermediate level of color quantization, where 48 color names were aggregated into 19 spectral macro-categories by an independent human expert. Adopted acronyms for the 19 spectral macro-categories: refer to Table 8-2.



2006 OAMTRX, Probabilities (%). Rows: SIAM™-WELD 2006, 19 spectral categories; Columns: NLCD 2006, 16 land cover classes.																		
NLCD code	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95		
NLCD class	OW	PIS	DOS	DLI	DMI	DHI	BL	DF	EF	MF	SS	GH	PH	CC	WW	EHW		
LCCD-DP1&2	B4	B4	B3->A1	B3->A1	B3->A1	B3->A1	B3	A1	A1	A1	A1->B3	A1->B3	A1	A1	A2	A2		
SIAM™ Intermediate Granularity, 19 Spectral Categories.	sV_HC	0.07%	0.00%	1.00%	0.13%	0.01%	0.00%	0.02%	9.51%	4.29%	1.73%	1.10%	0.77%	3.04%	8.35%	2.76%	0.31%	
	aV_HC	0.25%	0.00%	1.31%	0.82%	0.20%	0.02%	0.04%	1.38%	4.75%	0.33%	1.67%	1.66%	2.32%	3.71%	0.99%	0.50%	
	wV_HC	0.04%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%	0.04%	0.00%	0.02%	0.01%	0.00%	0.01%	0.00%	0.02%	
	sV_MC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	aV_MC	0.04%	0.00%	0.70%	0.28%	0.09%	0.01%	0.05%	0.48%	2.31%	0.06%	4.93%	6.66%	1.47%	2.55%	0.19%	0.23%	
	sV_LC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	aV_LC	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.07%	0.27%	0.02%	0.02%	0.00%	0.00%	
	wbV_MLC	0.01%	0.00%	0.07%	0.07%	0.09%	0.03%	0.04%	0.00%	0.31%	0.00%	2.54%	0.79%	0.02%	0.13%	0.01%	0.01%	
	wdV_MLC	0.02%	0.00%	0.02%	0.06%	0.06%	0.02%	0.03%	0.01%	0.38%	0.00%	0.71%	0.12%	0.01%	0.05%	0.00%	0.03%	
	sbS_1	0.01%	0.00%	0.09%	0.04%	0.04%	0.02%	0.57%	0.00%	0.01%	0.00%	2.88%	0.90%	0.03%	0.38%	0.01%	0.01%	
	sbS_2	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	
	smS_1	0.01%	0.00%	0.05%	0.01%	0.00%	0.00%	0.04%	0.01%	0.04%	0.00%	2.09%	1.94%	0.03%	0.43%	0.00%	0.01%	
	smS_2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	
	sdS	0.01%	0.00%	0.00%	0.00%	0.01%	0.01%	0.03%	0.00%	0.01%	0.00%	0.16%	0.01%	0.00%	0.00%	0.00%	0.00%	
	aS	0.01%	0.00%	0.08%	0.04%	0.04%	0.05%	0.17%	0.01%	0.12%	0.00%	5.47%	1.76%	0.02%	0.26%	0.01%	0.01%	
	wS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	
	SN	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	WA	1.10%	0.00%	0.00%	0.00%	0.01%	0.01%	0.04%	0.00%	0.00%	0.00%	0.03%	0.01%	0.00%	0.01%	0.00%	0.05%	
	O	0.14%	0.00%	0.04%	0.04%	0.03%	0.01%	0.09%	0.01%	0.08%	0.00%	0.34%	0.11%	0.03%	0.23%	0.01%	0.02%	
	1.71%	0.02%	3.36%	1.48%	0.59%	0.20%	1.21%	11.41%	12.35%	2.13%	22.19%	15.03%	6.99%	16.12%	3.99%	1.21%		

Table 8-4. OAMTRX instance generated from a wall-to-wall overlap between the test SIAM-WELD 2006 map of the CONUS with legend A = 19 spectral macro-categories and the reference NLCD 2006 map with legend B = 16 LC class names. Gray entry-pair cells identify the binary relationship  $R: A \Rightarrow B \subseteq A \times B$  chosen by the independent human expert to guide the interpretation process of the OAMTRX = FrequencyCount(A × B). Statistically independent TQ<sup>2</sup>Is are CVPAl2(R: A ⇒ B) = 0.6689 and OA(OAMTRX) = 96.88%. Adopted acronyms for reference LC classes and test spectral macro-categories are described in Table 8-1 and Table 8-2 respectively.

	NLCD Code (class acronym), 16 classes	41 (DF), 42 (EF), 43 (MF), 52 (SS) <-> B3, 71 (GH) <-> B3, 81 (PH), 82 (CC)		21 (DOS) <-> A1, 22 (DLI) <-> A1, 23 (DMI) <-> A1, 24 (DHI) <-> A1, 31 (BL)		11 (OW), 12 (PIS)		Sum per row
		≈ A1	≈ A2	≈ B3	≈ B4			
	LCCS-DP1&2 Code, 4 classes	Veg terstrl	Veg aqutc	Non-veg terstrl	Non-veg aqutc			
SIAM™ Intermediate Granularity, 19 Spectral Categories.	sV_HC	28.80%	3.07%	1.17%	0.07%			33.11%
	aV_HC	15.82%	1.49%	2.38%	0.25%			19.94%
	wV_HC	0.08%	0.03%	0.03%	0.04%			0.18%
	sV_MC	0.00%	0.00%	0.00%	0.00%			0.00%
	aV_MC	18.45%	0.42%	1.13%	0.05%			20.05%
	sV_LC	0.00%	0.00%	0.00%	0.00%			0.00%
	aV_LC	0.39%	0.00%	0.01%	0.00%			0.40%
	wbV_MLC	3.80%	0.02%	0.30%	0.01%			4.12%
	wdV_MLC	1.27%	0.04%	0.15%	0.02%			1.48%
	sbS_1	4.21%	0.01%	0.76%	0.01%			5.00%
	sbS_2	0.06%	0.00%	0.03%	0.00%			0.09%
	smS_1	4.53%	0.01%	0.10%	0.01%			4.65%
	smS_2	0.13%	0.00%	0.06%	0.00%			0.19%
	sdS	0.18%	0.00%	0.06%	0.01%			0.25%
	aS	7.63%	0.02%	0.38%	0.01%			8.04%
	wS	0.02%	0.00%	0.00%	0.00%			0.02%
	SN	0.00%	0.00%	0.00%	0.00%			0.01%
WA	0.06%	0.05%	0.07%	1.11%			1.28%	
O	0.80%	0.03%	0.21%	0.14%			1.19%	
	Sum per column	86.23%	5.20%	6.84%	1.73%			

Table 8-5. OAMTRX instance generated from a wall-to-wall overlap between the test SIAM-WELD 2006 map of the CONUS with legend C = 19 spectral macro-categories and the reference NLCD 2006 map with legend B = 2-level 4-class LCCS-DP names. Gray entry-pair cells identify the binary relationship  $R: C \Rightarrow B \subseteq C \times B$  chosen by the present authors to guide the interpretation process of the OAMTRX = FrequencyCount(C × B). Statistically independent TQ<sup>2</sup>Is





are  $CVP_{AI2}(R: C \Rightarrow B) = 0.7486$  and  $OA(OAMTRX) = 93.09\%$ . Adopted acronyms for reference LC classes and test spectral macro-categories are described in Fig. 8-1 and Table 8-2 respectively.

NLCD class	SIAM1	$P_{SIAM1 NLCD}$	SIAM2	$P_{SIAM2 NLCD}$	SIAM3	$P_{SIAM3 NLCD}$	SIAM4	$P_{SIAM4 NLCD}$	SIAM5	$P_{SIAM5 NLCD}$
Open Water	WA	0.64	aV_HC	0.15	O	0.08	sV_HC	0.04	wV_HC	0.03
Ice/Snow	SN	0.22	O	0.16	WA	0.10	aV_MC	0.10	aS	0.08
Developed, Open	aV_HC	0.39	sV_HC	0.30	aV_MC	0.21	sbS_1	0.03	aS	0.02
Developed, Low	aV_HC	0.55	aV_MC	0.19	sV_HC	0.09	wbV_MLC	0.05	sbS_1	0.03
Developed, Medium	aV_HC	0.33	wbV_MLC	0.15	aV_MC	0.15	wdV_MLC	0.10	aS	0.07
Developed, High	aS	0.24	wbV_MLC	0.16	wdV_MLC	0.11	sbS_1	0.11	aV_HC	0.08
Rock/Sand/Clay	sbS_1	0.48	aS	0.14	O	0.07	aV_MC	0.04	smS_2	0.04
Deciduous Forest	sV_HC	0.83	aV_HC	0.12	aV_MC	0.04	O	0.00	wdV_MLC	0.00
Evergreen Forest	aV_HC	0.38	sV_HC	0.35	aV_MC	0.19	wdV_MLC	0.03	wbV_MLC	0.03
Mixed Forest	sV_HC	0.81	aV_HC	0.16	aV_MC	0.03	O	0.00	wbV_MLC	0.00
Shrub/Scrub	aS	0.25	aV_MC	0.22	sbS_1	0.13	wbV_MLC	0.11	smS_1	0.09
Grassland/Herbaceous	aV_MC	0.44	smS_1	0.13	aS	0.12	aV_HC	0.11	sbS_1	0.06
Pasture/Hay	sV_HC	0.43	aV_HC	0.33	aV_MC	0.21	sbS_1	0.00	smS_1	0.00
Cultivated Crops	sV_HC	0.52	aV_HC	0.23	aV_MC	0.16	smS_1	0.03	sbS_1	0.02
Woody Wetlands	sV_HC	0.69	aV_HC	0.25	aV_MC	0.05	wbV_MLC	0.00	O	0.00
Herbaceous Wetlands	aV_HC	0.41	sV_HC	0.26	aV_MC	0.19	WA	0.04	wdV_MLC	0.03

Table 8-6. Class-conditional probability  $p(SIAM-WELD_t | NLCD_r)$ ,  $t = 1, \dots, TC = |A| = 19$  color names,  $r = 1, \dots, RC = |B| = 16$  LC class names. For each NLCD 2006 reference map's LC class, the five best-matching SIAM-WELD 2006 test map's spectral macro-categories, belonging to the finite set A of 19 spectral macro-categories, are shown as SIAM1 to SIAM5.

SIAM	NLCD 1	$P_{NLCD1 SIAM}$	NLCD 2	$P_{NLCD2 SIAM}$	NLCD 3	$P_{NLCD3 SIAM}$	NLCD 4	$P_{NLCD4 SIAM}$	NLCD 5	$P_{NLCD5 SIAM}$
sV_HC	Deciduous Forest	0.29	Cultivated Crops	0.25	Evergreen Forest	0.13	Pasture/Hay	0.09	Woody Wetlands	0.08
aV_HC	Evergreen Forest	0.24	Cultivated Crops	0.19	Pasture/Hay	0.12	Shrub/Scrub	0.08	Grassland/Herbaceous	0.08
wV_HC	Open Water	0.25	Evergreen Forest	0.20	Herbaceous Wetlands	0.12	Shrub/Scrub	0.12	Developed, Medium	0.08
sV_MC	Deciduous Forest	0.33	Cultivated Crops	0.17	Pasture/Hay	0.15	Grassland/Herbaceous	0.08	Woody Wetlands	0.06
aV_MC	Grassland/Herbaceous	0.33	Shrub/Scrub	0.25	Cultivated Crops	0.13	Evergreen Forest	0.12	Pasture/Hay	0.07
sV_LC	Shrub/Scrub	0.04	Evergreen Forest	0.03	Grassland/Herbaceous	0.01	Woody Wetlands	0.01	Cultivated Crops	0.01
aV_LC	Grassland/Herbaceous	0.68	Shrub/Scrub	0.18	Cultivated Crops	0.05	Pasture/Hay	0.04	Evergreen Forest	0.02
wbV_MLC	Shrub/Scrub	0.62	Grassland/Herbaceous	0.19	Evergreen Forest	0.08	Cultivated Crops	0.03	Developed, Medium	0.02
wdV_MLC	Shrub/Scrub	0.48	Evergreen Forest	0.26	Grassland/Herbaceous	0.08	Developed, Medium	0.04	Cultivated Crops	0.03
sbS_1	Shrub/Scrub	0.58	Grassland/Herbaceous	0.18	Rock/Sand/Clay	0.11	Cultivated Crops	0.08	Developed, Open	0.02
sbS_2	Shrub/Scrub	0.58	Rock/Sand/Clay	0.27	Developed, High	0.06	Developed, Medium	0.03	Grassland/Herbaceous	0.02
smS_1	Shrub/Scrub	0.45	Grassland/Herbaceous	0.42	Cultivated Crops	0.09	Developed, Open	0.01	Rock/Sand/Clay	0.01
smS_2	Shrub/Scrub	0.66	Rock/Sand/Clay	0.25	Grassland/Herbaceous	0.03	Developed, Low	0.01	Developed, Medium	0.01
sdS	Shrub/Scrub	0.63	Rock/Sand/Clay	0.13	Developed, High	0.05	Evergreen Forest	0.04	Grassland/Herbaceous	0.04
aS	Shrub/Scrub	0.68	Grassland/Herbaceous	0.22	Cultivated Crops	0.03	Rock/Sand/Clay	0.02	Evergreen Forest	0.01
wS	Grassland/Herbaceous	0.71	Shrub/Scrub	0.14	Cultivated Crops	0.06	Pasture/Hay	0.03	Evergreen Forest	0.02
SN	Ice/Snow	0.47	Rock/Sand/Clay	0.35	Grassland/Herbaceous	0.06	Shrub/Scrub	0.03	Cultivated Crops	0.03
WA	Open Water	0.86	Herbaceous Wetlands	0.04	Rock/Sand/Clay	0.03	Shrub/Scrub	0.03	Grassland/Herbaceous	0.01
O	Shrub/Scrub	0.28	Cultivated Crops	0.20	Open Water	0.12	Grassland/Herbaceous	0.09	Rock/Sand/Clay	0.07

Table 8-7. Class-conditional probability  $p(NLCD_r | SIAM-WELD_t)$ ,  $t = 1, \dots, TC = |A| = 19$  color names,  $r = 1, \dots, RC = |B| = 16$  LC class names. For each SIAM-WELD 2006 test map's spectral macro-category, the five best-matching NLCD 2006 reference map's LC classes, belonging to the finite set B of 16 LC classes, are shown as NLCD1 to NLCD5.



Wyoming Basin Ecoregion, 2006 OAMTRX. Probabilities (%). Rows: SIAM™-WELD 2006, 19 spectral categories; Columns: NLCD 2006, 16 land cover classes.																		
NLCD code	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95		
NLCD class	OW	PIS	DOS	DLI	DMI	DHI	BL	DF	EF	MF	SS	GH	PH	CC	WW	EHW		
LCCD-DP1&2	B4	B4	B3->A1	B3->A1	B3->A1	B3->A1	B3	A1	A1	A1	A1->B3	A1->B3	A1	A1	A2	A2		
sV_HC	0.00%	0.00%	0.03%	0.01%	0.00%	0.00%	0.00%	0.03%	0.02%	0.01%	0.06%	0.02%	1.04%	0.23%	0.09%	0.11%		1.65%
aV_HC	0.02%	0.00%	0.05%	0.03%	0.01%	0.00%	0.00%	0.07%	0.43%	0.02%	0.41%	0.13%	1.05%	0.16%	0.38%	0.35%		3.11%
wV_HC	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.01%	0.00%	0.00%	0.00%	0.01%	0.00%		0.05%
sV_MC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		0.00%
aV_MC	0.01%	0.00%	0.06%	0.02%	0.00%	0.00%	0.00%	0.08%	0.68%	0.01%	3.47%	1.21%	0.60%	0.06%	0.23%	0.42%		6.85%
sV_LC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		0.00%
aV_LC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%	0.01%	0.00%	0.00%	0.00%	0.00%		0.07%
wbV_MLC	0.01%	0.00%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.18%	0.00%	2.35%	0.55%	0.04%	0.02%	0.05%	0.07%		3.35%
wdV_MLC	0.01%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.14%	0.00%	0.57%	0.05%	0.01%	0.00%	0.01%	0.01%		0.81%
sbS_1	0.01%	0.00%	0.12%	0.04%	0.01%	0.00%	0.69%	0.00%	0.01%	0.00%	16.90%	5.66%	0.09%	0.04%	0.04%	0.11%		23.72%
sbS_2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%		0.03%
smS_1	0.00%	0.00%	0.12%	0.01%	0.00%	0.00%	0.03%	0.00%	0.06%	0.00%	32.71%	4.11%	0.04%	0.01%	0.02%	0.09%		37.20%
smS_2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.04%	0.02%	0.00%	0.00%	0.00%	0.00%		0.07%
sdS	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	0.01%	0.00%	0.00%	0.00%	0.00%		0.10%
aS	0.02%	0.00%	0.17%	0.07%	0.01%	0.00%	0.08%	0.00%	0.13%	0.00%	17.86%	3.25%	0.05%	0.02%	0.03%	0.10%		21.79%
wS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		0.00%
SN	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		0.00%
WA	0.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.06%	0.01%	0.00%	0.00%	0.01%	0.01%		0.64%
O	0.02%	0.00%	0.01%	0.01%	0.00%	0.00%	0.01%	0.00%	0.01%	0.00%	0.31%	0.06%	0.05%	0.03%	0.02%	0.03%		0.56%
	0.65%	0.00%	0.61%	0.22%	0.04%	0.01%	0.83%	0.18%	1.67%	0.04%	74.91%	15.10%	2.97%	0.57%	0.89%	1.30%		100.00%

Table 8-8. OAMTRX instance generated from a wall-to-wall overlap over the Wyoming Basin Ecoregion between the test SIAM-WELD 2006 map with legend A = 19 spectral macro-categories and the reference NLCD 2006 map with legend B = 16 LC class names. Gray squares identify the binary relationship  $R: A \Rightarrow B \subseteq A \times B$  chosen by the independent human expert to guide the interpretation process of the  $OAMTRX = \text{FrequencyCount}(A \times B)$ , same as in Table 8-4. The Wyoming Basin Ecoregion is predominantly desertic. It is classified as LC class "Scrub/Shrub" (SS) or LC class "Grassland/Herbaceous" (GH) in the NLCD 2006 reference map (refer to Table 8-1), and predominantly as spectral macro-categories of bare soil (sbS\_1, smS\_1, aS) in the SIAM-WELD 2006 test map (refer to Table 8-2). This phenomenon of large-scale "conceptual mismatch" between the NLCD 2006 and SIAM-WELD 2006 thematic maps is discussed thoroughly in Chapter 8.4.3.

## 9 Manuscript 6 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01940): Automated Near Real-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images

### Motivation and Contributions to the Dissertation

Among the original pair of expert systems (prior knowledge-based decision trees) for color naming presented in Chapter 3 (Technical report 1) and adopted by an Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the RGB Image Automatic Mapper™ (RGBIAM™) lightweight computer program was designed and implemented to accomplish true- or false-color RGB cube polyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time, i.e., in linear time complexity monotonically increasing with image size. In the present Chapter 9 (Manuscript 6) the RGBIAM lightweight computer program pipeline, including an original statistical model-based self-organizing color constancy algorithm required when an RGB image is not radiometrically calibrated, is presented and discussed in detail.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 9 (Manuscript 6) are color filled.

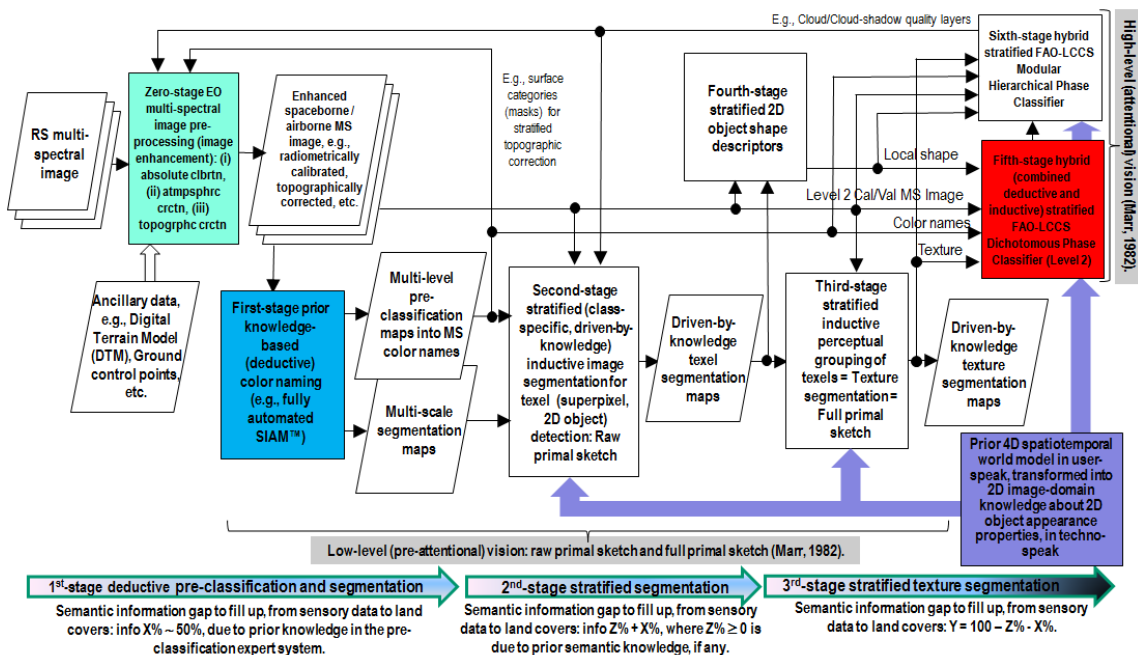


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 9 (Manuscript 6).



## Automated Linear-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images

Andrea Baraldi<sup>a,b,\*</sup>, Dirk Tiede<sup>b</sup>, and Stefan Lang<sup>b</sup>

<sup>a</sup> Department of Agricultural Sciences, University of Naples Federico II, Portici (NA), Italy.

<sup>b</sup> Department of Geoinformatics – Z\_GIS, University of Salzburg, Salzburg 5020, Austria.

\* Corresponding author. Email: andrea6311@gmail.com

### Abstract

In this methodological paper, provided with a relevant survey value, an original low-level computer vision (CV) software pipeline, called RGB Image Automatic Mapper™ (RGBIAM™), is presented and discussed. RGBIAM is a lightweight computer program capable of automated near real-time superpixel detection and quality assessment in an uncalibrated monitor-typical red-green-blue (RGB) image, depicted in either true- or false-colors. The RGBIAM system design consists of known CV modules, constrained by the Calibration/Validation (*Cal/Val*) requirements of the Quality Assurance Framework for Earth Observation (QA4EO) guidelines. In agreement with the QA4EO *Cal* requirements, to benefit from multi-source data harmonization and interoperability, RGBIAM requires as mandatory an uncalibrated RGB image pre-processing first stage, consisting of an automated (self-organizing) statistical model-based color constancy algorithm. The RGBIAM's *hybrid* (combined deductive/top-down and inductive/bottom-up) inference pipeline comprises: (I) a direct quantitative-to-nominal (QN) RGB variable transform, where RGB pixel values are mapped (quantized) onto a prior dictionary of color names, to be community-agreed upon in advance, equivalent to a static (non-adaptive to data) polyhedralization of the RGB cube. Prior color naming is the deductive counterpart of popular inductive vector quantization (VQ) algorithms, whose typical VQ error function to minimize is a root mean square error (RMSE). In the output multi-level color map domain, superpixels are automatically detected in linear time as connected sets of pixels featuring the same color label. (II) An inverse nominal-to-quantitative (NQ) RGB variable transform, where a superpixelwise-constant RGB image approximation is generated in linear time, which allows to assess a VQ error image in compliance with the QA4EO *Val* requirements. The hybrid direct and inverse RGBIAM-QNQ transform is: (i) general-purpose, i.e., data- and application-independent. (ii) Automated, i.e., it requires no user-machine interaction. In the hybrid RGBIAM pipeline, a deductive inference first stage, analogous to genotype, provides automatically inherently ill-posed inductive learning-from-data algorithms, equivalent to phenotype, with initial conditions. (iii) Near real-time, i.e., its computational complexity increases linearly with image size. (iv) Implemented in tile streaming mode, to cope with massive images. As a proof-of-concept, a realization of the RGBIAM pipeline was tested on three RGB images acquired by different imaging sensors and acquisition platforms. Collected outcome and process quantitative quality indicators, including degree of automation, computational efficiency, VQ rate and VQ error, are consistent with theoretical expectations and reveal that the RGBIAM lightweight computer program is suitable for low-level CV mobile software applications specifically designed to run on web browsers and mobile devices, such as tablet computers, smartphones and Unmanned Aerial Systems (UASs) mounting low-weight consumer-level color cameras.

### Index Terms

Cognitive science, color naming, contour detection, edge-preserving image smoothing, hybrid inference, image segmentation, inductive data learning, object-based image analysis (OBIA), outcome quality indicator, prior knowledge, process quality indicator, RGB cube, superpixel, texture element, texture segmentation, unmanned aerial vehicle, vector quantization.





## 9.1 Introduction

Outpaced by the rate of collection of images and videos of ever-increasing quality and quantity, such as those acquired by novel generations of spaceborne and airborne Earth observation (EO) imaging sensors in addition to images acquired by consumer-level color cameras mounted on mobile devices, such as tablet computers, smartphones and unmanned aerial vehicles (UAVs) [104], the computer vision (CV) and remote sensing (RS) communities appear unable to transform big image datasets into operational, comprehensive and timely information products, in compliance with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines [20]. The overarching goal of the present methodological paper is to contribute to invert this negative trend by fostering information-as-data-interpretation capabilities of a low-level CV system in operating mode, where sub-symbolic quantitative (numeric and unequivocal) variables in the (2D) image-domain are automatically transformed into nominal (categorical and inherently equivocal) variables of symbolic quality in the modeled world [21], eligible for use in symbolic human reasoning typically mimicked by fuzzy logic [83].

Formally a finite image is a function that assigns colors (e.g., coded by real numbers) to a finite, rectangular array of locations in space (e.g., coded by ordered pairs of integers) [101]. There is a vast CV literature dealing with low-level (pre-attentive) image feature extraction, image analysis and synthesis (coding and decoding), image pair similarity assessment, image-object segregation (segmentation), high-level (attentive) image understanding (classification), etc. [1]-[17]. Unfortunately, links of CV to biological vision remain extremely weak [5], [7]-[12]. On the one hand, human vision can be considered a huge puzzle with a lot of missing pieces to date [9], [10]. On the other hand, computational models of simple, complex and end-stopped cells located in the primal visual cortex of mammals have been proposed in the last 10 years [4], [5], [9]-[12], [18]. In general, “no claim is made about the pertinence or adequacy of CV models to human visual perception... This enigmatic situation arises because research and development in CV is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that *biological vision is currently the only measure of the incompleteness of the current stage of CV and illustrates that the problem is still open to solution*” [19]. For example, if we require that a CV system should be able to predict perceptual effects, such as the well-known Mach bands illusion where bright and dark bands are seen at ramp edges, then the number of published vision models becomes surprisingly small [12].

Our working hypothesis was that a diffuse inconsistency of existing CV systems with human visual perception may partly explain why the CV and RS communities are being outpaced by the rate of collection of images of ever-increasing quality and quantity. This conjecture is supported by several true-facts. Still now the percentage of EO images downloaded by users from the European Space Agency (ESA)’s EO image databases is estimated at about 10% or less. In addition, no ESA EO Level-2 prototype product has ever been generated systematically at the ground segment. The ESA definition of EO Level 2 product encompasses a multi-spectral (MS) image corrected for atmospheric, adjacency and topographic effects into surface reflectance values, stacked with its data-derived scene classification map (SCM) [105], [106].

Since the first key principle of *accessibility* to sensory data required by the QA4EO guidelines has greatly improved in recent years, a low exploitation rate of EO big data to be systematically transformed into Level 2 products may be due to an ongoing inadequacy of the second QA4EO key principle, specifically, *suitability* of data, information processing methods and information products, subject to a quantitative quality assurance ( $Q^2A$ ) policy to be community-agreed upon in advance [20]. To better understand their potential impact upon the RS community, the two QA4EO key principles of accessibility to and suitability of sensory data and data-derived information products can be related to the two information theories investigated by philosophical hermeneutics, specifically, quantitative/ unequivocal/ objective *information-as-thing* and qualitative/ equivocal/ subjective *information-as-data-interpretation* [21]. These two concepts of information correspond intuitively to the two well-known classes of variables, either quantitative/numeric or qualitative/nominal/categorical. Unequivocal *information-as-thing*, related to the popular Shannon’s mathematical theory of data communication/transmission, irrespective of the meaning of the transmitted message [22], is “easier to cope with” than equivocal *information-as-data-interpretation*, where the message receiver has a pro-active role in the message interpretation process [21]. This double meaning of word “information” is a major cause of equivocality at the root of information technology (IT) [23], [24]. It explains why to date the QA4EO requirement of making quantitative EO data accessible has been easier to deal with than assuring suitability of qualitative information products generated by EO data interpretation algorithms.

To increase the suitability of EO image value-adding processes and products available for operational, timely and comprehensive exploitation of EO image “big data” by a wide public of image service industries, providers and end users [20], our work pursues a holistic approach to the study of vision as part of cognitive science [21], [25]-[30], see Fig. 9-1.



Cognition is the transformation of ever-varying sensations in the real world into stable percepts/concepts in the modeled world [1], [2], [30]. To fill the semantic information gap from low-level sub-symbolic sensory data to high-level symbolic classes of objects in a *world model* [1], [2], cognitive processes, including vision, are inherently difficult, which means ill-posed in the Hadamard sense [31], [77], i.e., NP-hard [102], [103]. To become better posed for solution, they require *a priori* knowledge available in addition to sensory data [2], [32], [33]. In practice, cognitive systems are *hybrid* inference systems where bottom-up inductive learning-from-data capabilities, typically investigated by machine learning [32], [33], must be combined with top-down knowledge-by-rule inference, which is the traditional focus of attention of expert systems in artificial intelligence [6], [34]. A multidisciplinary cognitive approach to vision agrees with an increasing trend in scientific literature, where hybrid inference systems combine bottom-up statistical models with top-down physical models to take advantage of the unique features of each and overcome their shortcomings [23], [24], [35]. For the sake of completeness, pros and cons of deductive and inductive inference systems are summarized below.

- Relying exclusively on *a priori* knowledge available in addition to data, expert systems are static (non-adaptive to data) non-iterative (one-pass) decision trees that require neither user-defined parameters nor training data to run. Hence, they are called fully automatic. Their computation time is linear with the number of decision rules. In general, it takes a long time for human experts to learn physical laws of the real world-through-time and tune physical models. In addition, expert systems suffer from an intrinsic lack of flexibility, i.e., decision rules do not adapt to changes in the input dataset and in user needs, hence their knowledge base may soon become obsolete [34]. Finally, they suffer from an intrinsic lack of scalability, in particular static rule-based systems are impractical for complex problems. Nevertheless, once a syntactic inference system is set up and proved to be robust to changes in input data, then the effort pays off [23].
- Inherently ill-posed in the Hadamard sense [31], inductive learning-from-data algorithms require *a priori* knowledge in addition to data to become better posed for numerical solution [32]. It means that, in general, inductive data learning algorithms are semi-automatic and training data-specific [35]. They are unable to learn class grammars and semantic networks of concepts as nodes and inter-concept relations, such as part-of, subset-of, etc., as arcs between nodes [6]. For training and testing phases, they require an adequate reference dataset whose quality, availability and costs can be impossible or unaffordable to cope with, especially when ground truth must be collected across time and geographical space, such as in EO data mapping problems at global scale.

Quite strikingly, instead of focusing on hybrid CV solutions, the mainstream CV and RS communities concentrate their research and technological development (RTD) efforts on driven-without-knowledge inductive learning-from-data algorithms, such as support vector machines [36], random decision forest classifiers [37] and the increasingly popular deep convolutional neural networks [38], [39], whose free-parameters to be user-defined based on heuristic criteria include the network architecture [32], [33].

In the multidisciplinary scenario of cognitive science, sketched in Fig. 9-1, the goal of the present low-level vision software design and implementation project was to develop an off-the-shelf (ready-for-use) hybrid reversible (direct and inverse) quantitative-to-nominal-to-quantitative (Q<sup>2</sup>NQ) transform of an uncalibrated monitor-typical red-green-blue (RGB) image for automated *superpixel* detection and Q<sup>2</sup>A in linear time. To become better conditioned for numerical solution, the low-level CV software pipeline was constrained by the following process and outcome requirements.

(I) To comply with the QA4EO Calibration/Validation (*Cal/Val*) requirements [20]. The QA4EO *Val* principle requires an EO data processing pipeline to provide each processing stage with a set of community-agreed quantitative quality indicators (Q<sup>2</sup>Is), so that error propagation across the pipeline can be monitored in comparison with quality reference standards [20]. EO data *Cal* is the transformation of dimensionless digital numbers into a physical unit of radiometric measure, e.g., at-sensor reflectance, based on radiometric *Cal* metadata parameters [20]. In line with the QA4EO recommendations the RS community should regard as an indisputable fact that “the prerequisite for physical model-based, quantitative analysis of airborne and satellite sensor measurements in the optical domain is their calibration to spectral radiance” ([105]; p. 29). In general, availability of physical variables is a necessary not sufficient condition for sensory data analysis with physical models [6], [32], [35]. In addition to physical models, physical variables can be investigated by statistical models [6], [32], [35]. Hence, physical variables can be analyzed by hybrid inference systems. On the other hand, quantitative variables provided with no physical unit of measure can be investigated by statistical models exclusively [35]. Irrespective of this common knowledge, EO data *Cal* is largely ignored in the RS common practice [23], [24].

(II) To run without user-machine interaction, i.e., it requires neither user-defined parameters nor training data to run. Inference system automation can be accomplished by a hybrid inference system where a deductive inference first stage employs *a priori* knowledge, available in addition to data, to provide second-stage inference learning-from-data algorithms



with initial conditions without user-machine interaction. This hybrid inference approach complies with: (a) biological cognitive systems, where “there is never an absolute beginning” [29] because top-down genotype initializes bottom-up phenotype [30]. (b) The principle of statistical stratification. Well known in statistics, it states that “stratification will always achieve greater precision provided that the strata have been chosen so that members of the same stratum are as similar as possible in respect of the characteristic of interest” [40].

(III) To detect color-homogeneous texture elements, traditionally known as *texels/ textons* [6], [14]-[17], otherwise called *tokens* in agreement with the Marr definition of low-level vision’s raw primal sketch [4]. Color is the sole quantitative visual information available at the pixel resolution, i.e., color values are context-insensitive, whereas all remaining visual properties, such as texture (perceptual spatial grouping of texels), image-object shape and size, inter-object spatial relationships, etc., are spatial-context dependent. In general, both in the spatiotemporal 4D real world-through-time domain, described in user-speak by a world model [1], [2], and in the (2D) image domain described in techno-speak, spatial information dominates color information, irrespective of data dimensionality reduction from 4D to 2D [102], [103]. In fact, human panchromatic vision is nearly as effective as human color vision in the provision of a complete scene-from-image representation in our brain: from local syntax of individual objects to global gist and layout of objects in space, including semantic interpretations and even emotions [5], [9], [10], [13]. Since spatial reasoning is required in high-resolution image understanding where spatial information dominates color information, the geographic object-based image analysis (GEOBIA) paradigm has become increasingly popular in the RS community [50], while the CV community has increasingly adopted semi-automatic inductive data learning algorithms for color-homogeneous *superpixel* detection as a prior step in many high-level CV tasks [41]. To the best of these authors’ knowledge, the concept of superpixel is not novel, but highly related to the Julesz’s *texton/ texel* theory of pre-attentive human vision developed back in the 1970’s [6], [14]-[17]. If this conjecture holds, terms *superpixel* and *texel/ texton* are synonyms and related to the Marr’s raw primal sketch in pre-attentive vision [4].

(IV) To be general-purpose, i.e., data-, user- and application-independent. Selected by a user according to his/her own photointerpretation purposes, any true- or false-color three-band image can be uploaded onto a monitor-typical RGB cube. False-color image composites include 3-channel stacks of heterogeneous and/or dimensionless scalar 2-D gridded variables, such as image-derived spectral indexes (SIs, where each SI is equivalent to a 2-D membership function), such as a bare soil SI, a vegetation SI and a water SI uploaded onto the RGB channel R, G and B respectively [35].

(V) To run in linear time, i.e., the system complexity is  $O(N)$ , where  $N$  is the image size.

(VI) To be implemented in tile streaming mode, to cope with massive digital images by reducing central memory occupation.

These constraints make the proposed low-level vision RGB image QNQ transform a lightweight computer program eligible for use in a mobile software application. By definition a mobile software application, eventually provided with a mobile user interface, is a lightweight computer program specifically designed to run on web browsers and mobile devices, such as tablet computers and smartphones.

If existing image segmentation algorithms are categorized according to the well-known hierarchical vision theory developed in the late 1970’s by Marr [4], there is no rationale in comparing algorithms belonging to different information processing stages of the visual system pipeline. For example, the proposed automatic and data-independent low-level vision software system for *superpixel/ texton/ texel* detection fully complies with the Marr definition of raw primal sketch [4]. As a consequence, it is preliminary not alternative to any inductive semi-automatic and site-specific algorithm [35] for either: (a) texture segmentation (perceptual grouping) [1], [2], [13], equivalent to the Marr’s low-level vision full primal sketch [4], where texture is the visual effect generated by the spatial distribution of *texels* detected at the raw primal sketch, or (b) surface segmentation, equivalent to the Marr’s high-level 2½D sketch [4], e.g., refer to [42]. In addition to being inherently site-specific and semi-automatic, inductive data learning algorithms for texture segmentation or surface segmentation must be spatially context-sensitive, which boosts their computation time [13]. For example, in [42] spatial coherence is taken into account. Noteworthy, inherent functional properties of inductive data learning algorithms do not meet the aforementioned CV software project requirements.

Although the proposed low-level vision RGB image QNQ transform was developed within the RS community, its application domain extends to RGB images acquired by the whole CV discipline, which includes RS applications as a special case, see Fig. 9-1. Such an extended application domain of the proposed RGB image QNQ transform is a value added, i.e., it does not decrease the potential impact of the proposed solution upon EO image understanding applications.

The rest of this methodological paper is organized as follows. Chapter 9.2 provides this paper with a relevant survey value by presenting the problem background. Methods selected and proposed are discussed in Chapter 9.3. Materials



adopted in the experimental session are described in Chapter 9.4. Experimental results are presented in Chapter 9.5 and discussed in Chapter 9.6. Conclusions are reported in Chapter 9.7.

## 9.2 Problem Background

### 9.2.1 Human vision

The primary objective of any biological or artificial visual system is to back-project the information in the 2-dimensional (2D) image domain to that in the 3D viewed-scene domain, equivalent to a subset of the 4D real world-through-time observed at a given time [1], [2], [4]. To provide one or multiple plausible symbolic description(s) of a 3D viewed-scene, an image understanding system (IUS) is expected to find associations between ever-varying 2D quantitative (sub-symbolic) image features and stable symbolic concepts of 4D objects-through-time, e.g., houses, cats, etc. A finite and discrete ensemble of classes of 4D objects form a so-called 4D *world model* [2], or spatiotemporal ontology of the world-through-time [43]. It can be graphically represented as a semantic network, with nodes as classes of objects and arcs between nodes as inter-class relationships [2], [44]. This definition of vision means that image understanding is an ill-posed problem in the Hadamard sense [31]. First, vision is affected by data dimensionality reduction from 4D to 2D causing, for example, image occlusion phenomena, which are seamlessly filled in by human visual perception. Second, there is a well-known semantic information gap from sub-symbolic image features, either 0D points, 1D lines or 2D polygons [45], and plausible symbolic description(s) of a 3D viewed-scene. This is the same information gap thoroughly investigated by both philosophy and psychophysical studies of perception in cognitive science [2], [30]. In vision, we are always seeing objects we have never seen before at the sensation level, while we perceive familiar objects everywhere at the perception level [2]. In practice, the human visual system is able to construct on the basis of a brief glance a complete scene-from-image representation in our brain: from local syntax of individual objects to global gist and layout of objects in space, including semantic interpretations and even emotions [5].

To accomplish its cognitive tasks, the visual system of mammals comprises a pre-attentional vision first phase and an attentional vision second phase, summarized as follows.

- (1) Pre-attentive/low-level vision extracts picture primitives based on general-purpose image processing criteria independent of the scene under analysis. It acts in parallel on the entire image as a rapid ( $< 50$  ms) scanning system to detect variations in simple visual properties [7]-[12]. In each hypercolumn of the primary visual cortex (PVC), there are end-stopped cells, in addition to simple- and complex-cells [5], [10]. While simple- and complex-cells are thought to accomplish line and edge extraction [9], end-stopped cells respond to image singularities, such as line/edge crossings, vertices of image-objects and end-points of line segments [5], [10]. According to Marr [4], pre-attentive vision consists of two phases. (a) A *raw primal sketch* for token extraction, where tokens are 0D image singularities, 1D contours and 2D segments. (b) A *full primal sketch*, also known as perceptual grouping [14]-[17], where texture detection is accomplished based on the spatial distribution of tokens/textels. Noteworthy, in the words of Marr, “vision goes symbolic almost immediately, right at the level of zero-crossing (first-stage primal sketch), without loss of information” ([4]; p. 343).
- (2) Attentive/high-level vision operates as a careful scanning system employing a focus of attention mechanism based on end-stopped cells [10], [18]. Scene subsets, corresponding to a narrow aperture of attention, are observed in sequence and each step is examined quickly (20–80 ms) [7], [8]. The Marr high-level vision subsystem comprises an understanding of local surface properties, called *2½D sketch*, followed by the generation of one or several plausible symbolic descriptions of the 3D viewed-scene [4].

As reported in Chapter 9.1, human panchromatic vision is nearly as effective as human color vision in the provision of a complete scene-from-image representation in our brain: from local syntax of individual objects to global gist and layout of objects in space, including semantic interpretations and even emotions [5], [9], [10], [13]. On an *a posteriori* basis, this observation has two important implications. First, in the 4D real world-through-time, color information of 4D objects, e.g., cars and trees, is dominated by their spatiotemporal attributes, as properly stated by Adams *et al.* [93]. Second, the same consideration holds for a planar representation of the 4D world-through-time, where 2D spatial (contextual) information dominates color information. To cope with the dominant 2D spatial information in an image, the human visual system employs modular arrays of multiscale 2D local filters [13], capable of a topology-preserving mapping of an image onto a neural network [46]-[49].

In compliance with the Marr’s vision theory [4], a hierarchical ontology of a 3D viewed-scene could be obtained by formalizing a top-down language which should address all possible 3D model representations, down to an understanding of local surface properties, down to a description of the image properties [4]. On the other hand, intuitive bottom-up image





understanding starts from scratch with a bunch of pixels, to be partitioned into image-objects to be grouped according to inter-object spatial relations (layout), etc., in agreement with the increasingly popular object-based image analysis (OBIA) paradigm [50], whose final output product consists of symbolic image-objects in a geographic information system (GIS)-ready file format. Both top-down and bottom-up inference approaches are possible, but the human visual system employs them jointly all the time [5]. For example, Vecera and Farah proved that “image segmentation can be influenced by the familiarity of the shape being segmented... Results are consistent with the hypothesis that image segmentation is an interactive (hybrid) process, in which top-down knowledge partly guides lower level processing... If an unambiguous, yet unfamiliar, shape is presented, top-down influences are unable to overcome powerful bottom-up cues... While bottom-up cues are sometimes sufficient for processing, these cues do not act alone; top-down cues, on the basis of familiarity, also appear to influence perceptual organization” ([51]; p. 1294).

### 9.2.2 RGB Cube Partitioning into “Universal” Basic Colors Known in Advance

In a color model, each color is addressed by a unique coordinate in a color coordinate system. Based on various guiding principles, there are several coordinate arrangements that generate, in turn, color models. Terms like color space, color coordinate system and color model are generally used as synonyms [52]. Typically, color coordinate systems are categorized as objective or subjective. The subjective category is also termed “perceptual” because in perceptual color spaces, such as the Munsell system, color distances are closer to visual differences perceived by a human observer. Among objective color systems, the hardware-oriented RGB color model is the most commonly employed in color monitors [52]. It consists of a Cartesian coordinate system where colors are a linear additive mixing of the three RGB primary colors. Its gamut approximates well the gamut of surface colors, as the gamut of surface colors is not far from being parallelepiped in form, which is inevitable given the non-convexity of the spectral locus [53]. This may explain why seven of the RGB cube’s vertices coincide well with the foci of seven of the eleven Berlin and Kay’s human basic color (BC) categories [25], specifically, black, white, red, green, blue, yellow and purple [53], see Fig. 9-2. Central to this consideration is Berlin and Kay’s landmark study of color words in 20 human languages. On the basis of that study they claimed that the ‘basic color terms of any given language are always drawn’ from a universal inventory of eleven color names: black, white, gray, red, orange, yellow, green, blue, purple, pink and brown [25]. These perceptual BC categories are expected to be “universal”, i.e., users are not required to learn a new color representation for ever-varying sensory datasets, but they can apply the same universal color representation independently of the image at hand [54]. To summarize, in addition to be hardware-oriented, the RGB cube is a “natural” coordinate system, quite consistent with human BCs. If the RGB cube axes are discretized into known quantization levels, i.e., when the forward discrete color map is known, than an efficient inverse color map algorithm can replace each input RGB pixel with the center value of its assigned discrete color cell [55]. In [52], the reversibility of six geometric color spaces was investigated, where reversibility was defined as the capacity of a color model to recover an original 8-bit depth RGB image representation in  $(256)^3$  RGB combinations, after transforming the  $(256)^3$  RGB combinations into the target color space and then transforming back to the RGB cube. The conclusion was that the reversibility error in the forward and backward color space transformations was negligible for human visual interpretation purposes.

### 9.2.3 Color Constancy

In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions, so that they appear identical to a “canonical” (reference) image, subject to a “canonical” (known) light source (of controlled quality), e.g., under a white light source [56]. In short, solution of the color constancy problem is the recovery “of an illuminant-independent representation of the reflectance values in a scene” [57]. In practice color constancy supports image harmonization and interoperability when no radiometric calibration parameter is available. Its goal is analogous to that of inter-image relative calibration [35] and image-specific absolute radiometric *Cal*, considered mandatory by the QA4EO guidelines when radiometric calibration parameters are available [20].

Color constancy is intrinsically related to brightness in a more global (large scale) sense than, say, image-contour detection, which depends on local non-stationary image properties. Although the biological mechanisms involved with the color constancy ability are not yet fully understood [3], [56], it is speculated that a special type of retinal ganglion cells can play a role in the estimation of a global “background” brightness, on which lines, ramps and step edges [58] can be projected [5]. In principle, this special type of retinal ganglion cells can be involved with brightness perception because: (i) it features a very large receptive field. (ii) It is not connected to either rods or cones. (iii) It is connected to central brain areas for controlling the circadian clock (day-night rhythm). (iv) Via a feedback loop, it is connected to the eye’s iris (pupil size).



(vi) These special retinal cells also connect to at least the ventral area of the lateral geniculate nucleus [5].

Computational color constancy is a fundamental prerequisite of many CV applications, such as the RGB image QNQ transform proposed in this paper. Also because biophysical mechanisms of color constancy remain largely unknown, computational color constancy algorithms are typically unable to simulate color constancy effects observed in humans [19]. Computational color constancy is an under-constrained problem in the Hadamard sense [31]. Since it does not have a unique solution, it requires *a priori* knowledge in addition to data for numerical treatment [32]. For these reasons there has been a large number of alternative color constancy algorithms proposed in the CV literature in the last 30 years [59]. Early computational models were derived from works on human perceptual theory, resulted in the Retinex perceptual theory by Land [60], considered inadequate by now. In survey works such as [3], [56] and [59], computational color constancy approaches are divided into three categories. (1) Low-level statistical model-based methods. (2) Physical model-based methods. (3) Gamut-based methods. In statistical models, the best-known statistical assumption about color distributions is the so-called *Grey-World* assumption: the average reflectance in a scene under a neutral light source is achromatic, which means that the color of the light source can be estimated by computing the average color in the image. Another well-known assumption is the *White-Patch* assumption: the maximum response in the RGB channels is assumed to be caused by a perfect reflector. The assumption of perfect reflectance is alleviated by considering the color channels separately, resulting in the *max-RGB* assumption, where the illuminant is estimated as the maximum response in each color channel separately. In common practice, if an image contains few edges (corresponding to a “low” signal-to-noise ratio), then pixel-based (context-insensitive, 1st-order spatial distribution) methods, like *Grey-World* and *White-Patch*, are preferred. When the signal-to-noise ratio is “medium” or “high”, context-sensitive (edge-based) methods are preferred: for example, 1st- and 2nd-order local spatial derivatives are adopted in the so-called *Grey-Edge* method [61]. In [59], the eleven “universal” human BC categories, identified by Berlin and Kay [25], see Fig. 9-2, were adopted as a form of *a priori* knowledge, available in addition to and independent of data (refer to Chapter 9.2.2), to improve the computational efficiency of the color-by-correlation algorithm proposed in [57]. In the statistical color constancy algorithm implemented in the Environment for Visualizing Images (ENVI) commercial software toolbox [62], “ENVI does a special (‘ultimate’) stretch for the display case, which really can't be reproduced using the ENVI's default 2% stretch. If the histogram features more than three bins, the special stretch will calculate the left hand percent stretch on `hist[1:*]` and the right hand percent stretch on `hist[0:n_elements(hist)-2]`. If the histogram is a Gaussian (normal) shaped curve, then the difference between this and the ‘full’ histogram is negligible. However, if there is a large saturation of min or max values (such as an image with a lot of background, e.g., water pixels, and/or foreground, e.g., cloud pixels), then ENVI's default stretch will ignore the spike(s) and calculate the percent linear stretch, e.g., a 2% stretch, based on the rest of the “real” histogram. This allows ENVI to display, by default, many images which otherwise would not stretch well with a 2% linear stretch since they contain more than 2% background and/or foreground” [63].

For a complete survey of computational color constancy methods, the interested reader can refer to the existing literature, e.g., [3], [56], [59].

#### 9.2.4 Deductive and Inductive Vector Quantization

In his seminal work conceived to bridge independent studies on color naming conducted by linguistics and CV [53], Griffin verified the hypothesis that the best system of color categories for pragmatic purposes coincides with human BCs, see Fig. 9-2. In line with [53], additional relationships among independent studies conducted by linguistics, inductive machine learning and deductive artificial intelligence can be identified to shed new light on the problem of prior knowledge-based color space partitioning/ discretization/ quantization. In the multidisciplinary framework of cognitive science (see Fig. 9-1), first, prior knowledge-based color space discretization is equivalent to color naming in a natural language [3], [64], [65]. According to linguistics, a discrete and finite set of BC names (refer to Chapter 9.2.2) must be community-agreed upon in advance to become “universal”, which means consistently used by members of the community on a regular basis [53].

Second, prior knowledge-based color space quantization is the deductive counterpart of inductive unlabeled data learning algorithms for vector quantization (VQ) [32], [33] and data compression [52]. Two of the most popular and widely used VQ heuristics in unlabeled data analysis are the *k*-means VQ algorithm, also known as Lloyd's or Linde-Buzo-Gray's VQ algorithm [47], [66]-[70], and the Iterative Self-Organizing Data Analysis Technique (ISODATA) [71]. Typically, inductive VQ algorithms have to minimize a known VQ error function given a number of *k* discretization levels defined beforehand, such as the *k*-means VQ [47], whose practical time complexity is equal to  $O(N \cdot k \cdot I)$  [72], [73], where variable *I* is the number of iterations required to reach convergence. Vice versa, a user can fix the target VQ error value, so that it



is the parameter  $k$  to be dynamically learned from unlabeled data by the inductive VQ algorithm [66], [67], such as ISODATA [71].

It is worth mentioning here that unsupervised data learning VQ algorithms should never be confused with unsupervised data clustering algorithms [32]. The goal of unsupervised data clustering is to locate hidden “perceptual” (fuzzy) data structures (clusters, hypervolumes) of any possible shape in the unlabeled data set at hand [46]-[49]. Unsupervised data clustering is a non-predictive unlabeled data mapping task [32], i.e., it is expected to perform the best with the available input dataset, irrespective of possible unknown future samples. Unlike VQ error minimization problems, there is no known cost function to minimize in unsupervised data clustering [47]-[49]. Finally, to provide a topology-preserving map of each unlabeled data cluster, unsupervised data clustering algorithms model both centroids (centers of mass) of processing elements, similar to those estimated by inductive VQ algorithms, but also lateral connections (synapses) between pairs of processing elements [46], to form one connected network of processing elements per data cluster [47]-[49].

Back to deductive and inductive VQ algorithms for color space partitioning, when they are compared, their functional differences make them complementary rather than alternative in nature, as discussed below.

(1) *Degree of automation*, directly related to the ease of use and inversely related to the number of system’s free-parameters to be user-defined. On the one hand, prior knowledge-based color space partitioning is automatic and data-independent, refer to Chapter 9.1. On the other hand, inductive VQ algorithms are inherently ill-posed [32], semi-automatic and site-specific [35], refer to Chapter 9.1. For example, the  $k$ -means VQ algorithm requires the following free-parameters to be user-defined based on heuristics: (i) the number of vector quantization levels  $k$ , (ii) a convergence threshold, typically defined as the maximum number  $I$  of training iterations, and (iii) a dictionary (codebook) of  $k$  vector data centroids (centers of mass) to be initialized, e.g., by means of random data sampling.

(2) *Measurement space partitioning*. In inductive VQ algorithms a popular VQ error function to minimize is the root mean square error [32], [33],  $RMSE_b$ ,  $b = 1, \dots, B$ , defined as:

$$RMSE_b = \sqrt{\frac{\sum_{i=1}^N [P_b(i) - P_b^*(i)]^2}{N}}, \quad b = 1, \dots, B, \quad (1)$$

where  $N$  is the total number of pixels,  $P_b(i)$  is the scalar value of the  $i$ th-pixel in band  $b = 1, \dots, B$ ,  $P_b^*(i)$  is the post-quantization  $i$ th-pixel value, while the adopted metric distance is the Euclidean distance. When they adopt a Euclidean metric distance minimization criterion, such as Eq. (1), and they reach convergence, then inductive VQ algorithms, such as  $k$ -means and ISODATA, accomplish a Voronoi tessellation of the input data space, which is a special case of convex polyhedralization [46], [47]. On the contrary, the designer of a prior knowledge-based decision tree for color space discretization is free to adopt discretization levels of any possible shape and size, either convex or concave, either connected or not, see Fig. 9-2. Unfortunately, when the MS space dimensionality is superior to three, a prior dictionary of mutually exclusive and totally exhaustive hyper-polyhedra is difficult to think of and impossible to visualize.

(3) *Computational complexity*. Because color is the sole visual information to be context-independent [1], [2], [4] (refer to Chapter 9.1), then a color space discretization algorithm, either inductive or deductive, is expected to work with pixel values as vector data. A pixel-based VQ considers a (2D) image as a 1D vector stream, featuring no spatial information. In this case, an iterative suboptimal  $k$ -means VQ algorithm has a practical time complexity equal to  $O(N \cdot B \cdot k \cdot I)$  [72], [73], where variable  $I$  is the number of iterations required for convergence and  $B$  is the number of color channels. On the contrary, an expert system for color space discretization is a one-pass static decision tree, with one decision rule per target BC category. Hence, its complexity is equal to  $O(N \cdot B \cdot BC)$ , where  $BC$  is the number of basic color categories known *a priori*.

Based on the aforementioned comparison, in addition to considerations reported in Chapter 9.1 about the complementary nature of inductive and deductive inference, it is important to conclude that *deductive and inductive VQ algorithms should never be considered alternative, but complementary in nature*. For example, a deductive color space quantization first stage can be employed to initialize automatically, based on prior color knowledge, the number  $k$  of vector data centroids and their initial values in an inductive driven-without-knowledge  $k$ -means VQ algorithm. The resulting hybrid inference system would be an automatic driven-by-knowledge  $k$ -means VQ algorithm. A realization of this hybrid inference concept can be found in [74], where an expert system for RGB cube partitioning was employed to initialize the semi-automatic and site-specific inductive data learning algorithm for image segmentation [75] adopted by the popular eCognition commercial software product [76] for OBIA applications [50].

### 9.2.5 Two-Pass Connected-Component Multi-Level Image Labeling

Let’s define: (i) an image as a 2D gridded quantitative/numeric variable [17] (refer to Chapter 9.1), and (ii) a multi-level



image as a 2D gridded qualitative/categorical/nominal variable. Image segmentation [42], [75] is the dual problem of image contour detection [77]. These are both inherently ill-posed problems in the Hadamard sense [23], [24], [31], [75], [77]. They admit no unique solution and require *a priori* knowledge in addition to data to become better posed for numerical treatment [32]. On the contrary, a multi-level image, such as a classification map, can be deterministically partitioned into connected image-objects, consisting of 0D, 1D or 2D planar objects [45]. This is a well-posed planar segmentation problem, whose deterministic solution is unique. Whereas exactly one segmentation map can be derived from one multi-level image, the vice versa does not hold, i.e., the same segmentation map can be extracted from different multi-level images [78], see Fig. 9-3. Unfortunately, well-posed multi-level image segmentation algorithms are often confused with inductive ill-posed image segmentation algorithms. For example, a two-pass connected-component multi-level image labeling algorithm is automatic and its two-pass computational complexity  $O(2 \cdot N)$  is linear in the image size in pixels,  $N$  [6], [79]. In addition, to cope with massive images by reducing central memory occupation, it can be implemented in tile streaming mode [6], [79].

### 9.2.6 Superpixel Detection Equivalent to Texel Detection in the Pre-Attentional Raw Primal Sketch

To the best of these authors' knowledge the concept of superpixel, developed by the CV community in recent years, is equivalent to the Julesz's texton/texel theory of pre-attentive human vision [6], [14]-[17] (refer to Chapter 9.1). In CV, superpixel detection is an image pre-processing first stage employed for image simplification as input to high-level vision tasks. It is required to be fast to compute, memory efficient, simple to use by featuring few and intuitive input parameters to be user-defined, and capable of increasing the speed and quality of the higher-level vision tasks [41]. One popular inductive algorithm for superpixel detection is the simple linear iterative clustering (SLIC) [41], which is an adaptation of the popular  $k$ -means VQ algorithm [47], [66]-[70], refer to Chapter 9.2.4. The SLIC algorithm's free-parameters to be user-defined based on heuristics are  $k$ , the desired number of approximately equally-sized superpixels, and  $m$ , a compactness term in range [1, 40], set by default equal to 10. If parameter  $m$  increases, then detected superpixels tend to feature more regular size and shape [41]. The complexity of SLIC is linear in the total number of pixels  $N$ ,  $O(N)$ , irrespective of  $k$ . According to the SLIC authors, "the (user's) ability to specify the amount of superpixels, and to control the compactness of the superpixels" are important [41]. This would agree with a popular criterion of *good* segmentation requiring that "a good segmentation region should be formed by connected pixels with homogeneous colors whose shape should be as compact as possible" [42], [80]. However, in common practice, the SLIC dependence on two unknown parameters to be user-defined based on empirical criteria decreases the algorithm's degree of automation (ease of use) and has a negative impact on its robustness to changes in input parameters and to changes in input data. For example, how can a user predict a reasonable number  $k$  of equally-sized superpixels to be detected in massive images of complex real-world scenes? According to the present authors, by scoring low in degree of automation the inductive SLIC algorithm is little useful in common practice.

## 9.3 Methods

An original low-level vision software pipeline was implemented to accomplish a QNQ transform of a monitor-typical RGB image for superpixel detection and  $Q^2A$ , subject to the CV software project requirements listed in Chapter 9.1. Specifically, the low-level CV software realization was required to be automatic, linear-time, data-, user- and application-independent, and in tile streaming mode. Our system realization was a proof-of-concept. It proved that the target project admits solution(s), based on existing algorithms, but it does not claim to be the "best" solution, if any exists. Actually, *information-as-data-interpretation* problems [21], such as cognitive problems including vision, are inherently equivocal/subjective/ ill-posed (refer to Chapter 9.1), i.e., there is no absolute "best" solution to *information-as-data-interpretation* problems [21].

### 9.3.1 Software Design, Algorithm Selection and Implementation

The implemented RGB image analysis and synthesis software pipeline, consisting of six subsystems identified as block 1 to 6 in Fig. 9-4, is described below.

*RGB image analysis for superpixel/ texel detection, refer to blocks 1 to 5 in Fig. 9-4.*

1. *RGB image pre-processing for color constancy.* According to Chapter 9.2.3, color constancy is considered mandatory to guarantee uncalibrated image harmonization and interoperability when no calibration metadata parameters are available, such as in UAVs employing low-weight consumer-level color cameras [104]. By analogy with the QA4EO *Cal* requirements [20], uncalibrated image color constancy is considered a necessary not sufficient condition for sensory





data analysis with hybrid inference systems, where physical and statistical models are combined to take advantage of each and overcome their shortcomings. There is a wide variety of published algorithms for image color constancy [56], [57], [59], [61]. To comply with the software project requirements proposed in Chapter 9.1, we designed and implemented an original self-organizing statistical algorithm (never published, patent pending) for 1<sup>st</sup>-order histogram-based (non-contextual) image color constancy in linear time  $\leq O(I \cdot N \cdot B)$ , where the number of learning-from-data iterations  $I$  is  $\leq 3$ . It was inspired by the ENVI “ultimate” image stretching algorithm summarized in Chapter 9.2.3. Our solution analyzes each single channel to detect one-of-four 1<sup>st</sup>-order histogram distributions, described as follows (see Fig. 9-5). (i) Neither a background nor a foreground mode is present in addition to a central mode. (ii) A background mode with a long right tail can be identified. (iii) A foreground mode with a long left tail can be identified. (iv) One background and one foreground mode can be identified in addition to a central mode. Once background and foreground modes are detected, if any, they are mapped onto the minimum output gray value,  $\text{hist}[0]$ , and the maximum output gray value,  $\text{hist}[255]$ , respectively. Next, a standard linear stretching algorithm is applied per channel, to fill the histogram bins  $\text{hist}[1:254]$ , according to a traditional max-RGB criterion (refer to Chapter 9.2.3).

2. *RGB cube partitioning into static (non-adaptive to data) polyhedra corresponding to BC names known a priori.* To bridge independent studies on color naming conducted by linguistics and CV, Griffin verified the hypothesis that the best system of color categories for pragmatic purposes coincides with human BCs [53], refer to Chapter 9.2.2. By using a classification task to test this hypothesis, he obtained results consistent with it. In [53], the test dataset consisted of color RGB jpeg-format images collected by means of a web-based search-by-noun engine. A BC category system was generated with a multi-step process. First, the 267 Munsell coordinate-specified chips [81] were assigned with (gathered into) the eleven BC names. Next, color chips were mapped into the Commission Internationale de l'Eclairage (CIE)-Lab color space [52]. Finally, the eleven BC extents in CIE-Lab space were transformed into data expressed over a uniform 323 sampling of the monitor-typical RGB cube [52]. The final result was a hardware-oriented RGB cube partitioned into eleven mutually exclusive and totally exhaustive human-derived BC categories, as shown in Fig. 9-2.

In [54], if compared against inductive learning-from-data descriptors, static “universal” color descriptors are expected to cause a drop of quantization accuracy, counterbalanced by an increase in computational efficiency and degree of automation. Acknowledged that no single universal (“best”) color dictionary exists, because any color naming is a conventional *information-as-data-interpretation* process to be community-agreed upon in advance [21], three alternative RGB cube partitions, featuring 11, 25 and 50 color clusters respectively, were investigated in [54]. Each color cluster gathered color cells that must be connected in the  $L^*a^*b^*$  cube, starting from a total number of equally spaced color cells equal to  $m = 4000 = 10 \times 20 \times 20$ . In [54], no parent-child inter-cluster relationships exist at the different static color quantization levels.

We designed and implemented an original software solution for prior knowledge-based RGB cube partitioning in compliance with the project requirements listed in Chapter 9.1. Called RGB Image Automatic Mapper™ (RGBIAM™, never published, patent pending), it found inspiration in the existing Satellite Image Automatic Mapper™ (SIAM™), an expert system for automatic transformation of a radiometrically calibrated EO multi-spectral image onto a set of color maps whose legends are color dictionaries featuring parent-child relationships [23], [24], [82]. In particular, RGBIAM is a one-pass prior knowledge-based decision tree for RGB cube partitioning into static polyhedra, non-adaptive to data and not necessarily convex and/or connected. Any prior knowledge-based decision tree encompasses a structural and a procedural knowledge. The former relates to the adopted set of decision rules, the latter to their order of presentation. By changing either its structural or procedural knowledge, the decision tree realization changes [23], [24]. Like in SIAM [82], the RGBIAM’s decision rules define their individual domain of activation in the measurement space as one or more polyhedra, each one described by shape and intensity. First, it transforms each quantitative R, G and B variable into a qualitative variable consisting of fuzzy sets (FSs), e.g., low (L), medium-low (ML), medium-high (MH) and high (H), not necessarily uniform, according to the principles of fuzzy logic [83], [84]. Second, it identifies quantitative inter-channel relationships, called spectral rules (SRs), e.g.,  $SR1 = \max\{B, G\} < (0.5 * R)$ . Third, it defines color names, called spectral categories (SC), as polyhedra combining SRs for shape and FSs for intensity, e.g., Color 1 = Bright Dominant Red =  $SR1 \text{ AND } H\_R$ . Two color discretization levels were implemented: (a) a fine color discretization level, consisting of  $49+1 = 50$  color names, including class “unknown”, and (b) a coarse color discretization level, consisting of the 11 human BCs (refer to Chapter 9.2.2) plus 1 class “unknown” = 12 color names, generated as a fixed parent-child combination of the 50 color names available at the fine discretization level, see Fig. 9-6. The RGBIAM’s static decision tree computational complexity is  $\leq O(C1 \cdot N \cdot B + C2 \cdot N)$ , where  $C1 = 50 =$  cardinality of the fine-granularity color dictionary and  $C2 = 12 =$  cardinality of the



coarse-granularity color dictionary, generated as an aggregation of the  $C1$  color names, i.e., inequality  $C2 < C1$  must hold.

The (SIAM and) RGBIAM's output color maps are called image QuickMap™ products, as opposed to the traditional image QuickLook (typically, a true-color image in the jpg-file format) promoted by ESA in its user-driven EO image retrieval systems. Let's assume that the input RGB image is byte-coded, hence each pixel value requires 24 bits of memory space. An RGBIAM's map whose codebook consists of 50/12 color names requires 6/4 bits per pixel respectively. It means that RGBIAM works as an RGB data compression system at a given compression rate of 4:1 up to 6:1. In common practice, by featuring superior levels of data compression and semantics an image QuickMap can replace any QuickLook image employed in user-driven EO image retrieval systems.

3. *Well-posed extraction of connected components from a multi-level color map.* To transform a multi-level color map into a segmentation map, a well-posed two-pass connected component multi-level image labeling algorithm was implemented according to the existing literature [6], [79], refer to Chapter 9.2.5. In our software pipeline, a two-scale segmentation map, where inter-scale image-objects feature parent-child relationships, was generated with computational complexity  $O(2 \cdot N \cdot 2)$ , where a factor of 2 is due to the generation of two single-scale segmentation maps, one for each of the two RGBIAM's color maps featuring 50 and 12 color levels respectively. An original non-trivial tile streaming implementation of this algorithm was pursued to comply with the project requirements, refer to Chapter 9.1.

4. *Well-posed contour extraction from known image-objects.* The dual problem of image segmentation is contour detection [23], [24], i.e., contours are image-object perimeters. When segments are detected beforehand, their deterministic (non-equivocal) contours can be coded in either raster or vector format. For example, non-equivocal contours of detected superpixels are shown in [41]. In compliance with the software project requirements listed in Chapter 9.1, we extracted contours of detected superpixels in linear time and raster format by means of a well-posed one-pass 4- and 8-adjacency cross-aura estimate, implemented in tile streaming mode [85], see Fig. 9-7. The algorithm complexity is  $O((4 \cdot N + 8 \cdot N) \cdot 2)$ , where the factor 2 is due to the presence of two RGBIAM's color maps to extract contours from. Noteworthy, the 4-adjacency cross-aura measure is useful for high-level attentional OBIA applications [50], to be pursued in series with the RGB image QNQ transform, refer to processing block 7 in Fig. 9-4. For example, superpixel/texel contours can be used to partition an image into high- or low-texture areas, according to the empirical rule proposed in [1]. If a moving window of size  $W \times W$  centered on pixel  $p$  contains more than  $2W$  boundary pixels, then the central pixel  $p$  can be marked as belonging to the high-texture image layer. Moreover, the 4-adjacency cross-aura allows estimation of a scale-invariant shape index of compactness, also called roundness ( $Rndnss$ ) [7]. In particular,

$$Rndnss = (4 \times \text{sqrt}(A) / PL) \in [0, 1], \quad (2)$$

where  $A$  is the segment area and  $PL$  is the cumulative 4-adjacency cross-aura measure of the region's total boundary, where the total boundary takes into account contributions from inner holes, if any, i.e., total boundary = external (outer) boundary + inner boundary (due to holes). It can be easily proved (by induction) that this  $Rndnss$  formulation is scale invariant. For example, a 0D planar object consisting of a single pixel features  $PL = 4$ , then  $Rndnss = 4/4 = 1$  (maximum). For a 4-pixel square object,  $PL = 8$ , then  $Rndnss = 4 \cdot 2 / 8 = 1$ , etc. On the contrary, most of the existing formulations of  $Rndnss$  are not scale invariant in raster imagery [1], [2], [6], [86], [87].

5. *Segment description table allocation and initialization.* Positional, colorimetric, geometric and spatial attributes of raster image-objects can be stored in tabular format in a so-called segment description table (SDT) [1], to be dynamically allocated and initialized in central memory. For example, in their seminal works at the root of the OBIA paradigm [50], where they try to mimic the convergence-of-evidence approach adopted by human reasoning, Nagao and Matsuyama employed the tabular information of an SDT to classify image-objects based on converging colorimetric, geometric, textural and spatial evidence [1], [2]. Starting from the input RGB image and the two segmentation maps generated from the two RGBIAM's color maps of the input RGB image, the computational complexity of an SDT initialization phase is  $O((N \cdot B) \cdot 2)$ , with  $B = 3$ , where the factor of 2 is due to the presence of two SDTs, one per segmentation map. To assess the central memory occupation of an SDT, as an example, let us consider a big RGB image, say, rows =  $RW$  = columns =  $CL$  = 50000, bands  $B = 3$ , byte-coded, whose number of segments, detected by the RGBIAM expert system, is assumed to be equal to  $(RW \times CL / 5 \text{ pixels per segment as average}) = 5 \times 10^8$ . The expected SDT memory occupation per segment would be the following. Segment identifier (ID) = unsigned long int = 4 bytes, locational property (minimum enclosing rectangle, defined by the upper right and lower left corners) = unsigned long int  $\times$  4 = 16 bytes, RGBIAM's color label = unsigned char = 1 byte, area size = unsigned long int = 4 bytes, colorimetric mean = float  $\times$  3 bands = 12 bytes. Hence, the SDT memory occupation per segment is around 37 bytes. In this example, the SDT central memory occupation, equal to

the number of segments  $\times$  memory occupation per segment, would be  $5 \times 10^8 \times 37$  bytes = 18.5 GB. It means that, in general, the tabular representation of the raster image-object attributes in an SDT can be considered very demanding in terms of central memory occupation.

*RGB image synthesis (reconstruction) from the segmentation map and the SDT, refer to block 6 in Fig. 9-4.*

6. *Piecewise-constant input image approximation.* A one-pass piecewise-constant input image approximation, called “object-mean view” in the eCognition commercial software product [76], is equivalent to an edge-preserving smoothed image [1], such as that required in frames of a video sequence [88]. It was accomplished in near real-time by replacing each pixel scanned in the segmentation map with the SDT’s colorimetric mean value of the superpixel that pixel belongs to. In the reconstruction of any RGB image typically characterized by non-stationary local statistics, e.g., mean, standard deviation, spatial autocorrelation, etc., an “object mean view” approach is expected to be more accurate, because more sensitive to varying local statistics, than an inverse color mapping where each input RGB pixel is replaced by the center value of its assigned discrete color cell, such as that proposed in [55]. The complexity of the implemented superpixelwise-constant image approximation algorithm is  $O((N \cdot B) \cdot 2)$ , with  $B = 3$ , where the factor 2 is due to the presence of two STDs, corresponding to the two detected color maps whose color codebook is  $C1 = 50$  and  $C2 = 12$  respectively. Since edge-preserving image smoothing follows image segmentation into superpixels, it is completely alternative to traditional edge-preserving image smoothing via 2D spatial filtering, which is required before image segmentation applied to frames of a video sequence [88]. For robotic applications it is very important that the labels of image segments do not change throughout a video stream. Currently only very few segmentation algorithms running in real-time achieve this objective, among them the Metropolis algorithm [89]. The conclusion is that the proposed automatic edge-preserving image smoothing filter is expected to be particularly useful in reducing oversegmentation phenomena that typically affect the Metropolis algorithm in textured areas of a video sequence [88].

*Summary of blocks 1 to 6 in Fig. 9-4.*

The 6-stage low-level vision software pipeline shown in Fig. 9-4 is implemented in tile streaming mode. Its overall computational complexity is  $\leq O(N \cdot ((C1 \cdot B + 7 \cdot B) + C2 + 28))$ , which is linear in the image size  $N$ , number of bands  $B$  and number of color quantization levels  $C1$  and  $C2$ , with  $C2 < C1$ . This software pipeline implementation complies with the software project requirements listed in Chapter 9.1.

### 9.3.2 Quantitative Quality Assurance (Q<sup>2</sup>A) of the Low-level Vision Software Pipeline

For Q<sup>2</sup>A of the RGB image QNQ converter, a minimally redundant and maximally informative set of outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is), encompassing both outcome Q<sup>2</sup>Is and process Q<sup>2</sup>Is, must be selected, to be community agreed-upon in advance, in compliance with the QA4EO guidelines [20]. To be considered operational, an information processing system must score “high” in all of its OP-Q<sup>2</sup>I scores [23], [24]. In general, process is easier to measure, outcome is more important. Based on past related works [23], [24], the selected outcome and process Q<sup>2</sup>Is are the following. (i) Outcome Q<sup>2</sup>I: Effectiveness of the superpixel detection, estimated by means of two Q<sup>2</sup>Is to be jointly maximized. (a) A VQ error, e.g., estimated as an RMSE, see Eq. (1), to be minimized. Vice versa, the inverse of the RMSE is a Q<sup>2</sup>I to be maximized. (b) The number of image-objects, to be minimized by the image segmentation algorithm [90]. It is inversely related to a data compression rate, to be maximized. When the VQ error is zero, which means best case, because the number of segments is maximum and equal to the number of pixels, then segmentation is worst case because there is no pixel aggregation at all. An optimal compromise between these two mutually opposing Q<sup>2</sup>Is of effectiveness should be searched for. (ii) Process Q<sup>2</sup>I: Efficiency, specifically: (a) computation time, required to be linear in the image size, and (b) central memory occupation, required to be kept “low” to cope with massive images. (iii) Process Q<sup>2</sup>I: Degree of automation (ease of use), monotonically decreasing with the number of system’s free-parameters to be user-defined. Full automation is required (refer to Chapter 9.1). (iv) Process Q<sup>2</sup>I: Robustness to changes in the input dataset. Data-independence is required (refer to Chapter 9.1). (v) Process Q<sup>2</sup>I: Robustness to changes in input parameters, if any. (vi) Process Q<sup>2</sup>I: Scalability, to keep up with changes in users’ needs and sensor properties. (vii) Outcome Q<sup>2</sup>I: Timeliness, defined as the time between data acquisition and data-derived information product generation, to be minimized. (viii) Outcome Q<sup>2</sup>I: Costs, in (a) manpower and (b) computer power, to be minimized. It is noteworthy that in papers published in the CV and RS literature, a CV system Q<sup>2</sup>A policy typically estimates the product effectiveness and, in case, the process efficiency, although these Q<sup>2</sup>Is are *per se* insufficient to assess the system’s overall degree of operativeness.



### 9.3.3 Comparison with Alternative Approaches

The proposed RGB image QNQ transform for superpixel detection and Q<sup>2</sup>A fully complies with the Marr definition of raw primal sketch [4], refer to Chapter 9.2.1. As a consequence, it is preliminary not alternative to any inductive semi-automatic and site-specific algorithm [35] for either: (a) texture segmentation (perceptual grouping) [1], [2], [13], equivalent to the Marr's low-level vision full primal sketch [4], or (b) surface segmentation, equivalent to the Marr's high-level 2½D sketch [4], such as that presented in [42].

In addition, the proposed deductive automatic and data-independent superpixel detector should not be compared against any inductive semi-automatic and site-specific superpixel detection algorithm, such as the SLIC reviewed in Chapter 9.2.6 [41]. Belonging to two complementary rather than alternative families of inference algorithms for VQ, their process and outcome Q<sup>2</sup>Is are expected to feature complementary, rather than alternative behaviors, refer to Chapter 9.2.4. This theoretical expectation is verified below.

(i) The RGB image QNQ transform realization detects each superpixel as a connected set of pixels featuring the same color label in a multi-level color map, whose map legend is a color dictionary defined *a priori*. Provided with a segment ID together with a color label, this superpixel is a planar objects identified by the Geospatial Consortium (OGC) terminology as 0D Point (code 1), 1D LineString (code 2), 2D Polygon (code 3), MultiPoint (code 4), MultiLineString (code 5), MultiPolygon (code 6) [45]. Vice versa, superpixels without a color label, such as those detected by the SLIC, can be coded as an OGC's Point, LineString or Polygon, on theory. In practice, the SLIC requires superpixels to be equally-sized and compact, which means their OGC spatial type is expected to be Polygon (code 3) exclusively. A color label provides each image-object with a semi-symbolic information superior to that of sub-symbolic pixels, whose semantics is zero, but never superior to, i.e., always less than or equal to, semantics of target 4D objects, e.g., land cover classes depicted in an EO image. The automatic detection in near real-time of semi-symbolic image-objects agrees with the quote by Marr that "vision goes symbolic almost immediately, right at the level of zero-crossing (first-stage primal sketch), without loss of information" ([4]; p. 343), refer to Chapter 9.2.1.

(ii) In the implemented RGB image QNQ transform the shape of a superpixel can be any, e.g., an elongated superpixel, coded as LineString (code 2), features low compactness. In the SLIC algorithm, compact superpixels are parameterized by a compactness index  $m$  to be user-defined in range [1, 40].

(iii) In the implemented RGB image QNQ transform the superpixel size can be any, from a minimum superpixel area value equal to 1 up to a maximum superpixel area equal to the image size  $N$ . As a consequence, their total number can be any, from 1, i.e., there is one single superpixel across the whole image, up to  $N$ , i.e., there is one superpixel per pixel. In the SLIC algorithm, the number  $k$  of superpixels to be detected must be user-defined in advance and superpixels are forced to be equally-sized.

The conclusion is that there would be neither conceptual nor practical rationale in the direct comparison of process Q<sup>2</sup>Is and outcome Q<sup>2</sup>Is collected from the proposed deductive automatic superpixel detector and an inductive semi-automatic superpixel detection algorithm, such as the SLIC proposed in [41]. In practice, this comparison would be as pointless as comparing an inductive inference system with its initial conditions, to be provided by deductive inference, refer to Chapter 9.1.

## 9.4 Materials

Input RGB images were selected to test the RGB image QNQ transform implemented as a proof of concept. This proof-of-concept was required to comply with process and outcome requirements, e.g., data-independence and robustness to changes in the input dataset, provided with metrological/statistically-based Q<sup>2</sup>Is, refer to Chapter 9.3.2. Therefore, three uncalibrated RGB images acquired by different imaging sensors mounted on different acquisition platforms, specifically, terrestrial, airborne and spaceborne, were selected for testing purposes.

Although the implemented RGB image QNQ transform was conceived for RS applications, it applies to any RGB image, e.g., an image depicting an "environmental scene" or an "object view". By definition [100], an "object view" subtends 1 to 2 meters around the observer, a "view on a scene" begins when there is actually a larger space between the observer and the fixated point, usually after 5 meters. In a real-world "object view", the observer's familiarity with the depicted object works as "ground truth", i.e., the quality of the RGB image QNQ transform can be intuitively assessed visually. For this reason and to prove the claim that the RGB image QNQ transform applies to any RGB image, one RGB jpeg-format image of a human face acquired by a consumer-level Canon color camera (of course, provided with no radiometric calibration capability), featuring rows =  $RW = 230$  and columns =  $CL = 219$ , acquired by a consumer-level color camera, was collected





by means of a web-based search-by-noun engine, see Fig. 9-8. One non-calibrated false-color (R channel = Visible Red, G channel = Near Infrared, B channel = Visible Blue) three-band image of a public event in Munich, Germany, acquired in 2013 by a 700 m-high airborne platform with a 4-band LEICA Airborne Digital Scanner (ADS) 80, with  $RW = 2744$  and  $CL = 4616$ , was provided by the German Aerospace Center (DLR), see Fig. 9-9. One heterogeneous bi-temporal RGB synthetic aperture radar (SAR) image, with  $RW = 4480$  and  $CL = 5012$ , acquired by the COSMO-SkyMed SAR sensor, was provided by the University of Naples Federico II, Italy [91]. Such an RGB-SAR image consists of the following bi-temporal heterogeneous SAR variables. R channel: interferometric coherence value in range  $[0, 1]$  coded as float; to save memory space by a factor of 4:1, it can be byte-coded into range  $\{0, 255\}$  with a negligible discretization error, equal to  $((1./255)/2.) = 0.2\%$  (where factor 2 in the denominator is due to rounding the ratio  $1./255$  to the closer integer, either superior or inferior). G channel: Test image at time  $T1$  (e.g., rain season), SAR backscatter in range  $[0, 1]$ , byte-coded into range  $\{0, 255\}$ . B channel: Reference image at time  $T0 < T1$ , SAR backscatter in range  $[0, 1]$ , byte-coded into range  $\{0, 255\}$ , see Fig. 9-10.

## 9.5 Results

From a methodological standpoint (refer to Chapter 9.1 and Chapter 9.2.4), the goal of this experimental session was threefold. (a) Prove that the RGB image QNQ transform realization, where the software pipeline shown in Fig. 9-4 adopts the RGBIAM expert system as block 2 for VQ, is as a proof-of-concept able to meet the project requirements specified in Chapter 9.1. Hereafter, it is referred to as the RGBIAM-QNQ realization. (b) When different VQ approaches are employed as block 2 in Fig. 9-4 (refer to Chapter 9.2.4 and Chapter 9.3.3), prove that: (i) deductive VQ, inherently automatic and non-adaptive to data, in comparison with inductive VQ, inherently semi-automatic and site-specific, feature complementary, rather than alternative functional properties. (ii) A hybrid VQ system, where deduction initializes induction (refer to Chapter 9.1 and Chapter 9.2.4), overcomes limitations of each individual part.

To accomplish these experimental objectives, three VQ approaches were tested as block 2 in Fig. 9-4.

(I) A one-pass deductive RGBIAM discretization prototype with 49+1 quantization levels, including category “unknown”, actually empty.

(II) A traditional iterative unlabeled data learning  $k$ -means VQ algorithm, implemented in the ENVI commercial software toolbox [62], with free-parameters  $k = 49$ , max number of iterations  $I = 3$ , change threshold = 5%, with centroids initialized by random sampling of the training dataset (refer to Chapter 9.2.4). In agreement with an  $n$ -fold cross-validation approach adopted in [54], a three-fold cross-validation was applied to the  $k$ -means VQ algorithm. It was trained in each of the three test images, where the training error was assessed, refer to cells depicted in dark gray in Table 9-1 and Table 9-2. After reaching convergence, it was run upon the remaining two images, to assess its prediction error. The final VQ error was estimated as the sum of training and testing errors. To guarantee a fair accuracy assessment of the  $k$ -mean VQ algorithm adopted as block 2 in the software pipeline shown in Fig. 9-4, its VQ error was not estimated at the output of block 2, where each pixel is assigned to the closest center of mass in a codebook of  $k$  global (image-wide) data-centroids, but at the output of block 6, where a piecewise-constant approximation of the input RGB image was synthesized. Dealing with local rather than global statistics, the latter VQ error is never superior, i.e., it is always inferior or equal, to the former.

(III) The RGBIAM’s output 49-level color map of each test image was employed to initialize the  $k = 49$  data-centroids of the  $k$ -means VQ algorithm, with free-parameters to be user-defined equal to: max number of iterations  $I = 3$ , change threshold = 5%.

Table 9-1 and Table 9-2 show multiple heterogeneous OP-Q<sup>2</sup>Is, but computation time (refer to Chapter 9.3.2), collected from the three test algorithms ran upon each of the three test RGB images (refer to Chapter 9.4). Output products are shown in Fig. 9-8 to Fig. 9-10. To accomplish a multivariate analysis of a heterogeneous set of OP-Q<sup>2</sup>Is, featuring different unit of measure, range of change and sensitivity to changes in the input dataset, each univariate quantitative variable was z-scored (standardized) to feature zero mean and unit variance across test datasets per algorithm. Next, univariate non-dimensional z-scores were summed into a univariate “ultimate” score per algorithm, to allow inter-algorithm score comparisons, refer to Table 9-2. Hence, conclusions about inter-algorithm comparisons of process Q<sup>2</sup>Is and outcome Q<sup>2</sup>Is should stem from Table 9-2, generated from Table 9-1 by z-scoring.

In Table 9-3 the measured computation time of the RGBIAM-QNQ transform realization is shown as a dependent variable of the independent image size for each of the three input images. These time values include both a data processing time, relying on a fast central memory, and an I/O data time, involved with a slow secondary memory, when the software pipeline’s output products include two color maps, two segmentation maps, four contour maps and two piecewise-constant



RGB image approximations, refer to Chapter 9.3.1. Because the data processing time is estimated to be linear in the image size (refer to Chapter 9.3.1), the I/O data time is expected to increasingly dominate the data processing time when the image size increases.

Additional examples where the proposed RGBIAM-QNQ transform is employed for automatic analysis and regionalization of various 3-tuple combinations of either homogeneous or heterogeneous 2-D gridded variables are shown in Fig. 9-11 and Fig. 9-12 respectively.

## 9.6 Discussion

Introduced in Chapter 9.5, test products, shown in Fig. 9-8 to Fig. 9-10, process Q<sup>2</sup>Is and outcome (product) Q<sup>2</sup>Is, collected in Table 9-1 to Table 9-3, are discussed below.

(i) Outcome Q<sup>2</sup>I: Effectiveness, where two Q<sup>2</sup>Is (vice versa, cost variables) must be jointly maximized (vice versa, minimized), specifically: (a) the RMSE and (b) the number of segments or, vice versa, the mean image-object area, inversely related to the number of segments. See Fig. 9-8 to Fig. 9-10 for perceptual assessment of the three different VQ approaches adopted in block 2 of the software pipeline shown in Fig. 9-4. In Table 9-2, generated from Table 9-1 by z-scoring, best results are shown in bold in the bottom row. Noteworthy, they slightly differ from those shown in bold in Table 9-1. In Table 9-2, in perfect agreement with theoretical expectations (refer to Chapter 9.1), first, the mean image-object area is maximized by the data-independent RGBIAM expert system for VQ. Second, the RMSE is minimized across the three input datasets by the hybrid VQ approach, where the RGBIAM expert system initializes the learning-from-data *k*-means algorithm for VQ as block 2 in the software pipeline shown in Fig. 9-2.

(ii) Process Q<sup>2</sup>I: Efficiency of the RGBIAM-QNQ realization in terms of: (a) memory occupation and (b) computation time. The RGBIAM-QNQ software pipeline is implemented in tile streaming mode, suitable for massive data mapping problems, where the maximum dynamic memory occupation is fixed, irrespective of the image size. In these experiments the dynamic memory maximum size parameter was set equal to 800 MB of random access memory (RAM), which can be considered a “small” value for dynamic memory occupation in standard personal computers. In addition to these 800 MB of RAM, an estimate of the dynamic memory occupation required by the STD was provided in Chapter 9.3.1. About computation time, the computational complexity of the RGBIAM-QNQ realization was claimed to be linear in the image size (refer to Chapter 9.3.1). Therefore, data processing time was expected to be linear in the image size. When it was run on a Dell Power Edge 710 server with dual Intel Xeon @ 2.70 GHz processor with 64 GB of RAM and a 64-bit Linux operating system, the RGBIAM-QNQ realization required a computation time, including data processing and data I/O operations (refer to Chapter 9.5), shown in Table 9-3. It agrees well with the linear-time expectation. The system output rate, defined as the inverse of computation time, can be considered not inferior to a reasonable input rate. For example, the typical input rate of a massive EO image acquired by a geostationary spaceborne EO imaging sensor is one every 15 minutes. If this consideration holds true, then computation time of the proposed RGBIAM-QNQ realization can be considered near real-time.

(iii) Process Q<sup>2</sup>I: Degree of automation of the RGBIAM-QNQ realization. It requires neither user-defined parameter nor training dataset to run, i.e., it is fully automatic. Hence, its degree of supervision is zero. Vice versa, its degree of automation is maximum and cannot be surpassed by any alternative approach.

(iv) Process Q<sup>2</sup>I: Robustness to changes in the input dataset of the RGBIAM-QNQ realization. It is non-adaptive to input data, hence its robustness to changes in the input dataset is maximum and cannot be surpassed by alternative approaches.

(v) Process Q<sup>2</sup>I: Robustness to changes in input parameters of the RGBIAM-QNQ realization. It requires no user-defined parameter to run. Hence, its robustness to changes in input parameters is maximum and cannot be surpassed by alternative approaches.

(vi) Process Q<sup>2</sup>I: Maintainability/ scalability/ re-usability, to keep up with changes in users' needs and sensor properties. The proposed RGBIAM-QNQ realization can be applied to any existing or future planned RGB imaging sensor, including consumer-level color cameras mounted in smartphones or on board light-weight UAVs, whether or not radiometrically calibrated, in true colors or false colors (e.g., refer to Chapter 9.4), irrespective of the image size. Rather than as a standalone low-level vision subsystem, it should be employed at the first stage of an innovative hybrid feedback IUS architecture [23], [24], consisting of six stages, including stage 0 (zero) for image pre-processing, suitable for OBIA applications whose final output product consists of symbolic image-objects in a GIS-ready file format. Shown in Fig. 9-13, this novel IUS design is completely alternative to the mainstream inductive feedforward IUS architecture adopted by the large majority of the RS



and CV communities, e.g., support vector machines [36], random decision forest classifiers [37] and deep convolutional image neural networks [38], [39] (refer to Chapter 9.1).

(vii) Outcome Q<sup>2</sup>I: Timeliness, defined as the time span between data acquisition and product generation. The proposed RGBIAM-QNQ realization reduces timeliness from image acquisition to information product generation to almost zero, due to linear-time computation exclusively, because user's supervision is zero. On the contrary, timeliness of inductive semi-automatic and site-specific data learning algorithms, such as the inductive k-means VQ algorithm, is always superior to zero, due to the time spent by a human supervisor in selecting and collecting a training dataset representative of the complexity of the inductive learning-from-data problem, in addition to the computation time spent by the algorithm to learn-from-data before reaching convergence (stability) at the cost of plasticity.

(viii) Outcome Q<sup>2</sup>I: Costs, monotonically increasing with manpower and computer power. The RGBIAM-QNQ realization is prior knowledge-based and near real-time in a standard laptop computer. Its costs are zero in terms of user's supervision and almost negligible in terms of computational power. This does not hold for inherently semi-automatic and site-specific inductive data learning algorithms, such as the inductive k-means VQ algorithm. They require human supervision to define the system's free-parameters and to provide a testing dataset. It is worth mentioning that many existing real-time solutions in image and video segmentation are based on multi-scale window-based local statistic estimates. These algorithms are computationally very intensive on traditional central processing units (CPUs). In practice, they require more expensive information technology solutions, with parallel hardware to estimate subwindows independently and graphics processor units (GPUs) for acceleration purposes [88].

(ix) Process Q<sup>2</sup>I: Data compression rate, equal to 4:1 up to 6: 1, depending on the selected number of quantization regions, equivalent to color names.

According to the definition provided in Chapter 9.3.2, since it scores "high" in each of the selected OP-Q<sup>2</sup>Is, the RGBIAM-QNQ realization can be considered off-the-shelf, i.e., ready-for-use. In particular, it is: (a) eligible for inclusion in a CV software library of off-the-shelf low-level vision functions, such as OpenCV [92], and (b) suitable for use in mobile software applications, defined as lightweight computer programs specifically designed to run on web browsers and mobile devices, such as tablet computers and smartphones.

## 9.7 Conclusions

Outpaced by the rate of collection of images and videos of ever-increasing quality and quantity, such as those acquired by consumer- or commercial-level color cameras mounted in mobile devices or on unmanned aerial vehicles (UAVs), the CV and remote sensing (RS) communities are affected by an ongoing lack of *information-as-data-interpretation* capabilities [21]. To invert this trend, the present research and technological development (RTD) CV software project promotes the exploitation of *a priori* knowledge, available in addition to sensory data, to initialize inductive learning-from-data algorithms in agreement with biological cognitive systems, where genotype initializes phenotype. A multidisciplinary cognitive approach to vision (see Fig. 9-1), where image pre-processing and understanding are considered *hybrid* (combined deductive and inductive) inference problems, is in contrast with the mainstream RTD in computer vision and RS, focused on feedforward inductive image learning systems, such as support vector machines [36], random decision forest classifiers [37] and deep convolutional image neural networks [38], [39]. It is well known, but rather overlooked by the scientific community that inductive learning-from-data is an inherently ill-posed inference problem, which requires *a priori* knowledge in addition to data to become better posed for numerical treatment [32], [33].

In compliance with the Quality Assurance Framework for Earth Observation (QA4EO) requirements, the goal of the present low-level (pre-attentional) vision software design and implementation project was to develop an off-the-shelf hybrid reversible quantitative-to-nominal-to-quantitative (Q<sup>2</sup>NQ) transform of a monitor-typical red-green-blue (RGB) image for automatic superpixel detection and quantitative quality assurance (Q<sup>2</sup>A) in linear time. Any stack of three scalar 2-D gridded variables, for example, three heterogeneous geospatial variables such as a vegetation spectral index, a geospatial density function of population and a geospatial index of scholarization, can be selected by a user to be investigated as a monitor-typical RGB image. Whenever it is transformed in a monitor-typical RGB image, any three-channel combination of 2D gridded variables can be automatically investigated by the proposed expert system for RGB image Q<sup>2</sup>NQ transform. Irrespective of its implementation the proposed six-stage system design, shown in Fig. 9-4, simultaneously fills two traditional information gaps: from pixels to image-objects and from ever-varying sensory data to stable concepts, such as data-, user- and application-independent color names belonging to a color dictionary community-agreed upon in advance, equivalent to *a priori* visual knowledge. In the direct data coding phase, the input RGB image is



automatically partitioned into connected image-objects provided with a color label in linear time. These semi-symbolic planar entities feature some degree of semantics, e.g., color green is typical of vegetation. Known in the existing literature as texture elements, texels, tokens or superpixels, color-homogeneous planar entities belong to the raw primal sketch of the Marr's low-level vision system model [4]. In the inverse data decoding phase, a superpixelwise-constant RGB image reconstruction provides each pixel with an approximation error. By improving the structural and procedural knowledge of the static decision tree implemented as the deductive vector quantization (VQ) block 2 of the RGB image QNQ transform pipeline shown in Fig. 9-4, pixel-based approximation errors can be maintained below the target visual problem's VQ error requirement, e.g., see Fig. 9-8(t) to Fig. 9-8(v).

A realization of the low-level vision system design for automatic linear-time detection and  $Q^2A$  of semi-symbolic superpixels was presented as a proof-of-concept. (I) It complies with the quote by Marr that "vision goes symbolic almost immediately, right at the level of zero-crossing (first-stage primal sketch), without loss of information" ([4]; p. 343). (II) It is complementary not alternative to inductive ill-posed semi-automatic and site-specific data learning algorithms for either VQ, such as the  $k$ -means and ISODATA algorithms, or image segmentation algorithms related to the Marr's raw primal sketch, such as the SLIC [41], full primal sketch [13] or  $2\frac{1}{2}$  sketch [74] (refer to Chapter 9.2.1). For example, adopted as an automatic near real-time edge-preserving image smoothing filter, e.g., see Fig. 9-8(p) to Fig. 8(r), the proposed RGB image QNQ converter can enhance segmentation of a video sequence pursued by a Metropolis algorithm [88], [89]. (III) By scoring "high" in each of the selected outcome and process  $Q^2$ Is (OP- $Q^2$ Is), it can be considered worth to enrich a general-purpose low-level vision software library, such as OpenCV [92], because presumably useful to a broad audience. For example, it is suitable for use in mobile software applications, defined as lightweight computer programs specifically designed to run on web browsers and mobile devices, such as tablet computers and smartphones. (IV) Rather than being considered a standalone low-level vision module, any *a priori* knowledge-based RGB image QNQ transform can be adopted in the low-level vision first stage of a novel hybrid feedback IUS architecture, shown in Fig. 9-11. It is well known that color names "cannot always be inverted to unique land cover class names" [93]. To disambiguate one-to-many or many-to-many relationships of color names with classes of target objects belonging to the 4D real world-through-time, spatio-temporal properties of image-objects must be investigated, in addition to first-stage color properties. A target object class-specific spatio-temporal analysis conditioned by first-stage color analysis can be performed at the high-level information processing stages 2 and 3 of the novel IUS architecture, shown in Fig. 9-11, according to a stratified convergence-of-evidence classification approach consistent with human reasoning [1], [2], [23], [24]. Hybrid inference mechanisms provide this novel IUS architecture with feedback loops, from higher to lower information processing stages. It means that low-level qualitative/categorical/nominal information products, such as pre-classification color maps, are not exclusively useful to better condition high-level classification tasks. They can also be adopted for statistical stratification of inherently ill-posed image pre-processing tasks, such as Earth observation (EO) image atmospheric correction and topographic correction, image co-registration, image compositing, etc. [94]. For example, the proposed RGB image QNQ converter can be applied to a bi-temporal RGB-SAR image, see Fig. 9-10, to enhance estimation of the coherence input variable for high-coherence targets, such as vessels [98], as well as to improve detection of urban areas described as high-texture image areas [1].

Planned future applications of the proposed off-the-shelf RGB image QNQ transform will try to provide both traditional inherently ill-posed low-level image enhancement algorithms, such as SAR image despeckling [95], [96], and high-level image understanding algorithms, such as SAR image classification [97], [98], with novel solutions based on *a priori* knowledge in addition to data. For example, an innovative low-level vision project called "Automatic speckle model-free despeckling of bitemporal RGB-SAR imagery" is currently ongoing [99].

### Acknowledgment

To accomplish this work Andrea Baraldi was supported in part by the National Aeronautics and Space Administration (NASA) under Grant No. NNX07AV19G, issued through the Earth Science Division of the Science Mission Directorate. Dirk Tiede was supported in part by the Austrian Research Promotion Agency (FFG), in the frame of project AutoSentinel2/3, ID 848009. Andrea Baraldi thanks Prof. Raphael Capurro for his hospitality, patience, politeness and open-mindedness. He also thanks Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, for his friendship and support, together with Michael L. Humber, for his initial collaboration. The authors also wish to thank the Editor-in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.





## References in Chapter 9

- [1] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York, 1980.
- [2] T. Matsuyama and V. S. Hwang, *SIGMA: A Knowledge-Based Aerial Image Understanding System*. New York, NY: Plenum Press, 1990.
- [3] T. Gevers, A. Gijsenij, J. van de Weijer, J. M. Geusebroek, *Color in Computer Vision*. Hoboken, NJ, USA: Wiley, 2012.
- [4] D. Marr, *Vision*. New York, NY: Freeman and C., 1982.
- [5] H. du Buf and J. Rodrigues, Image morphology: from perception to rendering, in *IMAGE - Computational Visualistics and Picture Morphology*, 2007.
- [6] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall, 1994.
- [7] C. Mason and E. R. Kandel, “Central visual pathways,” in *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 1991, pp. 420–439.
- [8] E. R. Kandel, “Perception of motion, depth and form,” in *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 1991, pp. 441–466.
- [9] J. Rodrigues and J.M. Hans du Buf, “Multi-scale lines and edges in V1 and beyond: Brightness, object categorization and recognition, and consciousness,” *BioSystems*, vol. 1, pp. 1-21, 2008.
- [10] J. Rodrigues and J.M.H. du Buf, “Multi-scale keypoints in V1 and beyond: Object segregation, scale selection, saliency maps and face detection”, *BioSystems*, vol. 86, pp. 75–90, 2006.
- [11] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, “Simulation of neural contour mechanisms: from simple to end-stopped cells,” *Vision Res.*, vol. 32, no. 5, pp. 963–981, 1992.
- [12] L. Pessoa, “Mach Bands: How Many Models are Possible? Recent Experimental Findings and Modeling Attempts”, *Vision Res.*, Vol. 36, No. 19, pp. 3205–3227, 1996.
- [13] A. Jain and G. Healey, “A multiscale representation including opponent color features for texture recognition,” *IEEE Trans. Image Process.*, vol. 7, pp. 124–128, 1998.
- [14] J. I. Yellott, “Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture,” *Opt. Soc. Am.*, vol. 10, no. 5, pp. 777-793, May 1993.
- [15] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, “Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited,” *Perception*, vol. 2, pp. 391-405, 1973.
- [16] B. Julesz, “Texton gradients: The texton theory revisited,” in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, vol. 54, no. 4-5, Aug. 1986.
- [17] J. Victor, “Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: Comment,” *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1680-1684, May 1994.
- [18] D. Lowe, “Distinctive image feature from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, pp. 91–110, 2004.
- [19] Q. Iqbal and J. K. Aggarwal, “Image retrieval via isotropic and anisotropic mappings,” in *Proc. IAPR Workshop Pattern Recognit. Inf. Syst.*, Setubal, Portugal, Jul. 2001, pp. 34–49.
- [20] *A Quality Assurance Framework for Earth Observation*, version 4.0, Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS), 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- [21] R. Capurro and B. Hjørland, “The concept of information,” *Annual Review of Information Science and Technology*, vol. 37, pp. 343-411, 2003.
- [22] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.
- [23] A. Baraldi and L. Boschetti, “Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction,” *Remote Sens.*, vol. 4, pp. 2694-2735., 2012.
- [24] A. Baraldi and L. Boschetti, “Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system



- architecture, information/knowledge representation, algorithm design and implementation,” *Remote Sens.*, vol. 4, pp. 2768-2817, 2012.
- [25] B. Berlin and P. Kay, *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.
- [26] G. A. Miller, "The cognitive revolution: a historical perspective", in *Trends in Cognitive Sciences*, vol. 7, pp. 141-144, 2003.
- [27] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.
- [28] F. Capra and P. L. Luisi, *The Systems View of Life: A Unifying Vision*. Cambridge, UK: Cambridge University Press, 2014.
- [29] J. Piaget, *Genetic Epistemology*. New York, NY: Columbia University Press, 1970.
- [30] D. Parisi, "La scienza cognitive tra intelligenza artificiale e vita artificiale," in *Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza*, Bologna, Italy: Patron Editore, 1991
- [31] J. Hadamard, "Sur les problemes aux derivees partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.
- [32] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley, 1998.
- [33] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- [34] R. Laurini and D. Thompson, *Fundamentals of Spatial Information Systems*. London, UK: Academic Press, 1992.
- [35] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons, 2004.
- [36] M. Pontil and A. Verri, "Support vector machines for 3d object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 6, pp. 637-646, 1998.
- [37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Proc. Systems*, vol. 25, pp. 1106-1114, 2012.
- [39] A. Romero, *Assisting the training of deep neural networks with applications to computer vision*. PhD. Thesis, Dept. of Applied Mathematics and Analysis, Univ. of Barcelona, 2015.
- [40] N. Hunt and S. Tyrrell, *Stratified Sampling*. Coventry University, 2012. [Online] Available: <http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html>
- [41] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 1, pp. 1-8, vol. 6, no. 1, 2011.
- [42] E. Vazquez, R. Baldrich, J. van de Weijer, M. Vanrell, "Describing reflectances for color segmentation robust to shadows, highlights, and textures". *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 917-930, 2011.
- [43] F. Fonseca, M. Egenhofer, P. Agouris, and G. Camara, "Using ontologies for integrated geographic information systems," *Transactions in GIS*, vol. 6, pp. 231–257, 2002.
- [44] C. Liedtke, J. Buckner, O. Grau, S. Growe, and R. Tonjes, "AIDA: A system for the knowledge based interpretation of remote sensing data," in *3rd International Airborne Remote Sensing Conference*, 1997.
- [45] *OpenGIS implementation standard for geographic information - simple feature access - Part 1: Common architecture*, Open Geo-spatial Consortium Inc., May 2011. [Online]. Available: <http://www.opengeospatial.org/standards/sfa>
- [46] T. Martinetz, G. Berkovich, and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
- [47] B. Fritzke, "Some competitive learning methods," Draft document, 1997. [Online]. Available: <http://sund.de/netze/applets/gng/full/tex/DemoGNG/DemoGNG.html>
- [48] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part I," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 645-661, March 2002.
- [49] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part II," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 662 -677, March 2002.
- [50] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, "Geographic object-based image analysis - towards a new paradigm," *ISPRS J. Photogram. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.
- [51] S. Vecera and M. Farah, "Is visual image segmentation a bottom-up or an interactive process?," *Percept. Psychophys.*, vol. 59, pp.1280–1296, 1997.



- [52] Tian-Yuan Shih, "The reversibility of six geometric color spaces," *Photogram. Eng. Remote Sens.*, vol. 61, no. 10, pp. 1223-1232, 1995.
- [53] L. D. Griffin, "Optimality of the basic color categories for classification", *J. R. Soc. Interface*, vol. 3, pp. 71–85, 2006.
- [54] R. Khan, J. van de Weijer, F. Shahbaz Khan, D. Muselet, C. Ducottet, C. Barat, "Discriminative color descriptors", *CVPR* 2013.
- [55] S. W. Thomas, "Efficient Inverse Color Map Computation", in *Graphics Gems II*, Boston, MA: Academic Press. 1991.
- [56] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments", *IEEE Trans. Image Proc.*, vol. 20, no. 9, pp. 2475-2489, 2010.
- [57] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1209-1221, 2001.
- [58] A. Baraldi and F. Parmiggiani, "Combined detection of intensity and chromatic contours in color images," *Optical Eng.*, vol. 35, no. 5, pp. 1413-1439, May 1996.
- [59] J. Vazquez-Corral, M. Vanrell, R. Baldrich, F. Tous, "Color constancy by category correlation". *IEEE Trans. Image Proc.*, vol. 21, no. 4, pp. 1997-2007, 2012.
- [60] E. Land, "Recent advances in retinex theory," *Vision Research*, vol. 26, pp. 7–21, 1986.
- [61] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Proc.*, vol. 16, no. 9, pp. 2207–2214, 2007.
- [62] *ENVI EX User Guide 5.0*, ITT Visual Information Solutions, Dec. 2009. [Online]. Available: [http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI\\_EX\\_User\\_Guide.pdf](http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf)
- [63] *Exelis VIS Technical Support*, Personal communication, March 13, 2013.
- [64] J. van de Weijer, C. Schmid, J. Verbeek, D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1512 – 1523, 2009.
- [65] R. Benavente, R. M. Vanrell, and R. Baldrich, "Parametric fuzzy sets for automatic color naming," *J. Opt. Society of America A*, vol. 25, pp. 2582-2593, 2008.
- [66] G. Patanè and M. Russo, "The enhanced-LBG algorithm," *Neural Networks*, vol. 14, no. 9, pp. 1219–1237, 2001.
- [67] G. Patanè and M. Russo, "Fully automatic clustering system," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1285-1298, 2002.
- [68] B. Fritzke, "The LBG-U method for vector quantization - An improvement over LBG inspired from neural networks," *Neural Processing Lett.*, vol. 5, no. 1, 1997.
- [69] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–94, Jan. 1980.
- [70] D. Lee, S. Baek, and K. Sung, "Modified k-means algorithm for vector quantizer design," *IEEE Signal Processing Lett.*, vol. 4, pp. 2–4, Jan. 1997.
- [71] N. Memarsadeghi, D. Mount, N. Netanyahu, and J. Le Moigne, "A fast implementation of the ISODATA clustering algorithm," *Int. J. Comp. Geometry & Applications*, vol. 17, no. 1, pp. 71-103, 2007.
- [72] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982
- [73] C. Elkan, "Using the triangle inequality to accelerate k-means," *Int. Conf. Machine Learning*, 2003.
- [74] D. Tiede, S. d'Oleire-Oltmanns and A. Baraldi, "Geospatial 2D AND 3D object-based classification and 3D reconstruction of ISO-containers depicted in a LiDAR dataset and aerial imagery of a harbor," Proc. IGARSS 2015, Milan, Italy, 27-31 July 2015. Awarded with the 2nd Place in the IEEE GRSS 2015 Data Fusion Contest.
- [75] M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," *ISPRS J. Photogramm.*, vol. 58, pp. 12–23, 2000.
- [76] *eCognition® Developer 9.0 Reference Book*, Trimble, 2015.
- [77] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, Aug. 1988.
- [78] A. Baraldi, L. Bruzzone, and P. Blonda, "Quality assessment of classification and cluster maps without ground truth knowledge," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, 2005.
- [79] M. B. Dillencourt, H. Samet, and M. Tamminen, "A general approach to connected component labeling for arbitrary image representations," *J. Assoc. Computing Machinery*, vol. 39, pp. 253-280, 1992.
- [80] L. Macaire, N. Vandenbroucke and J.G. Postaire, "Color image segmentation by analysis of subset connectedness and color homogeneity properties," *Computer Vis. Image Understanding*, vol. 102, no. 1, pp. 105–116, 2006.



- [81] K. L. Kelly and D. B. Judd, "Color: universal language and dictionary of names," *National Bureau of Standards*, vol. 189, 1976.
- [82] A. Baraldi, P. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 2563-2586, 2006.
- [83] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338-353, 1965.
- [84] A. Baraldi, "Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, pp. 2113-2134, 2011.
- [85] I. M. Elfadel and R. W. Picard, "Gibbs random fields, cooccurrences, and texture modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 24-37, Jan. 1994.
- [86] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354-2363, 2003.
- [87] A. K. Shackelford, *Development of urban area geospatial information products from high resolution satellite imagery using advanced image analysis techniques*. Ph.D. dissertation, University of Missouri- Columbia, 2004.
- [88] S. Reich, A. Abramov, J. Papon, F. Wörgötter and B. Dellen, "A novel real-time edge-preserving smoothing filter," *8th Int. Conf. Computer Vision Theory and Applications*, Feb. 2013.
- [89] A. Abramov, K. Pauwels, J. Papon, F. Wörgötter, and B. Dellen, "Real-time segmentation of stereo videos on a portable system with a mobile gpu," *IEEE Trans. Circuits and Systems for Video Technology*, 2012.
- [90] Hui Li, Jianfei Cai, Thi Nhat Anh Nguyen, Jianmin Zheng, "A benchmark for semantic image segmentation," *IEEE ICME*, 2013.
- [91] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A New Framework for SAR multitemporal data RGB representation: Rationale and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117-133, 2015.
- [92] *Open source computer vision library (OpenCV)*, [Online] Available: <http://opencv.org/>
- [93] J. B. Adams, E. S. Donald, V. Kapos, R. Almeida Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie, "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon," *Remote Sens. Environ.*, vol. 52, pp. 137-154, 1995.
- [94] A. Baraldi, M. Gironde, and D. Simonetti, "Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 112-146, Jan. 2010.
- [95] G. F. De Grandi, M. Leysen, J. S. Lee, and D. Schuler, "Radar reflectivity estimation using multiple SAR scenes of the same target: Technique and applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 1997, pp. 1047-1050.
- [96] S. Quegan and Jiong Jiong Yu, "Filtering of multichannel SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2373-2379, 2001.
- [97] F. Bovenga, D. Derauw, F. M. Rana, C. Barbier, A. Refice, N. Veneziani, and R. Vitulli, "Multi-chromatic analysis of SAR images for coherent target detection," *Remote Sens.*, vol. 6, pp. 8822-8843, 2014.
- [98] S. Quegan, Jiong Jiong Yu, H. Balzter, and T. LeToan, "Combining unsupervised and knowledge-based methods in large-scale forest classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. 2000 (IGARSS 2000)*, vol. 1, pp. 426-428.
- [99] A. Baraldi, D. Tiede and S. Lang, "Automatic speckle model-free despeckling of bitemporal RGB-SAR imagery," Univ. Internal Report, 10-2015, Dept. Geoinformatics – Z\_GIS, University of Salzburg, Salzburg, Austria.
- [100] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Computer Vision*, vol. 42, no. 3, pp. 145-175, 2001.
- [101] C. Chubb and J.I. Yellott, "Every discrete, finite image is uniquely determined by its dipole histogram," *Vision Research*, vol. 40, pp. 485-492, 2000.
- [102] J. K. Tsotsos, "A complexity level analysis of vision", in *Proc. Int. Conf. on Computer Vision, Human and Computer Vision Workshop*, London, England, June 1987.
- [103] S. Frintrop, "Computational visual attention," in *Computer Analysis of Human Behavior, Advances in Pattern Recognition*, A. A. Salah and T. Gevers, Eds., Springer, 2011.
- [104] A. C. Watts, V. G. Ambrosia and E. A. Hinkle, "Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use," *Remote Sens.*, vol. 4, pp. 1671-1692, 2012.





- [105] *Sentinel-2 User Handbook*, European Space Agency, Standard Document, Issue 1 Rev 2, Date 24/07/2015.
- [106] *Venus Satellite Sensor Level 2 Product*, CNES. [Online]. Available:  
[https://venus.cnes.fr/en/VENUS/prod\\_l2.htm](https://venus.cnes.fr/en/VENUS/prod_l2.htm)

Figures and figure captions in Chapter 9

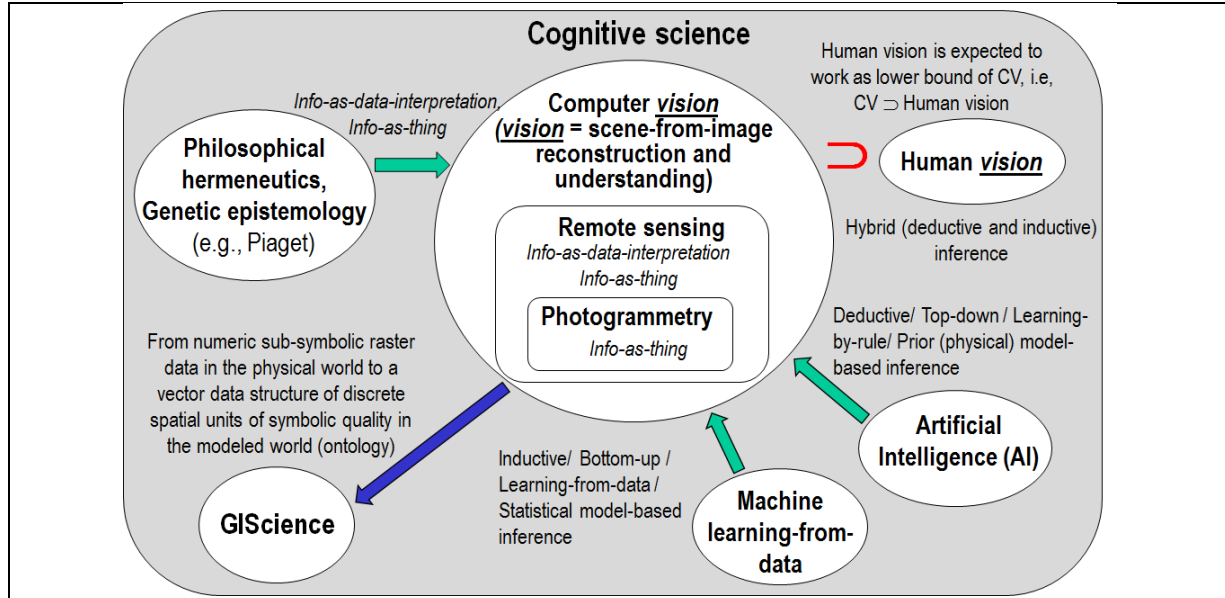


Fig. 9-1. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning [28]) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired from sensory data, processed and transferred within nervous systems (humans or other animals) and machines (e.g., computers) [26]-[28].

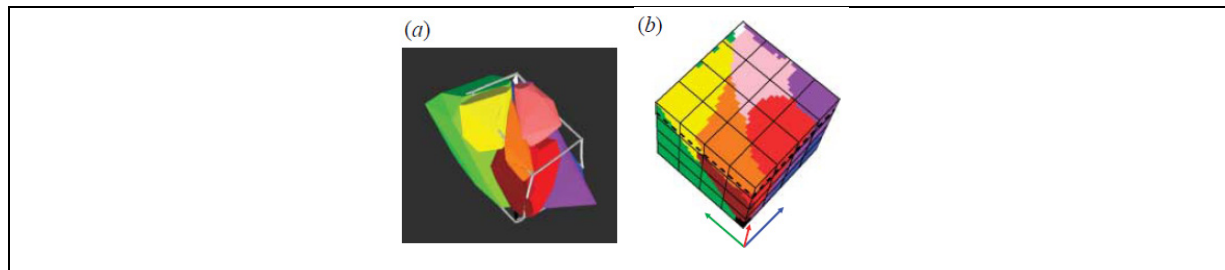


Fig. 9-2. Stages in mapping human basic colors (BCs) into the RGB cube. (a) CIE-Lab space with regions of eleven basic color labels as colored polyhedra and the edges of the monitor-typical RGB cube (in grey). (b) 323 quantization of the RGB space with the basic color extents from (a) mapped into it. The uniform 43 quantization of the RGB cube shown in (b) was adopted for representing color category systems whose classification performance was assessed. Images reproduced courtesy of [53].

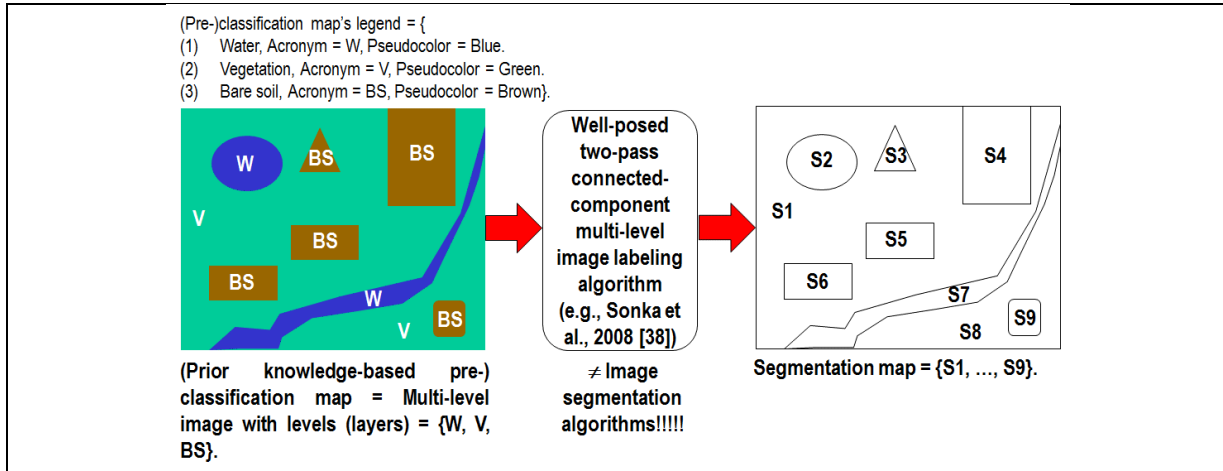


Fig. 9-3. One segmentation map is deterministically generated from one multi-level image, e.g., a classification map, but the vice versa does not hold, i.e., many multi-level images can generate the same segmentation map. In this example, stratum Vegetation, V, of the classification map consists of the two disjoint image objects, S1 and S9. Each image-object S1 to S9 consists of a connected set of pixels sharing the same label. Hence, the three spatial primitives labeled pixels, labeled segments and labeled strata (layers) co-exist in parent-child relationships.

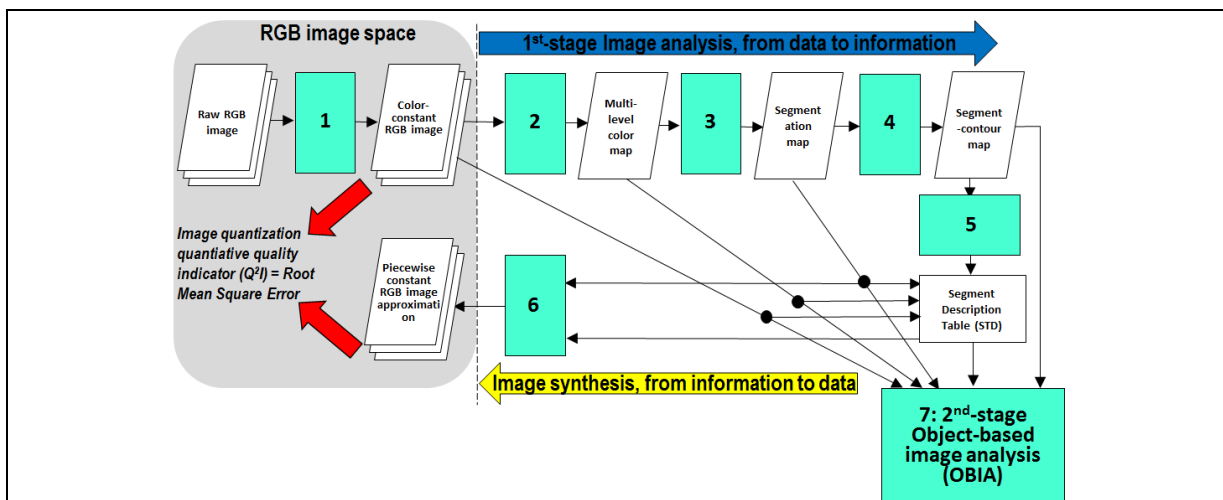


Fig. 9-4. The automated near real-time RGBIAM software toolbox for prior knowledge-based vector quantization, VQ, consists of subsystems 1 to 6. Phase 1-of-2 = Encoding phase/Image analysis - Stage 1: Self-organizing statistical algorithm for color constancy. Stage 2: Prior knowledge-based (static) RGBIAM decision tree for RGB cube partitioning (quantization, polyhedralization). Stage 3: Well-posed two-pass connected-component detection in the multi-level color map. Connected-components in the color map domain are connected sets of pixels featuring the same color label. These connected-components are also called image-objects, segments or superpixels. Stage 4: Well-posed superpixel-contour extraction. Stage 5: Well-posed Superpixel Description Table 9-(STD) allocation and initialization. Phase 2-of-2 = Decoding phase/Image synthesis - Stage 6: Superpixelwise-constant input image approximation (“object-mean view”) and per-pixel VQ error estimation. (Stage 7: in cascade to the RGBIAM’s superpixel detection, a high-level object-based image analysis (OBIA) approach can be adopted).

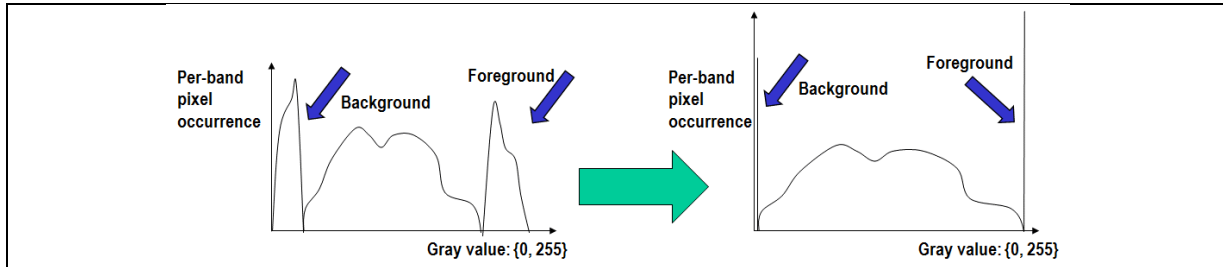


Fig. 9-5. Example of a gray-value univariate (one-channel) 1st-order distribution, to be stretched for color constancy. Four categories of univariate distributions can be considered: (i) neither a background nor a foreground mode is present in addition to a central mode; (ii) only a background mode with a right tail can be identified, (iii) only a foreground mode with a left tail can be identified, and (iv) both background and foreground modes are present in addition to a central mode.

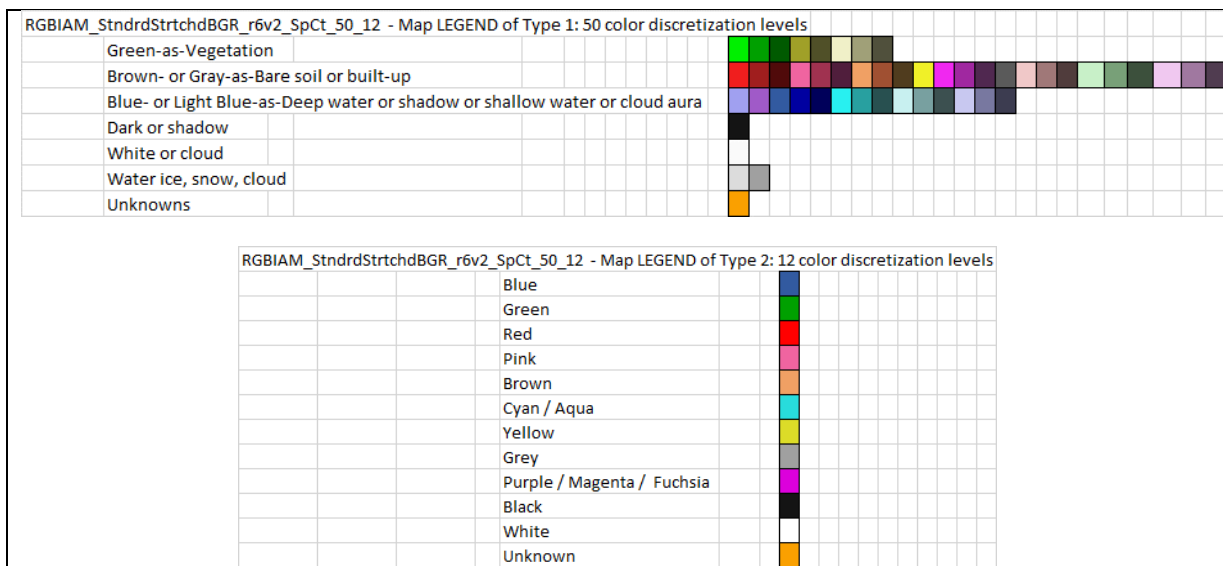


Fig. 9-6. Two-level RGB cube quantizer. Color map's legend at (left) fine (49 + 1 class unknown) and (right) coarse (11 + 1 class unknown) quantization levels, where the latter is a mutually exclusive and totally exhaustive combination of the former.

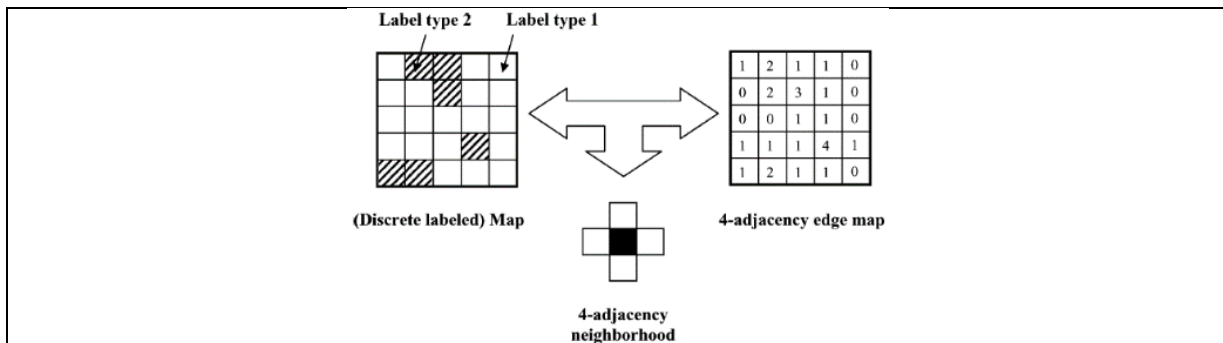
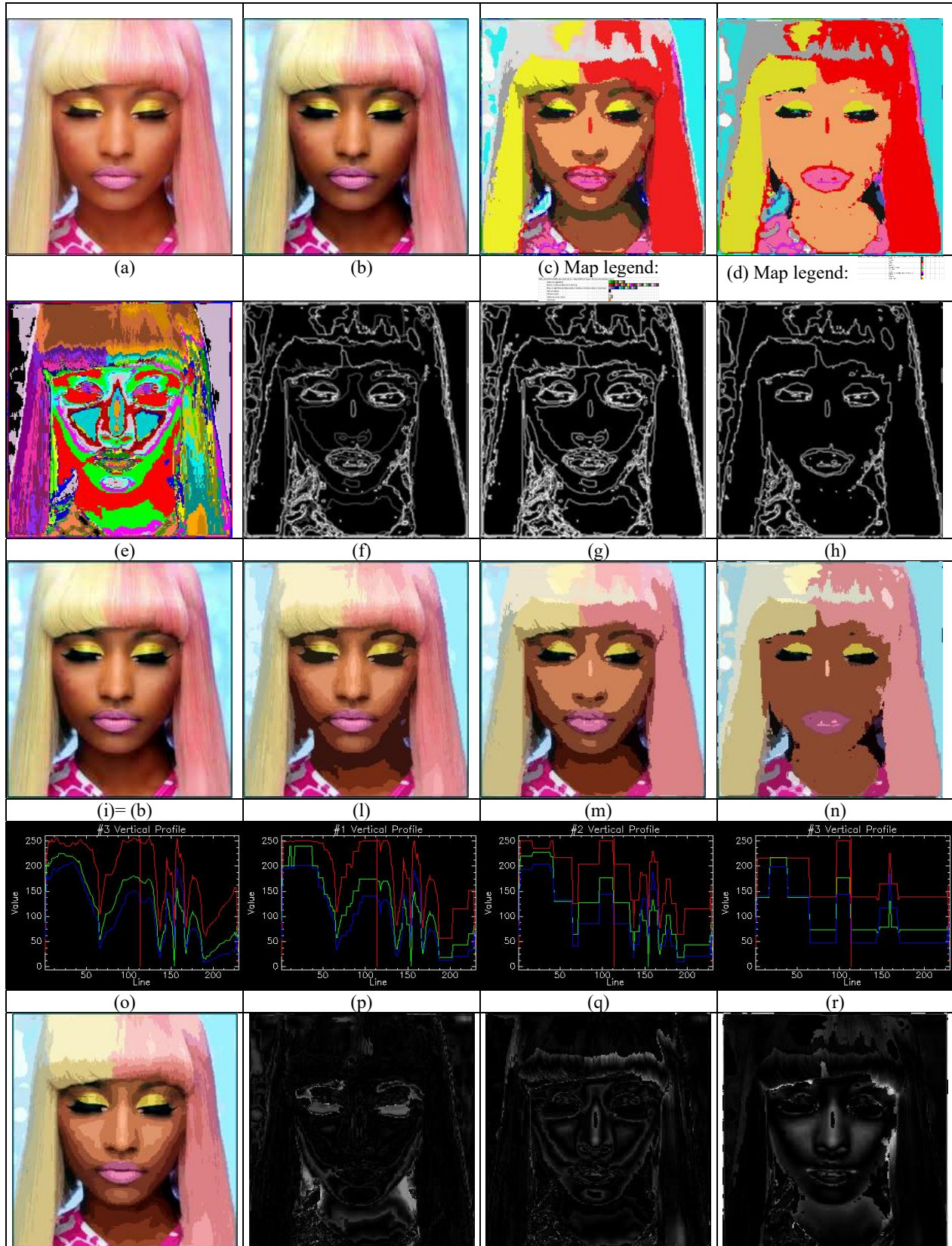


Fig. 9-7. Example of a 4-adjacency cross-aura map, shown at right, generated from a two-level image, shown at left [78], [85].







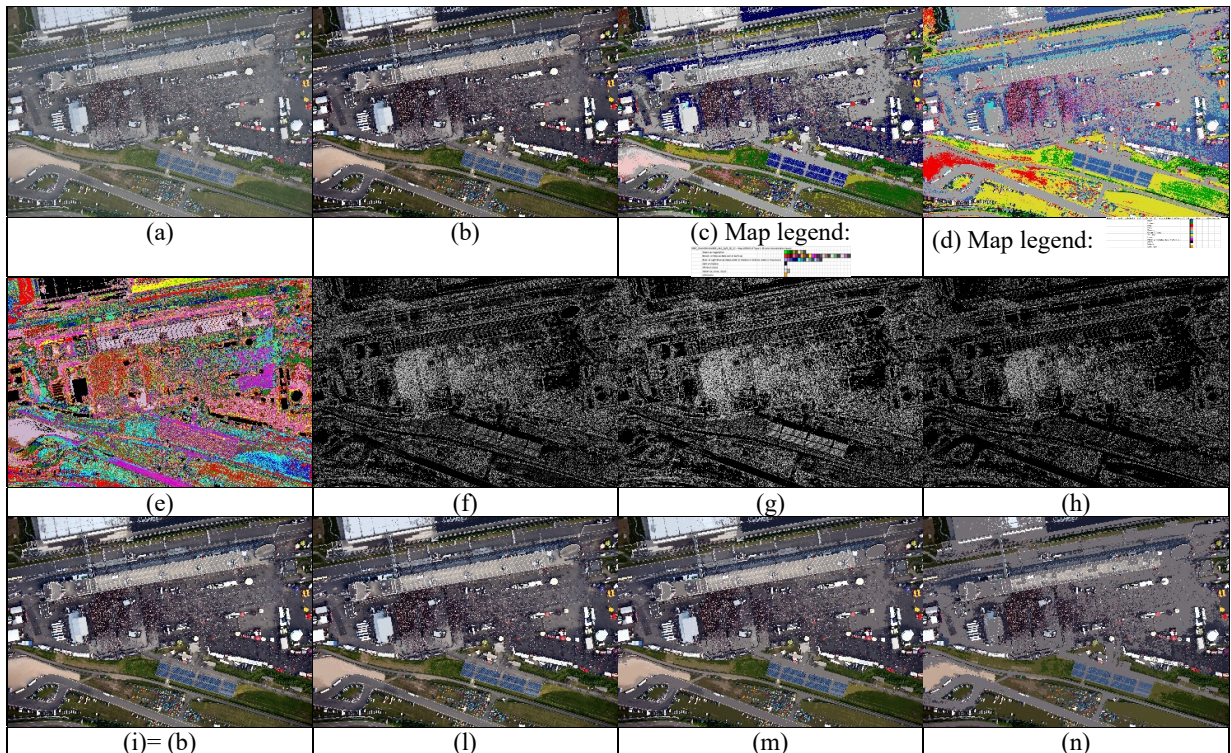
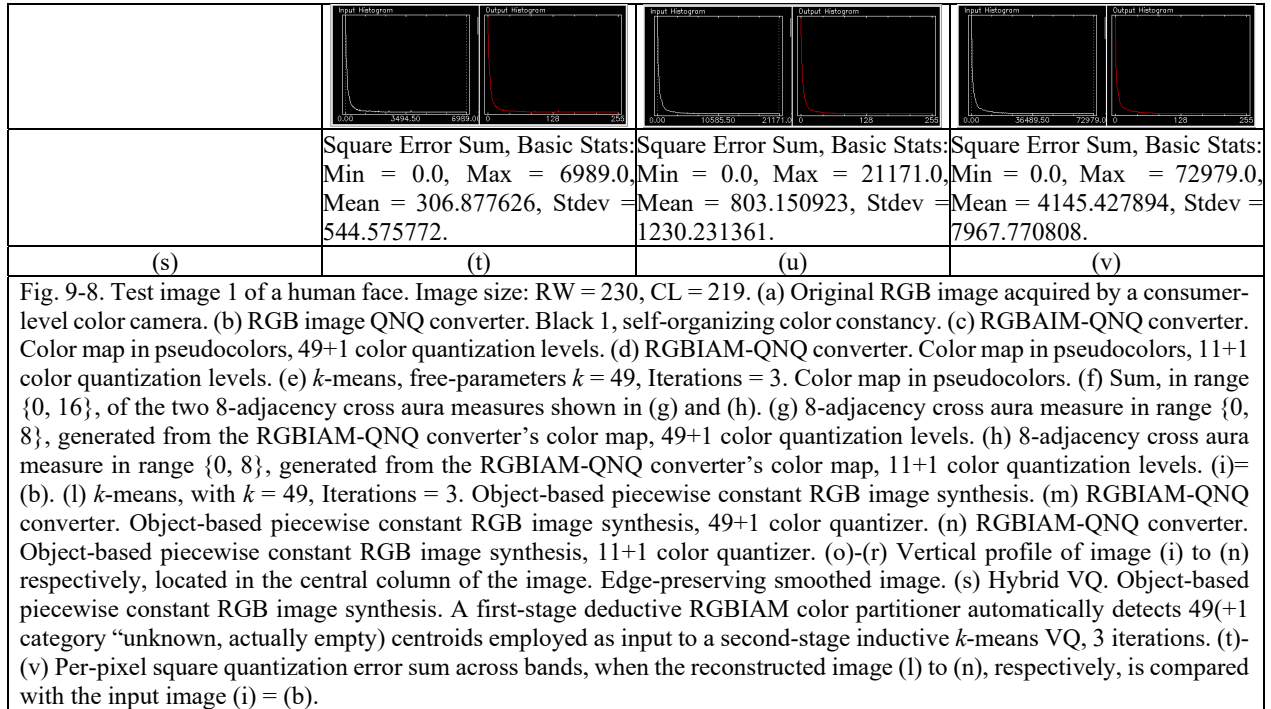


Fig. 9-9. Test image 2. 700 m-height aerial image of a public event in Munich, Germany (courtesy of DLR). Image size: RW = 2744, CL = 4616. (a) Original non-calibrated RGB image in false colors (R = Visible Red, G = Near Infrared, B = Visible Blue). (b) RGB image QNQ converter. Block 1, self-organizing color constancy. (c) RGBIAM-QNQ converter. Color map in pseudocolors, 49+1 color quantization levels. (d) RGBIAM-QNQ converter. Color map in pseudocolors, 11+1 color quantization levels. (e)  $k$ -means, free-parameters  $k = 49$ , Iterations = 3. Color map in pseudocolors. (f) Sum, in range  $\{0, 16\}$ , of the two 8-adjacency cross aura measures shown in (g) and (h). (g) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM-QNQ converter's color map, 49+1 color quantization





levels. (h) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM-QNQ converter's color map, 11+1 color quantization levels. (i) = (b). (l)  $k$ -means, with  $k = 49$ , Iterations = 3. Object-based piecewise constant RGB image synthesis. (m) RGBIAM-QNQ converter. Object-based piecewise constant RGB image synthesis, 49+1 color quantizer. (n) RGBIAM-QNQ converter. Object-based piecewise constant RGB image synthesis, 11+1 color quantizer.

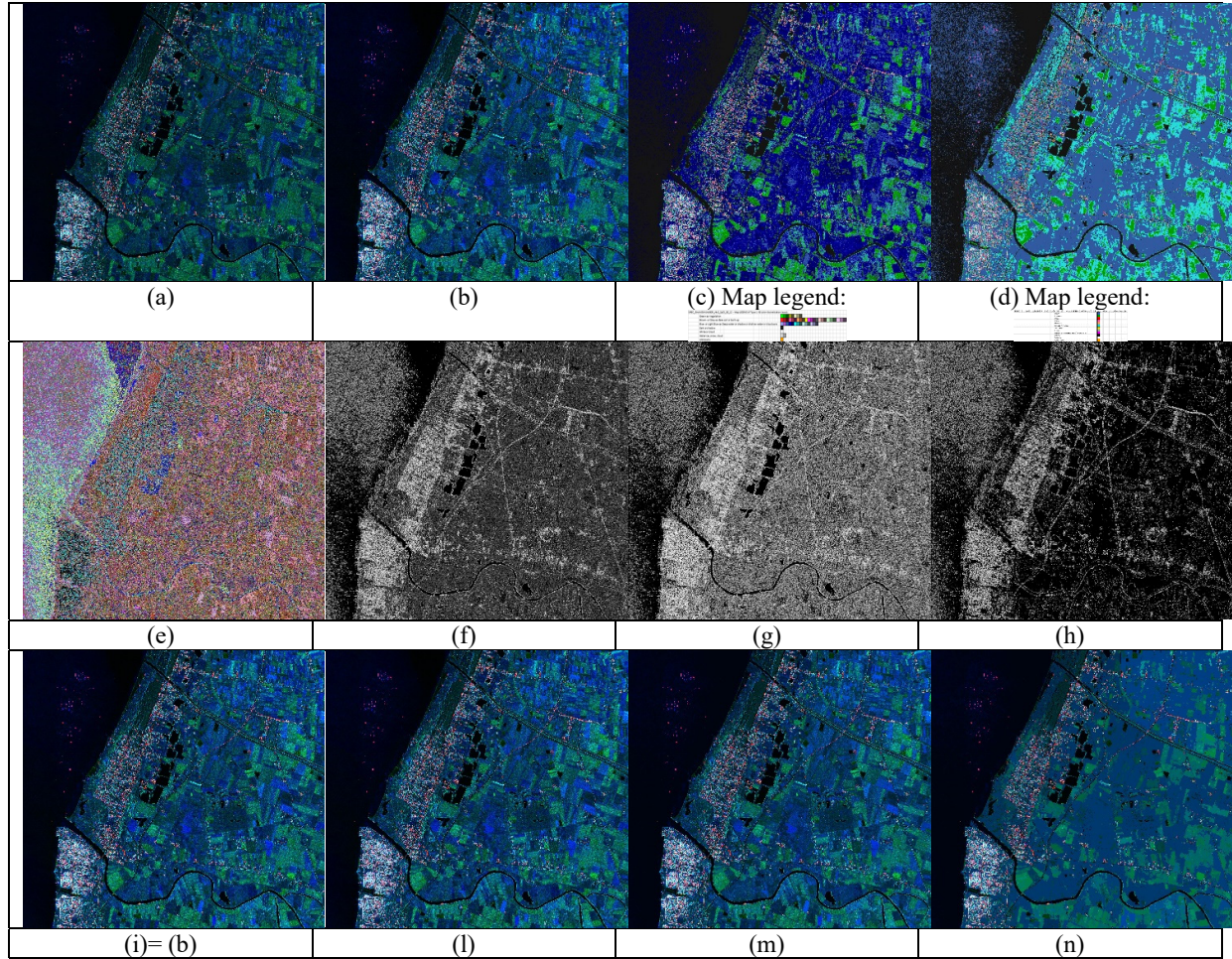
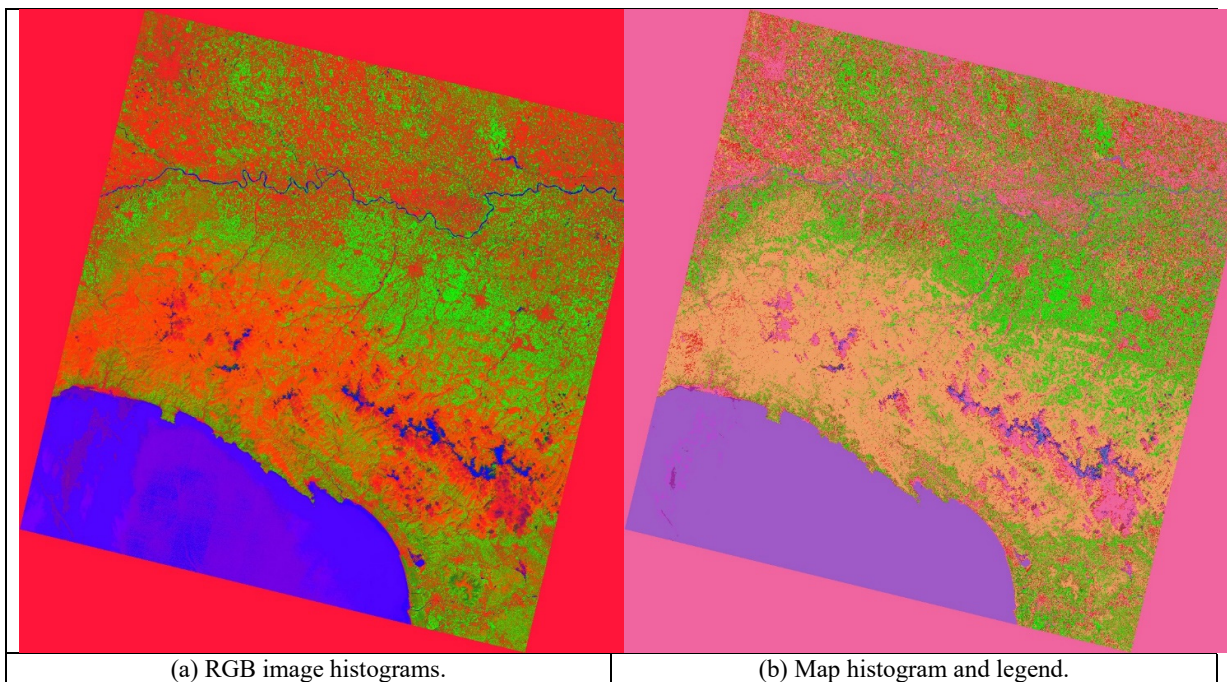
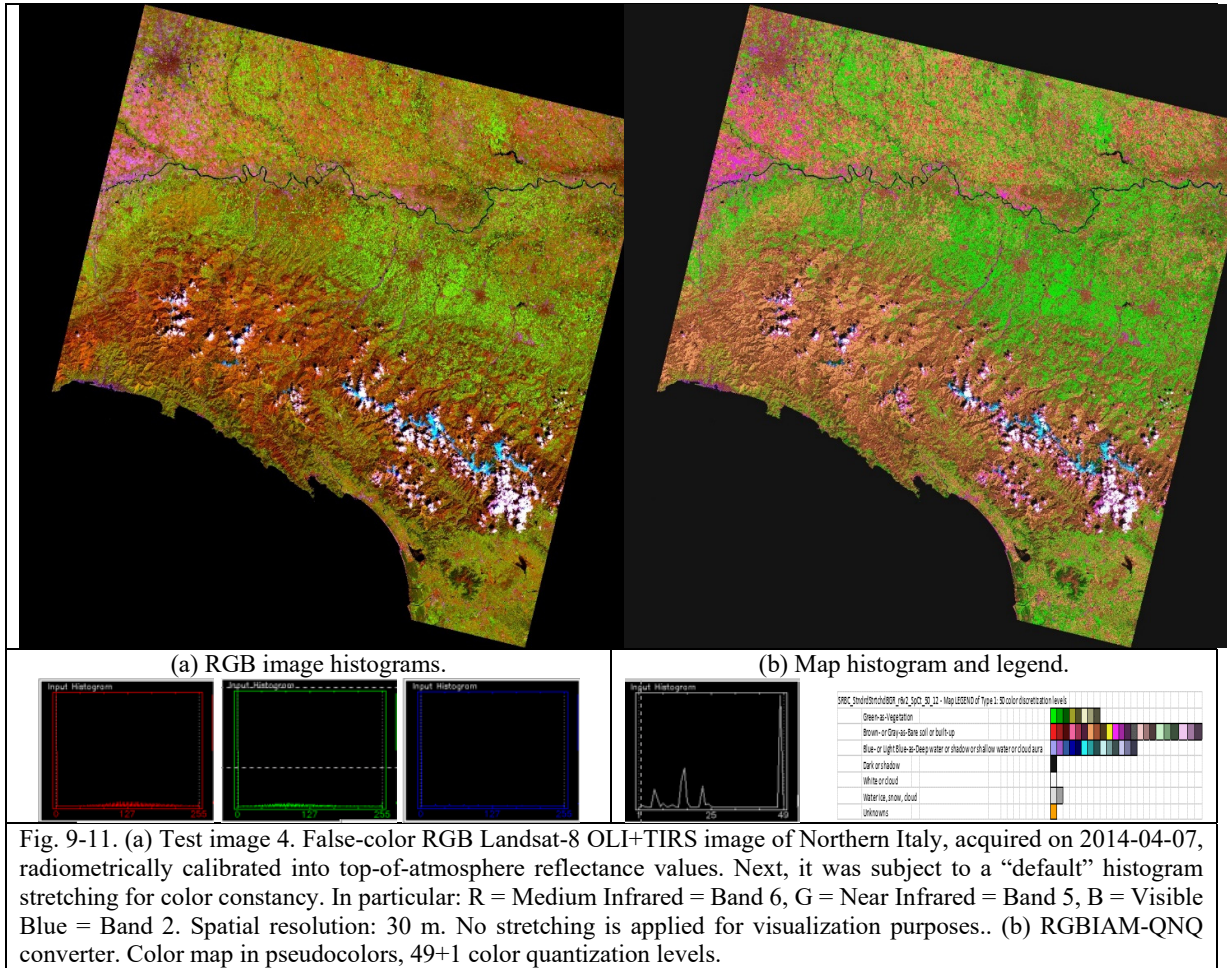


Fig. 9-10. Test image 3. Spaceborne bi-temporal RGB-SAR image of the Campania region, Italy (courtesy of University of Naples Federico II, Italy). Image size: RW = 4480, CL = 5012. (a) Original RGB-SAR image. (b) RGB image QNQ converter. Block 1, self-organizing color constancy. (c) RGBIAM-QNQ converter. Color map in pseudocolors, 49+1 color quantization levels. (d) RGBIAM-QNQ converter. Color map in pseudocolors, 11+1 color quantization levels. (e)  $k$ -means, free-parameters  $k = 49$ , Iterations = 3. Color map in pseudocolors. (f) Sum, in range  $\{0, 16\}$ , of the two 8-adjacency cross aura measures shown in (g) and (h). (g) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM-QNQ converter's color map, 49+1 color quantization levels. (h) 8-adjacency cross aura measure in range  $\{0, 8\}$ , generated from the RGBIAM-QNQ converter's color map, 11+1 color quantization levels. (i) = (b). (l)  $k$ -means, with  $k = 49$ , Iterations = 3. Object-based piecewise constant RGB image synthesis. (m) RGBIAM-QNQ converter. Object-based piecewise constant RGB image synthesis, 49+1 color quantizer. (n) RGBIAM-QNQ converter. Object-based piecewise constant RGB image synthesis, 11+1 color quantizer.







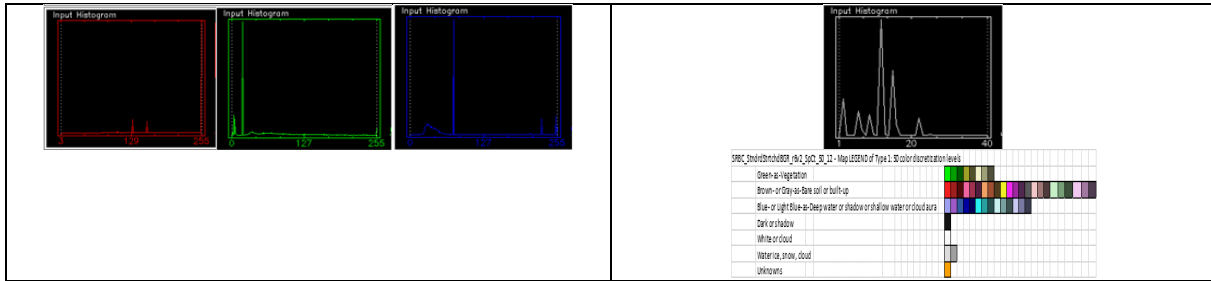


Fig. 9-12. (a) Test image 5. Landsat-8 OLI+TIRS image, acquired on 2014-04-07, Spatial resolution: 30 m, radiometrically calibrated into TOARF values. Band-ratio spectral indexes (SI), R = Bare Soil SI = Medium Infrared / Near Infrared, G = Vegetation SI = Near Infrared / Visible Red, B = Water SI = Visible Red / Medium Infrared, subject to histogram stretching for color constancy and mounted onto the RGB data space. These three SIs are equivalent to fuzzy membership functions and/or heterogeneous continuous input variables. No stretching is applied for visualization purposes.. (b) RGBIAM-QNQ converter. Color map in pseudocolors, 49+1 color quantization levels.

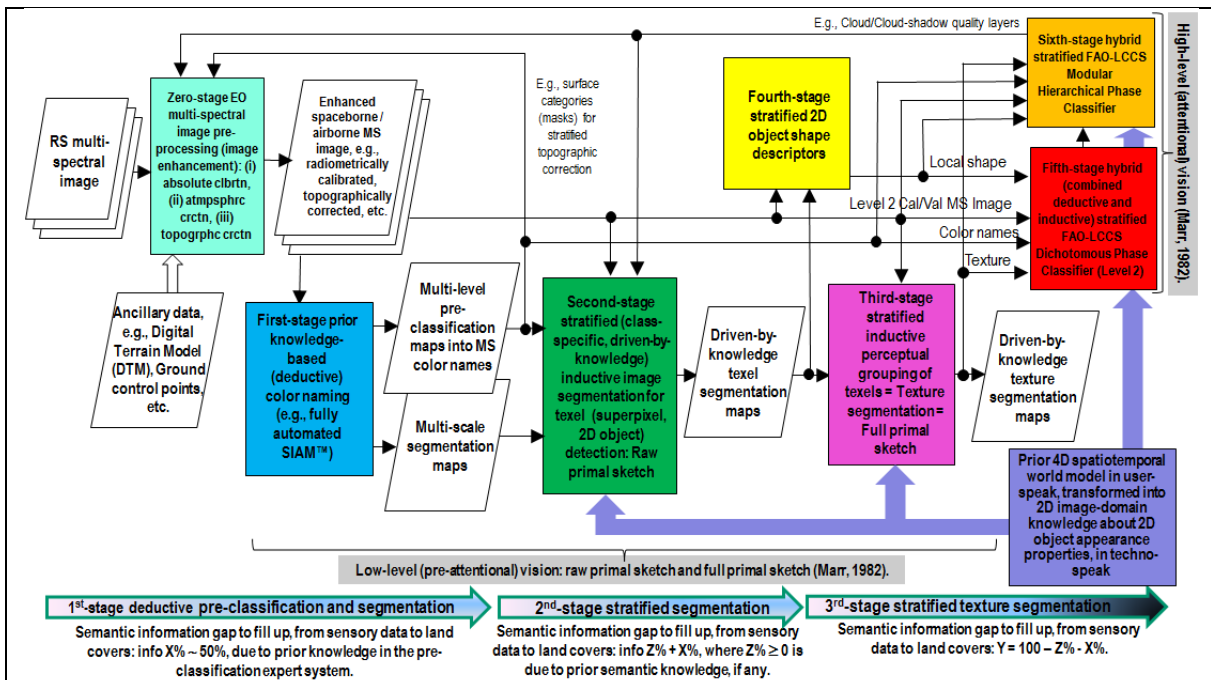


Fig. 9-13. Novel six-stage hybrid feedback image understanding system (IUS) architecture, proposed to the RS community in recent years. It includes Stage 0 (zero) for stratified image enhancement. According to experiments on EO image classification problems, approximately 50% of the semantic information gap from sensory data to land cover (LC) classes can be filled in by the first-stage expert system for automatic color naming [23], [24].

Tables and table captions in Chapter 9

RMSE, Tot. no. of segments, Mean area, 49 color labels, RGB image	Deductive (one-pass) RGBIAM, 49 + 1 color names	Inductive K-means VQ, learning phase in test image 1, k = 49, no. of iterations = 3.	Inductive K-means VQ, learning phase in test image 2, k = 49, no. of iterations = 3.	Inductive K-means VQ, learning phase in test image 3, k = 49, no. of iterations = 3.	Hybrid VQ: first-stage deductive RGBIAM and second-stage inductive k-means VQ, k = 49, no. of iterations = 3.	Mean of the variable across algorithms	StDev of the variable across algorithms
Test image 1, Human face, RW = 230, CL = 219.	R: 17.292217, G: 16.468367, B: 15.261817, Tot. no. of segments = 1008, Mean area (pixel) = 49.97	R: 12.978549, G: 8.689056, B: 7.933170, Tot. no. of segments = 2123, Mean area (pixel) = 23.73	R: 12.183979, G: 8.229305, B: 7.008564, Tot. no. of segments = 2949, Mean area (pixel) = 17.08	R: 11.104136, G: 8.186459, B: 6.905596, Tot. no. of segments = 3213, Mean area (pixel) = 15.68	R: 9.775062, G: 9.010200, B: 9.197877, Tot. no. of segments = 1662, Mean area (pixel) = 30.32	R: 12.67, G: 10.12, B: 9.26, Tot. no. of segments = 2191, Mean area (pixel) = 27.35	R: 2.85, G: 3.57, B: 3.48, Tot. no. of segments = 908.71, Mean area (pixel) = 13.92
Test image 2, Aerial optical, RW = 2744, CL = 4616.	R: 13.236089, G: 12.725385, B: 12.010231, Tot. no. of segments = 899416, Mean area (pixel) = 14.08	R: 11.771849, G: 10.646335, B: 10.953172, Tot. no. of segments = 1066650, Mean area (pixel) = 11.87	R: 6.356885, G: 6.009790, B: 5.960453, Tot. no. of segments = 3486527, Mean area (pixel) = 3.27	R: 10.565770, G: 9.723846, B: 9.675221, Tot. no. of segments = 1395591, Mean area (pixel) = 9.07	R: 5.825976, G: 5.208758, B: 5.193217, Tot. no. of segments = 2178923, Mean area (pixel) = 5.81	R: 9.55, B: 8.86, B: 8.76, Tot. no. of segments = 1805421.4, Mean area (pixel) = 8.89	R: 3.30, G: 3.17, B: 3.032, Tot. no. of segments = 1060685, Mean area (pixel) = 4.27
Test image 3, Satellite bi-temporal SAR, RW = 4480, CL = 5012.	R: 4.811451, G: 14.903615, B: 17.178292, Tot. no. of segments = 1859568, Mean area (pixel) = 12.07	R: 4.957717, G: 23.638669, B: 28.337016, Tot. no. of segments = 1192200, Mean area (pixel) = 18.83	R: 4.839931, G: 23.269672, B: 27.527727, Tot. no. of segments = 1379675, Mean area (pixel) = 16.27	R: 13.713457, G: 10.502449, B: 11.288297, Tot. no. of segments = 9713383, Mean area (pixel) = 2.31	R: 2.278091, G: 10.321090, B: 12.321693, Tot. no. of segments = 3475502, Mean area (pixel) = 6.46	R: 6.12, G: 16.58, B: 19.33, Tot. no. of segments = 3524066, Mean area (pixel) = 11.19	R: 4.39, G: 6.58, B: 8.17, Tot. no. of segments = 3574793, Mean area (pixel) = 6.82
<b>Total RMSE (to be ↓), Total no. of segments (to be ↓), Total Mean area (to be ↑)</b>	R = 35.339757, G = 44.097367, B = 44.45034, R+G+B = 123.89, Tot. no. of segments = 2759992, Tot. Mean area = <b>76.12</b>	R = 29.708115, G = 42.97406, B = 47.223358, R+G+B = 119.90, Tot. no. of segments = <b>2260973</b> , Tot. Mean area = 54.43	R = 23.380795, G = 37.508767, B = 40.496744, R+G+B = 101.38, Tot. no. of segments = 4869151, Tot. Mean area = 36.99	R = 35.383363, G = 28.434795, B = 27.019478, R+G+B = 91.66, Tot. no. of segments = 11112187, Tot. Mean area = 27.06	<b>R = 17.879129, G = 24.540048, B = 26.712787, R+G+B = 69.13</b> , Tot. no. of segments = 5656087, Tot. Mean area = 42.59		

Table 9-1. Quantitative comparison of three different realizations of the RGB image QNQ transform, shown in Fig. 9-4, whose block 2 for VQ is implemented as: (i) the deductive RGBIAM's static decision tree for RGB cube partitioning, (ii) the traditional inductive  $k$ -means VQ algorithm, where parameters  $k$  and maximum number of iterations must be user-defined, while the  $k$  vector centroids are initialized by random input data sampling [35], and (iii) the RGBIAM expert system initializes the inductive  $k$ -means VQ algorithm. Similar to [54], a three-fold cross-validation is adopted to provide a predictive estimate of the quantization error of the inductive  $k$ -means VQ algorithm. One-of-three datasets is employed for training (shown as a gray cell) and the remaining two for testing. In the last row, where pooling of each error indicator or quality index collected across datasets (rows) for each method (column) is accomplished by a sum, best overall results are shown in bold.

RMSE, Tot. no. of segments, Mean area, 49 color labels, RGB image	Deductive (one-pass) RGBIAM, 49 + 1 color names	Inductive K-means VQ, learning phase in test image 1, k = 49, no. of iterations = 3.	Inductive K-means VQ, learning phase in test image 2, k = 49, no. of iterations = 3.	Inductive K-means VQ, learning phase in test image 3, k = 49, no. of iterations = 3.	Hybrid VQ: first-stage deductive RGBIAM and second-stage inductive k-means VQ, k = 49, no. of iterations = 3.	Mean of the standardized variable across algorithms	StDev of the standardized variable across algorithms
Test image 1, Human face, RW = 230, CL = 219.	R: 1.62, G: 1.78, B: 1.72, Tot. no. of segments = -1.30, Mean area (pixel) = 1.62	R: 0.11, G: -0.40, B: -0.38, Tot. no. of segments = -0.07, Mean area (pixel) = -0.26	R: -0.17, G: -0.53, B: -0.65, Tot. no. of segments = 0.83, Mean area (pixel) = -0.74	R: -0.55, G: -0.54, B: -0.68, Tot. no. of segments = 1.12, Mean area (pixel) = -0.84	R: -1.01, G: -0.31, B: -0.02, Tot. no. of segments = -0.58, Mean area (pixel) = 0.21	R: 0, G: 0, B: 0, Tot. no. of segments = -0, Mean area (pixel) = 0	R: 1, G: 1, B: 1, Tot. no. of segments = 1, Mean area (pixel) = 1
Test image 2, Aerial optical, RW = 2744, CL = 4616.	R: 1.11, G: 1.22, B: 1.072, Tot. no. of segments = -0.85, Mean area (pixel) = 1.21	R: 0.67, G: 0.56, B: 0.72, Tot. no. of segments = -0.69, Mean area (pixel) = 0.69	R: -0.97, G: -0.90, B: -0.92, Tot. no. of segments = 1.58, Mean area (pixel) = 1.23	R: 0.31, G: 0.27, B: 0.30, Tot. no. of segments = -0.39, Mean area (pixel) = 0.042	R: -1.13, G: -1.15, B: -1.18, Tot. no. of segments = 0.35, Mean area (pixel) = -0.72	R: 0, G: 0, B: 0, Tot. no. of segments = -0, Mean area (pixel) = 0	R: 1, G: 1, B: 1, Tot. no. of segments = 1, Mean area (pixel) = 1



Test image 3, Satellite bi-temporal SAR, RW = 4480, CL = 5012.	R: -0.20, G: -0.25, B: 0.26, Tot. no. of segments = -0.46, Mean area (pixel) = 0.13	R: -0.26, G: 1.08, B: 1.10 Tot. no of segments = -0.65, Mean area (pixel) = 1.12	R: -0.29, G: 1.02, B: 1.0, Tot. no of segments = -0.60, Mean area (pixel) = 0.74	R: 1.73, G: -0.91, B: -0.98, Tot. no. of segments = 1.73, Mean area (pixel) = 1.30	R: -0.87, G: -0.94, B: -0.86, Tot. no of segments = -0.01, Mean area (pixel) = -0.69	R: 0, G: 0, B: 0, Tot. no of segments = -0, Mean area (pixel) = 0	R: 1, G: 1, B: 1, Tot. no of segments = 1, Mean area (pixel) = 1
Total RMSE (to be ↓), Total no. of segments (to be ↓), Total Mean area (to be ↑)	R = 2.44, G = 2.75, B = 2.53, <b>Tot. no. of segments = -2.62, Tot. Mean area = 2.97</b>	R = 0.52, G = 1.24, B = 1.44, Tot. no. of segments = -1.42, Tot. Mean area = 1.56	R = -1.43, G = -0.40, B = -0.57, Tot. no. of segments = 1.82, Tot. Mean area = -1.22	R = 1.49, G = -1.18, B = -1.36, Tot. no. of segments = 2.47, Tot. Mean area = -2.10	<b>R = -3.02, G = -2.40, B = -2.05</b> , Tot. no. of segments = -0.24, Tot. Mean area = -1.20		

Table 9-2. Z-scores extracted from Table 9-1. In the last row, where pooling of each error indicator or quality index collected across datasets (rows) for each method (column) is accomplished by a sum, best overall results are shown in bold.

Test image, Bands = B = 3	Rows, RW	Columns, CL	Image size in pixels, N	Computation time in seconds (including I/O operations)
Terrestrial, optical	219	230	50370	1
Airborne, optical	2744	4616	12666304	9
Spaceborne, SAR	4480	5012	22453760	19

Algorithm complexity analysis

Processing time (sec)

Number of pixels, with Number of spectral channels = 3

$y = 8E-07x + 0.341$   
 $R^2 = 0.9814$

Table 9-3. Computation time, including data processing and I/O operations, of the implemented RGB image QNQ transform employing the RGBIAM expert system for VQ. The computational complexity of the RGB image QNQ transform is linear in the image size. Irrespective of I/O operations, computation time complies well with the linear-time hypothesis.

## 10 Manuscript 7 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01941): Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis

### Motivation and Contributions to the Dissertation

In the convergence-of-evidence approach to computer vision (CV) proposed in Chapter 3 (Technical report 1), planar shape indexes are a source of spatial non-topological information in the (2D) image-domain, specifically, they are a source of spatial unit  $x$ -specific geometric information, where spatial unit  $x$  is either (0D) pixel, (1D) line or (2D) polygon. In the present Chapter 10 (Manuscript 7) an original minimally dependent and maximally informative (mDMI) set of planar shape indexes is presented and discussed to be considered eligible for use in the Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2). To guarantee that numeric variables are not dependent, i.e., to avoid cause-effect relationships between numeric variable pairs, the traditional Pearson's cross-correlation test is omitted because it is well known that, first, cross-correlation is sensitive to linear relationships exclusively and, second, that correlation does not imply causation. A Pearson's chi-square test of statistical independence is applied instead, whose inputs are two categorical variables. Hence, each numeric variable must be transformed into a categorical variable whose bins (quantization levels) are equiprobable and whose cardinality is appropriate according to a heuristic statistical criterion. If a feature-pair passes the Pearson's chi-square test of statistical independence, it can be hierarchically submitted to a Spearman's rank cross-correlation coefficient, showing whether two ranked random variables are monotonically increasing or decreasing, independent of linear relationships. Only if both tests are passed in comparison with any other feature a numeric variable is eligible for consideration in the mDMI set of features.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 10 (Manuscript 7) are color filled.

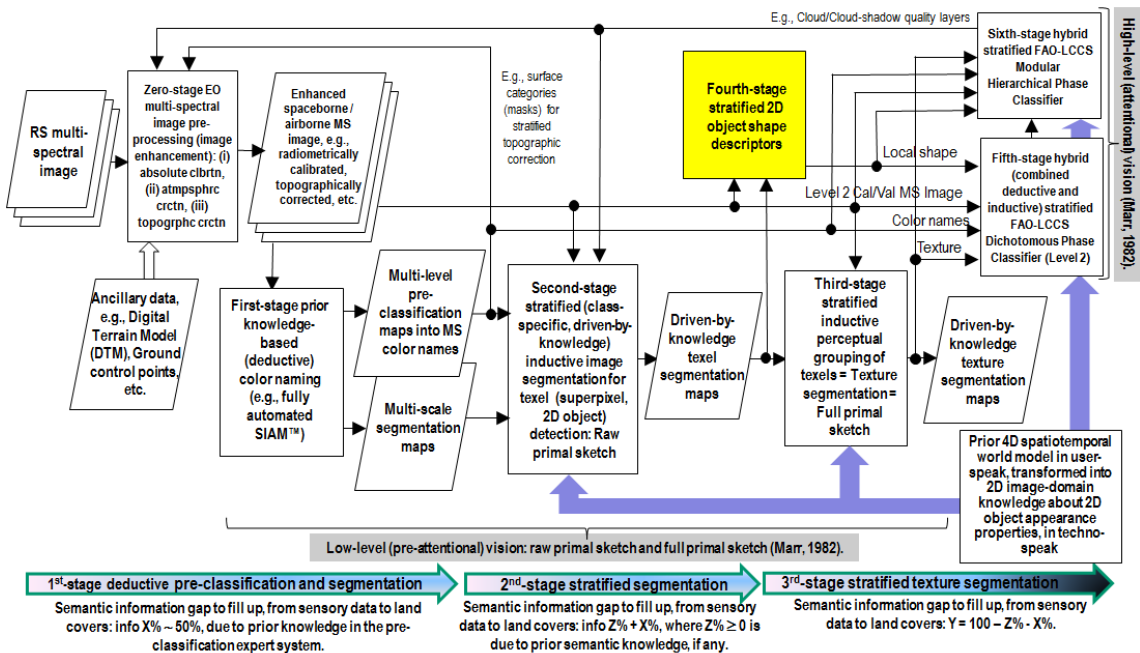


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted





by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 10 (Manuscript 7).



## Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis

Andrea Baraldi and João V. B. Soares

A. Baraldi was with the Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA. He is now with the Dept. of Agricultural and Food Sciences, University of Naples Federico II, Portici (NA), Italy. e-mail: andrea6311@gmail.com.

J. V. B. Soares is with the Department of Computer Science, University of Maryland, College Park, MD 20740 USA. e-mail: [joao@cs.umd.edu](mailto:joao@cs.umd.edu).

### Abstract

In recent years two sets of planar (two-dimensional, 2D) shape attributes, provided with an intuitive physical meaning, were proposed to the remote sensing community by, respectively, Nagao & Matsuyama and Shackelford & Davis in their seminal works on the increasingly popular geographic object-based image analysis (GEOBIA) paradigm. These two published sets of “simple” geometric features were selected as initial conditions by the present research and technological development software project, whose multi-objective goal was to accomplish: (i) a minimally dependent and maximally informative design (knowledge/information representation) of a general-purpose, user- and application-independent dictionary of 2D shape terms provided with a physical meaning intuitive to understand by human end users and (ii) an effective (accurate, scale-invariant, easy to use) and efficient implementation of 2D shape descriptors. To comply with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines, the proposed suite of geometric functions is validated by means of a novel quantitative quality assurance (Q<sup>2</sup>A) policy, centered on inter-feature dependence (causality) assessment. This innovative multivariate feature validation strategy is alternative to traditional feature selection procedures based on either inductive data learning classification accuracy estimation, which is inherently case-specific, or cross-correlation estimation, because statistical cross-correlation does not imply causation. The project deliverable is an original general-purpose software suite of seven validated off-the-shelf (ready-for-use) 2D shape descriptors intuitive to use. Alternative to existing commercial or open source software libraries of tens of planar shape functions whose informativeness remains unknown, it is eligible for use in (GE)OBIA systems in operating mode, expected to mimic human reasoning based on a convergence-of-evidence approach.

### Index terms

Causality, differential morphological profile, geometric (shape and orientation) features, human vision, image segmentation, object-based image analysis, Open Geospatial Consortium, Pearson’s chi-square test for statistical independence, Pearson’s cross-correlation coefficient, planar object, quality indicator, spatial autocorrelation, Spearman’s rank cross-correlation coefficient, statistical independence.

### 10.1 Introduction

Recent years have seen huge amounts of digital visual (two-dimensional, 2D, planar) data generated, transmitted, stored, retrieved and made accessible to a wide public of scientists and general users. Stemming *from* the ever increasing quality, quantity and accessibility of digital imagery, there is an urgent demand for effective and efficient computational vision tools, whose application domain ranges from low-level visual data representation and description to high-level image understanding (classification, i.e., three-dimensional (3D) scene reconstruction from (2D) imagery) [1], [2], [3], encompassing content-based image storage and retrieval (CBISR) system development [4], [5], [6], [7].

In human vision, 2D shape properties play a pivotal role [8], [9], [10], in combination with color [3], textural properties [3], [11], [12], [13], [14] and inter-object spatial relationships [15], [16], [17], [18]. When a 3D real-world object is projected onto a 2D image plane, one dimension of the 3D object information is lost. Due to dimensionality reduction [2], the 2D shape extracted from an image only partially represents the projected 3D object, i.e., 2D shapes can be affected by (self-)occlusion phenomena. In digital images, to make the 2D shape analysis and recognition problem even more complex, shapes of image-objects are also affected by spatial aliasing, due to the spatial resolution of the imaging sensor, and can be corrupted by photometric noise. In computational vision, effective and efficient solutions to the inherently difficult



geometric problem of quantitative 2D shape analysis and recognition would impact many scientific domains [19], such as cognitive science, computer vision, which includes medical imaging and remote sensing (RS) imaging, computer aided design, molecular biology, geographic information systems, etc., see Fig. 10-1. For example, shape is one of the primary low-level image features investigated in existing CBISR systems, not yet available in operating mode to a general public [4], [5], [6], [7], [23].

In their survey of 2D shape descriptors developed by computer vision, computational geometry and computational morphology [9], [23], Zhang and Lu categorized as “simple” those planar object features provided with an intuitive physical meaning, see Fig. 10-2. Related to human perception, “simple” 2D shape descriptors are eligible for use in computer-based decision systems capable of mimicking human reasoning [24]. Computational geometry is a relatively new and flourishing discipline of computer science, coping with the systematic study of algorithms and data structures for 2D and 3D geometric problems [20], [25], [26], [27], [28], with a focus on exact algorithms that are asymptotically fast. Software projects that provide easy access to efficient and reliable 2D and 3D geometric algorithms and data structures (e.g., Delaunay tetrahedralization, Voronoi tessellation, surface and volume mesh generation, convex hull algorithms, 2D- and 3D-object skeletonization, etc.) in the form of a software library are the Computational Geometry Algorithms Library (CGAL) [27] and the Library of Efficient Data Types and Algorithms (LEDA) [28]. Unfortunately, software libraries such as CGAL and LEDA do not provide any suite of “simple” 2D shape descriptors.

Computational morphology is the study of form or structure as in the case of automatic recognition of shape by machines [26], [29]. In the context of morphological image analysis [29], software libraries of “simple” geometric attributes of planar objects are made available by, for example, the eCognition commercial software product [30], the Open Source Computer Vision Library (OpenCV) [31] (refer to Appendix 1) and the ENvironment for Visualizing Images (ENVI) EX 5.0 commercial software toolbox [32] (refer to Appendix 2). Consisting of around forty-five, fifteen and fourteen geometric descriptors respectively, these three collections of basic geometric terms are completely different from one another. This means they adopt different multivariate feature representation and description optimization criteria; above all, these optimization criteria are unknown to the scientific community to date. As a consequence, although employed on a regular basis by computer vision and RS scientists and practitioners, these suites of geometric functions are not provided with any known validation (*Val*) policy for quantitative quality assurance (Q<sup>2</sup>A), to be community-agreed upon, as recommended by the Quality Assurance Framework for Earth Observation (QA4EO) guidelines [33] (refer to Appendix 3).

To contribute to filling the information gap from sensory “big data” to operational, comprehensive and timely information products subject to a *Val* policy for Q<sup>2</sup>A [33], the present interdisciplinary research and technological development (RTD) software project pursues an original multi-objective optimization of a dictionary of basic 2D shape terms. To make this inherently ill-posed software project better conditioned for numerical treatment, it is subject to the following constraints.

(1) Model design, regarded as “anything, but coding” [34]. The 2D shape representation consists of a discrete and finite set of quantitative variables, equivalent to a multivariate geometric indicator. It is expected to be: (i) general-purpose, i.e., data-, user- and application-independent [35], (ii) minimally redundant and (iii) maximally relevant [36], [37]. In the machine learning literature, these two latter conditions are known by acronym mRMR [36], [37]. For the sake of clarity and effectiveness, the mRMR criteria are reformulated in the present study. To account for the well-known dictum that “correlation does not imply causation”, quantitative summary statistics of a complex target phenomenon are required to be minimally dependent and maximally informative, mDMI. The collective mDMI feature design optimization criteria are considered related to the problem solving principle known as Occam’s razor<sup>1</sup> in the machine learning community [38], [39]. (iv) Made of “simple” perceptually significant 2D shape features. Provided with a physical meaning intuitive to understand [9], [23], these geometric features are eligible for use in the increasingly popular object-based image analysis (OBIA) paradigm, which includes geographic OBIA (GEOBIA) as a special case [40]. (GE)OBIA systems are expected to mimic human reasoning [1], [2], based on a convergence-of-evidence approach, where multiple fuzzy (eventually, weak) sources of converging perceptual evidence are combined to infer (eventually strong) conjectures [1], [2], in accordance with the general principles of fuzzy logic [41] and fuzzy decision trees [24], [42]. (v) Subject to a *Val* policy for Q<sup>2</sup>A, in agreement with the QA4EO guidelines (refer to Appendix 3) [33].

(2) Model implementation. Each variable descriptor (extractor) is expected to be delivered in operating mode. To be considered operational (off-the-shelf, ready-for-use), each individual 2D shape descriptor must be optimized to jointly

<sup>1</sup> A problem solver should always prefer a simpler model to more complex models and this preference should be traded off against the extent to which these alternative models fit the data.



score “high” in a set of community-agreed quantitative quality indicators (Q<sup>2</sup>Is) of operativeness (Q<sup>2</sup>IOs) [15], [16], subject to a *Val* policy for Q<sup>2</sup>A in compliance with the QA4EO guidelines (refer to Appendix 3) [33]. In the present work, for each geometric feature descriptor, adopted Q<sup>2</sup>IOs are: (a) accuracy, (b) efficiency and (c) robustness to changes in the input dataset, including invariance with respect to translations, rotations and scaling transformations [15], [16], [43].

The aforementioned list of project requirements is different, either totally or in part, from those found in related works on geometric feature design and implementation, such as [8], [23], [35], [43], [44].

The original contribution of the present RTD software project is twofold. Validated by a mandatory Q<sup>2</sup>A policy, the software project deliverable is a general-purpose suite of seven off-the-shelf 2D shape descriptors provided with an intuitive physical meaning. Pertaining to the fields of computational geometry [20], [25] and morphology [26], [29], “simple” 2D shape terms of certified quality can be integrated into software libraries of efficient and reliable computational geometry algorithms, such as CGAL [27]. Any general-purpose set of validated quantitative variables defined beforehand (equivalent to past knowledge) can be selected by a wide public of scientists and practitioners as a reliable initial condition (first choice), to be subject to further case-specific adjustments by means of data-, user- and application-dependent inductive data learning algorithms [39] for feature pruning [36], [37], feature mixture (e.g., principal component analysis [45]) or supervised data learning classification [38], [39]. For example, a popular statistical model for feature selection is stepwise regression. This is a greedy approach (globally suboptimal, but stepwise locally optimal) that adds the best feature (or deletes the worst feature) at each round according to an optimization criterion, such as mRMR [36], [37]. It is important to remind non-expert users that any inductive data learning algorithm is inherently ill-posed and requires *a priori* knowledge in addition to data to become better posed for numerical treatment [38], [39]. It means that no learning-from-examples algorithm (equivalent to phenotype in biological learning systems [46], [47]) should ever be confused with its initial conditions (equivalent to genotype in biological learning systems [46], [47]). In the words of genetic epistemology, for any (biological) cognitive system, “there is no absolute beginning” [46], [47]. By analogy, the application-independent dictionary of basic terms proposed in this study, equivalent to initial conditions driven from an *a priori* (deductive, top-down) knowledge base available in addition to data, should never be confused with any application-specific inductive (bottom-up) data learning algorithm, which explores the neighborhood of its initial conditions in a solution space based on evidence collected from the available dataset [47].

The second original contribution of the present study is the proposed *Val* strategy for Q<sup>2</sup>A of a set of quantitative random variables, whose dependence (causality [48]) must be minimized (mD). It comprises a three-level decision tree where the Pearson’s chi-square test for statistical independence [49], [50], [51], the Spearman’s rank cross-correlation coefficient (SRCC) [52], [53] and a local proof of the absence of monotonically increasing or decreasing pairwise feature relationships are estimated hierarchically. This multivariate feature Q<sup>2</sup>A policy is a viable alternative to statistical techniques for feature design and/or selection adopted by the mainstream RS community, including: (i) maximization of a multivariate classification accuracy [54], [55]. Any inductive data learning classification algorithm is inherently case-specific [38], [39], [56] and ignores other important Q<sup>2</sup>IOs of features descriptors, such as efficiency, whose relevance is high in other application domains, such as CBISR. (ii) Minimization of redundancy (mR), where redundancy is intended as inter-variable cross-correlation. Traditionally, cross-correlation is implemented by the Pearson’s cross-correlation coefficient (PCC) [35], sensitive to statistical linear relationships (collinearities). It is well known that “correlation does not imply causation” [52], [53]. It is also true that causation does not imply cross-correlation. For example, a non-linear inter-variable dependence (causal function) can cause PCC to be zero<sup>2</sup>. The hierarchical Q<sup>2</sup>A protocol for mD of a feature set proposed in this study can be applied to any multivariate feature analysis problem. For example, it could be adopted by software developers of the eCognition commercial software product [30], the OpenCV library [31] and the ENVI EX 5.0 commercial software toolbox [32], whose multi-objective optimization criteria for 2D shape index representation and implementation remain unknown, to comply with the *Val* requirements of the QA4EO guidelines.

The rest of this paper is organized as follows. Chapter 10.2 reviews the problem terminology. Related works are summarized in Chapter 10.3. Materials and methods are described in Chapter 10.4 and Chapter 10.5 respectively. Experimental results are presented in Chapter 10.6 and discussed in Chapter 10.7. Conclusions are proposed in Chapter 10.7I. For comparison purposes and to make the paper self-contained, Appendix 1 and Appendix 2 present, respectively, the list of geometric functions implemented in OpenCV [31] and in the ENVI 5.0 commercial software toolbox [32]. Provided with a relevant survey value, Appendix 3 summarizes the *Cal/Val* requirements of the QA4EO guidelines [33].

<sup>2</sup> For example, it is easy to prove that  $PCC(x,y) = \text{cov}(x,y) / (\sigma(x)\sigma(y)) = (E[xy] - E[x]E[y]) / \sigma(x)\sigma(y)$  is zero if  $y = x^2$  with  $x$  in  $[-1, 1]$ .





## 10.2 Terminology

Our investigation focuses on 2D shape descriptors exclusively. In this context, to avoid possible ambiguities in terms, let us introduce some terminology first. According to the Open Geospatial Consortium (OGC) nomenclature [57], a segmentation map is defined as a plane (image) partition consisting of mutually exclusive and totally exhaustive *plane entities*, where each part is identified by an integer value. As a consequence, a segmentation map is a multi-level image [42], where the number of levels equals the number of plane entities. Planar entities, termed (2D) *geometric objects* or *geometric primitives* in the OGC dictionary [57], are typically called (2D) tokens, segments, regions, patches or image-objects in the RS and computer vision literature [40], [42]. In the OGC terminology [57], the base *Geometry* class of geometric objects has four subclasses: (0D) *Point* (representing a single location in coordinate space; the boundary of a Point is the empty set), (1D) *Curve* (an open planar curve [43], defined as a connected sequence of Points), (2D) *Surface* (associated with an “exterior boundary” and zero or more “interior boundaries”) and *GeometricCollection*, which combines entities of the former subclasses. Examples of *atomic geometric types* of a geometric object, identified by an integer code, are: (0D) *Point* (code 1), (1D) *LineString* (code 2), (2D) *Polygon* (code 3), *MultiPoint* (code 4), *MultiLineString* (code 5), *MultiPolygon* (code 6, whose Polygon elements cannot intersect, but may touch), etc.

Spatial attributes of plane entities comprise positional and shape attributes, if any. In existing commercial or open source software libraries developed for computer vision and/or RS image processing (enhancement) and understanding applications, geometric attributes are extracted from image-objects whose atomic geometric type is either (1D) *LineString* (code 2) or (2D) *Polygon* (code 3), or, vice versa, from open or closed image contours. Noteworthy, image contour detection is the dual problem of image segmentation [2] and these are both inherently ill-posed problems [11] in the Hadamard sense [58], whose solution does not exist or is not unique or, if it exists, it is not robust to small changes in the input dataset.

## 10.3 Related Works

The taxonomy of 2D shape features proposed by Zhang and Lu [9], [23] comprises three dichotomous levels of feature categories, summarized as follows, see Fig. 10-2.

- Spatial and transformed domain representations. Some of the former, called “simple” shape descriptors [35], hold an intuitive physical meaning, like convexity, compactness, elongatedness, etc., such that their behavior can be intuitively predicted [43]. Other geometric descriptors in the spatial domain are not intuitive to interpret, but have the desirable properties of being invariant under translation, scaling and rotation, like the either area- or contour-based Hu geometric moments [59]. Traditionally, geometric operators in the spatial domain suffer from two main drawbacks: noise sensitivity and high dimensionality. These problems can be solved by analyzing shape in a transformed domain, like the spectral domain. Examples of spectral descriptors are the Fourier descriptor (FD) and the wavelet descriptor (WD) [8], [9], [10].
- Contour-based and region-based representations. Among geometric operators in the spatial domain, traditionally considered affected by high sensitivity to noise and high dimensionality, area-based descriptors are considered more robust, i.e. less sensitive to noise or shape deformations, while boundary-based descriptors are considered more sensitive to changes in the input dataset [23], [43].
- Continuous (global) as opposed to discrete (structural) representations. In the former, a planar shape is represented as a whole, such that the resulting representation is a quantitative feature vector. In the resulting multi-dimensional shape space [11], different shapes correspond to different points in this space and a quantification of shapes differences is accomplished using a metric distance between the acquired feature vectors, e.g., Hamming distance, Hausdorff distance, comparing skeletons and support vector machine pattern matching [60]. The latter represents a shape by sections, or primitives. For example, a shape boundary can be broken down into line segments by polygon approximation [9], [23]. In this case, the similarity measure between two shapes can be estimated by string matching or graph matching.

Peura and Iivarinen advocated the use of semantically “simple” shape descriptors whenever possible [35]. Their heuristic multi-objective feature representation and description optimization criteria are verbally, rather than quantitatively, expressed as follows. The feature set should be: (i) simple, intended as compact. We interpret this expression as minimally redundant, mR, because Peura and Iivarinen also stated that “some correlation between descriptors is acceptable”, where inter-feature (cross-)correlation was estimated as the PCC [51], [52], [53]. (ii) Generally applicable. To be general-purpose, it must be maximally relevant (MR), in any application domain. In qualitative terms, Peura and Iivarinen wrote that “combining descriptors should introduce a new perspective”. Finally, individual descriptors should be: (iii) computationally efficient and (iv) intuitive to understand, i.e., “each descriptor should be semantically simple”. For Q<sup>2</sup>A of a feature set of five “simple” shape descriptors, namely, convexity, ratio of principal axes, compactness, circular variance



and elliptic variance, Peura and Iivarinen selected a test set of 79 planar shape instances and estimated the PCC value for each feature pair.

In agreement with Peura and Iivarinen [35], Zhang and Lu [9], [23] observed that semantically “simple” shape descriptors are not suitable as standalone descriptors, but a combination of descriptors is necessary in order to accurately describe shapes. This domain-specific statement can be regarded as common knowledge to cope with the *non-injective property of any summary (gross) statistic* or  $Q^2I$ , which implies that no “universal”  $Q^2I$  can exist, because two different instantiations of the same target phenomenon can feature the same summary statistic [61]. Largely overlooked in common practice, the non-injective property of  $Q^2I$ s is inconsistent with a traditional search for universal image quality indexes (UIQIs), still ongoing by a relevant portion of the computer vision community [62], [63], [64], [65], [66]. By analogy, in economic studies, no economic univariate (scalar)  $Q^2I$ , such as the popular gross domestic product, should ever be considered *per se*, but in an mDMI combination with other  $Q^2I$ s [36], [37], such as the Gini index estimating the inequality of wealth, a pollution/environmental quality index, etc. [67]. To conclude, due to the non-injective property of summary statistics and in agreement with common sense, any quantitative investigation of a target complex phenomenon through summary statistics should avoid univariate analysis of one “universal” (scalar)  $Q^2I$ , which cannot exist in practice, in favor of a multivariate variable analysis, where an mDMI dictionary of  $Q^2I$ s must be designed and implemented in compliance with the Occam’s razor principle [38], [39]. For *Val* purposes, a multivariate feature  $Q^2A$  protocol for mDMI optimization can employ a multi-objective convergence-of-evidence approach [2], which is a key decision strategy in cognitive systems [24], [41], refer to Chapter 10.1.

In the domain of computational morphology, the eCognition commercial software toolbox [30], the OpenCV library [31] (refer to Appendix 1) and the ENVI EX 5.0 commercial software product [32] (refer to Appendix 2) include very different ensembles of “simple” geometric descriptors (refer to Chapter 10.1), whose atomic geometric type is either (1D) *LineString* (code 2) or (2D) *Polygon* (code 3, refer to Chapter 10.2), or, vice versa, open or closed image contours. Unfortunately, none of these available libraries of basic geometric functions was subject to any known *Val* strategy for feature selection/design and extraction/implementation)  $Q^2A$ .

In the RS literature, two compact sets of “simple” planar geometric attributes were proposed by, respectively, Nagao & Matsuyama [1], [2] and Shackelford & Davis [68], [69], whose seminal works pioneered the increasingly popular GEOBIA paradigm [40]. Encompassing both contour-based methods and region-based algorithms in the spatial domain, see Fig. 10-2, these two known sets of intuitive geometric properties are jointly reviewed hereafter.

- Oriented minimum enclosing rectangle (MER), parameterized as width (W), length (L) and orientation angle ( $\alpha$ ) of the L side, with  $L \geq W$ .
- Size, parameterized as area (A) in pixel unit.
- Elongatedness (*Elngtdnss*)  $\geq 1$ , estimated with several approaches, refer to [1], [2], [68] and [69] for further details.
- Compactness (*Cmpctns*) or circularity, estimated from the planar object’s area, A, and perimeter, P. Traditionally, it is estimated in a variety of different formulations [1], [35], [70], [71], [72], [73], e.g.,  $Cmpctns = 2\sqrt{A\pi}/P \in [0, 1]$  [35]. To increase confusion, some authors deal with noncompactness, e.g.,  $P^2/A \geq 1$  [1], [70], [73]. Most of these formulations are not scale invariant.
- Single-resolution (vice versa, single-scale) straightness of boundaries [1], [2], as an indicator of manmade Earth surface objects observed from space. This geometric feature is absent from open source and commercial software libraries of planar geometric functions, such as [30], [32] and [57].
- List of the object’s vertices after polygon skeletonization, where vertex attributes are position, angle and inter-endpoint distance [68], [69].
- Fuzzy rectangularity, particularly useful in RS images for building detection. An “approximately rectangular” image-object model of real-world buildings is defined as a 2D shape satisfying three properties: (i) about four endpoints with angles close to  $90^\circ$  and “large” separation (defined in this case larger than 5 m), (ii) about two or less endpoints with angles much larger than  $90^\circ$  and (iii) about two or less endpoints with angle much smaller than  $90^\circ$ . Because the attributes of an “approximately rectangular” image-object are imprecise and the shapes of buildings vary, fuzzy membership functions are adopted to measure how closely the endpoint angles and the line segment lengths between the endpoints match different physical model-based decision rules [68], [69].

Noteworthy, neither Nagao & Matsuyama [1], [2] nor Shackelford & Davis [68], [69] discussed their quantitative multi-objective optimization criteria for geometric feature design and implementation.

In the Moving Pictures Expert Group (MPEG)-7 Visual Standard proposed to search, identify, filter and browse



audiovisual content, several principles to measure a shape descriptor have been set: good classification/retrieval accuracy, compact features, general application domain, low computation complexity, retrieval performance robust to noisy data and hierarchical coarse-to-fine representation [8], [9], [10]. In MPEG-7, the either area- or contour-based multiscale generic Fourier descriptor (GFD) was considered to be the single best choice to accomplish the six principles set by MPEG-7 [8], [9], [10]. Although GFD has a physical interpretation and is particularly effective for quantitative shape matching and retrieval, it is not provided with an intuitive meaning. Hence, it is not suitable for implementation in computer vision systems expected to mimic human reasoning based on intuitive decision rules, such as GEOBIA systems [1], [2], [68], [69].

#### 10.4 Materials

To comply with standard evaluation criteria requiring a minimum of two real or realistic test datasets [74], three populations of geometric polygons were selected for testing purposes.

(i) One synthetic dataset of around thirty 2D shapes, shown in Fig. 10-4, provided an environment of controlled complexity suitable for human decision makers (HDMs) to test whether geometric operators satisfy quantitative theoretical expectations and qualitative human perception.

(ii) A first real-world dataset consisting of around 8000 geometric objects automatically detected in a very high resolution (VHR) satellite optical image by the Satellite Image Automatic Mapper™ (SIAM™), a prior knowledge-based vector quantizer capable of automatic MS image preliminary classification (pre-classification) and segmentation, proposed to the RS community in recent years [75], [76], [77]. Image-objects detected by SIAM are spectrally uniform, equivalent to textons (texture elements) detected at the raw primal sketch in low-level vision [3]. Spectrally uniform VHR image-objects are typically affected by irregular shapes and by the presence of inner holes. It means that, although all of these 2D segments are individually connected, only some of them are simply connected [75], [77].

(iii) A second real-world dataset consisting of around 300 images of individual leaves, acquired against a white background by a consumer-level digital camera, was automatically segmented into a binary map according to the near real-time segmentation algorithm proposed in [78], [79].

No segmentation quality assessment was pursued in this paper, since this task would go far beyond the goal of the present RTD project, reported in Chapter 10.1. In practice, segments were considered equivalent to an *a priori* knowledge which is, by definition, available beforehand in addition to data (observables, true-facts).

#### 10.5 Methods

In line with the words by Piaget (in cognitive systems “there is never an absolute beginning” [46], [47]), the proposed RTD software project moved from the legacy of past works by Nagao & Matsuyama [1], [2] and Shackelford & Davis [68], [69], summarized in Chapter 10.3 and selected as initial conditions, to design and implement an original software dictionary of “simple” 2D shape descriptors, where the atomic type of target geometric objects is either (1D) LineString (code 2) or (2D) Polygon (code 3), refer to Chapter 10.2, subject to multiple objectives to be jointly optimized in agreement with the Occam’s razor, refer to Chapter 10.1.

A formal analysis of multi-objective problems, where many possible courses of action are competing for attention, is due to the Italian civil engineer, economist and sociologist Vilfredo Pareto [80]. In his terminology, given a set of  $Q^2$ Is to be maximized, called Pareto dimensions, the so-called Pareto efficient frontier (PEF) is the set of choices that are Pareto efficient [80]. A multi-objective solution is called non-dominated, non-inferior, Pareto optimal or Pareto efficient if none of the objective functions can be improved in value without degrading some of the other objective values. By restricting attention to the set of choices that are Pareto efficient, an HDM can make subjective tradeoffs within the set of Pareto-efficient solutions, rather than considering the full range of every parameter. Without additional preference information by an HDM, which means without subjective (equivocal, qualitative) prior knowledge available in addition to quantitative (objective, unequivocal) data, all Pareto-efficient solutions must be considered equally good, i.e., in general, a multi-objective optimization problem is inherently ill-posed in the Hadamard sense [58]. When decision making is emphasized, i.e., when one single model solution must be chosen, an HDM has to select the most preferred Pareto optimal solution along the PEF according to his/her own preference (*information-as-data-interpretation*, also refer to Appendix 3).

Belonging to the class of inherently ill-posed multi-objective optimization problems subject to the Pareto’s formal analysis, our software design and implementation project is expected to admit no single “best” solution based on quantitative data analysis (*information-as-thing*, also refer to Appendix 3) exclusively.



### 10.5.1 Qualitative Maximization of Informativeness of a Feature Set

To maximize informativeness (MI) of a feature set in a general-purpose application domain as required by Chapter 10.1, Peura and Iivarinen wished that “combining descriptors should introduce a new perspective”. To our best knowledge, no quantitative case-independent feature design/selection strategy capable of accomplishing the MI constraint can be found in existing literature [8], [23], [35], [43], [44]. In computational morphology, we may distinguish between two phases of a shape detection process [26]. First, a multivariate shape parameterization (analysis) occurs, where a finite and discrete set of geometric features is selected and extracted. Second, any quantitative instantiation of a multivariate shape variable is mapped onto a discrete and finite set of mutually exclusive and totally exhaustive classes [81]. Typically, a multivariate data classification/retrieval accuracy index is adopted as an “overall” estimate of the effectiveness of the geometric feature representation and description sequence. However, as reported in Chapter 10.1, any inductive data learning classifier is inherently case-sensitive [35], [56]. It means that, to validate the MI property of a feature set in terms of classification accuracy on a general-purpose basis, a large variety of inductive classification algorithm implementations should be tested by many different users upon a large ensemble of different input datasets. In addition, the effectiveness (accuracy) of features employed to retrieve similar shapes from a designated database is not sufficient to guarantee the  $Q^2A$  of a shape selection and representation approach, because other  $Q^2IOs$ , such as efficiency, can be important in other application domains, such as CBISR [23].

To explain the ongoing lack of a standard community-agreed quantitative assessment of the MI property of a feature set in an application-independent scenario, we observe that the degree of informativeness of a set of features is an inherently equivocal (subjective, qualitative) cognitive problem (*information-as-data-interpretation*). As such, it does not pertain to the quantitative unequivocal domain of *information-as-thing* [4], refer to Appendix 3. If this consideration holds true, then this paper can provide no realistic  $Q^2I$  value to answer questions like: in general, does *Roundness* bear some useful information in addition to *Convexity*? Can *Rectangularity* be considered useful in the computer vision and RS common practice? Does feature *Straightness-of-boundaries*, proposed by Nagao & Matsuyama, but not implemented by any open source or commercial software library, such as [30], [32] and [57], “introduce any new perspective” [35]? These difficult (ill-posed) questions admit more than one (subjective) solution. Because estimating the collective degree of informativeness of a feature dictionary in a general-purpose application domain is an inherently equivocal *information-as-data-interpretation* problem [4], our realistic conclusion is that only “relative” (qualitative, subjective, equivocal) decisions, rather than “absolute” (quantitative, objective, unequivocal) decisions can be drawn by an HDM, based on his/her own preference or past knowledge, independent of and in addition to available observables (true-facts), if any [38], [39].

### 10.5.2 Quantitative Minimization of Dependence of a Feature Set

To account for the well-known dictum that “correlation does not imply causation”, the present work replaces the traditional multivariate feature optimization criterion of minimization of redundancy (mR), where redundancy is estimated as the PCC, which is equivalent to collinearity, with the criterion of minimizing an inter-feature degree of dependence (mD), where dependence means causality, refer to Chapter 10.1. The multivariate feature mD criterion can employ a goodness-of-fit (GOF) test well known in statistics, such as the Pearson’s chi-square test of independence between two categorical random variables [49], [50]. Unlike inherently ill-posed inductive data learning algorithms, which require prior knowledge in addition to data to become better posed for numerical treatment [38], [39], GOF tests employ no prior knowledge, but a reference model of a statistical distribution. On the one hand, GOF tests are, so to speak, “more inductive” (data-driven) and “less deductive” (prior knowledge-based, to become better posed) than inductive data learning algorithms for feature selection and data classification. On the other hand, due to the inherent subjectivity of any *information-as-data-interpretation* process (refer to Appendix 3), GOF test results, just like outcomes collected from any inductive data learning algorithm, must be carefully scrutinized for interpretation by human experts, as clearly acknowledged by statisticians in the following quote [60].

“Care must be taken not to over-interpret or over-rely on the findings of goodness-of-fit (GOF) tests. It is far too tempting ... to run GOF tests against a generous list of candidate distributions, pick the distribution with the ‘best’ goodness-of-fit statistic, and claim that the distribution that fit ‘best’ was not rejected at some specific level of significance. This practice is statistically incorrect and should be avoided. GOF tests have notoriously low power and are generally best for rejecting poor distribution fits rather than for identifying good fits... GOF tests may, at best, simply serve to confirm what the analyst has found through visual





inspection of the probability plots and other comparisons ... The human eye and brain are able to interpret and understand data anomalies far beyond the ability of any computer program or GOF test” [60].

To implement a multivariate feature mD criterion, we started from the formal analysis of causal models proposed by Pearl [48]. A causal model is a directed acyclic graph  $G$ , meaning a (directed) graph which has no loops in it. In a causal graph  $G$ , each vertex (node) has an associated random variable, all root vertices (i.e., vertices with no parents) are labelled by independent random variables [48] and each oriented arrow indicates the possibility of a direct causal influence between two random variables. In a causal graph  $G$ , causality means that for any particular variable  $X_j$ , the only random variables which directly influence the value of  $X_j$  are the parents of  $X_j$ , i.e., the collection of nodes  $X_{pa(j)}$  of random variables which are connected directly to  $X_j$  [48]. If two random variables  $X$  and  $Y$ , computed as deterministic functions (features) of an individual entity  $E$  (e.g., a plane entity) belonging to a sampled population  $P$  of entities, such that  $E \in P$ , where  $X = F1(E)$  and  $Y = F2(E)$ , are non-causal, i.e.,  $X$  does not belong to set  $pa(Y)$  and  $Y$  does not belong to set  $pa(X)$ , then it would be impossible to indicate which variable predicts or causes the other, see Fig. 10-3.

To reduce the risk that, in a target set of random variables, a causal relationship exists between any possible pair of random variables we adopted a hierarchical combination of sufficient but not necessary conditions for independence, collected at different levels of detail, from “global” statistics (properties of a population  $P$  of entities  $Es$ ) to “local” properties of individual entities,  $Es$ . The rationale of the adopted convergence-of-evidence approach is discussed below.

1) *First sufficient not necessary criterion for pairwise feature non-causality: Statistical independence*

The popular PCC, traditionally adopted to investigate pairwise feature redundancy considered equivalent to collinearity [35], [62], [63], [64], [65], [66] (refer to Chapter 10.1), is replaced by the Pearson’s chi-square test for random variable independence, well known in statistics [49], [50], [51], [52], [53]. By definition, two categorical random variables  $X$  and  $Y$  are statistically independent if the occurrence of one does not affect the probability of the other, i.e.,  $p(X \cdot Y) = p(X) \cdot p(Y)$ . It means that if two variables are statistically dependent, then knowledge of the level (category) of  $X$  can help predicting the level of variable  $Y$ , or vice versa. In other words, if two variables are statistically independent, it is impossible to indicate which variable predicts or causes the other. If there is a causal function of any degree between two random variables, then they are statistically dependent, although their PCC value can be zero, refer to footnote 2. On the other hand, the vice versa does not hold, i.e., if two variables are statistically dependent, they are not necessarily linked by a causal relationship. For example, let us consider two “simple” shape descriptors, such as *Compactness (Roundness)* and *Convexity*. Based on object-pair examples at a “local” scale of analysis (irrespective of “global” trends), it is clear that if *Roundness* is high then *Convexity* is also high. When *Roundness* is low, *Convexity* can be either high or low. Vice versa, when *Convexity* is high then *Roundness* can be either high or low. When *Convexity* is low, *Roundness* is also low. Hence, given a generic population of geometric objects, variables *Roundness* and *Convexity* are expected to perform as dependent variables to some (high) degree, because when the level of *Roundness* is known, it can help predicting the level of *Convexity* and vice versa. But (high) dependence does not mean there is a causal relationship between the two variables. To summarize, statistical independence is a sufficient, but not necessary condition for non-causality. If statistical dependence is “high”, then causality does not necessarily holds. When the Pearson’s chi-square test for independence is adopted to accomplish the mD feature constraint, the traditional project requirement specification that “some correlation between descriptors is tolerated” [35] becomes: “some statistical dependence between descriptors is tolerated”, if further investigations are conducted to avoid the risk of inter-variable causality. As such, statistical (in)dependence is far more informative (useful) than PCC for investigating inter-variable causality.

In the RS and computer vision common practice, the reason why the PCC is traditionally adopted for bivariate feature redundancy assessment, while the Pearson’s chi-square test for independence is almost never investigated, is that the former is input with two quantitative variables, which are very common in sensory data applications, while the latter requires as input two qualitative (categorical, nominal) variables [49], [50]. To transform a quantitative variable into a nominal one, it must be quantized (discretized) into a finite and discrete set of  $k$  levels (buckets, bins, categories, intervals), at the cost of a superior degree of quantitative data pre-processing. To provide a reasonable partition of the range of chance of the quantitative variable at hand while accounting for the variable distribution, these  $k$  buckets are recommended to be equiprobable and their total number is selected based on heuristic criteria [82]. A popular choice is:

$$k = 2 * N^{(2/5)}, \tag{1}$$

where  $N$  is the finite size of the sample dataset [82].



The proposed statistical test for independence consists of four steps [49]: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, refer to further Chapter 10.6, and (4) interpret results, refer to further Chapter 10.7.

- (1) State the null and alternative hypotheses. Null hypothesis,  $H_0$ : Categorical variables  $X$ , featuring  $R$  levels (like rows), and  $Y$ , featuring  $C$  levels (like columns), are independent. The alternative hypothesis,  $H_a$ , is that knowing the level of categorical variable  $X$  can help predicting the level of categorical variable  $Y$  and/or vice versa. If this inter-variable relationship (dependence) holds, it is not necessarily causal.
- (2) Statistical analysis plan. To describe how to use sample data to accept or reject the null hypothesis  $H_0$ , the statistical plan specifies:
  - a) The significance level  $\alpha$ , such that the degree of confidence is  $(1 - \alpha)$  [83]. Typically  $\alpha$  is selected equal to the probability value 0.05, hence the degree of confidence is 0.95. There is one critical value (CV) of the target statistical distribution that corresponds to a given significance level. The probability that values of the statistic distribution are equal or superior to the CV is equal to the significance level.
  - b) The test method. The well-known Pearson's chi-square test for independence is adopted. To be considered appropriate, it requires the following conditions to be met [49], [50]: (i) the sampling method is simple random sampling, (ii) each population is at least 10 times as large as its respective sample, (iii) the variables under study are both categorical, and (iv) the expected frequency count for each cell of the two-way contingency table is at least 5.
- (3) Analyze sample data. Applying the Pearson's chi-square test for independence to sample data requires to compute the degrees of freedom, the expected frequency counts and the chi-square test statistic output value. The P-value is defined as the probability of observing a sample statistic value equal or superior (i.e., not inferior) to the obtained test statistic value.
- (4) Interpret result. If the P-value is below the chosen level of significance  $\alpha$ , then  $H_0$  is dismissed at the level of confidence  $(1 - \alpha)$ . Similarly, given some significance level  $\alpha$  and its corresponding CV, if the test statistic value is higher than the CV, then  $H_0$  is dismissed at the level of confidence  $(1 - \alpha)$ , i.e., there is an  $\alpha$  probability that  $H_0$  is rejected by mistake [83] (refer to further Chapter 10.7).

For statistical data analysis, the popular Windows Excel software toolbox was adopted to quantize a quantitative random variable into  $k$  equiprobable bins, with  $k$  defined by Equation (1). Since the test statistic is a chi-square, we use the Windows Excel CHITEST() function to assess the P-value associated with the estimated chi-square test statistic value at the known value of degrees of freedom and significance level (refer to further Chapter 10.6).

### 2) *Second sufficient not necessary criterion for pairwise feature non-causality: Inter-feature statistical non-monotonicity (at the population level)*

If the statistical dependence between two random variables is "high", then we hierarchically consider a sufficient, although not necessary, condition for two geometric variables  $X = F_1(P)$  and  $Y = F_2(P)$ , generated from the same population  $P$ , to be non-dependent, such that neither function  $X = F_3(Y)$  nor function  $Y = F_4(X)$  exists, see Fig. 10-3, if those two random variables  $X$  and  $Y$  are neither monotonically increasing nor decreasing. In statistics, the SRCC is a widely adopted nonparametric measure of statistical dependence between two ranked variables, refer to Chapter 10.1. In greater detail,  $SRCC \in [-1, 1]$  assesses how well the relationship between two ranked quantitative variables can be described by a monotonically increasing or decreasing function. If there are no repeated data values, a perfect SRCC of +1 or -1 occurs when each of the variables is a monotonically increasing or decreasing function of the other, even if their relationship is not linear, which makes SRCC quite different from PCC [51]. Traditionally, a cross-correlation coefficient greater than 0.80 represents strong agreement, between 0.40 and 0.80 describes moderate agreement, and below 0.40 represents poor agreement [81]. To summarize, when the statistical dependence between two random variables is "high", then we require their SRCC value to be "not high", e.g.,  $SRCC < 0.8$ , to reduce the risk of inter-feature causality.

In our experiments, the popular Windows Excel software toolbox was adopted to generate ranked versions of two random variables to be input to the SRCC formulation.

### 3) *Third sufficient not necessary criterion for pairwise feature non-causality: Inter-feature local non-monotonicity at the individual level*

Last but not least, we must take into account the typical spatial nonstationarity of image statistics. In an image, "a localized image quality measurement can provide a spatially varying quality map of the image, which delivers more information about the quality of the image and may be useful in some applications" ([62], p. 7). It is like saying that (local)



counterexamples are known to be helpful because they quickly show that certain (general) conjectures, or ideas or theorems, are false. This allows scientists to save time and focus their efforts on ideas to produce provable theorems.

To try to avoid inter-feature causality, if the statistical dependence between two random variables X and Y is “high” at global (image-wide) scale and their SRCC value is also “high” at global scale, then we require to find “local” evidence that the relationship between the two random variables X and Y cannot be “locally” described by a monotonically increasing or decreasing function. Hence, if the statistical dependence between two planar shape features is “high”, but there is either global or local evidence that those two shape features are neither monotonically increasing nor monotonically decreasing, then those two features cannot be considered affected by a causal relationship, i.e., they cannot be considered the same feature, because it would be impossible to indicate which variable predicts or causes the other. In practice, the “local” absence of a monotonically increasing or decreasing function between two shape features  $X = F1(E)$ , e.g., roundness, and  $Y = F2(E)$ , e.g., convexity, of a planar entity E is verified if, in the target population of planar objects, at least one 3-tuple of entities, identified as R, E1 and E2, can be found such that, with respect to the reference instance, R, either one of the following two conditions (A) and (B) holds true.

Condition (A):

$$\{ [(F1(E1) > F1(R)) \text{ AND } (F2(E1) \geq F2(R))] \text{ AND} \\ \{ [(F1(E2) \geq F1(R)) \text{ AND } (F2(E2) < F2(R))] \text{ OR} \\ [(F1(E2) \leq F1(R)) \text{ AND } (F2(E2) > F2(R))] \} \}$$

OR Condition (B):

$$\{ [(F1(E1) < F1(R)) \text{ AND } (F2(E1) \leq F2(R))] \text{ AND} \\ \{ [(F1(E2) \leq F1(R)) \text{ AND } (F2(E2) > F2(R))] \text{ OR} \\ [(F1(E2) > F1(R)) \text{ AND } (F2(E2) \leq F2(R))] \} \}$$

## 10.6 Experimental Results

The objective of the experimental session is twofold. First, to comply with Chapter 10.5, the experimental session considers a pair of 2D geometric descriptors as minimally dependent (mD) if there is no multi-level evidence of their causal relationship. This occurs if: (i) their statistical dependence is not “high” (the null hypothesis, H0, is accepted, refer to Chapter 10.5.2.1), or (ii) their statistical dependence is “high” (H0 is rejected), but their SRCC value is not “high” ( $< 0.8$ , refer to Chapter 10.5.2.2), or (iii) their statistical dependence is “high” (H0 is rejected), their SRCC value is “high” ( $\geq 0.8$ ), but there is local evidence they are neither monotonically increasing nor decreasing (at any degree of relationship, whether linear or not), refer to Chapter 10.5.2.3.

Second, to comply with Chapter 10.1, the experimental session selects, for any proposed geometric attribute, the “best” descriptor, among possible alternative algorithms and implementations, that jointly maximizes the selected battery of Q<sup>2</sup>IOs, consisting of: (I) invariance with respect to translations, rotations and scaling transformations, (II) robustness to noisy data, (III) minimization of computation time and memory occupation.

### 10.6.1 Original Representation and Implementation of 2D Shape Features whose Q<sup>2</sup>IOs Must Score High

Inspired by the two sets of 2D shape descriptors proposed by Nagao & Matsuyama [1], [2] and Shackelford & Davis [68], [69], summarized in Chapter 10.3, the following original set of contour-based 2D shape attributes is proposed in the spatial domain.

- Angle (orientation) of the segment’s minimum enclosing rectangle [1], [84], not further discussed.



- Convexity, which is decreasing with the presence of holes ( $CnvxtyAndNoHole$ )  $\in [0, 1]$ , measured from the convex hull estimated from the outer boundary [42], refer to Chapter 10.6.1.1. It is absent from the two sets of attributes proposed in [1], [2] and [68], [69].
- Fuzzy rule-based rectangularity ( $FuzzyRuleBsdRctnglRty$ )  $\in [0, 1]$ , refer to Chapter 10.6.1.3. It employs a polygonal 2D shape representation alternative to the skeletonization algorithm adopted in [68], [69], refer to Chapter 10.6.1.2.
- Multiscale straightness of boundaries ( $MltScslStrghtnsOfBndrs$ )  $\in [0, 1]$ , alternative to the single-scale formulation proposed in [1]. Refer to Chapter 10.6.1.5.

Typically considered more robust to changes in the input dataset than contour-based attributes, area-based 2D shape attributes are estimated in the spatial domain according to the following original combination of features.

- Area (size), in pixel unit, not further discussed.
- Average per-segment characteristic scale ( $DMPmltScslChrctrstc$ )  $\geq 1$ , computed from the pixel-based morphological multiscale characteristic of the differential morphological profile (DMP) [85]. This characteristic scale value is proposed as a local spatial autocorrelation estimation, belonging to the class of local indicators of spatial association (LISA) [86]. This geometric feature is absent from the two sets of attributes proposed in [1], [2] and [68], [69]. See Fig. 10-4 and refer to Chapter 10.6.1.6.
- Scale invariant roundness (compactness, circularity), decreasing with the presence of holes ( $RndnssAndNoHole$ )  $\in [0, 1]$ . An original scale invariant compactness formulation is  $RndnssAndNoHole = ((4 \times \text{sqrt}(A)) / PL) \in [0, 1]$ , where PL is the 4-adjacency cross-aura measure estimated from the total perimeter, defined as the outer perimeter plus inner perimeter of holes, if any, see Fig. 10-5. To our best knowledge, this compactness equation is different from alternative formulations found in [1], [2], [35], [70], [71], [72]. Refer to Chapter 10.6.1.4.
- Elongatedness, increasing with the presence of holes ( $ElngtdnssAndNoHole$ )  $\geq 1$ . A new measure of  $ElngtdnssAndNoHole$  is presented to overcome operational limitations of the three formulations proposed by Nagao & Matsuyama, each at an increasing level of sophistication [1]. Refer to Chapter 10.6.1.7.
- (Combined) Simple connectivity, decreasing with the presence of holes ( $CombndSmplCnctvty$ )  $\in [0, 1]$ . This geometric feature is absent from the two sets of attributes proposed in [1], [2] and [68], [69]. It accounts for the dependence of region-based shape attributes on inner holes, whose presence may be an indicator of image noise. Refer to Chapter 10.6.1.8.

Noteworthy, planar shape operators belonging to range  $[0, 1]$  can be byte-coded, with a quantization error of  $(1. / 255) / 2 = 0.2\%$  when rounding to the nearest integer is adopted. For these operators, memory occupation is minimized.

Overall, seven 2D shape features are proposed, excluding area and orientation of the minimum enclosing rectangle, which are taken for granted. Hence, there are  $\binom{7}{2} = 7! / 2! (7-2)! = 21$  pairwise feature combinations to test for independence. The seven proposed descriptors and their related implementation issues are described hereafter.

#### 1) Area-based convexity sensitive to within-segment holes

A commonly used shape property is convexity, also known as convexity ratio or solidity. It is defined as follows:

$$CnvxtyAndNoHole = A / Aconvex, \quad (2)$$

where the area value,  $A$ , is defined as the number of pixels that belong to the region, excluding those belonging to holes, if any, while  $Aconvex$  is the area of the convex hull of the region, which by definition ignores (includes) inner holes. Since inequality  $A \leq Aconvex$  always holds true, then  $CnvxtyAndNoHole \in [0, 1]$ . The name of index  $CnvxtyAndNoHole$  is chosen to remind potential users that it takes on “high” values when the original segment is convex or close to being convex and, at the same time, it does not present holes. The first step in computing the area of the convex hull ( $Aconvex$ ) is to find a representation of the hull, e.g., refer to the solutions proposed by the CGAL [27]. To detect the convex hull, we began by tracing the outer boundary of the original region using the standard boundary tracing algorithm attributed to Moore [87], [88]. This results in a sequence of pixels describing a walk along the region’s boundary. From this sequence, the subset of the boundary pixels that form the set of vertices of the region’s convex hull was selected using the algorithm independently discovered by Melkman [89] and Tor and Middleditch [90], which runs in  $O(n)$  time, with  $n$  equal to the original number of boundary points.

After finding the vertices of the convex hull, the area  $Aconvex$  can be computed from the vertices using the algebraic surveyor’s (or shoelace) formula [91]. In digital imagery, this algebraic approach can lead to undesirable values of the  $CnvxtyAndNoHole$  variable. For example, a direct algebraic calculation of variable  $Aconvex$  for a region consisting of a straight line would be 0, whereas variable  $A$  would be the number of pixels in the line. Rather than trying to adapt the





surveyor's formula to cope with 2D signal aliasing effects, we decided to estimate *Aconvex* by means of a discretization procedure, where a binary discrete image is overlapped with the algebraic convex hull, such that discrete pixels are set to 1 if the major portion of their area overlaps with the convex hull and 0 otherwise. *Aconvex* is then estimated as the number of pixels whose values is equal to 1.

## 2) Polygonal representation of a 2D shape alternative to skeletonization

Shackelford & Davis adopted a simplified polygonal representation of image-objects, to which a set of fuzzy logic rules was applied to decide about their shapes, e.g., approximate rectangularity, refer to Chapter 10.3 [68]. They obtained the polygonal representation of a geometric object from the endpoints of the region's skeleton, computed via a morphological thinning skeletonization, e.g., refer to the solutions proposed by the CGAL [27]. In our tests, this approach turned out to be problematic when the endpoints do not give a good description of an object's boundary, see Fig. 10-6. Inaccuracy of the endpoint-based description arose independently of the adopted skeletonization algorithm. To avoid this problem, a different strategy was implemented. It is common to approximate a contour representing a polygon with another contour having fewer vertices. This simplified polygon representation is typically obtained by means of the Ramer-Douglas-Peucker (RDP) approximation [92], [93], e.g., refer to the routine `cvApproxPoly()` in the OpenCV software library in Appendix 1 [31]. In addition to giving a better approximation of the region's boundary, this algorithm is also faster to run than skeletonization. At each step of the RDP algorithm, two points are considered. The points determine a line segment, which can be thought of as a rough representation of some part of the polygon. The point in the original polygon which is the farthest from this line segment and which lies in between the two points that determine the segment is then found. If the distance from this point to the line segment is smaller than a tolerance value  $\epsilon$ , then the line segment is accepted as is. Otherwise, the problematic point is added to the representation, generating two new line segments which are recursively analyzed using the same procedure. We set the approximation tolerance  $\epsilon$  equal to the scale corresponding to the maximum value of *Straightness<sub>s</sub>*, refer to further Chapter 10.6.1.5. In other words,  $\epsilon = \text{argmax}_s \text{Straightness}_s$ , with  $s \in \{1, 2, 4, 8, 16, 32\}$ , where the observation scale parameter  $s$  in dyadic pixel units was used to compute the *Straightness<sub>s</sub>* measure. Fig. 10-6 shows toy regions for which the skeleton-based polygon does not provide a good representation, whereas simplification with the RDP algorithm does.

## 3) Contour-based fuzzy rectangularity

Several measures of a geometric object's rectangularity have been proposed in the RS and computer vision literature. One standard measure is the area of the region relative to that of its oriented minimum enclosing rectangle [1]. Rosin further improved this measure by developing robust rectangle fitting procedures [42]. Additionally, he also presented a measure of rectangularity defined using the difference in moments between the region and its best fitting rectangle. In place of a traditional area-based rectangularity index, Shackelford & Davis defined a measure of "approximate rectangularity" using a set of fuzzy rules [66]. These rules were applied to a polygonal representation obtained from a region's skeleton, as reviewed in Chapter 10.6.1.2. By using fuzzy rules to specify the expected number of vertices of a rectangle, their angles, and relative distance, the rectangularity measure becomes robust to a series of common shape variations, whether or not due to image noise. In the present study, the fuzzy rule-based rectangularity measure of Shackelford & Davis was selected due to its capability of modelling the within-class variance of rectangular shapes. Its implementation, identified as variable *FuzzyRuleBsdRctnglrty*  $\in [0, 1]$ , was subject to two important changes. First, the polygonal approximation (simplification) of the geometric object at hand to which the rule set is applied was not obtained by area-based skeletonization, but with the contour-based RDP algorithm [90], described in Chapter 10.6.1.2. Second, to cope with noisy shapes and/or highly irregular segment boundaries, several of the Shackelford & Davis fuzzy rules' free-parameters were relaxed. In practice, the same set of fuzzy rules, consisting of S-, Z- and  $\Pi$ -membership functions [91], proposed by Shackelford & Davis was adopted [66], [67]. However, their implementation was relaxed to become less strict, which produced results more in line with human photointerpretation of 2D shapes affected by noise. For example, in the original fuzzy rule set proposed by Shackelford & Davis, to determine if a within-segment angle is around  $90^\circ$ , a  $\Pi$ -membership function is implemented with a so-called bandwidth, equal to the difference between the minimum and maximum acceptable values, set to  $80^\circ$  and  $100^\circ$  respectively. In our implementation suitable for noisy data, these two minimum and maximum parameters were relaxed to  $60^\circ$  and  $120^\circ$ . To summarize, our new implementation is more robust to changes in input data and computationally more efficient than its original counterpart by Shackelford & Davis.



#### 4) Area-based roundness (compactness) sensitive to within-segment holes

An original *RndnssAndNoHole* (compactness) index formulation is proposed to be scale invariant and computationally efficient. A popular measure of a region's compactness, also called circularity or complexity, is  $A/P^2$ , with  $A$  denoting the region's area and  $P$  its perimeter [70]. The measure takes its maximum value of  $1/4\pi$  when the region in question is a circle without inner holes. This consideration motivates the definition of a measure of *Roundness*, or circularity, as  $4\pi A/P^2$ , which always lies in the range  $[0, 1]$  [35]. Additionally, some authors adopt a measure of noncompactness, e.g.,  $P^2/A$  [1], [70], [73]. In general, a segment's area,  $A$ , is defined as the number of pixels in the region, where pixels belonging to holes, if any, are excluded from the estimation. However, in common practice, compactness is estimated after holes have been filled in, i.e., it is assumed the region is simply connected. A decision must be made on how to compute the perimeter  $P$ , since different definitions of a region's perimeter exist when dealing with digital images [42], [73]. In our original formulation of geometric index compactness, the perimeter length  $PL$  is the computationally efficient 4-adjacency cross-aura measure (see Fig. 10-5 [61]) of the region's total boundary, where the total boundary takes into account contributions from holes, i.e., total boundary = external (outer) boundary + inner boundary (due to holes) [42]. Starting from these definitions of area,  $A$ , and perimeter length,  $PL$ , the proposed formulation of *Roundness* becomes

$$RndnssAndNoHole = (4 \times \text{sqrt}(A) / PL) \in [0, 1]. \quad (3)$$

Reflected in the way both  $A$  and  $PL$  are computed, this measure treats holes as intrinsic properties of the geometric object, rather than filling them up. It scores high for regions that are round and, at the same time, do not have holes. This behavior justifies its name, *RndnssAndNoHole*. It can be easily proved (by induction) that index *RndnssAndNoHole* is scale invariant. For example, for an isolated 1-pixel object,  $PL = 4$ , then  $RndnssAndNoHole = 4/4 = 1$  (maximum). For a 4-pixel square object,  $PL = 8$ , then  $RndnssAndNoHole = 4*2/8 = 1$ , etc. Although measures of compactness or roundness should be maximum for circles, Rosenfeld pointed out that, in digital images, depending on how the perimeter is measured, compactness measures could turn out to be larger for squares or octagons than for digitized circles [73]. Our measure is no exception and takes maximum values for squares, which have maximum area for a given fixed value of the 4-adjacency cross-aura measure.

#### 5) Contour-based multiscale straightness of boundaries

A measure of the straightness of a region's outer boundary is especially discriminative for the analysis of RS images of buildings in urban areas or crop fields in agricultural land. In general, manmade Earth surface structures tend to present straight boundaries, independently of whether their overall shape is simple or more complex. Following the procedure proposed by Nagao & Matsuyama [1], an estimation of the straightness of a region's boundary was implemented as follows. First, the boundary of the region was traced by means of the standard boundary tracing algorithm attributed to Moore [87], [88]. This results in a sequence of pixels describing a closed walk along the region's outer boundary, denoted as  $p_i$ , with  $i = 1, \dots, n$ , where  $p_1 = p_n$  (since the boundary is closed), so that the total number of boundary pixels is  $n - 1$ . For each pixel  $p_i$  on the boundary, the angle  $\Delta_i$  between the two lines connecting  $p_i$  with  $p_{i-s}$  and  $p_{i+s}$  was calculated, where values  $(i + s)$  and  $(i - s)$  are modulo  $n$ , i.e., they belong to range  $\{0, \dots, n-1\}$ . Variable  $s$  is referred to as the step size. A pixel  $i$  was counted as "straight" if  $|\Delta_i| \leq \alpha$ , for some angle threshold  $\alpha$ , which is given as a parameter to the method. Let  $n_s \in \{0, \dots, n-1\}$  denote the number of straight pixels in the boundary, measured using a step size  $s$ . Then, the straightness of boundary for step size  $s$  is  $Straightness_s = n_s / (n - 1) \in [0, 1]$ . In practice, the step size  $s$  acts as an observation scale. To deal with images of different resolution, as well as geometric objects of different size, it is important to choose an appropriate value of the step size  $s$ . Whereas Nagao & Matsuyama worked with a single step size, the following heuristic criterion was adopted to infer a step size adaptively. First, variable  $Straightness_s$  was computed for all values of  $s \in \{1, 2, 4, 8, 16, 32\}$ , where dyadic (power of 2) scale values in pixel unit were chosen for selection. Next, the final straightness measure was taken as the one with the maximum value, such that:

$$MltSclStrghtnsOfBndrs = \max_s \{Straightness_s \in [0, 1], \text{ with } s = 1, 2, 4, 8, 16, 32\}. \quad (4)$$

To summarize, due to its multi-scale implementation, the proposed *MltSclStrghtnsOfBndrs* estimator proved to be less sensitive to segmentation noise than its traditional single-scale counterpart, proposed by Nagao & Matsuyama [1].

#### 6) Pixel- and segment-based morphological multiscale characteristic

Spatial autocorrelation is of fundamental importance for human low-level vision, with special regard to the full primal sketch [3], which is responsible of image texture segmentation (perceptual grouping) [11], [12], [13], [14]. As a

consequence, spatial autocorrelation is or should be important in the computer vision and RS common practices [95], [96]. For example, in the RS literature spatial autocorrelation has been proposed as one quality term in a multi-objective Q<sup>2</sup>A function for image segmentation [97]. Last but not least, spatial autocorrelation is equivalent to the Tobler's first law of geography (TFLG) [98], considered the foundation of geostatistics and geographical sciences [99], [100], [101]. The TFLG states that, in the geospatial domain, "everything depends on everything else, but closer things more so" [98], although certain phenomena clearly constitute exceptions [102]. In practice, the TFLG implies that a spatial distance decay function exists, such that even though all geospatial observations have an influence on all other observations, after some distance threshold that influence can be neglected. Unfortunately, according to Lembo, "many geographers would say 'I don't understand spatial autocorrelation'; actually, they don't understand the mechanics, they do understand the concept" [91].

In their image segmentation algorithm, Pesaresi and Benediktsson computed a pixel-wise morphological multiscale characteristic, defined as the morphological scale where each pixel's DMP scores its maximum [85]. In our understanding, the morphological multiscale characteristic belongs to the class of local indicators of spatial association (LISA) [86], where spatial association is synonym of spatial autocorrelation. Unlike existing LISA proposed in the computer vision literature [86], the morphological multiscale characteristic formulated by Pesaresi and Benediktsson is simultaneously pixel-specific and edge sensitive [85]. In practice, it provides a per-pixel estimate of the local size (in pixel unit) of the image-object that pixel belongs to, without requiring spatial units (image-objects) to be detected beforehand by means of an image segmentation algorithm. Any inductive data learning algorithm for image segmentation (or, vice versa, image-contour detection) is inherently ill-posed [11]; hence, it is semi-automatic (where its degree of automation is monotonically decreasing with its number of system's free-parameters, to be user-defined based on heuristics) and site-specific (data-dependent) [56], refer to Chapter 10.5.1. Since it requires no image segmentation first stage, the multiscale morphological characteristic, adopted as a LISA, is expected to feature Q<sup>2</sup>IO values, including computation efficiency and degree of automation, higher than those of traditional LISA, such as Moran's I and Geary's C, reviewed hereafter for comparison purposes. The well-known Moran's I spatial autocorrelation index is either global (image-wide) or a LISA, which means a spatial autocorrelation value specific for each so-called "spatial unit" belonging to an image partition (segmentation) defined beforehand [86].

$$\begin{aligned} \text{Moran's } I &= \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{i,j}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2} \\ &= \frac{N}{2W} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2} \in [-1, 1], \end{aligned} \quad (5)$$

$$\text{Local Moran's } I_i = \frac{1}{\sum_{j=1}^N w_{i,j}} \frac{\sum_{j=1}^N w_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}, \quad i = 1, \dots, N, \quad (6)$$

where  $N$  is the number of spatial units indexed by  $i$  and  $j$ ,  $X$  is the variable of interest,  $\bar{X}$  is the mean of  $X$ ,  $w_{i,j}$  is a matrix of spatial weights monotonically non-increasing with distance between spatial units  $i$  and  $j$ . For example,  $w_{i,j} = 1$  if spatial units  $i$  and  $j$  are adjacent and  $w_{i,j} = 0$  otherwise. Coefficient  $W$  is the sum of all spatial weights  $w_{i,j}$ . Moran's I ranges from -1 (perfect inverse autocorrelation) to +1 (perfect positive autocorrelation). A zero value indicates a random spatial pattern. For example, in a black-and-white chessboard, where each square, either black or white, is a spatial unit and where the matrix of spatial weights consists of 1s in a 4-adjacency neighborhood and 0s otherwise, then the Moran's I value is equal to -1. According to Equation (6), there is one LISA estimate for each  $i$ -th spatial unit in the dataset. The sum of LISA values for all spatial units in an image region of interest is proportional to a corresponding global indicator of spatial association for that dataset. By "decomposing" a global autocorrelation result into its local parts, LISA values are very useful to uncover hidden, local patterns in data that the global statistics average over. For example, LISA values can detect when a significant global autocorrelation statistic at a given spatial lag may hide large spatial patches of no autocorrelation, or when an insignificant global autocorrelation statistic may hide patches of autocorrelation.

Moran's I is inversely related to another well-known global spatial autocorrelation index, Geary's C.

$$\begin{aligned} \text{Geary's } C &= \frac{N}{2 \sum_{i=1}^N \sum_{j=1}^N w_{i,j}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} (X_i - X_j)^2}{\sum_{i=1}^N (X_i - \bar{X})^2} \\ &= \frac{N}{2W} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} (X_i - X_j)^2}{\sum_{i=1}^N (X_i - \bar{X})^2} \in [0, 2), \end{aligned} \quad (7)$$

$$\text{Local Geary's } C_i = \frac{1}{2 \sum_{j=1}^N w_{i,j}} \frac{\sum_{j=1}^N w_{i,j} (X_i - X_j)^2}{\sum_{i=1}^N (X_i - \bar{X})^2}, i = 1, \dots, N. \quad (8)$$

The value of Geary's  $C$  lies between 0 and values  $\geq 2$ . Value 1 means no spatial autocorrelation. Values lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values higher than 1 illustrate increasing negative spatial autocorrelation. A simple reversed Geary's  $C$  formulation, such that

$$\text{Reversed Geary's } C = 1 - \min\{\text{Geary's } C, 2\}, \quad (9)$$

ranges from -1 (indicating inverse correlation) to +1 (perfect autocorrelation), like the Moran's  $I$  index.

Our understanding of the morphological multiscale characteristic as a viable alternative to existing LISA estimates, such as Equation (6) and Equation (8), agrees with some statements reported by Pesaresi and Benediktsson in their original paper [85]. An illustrative example of the behavior of the morphological multiscale characteristic is presented in Fig. 10-4 where, for each geometric object, the average of the characteristic scale over its pixels is adopted as an attribute describing the within-segment average of per-pixel LISA. The definition of a per-pixel DMP requires definitions of opening and closing by reconstruction for a grayscale image  $I$ . An *opening by reconstruction* is defined as  $\gamma^*_{\lambda} I = \text{Rec}(\varepsilon_{\lambda} I, I)$ , where  $\varepsilon_{\lambda} I$  denotes an erosion of  $I$  with a structuring element (SE) of size  $\lambda$  and  $\text{Rec}(\varepsilon_{\lambda} I, I)$  denotes the reconstruction by dilation of  $I$  from  $\varepsilon_{\lambda} I$ . For a detailed formal definition of the erosion and reconstruction operations, refer to [85], [103]. The traditional morphological opening of an image by a SE of size  $\lambda$  is used to filter out bright structures that are smaller than  $\lambda$ . The opening by reconstruction operator also filters out bright structures smaller than  $\lambda$ , but without affecting the fine-scale details of larger structures, since these are recovered in the reconstruction step. Analogously, a *closing by reconstruction* is defined as  $\phi^*_{\lambda} I = \overline{\text{Rec}}(\delta_{\lambda} I, I)$ , where  $\delta_{\lambda} I$  denotes a dilation of  $I$  with an SE of size  $\lambda$  and  $\overline{\text{Rec}}(\delta_{\lambda} I, I)$  denotes the reconstruction by erosion of  $I$  from  $\delta_{\lambda} I$ . Analogously to the opening by reconstruction, a closing by reconstruction filters out dark structures that are smaller than  $\lambda$ . Starting from these definitions, the opening profile of an image  $I$  is composed of a series of openings by reconstruction with a dyadic sequence of SE sizes,  $\lambda_i$  for  $i = 0, \dots, n$ , such that  $\lambda_0 = 0$ ,  $\lambda_1 = 1$ , and  $\lambda_i = 2^{i-1} + 1$  for  $i = 2, \dots, n$ . As an example of the sequence of sizes, if  $n = 4$  then the resulting sequence is  $\lambda_i = 0, 1, 3, 5, 9$ . Differently from Pesaresi and Benediktsson [85], we chose dyadic SE sizes  $\lambda_i$ , considered to be more practical (and biologically more plausible [104]), so that the resulting profile may be computed in a reasonable amount of time while being able to handle image structures of varying sizes. Given this sequence of spatial scales, the *opening profile* at pixel  $x$  is defined as the vector

$$\Pi\gamma(x) = \{\Pi\gamma_{\lambda_i} : \Pi\gamma_{\lambda_i} = \gamma^*_{\lambda_i}(x), i = 0, \dots, n\}. \quad (10)$$

Correspondingly, the *closing profile* at pixel  $x$  is

$$\Pi\phi(x) = \{\Pi\phi_{\lambda_i} : \Pi\phi_{\lambda_i} = \phi^*_{\lambda_i}(x), i = 0, \dots, n\}. \quad (11)$$

The per-pixel DMP records the rate of change in the opening and closing profiles. For each pixel, its DMP provides an estimate of the importance of structures of size  $\lambda$  to which the pixel might belong. The *derivative of the opening profile*  $\Delta\gamma(x)$  and the *derivative of the closing profile*  $\Delta\phi(x)$  are defined respectively as

$$\Delta\gamma(x) = \{\Delta\gamma_{\lambda_i} : \Delta\gamma_{\lambda_i} = \frac{|\Pi\gamma_{\lambda_i} - \Pi\gamma_{\lambda_{i-1}}|}{(\lambda_i - \lambda_{i-1})}, i = 1, \dots, n\}, \quad (12)$$





$$\Delta\phi(x) = \{\Delta\phi_{\lambda_i} : \Delta\phi_{\lambda_i} = \frac{|\Pi\phi_{\lambda_i} - \Pi\phi_{\lambda_{i-1}}|}{(\lambda_i - \lambda_{i-1})}, i = 1, \dots, n\}. \quad (13)$$

The complete DMP is obtained by concatenating the two DMP components expressed above. In order to define the *morphological multiscale characteristic* [85], the *multiscale opening characteristic* and the *multiscale closing characteristic* must be first defined. The *multiscale opening characteristic*  $\Phi\gamma(x)$  of an image  $I$  at pixel  $x$  is the SE size at which the opening DMP takes on the largest value,

$$\Phi\gamma(x) = \{\lambda : \Delta\gamma_{\lambda}(x) = \vee \Delta\gamma(x)\}, \quad (14)$$

where  $\vee$  denotes the supremum. Analogously, the *multiscale closing characteristic* is the SE size for which the closing DMP has its maximum value,

$$\Phi\phi(x) = \{\lambda : \Delta\phi_{\lambda}(x) = \vee \Delta\phi(x)\}. \quad (15)$$

The morphological multiscale characteristic  $\Phi(x)$  is chosen as the scale at which the DMP is maximum, whether taken from the opening or closing profile. It can be defined as

$$\Phi(x) = \begin{cases} \Phi\gamma(x) : \vee \Delta\gamma(x) > \vee \Delta\phi(x) \\ \Phi\phi(x) : \vee \Delta\gamma(x) < \vee \Delta\phi(x) \\ 0 : \vee \Delta\gamma(x) = \vee \Delta\phi(x) \end{cases}. \quad (16)$$

In the rest of this paper, variable  $\Phi(x)$  is identified as *DMPmltScIChrctrstc*. An illustrative example of applying this definition to all pixels of an image is presented in Fig. 10-4. Hence, *DMPmltScIChrctrstc*, which is computed pixel-wise, can be averaged per 2D shape.

A practical challenge in using the pixel-based morphological multiscale characteristic value is that computing openings and closings by reconstruction for large images can be prohibitively slow. In order to speed up this process, two techniques were selected from the existing literature: decomposable filters for dilation and erosion [105], and the downhill filter for reconstruction [106]. Hereafter, the sole case of opening by reconstruction is discussed, because its dual problem is closing by reconstruction for which the same observations hold. The initial erosion operation  $\varepsilon_{\lambda}I$ , that occurs before performing the reconstruction by dilation  $\text{Rec}(\varepsilon_{\lambda}I, I)$ , can be very fast depending on the structuring element that is used. Instead of using a disk, which seems like the natural choice, it is much faster to approximate the result that would be obtained from a disk by using a regular polygon [107]. In particular, erosion with a square structuring element is extremely fast, since it can be decomposed into single horizontal and vertical operations [105]. The purpose of the initial erosion operation is to eliminate bright structures that are smaller than the structuring element size. The reconstruction step that follows will guarantee that finer scale details of the remaining structures are reconstructed. In practice, the final results obtained from opening by reconstruction were found to be similar whether using a square or a disk for the initial erosion operation. To recapitulate, in the present study, the use of a square structuring element was preferred due to its superior implementation simplicity and lower computation time.

Due to computation time considerations, special attention must be given to the reconstruction by dilation operation,  $\text{Rec}(\varepsilon_{\lambda}I, I)$ . Vincent defined the notion of morphological reconstruction for grayscale images [108] as well as fast algorithms to speed up execution times considerably [103], [108]. Later, Robinson and Whelan presented the downhill filter for image reconstruction [106], with a precondition to be satisfied by input images in order to guarantee correctness. In the weak form, the precondition requires that, in the case of reconstruction by dilation, the marker image be everywhere less than or equal to the mask image. Fortunately, in the reconstruction by dilation  $\text{Rec}(\varepsilon_{\lambda}I, I)$ , this condition always holds true, since  $\varepsilon_{\lambda}I \leq I$ . The downhill filter computes the reconstruction in a single pass, guaranteeing a fast and consistent execution time. In experimental comparisons with the algorithms proposed by Vincent [103] over a range of input images, the downhill filter showed a consistent and oftentimes large speedup [106]. Hence, we adopted the downhill filter to compute the required reconstructions.

To summarize, the implemented morphological *DMPmltScIChrctrstc* function is proposed as a computationally efficient pixel-wise LISA sensitive to the presence of local edges. Unlike existing LISA estimates, such as the popular local Moran's I and Geary's C formulations, see Equation (6) and Equation (8), *DMPmltScIChrctrstc* requires no *a priori* partition (segmentation) of an image into spatial units (image-objects), which improves Q<sup>2</sup>IO values of the LISA estimator, including degree of automation and robustness to changes in the input image.



### 7) Area-based elongatedness sensitive to within-segment holes

A new measure of elongatedness is presented to overcome operational limitations of the three formulations proposed by Nagao & Matsuyama, each at an increasing level of sophistication [1]. Their most sophisticated solution is summarized below, to be compared with our new measure. To define a measure of elongatedness, Nagao & Matsuyama proposed to estimate the longest path along a region, together with the region's average width along that path [1]. First, holes of the region are filled in, if any, after which the region's skeleton is computed via thinning. Next, the longest path along the skeleton is detected. Additionally, for each point along the longest path, the local width of the region is computed and the average of these widths is taken. The elongatedness measure defined by Nagao & Matsuyama (NM) is

$$Elngt d n s s_{NM} = L_{NM} / W_{NM}, \quad (17)$$

where  $L_{NM}$  is the length of the longest path and  $W_{NM}$  is the average width along that longest path. A practical limitation of the  $Elngt d n s s_{NM}$  operator is that filling in the holes of a region before computing its elongatedness may not always be appropriate. Fig. 10-7 shows examples of segments of roads and rivers, extracted from a satellite image, where perception of elongatedness by a human expert scores "high", whereas index  $Elngt d n s s_{NM}$  does not, due to preliminary within-segment hole filling required by Nagao & Matsuyama's procedure. This limitation justifies the introduction of an improved measure of elongatedness, different from Nagao & Matsuyama's, discussed hereafter. Within-segment holes, if any, are not filled in beforehand. After computing the region's complete skeleton (without removing holes), our novel measure of elongatedness is defined as

$$Elngt d n s s_{AndNoHole} = L / W \geq 1, \quad (18)$$

where  $L$  is a measure of the length of the complete skeleton (without removing holes), i.e.,  $L$  is the total number of skeleton pixels, while  $W$  is the region's average width across each pixel along the skeleton. The novel measure of  $Elngt d n s s_{AndNoHole}$  captures a different notion of elongatedness from index  $Elngt d n s s_{NM}$ . First, in the former, the length measure  $L$  is estimated as the number of pixels in the whole skeleton, as opposed to just its longest path. This way,  $L$  takes into consideration every part of a region. This difference is illustrated in Fig. 10-8. Second, by not filling in possible holes, variable  $Elngt d n s s_{AndNoHole}$  treats holes as intrinsic characteristics of image-objects, as opposed to ignoring them. Fig. 10-9 shows how the two measures of elongatedness behave differently on a toy region with a hole. Another important difference between variables  $Elngt d n s s_{AndNoHole}$  and  $Elngt d n s s_{NM}$  is the algorithm used to compute the region's skeleton. Rather than a traditional thinning-based skeletonization algorithm, like the one adopted by Nagao & Matsuyama [1], we adopted an algorithm for the estimation of the region's filtered Euclidean skeleton. Many thinning-based skeletonization algorithms were proposed and analyzed over the years [109]. In general, direct thinning skeletonization procedures, though formulated with certain desirable properties in mind, present results that are not always easily predictable, i.e., results not in line with expectations. In recent years, new developments have been achieved in techniques for fast computation of Euclidean skeletons. These new techniques ensure a correct topology of the detected skeleton and allow for filtering out eventual noise [110]. An important advantage of using Euclidean skeletons is that they are better defined: in principle, a Euclidean skeleton can be defined simply as the set of points centered in the shape with respect to the Euclidean distance<sup>3</sup>. In addition, in common practice, Euclidean skeletonization algorithms compute faster than traditional region thinning algorithms. We obtained a filtered Euclidean skeleton by selecting Couprie *et al.*'s method [110]. A problem with the resulting skeleton is that it may be 4-adjacency connected. For our purposes, to obtain more consistent measurements of the total length  $L$  of the skeleton, skeleton pixels should be 8-adjacency connected. To obtain the final 8-adjacency connected skeleton we applied a cycle of the Cychosz's [111] fast implementation of Rosenfeld's parallel thinning algorithm [112], [113]. As shown in proposition 6 of the paper by Rosenfeld [112], this guarantees the final result is 8-adjacency connected. Finally, to compute the average width  $W$  of the skeleton, an original implementation was adopted to estimate the region's width across each pixel along the skeleton. As an intermediate product necessary to compute the Euclidean skeleton, Couprie *et al.*'s method computes the Euclidean distance transform of the region by using the linear-time method proposed by Meijster *et al.* [114], [115]. To estimate  $W$  in near real-time, the required per-pixel distances were read from the pre-computed values of the Euclidean distance transform.

In summary, in our experiments the proposed  $Elngt d n s s_{AndNoHole}$  index formulation and implementation improved robustness to noisy data and reduced computation time in comparison with its existing counterparts discussed by Nagao & Matsuyama [1].

<sup>3</sup> In spite of the simplicity of its mathematical definition, in practice, computing Euclidean skeletons is quite complicated due to the discrete nature of images.



8) *Area-based simple connectivity as a measure of the presence of within-segment holes*

Within-segment holes can be considered a shape property important in the identification of different classes of real-world objects depicted in images. From a perceptual standpoint, Bertamini showed that it is important to consider holes as constituent features of their enclosing objects [116]. Besides being a perceptually relevant geometric feature *per se*, holes affect area-based geometric attributes estimated in the spatial domain. On the one hand, contour-based shape descriptors are traditionally considered more sensitive to noise and variations than area-based geometric descriptors, assuming the latter are employed once within-segment holes, if any, have been filled in. On the other hand, unlike contour-based geometric attributes, area-based shape descriptors are affected by holes. If an object-specific simple connectivity value scores low, then “simple” (intuitive to use) area-based geometric indexes in the spatial domain are biased by the presence of holes and should be dealt with special care by an OBIA system. For example, inner holes in an image-objects increase geometric feature *ElngtDnssAndNoHole* and decrease geometric features *Convexity* and *Roundness*. In common practice, within-segment holes may be due to segmentation errors (e.g., undersegmentation phenomena) occurring in the inherently ill-posed image segmentation first stage (raw primal sketch [3]). In this case, segment holes can be considered an indication of segmentation noise.

To quantify the extent to which a segment is simply connected, i.e., the degree to which the segment is free of holes, an original normalized simple connectivity measure was designed and implemented as an indicator of the degree of confidence of area-based geometric indexes in the spatial domain [117]. Unfortunately, there is a limited amount of previous works on measures that assess the presence of holes. The simplest measure is the absolute number of holes of a region or, otherwise, its Euler number [7], [88]. Although the number of within-segment holes is informative, it gives no clue about the holes’ individual size and position and their overall extent and spatial distribution. Soffer & Samet [118] and Wentz [119] defined geometric measures based on the area of holes relative to that of the region. A disadvantage of area-based measures of simple connectivity is their independence from the holes’ spatial distribution. This justifies our presentation of a novel simple connectivity measure estimated as a fuzzy-AND (minimum) combination of two quantitative terms, one contour- and one area-based. The first term, called *SmplCnctvty4Adjcncty*, is contour-based. It quantifies the presence of holes by relating the length of the boundaries of the holes to the total length of all boundaries (equal to outer boundary + inner boundaries, if any) of the region. It is defined as:

$$\text{SmplCnctvty4Adjcncty} = \frac{\text{4-adjacency cross-aura measure of the external boundary}}{\text{4-adjacency cross-aura measure of the total boundary}}, \quad (19)$$

whose numerator, the “4-adjacency cross-aura measure of the external boundary” does not take into account inner boundaries of holes, if any. On the contrary, the denominator “4-adjacency cross-aura measure of the total boundary” does take into account contributions from holes. Hence, *SmplCnctvty4Adjcncty* belongs to range [0, 1]. A detailed explanation on how these boundary lengths are computed is presented in [117]. The disadvantage of the *SmplCnctvty4Adjcncty* term is its lack of sensitivity to the presence of one or few holes with a large area, but small boundary lengths. In this situation, where a desirable simple connectivity variable is expected to score low, *SmplCnctvty4Adjcncty* scores some intermediate value, which is perceptually counter-intuitive, see Fig. 10-10. This problem motivates the fuzzy-AND (minimum) combination of the *SmplCnctvty4Adjcncty* measure with an area-based geometric measure, called *FilledAreaRatio*, defined as

$$\text{FilledAreaRatio} = \text{Area} / \text{FilledArea}, \quad (20)$$

where *FilledArea* is the area of the region with its holes filled in. Hence, *FilledAreaRatio* belongs to range [0, 1]. The disadvantage of this second term of the simple connectivity variable is its low sensitivity to the presence of multiple small holes, i.e., holes featuring overall a large total boundary but a small total area, see Fig. 10-10. The proposed final measure of simple connectivity is the conservative (fuzzy-AND) combination of the two previously defined fuzzy membership values,

$$\text{CombndSmplCnctvty} = \text{Fuzzy-AND}\{ \text{SmplCnctvty4Adjcncty}, \text{FilledAreaRatio} \} = \text{Min}\{ \text{SmplCnctvty4Adjcncty}, \text{FilledAreaRatio} \}, \quad (21)$$

where *CombndSmplCnctvty* belongs to range [0, 1]. When there is no hole in the region, *CombndSmplCnctvty* scores 1. Fig. 10-10 shows how this ultimate simple connectivity formulation performs in better agreement with our perception of holes in a region.



### 10.6.2 Pairwise Feature Test of Statistical Independence

Seven geometric measures, featuring twenty-one pairwise combinations (refer to Chapter 10.6.1) were analyzed for statistical independence by means of the Pearson's chi-square test for independence, in compliance with statistical test constraints listed in Chapter 10.5.2.1. First, the two test populations of real-world geometric objects, described in Chapter 10.4, were subject to simple random sampling by a decimation factor (1/10). Next, the two sample sets were added to the third synthetic dataset to form a single test set, consisting of 745 samples. According to Equation (1), an empirical choice to transform a quantitative variable into a qualitative variable, eligible for being input to the Pearson's chi-square test for independence, is to choose  $k = \text{number of equiprobable buckets} = 2 * N^{(2/5)}$ , where  $N = \text{finite size of the sample dataset} = 745$ , therefore  $k = 28$  [82]. Hence, the cumulative distribution of each sample variable was estimated to detect twenty-eight equiprobable buckets per variable. Twenty-one two-way contingency tables were computed between pairs of the seven quantized geometric features, together with their expected frequency counts. The expected frequency count for each cell of the contingency tables was at least 5, to comply with the statistical analysis requirements, refer to Chapter 10.5.2.1. Finally, the chi-square test statistic was computed with the Windows Excel CHITEST() function, whose output is the P-value (refer to Chapter 10.5). In addition to the Pearson's chi-square test for independence, the normalized Pearson's chi-square index [50], also known as Cramer's V index (CVI) [52], was estimated, in agreement with [120]. It is defined as follows.

$$\text{CVI} = \text{Pearson's chi-square index} / \text{Maximum of the Pearson's chi-square index} = \text{Pearson's chi-square index} / [N * (\min(\text{row}, \text{column in the contingency table}) - 1)], \text{CVI} \in [0, 1], \quad (22)$$

where  $N$  is the number of samples [120]. Results are shown in Table 10-1 and Table 10-2 respectively. About the P-value and the CVI results, the following considerations hold.

(i) The P-value is the probability that a chi-square statistic having  $(\text{row} - 1) * (\text{column} - 1)$  degrees of freedom is equal or superior to the chi-square statistic value computed in the sample test for independence. When the P-value is less than the significance level  $\alpha$ , fixed equal to 0.05, then the null hypothesis of independence,  $H_0$ , cannot be accepted at a level of confidence equal to 0.95 (refer to Chapter 10.5). The P-value features an inductive inference value, i.e., it is suitable for use in an inductive (bottom-up, data-driven) inference framework, to make predictions upon the entire population based on a population sample. In common practice, chi-squared values tend to increase (showing increasing dependence) with the number of contingency cells. This may be due to any of the accidental or systematic errors listed in [50].

(ii) The CVI may be considered as the association (dependence) between two variables as a percentage of their maximum possible association. CVI varies from 0, corresponding to no association (independence) between the two variables, to 1, meaning complete association [52], [120], [121]. As such, the CVI holds a somehow "absolute" (data-independent) value in the normalized range of change  $[0, 1]$ . In practice, to gain normalization of its domain of change the CVI sacrifices its sensitivity to changes in the input dataset [52]. It is known that a discrepancy may arise in statistical sampling when the chi-square P-value turns out to be very low, such that the null hypothesis of independence between the two categorical input variables is dismissed based on evidence collected from sample data. In this situation, the CVI can border zero as an expression of an independence condition, in disagreement with the P-value [120]. On the other hand, in common practice, the greater the difference between rows and columns in the contingency table, which means the smaller the denominator of the CVI formulation, the more likely CVI will increase tending to 1, i.e., it can show increasing dependence although no chi-square statistical dependence at a given significance level is detected. In general, CVI may not be completely accurate for comparing the degree of association in different contingency tables [52]. As a consequence of these known drawbacks, the CVI was considered as yet-another source of weak statistical evidence, but little useful in practice. In fact, no known cut-off value was applied to CVI in Table 10-2 to mark pairs of random variables as independent.

For the sake of completeness, we mention that, in line with theoretical expectations, the two shape properties omitted from Table 10-1 and Table 10-2, namely, area and orientation, proved to be statistically independent from any other proposed variable.

### 10.6.3 Monotonically Increasing or Decreasing Relationship between Pairs of Planar Geometric Features at the Population and Individual Levels of Analysis

At the level of analysis of an entire population of 2D shape instances, Table 10-3 presents the SRCC values estimated in the twenty-one pairwise comparisons of seven ranked variables generated from the planar shape features discussed in Chapter 10.6.2.





For each pair of geometric features considered in Table 10-3, with special regard to cells depicted in dark gray which deserve further investigation to avoid inter-feature causality, one or more 3-tuples of planar objects, R, E1 and E2, were found in the test dataset, capable of fulfilling the third sufficient (and necessary?) condition for feature pair non-causality, refer to Chapter 10.5.2.3.

## 10.7 Discussion

### 10.7.1 Qualitative Analysis of 2D Shape Indexes

A graphical user interface (GUI) was specifically developed to allow an expert photointerpreter to assess qualitatively whether geometric indexes, estimated from hundreds of different image-objects, comply with theoretical and perceptual expectations. Screenshots of the GUI employed in real-world vision problems, ranging from satellite Earth observation (EO) image understanding to leaf image classification, are shown in Fig. 10-11 to Fig. 10-13. These examples illustrate how a discrete and finite dictionary of quantitative geometric terms, provided with an intuitive physical meaning, can be combined together, in addition to per-object photometric features and inter-object spatial relationships, to develop an application-specific physical model-based decision tree, eligible for use in an OBIA system capable of mimicking human reasoning.

Fig. 10-11 shows an EO image subset where geometric objects, automatically detected by the SIAM expert system in the 2 m resolution spaceborne MS test image (refer to Chapter 10.4), are replaced by values of their geometric attributes, specifically *ElngtdnssAndNoHole*, *CombndSmplCnctvty* and *MltSclStrghtnsOfBndrs*. Based on theoretical considerations (refer to Chapter 10.6.1.8), if *CombndSmplCnctvty* scores low then area-based geometric indexes in the spatial domain, including *ElngtdnssAndNoHole*, *CnvxtyAndNoHole* and *RndnssAndNoHole*, should be considered biased. This is correctly shown in Fig. 10-11, where segments containing holes and scoring “low” in *CombndSmplCnctvty*, refer to Fig. 10-11(c), present somewhat “high” values of *ElngtdnssAndNoHole* in Fig. 10-11(b), even though they may look relatively compact. The *MltSclStrghtnsOfBndrs* estimate, shown in Fig. 10-11(d), appears suitable for capturing manmade surface structures, which usually present straight boundaries irrespective of their shape, either simple or complex.

Fig. 10-12 presents a screenshot of the GUI showing some of the geometric objects detected by the SIAM expert pre-classifier in the spaceborne image depicted in part in Fig. 10-11(a). Segments numbered 1 through 6 are buildings or parts of buildings, while segments 7 through 9 are pieces of roads. To distinguish buildings from roads in a GEOBIA framework, some general decision rules can be inferred. The most discriminative geometric attribute appears to be *ElngtdnssAndNoHole*. Though buildings can be somewhat elongated, roads usually present very high values of *ElngtdnssAndNoHole*. Additionally, roads are typically non-compact, resulting in lower values of *RndnssAndNoHole*. Regarding appearance, bright segments (e.g., provided with high values of a photometric attribute called “mean panchromatic intensity”, complementary not alternative to geometric attributes) are usually indicative of buildings. Though buildings can appear as either bright or dark structures, asphalt roads are always dark, so that bright segments are highly indicative of non-road structures. Noticeably, not all rectangular segments are buildings, as exemplified by the mostly rectangular road segment 9. Some more specific decision rules employing different geometric attributes can be inferred from Fig. 10-12. Segments 4 and 5 represent buildings whose change in brightness generate holes in their surface area, lowering their simple connectivity values. Nonetheless, both segments have high values of indexes fuzzy rectangularity and straightness of boundaries, which are both contour-based. Segment 6 is particularly interesting, since it represents a building, yet it has an overall concave shape. This causes low values in *CnvxtyAndNoHole* and *RndnssAndNoHole*, though the segment scores high in *MltSclStrghtnsOfBndrs*. Segment 7 depicts a piece of a road network, which results in a very low value of *CnvxtyAndNoHole*. This is not the case of segment 8, which belongs to a single straight piece of road. In common, both segments share a very high value of index *ElngtdnssAndNoHole*, distinguishing them from other geometric objects. Finally, segment 9, though originating from a piece of road, is very hard to be distinguished from a building in general. It is clear that segment 9 depicts a piece of road only if its spatial neighborhood (not shown in Fig. 10-12) is observed in the image domain: this spatial context consists of road-like segments that possess the same overall orientation and cast no shadow, whereas neighboring buildings tend to cast shadows.

Fig. 10-13 presents a GUI screenshot of leaves, each from a different tree species. For example, in segment 1, the cluster of pine leaves stands out for having a very small area, high *ElngtdnssAndNoHole*, and low *RndnssAndNoHole*. Segments 2 through 4, which depict compound leaves, share high values of *ElngtdnssAndNoHole*. In particular, segment 2, featuring very thin leaflets, presents the highest *ElngtdnssAndNoHole* value. Compound leaves also tend to present holes formed from the overlap of separate individual leaflets, as exemplified by segments 2 and 3. This results in a decrease of their



*CombndSmplCnctvty* index. Segments 6 and 7 represent two different species of oak trees, though the marked protrusions in segment 7 give rise to lower values of *Convexity*, *RndnssAndNoHole* and *MltSclStrghtnsOfBndrs*. The sycamore leaf in segment 9 can be distinguished from the maple in segment 8 mainly by its low *MltSclStrghtnsOfBndrs* value. Finally, segment number 5 is a simple leaf with a smooth boundary, shown here to provide yet another reference set of attribute values, eligible for use in quantitative shape matching and retrieval.

### 10.7.2 Qualitative Interpretation of Quantitative Statistical (In)dependence Results

According to Chapter 10.1, care must be taken not to over-interpret or over-rely on the findings of GOF tests, like the Pearson's chi-square test for independence. GOF tests have notoriously low power and are generally best for rejecting poor distribution fits rather than for identifying good fits [60]. That said, Table 10-1 shows twenty-one inter-feature P-values computed with the Microsoft Excel CHITEST() function, adopted to finalize the Pearson's chi-square test for independence. If a P-value = Probability(chi-square value > test chi-square value) is less than the selected level of significance  $\alpha = 0.05$ , then the null hypothesis  $H_0$  is rejected at a 95% level of confidence (refer to Chapter 10.5), i.e., the two random variables are statistically dependent based on inductive inference. If two variables are statistically dependent, then knowing the value of one variable *can* help inferring the value of the second variable. For example, if contour-based *FuzzyRuleBsdRctnglRty*, independent of holes, is "high", then area-based *RndnssAndNoHole* and *ElngtdnssAndNoHole* are expected to be low; if *RndnssAndNoHole* and/or *Elongatedness* are high, then *FuzzyRuleBsdRctnglRty* is expected to be low. As another example, area-based *ElngtdnssAndNoHole*, increasing with holes, and area-based *CnvxtyAndNoHole*, decreasing with holes, are expected to be inversely related irrespective of holes, because an elongated region is likely to be bent, which causes convexity to score low. Based on these preliminary theoretical considerations, all statistical occurrences of (in)dependence highlighted by Table 10-1 appear intuitive to explain. In Table 10-1, two-of-seven features, the *DMPmltSclChrtrstc* and the *MltSclStrghtnsOfBndrs*, are statistically independent from any other geometric index. According to these authors' opinion, in Table 10-1, only one pairwise result of independence was somehow counterintuitive: area-based *CnvxtyAndNoHole*, decreasing with holes, and *CombndSmplCnctvty*, also decreasing with holes, resulted to be statistically independent, whereas we expected this feature pair to be statistically dependent.

It is known that a discrepancy between the chi-square P-value and the CVI may arise in statistical sampling when the chi-square P-value, which has a "relative" data-dependent inference value, turns out to be very low, meaning that the two categorical input variables are dependent. In this situation, the CVI, which belongs to an "absolute" domain of change, can border zero as an expression of a condition of independence, in disagreement with the P-value [120], refer to Chapter 10.6.2. This is exactly what happens in Table 10-2, where all pairwise feature tests show independence in "absolute" terms, irrespective of the "relative" evidence of dependence shown in Table 10-1. Although the result shown by Table 10-2 is highly desirable (all variable pairs are independent), it cannot be considered "ultimate", since the findings of GOF tests must always be scrutinized with care by an HDM, based on visual inspection of the probability plots and other comparisons [60]. For example, according to [52], the use of CVI is not recommended because, in the normalization of its domain of change, it sacrifices sensitivity to changes in the input dataset.

### 10.7.3 Qualitative Interpretation of Quantitative SRCC results

Table 10-3 shows SRCC values perfectly in line with theoretical expectations. There are five-of-twenty-one SRCC instances to be considered either "average" or "high", i.e., superior to 0.4 (refer to Chapter 10.6.2). All these cases regard area-based geometric features affected by the presence (or absence) of holes, estimated via parameter *CombndSmplCnctvty*, specifically, *ElngtdnssAndNoHole* (monotonically decreasing with *CombndSmplCnctvty*) and *RndnssAndNoHole* (monotonically increasing with *CombndSmplCnctvty*), whereas *RndnssAndNoHole* is monotonically decreasing with *ElngtdnssAndNoHole* and monotonically increasing with *CnvxtyAndNoHole*, which implies that the *CnvxtyAndNoHole* index too is monotonically decreasing with *ElngtdnssAndNoHole*.

Five-of-seven variable pairs considered statistically dependent in Table 10-1 feature an "average" or "high" SRCC value in Table 10-3, but only two-of-seven variable pairs score "high" in the SRCC, with three geometric features involved: *ElngtdnssAndNoHole*, *RndnssAndNoHole* and *CnvxtyAndNoHole*. Before the third and last level of the hierarchical test for pairwise variable causality proposed in Chapter 10.6, these three variables are eligible for being considered the same dependent variable. If a population-wide distribution is collected (at large spatial extent) from local estimates, then local non-stationary statistics may not survive the averaging process. To account for this consideration, a "local" scrutiny by an HDM of the pairwise feature pair at risk is recommended at the third and last level of the proposed analysis for causality. Actually, it is easy to provide evidence, e.g., based on synthetic examples of individual 2D objects, where the three intuitive



geometric variables *ElngrdnssAndNoHole*, *RndnssAndNoHole* and *CnvxtyAndNoHole* show their “local” absence of a monotonically increasing or decreasing pairwise feature relationship. More in general, investigations at a local spatial scale revealed that none of the feature pairs in the selected set of seven geometric features is either monotonically increasing or decreasing, as recommended by Chapter 10.6.3. This local-scale conclusion does not contradict large-scale SRCC results shown in Table 10-3, but provides insights on the behavior of planar geometric features at multiple spatial scales of analysis, from local scale, at the level of individuals, to global spatial extent, at the level of population of individuals.

Since implementation of the three levels of sufficient but not necessary criteria, required by Chapter 10.6 to investigate pairwise feature non-causality, brought no evidence of causal relationship, the selected set of seven 2D shape attributes can be considered minimally dependent (mD), until proved otherwise.

## 10.8 Conclusions

The present research and technological development (RTD) software project moves from the seminal works by Nagao & Matsuyama [1], [2] and Shackelford & Davis [68], [69], published in the geographic object-based image analysis (GEOBIA) literature, to design, implement and validate, for quantitative quality assurance (Q<sup>2</sup>A) purposes, an original general-purpose dictionary of seven off-the-shelf 2D shape descriptors in the spatial domain, provided with an intuitive physical meaning to be suitable for use in (GE)OBIA systems in operating mode, required to mimic human reasoning. This is an inherently ill-posed multi-objective optimization (cognitive) problem. To become better posed for numerical solution, it adopts the following original combination of constraints. The ensemble of planar geometric indexes is required to be: (I) general purpose, (II) minimally dependent (mD) and (III) maximally informative (MI), in compliance with the Occam’s razor principle, familiar to the machine learning community (refer to footnote 2), (IV) intuitive to use in a (GE)OBIA paradigm, expected to mimic human reasoning, and (V) subject to a *Val* policy for Q<sup>2</sup>A, in compliance with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines (refer to Appendix 3). In addition to the ensemble mDMI optimization criteria, each individual 2D shape descriptor is expected to optimize a set of community-agreed quantitative quality indicators (Q<sup>2</sup>I) of operativeness (Q<sup>2</sup>IOs), including: (VI) accuracy, (VI) computational efficiency, (VII) invariance with respect to translations, rotations and scaling transformations, and (VIII) robustness to the presence of noise in the data. Also the operator-specific maximization of Q<sup>2</sup>IOs is subject to a *Val* policy for Q<sup>2</sup>A, in compliance with the QA4EO guidelines (refer to Appendix 3).

Unfortunately, but realistically and in line with existing literature, no “objective” (quantitative) assessment of the qualitative degree of informativeness (MI) of the proposed set of application-independent descriptors is claimed in this study, to be rather accomplished by a (subjective) human decision maker (HDM) in his/her own data- and application-specific domain of interest.

Of potential interest to a broad audience of computer vision and RS scientists and practitioners, conclusions of this software project are twofold. First, the proposed general-purpose dictionary of seven off-the-shelf 2D geometric descriptors in the spatial domain (plus area and orientation, refer to Chapter 10.6), specifically, *CnvxtyAndNoHole*, *FuzzyRuleBsdRctnglrty*, *RndnssAndNoHole*, *MltScIStgrhtnsOfBndrs*, *DMPmltScIChrtrstc*, *ElngrdnssAndNoHole* and *CombndSmplCncvty*, provided with an intuitive physical meaning and subject to a known *Val* policy for Q<sup>2</sup>A, can be integrated into software libraries of efficient and reliable computational geometry algorithms, such as CGAL [27], where “simple” 2D shape descriptors are absent. This compact feature set of certified quality is alternative to the large number and variety of 2D shape functions implemented in existing commercial or open source software libraries, such as eCognition’s [30], OpenCV [31] (refer to Appendix 1) and ENVI’s [32] (refer to Appendix 2), whose multi-objective geometric feature representation and description criteria remain unknown for *Val* purposes. At the levels of understanding of system design and knowledge/information representation [3], [15], three of the seven proposed geometric features, specifically, *MltScIStgrhtnsOfBndrs*, *DMPmltScIChrtrstc* and *CombndSmplCncvty* are totally new, i.e., they are absent from the three aforementioned commercial or open source software libraries and from the reference works by Nagao & Matsuyama [1], [2] and Shackelford & Davis [68], [69]. At the levels of understanding of algorithms and implementation [3], [15], all the proposed geometric descriptors feature several degrees of novelty, introduced to optimize their Q<sup>2</sup>IO values in comparison with alternative solutions.

The second original contribution of the present study of potential interest to a wide scientific audience is the proposed hierarchical *Val* strategy for Q<sup>2</sup>A of a set of quantitative random variables whose dependence (causality [48]) must be minimized (mD). At the first sufficient level of investigation for independence, in place of the popular Pearson’s cross-correlation coefficient (PCC), which is sensitive to bivariate linear relationships exclusively, a more general Pearson’s chi-square test for independence is proved to be feasible and more adequate than PCC to quantitatively investigate the degree



of pairwise feature dependence. This statistical improvement comes at the cost of a preliminary transformation of each input quantitative random variable into a qualitative (nominal) variable featuring an adequate number of equiprobable levels of discretization, according to heuristic statistical criteria, see Equation (1). In common practice, the Pearson's chi-square test for independence is implemented by the well-known Windows Excel CHITEST() function, capable of coping with a chi-square distribution whose number of degrees of freedom is very high, equal to  $(\text{row} - 1) * (\text{column} - 1)$ , where  $\text{row} = \text{column} = \text{number of equiprobable levels of discretization} = 28$  in our experiments (refer to Chapter 10.6.2). At the second sufficient level of investigation, to be applied when the Pearson's chi-square test for independence reveals statistical dependence of a variable pair, the Spearman's rank cross-correlation coefficient (SRCC) value is estimated. It shows whether two ranked random variables are monotonically increasing or decreasing, independent of linear relationships. Finally, at the third level of investigation, to be applied if the first two sufficient tests for independence reveal pairwise variable dependence, a local proof of the absence of monotonically increasing or decreasing pairwise feature relationships is required to account for the typical nonstationarity of planar statistics. Overall, convergence-of-evidence stemming from these three levels of analysis is perfectly in line with theoretical expectations.

Future applications of the implemented set of general-purpose 2D shape descriptors in operating mode will regard specific image domains and (GE)OBIA classification systems, in comparison with alternative software libraries of geometric functions.

### Appendix 1 – OpenCV library

The OpenCV Library [31] supports the estimation of a wide set of quantitative geometric indexes as summary characteristics of contours, rather than area-based polygons, so that two (1D) curves or (2D) polygons, equivalent to, respectively, open or closed contours, can be compared (matched) for (dis)similarities by simply computing their summary characteristics and estimating their difference according to a chosen (dis)similarity criterion. As reported in Chapter 10.2, area-based descriptors are traditionally considered more robust, i.e., less sensitive to noise or shape deformations, while boundary-based descriptors are considered more sensitive. The OpenCV contour-specific geometric descriptors and matching functions are summarized below.

1. If we are drawing a contour (in vector data format) or are engaged in shape analysis, it is common to approximate a contour representing a polygon with another contour having fewer vertices. This is accomplished with the routine `cvApproxPoly()`, implemented as the RDP approximation [93].
2. Closely related to the contour/polygon approximation is the process of finding dominant points, with the routine `cvFindDominantPoints()`.
3. Contour length - `cvArcLength()`; Contour (polygon) area – `cvContourArea()`;
4. Positional attributes as bounding boxes, specifically, `cvBoundingRect()`: horizontal bounding rectangle; `cvMinAreaRect2()`: to handle rectangles of any inclination; Enclosing circles and ellipses; Convex hull and convexity defects.
5. One of the simplest ways to compare two contours is to compute contour moments: Moments, Normalized moments (scale-invariant) and the seven Hu invariant moments (scale-invariant, rotation-invariant) – `cvGetHuMoments()`. Unfortunately, contour moments lack an intuitive meaning.
6. Pairwise geometrical histograms (PGH) as a generalization of a chain code histogram representation of a contour – `cvCalcPGH()`.
7. Several functions provide hierarchical contour trees and accomplish hierarchical matching of contour trees.

### Appendix 2 – ENVI EX 5.0

In the ENVI EX commercial software product [32], the feature extraction phase adopts an OBIA approach to classification, as opposed to pixel-based classification. Image-objects can be depicted with a variety of spatial, spectral, and texture attributes, to be selected interactively by the user. If the user chooses to compute spatial attributes, ENVI EX performs an internal raster-to-vector operation and computes spatial attributes from the vectors. In particular, ENVI EX calculates all of its spatial attributes based on a smoothed version of the geometry, not the original geometry, according to the RDP approximation [93]. Performing calculations on a smoothed geometry ensures that shape measurements are less sensitive to object rotation and image noise. The list of spatial descriptors supported by the ENVI EX 5.0 commercial software toolbox is provided in Table 10-4 [32].





### Appendix 3 – QA4EO guidelines

According to the QA4EO recommendations [33], delivered by the intergovernmental Group on Earth Observations (GEO), the visionary goal of a timely, comprehensive and operational transformation of massive amounts of spaceborne/airborne EO images into information products requires the successful implementation of two key principles: Accessibility/Availability and Suitability/Reliability of RS data, processes and outcomes. The QA4EO key principle of Suitability/Reliability relies on mandatory calibration and validation (*Cal/Val*) activities, whose implementation becomes critical to the quantitative (metrological/statistically-based) Q<sup>2</sup>A of data, processes and products. *Cal* is the transformation of sensory data into a physical unit of radiometric measure [122], which guarantees harmonization and interoperability of multi-source and/or multitemporal datasets. *Val* is the process of assessing, by independent means to be community-agreed upon, the “standard” quality of process and outcome. It requires each data processing stage and output product to be assigned with Q<sup>2</sup>Is, to be community-agreed upon, featuring a degree of uncertainty in measurement at a known degree of statistical significance. Hence, *Val* provides a documented traceability of the propagation of errors through the information processing chain, in comparison with established “community-agreed reference standards” [33].

Quite strikingly, the GEO’s *Cal/Val* requirements included in the QA4EO guidelines, although regarded as common knowledge, are neglected or ignored in the RS common practice [15], [16]. About the *Cal* requirement, it is an unquestionable fact that the word “calibration” is absent from a large portion of papers published in the RS literature. For example, when popular spectral indexes, e.g., vegetation indexes, are computed from raw digital numbers not transformed into a radiometric unit of measure, such as top-of-atmosphere reflectance, published conclusions lack any physical meaning, e.g., refer to ([123], p. 3027, Fig. 10-1). Another unquestionable fact is that, in popular RS image processing commercial software products [30], [32], RS data *Cal* is not a pre-requisite [15], [16]. It means that these popular RS image processing commercial software products are collections of statistical (inductive, bottom-up) rather than physical (deductive, top-down) model-based algorithms. The former are inherently ill-posed [38], [39], semi-automatic and site-specific [56], but do not consider data *Cal* as mandatory, although their robustness to changes in the input dataset may benefit from data harmonization accomplished by a data *Cal* policy. In the RS common practice, EO image understanding systems (EO-IUSs) relying exclusively on inductive learning-from-data algorithms are expected to score low in community-agreed Q<sup>2</sup>IOs, such as degree of automation, robustness to changes in the input dataset, robustness to changes in input parameters, scalability, transferability and timeliness (from data acquisition to product generation) [15], [16]. About the *Val* requirement, it is a fact that, in the majority of papers published in the RS literature, classification accuracies are not provided with a mandatory degree of uncertainty in measurement [124]; as a consequence, these accuracy values feature no statistical meaning. For example, in the framework of the increasingly popular GEOBIA paradigm [40], a typical EO-IUS implementation, such as that proposed in [125], employs no *Cal* policy; no *Val* strategy is applied to a multi-spectral (MS) image panchromatic sharpening pre-processing and to a low-level inherently ill-posed image segmentation first stage [11]; accuracy estimation of the high-level classification second stage is provided with no degree of uncertainty in measurement, etc. The consequence of this ongoing lack of community-agreed *Val* initiatives, where Q<sup>2</sup>IOs are estimated in compliance with the QA4EO guidelines, is that the application domain of a great majority of the EO-IUSs published in the RS literature remains unknown to date [15], [16].

In an interdisciplinary framework such as that sketched in Fig. 10-1, the first QA4EO requirement of *Accessibility/Availability* is related to the well-known Shannon’s information theory of data communication [126], called unequivocal (“easy”, quantitative) information-as-thing theory by philosophical hermeneutics [4]. The second QA4EO constraint of *Suitability/Reliability* is related to the inherently equivocal (“difficult”, qualitative) information theory known as information-as-data-interpretation [4], which is the interdisciplinary focus of attention of epistemology [46], [47] and cognitive science [21], [22], see Fig. 10-1. Not surprisingly, the first (“easy”) QA4EO key principle has recorded significant improvements by the RS community in recent years [127]. Unfortunately, the second (“difficult”) QA4EO key principle remains to a large degree an open problem. This may explain why, still now, the percentage of data downloaded by RS stakeholders from the European Space Agency EO databases is estimated at about 10% or less [128].

### Acknowledgments

This work was supported in part by the National Aeronautics and Space Administration under Grant No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. The authors are very grateful to Prof. David W. Jacobs for helpful discussions and for revising the paper. A. Baraldi thanks Prof. Christopher Justice, Chair of the Department of Geographical Sciences at the University of Maryland, and Prof. Luigi Boschetti, currently at the Department of Forest, Rangeland and Fire Sciences, University of Idaho, for their support. The authors also wish to thank the Editor-



in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.

## References in Chapter 10

- [1] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York, 1980.
- [2] T. Matsuyama and V. S.-S. Hwang, *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Plenum Press, New York, 1990.
- [3] D. Marr, *Vision*. New York: Freeman and C., 1982.
- [4] R. Capurro and B. Hjørland, "The concept of information," *Annual Review of Information Science and Technology*, vol. 37, pp. 343-411, 2003.
- [5] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, "Describable visual attributes for face verification and image search," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 33, no. 10, pp. 1962-1977, 2011.
- [6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)*, vol. 40, no. 2, pp. 5-10, 2008.
- [7] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, "Shape measures for content based image retrieval: a comparison," *Info. Proc. Management*, vol. 33, no. 3, pp. 319-337, 1997.
- [8] S. Jeannin (Ed.), *MPEG-7 Visual part of experimentation model version 5.0*, ISO/IEC JTC1/SC29/WG11/N3321, Nordwijkerhout, March, 2000.
- [9] D. Zhang and G. Lu, "Evaluation of MPEG-7 shape descriptors against other shape descriptors," *Multimedia Systems*, vol. 9, pp. 15-30, 2003.
- [10] T. Sikora, "The MPEG-7 Visual Standard for Content Description - An Overview," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 11, no. 6, pp. 696-702, 2001.
- [11] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. of the IEEE*, vol. 76, no. 8, pp. 869-889, Aug. 1988.
- [12] *Perceptual Grouping*, Purdue University. [Online]. Available: <http://cs.iupui.edu/~tuceryan/research/ComputerVision/perceptual-grouping.html>
- [13] K. A. Stevens, Computation of locally parallel structure, *Bio. Cybernetics*, vol. 29, pp. 19-28, 1978.
- [14] A. Baraldi and F. Parmiggiani, "Combined detection of intensity and chromatic contours in color images," *Optical Engin.*, vol. 35, no. 5, pp. 1413-1439, May 1996.
- [15] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond geographic object-based and objectoriented image analysis (GEOBIA/GEOOIA). Part 1: Introduction," *Remote Sens.*, vol. 4, no. 9, pp. 2694-2735, Sep. 2012.
- [16] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation," *Remote Sens.*, vol. 4, pp. 2768-2817. 2012.
- [17] C. Liedtke, J. Buckner, O. Grau, S. Growe, and R. Tonjes, "AIDA: A system for the knowledge based interpretation of remote sensing data," in *3rd Int. Airborne Remote Sensing Conf.*, 1997.
- [18] J. B"uckner, M. Pahl, O. Stahlhut, and C. Liedtke, "geoAIDA – A knowledge based automatic image data analyser for remote sensing data," in *ICSC Congress on Computational Intell. Methods Applic. (CIMA)*, 2001.
- [19] A. Srivastava1, W. Mio, E. Klassen, and S. Joshi, "Geometric analysis of continuous, planar shapes," *Proc. 4th Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2003.
- [20] M. De Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry*. Heidelberg, Germany: Springer, 2008.
- [21] G. A. Miller, "The cognitive revolution: a historical perspective", in *Trends in Cognitive Sciences*, vol. 7, pp. 141-144, 2003.
- [22] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press, 1991.
- [23] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Rec.*, vol. 37, no. 1, pp. 1-19, 2004.
- [24] R. Laurini and D. Thompson, *Fundamentals of Spatial Information Systems*. London, UK: Academic Press, 1992.
- [25] S. Skiena, *The Algorithm Design Manual*. London, UK: Springer-Verlag, 2008
- [26] G. T. Toussaint, "Computational geometry and morphology," in *Proc. 1st Int. Symposium for Science on Form*, Tokyo, 1986, pp. 395-403.
- [27] The Computational Geometry Algorithms Library (CGAL) [Online] Available: <http://www.cgal.org/>
- [28] Library of Efficient Data Types and Algorithms (LEDA) [Online] Available: [http://www.algorithmic-solutions.com/leda/about/changes\\_archiv.htm](http://www.algorithmic-solutions.com/leda/about/changes_archiv.htm)
- [29] P. Soille, *Morphological Image Analysis*, Berlin: Germany: Springer-Verlag, 2003.
- [30] *eCognition® Developer 9.0 Reference Book*, Trimble, 2015.
- [31] *Open Source Computer Vision Library (OpenCV)*. [Online]. Available: <http://opencv.org/>.



- [32] *ENVI EX User Guide 5.0*, ITT Visual Information Solutions, Dec. 2009. [Online]. Available: [http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI\\_EX\\_User\\_Guide.pdf](http://www.exelisvis.com/portals/0/pdfs/enviex/ENVI_EX_User_Guide.pdf)
- [33] A Quality Assurance Framework for Earth Observation, version 4.0, Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS), 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- [34] M. Page-Jones, *The Practical Guide to Structured Systems Design*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
- [35] M. Peura and J. Iivarinen, "Efficiency of simple shape descriptors," in Proc. 3rd Int. Workshop on Visual Form, 1997, pp. 443–451.
- [36] Si Liu, Hairong Liu, L. J. Latecki, Shuicheng Yan, Changsheng Xu, Hanqing Lu, "Size adaptive selection of most informative features," *Assoc. Advanc. Artificial Intel.*, 2011.
- [37] H. Peng, H. F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, pp. 1226–1238, 2005.
- [38] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- [39] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*. Hoboken, NJ, USA: Wiley, 1998.
- [40] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, "Geographic object-based image analysis - towards a new paradigm," *ISPRS J. Photogram. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.
- [41] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–353, 1965.
- [42] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall, 1994.
- [43] J. Žunic, J. Pantovic and P. L. Rosin, "Measuring linearity of planar curves." A. Fred and M. De Marsico (eds.), in *Pattern Recognition Applications and Methods, Advances in Intelligent Systems and Computing* 318, DOI 10.1007/978-3-319-12610-4\_16.
- [44] P. L. Rosin, "Measuring shape: ellipticity, rectangularity, and triangularity," *Machine Vision and Applications*, vol. 14, no. 3, pp. 172–184, Jul. 2003.
- [45] P. Mather, *Computer Processing of Remotely-Sensed Images - An Introduction*. Chichester, U.K.: Wiley, 1994.
- [46] J. Piaget, *Genetic Epistemology*, New York: Columbia University Press, 1970.
- [47] D. Parisi, "La scienza cognitive tra intelligenza artificiale e vita artificiale," in *Neuroscienze e Scienze dell'Artificiale: Dal Neurone all'Intelligenza*, Bologna, Italy: Patron Editore, 1991.
- [48] J. Pearl, *Causality: Models, Reasoning and Inference*. New York (NY): Cambridge University Press, 2009.
- [49] *Chi-Square Test for Independence*. Available online: <http://stattrek.com/chi-square-test/independence.aspx>. Accessed on 27 Feb. 2015.
- [50] K. L. Delucchi, "The use and misuse of Chi-Square: Lewis and Burke revisited," *Psychological Bulletin*, vol. 94, no. 1, pp. 166–176, 1983.
- [51] E. Kreyszig, *Applied Mathematics*. Wiley Press, 1979.
- [52] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL: Chapman & Hall/CRC, 2000.
- [53] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, Sixth Edition. Harlow, England: Pearson, 2014.
- [54] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 24, no. 4, pp. 509–522, 2002.
- [55] S. Belongie, G. Mori, and J. Malik, "Matching with shape contexts," in *Statistics and Analysis of Shapes*. Berlin/Heidelberg, Germany: Springer, 2006, pp. 81–105.
- [56] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: Wiley, 2004.
- [57] *Open source computer vision library (OpenCV)*, [Online]. Available: <http://opencv.org/>. Accessed on March 20, 2015.
- [58] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.
- [59] M.K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory IT*, vol. 8, pp. 179–187, 1962.
- [60] *Distribution Selection - Cautions Regarding Goodness-of-Fit Tests*, United States Environmental Protection Agency (EPA). [Online]. Available: [www.epa.gov/scipoly/sap/meetings/1998/march/attach3.pdf](http://www.epa.gov/scipoly/sap/meetings/1998/march/attach3.pdf). Accessed on 27 Feb. 2015.
- [61] A. Baraldi, L. Bruzzone, and P. Blonda, "Quality assessment of classification and cluster maps without ground truth knowledge," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, 2005.
- [62] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 1–14, 2004.
- [63] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images", *Photogram. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [64] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Proc. Letters*, vol. 9, no. 3, pp. 81–84, March 2002.



- [65] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Letters*, vol. 1, no. 4, pp. 313-317, Oct. 2004.
- [66] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, and F. Nencini, "A new method for MS+Pan image fusion assessment without reference," *IEEE Proc.*, 2006.
- [67] *World Happiness Report, 2012/2013*, Columbia University, Canadian Institute for Advanced Research, London School of Economics.
- [68] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354–2363, 2003.
- [69] A. K. Shackelford, *Development of urban area geospatial information products from high resolution satellite imagery using advanced image analysis techniques*. Ph.D. dissertation, University of Missouri- Columbia, 2004.
- [70] R. Montero, "State of the art of compactness and circularity measures," *International Mathematical Forum*, vol. 4, no. 27, pp. 1305 – 1335, 2009.
- [71] A. M. Maceachren, "Compactness of geographic shape: Comparison and evaluation of measures," *Geografiska Annaler. Series B, Human Geography*, vol. 67, no. 1, pp. 53-67, 1985.
- [72] Wenwen Li, M. F. Goodchild, and R. L. Church, "An efficient measure of compactness for 2D shapes and its application in regionalization problems," *Int. J. of Geographical Information Science*, vol. 27, no. 6, pp. 1227-1250, 2013.
- [73] A. Rosenfeld, "Compact figures in digital pictures," *IEEE Trans. Systems, Man and Cyber.*, vol. SMC-4, no. 2, pp. 221–223, 1974.
- [74] L. Prechelt, "A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice," *Neural Networks*, vol. 9, 1996.
- [75] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral-rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 imagery – Part I: System design and implementation," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1299–1325, 2010.
- [76] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2563–2586, 2006.
- [77] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1326–1354, 2010.
- [78] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *European Conf. Computer Vision (ECCV)*, 2012, pp. 502–516.
- [79] J. V. B. Soares and D. W. Jacobs, "Efficient segmentation of leaves in semi-controlled conditions," *Machine Vision and Applications*, vol. 24, no. 8, pp. 1623–1643, Nov. 2013.
- [80] L. Boschetti, S.P. Flasse, and P.I. Brivio, "Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto boundary," *Remote Sens. Environ.*, vol. 91, pp. 280–292, 2004.
- [82] R. B. D'Agostino and M. A. Stephens (Eds.), *Goodness-of-Fit Techniques*. New York: Marcel Dekker, Inc. Welch, B. L., 1986.
- [83] R. S. Lunetta and C. D. Elvidge, *Remote sensing and Change Detection: Environmental Monitoring Methods and Applications*. Chelsea, MI, USA: Ann Arbor Press, 1998.
- [81] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, Lewis Publishers: Boca Raton, 1999.
- [84] D. S. Arnon and J. P. Giesemann, "A linear time algorithm for the minimum area rectangle enclosing a convex polygon," *Tech. Rep.*, Purdue University, 1983.
- [85] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, 2001.
- [86] Y. Chen, "New approaches for calculating Moran's index of spatial autocorrelation," *PLoS ONE*, vol. 8, no. 7, p. e68336, Jul. 2013.
- [87] G. A. Moore, "Automatic scanning and computer processes for the quantitative analysis of micrographs and equivalent subjects," in *Pictorial Pattern Recog.*, Washington, DC, 1968, pp. 275–326.
- [88] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Pearson/Prentice Hall, 2008.
- [89] A. A. Melkman, "On-line construction of the convex hull of a simple polyline," *Info. Proces. Letters*, vol. 25, pp. 11–12, 1987.
- [90] S. Tor and A. Middleditch, "Convex decomposition of simple polygons," *ACM Trans. on Graphics (TOG)*, vol. 3, no. 4, pp. 244–265, 1984.
- [91] B. Braden, "The surveyor's area formula," *The College Mathematics Journal*, vol. 17, no. 4, pp. 326–337, 1986.





- [92] P. S. Heckbert and M. Garland, "Survey of polygonal surface simplification algorithms," in *Multiresolution Surface Modeling Course Notes*. ACM SIGGRAPH, 1997.
- [93] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica*, vol. 10, no. 2, pp. 112-122, 1973.
- [94] A. Baraldi, "Fuzzification of a crisp near-real-time operational automatic spectral-rule-based decision-tree preliminary classifier of multisource multispectral remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2113 - 2134, June 2011.
- [95] C. E. Woodcock, A. H. Strahler, and D. L. B. Jupp, "The use of variograms in remote sensing: I. Scene models and simulated images," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 323-348, Aug. 1988.
- [96] C. E. Woodcock, A. H. Strahler, and D. L. B. Jupp, "The use of variograms in remote sensing: II. Real digital images," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 349-379, Aug. 1988.
- [97] G. M. Espindola, G. Camara, I. A. Reis, L. S. Bins, A. M. Monteiro, "Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation", *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3035-3040, 2006.
- [98] W. R. Tobler, "A computer movie simulating urban growth in the Detroit Region," *Economic Geography*, vol. 46, pp. 234-240, 1970.
- [99] A. Balaguer, L. A. Ruiz, T. Hermosilla, and J. A. Recio, "Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification," *Computers & Geosciences*, vol. 36, no. 2, pp. 231-240, Feb. 2010.
- [100] M. Armstrong, *Basic Linear Geostatistics*. Berlin/Heidelberg, Germany: Springer, 1998.
- [101] P. A. Longley, M. F. Goodchild, D. J. Maguire D. W. Rhind, *Geographic Information Systems and Science*, Second Edition. New York: Wiley, 2005.
- [102] A. J. Lembo Jr., *Spatial Autocorrelation: Moran's I and Geary's C*, Salisbury University, 2000. [Online] Available: [faculty.salisbury.edu/~ajlembo/419/sa.pdf](http://faculty.salisbury.edu/~ajlembo/419/sa.pdf)
- [103] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Proc.*, vol. 2, no. 2, pp. 176-201, 1993.
- [104] H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vis. Res.*, vol. 19, no. 1, pp. 19-32, 1979.
- [105] J. Gil and M. Werman, "Computing 2D min, median, and max filters," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 15, no. 5, pp. 504-507, 1993.
- [106] K. Robinson and P. F. Whelan, "Efficient morphological reconstruction: a downhill filter," *Pattern Recog. Let.*, vol. 25, no. 15, pp. 1759-1767, 2004.
- [107] R. Adams, "Radial decomposition of disks and spheres," *CVGIP: Graphical Models and Image Proc.*, vol. 55, no. 5, pp. 325-332, 1993.
- [108] L. Vincent, "Morphological grayscale reconstruction: definition, efficient algorithm and applications in image analysis," in *Computer Vision and Pattern Recog. (CVPR)*, 1992, pp. 633-635.
- [109] L. Lam, S.-W. Lee, and C. Y. Suen, "Thinning methodologies – a comprehensive survey," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 14, no. 9, pp. 869-885, 1992.
- [110] M. Couprie, D. Coeurjolly, and R. Zrour, "Discrete bisector function and Euclidean skeleton in 2D and 3D," *Image and Vision Comput.*, vol. 25, no. 10, pp. 1543-1556, 2007.
- [111] J. M. Cychosz, "Efficient binary image thinning using neighborhood maps," in *Graphics gems IV*, P. S. Heckbert, Ed. Academic Press Professional, Inc., 1994, pp. 465-473.
- [112] A. Rosenfeld, "A characterization of parallel thinning algorithms," *Information and Control*, vol. 29, no. 3, pp. 286-291, 1975.
- [113] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. Academic Press, Inc., 1982.
- [114] M. A. Butt and P. Maragos, "Optimum design of Chamfer distance transforms", *IEEE Trans. Image Proc.*, vol. 7, no. 10, pp. 1477-1484, 1998.
- [115] A. Meijster, J. B. T. M. Roerdink and W. H. Hesslink. "A general algorithm for computing distance transforms in linear time," in *Mathematical Morphology and its Applications to Image and Signal Processing*, Berlin/Heidelberg, Germany: Springer, 2000, pp. 331-340.
- [116] M. Bertamini, "Who owns the contour of a visual hole?" *Perception*, vol. 35, pp. 883-894, 2006.
- [117] J. V. B. Soares, A. Baraldi, and D. W. Jacobs, "Segment-based simple-connectivity measure design and implementation," *Tech. Rep.*, University of Maryland, College Park, 2014. [Online]. Available: <http://hdl.handle.net/1903/15430>.
- [118] A. Soffer and H. Samet, "Negative shape features for image databases consisting of geographic symbols," in *Proc. 3rd Int. Workshop on Visual Form*, 1997.
- [119] E. A. Wentz, "A shape definition for geographic applications based on edge, elongation, and perforation," *Geographical Anal.*, vol. 32, no. 2, pp. 95-112, 2000.



- [120] G. Portoso, “On the maximum chi-square range of variation,” *Dipartimento Scienze Economiche e Metodi Quantitativi*, Univ. degli Studi del Piemonte Orientale, 2008. [Online]. Available: <http://semeq.unipmn.it/files/0814%20-%20Quaderno%20-%20Portoso.pdf>. Accessed on: March 13, 2015.
- [121] L. Hatcher, *Step-by-Step Basic Statistics Using SAS: Student Guide*. SAS Institute, 2003.
- [122] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik, “Reflectance quantities in optical remote sensing - Definitions and case studies,” *Remote Sens. Environ.*, vol. 103, pp. 27–42, 2006.
- [123] Hanqiu Xu, “Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery,” *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025-3033, 2006.
- [124] A. Baraldi, L. Boschetti, L., and M. Humber, “Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, Part: 2, pp. 701-760, Jan. 2014.
- [125] A. Hamedianfara and H. Z. Mohd Shafr, “Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images,” *Int. J. Remote Sens.*, pp. 3380-3396, 2015.
- [126] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.
- [127] Group on Earth Observations, *GEO Announces Free and Unrestricted Access to Full Landsat Archive*. [Online] Available: [www.fabricadebani.ro/userfiles/GEO\\_press\\_release.doc](http://www.fabricadebani.ro/userfiles/GEO_press_release.doc)
- [128] S. D’Elia, European Space Agency, personal communication, 2002.

Figures and figure captions in Chapter 10

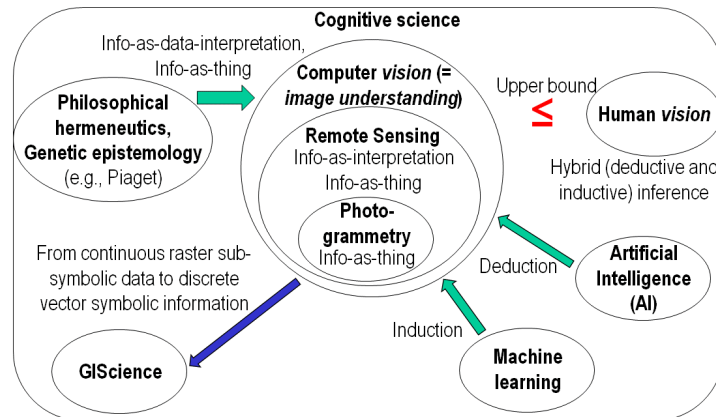


Fig. 10-1. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world [20]. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred within nervous systems (humans or other animals) and machines (e.g., computers) [21], [22].

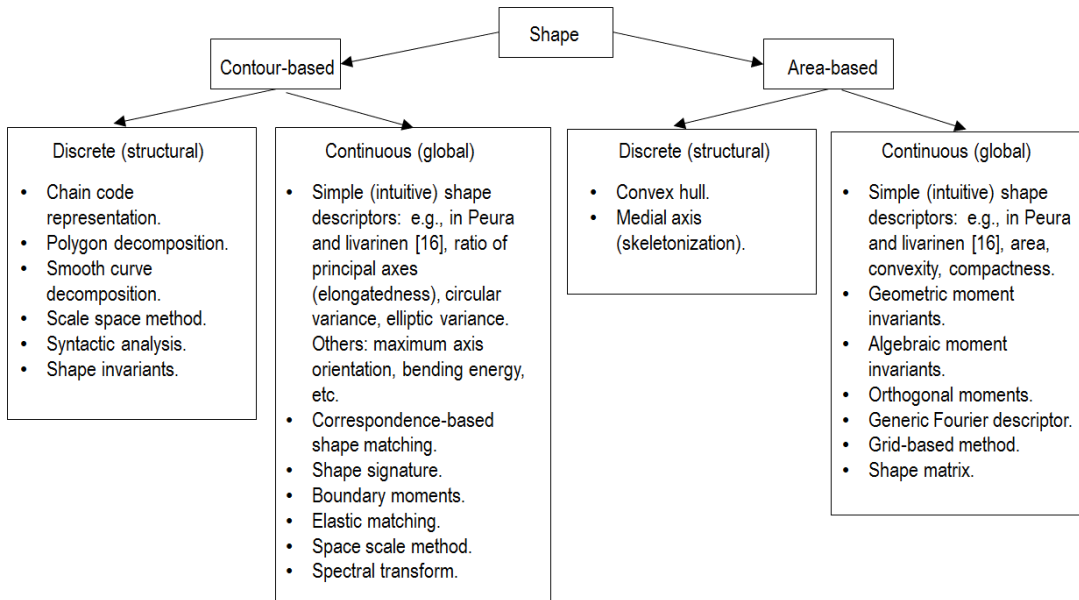


Fig. 10-2. Taxonomy of 2D/3D shape representation (modeling) and description (estimation) techniques, according to Zhang and Lu [9], [23]. Continuous approaches (global) do not divide shape into sub-parts. In the resulting multi-dimensional shape space [19], different shapes correspond to different points in the shape space and a quantification of shapes differences is accomplished using metrics between the acquired feature vectors based on geodesic lengths, e.g., Hamming distance, Hausdorff distance, comparing skeletons and support vector machine pattern matching, etc. Discrete approaches (structural) break the shape boundary into segments, called primitives using a particular criterion. The final representation is usually a string or a graph (or tree) and the similarity measure is accomplished by string matching or graph matching.

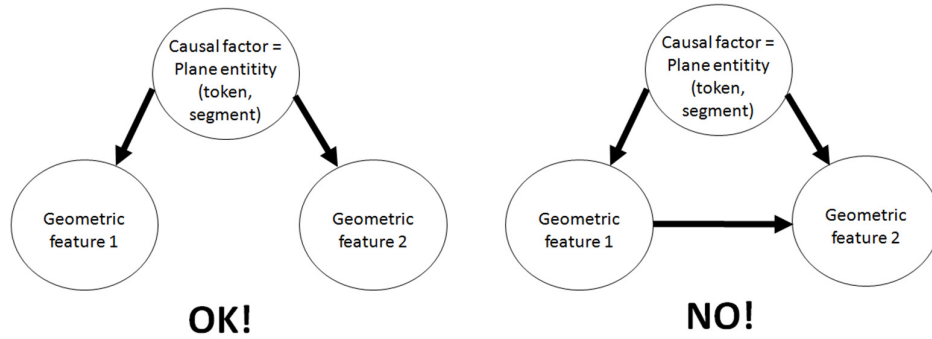


Fig. 10-3. A causal model is a directed acyclic graph  $G$ , which has no loops in it. Each vertex (node) has an associated random variable. All the arrows in a causal model indicate the possibility of a direct causal influence. All root vertices (with no parents) in the graph are labeled by independent random variables. No causal relationship between two geometric features is accepted [48].

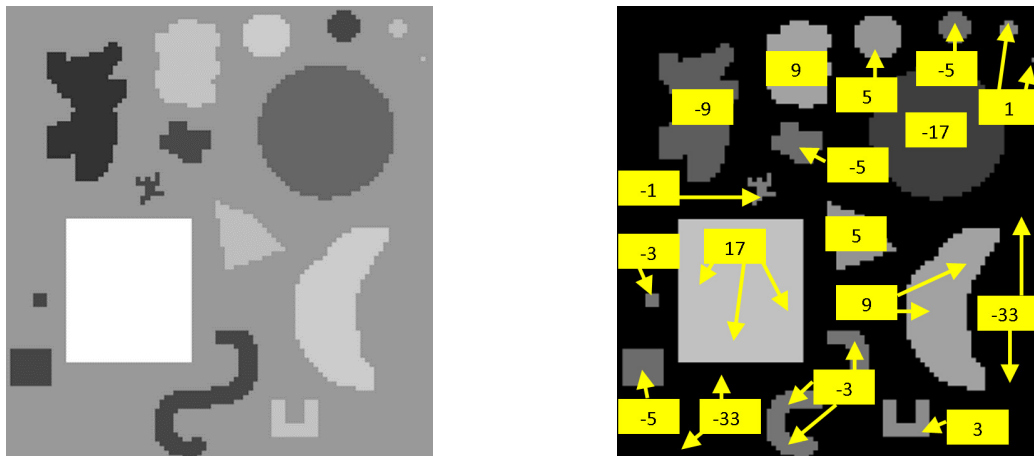


Fig. 10-4. (a) Left: Subset of a synthetic grayscale test image of known complexity, featuring several planar shapes of different size, selected to test the degree of match of the morphological multiscale characteristic with known object properties; (b) Right: Per-pixel morphological multiscale characteristic values. A negative (vice versa, positive) value indicates that the closing (vice versa, opening) differential morphological profile (DMP) contains the largest DMP response [85]. This occurs when the segment that pixel belongs to is darker (vice versa, lighter) than its background. The per-pixel characteristic scale is defined as the scale at which the DMP response is maximum. The characteristic scale thus provides an estimate of the local size (in pixel unit) of the image-object that pixel belongs to. A segment-based average of the per-pixel characteristic scale is an estimate of the size of the image-object. It is shown in yellow highlight. In this experiment, the adopted dyadic scales were  $s = 0, 1, 3, 5, 9, 17, 33$  in pixel unit (see text).



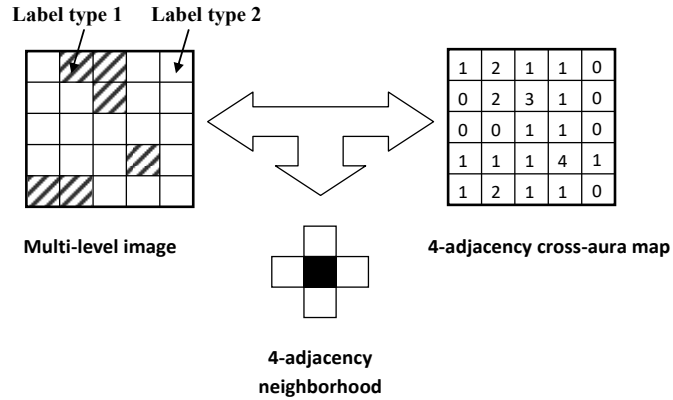


Fig. 10-5. Four-adjacency cross-aura map computed from a multi-level image. Every pixel value in the four-adjacency cross-aura map is equivalent to the number of four-adjacency neighboring pixels that do not belong to the same label type assigned to the central pixel.

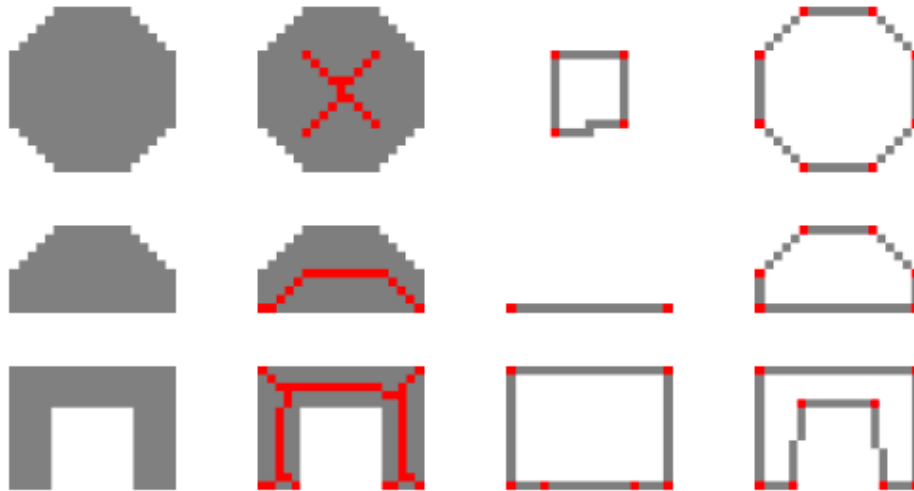


Fig. 10-6. Regions where polygon approximation by skeleton endpoints fails. From left to right, columns represent: (i) original region, (ii) skeleton (depicted in red) obtained by region thinning, (iii) polygonal approximation from skeleton endpoints (depicted in red), (iv) polygonal approximation using the Ramer-Douglas-Peucker (RDP) algorithm [92], [93]. In column (iii), the skeleton-based polygonal approximation shows problems in the first and second rows due to the rounded region boundaries where no skeleton endpoint is located. The third row appears affected by a similar effect due to concavities.



Fig. 10-7. Spaceborne VHR test image. (a) Left: A segment of a river, containing large holes. (b) Right: A segment of a road, containing large and small holes. If these holes were filled in, these segments would score low in *elongatedness*, whereas human experts would expect segments of rivers and roads to score high in *elongatedness*.

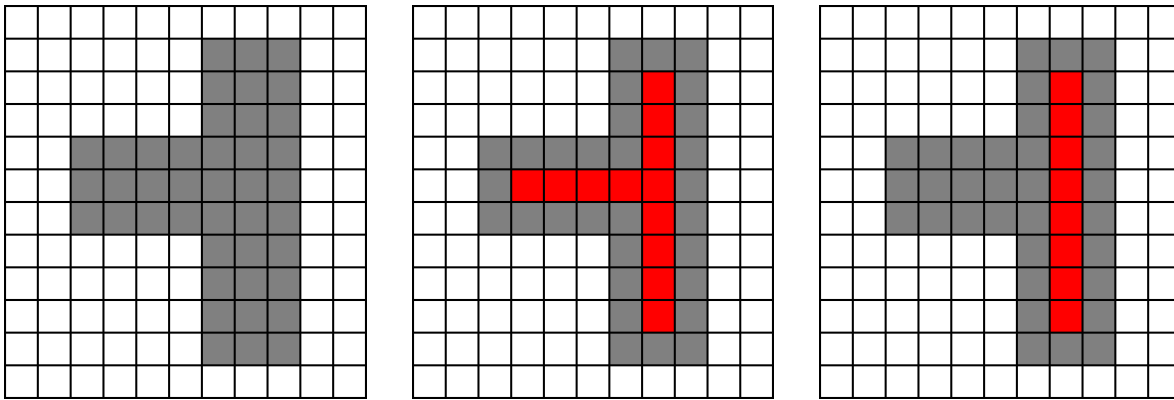


Fig. 10-8. (a) Left: Test plane object number 1. (b) Center: Skeleton of the test region. (c) Right: The longest path along the skeleton, estimated according to Nagao & Matsuyama [1]. This test illustrates how popular alternative formulations of elongatedness perform differently. For example, there is a large difference in results if the region's elongatedness is estimated using either the whole skeleton, like in Equation (18), where  $L = 12$  pixels, or just its longest path, like in Equation (17), where  $L_{NM} = 8$ . Although different methods compute local widths differently, for the purposes of illustration, let us assume the width computed at every point on the skeleton is equal to 3 pixels. Therefore,  $W = 3$  in Equation (18) and  $W_{NM} = 3$  in Equation (17). Hence,  $Elngtdnss_{AndNoHole} = \text{Equation (18)} = L / W = 12/3 = 4$ , whereas  $Elngtdnss_{NM} = \text{Equation (17)} = L_{NM} / W_{NM} = 8/3 = 2.67$ .

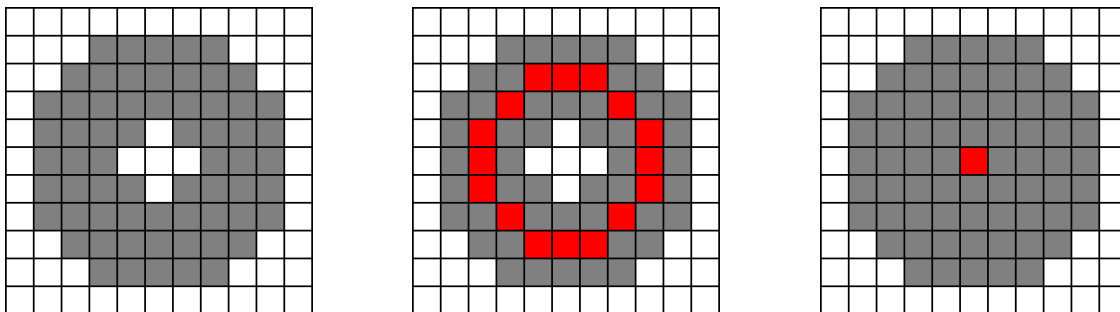


Fig. 10-9. (a) Left: Test plane object number 2. (b) Center: Skeleton of the test region. (c) Right: Skeleton of the test region whose hole was filled in, as required in [1]. When the hole is not filled in, the length value of the skeleton is

larger, while the segment's average width is smaller. The elongatedness measure we propose here uses the skeleton depicted in the center, so that  $ElngtdnssAndNoHole = L / W = 16.0 / 4.1 = 3.9$ . On the other hand, the method by Nagao & Matsuyama [1] will use the skeleton depicted on the right, resulting in  $Elngtdnss_{NM} = L_{NM} / W_{NM} = 1 / 10 = 0.1$ .




	Segment	FilledAreaRatio	Simple-Connectivity 4-Adjacency	Simple-Connectivity = $\min\{\text{FilledAreaRatio}, \text{Simple-Connectivity 4-Adjacency}\}$
Boundary of holes increases		0.84	0.71	0.71
		0.84	0.56	0.56
Area of holes increases		0.31	0.55	0.31

Fig. 10-10. Examples of simple connectivity measures. Since the total area of the holes in the top and middle segments shown in the left column is the same, then their *FilledAreaRatio* term is the same (equal to 0.84). However, in these two segments the spatial distribution of holes and the inner boundary length are different. The *SmplCnctvty4Adjcncty* term captures these differences. In the middle and bottom segments of the left column, the boundary lengths are similar, but the total areas of the holes are different. In this case, the *FilledAreaRatio* term is able to capture this difference (presenting values, respectively, of 0.84 and 0.31), while the *SmplCnctvty4Adjcncty* is not. The final measure, *CombndSmplCnctvty*, where a fuzzy-AND (minimum) operator combines the two membership functions *FilledAreaRatio* and *SmplCnctvty4Adjcncty*, behaves in good agreement with human perception.

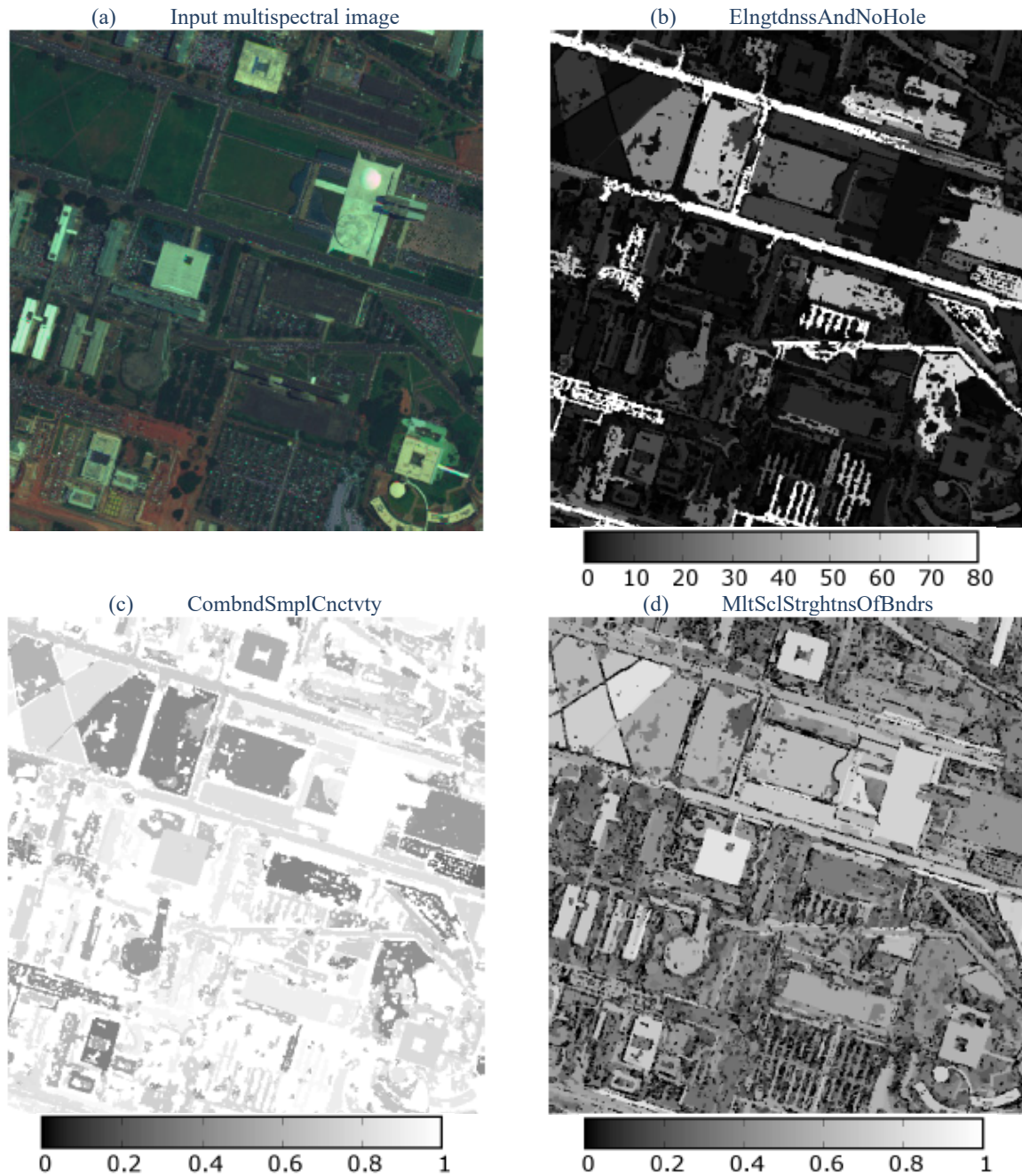


Fig. 10-11. (a) Zoom in of a WorldView-2 multispectral image of Brasilia, Brazil, depicted in false colors (R channel: Visible Red; G channel: Near Infrared; Blue channel: Visible Blue). Spatial resolution: 2 m. Acquisition date and time: 2010-08-04, 13:32 p.m.; (b) *ElngtdnssAndNoHole* planar feature values; (c) *CombndSmplCnctvty* planar feature values. When simple connectivity scores low, then area-based geometric indexes, including *ElngtdnssAndNoHole*, *CnvtvtyAndNoHole* and *RndnssAndNoHole*, should be considered biased by the presence of holes; (d) *MltScIStrghtnsOfBndrs* planar feature values.





Segment Number	Chromatic	Panchromatic	Segment	Convexity and No Hole	Elongatedness	Polygon-Based Approximate Rectangularity	Roundness and No Hole	Simple-Connectivity	Straightness of Boundary	Angle of MER (in degrees)	Area (in pixels)	Average Contrast Along Boundary	Morphological multiscale characteristic	M Panch Int
1				0.96	1.10	1.00	0.90	1.00	0.68	90.00	81	30.53	5.59	94
2				0.85	2.92	0.95	0.66	1.00	0.63	75.07	666	1.91	15.14	57
3				0.86	4.68	1.00	0.62	1.00	0.77	-16.50	237	36.07	5.12	11
4				0.87	8.72	1.00	0.53	0.72	0.83	74.05	1406	27.24	7.27	88
5				0.78	4.86	1.00	0.58	0.89	0.89	73.30	1812	27.00	15.62	76
6				0.48	9.24	0.78	0.44	1.00	0.79	155.85	461	7.83	16.31	59
7				0.35	50.42	0.72	0.22	1.00	0.89	-105.95	727	17.72	9.64	54
8				0.67	22.35	0.05	0.33	1.00	0.85	-5.57	340	20.51	8.87	53
9				0.84	9.61	1.00	0.54	0.93	0.85	167.83	555	20.22	8.67	49

Fig. 10-12. Screenshot of the GUI specifically developed to show a human expert values of the proposed set of geometric attributes. In this GUI, darker cells correspond to: (i) higher values of geometric attributes and (ii) lower values of photometric attributes, like the panchromatic mean intensity shown at the rightmost column. In this figure, for reasons of readability only nine segments are shown simultaneously for comparison. Detected in the spaceborne VHR test image of an urban area, segments 1 through 6 correspond to buildings or parts of buildings while segments 7 through 9 belong to roads. These two families of segments appear easy to discriminate based on different combinations of ranges of change of their geometric attributes.



Segment Number	Chromatic	Panchromatic	Segment	Convexity and No Hole	Elongatedness	Polygon-Based Approximate Rectangularity	Roundness and No Hole	Simple-Connectivity	Straightness of Boundary	Angle of MER (in degrees)	Area (in pixels)	Average Contrast Along Boundary	Morphological multiscale characteristic	M Panch int
1				0.34	46.14	0.00	0.24	0.91	0.93	-3.37	356	39.65	1.39	58
2				0.68	125.64	0.00	0.16	0.70	0.70	-63.00	4502	32.39	3.05	10
3				0.51	29.78	0.00	0.23	0.88	0.83	91.15	3444	42.87	7.49	48
4				0.59	14.52	0.00	0.36	0.94	0.88	177.61	6004	59.83	15.34	68
5				0.95	1.31	0.82	0.75	1.00	0.95	116.03	5112	43.14	30.41	43
6				0.85	3.47	0.09	0.65	1.00	0.82	157.46	2765	32.60	29.69	33
7				0.39	4.21	0.00	0.30	1.00	0.80	-118.86	2462	45.73	7.61	74
8				0.63	2.11	0.32	0.48	1.00	0.80	122.59	2692	45.92	14.87	23
9				0.61	3.43	0.00	0.35	1.00	0.69	37.57	3927	42.51	28.30	49

Fig. 10-13. Screenshot of the GUI specifically developed to show a human expert values of the proposed set of geometric attributes. In this GUI, darker cells correspond to: (i) higher values of geometric attributes and (ii) lower values of photometric attributes, like the panchromatic mean intensity shown at the rightest column. In this figure, for reasons of readability only nine segments are shown simultaneously for comparison. Segments 1 through 9 are examples of segments extracted from pictures of leaves of different tree species, to be discriminated by different combinations of values of their geometric attributes.



Tables and table captions in Chapter 10

Two-way contingency table, Pearson's chi-square test for independence		1	2	3	4	5	6	7
Feature ID	Feature Name	DMPmitSciC hrctrstc	FuzzyRuleBs dRctnglrty (independent of holes)	RndnsAndNo Hole	MltSciStrght nsOfBndrs (independent of holes)	ElngtdnssAndNoHole	CombndSmp ICnctvty	CnvxtyAndNoHole
1	DMPmitSciC hrctrstc	-	0.9999464464 95493 INDEP	0.999983933 92187 INDEP	0.9999916290 40779 INDEP	0.9527866856 42743 INDEP	0.9999818902 58323 INDEP	0.9999999642 92355 INDEP
2	FuzzyRuleBs dRctnglrty (independent of holes)	-	-	0.0008401199 80309 INV RELATED	0.9953849722 79882 INDEP	0.0000175247 985527093 INV RELATED	0.9212526080 90429 INDEP	0.6983218414 26978 INDEP
3	RndnsAndNo Hole	-	-	-	0.9999970291 44091 INDEP	8.3279525731 8881E-235 INV RELATED	0.0005878597 0797259 DIR RELATED	1.0891577460 9503E-83 DIR RELATED
4	MltSciStrght nsOfBndrs (independent of holes)	-	-	-	-	0.9436574494 26006 INDEP	0.9999985403 33204 INDEP	0.9855412088 0859 INDEP
5	ElngtdnssAndNoHole	-	-	-	-	-	2.1712600124 6908E-25 INV RELATED	4.9257220484 201E-36 INV RELATED
6	CombndSmp ICnctvty	-	-	-	-	-	-	0.7049899842 08181 INDEP
7	CnvxtyAndNoHole	-	-	-	-	-	-	-

Table 10-1. Instances of the probability value (P-value) for the Pearson's chi-square test for independence. If the P-value for the chi-square test for independence, such that  $P\text{-value} = \text{Probability}(\text{chi-square value} > \text{test chi-square value})$ , is less than the adopted level of significance  $\alpha = 0.05$ , then the null hypothesis  $H_0$  (the two random variables are independent) is rejected at  $(1 - \alpha) = 95\%$  level of confidence. Cells in gray deserve further investigation to avoid inter-feature causality.

Two-way contingency table, Cramer's V index (CVI)		1	2	3	4	5	6	7
Feature ID	Feature Name	DMPmitSciC hrctrstc	FuzzyRuleBs dRctnglrty (independent of holes)	RndnsAndNo Hole	MltSciStrght nsOfBndrs (independent of holes)	ElngtdnssAndNoHole	CombndSmp ICnctvty	CnvxtyAndNoHole
1	DMPmitSciC hrctrstc	-	0.0387399623 143401 INDEP	0.0422184875 974543 INDEP	0.0427464555 595364 INDEP	0.0464152995 186267 INDEP	0.0383591561 811503 INDEP	0.0399422366 815894 INDEP
2	FuzzyRuleBs dRctnglrty (independent of holes)	-	-	0.0604386176 657904 INDEP	0.0414460006 060964 INDEP	0.0444503290 568624 INDEP	0.0248705744 71763 INDEP	0.0490802657 687802 INDEP
3	RndnsAndNo Hole	-	-	-	0.0421184979 760391 INDEP	0.1975953799 24897 INDEP	0.0617587354 069437 INDEP	0.1076161185 68181 INDEP
4	MltSciStrght nsOfBndrs (independent of holes)	-	-	-	-	0.0468008697 73228 INDEP	0.0374046540 182519 INDEP	0.0458727712 064827 INDEP



5	ElngtdnssAndNoHole	-	-	-	-	-	0.0690844377 9571 INDEP	0.0942004574 163843 INDEP
6	CombndSmpICnctvty	-	-	-	-	-	-	0.0480834141 993999 INDEP
7	CnvxtyAndNoHole	-	-	-	-	-	-	-

Table 10-2. Values of the Cramer's V index =  $CVI = \text{Pearson's chi-square index} / \text{Maximum of the Pearson's chi-square index}$ , equal to  $[N * (\min(\text{row}, \text{column in the contingency table}) - 1)] = \text{Normalized chi-square test for independence} \in [0, 1]$ . Intuitively, if CVI tends to zero, then statistical independence holds. No known cut-off value is adopted though, to consider the CVI as "low" for independence.

Two-way contingency table, Cramer's V index (CVI)		1	2	3	4	5	6	7
Feature ID	Feature Name	DMPmitSciChrcrtrstc	FuzzyRuleBs dRctnglrty (independent of holes)	RndnsAndNo Hole	MitSciStrghtnsOfBndrs (independent of holes)	ElngtdnssAndNoHole	CombndSmpICnctvty	CnvxtyAndNoHole
1	DMPmitSciChrcrtrstc	-	-0.096794295	-0.095800252	0.164210063	0.075797658	-0.018788534	-0.111445368
2	FuzzyRuleBs dRctnglrty (independent of holes)	-	-	0.344483351	-0.013300685	-0.27239941	0.061261282	0.28369982
3	RndnsAndNo Hole	-	-	-	-0.270656762	-0.935916078	0.561271973	0.884509053
4	MitSciStrghtnsOfBndrs (independent of holes)	-	-	-	-	0.315012068	-0.082310281	-0.203619857
5	ElngtdnssAndNoHole	-	-	-	-	-	-0.612848806	-0.795742144
6	CombndSmpICnctvty	-	-	-	-	-	-	0.335519155
7	CnvxtyAndNoHole	-	-	-	-	-	-	-

Table 10-3. The Spearman's rank cross-correlation coefficient (SRCC) assesses how well the relationship between two ranked variables can be described using a monotonically increasing or decreasing function, even if their relationship is not linear, unlike the Pearson's cross-correlation coefficient, PCC. Traditionally, in absolute values, (i) a cross-correlation coefficient  $\geq 0.80$  represents strong agreement (cells in dark gray), (ii) between 0.40 and 0.80 describes moderate agreement (cells in light gray), and (iii)  $\leq 0.40$  represents poor agreement. Cells in dark gray deserve further investigation to avoid inter-feature causality.

Attribute type	Attribute name	Description
Spatial		Formulas for calculating COMPACTNESS, CONVEXITY, SOLIDITY, ROUNDNESS, and FORM FACTOR are from Russ, J. C. (2002). The Image Processing Handbook, Fourth Edition. Boca Raton, FL: CRC Press.
1	AREA	Total area of the polygon, minus the area of the holes. Values are in map units.
2	LENGTH	The combined length of all boundaries of the polygon, including the boundaries of the holes. This is different than the MAXAXISLEN attribute. Values are in map units.
3	COMPACTNESS	A shape measure that indicates the compactness of the polygon. A circle is the most compact shape with a value of $1 / \pi$ . The compactness value of a square is $1 / 2(\sqrt{\pi})$ . $COMPACT = \text{Sqrt}(4 * \text{AREA} / \pi) / \text{outer contour length}$
4	CONVEXITY	Polygons are either convex or concave. This attribute measures the convexity of the polygon. The convexity value for a convex polygon with no holes is 1.0, while the value for a concave polygon is less than 1.0. $CONVEXITY = \text{length of convex hull} / \text{LENGTH}$





5	SOLIDITY	A shape measure that compares the area of the polygon to the area of a convex hull surrounding the polygon. The solidity value for a convex polygon with no holes is 1.0, and the value for a concave polygon is less than 1.0. $SOLIDITY = AREA / \text{area of convex hull}$
6	ROUNDNESS	A shape measure that compares the area of the polygon to the square of the maximum diameter of the polygon. The "maximum diameter" is the length of the major axis of an oriented bounding box enclosing the polygon. The roundness value for a circle is 1, and the value for a square is $4 / \pi$ . $ROUNDNESS = 4 * (AREA) / (\pi * MAXAXISLEN^2)$
7	FORM FACTOR	A shape measure that compares the area of the polygon to the square of the total perimeter. The form factor value of a circle is 1, and the value of a square is $\pi / 4$ . $FORMFACTOR = 4 * \pi * (AREA) / (\text{total perimeter})^2$
8	ELONGATION	A shape measure that indicates the ratio of the major axis of the polygon to the minor axis of the polygon. The major and minor axes are derived from an oriented bounding box containing the polygon. The elongation value for a square is 1.0, and the value for a rectangle is greater than 1.0. $ELONGATION = MAXAXISLEN / MINAXISLEN$
9	RECTANGULAR FIT	A shape measure that indicates how well the shape is described by a rectangle. This attribute compares the area of the polygon to the area of the oriented bounding box enclosing the polygon. The rectangular fit value for a rectangle is 1.0, and the value for a non-rectangular shape is less than 1.0. $RECT\_FIT = AREA / (MAXAXISLEN * MINAXISLEN)$
10	MAIN DIRECTION	The angle subtended by the major axis of the polygon and the x-axis in degrees. The main direction value ranges from 0 to 180 degrees. 90 degrees is North/South, and 0 to 180 degrees is East/West.
11	MAJAXISLEN	The length of the major axis of an oriented bounding box enclosing the polygon. Values are map units of the pixel size. If the image is not georeferenced, then pixel units are reported.
12	MINAXISLEN	The length of the minor axis of an oriented bounding box enclosing the polygon. Values are map units of the pixel size. If the image is not georeferenced, then pixel units are reported.
13	NUMHOLES	The number of holes in the polygon. Integer value.
14	HOLESOLRAT	The ratio of the total area of the polygon to the area of the outer contour of the polygon. The hole solid ratio value for a polygon with no holes is 1.0. $HOLESOLRAT = AREA / \text{outer contour area}$

Table 10-4. Geometric descriptors implemented in the ENVI EX 5.0 commercial software product.

# 11 Manuscript 8 (never submitted to a peer-reviewed process, made available in the public archive arXiv:1701.01942): Multi-spectral Image Panchromatic Sharpening – Outcome and Process Quality Assessment Protocol

## Motivation and Contributions to the Dissertation

To date no “universal” perceptual visual quality metric exists between a reference image and a test image. A special case of this perceptual image-pair comparison problem is when the test image is a multi-spectral (MS) image generated by panchromatic sharpening from a coarse-spatial resolution MS image and a fine-spatial resolution panchromatic image. In the present Chapter 11 (Manuscript 8) an original outcome and process quality assessment protocol for multi-spectral (MS) image panchromatic sharpening is proposed to comply with the Quality Assurance Framework for Earth Observation (QA4EO) *Cal/Val* requirements. In this application framework, reference images are a coarse-spatial resolution MS image and a fine-spatial resolution panchromatic image, while the test image is a fused panchromatic-sharpened MS image at fine-spatial resolution. Typically no process quality assessment is involved with the comparison of MS pan-sharpened images. Alternative to a traditional normalization of quality indexes adopted before index comparison and combination, one important strategy adopted by the proposed quantitative quality assessment protocol is that, before comparing and combining multiple quality indexes estimated from the same population, each individual quality index is standardized through the population to feature zero mean and unit variance, while its range of change remains unbounded above and below. As a future development of the present Chapter 11 (Manuscript 8), an innovative perceptual image-pair quality/similarity/ dissimilarity index/ metric was proposed in Chapter 3.13 (Technical report 1).

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 11 (Manuscript 8) are color filled.

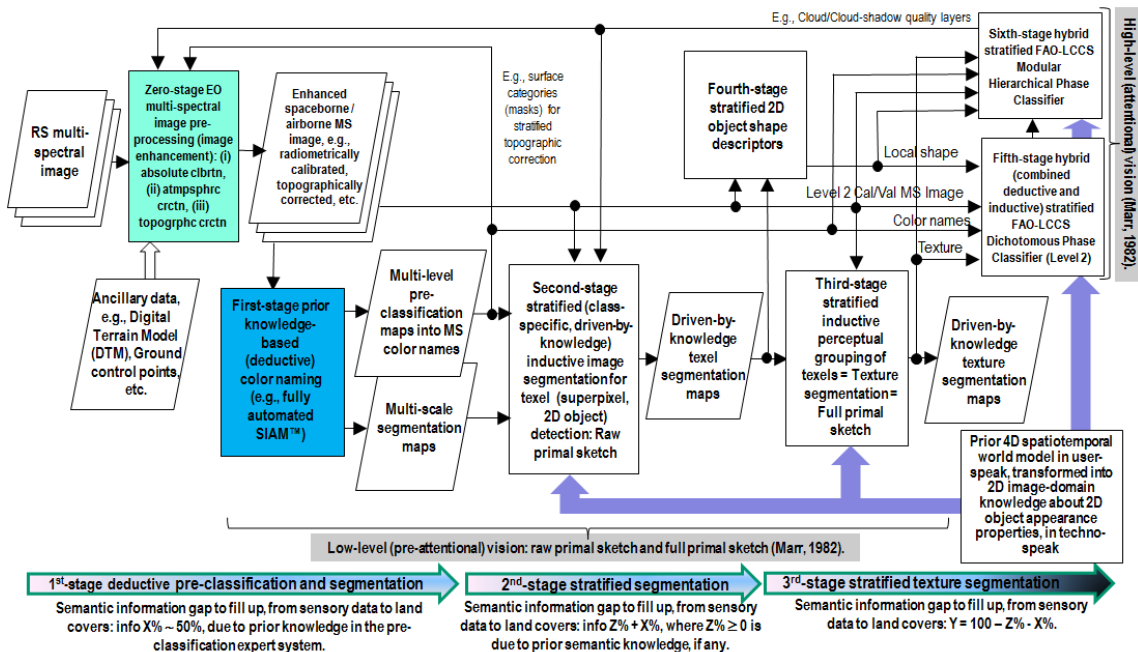


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 11 (Manuscript 8).



## Multi-spectral Image Panchromatic Sharpening – Outcome and Process Quality Assessment Protocol

Andrea Baraldi<sup>a</sup>, Francesca Despini<sup>b</sup>, and Sergio Teggi<sup>b</sup>

<sup>a</sup> A. Baraldi was with the Dept. of Geographical Sciences, University of Maryland, College Park, MD 20742, USA. He is now with the Dept. of Agricultural and Food Sciences, University of Naples Federico II, Portici (NA), Italy (e-mail: [andrea6311@gmail.com](mailto:andrea6311@gmail.com)). A. Baraldi was supported in part by the National Aeronautics and Space Administration under Grant/Contract/Agreement No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate

<sup>b</sup> Francesca Despini and Sergio Teggi were funded by the Agenzia Spaziale Italiana (ASI), in the framework of the project "Analisi Sistema Iperspettrali per le Applicazioni Geofisiche Integrate - ASI-AGI" (n. I/016/11/0). F. Despini (email: [francesca.despini@unimore.it](mailto:francesca.despini@unimore.it)) and S. Teggi (email: [sergio.teggi@unimore.it](mailto:sergio.teggi@unimore.it)) are with the Dept. of Engineering "Enzo Ferrari" (DIEF), University of Modena and Reggio Emilia, Italy.

### Abstract

Multi-spectral (MS) image panchromatic (PAN)-sharpening algorithms proposed to the remote sensing community are ever-increasing in number and variety. Their aim is to sharpen a coarse spatial resolution MS image with a fine spatial resolution PAN image acquired simultaneously by a spaceborne/airborne Earth observation (EO) optical imaging sensor pair. Unfortunately, to date, no standard evaluation procedure for MS image PAN-sharpening outcome and process is community-agreed upon, in contrast with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines proposed by the intergovernmental Group on Earth Observations (GEO). In general, process is easier to measure, outcome is more important. The original contribution of the present study is fourfold. First, existing procedures for quantitative quality assessment (Q<sup>2</sup>A) of the (sole) PAN-sharpened MS product are critically reviewed. Their conceptual and implementation drawbacks are highlighted to be overcome for quality improvement. Second, a novel (to the best of these authors' knowledge, the first) protocol for Q<sup>2</sup>A of MS image PAN-sharpening product and process is designed, implemented and validated by independent means. Third, within this protocol, an innovative categorization of spectral and spatial image quality indicators and metrics is presented. Fourth, according to this new taxonomy, an original third-order isotropic multi-scale gray-level co-occurrence matrix (TIMS-GLCM) calculator and a TIMS-GLCM texture feature extractor are proposed to replace popular second-order GLCMs.

### Index Terms

Gray level co-occurrence matrix, human vision, multi-spectral image spatial and spectral qualities, panchromatic sharpening, standard score of a raw score, third-order spatial statistics.

### 11.1 Introduction

The goal of this multidisciplinary investigation is the design, implementation and validation by independent means of a novel (to the best of these authors', the first) quantitative evaluation procedure for Earth observation (EO) multi-spectral (MS) image panchromatic (PAN)-sharpening outcome and process, in compliance with (conditioned by): (i) human vision, considered as a reference baseline, and (ii) the Quality Assurance Framework for Earth Observation (QA4EO) guidelines, delivered by the intergovernmental Group on Earth Observations (GEO) [1]. This technological research and development (TRD) project is of potential interest to the computer vision discipline and to the relevant segment of the remote sensing (RS) community whose demand for effective, efficient and easy-to-use EO image understanding systems (EO-IUSs) is ever-increasing with the quality and quantity of spaceborne/airborne EO images [2], [3].

According to the ongoing Global Earth Observation System of Systems (GEOSS) implementation plan for years 2005-2015 [4] and to the QA4EO guidelines [1], both delivered by GEO, the visionary goal of providing "the right (geospatial) information, in the right format, at the right time, to the right people, to make the right decisions" requires two necessary



and sufficient key principles to be met: “accessibility” and “suitability/reliability” of input RS data, processes and output information products. According to philosophical hermeneutics, information is meant to be either quantitative (non-equivocal) *information-as-thing* [5], e.g., values of a leaf area index are estimated from sensory data [6], or qualitative (equivocal) *information-as-data-interpretation* [7], e.g., land cover (LC) classification and LC change (LCC) detection maps are derived from EO images [8], [9]. In greater detail, the GEO’s key principle of “suitability/reliability” relies on mandatory calibration and validation (*Cal/Val*) activities, whose implementation becomes critical to sensory data, process and product quality assurance.

(i) *Cal* activities. An appropriate coordinated program of calibration activities throughout all stages of a spaceborne/airborne mission, from sensor building to end-of-life, is considered mandatory to ensure the harmonization and interoperability of multi-source multi-temporal remote sensing (RS) data [1]. By definition, radiometric calibration is the transformation of dimensionless digital numbers (DNs) into a community-agreed physical unit of radiometric measure, e.g., top-of-atmosphere (TOA) radiance (TOARD), TOA reflectance (TOARF), surface reflectance (SURF), etc.

(ii) *Val* activities. By definition, validation (not to be confused with testing, suitable for internal use) is the process of assessing, by independent means to be community-agreed upon, the “standard” quality of process and outcome [10]. In greater detail, each RS data processing stage and output product must be assigned with metrological/statistically-based (quantitative) quality indicators (QIs), to be community-agreed upon, featuring a degree of uncertainty in measurement at a known degree of statistical significance, to comply with the general principles of statistics and provide a documented traceability of the propagation of errors through the information processing chain, in comparison with established “community-agreed reference standards” [1].

Quite strikingly, the GEO’s *Cal/Val* recommendations, although based on common knowledge, are neglected or ignored in the RS common practice [8], [9]. About *Cal* activities, on the one hand, the RS community regards as baseline knowledge that “the prerequisite for physically based, quantitative analysis of airborne and satellite sensor measurements in the optical domain is their calibration to spectral radiance” ([11], p. 29). More explicitly, according to related works [8], [9], [12], [13], [14], radiometric calibration is a necessary not sufficient condition for automatic interpretation of (for physical model-based inference from) EO imagery. On the other hand, in common practice, first, the word “calibration” is absent from a large portion of papers published in the RS literature. Second, the large majority of selectable algorithms implemented in commercial EO image processing software products does not consider radiometric calibration as mandatory [14]. Relaxation of the GEO’s *Cal* constraint implies that the RS community heavily relies on statistical (inductive inference) systems, which are inherently ill-posed [15], semi-automatic and site-specific [16], whereas physical (deductive inference) models are largely neglected. Although statistical systems do not require as input sensory data provided with a physical meaning, they may benefit from *Cal* activities in terms of augmented robustness to changes in the input dataset. This is tantamount to saying that, whereas dimensionless sensory data, provided with no physical unit of measure, are eligible for use as input to statistical models exclusively, on the contrary, numerical data provided with a physical unit of measure can be input to both physical and statistical models [14]. In compliance with the QA4EO guidelines, the present work considers *Cal* activities mandatory in its further experimental Chapter 11.5, in spite of the fact it deals with statistical systems exclusively.

With regard to the GEO’s *Val* requirements, to date the RS community appears affected by a lack of standard (recognized) evaluation procedures, whose application domain ranges from (qualitative, categorical) RS image classification [12] to (quantitative) RS data fusion [17], [18], [19], [20]. In agreement with other authors [17], [21], Wald defines image fusion as “a formal framework in which are expressed means and tools for the alliance of images originating from different sources. It aims at obtaining information of a greater quality, although the exact definition of ‘greater quality’ will depend on the application” [22].

The present paper copes with the ongoing lack of a standard *Val* procedure for multi-spectral (MS) image panchromatic (PAN)-sharpening product and process [18]. MS image PAN-sharpening (merging, synthesizing) algorithms aim at taking advantage of the complementary spatial and spectral properties of MS and PAN imaging sensors [23]. Their goal is to deliver as output a fused PAN-sharpened MS image,  $MS^*_{i,b}$ , by injecting into a coarse spatial resolution MS image,  $MS_{l,b}$ , with  $b = 1, \dots, B$ , where  $B$  is the number of spectral channels and  $l$  stands for low scale factor (by definition, scale factor =  $1 / \text{spatial resolution}$  [24]), the high-pass spatial details conveyed from a fine spatial resolution PAN image,  $P_h$ , where  $h > l$  stands for high scale factor and where the two sensory images,  $MS_l$  and  $P_h$ , are assumed to be acquired (nearly) simultaneously and to depict the same Earth surface. Typical spatial resolutions of a spaceborne MS and PAN imaging sensor pair range from low ( $> 500$  m, e.g., Meteosat Second Generation, MSG) to medium (from 30 m to 500 m, e.g., Landsat-8), high ( $< 30$  m to 5 m, e.g., SPOT-4, SPOT-5, IRS-1C/D LISS III, EO-1 ALI) and very high ( $< 5$  m, e.g.,





IKONOS-2, QuickBird-2, GeoEye-1, WorldView-2, WorldView-3, PLEIADES-1A/B, SPOT-6/7, FORMOSAT-2) for the MS imaging sensor, while its PAN counterpart features a spatial resolution finer by a factor of two (e.g., Landsat-8, SPOT-4, SPOT-5), three (e.g., MSG, EO-1 ALI) or four (e.g., IKONOS-2, QuickBird-2, GeoEye-1, WorldView-2, WorldView-3, PLEIADES-1A/B, SPOT-6/7, FORMOSAT-2, IRS-1C/D LISS III).

An  $MS_h^*$  image synthesized at fine spatial resolution and featuring ‘high spectral quality’ (whatever this definition means, in line with Wald [22]) is considered crucial for most RS image applications based on the analysis of spectral signatures, from stratigraphic and lithologic mapping [25], to soil and vegetation analysis [26], [27], to digital surface model (DSM) correction [23]. For example, in [23], a 2D elevation map is generated from a stereo PAN image, then it fits also on the PAN-sharpened MS image. In the 2D elevation map, some image-objects (planar segments) are typically affected by no-data, e.g., due to occlusion phenomena. To correct each no-data pixel value in the height map, a neighboring pixel is searched for in the PAN-sharpened MS image which has the most similar color (in all spectral bands) and a non no-data value in the height map. This colorimetric best fitting neighbor’s height value is used to fill the missing value in the DSM.

Unfortunately, the peculiar nature of the MS image PAN-sharpening problem requires that no sensory (non-synthesized) MS image “truth” at high spatial resolution,  $MS_h$ , exists for comparison with the PAN-sharpened MS outcome,  $MS_h^*$ . If it were not so, the MS image PAN-sharpening problem would cease to exist. It means that MS image PAN-sharpening is an inherently ill-posed problem in the Hadamard sense, whose solution does not exist or is not unique or, if it exists, it is not robust to small changes in the input dataset [28]. As such, it is difficult to solve and requires *a priori* knowledge, in addition to sensory data,  $P_h$  and  $MS_i$ , to become better posed for numerical treatment [15], [29]. Since MS image PAN-sharpening is inherently ill-posed, so it is the quantitative quality assessment ( $Q^2A$ ) of PAN-sharpened MS imagery. This explains why the latter, too, is a much debated issue [18], [19], [20]: due to a lack of interdisciplinary background, the RS community may keep looking for a single “best” quantitative (objective) solution of an inherently ill-posed (visual, cognitive, qualitative, equivocal) problem, where no single “best” solution exists.

To recapitulate, the objective of the present study is to fill the information gap in  $Q^2A$  of MS image PAN-sharpening outcome and process, subject to (conditioned by) the following constraints, required to make the inherently difficult (ill-posed) problem at hand better posed for numerical treatment: problem solution(s), if any, must comply with, first, well-known functional principles of human vision, considered as a reference baseline [30]; second, with perceptual visual quality, assessed by human subjects under controlled experimental conditions; third, with the *Cal/Val* requirements, proposed by GEO in the QA4EO guidelines to be enforced by the RS community [1]. In general, process is easier to measure, outcome is more important. Provided with a relevant survey value, the proposed multidisciplinary investigation is of potential interest to the computer vision community, which includes RS scientists and practitioners involved with EO-IUS activities, see Fig. 11-1 [31], [32].

The rest of this paper is organized as follows. Chapter 11.II provides the problem background. In Chapter 11.3, existing PAN-sharpened MS image quality estimation procedures are critically revised. Materials and methods adopted in the experimental session are described in Chapter 11.4, where a novel protocol for  $Q^2A$  of the MS image PAN-sharpening outcome and process is proposed. Experimental results are presented in Chapter 11.5 and discussed in Chapter 11.6. Conclusions are reported in Chapter 11.7.

## 11.2 Problem Background

According to Chapter 11.I, MS image PAN-sharpening algorithms form a subset of the parent-class of inductive (bottom-up) data learning algorithms for function regression [15], [33], where no target 2D function,  $MS_h$ , exists. Hence, an output product,  $MS_h^*$ , must be synthesized (extrapolated) based on *a priori* knowledge (assumptions) in addition to sensory data. In the machine learning discipline, it is common knowledge that inductive data learning problems (either supervised data learning for function regression or classification [15], [29], or unsupervised data learning for vector quantization [34], [35], [36], [37], [38], vector clustering [39], [40], [41], [42], density function estimation or entropy maximization [39]) are inherently ill-posed in the Hadamard sense [28]. It means they are difficult to solve and require prior (top-down, deductive) knowledge in addition to data to become better posed for numerical treatment [15], [29]. By definition, *a priori* knowledge is any knowledge available in addition (“from the earlier”, top-down) to the (quantitative) dataset at hand. In common practice, inductive data learning algorithms are semi-automatic (depending on system’s free-parameters to be user-defined) and site-specific (depending on training data to learn from, by induction) [16], [43]. Typically, an inherently ill-posed inductive data learning algorithm is provided with prior knowledge in three forms [15]: (i) at the level of understanding of system design (architecture), where an inference function (e.g., Bayesian inference) is selected for



maximization/minimization purposes, (ii) at the level of understanding of system algorithm, where a class of approximating functions (e.g., radial basis function, polynomial function, etc.) is selected together with a model complexity term, e.g., a term capable of regularizing (smoothing) the function regression solution to avoid (exact) function interpolation, (iii) in the initialization phase, when the system's free-parameters are user-defined based on heuristic (qualitative) criteria, which decreases the degree of automation of the statistical model. In addition to these traditional forms of prior knowledge by the parent-class of inductive data learning systems, the special subcategory of MS image PAN-sharpening algorithms requires an *a priori* model of the target 2D function,  $MS_h$ , to be approximated by the fused image,  $MS^*_h$ . To recapitulate, the subcategory of MS image PAN-sharpening algorithms is “more” ill-posed than traditional inductive data learning algorithms for function regression. Since it lacks a quantitative “truth” to approximate, the former subcategory rather belongs to the class of inherently ill-posed cognitive problems, like vision (image understanding) in general [44], [45], and early-vision in particular [46], e.g., image segmentation [47], where there is no known cost function to minimize. This consideration justifies the interdisciplinary scenario sketched in Fig. 11-1. If these inter-disciplinary relationships hold, but are not fully acknowledged by individual scientific communities, consequences may be dreadful. For example, due to an underestimation of the inherent complexity (ill-posedness) of cognitive problems, an ever-increasing number of alternative MS image PAN-sharpening algorithms is expected to be submitted for consideration for publication in the RS and computer vision literature in the close future, exactly like tens of “novel”, supposedly “better”, inherently ill-posed image segmentation and contour detection algorithms are being published each year. Yet-another “better” solution in a class of (inherently ill-posed) inductive data learning algorithms, where no “single best solution” exists, means that alternative solutions differ one another in the degree of prior knowledge employed to become better conditioned for numerical treatment. Hence, when dealing with inductive learning-from-data algorithms, the focus of scientific attention for discrimination and quality improvement should shift from algorithms to initial conditions, consisting of an *a priori* (deductive) knowledge available in addition to data.

As reported in Chapter 11.I, the recognition by the RS community of standard procedure(s) for Q<sup>2</sup>A of PAN-sharpened MS images is a controversial problem whose solution would be of the utmost importance [18], [19], [20], in accordance with the QA4EO recommendations [1]. In general, it is well established that any data enhancement process (data pre-processing stage), including image fusion, whose input and output variables are quantitative (*information-as-thing*, refer to Chapter 11.I), is required to assess the quality of the output data (expected to be of “greater quality” [22]) in comparison with the quality of the input dataset(s), according to a (dis)similarity metric [44]. Unfortunately, in the specific case of Q<sup>2</sup>A of PAN-sharpened MS imagery, this comparison is particularly difficult because, in the absence of a full-resolution image “truth”,  $MS_h$ , the sensory image pair,  $P_h$  and  $MS_i$ , and the output data product,  $MS^*_h$ , to be compared feature a different spatial or spectral resolution [48]. About image QIs and quality metrics, the following general considerations hold.

### 11.2.1 Quantitative Image Quality Metrics: Signal Fidelity Measures and Perceptual Visual Quality Metrics

In the words of Iqbal and Aggarwal: “frequently, no claim is made about the pertinence or adequacy of the digital models as embodied by computer algorithms to the proper model of human visual perception... This enigmatic situation arises because research and development in computer vision is often considered quite separate from research into the functioning of human vision. A fact that is generally ignored is that biological vision is currently the only measure of the incompleteness of the current stage of computer vision, and illustrates that the problem is still open to solution” [30].

Objective (quantitative) quality evaluation for images and video can be classified into two board types: signal fidelity measures and perceptual visual quality metrics (PVQMs) [49], [55].

The signal fidelity measures refer to the traditional MAE (mean absolute error), MSE (mean square error), SNR (signal-to-noise ratio), PSNR (peak SNR), etc. Although they are simple, well defined, with clear physical meanings and widely accepted, signal fidelity measures can be a poor predictor of perceived visual quality, especially when the noise is not additive. For example, MAE and MSE are pixel-by-pixel differences, i.e., these statistics are non-contextual and position-dependent. Since they consider a (2D) image as a (0D) string of pixels, i.e., they ignore contextual image information, therefore they are inconsistent with visual perception. In addition, being image position-dependent, they are sensitive to image rotations.

According to a relevant portion of the computer vision literature, *the primary use of image quality metrics is to quantitatively measure an image quality that correlates with perceptual visual quality*. So-called perceptual visual quality metrics, PVQMs, are *objective models for predicting subjective visual quality scores*, like the resultant mean opinion score (MOS) obtained by many observers through repeated viewing sessions [47], [50], [51], [55]. In spite of the recent progress in related fields, objective evaluation of picture quality in line with human perception is still a long and difficult odyssey due to the complex, multi-disciplinary nature of the problem (related to physiology, psychology, vision research and



computer science) [55]. For example, cognitive understanding, prior knowledge and interactive visual processing (e.g., eye movements) influence the perceived quality of images; this is the so-called cognitive interaction problem [61]. A human observer will give different quality scores to the same image if s/he is provided with different instructions. Prior information regarding the image content, or attention and fixation, may also affect the evaluation of the image quality. But most image quality metrics do not consider these effects, they are difficult to quantify and not well understood [61]. It is clear that, unlike so-called signal fidelity measures, PVQMs have to quantify the spatial difference (e.g. Position difference in image contours) together with the spectral difference (e.g., image-wide difference in spectral means) between a reference and test image pair [12], [52], [53], [54]. There are two major categories of PVQMs with regard to reference requirements: double-ended and single-ended. Double-ended metrics require both the reference (original) signal and the test (processed) signal, and can be further divided into two subclasses: reduced-reference (RR) metrics that need only part of the reference signal and full-reference (FR) ones that need the complete reference signal. Single-ended metrics use only the processed signal, and are therefore also called no-reference (NR) ones. Most existing PVQMs are FR ones [55], e.g., the popular univariate (one-channel) “universal” (scalar) image quality index (UIQI), or Q index for brevity [60], which was further generalized into the so-called structural similarity (SSIM) index [55], [61]. Noteworthy, although SSIM is considered a PVQM, it does not appear to be provided with a perceptual relevance on a strong theoretical ground [55], in fact SSIM bears both a statistical link and a formal connection with traditional signal fidelity measures, such as the conventional pixel-based MSE [137]. Important conclusions reported in [137] are quoted as follows: “In both an empirical study and a formal analysis, evidence of a relationship between the increasingly popular SSIM and the conventional MSE is uncovered. This research is perhaps the first to uncover a statistical link of this nature and likely the only in which a formal connection is established... Collectively, these findings suggest that the performance of the SSIM is perhaps much closer to that of the MSE than some might claim. Consequently, one is left to question the legitimacy of many of the applications of the SSIM. Ultimately, this investigation once again illustrates the enormous gap that continues to exist between an automated measure of image quality and that of the human mind. Until a more radical approach is considered, this problem will likely continue to confound researchers in the field.”

To recapitulate, in a PVQM, quantitative spatial and spectral (2D) image QIs must to be estimated jointly, to be validated by the MOS collected from a group of human subjects [55], [61], e.g., refer to [18] for a detailed description of a visual analysis of PAN-sharpened MS images.

### 11.2.2 Non-Injective Property of Summary (Gross) Characteristics

It is common knowledge that any QI (or summary statistic) is inherently non-injective [56]. The non-injective property of summary statistics or (gross) QIs means that no “universal” QI can exist, because two different instantiations of the same target complex phenomenon can feature the same summary statistic. For example, Zhang and Lu duly observe that semantically “simple” (intuitive to use) planar (2D) shape descriptors are not suitable as standalone descriptors, but a combination of descriptors (feature/error pooling [55], [61], [134]) is necessary in order to accurately describe planar shapes [57]. In economic studies, the popular gross domestic product should never be considered *per se*, but in a minimally dependent and maximally informative (mDMI) [58] combination with other QIs like, for example, the Gini index, estimating the inequality of income or wealth, the pollution/environmental quality, etc. [59]. In practice, the design and development of an mDMI vocabulary of QIs allows to enforce a convergence-of-evidence approach, which is a key decision strategy in cognitive systems [44], [58]. The apparently well-known non-injective property of any QI is in contrast with a search for “universal” QIs traditionally pursued by significant portions of the scientific community. For example, the popular univariate (one-channel) “universal” (scalar) image quality index (UIQI), or Q index for brevity [60], which was further generalized into the so-called structural similarity (SSIM) index [55], [61], were both developed by the computer vision community. Evaluation procedures for PAN-sharpened MS outcome, based on the “universal” Q index, are widely adopted by the RS community [17], [48], [52], [62], [63], [64]. In greater detail, the four-channel “universal” image quality index Q4 and its extension to 2n bands, Q2<sup>n</sup> [64], are multivariate generalization of the popular univariate Q index [55], [60]. Like Q, its Q4/Q2<sup>n</sup> extensions are logical AND-combinations of three different factors [17], [52], [62], [63], [64]. The first is the modulus of the hypercomplex (multivariate pixel-based) correlation coefficient in range [1-, 1]. The second and third terms measure, respectively, the normalized degree of similarity (in range [0, 1]), estimated across spectral bands, between two univariate (one-channel) means and two univariate standard deviations. In practice, any so-called “universal” Q/Q4/Q2<sup>n</sup> index is a (weighted) mixture (e.g., a logical AND-combination) of heterogeneous QIs, each featuring its own unit of measure (if any), domain of change and sensitivity to changes in input data, into one “ultimate” (universal) scalar QI, to be dealt with by univariate analysis. While pursuing dimensionality reduction, the Q/Q4/Q2<sup>n</sup> indexes can cause an information loss. To comply with the non-injective property of QIs, a viable strategy alternative to



searching for a “universal” QI (which cannot exist) is to develop an mDMI set of individual QIs (independent random variables) to be dealt with by multivariate analysis [58]. In this context, multivariate analysis of heterogeneous QIs is intended as synonym of a convergence-of-evidence approach [44], such that converging sources of weak (fuzzy), but independent evidence allow to infer strong conjectures [43], in accordance with the general principles of fuzzy logic [65], [66].

### 11.2.3 Multi-scale Image Statistics

Perceptual image quality is inherently multi-scale in the 2D spatial domain, known that human pre-attentive vision adopts at least four spatial scales of analysis to capture non-stationary planar (2D) statistics [67], [68], [69], [70]. It means that no “best” spatial scale exists in vision. Rather, a single well-designed battery of multi-scale spatial filters is necessary and sufficient to solve any possible visual problem [49]. Due to the central limit theorem, any “big data” distribution, like a summary (gross) statistic estimated image-wide from non-stationary local statistics, tends to have a Gaussian shape, where individual contributions of independent (non-stationary) random variables (like basis functions) become indistinguishable from the whole [13]. In other words, global (image-wide) statistics are likely to average over non-stationary local patterns in data. For example, global (image-wide) bivariate Pearson’s correlation coefficient (PCC) values are scale invariant only when the original image pair is strongly correlated. Otherwise, PCC values may change with spatial scale (e.g., simulated by image resampling) by more than 20%. In [52], [54], [60], it was observed that if the image-wide PCC statistic is replaced with the average of spatially local PCC values, then the latter computation is much less sensitive to changes in scale. Noteworthy, the size of the moving window required to estimate spatially local PCC/Q/Q4/ Q2<sup>n</sup> values is one system’s free scale parameter. Unfortunately, first, it must be user-defined based on heuristics. Second, no single spatial scale is sufficient to solve visual problems, different from toy problems [67], [68], [69], [70].

Unlike bivariate PCC statistics, popular univariate summary statistics, like image mean and standard deviation also employed in the Q/Q4/ Q2<sup>n</sup> indexes [60], are more robust to changes in scale, which is easy to prove when the resampling algorithm is the nearest-neighbor.

### 11.2.4 Yellott's Theory of Low-Level Vision for Texture Discrimination: The Triple Autocorrelation Uniqueness (TAU) Theorem

The long-disproved Julesz conjecture concerning texture discrimination in biological vision states that pre-attentive discrimination of textures is possible only for textures that have different 2nd-order autocorrelation statistics (univariate statistics of the 2nd-order in the spatial domain). Many counter-examples to this theorem have subsequently been discovered by Julesz and co-workers as well as by other independent researchers [71], [72], [73], [74]. In other words, it is possible to construct pairs of physically distinct texture images whose 2nd-order univariate statistics are exactly identical. This simple background knowledge found in existing literature has an important practical consequence: it implies that popular 2<sup>nd</sup>-order spatial statistics, extracted from a gray-level co-occurrence matrix (GLCM) implemented in nearly all existing RS image processing software toolboxes, are inadequate for texture assessment and comparison purposes [75], [76]. Actually, in a more recent paper Yellott appeared to reintroduce the validity of 2<sup>nd</sup>-order spatial statistics, by proving that every discrete, finite image is uniquely determined by its two-dimensional dipole histogram [135].

In the context of more recent re-thinking on this subject, Julesz synthesized his studies of pre-attentive texture discrimination as follows: “In essence, we found that texture segmentation is not governed by global (statistical) rules, but rather depends on local, nonlinear features (textons).” As a consequence, “contrary to common belief, texture segmentation cannot be explained by differences in power spectra” (which are image-wide statistics, rather than local statistics). In other words, in biological vision, the neural computations are inherently local in the 2D spatial domain; next, a spatial average is superimposed on the local computational processes. For example, the overall amount of contrast is a visually salient feature which survives this averaging process, although the precise position of each contrast element does not survive the averaging process [74].

In a more recent paper, Yellott stated the following [71].

- Given a discrete image (2D) array,  $I(c, r)$ ,  $c = 1, \dots, C$ ,  $r = 1, \dots, R$ , consisting of  $C$  columns and  $R$  rows, the discrete image-wide 1st-, 2nd-, and 3rd-order spatial statistics are defined respectively as:

$$a_{1,l} = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r), \quad (1)$$





$$a_{2,l}(n, m) = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r) I(c + n, r + m)$$

$$= \frac{1}{C \cdot R} \cdot \text{AutocorrlnFnctn},$$

(2)

$$a_{3,l}(n_1, m_1, n_2, m_2) = \frac{1}{C \cdot R} \sum_{c=1}^C \sum_{r=1}^R I(c, r) I(c + n_1, r + m_1) I(c + n_2, r + m_2) = \frac{1}{C \cdot R} \cdot \text{TripleAutoCrrlnFnctn}$$

(3)

where Eq. (2) is the so-called continuous autocorrelation function (up to a multiplicative factor), while Eq. (3) is known as the third-order continuous autocorrelation function (up to a multiplicative factor).

- In a black and white (binary) image of finite size, the image-wide third-order statistics are equivalent to the image-wide triple autocorrelation function, which is a generalization of the ordinary image-wide autocorrelation function.
- In a black and white (binary) image of finite size, the image-wide second-order statistics are equivalent to its image-wide autocorrelation function. For images with more than two gray levels, this equivalence breaks down, i.e., two images can have the same autocorrelation function, but different 2nd-order statistics.
- Discrimination between textured images of finite size becomes increasingly difficult as their image-wide third-order statistics become more similar.
- The Yellott's Triple Autocorrelation Uniqueness (TAU) theorem states that every panchromatic (one-channel multi-gray leveled) image of finite size is uniquely determined (up to spatial translation) by its image-wide third-order statistics. Let's consider the two following statements.
  - o Statement A: Two panchromatic images of finite size are visually identical (up to spatial translation).
  - o Statement B: Two panchromatic images feature identical image-wide third-order statistics.

The TAU principle affirms that statement B is a necessary condition of statement A, i.e., statement A implies statement B. On the other hand, statement B is a sufficient condition of statement A, i.e., statement B implies statement A.

- Identical image-wide third-order statistics imply identical image-wide 2nd-order statistics. In commenting Yellott's work, Victor observes the following [74].
  - The TAU theorem is computed image-wide, i.e., it applies to images of finite size, while the Julesz conjecture applies to textures conceived as a single infinite image or as an infinite ensemble of finite images (which relates to the property of ergodic textures, such that averages performed over the infinite ensemble of textures can be replaced by spatial averages over a single spatially infinite image extracted from the ensemble). Thus, the TAU theorem does not apply to texture ensembles, i.e., it does not trivialize the Julesz conjecture based on local, rather than global statistics. In practice, TAU, which refers to image-wide third-order statistics in images of finite size, does not hold true.
  - Biological vision consists of a set of ill-posed problems, such as shape from shading, shape from texture, structure from motion, etc. [46]. Due to the inherent ill-posedness of the (3-D) scene reconstruction from (2D) imagery, the visual system necessarily makes inferences from partial (incomplete) information, and the discovery of how these inferences are made is what the study of biological vision is all about [44].

By combining the TAU theorem with the inherently ill-posed problem of texture segmentation in pre-attentive vision whose neural computations are inherently local [49], [67], [68], [69], [70], [74], a new version of the Julesz conjecture, hereafter referred to as the Enhanced TAU (ETAU) theorem, is formulated as follows.

*“Two images of either finite or infinite size are visually identical (up to spatial translation) if their local, non-linear, non-specific elements (textons) of texture perception (“tokens” in the Marr’s terminology [77], where tokens are detected in the raw primal sketch of early vision) have identical third-order spatial statistics; if this occurs, it means that two different textures (homogeneous spatial distributions of tokens, detected in the full primal sketch of early vision [77]) are the same texture.”*

In this latter statement, concepts like texture element/ texton/ token and texture, where texture is defined as the visual effect generated by a spatial distribution of tokens, are necessarily vague (fuzzy), to account for the inherent ill-posedness of pre-attentive vision [46], [77]. Analogously, the same vagueness holds in the inherently ill-posed early-vision process of texture detection (texture segmentation), dealt with by the pre-attentive visual second stage, known as full primal sketch [46], [77].

A simple relationship between the aforementioned ETAU thesis and biological vision reinforces the former speculation.



To date, the human visual system can be seen as a huge puzzle with a lot of missing pieces. Even in the first processing layers of the primary visual cortex (PVC, area V1 of the visual cortex, striate cortex) there remain many gaps, in spite of knowledge acquired by neuroscience [67], [68], [69], [78]. In part, these information gaps are being filled by developing and studying computational models. For example, models of simple, complex and end-stopped cells have been implemented in the last 10 years [79], [80], [81]. However, if we require that a computational model of vision should be able to predict perceptual effects, like the Mach bands illusion, where bright and dark bands are seen at ramp edges, then the number of published vision models becomes surprisingly small [82]. In a rather schematic summary, V1 is the input layer of the visual cortex in both left and right hemispheres of the brain. It is organized in so-called cortical hypercolumns, with neighboring left-right regions which receive input—via the optic chiasm and the lateral geniculate nucleus (NGL)—from the left and right eyes, with small “islands,” called the “chromatic” blobs [67], [68], [78]. Traditionally, blobs are believed to consist of color-sensitive cells, called double-opponent cells, (apparently) non-oriented, but sensitive to colors [49]. More recent studies found that many color cells in V1 are also orientation tuned [83]. Differently from double-opponent cells in blob areas, most cells in the large interblob areas are (apparently) selective for orientation, but are not chromatic. In the interblob hypercolumns there are simple (S-)cells, complex (C-)cells and end-stopped cells. Complex cells are thought to receive convergent excitatory connections from several simple cells [67]. A major difference between S- and C-cells is that the former are quasilinear while the latter exhibit a clear second-order nonlinearity [84], [85]. There is general agreement that S- and C-cells serve for line and edge extraction, to accomplish object segregation, categorization and recognition [79], [80]. Unfortunately, there are tens of different computational models trying to explain how S- and C-cells interact for line and edge extraction, e.g., refer to [79], [80], [81], [82], [84], [85].

About end-stopping, there seems to be no sharp distinction between end-stopped and not end-stopped cell populations. Furthermore, end-stopped cells show the well-known characteristics of either simple or complex cells. All this suggests that end-stopping is an attribute added to the simple and complex types [79], [80]. End-stopped cells respond to singularities, like line/edge crossings, vertices and end points. The so-called multi-scale keypoint representation [86], accomplished via end-stopped cells, serves as Focus-of-Attention (FoA) [79], [80]. The information represented at the keypoints complements the edge representation. The edge signal is weak or undefined at points of strong 2D intensity variations such as corners or terminations produced by occlusion. The result are gaps in the contours, and false connections between foreground and background, which make an interpretation of an edge map difficult. One can see that the representation of keypoints indicates precisely these critical locations, like terminations, corners and junctions. Typically, many of the keypoints are located on occluding contours [79], [80].

Since they have a proactive role in contour detection, where they are claimed to be sufficient for edge detection by zero-crossing [77], [87], and since they provide inputs to both C-cells (whatever this cell type does) and end-stopped cells suitable for keypoint detection, S-cells are of key relevance in pre-attentive vision. Typically, they are modelled by complex Gabor (wavelet) functions, or quadrature filters with a real cosine and an imaginary sine component, both with a Gaussian envelope, see Fig. 11-2 and Fig. 11-3. If the even-symmetric (real) part of a Gabor local filter is implemented like a second-order derivative of an oriented Gaussian shape, like that shown in Fig. 11-3(a), then it is: (i) suitable for detecting image contours as zero-crossings of the even-symmetric filtered image, in agreement with the Marr’s theory of early vision [77], [87], and (ii) eligible for collecting 3rd-order spatial statistics, like those envisaged by the Yellott’s ETAU principle.

To conclude, the ETAU speculation finds a physical justification in the multi-scale model of even-symmetric S-cells found in the interblob hypercolumns, as those shown in Fig. 11-2 and Fig. 11-3(a), in agreement with the Marr’s theory of early vision [75], [77]. Noteworthy, the odd-symmetric (imaginary) part of the same Gabor filter, which is equivalent to a first-order derivative of an oriented Gaussian shape, shown in Fig. 11-3(b), would be eligible for collecting 2nd-order spatial statistics, like those envisaged by the long-disproved Julesz conjecture about texture discrimination.

### 11.2.5 Criteria for Quality Improvement of Existing PAN-Sharpened MS Image Estimation Procedures

Well-grounded in common knowledge and in the existing literature (refer to Chapter 11.II.1 to Chapter 11.II.4), four criteria are proposed to be adopted in the further Chapter 11.3 for quality improvement of existing estimation procedures for PAN-sharpened MS outcome.

(I) Quantitative planar (2D) spatial and spectral QIs are estimated together with perceptual (qualitative) image quality values collected from a group of human subjects: yes/no. If no, the estimation procedure is lacking in terms of reference (prior) knowledge to be considered as “truth”.

(II) The same (homogeneous) multi-scale image statistic, e.g., a spectral local mean, is combined across spatial scales: yes/no. If yes, on theory, this combination of information (feature/error pooling [55], [61], [134]) is acceptable, because an overall information gain can be accomplished. For example, appropriate multi-scale spatial filter combinations allow



detection of color image contours [49]. In practice, any multi-scale combination of homogenous information ought to be further scrutinized at the level of understanding of algorithm implementation, to check whether or not this combination of information sources leads to an information gain. If spatial statistics are collected either pixel-based (1<sup>st</sup>-order spatial statistics) or at a single spatial scale (like local PCC/Q/Q4/Q2<sup>n</sup> indexes [55], [60], [64]), then these statistics are likely to be inadequate to capture inter-image similarities featuring up to 3<sup>rd</sup>-order spatial autocorrelation properties, according to the Yellott's ETAU principle (refer to Chapter 11.II.4).

(III) Multiple heterogeneous statistics, e.g., image-wide spectral mean and variance, each featuring its own unit of measure (if any), domain of variation and sensitivity to changes in input data, are combined into a "universal" QI, which cannot exist, due to the non-injective property of summary statistics: yes/no. This mixture of heterogeneous random variables into one scalar "universal" QI is, in general, subject to a loss of information, due to dimensionality reduction; hence, in general, it is theoretically inconvenient, in particular when either of the following conditions occurs.

- It is based on a heuristic (subjective, equivocal) weighted combination of individual terms, where weights are user-defined based on empirical criteria. These weights increase the number of system's free-parameters to be user-defined. Their total number is monotonically decreasing with the system's degree of automation (ease of use) [8], [9].
- Heterogeneous terms are combined without harmonization of their units of measure, domains of variation and sensitivities to changes in input data. For example, in [88], a multi-source geospatial index of climate change adopts a linear min-max normalization function to render each input dataset comparable in (normalized) range of change and (dimensionless) unit of measure. Unfortunately, a linear min-max normalization function applied to different data sources (random variables) does not harmonize their sensitivities to changes in the input dataset. The so-called z-score (standard score of a raw score, standardized variable) would be a better solution [89].

(IV) The popular bivariate PCC in range [-1, 1] is adopted as an image QI, irrespective of its local spatial scale of analysis, refer to the aforementioned point (II), whether or not it is combined with other QI indexes, refer to the aforementioned point (III): yes/no. In general, exploitation of PCC as an inter-image QI and metric should be considered theoretically inconvenient. The well-known sensitivity of the PCC to linear transformations of the two random variables means that PCC is maximum (in absolute terms) between two images that are either identical or one the linear transformation of the other although, in this latter case, they can look (perceptually) very different. Its macroscopic inconsistency with (its independence from) visual perception should discourage perceptual image (dis)similarity metrics from using the PCC as an input variable. Similar considerations led to neglect the estimation of correlation as a viable texture feature from popular 2<sup>nd</sup>-order GLCMs [75].

### 11.3 Critical Review of Existing Procedures for Q<sup>2</sup>A of PAN-Sharpended MS Images

State-of-the-art procedures for Q<sup>2</sup>A of PAN-sharpened MS outcomes belong to two families, depending on whether or not the sensory MS<sub>l</sub> image is adopted as a reference dataset.

#### 11.3.1 Abstract Three-Statement Wald's Protocol, where an Ideal Reference Image at Fine Resolution is Available for Comparison Purposes

Let us assume that together with the sensory P<sub>h</sub> and MS<sub>l</sub> images acquired simultaneously by a PAN and MS sensor pair at spatial scales  $h$  and  $l$  respectively, with  $h > l$ , an ideal reference MS<sub>h</sub> is also available as "truth", like it were acquired by the same MS sensor capable of working at high and low spatial scales simultaneously. The Wald's protocol is based on the following three PAN-sharpened MS image quality assessment criteria [48].

1. Any fused image, MS<sub>h</sub><sup>\*</sup>, if spatially degraded from scale  $h$  to  $l$ , identified as MS<sub>h->l</sub><sup>\*</sup>, should be as nearly identical as possible to the original MS<sub>l</sub> image. For example, a channel-specific difference between images MS<sub>h->l,b</sub><sup>\*</sup> and MS<sub>l,b</sub>,  $b = 1, \dots, B$ , can be computed on a per-pixel basis ([48], p. 694). This property, called the consistency property, is a necessary, but not sufficient condition for image fusion, i.e., its fulfillment does not imply a correct fusion [18]. In practice, there is an influence of the downsampling strategy upon the results of comparison between MS<sub>h->l</sub><sup>\*</sup> and MS<sub>l</sub>, but this influence can be kept small, provided the MS image downsampling operator is such that, first, an appropriate low-pass filter (LPF), whose transfer function has to match the average modulation transfer function (MTF) of the MS sensor [90], is applied to the MS image ([48], p. 694). Second, a decimation operator, characterized by a sampling factor equal to the spatial scale ratio ( $h: l$ ) between the two native scales of images [63], is applied to the low-pass filtered MS image.
2. Each band of the synthetic image MS<sub>h,b</sub><sup>\*</sup>,  $b = 1, \dots, B$ , should be as identical as possible to its ideal reference counterpart MS<sub>h,b</sub>,  $b = 1, \dots, B$ . Since this property does not cope with the entire set of channels simultaneously, then a third consistency property is required.
3. As a whole (i.e., when all channels are examined simultaneously), the synthetic MS<sub>h</sub><sup>\*</sup> image should be as identical as



possible to the ideal reference image  $MS_h$ .

### 11.3.2 Quantitative Analysis with the Sensory $MS_l$ Image Adopted as Reference: Revised Two-Statement Wald's Protocol and Its One-Statement Simplified Version

In [48], the second and third virtual properties of the abstract Wald's protocol are implemented as follows.

- Second property proposed in Chapter 11.3.1. If the original input images,  $P_h$  and  $MS_l$ , are degraded as  $P_{h \rightarrow s}$  and  $MS_{l \rightarrow s}$ , where the spatial scale  $s$  is such that  $s < l < h$ , and if a PAN-sharpened  $MS$  image,  $MS^*_l$ , is synthesized at the native spatial scale  $l < h$  starting from degraded images  $P_{h \rightarrow s}$  and  $MS_{l \rightarrow s}$ , then the fused image,  $MS^*_l$ , should be as nearly identical as possible to the original  $MS_l$  image considered as reference. It is recommended that spatial scale ratio ( $l : s$ ) is chosen equal to ( $h : l$ ). This is a realistic strategy to check in practice the synthesis property [18], [20], [48]. To obtain  $MS_{l \rightarrow s}$ , the same constraints about the  $MS$  image degradation filter listed in Chapter 11.3.1 hold. In addition, the PAN image degradation filter used to generate  $P_{h \rightarrow s}$  is typically designed as an ideal filter [90].
- Third property proposed in Chapter 11.3.1. Assuming that the high-frequency (fine resolution) spatial information is conveyed into the  $MS$  image to be synthesized at fine resolution,  $MS^*_h$ , by the sensory PAN image,  $P_h$ , a realistic spatial quality assessment of the fused image requires the difference to be ideally null between: (i) bivariate PCCs computed between the downsampled PAN image,  $P_{h \rightarrow l}$ , and each band of the synthesized image,  $MS^*_{l,b}$ ,  $b = 1, \dots, B$ , and (ii) bivariate PCCs computed between the same downsampled PAN image,  $P_{h \rightarrow l}$ , and each band of the original image  $MS_{l,b}$ , with  $b = 1, \dots, B$  (see [48], p. 695).

A typical simplified implementation of the Wald's protocol consists of the aforementioned second property exclusively.

Unfortunately, the practical choice of the pair of LPFs applied to the sensory  $MS_l$  and  $P_h$  images for downsampling is crucial in these realistic adaptations of the ideal Wald's protocol. An erroneous choice of these LPFs may lead to mismatches between the  $Q^2A$  of the image fusion at reduced resolution,  $MS^*_l$ , and the quality, perceived by visual inspection exclusively (due to the absence of an  $MS_h$  image "truth"), of the image fusion outcome at full resolution,  $MS^*_h$  [91]. This holds true particularly in the case of  $MS$  image PAN-sharpening methods exploiting spatial filters [92].

The proposed realistic adaptation of the abstract three-statement Wald's protocol requires exploitation of both so-called "scalar" QIs (i.e., statistics estimated in a single channel) and so-called "vector" QIs (i.e., statistics estimated in all spectral channels simultaneously), together with (dis)similarity metrics [63]. In [18], one-channel "scalar" statistics are otherwise called "unimodal", whereas multi-channel "vector" statistics are otherwise called "multimodal". In the rest of the present work, these expressions are replaced by terms "univariate" and "multivariate" respectively. Examples of univariate statistics are relative bias (difference in mean), difference in variances, relative difference in standard deviation, and many others [50]. The well-known PCC is a bivariate statistic. The popular UIQI, typically identified as  $Q$  index, combines into one scalar value several heterogeneous statistics including PCC [60]; hence, the  $Q$  index is also a bivariate statistic. Well-known examples of multivariate summary statistics are  $Q4$ , as a generalization of  $Q$  [62], and  $Q2^n$  as a generalization of  $Q4$  [64], the relative dimensionless global error (ERGAS, Erreur Relative Globale Adimensionnelle de Synthèse) [93], the average spectral angle mapper (SAM) cost index [94] and many others [50]. In [53], image QIs are divided into either spectral or spatial, where examples of the latter category are the Zhou spatial correlation coefficient (ZCC) [95] and the true edge (TE) detector [96].

In the simplified implementation of the Wald's protocol, consisting of the aforementioned second property exclusively, univariate QIs can be omitted, i.e., multivariate QIs can be adopted exclusively.

To make this paper self-contained, the popular univariate  $Q$  index and the multivariate SAM and ERGAS cost indexes are presented hereafter. The SAM formulation computes the inter-vector angle between two data vectors  $\vec{x}$  and  $\vec{y}$  as:

$$SAM(\vec{x}, \vec{y}) = \arccos \left( \frac{\langle \vec{x}, \vec{y} \rangle}{|\vec{x}| \cdot |\vec{y}|} \right),$$

(4)

where  $|\cdot|$  and  $\langle \cdot \rangle$  indicate, respectively, the Euclidean norm and the scalar vector product [94]. In the application domain of  $MS$  image PAN-sharpening, SAM is adopted to quantify the inter-image pixel-specific  $MS$  difference (distorsion), irrespective of the pixel-pair difference in modula (color intensities). An image-wide SAM statistic is the average of the pixel-based SAM values [63], [94]. If there is no inter-image spectral distorsion, then average SAM is zero. It means that, in  $MS$  image PAN-sharpening applications, average SAM is a cost (error) index to be minimized.

The ERGAS cost (error) index is a heuristic multivariate estimate of a dimensionless pixel-based inter-image difference adopted by several procedures for  $Q^2A$  of PAN-sharpened  $MS$  outcome [18], [63], [97]. It is defined as:





$$ERGAS = 100 \cdot \frac{h}{l} \sqrt{\frac{1}{B} \sum_{b=1}^B \frac{RMSE(MS_b)^2}{(Mean_b)^2}}, \quad (5)$$

where  $Mean_b$  is the mean value of the sensory image  $MS_{l,b}$ ,  $b = 1, \dots, B$ , while the root mean square error term,  $RMSE(MS_b)$ , is defined as:

$$RMSE(MS_b) = \sqrt{\frac{\sum_{i=1}^{NP} (MS_{l,b}(i) - MS_{l,b}^*(i))^2}{NP}}, \quad (6)$$

where  $i$  is a pixel identifier and  $NP$  is the total number of pixels. According to Wald, ERGAS exhibits a strong tendency to decrease as the inter-image similarity increases. Typical ERGAS values of “good inter-image quality” range below 3 [93].

In the computer vision literature, Wang and Bovik [60] proposed a so-called “universal” (combined scalar value from heterogeneous statistics) image quality index, UIQI, identified as  $Q$ . This is a similarity metric instantiated as an AND-combination of heterogeneous bivariate and univariate statistics extracted from a pair of one-channel images  $x$  and  $y$ , such that it is maximum when the two one-channel images are the same. In particular:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}, \quad (7)$$

where  $\sigma_{xy} = \frac{1}{NP-1} \sum_{i=1}^{NP} (x_i - \bar{x})(y_i - \bar{y})$  is the covariance between one-channel images  $x$  and  $y$ , with  $\bar{x} = E[x]$  and standard deviation  $\sigma(x) = \sigma_x \geq 0$ . In Eq. (7), the first term is the popular Pearson’s cross-correlation coefficient,  $PCC = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

$\in [-1, 1]$ , whose absolute value increases if there is a linear relationship between the two univariate random variables  $x$  and  $y$ . The second term is the normalized similarity (belonging to range  $[0, 1]$ ) between the image means (called luminance), which is equal to one if and only if the two univariate means are the same. The third term is the normalized similarity (in range  $[0, 1]$ ) of the two image standard deviations (called contrast), which is equal to one if and only if the two univariate standard deviations are the same. Although it is indeed practical to estimate an inter-image similarity as an overall (image-wide) scalar QI value, an inter-image (dis)similarity measure is typically space variant because image signals are generally nonstationary in the 2D spatial domain. Hence, it is more appropriate to measure the  $Q$  index values locally, e.g., based on a non-overlapping moving window of size  $W1 \times W2$  in pixel unit, to accomplish an image partition into blocks, and then combine these local values at an image-wide spatial scale, e.g., by averaging the sum of local values [60]. This is equivalent to implementing the  $Q$  index at a single-scale of analysis, whereas human vision is known to adopt at least four-scale spatial filters (refer to Chapter 11.II.3). In more recent years, the so-called SSIM index was proposed as a generalization of the bivariate  $Q$  index [61]. A multi-scale implementation of the SSIM index was proposed by the same authors [136].

Although SSIM is considered a PVQM, it does not appear to be provided with a perceptual relevance on a strong theoretical ground, in fact SSIM bears certain similarities with traditional signal fidelity measures, such as the MSE [55]. This is clearly explained in [137] whose conclusions are quoted in Chapter 11.II.1.

To account for widespread criticisms about the SSIM such as those reported in [137], Simoncelli et al. have recently proposed a PVQM based on a normalized Laplacian pyramid for image analysis and synthesis as a viable alternative to the SSIM. The proposed PVQM formulation is:

$$D(I_R, I_T) = \frac{1}{S} \sum_{s=0}^{S-1} \frac{1}{\sqrt{SF_s}} \|\hat{I}_{R,s} - \hat{I}_{T,s}\|^2,$$

where  $I_R$  and  $I_T$  are the reference and the test image respectively,  $\hat{I}_{R,s}$  and  $\hat{I}_{T,s}$  denote vectors containing the transformed reference and distorted image data at scale  $s = 0, \dots, S-1$ , respectively, and where  $SF_s$  is the number of spatial filters in the sub-band at scale  $s$ . In this equation a root mean squared error is computed for each scale, and then averaged over these scales giving larger heuristic weights to the lower frequency coefficients (which are fewer in number, due to subsampling).

Since the “universal”  $Q$  index is bivariate, i.e., it applies to two one-channel images exclusively, Alparone *et al.* [62] proposed  $Q4$  as a “universal”  $Q$  index extended to four-band image pairs. Like  $Q$ ,  $Q4$  is computed per image block and, next, averaged across blocks. Like  $Q$ , also  $Q4$  is a similarity index, to be maximized in MS image PAN-sharpening



applications, made of three factors. The first is the modulus of the multivariate (multi-band) hypercomplex PCC in range  $[-1, 1]$ . The second and third terms measure, respectively, the normalized difference across bands of univariate (one-band) mean pairs and standard deviation pairs. Typically, an image-wide Q4 value is computed as average of one-scale local Q4 values estimated across an image partition of  $16 \times 16$  or  $32 \times 32$  blocks. In recent years, Garzelli and Nencini [64] presented a “universal”  $Q^2$  similarity index as a generalization of Q4 for image pairs of more than four bands by a power of 2.

### 11.3.3 Quantitative Analysis at High Spatial Scale $h$ , Without Reference Image

In [52], [54], a new “universal” (combined from heterogeneous statistics) QI, called “quality with no reference”, QNR, is proposed for  $Q^2A$  of PAN-sharpened MS outcome. This second approach exploits no reference image, but relationships among the two sensory images,  $P_h$  and  $MS_l$ , and the synthesized  $MS_h^*$  product exclusively. Hence, it is appealing because its inputs are the two sensory images in addition to the output fused image, at their native scales. Unfortunately, this second approach strongly depends on the choice of QIs and quality metrics. In line with [60], summary statistics should be computed as image-wide averages of one-scale local estimates, to better capture the non-stationarity of image statistics. This is especially true for pixel-based bivariate PCC values that may change with scale by more than 20% (e.g., simulated by image resampling, refer to Chapter 11.II.3).

In [52], [54], [63], the “universal” QNR similarity metric is implemented as an AND-combination of a spatial similarity index with a spectral similarity index:

$$QNR = (1 - D_\lambda)^\alpha \cdot (1 - D_S)^\beta, \quad (8)$$

where  $D_\lambda$  is the spectral distortion,  $D_S$  is the spatial distortion and  $\alpha$  and  $\beta$  are two coefficients to be user-defined based on heuristics to weight the two terms. The spectral distortion is computed as:

$$D_\lambda = \sqrt[p]{\frac{1}{B(B-1)} \sum_{i=1}^B \sum_{j=1, j \neq i}^B |d_{i,j}(MS_l, MS_h^*)|^p}, \quad (9)$$

where  $p$  is a metric parameter to be user-defined based on empirical criteria (it is usually set to 1 [54]) and term

$$d_{i,j}(MS_l, MS_h^*) = Q(MS_{l,i}, MS_{l,j}) - Q(MS_{h,i}^*, MS_{h,j}^*) \quad (10)$$

is a dissimilarity measure between two “universal” one-band Q index values. The spectral cost function (9) is minimized when the inter-band heterogeneous Q combination of spectral properties at high spatial scale  $h$ , within the synthetic  $MS_h^*$  image, are the same of their spectral counterparts at low spatial scale  $l$ , within the sensory  $MS_l$  image. The spatial distortion function is defined as:

$$D_S = \sqrt[q]{\frac{1}{B} \sum_{i=1}^B |Q(MS_{h,b}^*, P_h) - Q(MS_{l,b}, P_{h \rightarrow l})|^q}, \quad (11)$$

where  $q$  is a metric parameter to be user-defined based on empirical criteria (it is usually set to 1 [54]). The spatial distortion Eq. (11) is minimized when the similarity Q index computed at high spatial scale  $h$  between each band of the fused  $MS_h^*$  image and the  $P_h$  image is equal to the similarity Q index computed at low spatial scale  $l$  between each band of the sensory  $MS_l$  image and the downsampled  $P_{h \rightarrow l}$  image. To recapitulate, QNR is a PAN-sharpened MS image QI in range  $[0, 1]$ . To maximize QNR, the spectral distortion term,  $D_\lambda$ , and the spatial distortion term,  $D_S$ , must be minimized to zero. It means that QNR is monotonically increasing with the combined spatial and spectral qualities of the fused  $MS_h^*$  product.

Alternative formulations of the spatial and spectral cost functions (9) and (11) employ QIs and quality metrics like standard deviation, entropy (He), cross entropy (CE), spatial frequency (SF), fusion mutual information (FMI), fusion quality index (FQI), fusion similarity metric (FSM), etc. [50].

## 11.4 Materials and methods

To design, implement and validate by independent means an innovative procedure for  $Q^2A$  of MS image PAN-sharpening process and outcome, the following materials and methods were selected.

### 11.4.1 Validation Dataset

According to standard review quality criteria adopted by peer-reviewed journals in computer science, experimental results are expected to be shown for a sufficient number of real and standard/appropriate data sets, typically two or more [98], [99]. For example, to assess the best among alternative MS image PAN-sharpening algorithms in terms of QIs of operativeness (QIOs), encompassing accuracy, efficiency, degree of automation, robustness to changes in the input dataset, etc. [8], [9], at least two input dataset should be considered mandatory.

Rather than selection of a best algorithm among alternative solutions, the goal of the experimental session of the present study is validation of an evaluation procedure, refer to Chapter 11.I. By definition, validation (not to be confused with



testing) is the process of assessing, by independent means, the quality of the information processing system's outputs [10]. In this work, the system under investigation is an evaluation procedure whose outputs are quantitative ranks of PAN-sharpened MS images to be validated against qualitative ranks collected from human subjects independent of the authors of the procedure under validation.

For validation of the proposed evaluation protocol and for the sake of paper brevity, only one validation sensory dataset was selected, to be representative of the complexity of the target phenomenon under investigation, namely, Q<sup>2</sup>A of MS image PAN-sharpening outcome and process. The selected sensory dataset consists of a very high resolution (VHR) spaceborne MS and PAN image pair, acquired by the QuickBird-2 imaging sensors on 2004-06-13 at 09:58 a.m. over the Campania region, Italy. The QuickBird-2 imaging sensors simultaneously acquire a PAN image at 0.61 m spatial resolution and a four-band image at 2.44 m resolution, whose spectral channels are: visible blue (B), green (G), red (R) and near-infrared (NIR). Based on ancillary calibration metadata files, both PAN and MS images were radiometrically calibrated into TOARF values, in compliance with the QA4EO requirements (refer to Chapter 11.I). The radiometrically calibrated MS and PAN images are shown in Fig. 11-4. This VHR image pair was considered appropriate because, first, it depicts a wide variety of land surface classes, ranging from urban areas to forests and agricultural fields, but also includes real-world RS image noise, where chromatic information is saturated (in white-color image areas, like clouds), null (in black-color image areas, like cloud-shadows) or fuzzy, e.g., image areas affected by haze. Second, it consists of four bands. Hence, this four-band sensory MS<sub>1</sub> image and its synthesized MS<sub>1</sub><sup>s</sup> versions were not too difficult to be visually assessed by a pool of human subjects, who could rely exclusively on a three-channel RGB monitor for image comparison and quality assessment. Last but not least, this four-band validation image allowed estimation of the popular Q4 index [62].

According to Table 11-1 (also refer to the further Chapter 11.4.2), there were fourteen PAN-sharpened MS image instances to be ranked for validation purposes by the proposed quantitative evaluation procedure in comparison with a qualitative (perceptual) assessment by human subjects, adopted as a reference ("truth") and expected to be mimicked (matched) by the proposed quantitative PVQM approach.

Actually, the proposed validation dataset of fourteen PAN-sharpened MS images, provided with perceptual visual quality ranks as "truth", was sufficient to act as counter-example where popular multivariate scalar QIs (or cost indexes), like average SAM, ERGAS and Q4, adopted by state-of-the-art procedures for PAN-sharpened MS image quality estimation [18], [63], were correlated one another, but fail to be uncorrelated with perceptual visual quality assessed by human subjects under controlled experimental conditions.

For the sake of completeness, we mention that an internal testing phase of the proposed evaluation procedure predated the validation phase, documented in this paper. During tests conducted on several PAN-sharpened MS images, the proposed quantitative estimation procedure was compared with a visual assessment (for acquisition of "truth") by the same authors of the estimation procedure. These same authors considered the experimental degree of match of quantitative test results with their own visual assessment in agreement with theoretical expectations (refer to Chapter 11.II), to be further confirmed in a validation phase by independent means.

For Q<sup>2</sup>A of MS image PAN-sharpening outcome we adopted a reduced resolution approach following the simplified one-statement Wald's protocol [93], refer to Chapter 11.3.2. In this evaluation framework, according to Vivone *et al.* [63], the original QuickBird-2 PAN and MS images were downsampled, by means of a Gaussian low-pass filter (LPF) and a decimation operator, to a spatial resolution of 2.44 m and 9.76 m respectively, to maintain the same fusion ratio (1:4) as in the original QuickBird image pair. The implemented LPF matched the average modulation transfer function (MTF) of the MS and PAN imaging sensors [90], in agreement with Chapter 11.3.2.

#### 11.4.2 MS Image PAN-Sharpening Algorithms Selected for Testing

For testing purposes we surveyed popular MS image PAN-sharpening algorithms available in several RS image processing commercial software toolboxes, specifically, ERDAS Imagine (licensed by ERDAS, Inc.) [100], Environment for Visualizing Images (EN6, licensed by ITT Industries, Inc.) and Interactive Data Language (IDL, licensed by ITT Industries, Inc.) [101]. This survey led to a selection of eight algorithms.

- Principal Component (PC) transform [102], implemented by EN6.
- Gram-Schmidt (GS) transform [101], [103], implemented by EN6.
- Color Normalized Spectral Sharpening (CN) transform [104], implemented by EN6.
- Discrete Wavelet (DWT) transform [105], [106], implemented by IDL.
- A Trous Wavelet (ATW) transform [106], [107], [108], implemented by IDL.
- Hyperspherical Color Space (HCS) [109], implemented by ERDAS.



- Ehlers (EH) transform [110], implemented by ERDAS.
- Resolution Merge (RM), [111], implemented by ERDAS.

A description of these eight algorithms is beyond the scope of this paper; interested readers can refer to literature.

For each of these algorithms, there were one or more system's free-parameters to be user-defined. One input parameter was selected for discriminative purposes, i.e., its changes in value led to different runs of the same algorithm with different outcomes. Other input parameters, if any, were kept fixed in the different runs by the same algorithm. Table 11-1 shows that different resampling methods were selected as input parameter by some of the eight algorithms, which increased to fourteen the total number of alternative MS image PAN-sharpening system implementations to be compared.

When applied to the downsampled version of the validation sensory dataset shown in Fig. 11-4, these test algorithms generated a fused MS\*<sub>1</sub> image at low spatial scale *l*, as shown in Fig. 11-5 and Fig. 11-6.

#### 11.4.3 State-of-the-Art Multivariate Scalar QIs Selected for Comparison Purposes

In order to compare image quality estimates by the new evaluation protocol with existing QIs, three multivariate (all-band) scalar QIs were selected as the most widely adopted by the scientific community in recent years.

- SAM dissimilarity index [94].
- ERGAS dissimilarity index [97].
- Q4 similarity index [62].

A review of these indexes is provided in Chapter 11.3.2. For further details, refer to the literature. About Q4, it was calculated as averages on  $BL \times BL$  image blocks, with  $BL = 8$ . Hence, Q4 depends on BL too, denoted as  $Q4_{BL}$ . Finally,  $Q4_{BL}$  was averaged over the whole image to yield the global score index Q4, in agreement with [62].

#### 11.4.4 Perceptual Image Quality Assessment

According to Chapter 11.II.1, human visual analysis is indispensable to provide the inherently ill-posed MS image PAN-sharpening problem with a reference baseline for quality estimation and comparison, also refer to Chapter 11.I. Although stated in other terms, this concept is widely acknowledged in works like [18], [50], [91]. Unfortunately, visual quality assessment of multiple images is a difficult and lengthy task to handle because, first, the human visual system is not equally sensitive to various types of distortion or color contrast in an image. Second, the perceived image quality is strongly dependent upon the observer and the thematic application at hand (*information-as-data-interpretation*, refer to Chapter 11.I). Third, technical factors, such as difficulties in image representation or evaluation, e.g., when MS images have more than three spectral bands, may undermine the validity of the experiment with human subjects. To minimize accidental and systematic errors in visual evaluation, protocols for visual quality assessment have been proposed in the fields of television and image compression [55], image segmentation [47] and EO MS image PAN-sharpening [93].

The goal of the following procedure is to estimate the resultant MOS obtained by many observers through repeated viewing sessions [55], [61]. In agreement with [47], sixteen Modena and Reggio Emilia University staff and students served as subjects, equally split by gender. None was paid for his/her participation. All were native Italian speakers and reported having normal or corrected-to-normal vision. Two categorical variables, identified as Spatial QI and Spectral QI, were assigned with seven levels, from A to G, corresponding to numbers 1 to 7, see Table 11-2. The original PAN and MS image pair and each test PAN-sharpened MS image were partitioned into  $BL \times BL$  non-overlapping blocks, with  $BL = 8$ . performance of the current quality assessment algorithms.

To account for the cognitive interaction problem, where the perceived quality of images is influenced by prior information [61], before beginning the experimental trials, the subjects received six practice trials, each consisting of the same randomly selected block extracted from all the PAN-sharpened MS images in comparison with the same block extracted from the original PAN and MS image pair, to get familiar with concepts like spatial and spectral qualities (similarities) in four-band images shown in a three-channel (RGB) monitor. The test image is required to be as identical as possible to the reference image, neither better nor worse in perceptual terms [60]. Following this practice, the subjects started on the remaining 58 blocks for spatial and spectral quality (similarity) assessment. Each PAN-sharpened MS image-block was ranked (sorted) by each subject, who was free to change band combinations shown in the RGB monitor. The two spatial QI and spectral QI distributions were estimated per image. Each distribution was standardized (to feature zero mean and unit variance) [89] and the standardized range of change was split into seven bins of the same width, labeled A to G again. Finally, a winner-take-all strategy was adopted. If the difference in score between the first-best and the second-best level was less than 10% of the first-best score, than both levels were considered winners, as shown in Table 11-2. This MOS procedure is very different from that proposed in [61], where subjects were asked to rank an ensemble of images compared with the same reference image, but were asked to provide their perception of quality of each pairwise image





comparison on a continuous linear scale that was divided into five equal regions marked with adjectives “Bad”, “Poor”, “Fair”, “Good” and “Excellent”. Raw scores for each subject were normalized by the mean and variance of scores for that subject (i.e., raw values were converted to Z-scores [54]) and then the entire data set was rescaled to fill the range from 1 to 100. MOSs were then computed for each image, after removing outliers.

#### 11.4.5 Expert System in Operating Mode for Prior Knowledge-Based MS Data Space Discretization (Partitioning)

Prior knowledge-based (top-down, deductive, physical model-based) preliminary classification (pre-classification) has an important role in the operational, comprehensive and timely generation of information products from EO “big data” [112], in compliance with the QA4Eo guidelines [1]. Documented applications of prior knowledge-based image mapping systems date back to the early 1980s [44], [45] and span from RS image enhancement (data pre-processing), like automatic stratified (conditioned, pre-classified) image co-registration, topographic correction [113], [114] and cloud masking [115], [116], [117], to second-stage (high-level) stratified (conditioned) LC classification and LCC detection [8], [9], [12], [13], [14], [44], [45], [118]. *Equivalent to color naming in a natural language* [119], [120], [121], *prior knowledge-based continuous color space discretization (compression, quantization, partitioning) is the automatic deductive counterpart of semi-automatic inductive vector quantization algorithms, like the popular k-means algorithm (also known as Linde-Buzo-Gray algorithm, LBG)* [15], [33], [34], [35], [36], [37], [38], *not to be confused with unsupervised data clustering algorithms, where termination is not based on optimizing any model of the process or its data* [39], [40], [41], [42], [122], [123], [124]. In unsupervised (unlabeled) data quantization problems, the target cost function to minimize is known and equal to a data quantization error (typically, a mean square error). In the machine learning literature, it is common knowledge that any inductive data learning problem is inherently ill-posed in the Hadamard sense [28] and requires *a priori* knowledge in addition to data to become better-posed for numerical solution [15], [33]. More specifically, in a generic data quantization error minimization problem, the quantization error is expected to be monotonically decreasing with the number of quantization levels; hence, no number of quantization levels is “optimum” *per se*. As a consequence, the number of quantization levels is typically user-defined based on subjective criteria, like in the well-known k-means unsupervised data learning algorithm, where the number of quantization levels  $k$  is an input parameter to be user-defined. Vice versa, it is possible to provide the target data quantization error as a user-defined input parameter, such that it is the system’s free-parameter  $k$  to be dynamically learned from data by the inductive data quantization algorithm [34], [35].

The expert system for prior knowledge-based MS data quantization selected in this study was the Satellite Image Automatic Mapper (SIAM), proposed to the RS community in recent years [8], [9], [12], [13], [14], [125]. Since it is based on prior knowledge, SIAM is: (i) independent of (non-adaptive to) data and (ii) fully automatic, i.e., it requires neither input parameters nor training data to run. In agreement with the GEOSS implementation plan [4], the SIAM software product is implemented as an integrated system of four subsystems, including one “master” 7-band Landsat-like subsystem, L-SIAM (whose spectral resolution comprises bands visible blue: B, visible green: G, visible red: R, near infra-red: NIR, medium infra-red 1: MIR1, medium infra-red 2: MIR2, and thermal infra-red: TIR) plus three “slave” (downscale) subsystems, namely, a 4-band Satellite Pour l’Observation de la Terre (SPOT)-like (G, R, NIR and MIR1), S-SIAM, a 4-band Advanced Very High Resolution Radiometer (AVHRR)-like (R, NIR, MIR1 and TIR), AV-SIAM, and a 4-band QuickBird-like (B, G, R and NIR), Q-SIAM, whose spectral resolutions overlap with Landsat’s, but are inferior to Landsat’s. Noteworthy, an expression like “Landsat-like MS image” adopted in this paper means: “an MS image whose spectral resolution mimics the spectral domain of the 7 bands of the Landsat family of imaging sensors”, i.e., a spectral resolution where bands visible blue (B), visible green (G), visible red (R), near infra-red (NIR), medium infra-red 1 (MIR1), medium infra-red 2 (MIR2) and thermal infra-red (TIR) overlap (which does not mean coincide) with Landsat’s. According to these four families of spectral resolution specifications, the SIAM software product can pre-classify any radiometrically calibrated MS image acquired by past, existing or future-planned optical imaging sensors, either spaceborne or airborne, refer to Table 11-3. To realistically cope with the fact that there is no “fixed” number of quantization levels which is “optimal” in general, since this number is user- and application-specific (refer to this Chapter 11.above), each SIAM subsystem delivers as output four pre-classification maps at different levels of color quantization, which are not alternative, but co-exist (like different hierarchical levels of detail co-exist in ontologies of the world [133], like LC class taxonomies [126], [127]): fine, intermediate, coarse and “shared”, see Table 11-3. The latter provides a pre-defined vocabulary of color names “shared” by the four SIAM subsystems. Hence, this “shared” color vocabulary can be employed for inter-sensor post-classification change/no-change detection.

Since it is a physical model, the sole requirement of the prior knowledge-based SIAM color quantizer is to be input with MS data provided with a physical unit of radiometric measure, namely, digital numbers (DNs) radiometrically calibrated into TOARF or SURF values, in agreement with the QA4EO recommendations [1], refer to Chapter 11.I. Noteworthy,



TOARF  $\supseteq$  SURF, i.e., SURF is a special case of TOARF in clear sky and flat terrain conditions [128]. In practice, (noisy) TOARF  $\approx$  (noiseless) SURF + atmospheric noise + non-flat terrain effects. If SIAM is successful in mapping a MS data space of (noisy) TOARF values into fixed non-overlapping hypervolumes (discrete color names as mutually exclusive and totally exhaustive buffer zones or domains of activation), then (noiseless) SURF values fall around the center of these hypervolumes, see Fig. 11-7. Examples of the SIAM output products, to be employed in the further experimental Chapter 11.5, are shown in Fig. 11-8, Fig. 11-9 and Fig. 11-10.

#### 11.4.6 New Protocol for Q<sup>2</sup>A of MS Image PAN-sharpening Outcome and Process

Summarized in Chapter 11.II.5, preliminary considerations about existing PAN-sharpened MS image quality estimation procedures, surveyed in Chapter 11.3, are taken into account to design and implement a novel (to the best of these author's knowledge, the first) procedure for Q<sup>2</sup>A of MS image PAN-sharpening outcome and process. The proposed estimation procedure belongs to the class of simplified one-statement Wald's protocols with reference image MS<sub>l</sub> at low spatial scale  $l < h$ , refer to Chapter 11.3.2. In spite of this, proposed findings in QI selection and quality metrics can be extended to the second class of procedures for Q<sup>2</sup>A of PAN-sharpened MS images at high spatial scale  $h$ , without reference image, to replace the bivariate heterogeneous UIQI metric,  $Q$ , typically employed in the QNR formulation, refer to Chapter 11.3.3.

The first step in the design of the novel procedure was the definition of an original taxonomy of PAN-sharpened MS image QIs and quality metrics, which are mapped onto four nominal scales, i.e., each QI is assigned with four categorical variables, see Table 11-4, alternative to traditional categorizations, like those proposed or surveyed in [18], [50], [53], [55], [95].

- I. Homogeneous versus heterogeneous (claimed to be "universal") combinations of statistics, refer to Chapter 11.II.5. In general, the latter should be discouraged as a possible cause of non-redundant information loss, due to dimensionality reduction.
- II. Univariate (one-channel) / bivariate / multivariate (multi-channel) analysis.
- III. 1<sup>st</sup>-, 2<sup>nd</sup>- or 3<sup>rd</sup>-order statistics in the spatial domain, in line with the ETAU principle, refer to Chapter 11.II.4.
- IV. QI (or, vice versa, cost index) categories 1 to 4.
  1. SPCTRL: Context-insensitive (pixel-based) Position (row and column)-independent Spectral cost indexes.
  2. SPCTRL & SPTL1: Context-insensitive Position-dependent Spectral cost indexes.
  3. SPCTRL & SPTL2: Context-sensitive Position-independent Spectral cost indexes.
  4. SPCTRL & SPTL1 & SPTL2: Context-sensitive Position-dependent Spectral cost indexes.

For example, this fourfold QI taxonomy, I to IV, reveals that traditional multivariate QIs (or, vice versa, cost indexes), like Q4, ERGAS and average SAM, are all 1<sup>st</sup>-order (non-contextual) statistics in the spatial domain and all belong to the product QI category 2, SPCTRL & SPTL1 - Context-insensitive Position-dependent. Hence, from a statistical standpoint, due to their degree of similarity, they are expected to be correlated in the RS common practice.

In the rest of this section, first, product QIs are described and implemented, in agreement with Table 11-4. Some of these product QIs are extracted from the SIAM output products, generated automatically and in near real-time from the fused MS\*<sub>l</sub> and the reference MS<sub>l</sub> image. Second, process QIs are selected and instantiated. Finally, intra- and inter-category quantitative QI combination and ranking are discussed.

#### 4) First category of product QIs: SPCTRL - Context-insensitive (pixel-based) and Position (row and column)-independent (Rotation invariant)

The first category of product QIs and quality metrics consists of traditional multidimensional (multi-band) absolute differences (e.g., implemented via the Minkowski distance of order 1 [61]) between univariate (one-channel) global (image-wide) gross (summary) characteristics of the 1<sup>st</sup>-order in the spatial domain, estimated from corresponding pairs of bands of the sensory and fused MS images at low spatial scale  $l$ , MS<sub>l,b</sub> and MS\*<sub>l,b</sub>,  $b = 1, \dots, B$ . Given the univariate 1<sup>st</sup>-order (in the spatial domain, pixel-based) histogram (distribution),  $H(x)$ , and its probability density function,  $p(x)$ , with  $\sum_{x \in GL} p(x) =$

1, of a one-channel (univariate) image  $x$  (random variable), with scalar pixel values belonging to the set of gray levels  $GL$ , the implemented scalar summary statistics of  $H(x)$  are (see Table 11-4):

- Mean of  $H(x)$ , 1st-degree moment, MeanUnvrt =  $E[x] = \bar{x} \geq 0$ .

$$(12)$$

- Standard deviation of  $H(x)$ , 2nd-degree moment about the mean, StDvUnvrt =  $\sigma(x) = \sigma_x \geq 0$ .

$$(13)$$



- Skewness of  $H(x)$ , 3rd-degree moment about the mean,  $SkwnsUnvrt = \frac{1}{\sigma_x^3} E[(x - \bar{x})^3]$ .

(14)

- Kurtosis of  $H(x)$ , fourth-degree moment about the mean,  $KrtsUnvrt = \frac{1}{\sigma_x^4} E[(x - \bar{x})^4] \geq 0$ .

(15)

- Entropy of  $H(x)$  [33],  $EnpryUnvrt = -\sum_{x \in GL} p(x) \log_2(p(x)) \in [0, \log_2 GL]$ ,

(16)

where  $EnpryUnvrt$  is maximum (in case of an equiprobable distribution) when  $Energy$  ( $EnrgyUnvrt$ ) is minimum, with  $EnrgyUnvrt = \sum_{x \in GL} p(x)^2 \in [0, 1]$ , such that an alternative formulation of  $EnpryUnvrt$  can be  $EnpryUnvrt = (1 -$

$EnrgyUnvrt) \in [0, 1]$  [75]. In [75], in the domain of 2<sup>nd</sup>-order spatial statistics extracted from a GLCM, it was proved that the two highly correlated measures of image energy and entropy tend to be poorly correlated with features like image contrast and standard deviation, which are in turn highly correlated one another.

### 5) Second category of product QIs: SPCTRL & SPTL1 - Context-insensitive and Position-dependent (Sensitive to Rotation)

As mentioned in the introduction to Chapter 11.4.6, popular multivariate image QIs, like average SAM, ERGAS and Q4, belong to this second category of product QIs.

In our experiments, two features belonging to this category were implemented, one traditional and one innovative, see Table 11-4. The traditional feature is the bivariate PCC, computed pixel-based and band-specific between each pair of bands  $MS_{1,b}$  and  $MS_{1,b}^*$ ,  $b = 1, \dots, B$ . The inverse PCC parameter,  $InvrCrlnBivrt = 1 - PCC$ , is a cost function, in range  $[0, 1]$ , to be minimized for image quality improvement. Band-specific  $InvrCrlnBivrt$  values are averaged across bands. In compliance with Chapter 11.II.5, since PCC is sensitive to collinearities between the two random variables, the proposed evaluation procedure was planned to be validated with and without the contribution of PCC. The latter evaluation case was expected to be more in line with human photointerpretation results.

The innovative feature was the SIAM-based multivariate  $PostClChngDctnMvrt$  statistic, mentioned in Table 11-4. The Q-SIAM prior knowledge-based color space quantizer was run automatically and in near real-time on the fused  $MS_1^*$  and the sensory  $MS_1$  image. Table 11-3 shows that the Q-SIAM subsystem delivers as output one pre-classification map at a so-called “shared” number of color levels. This “shared” color map vocabulary can be employed for automatic inter-sensor post-classification change/no-change detection, as shown in Fig. 11-10. The cumulative number of pixels featuring a change in the multivariate SIAM-based post-classification mapping provided the  $PostClChngDctnMvrt$  statistic.

### 6) Third category of product QIs: SPCTRL & SPTL2 - Context-sensitive Position-independent (Rotation invariant)

Implemented features belonging to this category have no counterpart in state-of-the-art evaluation procedures with or without reference image, like those proposed in [18], [54], [63], [92].

#### a) Original TIMS-GLCM calculator and TIMS-GLCM texture feature extractor

To account for the ETAU principle presented in Chapter 11.II.4 and inspired by the third-order GLCM proposed in [76], a novel third-order isotropic multi-scale GLCM (TIMS-GLCM) was designed and implemented as an upper triangular three-dimensional array (where the typical symmetry of a GLCM is exploited to reduce memory size and computation time), transformed into a probability distribution, such that  $\sum_{i=0}^{GL-1} \sum_{j=i+1}^{GL-1} \sum_{k=0}^{GL-1} p(i, j, k) = 1$ , as shown in Fig. 11-11 and Fig.

11-12.

To reduce computation time (at the cost of a loss in sensitivity), pixel values were discretized into  $GL = 32$  gray levels (histogram bins), in agreement with [76]. The maximum size  $S$  of the square moving window was fixed equal to 7 pixels, corresponding to  $2.44 \times 7 \approx 18$  m in the  $MS_1$  image, investigated in a three-scale TIMS-GLCM instance featuring ray  $r =$  displacement  $D = 1, 2, 3$  in pixel unit, see Fig. 11-11.

The moving window was centered on each pixel of the fused and reference images,  $MS_1^*$  and  $MS_1$ . One univariate TIMS-GLCM was instantiated per image channel, where 3-tuples collected from the pixel-centered moving window were cumulated. The  $mDMI$  set of univariate texture features extracted from a single-band TIMS-GLCM instance, transformed



into a probability distribution, such that  $\sum_{i=0}^{GL-1} \sum_{j=i+1}^{GL-1} \sum_{k=0}^{GL-1} p(i, j, k) = 1$ , were contrast and energy, according to [75], and the large number emphasis (LNE), according to [76].

- Third-order second-degree Contrast =  $\sum_{i=0}^{GL-1} \sum_{j=i+1}^{GL-1} \sum_{k=0}^{GL-1} [(i-j)^2 + (j-k)^2 + (i-k)^2] p(i, j, k)$ .

(17)

- Third-order second-degree Energy (Angular second moment) =  $\sum_{i=0}^{GL-1} \sum_{j=i+1}^{GL-1} \sum_{k=0}^{GL-1} p^2(i, j, k)$ .

(18)

- Third-order second-degree Large Number Emphasis (LNE) =  $\sum_{i=0}^{GL-1} \sum_{j=i+1}^{GL-1} \sum_{k=0}^{GL-1} (i^2 + j^2 + k^2) p(i, j, k)$ .

(19)

The absolute differences (Minkowski distances of order 1 [61]) between per band-specific 3<sup>rd</sup>-order texture statistics were computed between each pair of corresponding bands in the fused MS<sub>1</sub><sup>\*</sup> and reference MS<sub>1</sub> images.

**b) Multivariate SIAM-based multi-level 8-adjacency cross-aura contour measure**

From each of the two multi-level SIAM maps, featuring fine, intermediate and coarse color discretization levels (refer to Table 11-3), automatically generated in near real-time from the test and reference images, MS<sub>1</sub><sup>\*</sup> and MS<sub>1</sub>, a three-level sum of 8-adjacency cross-aura measures was computed, so that the cross-aura value of each pixel ranges in interval {0, 24 = 8 × 3}, see Fig. 11-8 and Fig. 11-9. This SIAM-based three-level sum of 8-adjacency cross-aura measures provides a pixel-specific contour intensity, in range {0, 24}, increasing if the pixel is an isolated contour pixel (see Fig. 11-9) or if its color contrast is persistent through reductions of the color quantization levels (refer to Table 11-3). Next, the image-wide per-pixel average statistic was collected. The absolute difference (Minkowski difference of order 1 [61]) between these two image-wide multivariate statistics collected from images MS<sub>1</sub><sup>\*</sup> and MS<sub>1</sub> was computed.

**7) Fourth category of product QIs: SPCTRL & SPTL1 & SPTL2 - Context-sensitive Position-dependent (Sensitive to rotation)**

Same as in Chapter 11.4.6.3, but the multi-level sum of cross-aura values was binarized into a binary contour pixel value: yes/no, coded as either 1 or 0, to avoid considering contour intensity values superior for isolated pixels. Next, an inter-image absolute difference was computed pixel-wise and, finally, averaged image-wise. In practice, this is an inter-image edge difference. This category of features has no counterpart in state-of-the-art evaluation procedures, with or without reference image, like those proposed in [18], [54], [63], [92].

**8) Process QIs**

In addition to the aforementioned product QIs, two process QIs, also called QIs of operativeness (QIOs) [8], [9], were considered for each MS image PAN-sharpening algorithm: the processing time and the number of system's free parameters, to be user-defined based on heuristics. The latter is monotonically decreasing with ease of use, i.e., it is inversely related to the system's degree of automation [8], [9].

**9) Quantitative QI combination and ranking**

To combine two (or more) quantitative variables whose units of measure, domains of change and/or sensitivity to changes in input data are different, there are two possible strategies: (i) quantitative variables are transformed into ranked variables, equivalent to categorical variables (affected by a quantization error), then ranks are combined, or (ii) their units of measure, domains of change and sensitivities are harmonized, so that the two harmonized quantitative variables can be combined before or after ranking (categorization). In line with past works [129], [130], in the present study heterogeneous quantitative variables were standardized for harmonization before combination. By definition [89], the standard score  $z$  of a raw population  $x$  is:

$$z = \text{stdrd}(x) = \frac{(x - \bar{x})}{\sigma(x)},$$

(20)

such that  $E[z] = 0$  and  $\sigma(z) = 1$ . Since the standardized variable  $z$  represents the distance between the raw score and the population mean in units of the standard deviation, then  $z$  is negative when the raw score is below the population mean, positive when above. It is worth mentioning that the sum of  $T$  standardized variables is a variable with zero mean and variance equal to  $T$ .





In the present study, standardized variables were combined (summed) before ranking when they were belonging to the same product QI category 1 (SPCTRL) to 4 (SPCTRL & SPTL1 & SPTL2), refer to Chapter 11.4.6.1 to Chapter 11.4.6.4, as shown in Table 11-7. Otherwise, to account for inter- rather than intra-category differences, e.g., to combine statistical apples with statistical oranges, inter-category statistics were combined after ranking, see Table 11-9. The advantage of summing up quantitative standardized variables in place of ranked (categorical) variables is that the domain of change of the former is continuous rather than discrete, i.e., standardized variables are not affected by any discretization error.

To recapitulate, in any quantitative evaluation procedure defined beforehand, i.e., prior to looking at the dataset at hand, the arbitrary and application-specific choice of similarity (quality) or dissimilarity (distorsion) QIs and quality metrics does not allow to reach any “ultimate” conclusion about the “absolute” quality of outcomes or processes being assessed. In other words, any quantitative evaluation procedure provides nothing more than relative (subjective) conclusions about alternative solutions. Nonetheless, when a pool of QIs, individually equivalent to weak sources of evidence, is mDMI, then a multivariate convergence-of-evidence approach can be applied, to infer strong conjectures from weak sources of univariate evidence [58], refer to Chapter 11.II.5.

### 11.5 Results

In agreement with Chapter 11.4, the sensory image pair,  $P_h$  and  $MS_i$ , was radiometrically calibrated into TOARF values (see Fig. 11-4) and the reference  $MS_i$  image was automatically mapped by the Q-SIAM expert system for color quantization (see Table 11-3). The radiometrically calibrated downsampled  $P_{h \rightarrow s}$  and  $MS_{i \rightarrow s}$  image pair, with spatial scale  $s = 1/4 = 1/(2.44 \text{ m} \times 4)$ , such that the “artificial” spatial scale ratio ( $1 : s$ ) is equal to the “native” spatial scale ratio ( $h : 1$ ), where spatial scale  $h = 1/0.61 \text{ m}$ , was generated through LPF filtering and filtered image downsampling in accordance with Chapter 11.3.2. The fourteen alternative MS image PAN-sharpening algorithms selected for testing (refer to Table 11-1) were input with the radiometrically calibrated downsampled  $P_{h \rightarrow s}$  and  $MS_{i \rightarrow s}$  image pair to deliver as output fourteen alternative fused  $MS^*_1$  images. Each fused  $MS^*_1$  image was mapped automatically by the prior knowledge-based SIAM pre-classifier. In parallel, according to Chapter 11.4.4 and Table 11-2, a perceptual visual assessment of the fourteen fused  $MS^*_1$  images was conducted by sixteen human subjects, which led to the development of Table 11-5 as a reference baseline.

Summarized in Table 11-4, product QIs were collected from each fused  $MS^*_1$  image in comparison with the reference  $MS_i$  image, as shown in Table 11-6. According to Chapter 11.4.6.6, standardized QIs were combined (summed) per product QI category 1 (SPECL) to 4 (SPCTRL & SPTL1 & SPTL2) and ranked as shown in Table 11-7. The inter-category combination of per-category product ranks was accomplished in Table 11-9. This pool of per-category product ranks was combined with process ranks, shown in Table 11-8, as summarized in Table 11-10. To compare Table 11-10, delivered by the proposed evaluation procedure, with standard QIs (or cost indexes), the popular ERGAS, average SIAM and Q4 statistics were collected from each fused  $MS^*_1$  image compared against the reference  $MS_i$  image, as shown in Table 11-11. In particular, the final Q4 estimate was averaged over Q4 values estimated per image-block in a regular-grid image partition of  $BL \times BL$  image blocks, with  $BL = 8$ , refer to Chapter 11.4.3. Table 11-12 summarizes final ranks of: (i) alternative fused  $MS^*_1$  images, scored by visual inspection (refer to Table 11-5), (ii) products and products & processes, scored by the proposed evaluation procedure (refer to Table 11-9 and Table 11-10 respectively), and (iii) products, assessed in quality by traditional QIs (refer to Table 11-11).

Table 11-13 presents the Spearman's rank correlation coefficients (SRCCs) generated from pairwise comparisons of ranked variables generated in case of ERGAS, average SAM, Q4, Product Case C and Product & Process Case D, selected from Table 11-12. The SRCC index in range  $[-1, 1]$  is a nonparametric measure of statistical dependence between two ranked variables. It assesses how well the relationship between two ranked variables can be described by a monotonically increasing or decreasing function. If there are no repeated data values, a perfect Spearman correlation of  $+1$  or  $-1$  occurs when each of the variables is a monotonically increasing or decreasing function of the other, even if their relationship is not linear, which makes it quite different from the popular PCC.

### 11.6 Discussion

In this experiment, according to Table 11-5, the two qualitatively “best” PAN-sharpened MS images selected by human subjects were those generated by the PC2\_B and HCS3\_NN algorithm implementations.

In agreement with Chapter 11.II.5, the well-known sensitivity of the bivariate PCC to collinearities between the two input random variables, which led to neglect the estimation of PCC as a viable texture feature from popular 2<sup>nd</sup>-order GLCMs [75], recommended the computation of a product rank Case C, where the *InvrCrlnBivrt* cost term was omitted from the product QI category 2, SPCTRL & SPTL1 - Context-insensitive Position-dependent Spectral cost indexes, as a viable alternative to the product rank in Case A, see Table 11-7 and Table 11-9. The same omission of the *InvrCrlnBivrt* cost term accounts for the product & process rank in Case D, alternative to the product & process rank in Case B, see Table



11-10. The conclusion is that, based on theoretical considerations in addition to experimental evidence summarized in Table 11-9 and Table 11-10, the removal of PCC from inter-image QIs is strongly recommended. In addition, in line with theoretical expectations, Table 11-9 and Table 11-10 show that the proposed convergence-of-evidence approach is robust to changes in one information source, like *InvrCrlnBivrt*, in both product and product & process quality assessments.

The final summary Table 11-12 shows that, in this experiment, the first- and second-best choices of the novel quantitative estimation procedure comply with perceptual ranking by human subjects of fourteen alternative PAN-sharpened MS outcomes. Noteworthy, traditional multivariate QIs, like *Q4*, average SAM and ERGAS, completely fail detecting one-of-two best choices by human subjects, specifically, the outcome of the *HCS3\_NN* algorithm implementation, see Fig. 11-13.

Although the goal of this experiment is validation by independent human subjects of an evaluation procedure, rather than selection of a “best” MS image PAN-sharpening algorithm, the conclusion that, according to the proposed product and product & process evaluation procedures, the implemented *HCS3\_NN* algorithm [109] is first-best in both scores is of potential interest to those RS scientists and practitioners who, in the RS common practice, are recommended to comply with the QA4EO guidelines (refer to Chapter 11.I).

Table 11-9 reveals that, in this experiments, the sole QI belonging to category 4, *SPCTRL* & *SPTL1* & *SPTL2* - Context-sensitive Position-dependent Spectral cost indexes, specifically, the *BinaryCntourMvrt* index (see Fig. 11-8 and Fig. 11-9), is the individual indicator that best approximates (which is highly correlated with) the Product final ranks, either Case A or Case C, although no single “universal” QI can exist on a theoretical basis (refer to Chapter 11.II.2).

As observed in Chapter 11.4.6, since they belong to the same product QI category 2, *SPCTRL* & *SPTL1* - Context-insensitive Position-dependent, traditional multivariate QIs (or cost indexes), like *Q4*, ERGAS and average SAM, are likely to be highly correlated (little informative, highly redundant) in many datasets. Traditionally, a correlation coefficient greater than 0.80 represents strong agreement, between 0.40 and 0.80 describes moderate agreement, and below 0.40 represents poor agreement [131]. This theoretical expectation about *Q4*, ERGAS and average SAM is clearly observable in works by other authors, like [18] and [63]. In line with theory, Table 11-13 shows that, also in our experiment, these traditional QIs feature high values of the SRCC, which means their pairwise relationships are nearly monotonically increasing or decreasing.

## 11.7 Conclusions

Provided with a relevant survey value, this paper reports on the design, implementation and validation by independent human subjects of a novel (to the best of these authors’ knowledge, the first) procedure for perceptual visual quality assurance of MS image PAN-sharpening outcome and process. This is an inherently difficult (ill-posed) inductive learning-from-data problem, open to better-posed solutions in compliance with the QA4EO guidelines and the principles of human vision. Several conclusions of potential interest to a large segment of the RS and computer vision communities can be inferred from the high degree of convergence between theoretical considerations and experimental evidence highlighted in this paper.

The first conclusion is that, based on a critical analysis of existing literature, traditional PAN-sharpened MS image quality estimation procedures appear affected by conceptual and implementation drawbacks, which undermine their effectiveness in quality assurance.

- (1) Belonging to the product QI category 2, *SPCTRL* & *SPTL1* - Context-insensitive Position-dependent, “universal” multivariate  $Q/Q4/Q2^n$  and SSIM indexes are implemented as heuristic mixtures, specifically, logical AND-combinations, of “heterogeneous” scalar QIs of signal fidelity, not to be confused with PVQMs. The same consideration holds for the fused image quality assessment without reference, QNR, where the “universal” Q metric is adopted. These AND-combinations of heterogeneous random variables featuring different statistical properties, domain of change and sensitivity to changes in input data, such as the AND-combination of pixel-based bivariate cross-correlation with differences in local univariate mean and standard deviation, are pursued in the erroneous attempt to find a “universal” summary (gross) statistic which cannot exist, due to the non-injective property of QIs, to be regarded as common knowledge by the scientific community. It means that for model comparison purposes, heuristic mixtures of heterogeneous QIs should be avoided before qualitative (categorical) ranking of each individual QI takes place. Otherwise, if any combination of heterogeneous QIs occurs before ranking, then each single QI in the combination should be standardized (z-scored) in advance to feature zero mean and unit variance, which accomplishes inter-QI harmonization of units of measure, domains of variation and sensitivities to changes in input data.
- (2) Popular univariate (one-channel) QIs of signal fidelity [18], like relative bias in percent, relative difference of variances, relative standard deviation, etc., are all 1<sup>st</sup>-order statistics in the spatial domain and all belong to the product QI category 1, *SPCTRL* - Context-insensitive Position-independent, exclusively. The bivariate PCC, which employs as input



two spectral bands, together with popular multivariate QIs [18], like Q4, ERGAS and average SAM, are all 1<sup>st</sup>-order statistics in the spatial domain (pixel-based, context-insensitive) and all belong to the MS image PAN-sharpening outcome QI category 2, SPCTRL & SPTL1 - Context-insensitive Position-dependent, exclusively. The conclusion is that existing MS image PAN-sharpening product evaluation procedures, e.g., refer to [18] and [63], employ no QI belonging to either category 3 (SPCTRL & SPTL2 - Context-sensitive Position-independent) or category 4 (SPCTRL & SPTL1 & SPTL2 - Context-sensitive Position-dependent). Consequences are twofold.

- (i) Since they are pixel-based (context-insensitive), traditional MS image PAN-sharpening product quality estimation procedures consider a planar 2D dataset, i.e., a 2D array of vector data, equivalent to a 1D string of vector data. In practice, they adopt a pixel-based image analysis approach, which is a special case of 1D image analysis where the 2D spatial non-topological and topological properties of images are ignored. Image statistics are typically non-stationary; according to human pre-attentive vision, they should be investigated on a multi-scale basis (at least four spatial scales) up to third-order statistics in the 2D spatial domain [49], [71]. Actually, the recommended block-based implementation of  $Q/Q4/Q2^n$ , does not counterbalance this lack, at the cost of introducing yet-another system's free-parameter, the block size, to be user-defined based on heuristics, because  $Q/SSIM$  is a signal fidelity measure featuring both a statistical link and a formal connection with the conventional pixel-based mean squared error [137]. In image analysis, no "single" best spatial scale can exist [70], which means that no single-scale  $Q/Q4/Q2^n$  can be robust to changes in block size. In convergence with these theoretical considerations, our experimental results provide a significant counter-example where the three popular Q4, ERGAS and average SAM indexes simultaneously fail detecting one-of-two best choices by human subjects, see Table 11-12.
- (ii) Since they belong to the same MS image PAN-sharpening product QI category 2, SPCTRL & SPTL1 - Context-insensitive Position-dependent, popular QIs of signal fidelity, such as Q4, ERGAS and average SAM, are likely to be highly correlated (little informative, highly redundant) in common practice, in agreement with [137]. This theoretical expectation is clearly observable in works by other authors, like [18] and [63]. It is confirmed by the present work, see Table 11-12 and Table 11-13.
- (3) The bivariate PCC statistic is widely adopted in existing estimation procedures, e.g., in the  $Q/Q4/Q2^n$  indexes. The well-known sensitivity of PCC to linear transformations means that correlation is maximum between two images that are either identical or one the linear transformation of the other although, in this latter case, they can look (perceptually) very different. Its macroscopic inconsistency with PVQMs strongly discourages image (dis)similarity metrics, such as those surveyed in [55], from using the PCC as an input variable. Similar considerations led to neglect the estimation of correlation as a viable texture feature from popular GLCMs [75]. This recommendation is successfully put into practice in the present study, see Table 11-9 and Table 11-10.

The second conclusion is that four nominal taxonomies of image-pair QIs are proposed, alternative to existing categorizations, like the univariate/multivariate QI categorization adopted in [18], or various QI taxonomies discussed in [50], [55] and [95]. The four categorization criteria for fused image QIs proposed in this paper are summarized below.

- I. Homogeneous versus heterogeneous ("universal") statistics, like those proposed in [60].
- II. Univariate (one-channel) or multivariate (multi-channel) statistics (where bivariate cross-correlation must be avoided).
- III. 1<sup>st</sup>-order (mean, stdev, skewness, kurtosis, entropy) or 3<sup>rd</sup>-order (contrast, energy, large number emphasis) statistics in the spatial domain.
- IV. Inter-image comparison of type 1: SPCTRL = Context-insensitive (pixel-based) Position (row and column)-independent Spectral QI; 2: SPCTRL & SPTL1 = Context-insensitive Position-dependent Spectral QI; 3: SPCTRL & SPTL2 = Context-sensitive Position-independent Spectral QI; 4: SPCTRL & SPTL1 & SPTL2 = Context-sensitive Position-dependent Spectral QI.

These four nominal taxonomies of fused image QIs and quality metrics, I to IV, help users to select MS image PAN-sharpening product QIs that are minimally dependent and maximally informative (mDMI), in order to adopt a multivariate convergence-of-evidence approach [58]. For example, popular image-pair QIs, such as average SAM, ERGAS and  $Q/Q4/Q2^n$ , which are typically assessed together [18], [63], all belong to the aforementioned product QI category 2, SPCTRL & SPTL1, which means they together tend to be maximally redundant and minimally informative, refer to this Chapter 11.above.

The third conclusion is that, in the proposed experiment, the implemented evaluation procedure, based on the analysis of multiple independent sources of converging evidence, successfully agrees with perceptual ranking by human subjects of fourteen alternative PAN-sharpened MS outcomes, whereas traditional product QIs, like Q4, SAM and ERGAS, completely fail detecting one-of-two best choices by human subjects, specifically, the outcome of the HCS3\_NN



algorithm's implementation, see Table 11-12. Noteworthy, in our experiment, the sole MS image PAN-sharpening product QI belonging to category 4, SPCTRL & SPTL1 & SPTL2, is the individual indicator that best approximates (which is highly correlated with) the product final ranks A and C, see Table 11-9, although it does not mean a single "universal" quality indicator can exist (refer to Chapter 11.II.2).

Despite the goal of this experiment is validation of an evaluation procedure by human subjects, rather than selection of a "best" MS image PAN-sharpening algorithm, the fourth conclusion is of potential interest to RS scientists and practitioners, required to comply with the QA4EO guidelines in their RS common practice: according to the proposed product and product & process evaluation procedures, the implemented HCS3\_NN algorithm [109] is first-best in both scores, see Fig. 11-13.

The fifth conclusion is that, although the proposed procedure for Q<sup>2</sup>A of PAN-sharpened MS outcome and process is a simplified one-statement instantiation of the Wald's three-statement protocol, i.e., the proposed evaluation procedure adopts the raw MS<sub>i</sub> image as a reference benchmark, all the proposed 1<sup>st</sup>- and 3<sup>rd</sup>-order statistics in the spatial domain can be employed in Eq. (9) and Eq. (11) of the MS image PAN-sharpening "quality with no reference" index, QNR. It means that Eq. (8), where the "universal" univariate Q metric is typically adopted by QNR to calculate the dissimilarity between couples of bands, should be reformulated, refer to Chapter 11.3.

The sixth conclusion is that, in compliance with the Yellott's (E)TAU principle (refer to Chapter 11.II.4), the proposed third-order isotropic multi-scale gray-level co-occurrence matrix (TIMS-GLCM) calculator and the TIMS-GLCM texture feature descriptor are expected to outperform the traditional 2<sup>nd</sup>-order GLCM [75], still popular in the RS community and implemented in commercial RS image processing software toolboxes, like Exelis EN6 [101], Trimble eCognition [132] and ERDAS Imagine [100].

Last but not least, this study reveals another RS data application domain, specifically, Q<sup>2</sup>A of PAN-sharpened MS images, where a prior knowledge-based MS data space quantization (partitioning) algorithm in operating mode, such as SIAM, is eligible for use. Documented applications of prior knowledge-based image mapping systems date back to the early 1980s [44], [45] and span from RS image enhancement (pre-processing) [113], [114] to RS image understanding [8], [9], [12], [13], [14], [44], [45], [115], [116], [117], [118], refer to Chapter 11.4.5.

Planned future developments of this work will regard the validation of the proposed evaluation procedure for MS image PAN-sharpening outcome when input three-band test images, acquired by terrain-level cameras, depict natural landscapes, whose photointerpretation and perceptual quality assessment by independent human subjects is expected to be more intuitive and, therefore, reliable. These future developments are expected to agree with the PVQM proposed in [138] as a viable alternative to the signal fidelity measure SSIM by one of its authors. It is summarized as follows.

$$D(I_R, I_T) = \frac{1}{S} \sum_{s=0}^{S-1} \frac{1}{\sqrt{SF_s}} \|\hat{I}_{R,s} - \hat{I}_{T,s}\|^2, \quad (21)$$

where  $I_R$  and  $I_T$  are the reference and the test image respectively,  $\hat{I}_{R,s}$  and  $\hat{I}_{T,s}$  denote vectors containing the transformed reference and distorted image data at scale  $s = 0, \dots, S-1$ , respectively, and where  $SF_s$  is the number of spatial filters in the sub-band at scale  $s$ . In this equation a root mean squared error is computed for each scale, and then averaged over these scales giving larger weight to the lower frequency coefficients (which are fewer in number, due to subsampling). Per se, this scale-dependent weighting policy is not supported by any perceptual plausibility.

Alternative to the normalized Laplacian pyramid proposed in [138], an even-symmetric multi-scale filter bank proposed in [139] provides a near-orthogonal image decomposition and a zero-crossing (ZX) image-contour detection. For the sake of simplicity, in 1D signal processing, it is such that:

$$\text{Renstrect}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G / \partial x^2] / 2, \quad (22)$$

where  $G(x)$  is a 1D Gaussian low-pass filter and  $\partial^2 G / \partial x^2$  is the second-derivative of a 1D Gaussian function which mimics an even-symmetric spatial filter. Noteworthy, according to [140], the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution [88]. Therefore, the filtered image ( $\partial^2 G / \partial n^2 \circ I$ ) is different from the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  applied to the low-pass image adopted by both Canny [141] and Bertero, Torre and Poggio [46]. Hence, the inequality

$$(\partial^2 G / \partial n^2 \circ I) \neq \partial^2 / \partial n^2 (G \circ I) \quad (23)$$

always holds true. As a consequence, in Eq. (22), term  $\{f(x) \circ G(x)\} \in [0, \text{MaxGrayValue}]$  and  $\{[f(x) \circ \partial^2 G / \partial x^2] / 2\} \in [-\text{MaxGrayValue} / 2, \text{MaxGrayValue} / 2]$ . According to these properties, the PVQM proposed in [139] is (attention: for the time being, this index ignores the multiple spatial scales):





$$D(I_R, I_T) = |f_R(x) \circ G(x) - f_T(x) \circ G(x)| + \left| \frac{f_R(x) \circ \frac{\partial^2 G}{\partial x^2}}{2} - \frac{f_T(x) \circ \frac{\partial^2 G}{\partial x^2}}{2} \right|, \quad (24)$$

where  $|f_R(x) \circ G(x) - f_T(x) \circ G(x)| \in [0, \text{MaxGrayValue}]$  and  $|\frac{f_R(x) \circ \partial^2 G / \partial x^2}{2} - \frac{f_T(x) \circ \partial^2 G / \partial x^2}{2}| \in [0, \text{MaxGrayValue}]$ . In mathematical terms, this is a Minkowski distance with degree  $d$  equal to 1. If appropriate,  $d$  can be set equal 2 (to apply a Euclidean distance) or superior.

### Acknowledgments

To accomplish this work Andrea Baraldi was supported in part by the National Aeronautics and Space Administration under Grant No. NNX07AV19G issued through the Earth Science Division of the Science Mission Directorate. Francesca Despini and Sergio Teggi were funded by the Agenzia Spaziale Italiana (ASI), in the framework of the project "Analisi Sistema Iperspettrali per le Applicazioni Geofisiche Integrate - ASI-AGI" (n. I/016/11/0). The authors are grateful to the Modena and Reggio Emilia University staff and students who voluntarily participated in this work through the perceptual image quality assessment experiment. Andrea Baraldi thanks Prof. Raphael Capurro for his hospitality, patience, politeness and open-mindedness. He also thanks Prof. Christopher Justice, Chair of the Department of Geographical Sciences, University of Maryland, for his friendship and support. The authors also wish to thank the Editor-in-Chief, Associate Editor and reviewers for their competence, patience and willingness to help.

### References in Chapter 11

- [1] *A Quality Assurance Framework for Earth Observation, version 4.0*, Group on Earth Observation / Committee on Earth Observation Satellites (GEO/CEOS), 2010. [Online]. Available: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf)
- [2] M. C. Hansen and T. R. Loveland, "A review of large area monitoring of land cover change using Landsat data," *Remote Sens. of Env.*, vol. 122, pp. 66–74, 2012.
- [3] A. Chen, G. G. Leptoukh, and S. J. Kempner, "Using KML and virtual globes to access and visualize heterogeneous datasets and explore their relationships along the A-Train tracks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 3, pp. 352–358, 2010.
- [4] *The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan*, Group on Earth Observation (GEO), 2005. [Online]. Available: <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf>
- [5] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.
- [6] I. Herrmann, A. Pimstein, A. Karnieli, Y. Cohen, V. Alchanatis, D.J. Bonfil, "LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands," *Remote Sens. of Env.*, vol. 115, pp. 2141–2151, 2011.
- [7] R. Capurro and B. Hjørland, "The concept of information," *Annual Review of Inf. Science and Tech.*, vol. 37, pp. 343–411, 2003.
- [8] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction," *Remote Sensing*, vol. 4, pp. 2694–2735, 2012.
- [9] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation," *Remote Sens.*, vol. 4, pp. 2768–2817, 2012.
- [10] *Land Product Validation (LPV)*, Committee on Earth Observation Satellites (CEOS) - Working Group on Calibration and Validation (WGCV), [Online]. Available: <http://lpvs.gsfc.nasa.gov/>. Accessed on March 20, 2015.
- [11] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik, "Reflectance quantities in optical remote sensing - Definitions and case studies," *Remote Sens. Environ.*, vol. 103, pp. 27–42, 2006.
- [12] A. Baraldi, L. Boschetti, L., and M. Humber, "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, Part: 2, pp. 701–760, Jan. 2014.



- [13] A. Baraldi, M. L. Humber, D. Tiede, and S. Lang, "Automatic pre-classification of Landsat image composites of the conterminous United States – Process and outcome quality indicators", *Remote Sens.*, submitted for consideration for publication, Jan. 2015.
- [14] A. Baraldi and M. Humber, "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 1 – Theory," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 1307-1329, March 2015.
- [15] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*. New York, NY: Wiley, 1998.
- [16] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons, 2004.
- [17] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Qian Du, "Multi-Modal Change Detection, Application to the Detection of Flooded Areas: Outcome of the 2009–2010 Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331-342, 2012.
- [18] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, et al., "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data Fusion Contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012-3021, 2007.
- [19] J. Li, "Spatial quality evaluation of fusion of different resolution images," *ISPRS Int. Arch. Photogramm. Remote Sens.*, vol. 33, no. B2-2, pp. 339–346, 2000.
- [20] C. Thomas and L. Wald, "Assessment of the quality of fused products," in *Proc. 24th EARSeL Annu. Symp. New Strategies Eur. Remote Sens.*, Dubrovnik, Croatia, May 25–27, 2004. M. Oluic, Ed., Rotterdam, The Netherlands: Balkema, 2005, pp. 317–325.
- [21] Z. Wang, D. Ziou, C. Armenakis, D. Li and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, issue 6, pp. 1391-1402, 2005.
- [22] L. Wald, "Some terms of reference in data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, May 1999.
- [23] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Spectral information extraction by means of MS+PAN fusion," *ESA-EUSC*, 2004.
- [24] D. A. Quattrochi and M. F. Goodchild, Eds., *Scale in Remote Sensing and GIS*. CRC Press, 1997.
- [25] A. Dadon, E. Ben-Dor, M. Beyth, and A. Karnieli, "Examination of spaceborne imaging spectroscopy data utility for stratigraphic and lithologic mapping," *J. Appl. Remote Sens.*, vol. 5, no. 1, pp.-053507, 2011.
- [26] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.
- [27] B. Garguet-Duport, J. Girel, J.-M. Chassery, and G. Pautou, "The use of multi-resolution analysis and wavelets transform for merging SPOT panchromatic and multi-spectral image data," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1057–1066, 1996.
- [28] J. Hadamard, "Sur les problemes aux derivees partielles et leur signification physique," *Princet. Univ. Bull.*, vol. 13, pp. 49–52, 1902.
- [29] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. London, U.K.: Chapman & Hall, 1994.
- [30] Q. Iqbal and J. K. Aggarwal, "Image retrieval via isotropic and anisotropic mappings," in *Proc. IAPR Workshop Pattern Recognit. Inf. Syst.*, Setubal, Portugal, Jul. 2001, pp. 34–49.
- [31] G. A. Miller, "The cognitive revolution: a historical perspective", in *Trends in Cognitive Sciences 7*, pp. 141-144, 2003.
- [32] F. J. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press, 1991.
- [33] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
- [34] G. Patané and M. Russo, "The enhanced-LBG algorithm," *Neural Networks*, vol. 14, no. 9, pp. 1219–1237, 2001.
- [35] G. Patané and M. Russo, "Fully automatic clustering system," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1285–1298, 2002.
- [36] B. Fritzke, "The LBG-U method for vector quantization—An improvement over LBG inspired from neural networks," *Neural Processing Lett.*, vol. 5, no. 1, 1997.
- [37] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–94, Jan. 1980.



- [38] D. Lee, S. Baek, and K. Sung, "Modified k-means algorithm for vector quantizer design," *IEEE Signal Processing Lett.*, vol. 4, pp. 2–4, Jan. 1997.
- [39] B. Fritzke, *Some competitive learning methods*. Draft document, 1997. [Online]. Available: <http://www.demogng.de/JavaPaper/t.html>
- [40] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part I," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 645–661, March 2002.
- [41] A. Baraldi and E. Alpaydin, "Constructive feedforward ART clustering networks - Part II," *IEEE Trans. Neural Networks*, vol. 13, no. 3, pp. 662–677, March 2002.
- [42] T. Martinetz, G. Berkovich, and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
- [43] R. Laurini and D. Thompson, *Fundamentals of Spatial Information Systems*. London, UK: Academic Press, 1992.
- [44] T. Matsuyama and V. Shang-Shouq Hwang, *SIGMA – A Knowledge-based Aerial Image Understanding System*, Plenum Press, 1990.
- [45] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. New York, NY, USA: Plenum, 1980.
- [46] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, pp. 869–889, 1988.
- [47] S. P. Vecera and M. J. Farah, "Is visual image segmentation a bottom-up or an interactive process?," *Perception and Psychophysics*, vol. 59, pp. 1280–1296, 1997.
- [48] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images", *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [49] Jain A. and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 124–128, Jan. 1998.
- [50] J. Pa and A. V. Hegdeb, "A Review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, pp. 133 – 142, 2015.
- [51] Y. Zhang and R. K. Mishra, "A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion", in *IEEE Geosci. and Remote Sens. Symposium (IGARSS) 2012, 22-27 July 2012*, pp. 182–185.
- [52] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, and F. Nencini, "A new method for MS+Pan image fusion assessment without reference," *IEEE*, 2006.
- [53] M. C. El-Mezouar, N. Taleb, K. Kpalma, and J. Ronsin, "A new evaluation protocol for image pan-sharpening methods," *ICCIT 2012*, pp. 144–148.
- [54] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, 2008.
- [55] Weisi Lin, C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image R.*, vol. 22, pp. 297–312, 2011.
- [56] A. Baraldi, L. Bruzzone, and P. Blonda, "Quality assessment of classification and cluster maps without ground truth knowledge," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, Apr. 2005.
- [57] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.
- [58] A. Baraldi and J. V. B. Soares, "Software library of two-dimensional shape descriptors in object-based image analysis," *IEEE Trans. Image Proc.*, submitted, 2015.
- [59] *World Happiness Report, 2012/2013*, Columbia University, Canadian Institute for Advanced Research, London School of Economics, 2014.
- [60] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Proc. Letters*, vol. 9, no. 3, pp. 81–84, March 2002.
- [61] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, p. 1–14, 2004.
- [62] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Letters*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
- [63] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Liardi, R., Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565 – 2586, 2015.



- [64] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyper-spectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.
- [65] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–353, 1965.
- [66] B. Kosko, *Fuzzy Thinking*. Flamingo, London, UK, 1994.
- [67] C. Mason and E. R. Kandel, "Central visual pathways," in *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 1991, pp. 420–439.
- [68] P. Gouras, "Color vision," in *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 1991, pp. 467–479.
- [69] E. R. Kandel, "Perception of motion, depth and form," in *Principles of Neural Science*, E. Kandel and J. Schwartz, Eds. Norwalk, CT, USA: Appleton and Lange, 1991, pp. 441–466.
- [70] H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vis. Res.*, vol. 19, no. 1, pp. 19–32, 1979.
- [71] J. I. Yellott, "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," *Opt. Soc. Am.*, vol. 10, no. 5, pp. 777–793, May 1993.
- [72] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited," *Perception*, vol. 2, pp. 391–405, 1973.
- [73] B. Julesz, "Texton gradients: The texton theory revisited," in *Biomedical and Life Sciences Collection*, Springer, Berlin / Heidelberg, vol. 54, no. 4-5, Aug. 1986.
- [74] J. Victor, "Images, statistics, and textures: Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: Comment," *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1680–1684, May 1994.
- [75] A. Baraldi and F. Parmiggiani, "An investigation of textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 293–304, March 1995.
- [76] H. Anys and D. C. He, "Evaluation of textural and multipolarization radar features for crop classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1170–1181, 1995.
- [77] D. Marr, *Vision*. New York, NY, USA: Freeman and C., 1982.
- [78] R. M. Boynton, "Human color perception," in *Science of Vision*, K. N. Leibovic, Ed., pp. 211–253, Springer-Verlag, New York, 1990.
- [79] J. Rodrigues and J.M. Hans du Buf, "Multi-scale lines and edges in V1 and beyond: Brightness, object categorization and recognition, and consciousness," *BioSystems*, vol. 1, pp. 1–21, 2008.
- [80] J. Rodrigues and J.M.H. du Buf, "Multi-scale keypoints in V1 and beyond: Object segregation, scale selection, saliency maps and face detection", *BioSystems*, vol. 86, pp. 75–90, 2006.
- [81] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, "Simulation of neural contour mechanisms: from simple to end-stopped cells," *Vision Res.*, vol. 32, no. 5, pp. 963–981, 1992.
- [82] L. Pessoa, "Mach Bands: How Many Models are Possible? Recent Experimental Findings and Modeling Attempts", *Vision Res.*, Vol. 36, No. 19, pp. 3205–3227, 1996.
- [83] H. du Buf and J. Rodrigues, Image morphology: from perception to rendering, in *IMAGE - Computational Visualistics and Picture Morphology*, 2007.
- [84] D. C. Burr and M. C. Morrone, "A nonlinear model of feature detection," in *Nonlinear Vision: Determination of Neural Receptive Fields, Functions, and Networks*, R. B. Pinter and N. Bahram, Eds., pp. 309–327, CRC Press, Boca Raton, FL, 1992.
- [85] E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of motion," *J. Opt. Soc. Am. A*, vol. 2, pp. 284–299, 1985.
- [86] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comp. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [87] A. Baraldi and F. Parmiggiani, "Combined detection of intensity and chromatic contours in color images," *Optical Engineering*, vol. 35, no. 5, pp. 1413–1439, May 1996.
- [88] M. Hagenlocher, S. Lang, D. Hölbling, D. Tiede, and S. Kienberger, "Modeling hotspots of climate change in the Sahel using object-based regionalization of multidimensional gridded datasets," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 229–234, Jan. 2014.
- [89] E. Kreyszig, *Applied Mathematics*. Wiley Press, 1979.
- [90] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.





- [91] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, "Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images," *Inf. Fusion*, vol. 6, no. 3, pp. 193–212, Sep. 2005.
- [92] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*, 2nd ed., C.-H. Chen, Ed. Boca Raton, FL, USA: CRC Press, 2012, pp. 533–548.
- [93] L. Wald, *Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses, Ecole des Mines de Paris, 2002.
- [94] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
- [95] C. Thomas and L. Wald, "Comparing distances for quality assessment of fused products," in *Proc. 26th EARSeL Annu. Symp. New Develop. Challenges Remote Sens.*, Warsaw, Poland, May 29–31, 2006. Z. Bochenek, Ed., Rotterdam, The Netherlands: Balkema, 2007, pp. 101–111.
- [96] P.S. Pradhan, R.L. King, N.H. Younan, and D.W. Holcomb, "Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 3674 – 3683, May 2006.
- [97] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, 2000.
- [98] P. Lukowicz, E. Heinz, L. Prechelt, and W. Tichy, "Experimental evaluation in computer science: a quantitative study," *Tech. Rep. 17/94*, Univ. Karlsruhe (Germany), 1994.
- [99] L. Prechelt, "A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice," *Neural Networks*, vol. 9, 1996.
- [100] *ERDAS IMAGINE 2015 User's Guide*, Hexagon Geospatial. [Online] Available: <http://www.hexagongeospatial.com/products/remote-sensing/erdas-imagine>.
- [101] *EN6 EX User Guide 5.0*, ITT Visual Information Solutions, Dec. 2009. [Online]. Available: [http://www.exelisvis.com/portals/0/pdfs/enviex/EN6\\_EX\\_User\\_Guide.pdf](http://www.exelisvis.com/portals/0/pdfs/enviex/EN6_EX_User_Guide.pdf)
- [102] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.
- [103] C. A. Laben and B. V. Brower, "Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpener," *US Patent 6.011.875*, 1998.
- [104] P. A. Brivio, G. Lechi, and E. Zilioli, *Principi e metodi di Telerilevamento*. CittàStudi Edizioni, 2006.
- [105] K. Amolins, Y. Zhang, and P. Dare. "Wavelet based image fusion techniques - An introduction, review and comparison," *ISPRS J. Photogram. & Remote Sens.*, vol. 62, pp. 249–263, 2007.
- [106] M. J. Canty, *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for EN6/IDL and Python*. Crc Press, 2014.
- [107] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multi-resolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.
- [108] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, (1999). Multiresolution- based image fusion with additive wavelet decomposition, *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [109] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood, "WorldView-2 pan-sharpening," in *American Society for Photogrammetry and Remote Sensing*, 2010.
- [110] M. Ehlers, "Spectral characteristics preserving image fusion based on Fourier domain filtering," in *Proc. of SPIE*, Maspalomas, Spain, 2004, vol. 5574, pp. 1–13.
- [111] P.S. Chavez, S.C. Sides, J.A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT Panchromatic," *Photogram. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.
- [112] *The 5th International Conference on Computing for Geospatial Research and Application (COM.Geo 2014)*, Call For Presentation. [Online]. Available: <http://www.com-geo.org/conferences/2014/topics.htm>. Accessed on March 20, 2015.



- [113] A. Baraldi, M. Humber and L. Boschetti, "Quality assessment of pre-classification maps generated from spaceborne/airborne multi-spectral images by the Satellite Image Automatic Mapper™ and Atmospheric/Topographic Correction™-Spectral Classification software products: Part 2 – Experimental results," *Remote Sens.*, vol. 5, pp. 5209-5264, Oct. 2013.
- [114] A. Baraldi, M. Girona, and D. Simonetti, "Operational three-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 112-146, Jan. 2010.
- [115] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophys. Res.*, vol. 103, D24, no. 32, pp. 141-157, 1998.
- [116] Y. Luo, A. P. Trishchenko and K. V. Khlopenkov, "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sens. of Env.*, vol. 112, pp. 4167-4185, 2008.
- [117] Zhe Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. of Env.*, vol. 118, pp. 83-94, 2012
- [118] D. Simonetti, E. Simonetti, Z. Szantoi, A. Lupi and H. D. Eva, "First results from the phenology-based synthesis classifier using Landsat 8 imagery," *IEEE Geosci. Remote Sens. Letters*, accepted for publication, Feb. 2015.
- [119] J. van de Weijer, C. Schmid, J. Verbeek, D. Larlus, Learning color names for real-world applications, *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1512 – 1523, 2009.
- [120] R. Benavente, M. Vanrell, and R. Baldrich, "Parametric fuzzy sets for automatic color naming," *J. Opt. Soc. Am. A*, vol. 25, pp. 2582-2593, 2008.
- [121] T. Gevers, A. Gijzenij, J. van de Weijer, J. M. Geusebroek, *Color in Computer Vision*. Hoboken, NJ, USA: Wiley, 2012
- [122] E. C. Tsao, J. C. Bezdek, and N. R. Pal, "Fuzzy Kohonen clustering network," *Pattern Recognition*, vol. 27, no. 5, pp. 757-764, 1994.
- [123] E. Erwin, K. Obermayer, and K. Schulten, "Self-organizing maps: Ordering, convergence properties and energy functions," *Biol. Cybern.*, vol. 67, pp. 47-55, 1992.
- [124] S. P. Luttrell, "A Bayesian analysis of self-organizing maps," *Neural Comput.*, vol. 6, pp. 767-794, 1994.
- [125] A. Baraldi, P. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 2563-2586, 2006.
- [126] *OWL Web Ontology Language*, World Wide Web Consortium (W3C). [Online] Available: <http://www.w3.org/TR/owl-features/>
- [127] H. Couclelis, "Ontologies of geographic information," *Int. J. Geo. Info. Science*, vol. 24, no. 12, pp. 1785-1809, 2010.
- [128] P. S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, pp. 459-479, 1988.
- [129] A. Baraldi, L. Bruzzone, P. Blonda, and L. Carlin, "Badly-posed classification of remotely sensed images - An experimental comparison of existing data labeling systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 214-235, Jan. 2006.
- [130] A. Baraldi, L. Bruzzone and P. Blonda, "A multi-scale Expectation-Maximization semisupervised classifier suitable for badly-posed image classification," *IEEE Trans. Image Proc.*, vol. 15, no. 8, pp. 2208-2225, Aug. 2006.
- [131] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, Boca Raton, FL: Lewis Publishers, 1999.
- [132] *eCognition® Developer 9.0 Reference Book*, Trimble, 2015.
- [133] H. Couclelis, "Ontologies of geographic information," *Int. J. Geo. Info. Science*, vol. 24, no. 12, pp. 1785-1809, 2010.
- [134] G. Ginesu, F. Massidda, and D. Giusto, "A multi-factors approach for image quality assessment based on a human visual system model," *Signal Processing: Image Communication*, vol. 21, pp. 316-333, 2006.
- [135] C. Chubb and J. I. Yellott, "Every discrete, finite image is uniquely determined by its dipole histogram," *Vision Research*, vol. 40, pp. 485-492, 2000.
- [136] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pp. 1398-1402, 2003.



- [137] R. Dosselmann and Xue Dong Yang, “A comprehensive assessment of the structural similarity index,” SIViP 2011, vol. 5, pp. 81–91, 2011.
- [138] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, Perceptual image quality assessment using a normalized Laplacian pyramid, Proc. IS&T Int’l Symposium on Electronic Imaging, Conf. on Human Vision and Electronic Imaging, vol. 2016(16), Feb 2016.
- [139] A. Baraldi, Operational automatic stratified multi-scale image-contour, keypoint, texel and texture-boundary detection in panchromatic and color images: Developments and open challenges, *Preliminary report No. 1, version 10.1*, Baraldi Consultancy in Remote Sensing of Andrea Baraldi, August 2016.
- [140] V. Torre and T. Poggio, “On edge detection,” *IEEE Trans. Pattern Anal. and Mach. Intell.* vol. 8, no. 2, pp. 147–163, 1986.
- [141] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. and Mach. Intell.* vol. 8, no. 6, pp. 679–698, 1986.

Figures and figure captions in Chapter 11

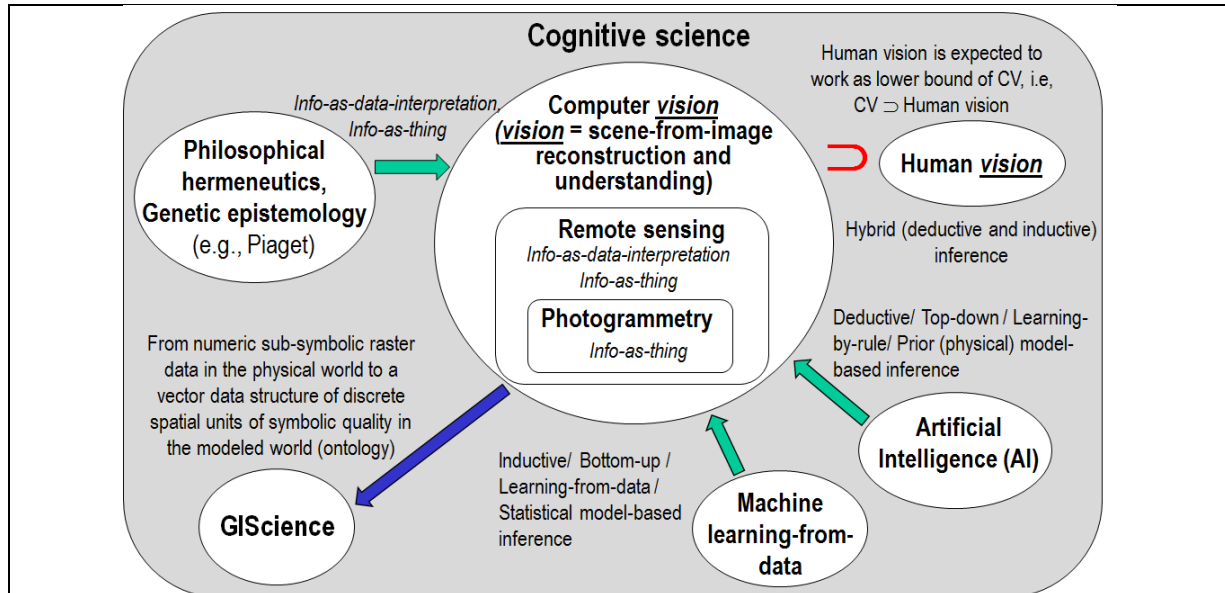


Fig. 11-1. Like engineering, remote sensing (RS) is a metascience, whose goal is to transform knowledge of the world, provided by other scientific disciplines, into useful user- and context-dependent solutions in the world [133]. Cognitive science is the interdisciplinary scientific study of the mind and its processes. It examines what cognition (learning) is, what it does and how it works. It especially focuses on how information/knowledge is represented, acquired, processed and transferred within nervous systems (humans or other animals) and machines (e.g., computers) [31], [32].

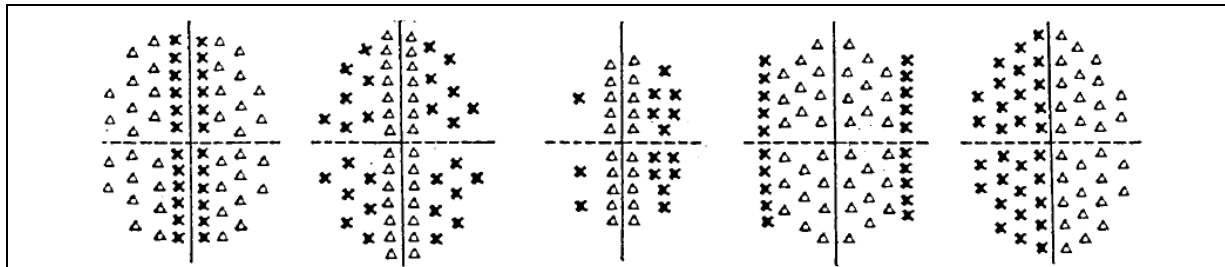


Fig. 11-2. Excitatory and inhibitory terms, shown as triangles and crosses, in activation domains of even- and odd-symmetric S-cells found in the PVC of mammals by Mason and Kandel in their seminal work on neuroscience [67].

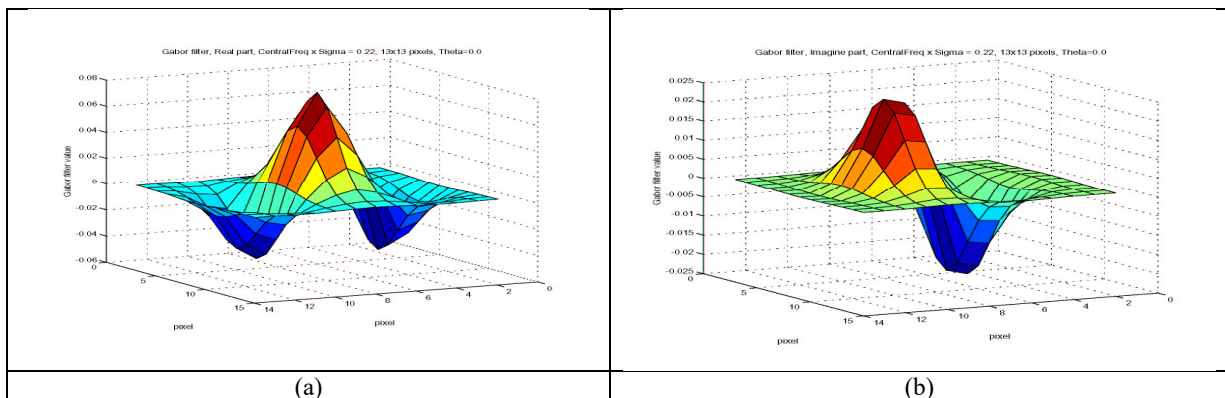






Fig. 11-3. Multi-scale multi-orientation filter bank consisting of Gabor wavelets such that: (I) A Gabor filter is a Gaussian function modulated by a complex sinusoid. (II) The real part of an oriented Gabor mother-wavelet is selected as an even-symmetric 2<sup>nd</sup>-order derivative of a Gaussian function, equivalent to a 3<sup>rd</sup>-order spatial statistic, in line with the works by Yellot [71] and Victor [74], see Fig. 11-3(a). According to [77], [87], this local filter is necessary and sufficient to detect any sort of image contours, namely, step edge, roof, line (ridge) and ramps (in compliance with the Mach bands illusion [82]), as zero-crossings of the even-symmetric filtered image. (III) The imaginary part of an oriented Gabor mother-wavelet, shown in Fig. 11-3(b), provides an odd-symmetric 1<sup>st</sup>-order derivative of a Gaussian function, equivalent to a 2<sup>nd</sup>-order spatial statistic. According to the first author of the present study, in the raw primal sketch, this odd-symmetric filter is not employed in image contour detection, in agreement with [77], [87], but in multi-scale keypoint extraction exclusively, in line with [120], [121]. (IV) The oriented Gabor mother-wavelet is designed with a zero DC-component (to be insensitive to ramps and constant offsets), in line with [87]. (V) To provide the best compromise between computation time and the quality of the image decomposition/synthesis the following filter bank design is selected [87]. (i) Four dyadic spatial scales (one octave apart), with filter size equals to 3, 3×2, 3×2<sup>2</sup>, 3×2<sup>3</sup> pixels, in agreement with [70]. (ii) Two spatial orientations: 0 and 90 degrees.

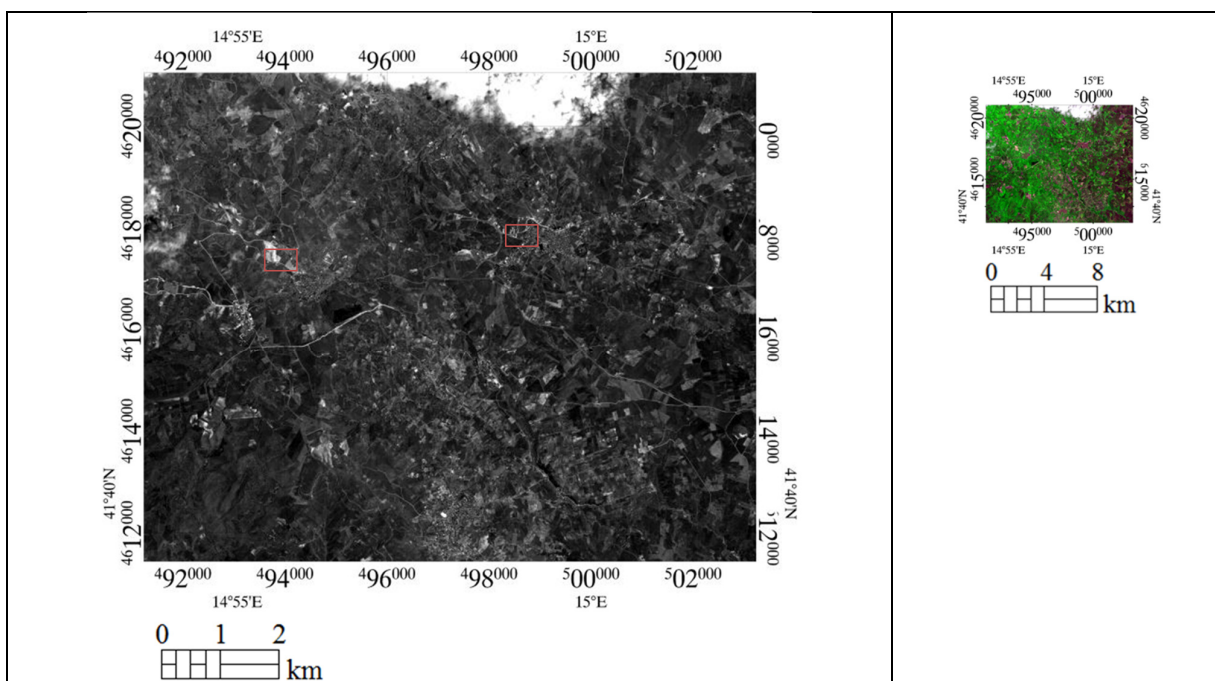
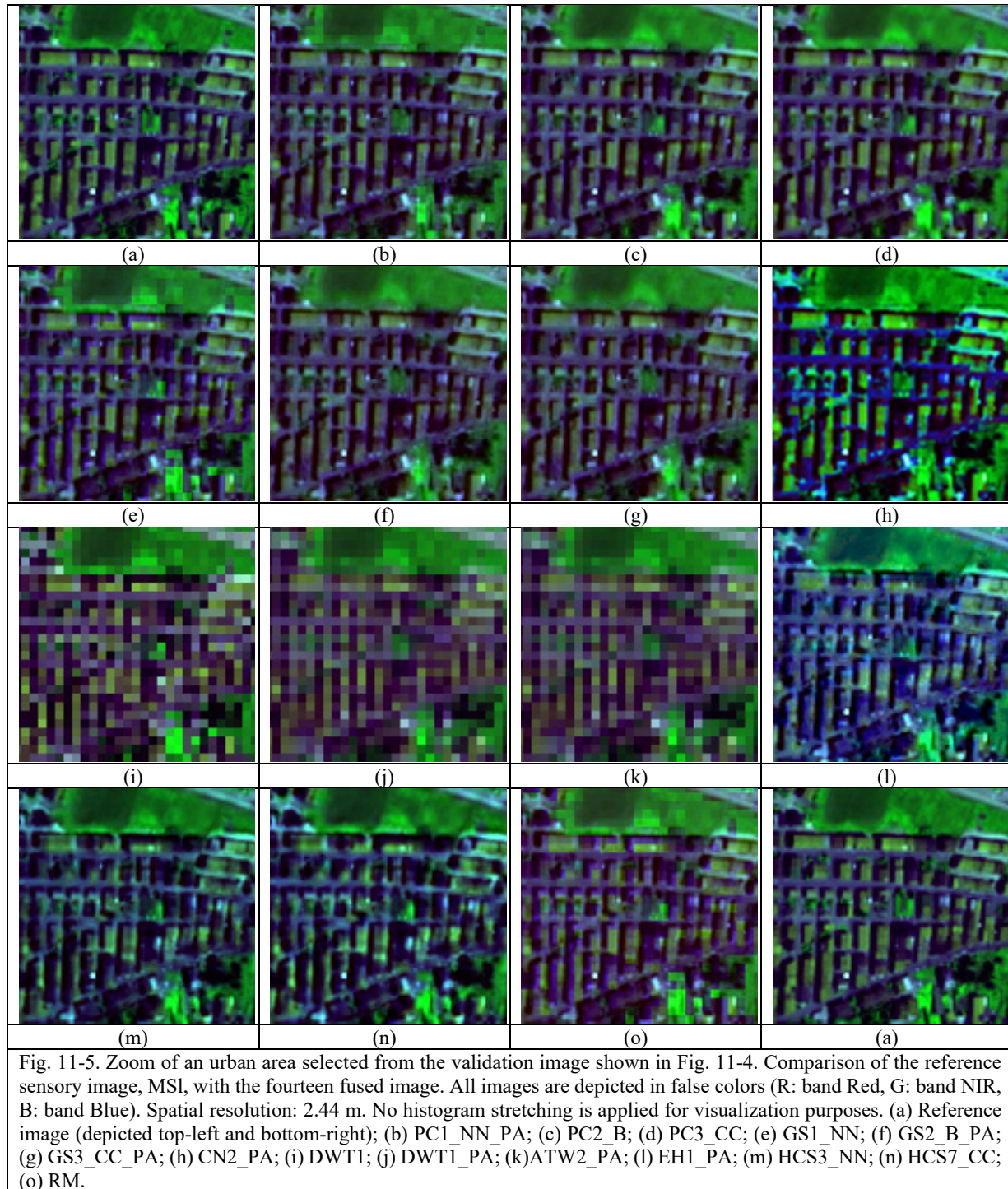


Fig. 11-4. (a) Right. Radiometrically calibrated QuickBird-2 MS image of the Campania region, Italy, acquired on 2004-06-13 at 09:58 a.m., depicted in false colors (R: band Red, G: band NIR, B: band Blue). Spatial resolution: 2.44 m. Default EN6 2% linear histogram stretching applied for visualization purposes. (b) Left. Radiometrically calibrated QuickBird-2 PAN image of the Campania region, Italy, acquired on 2004-06-13 at 09:58 a.m. Spatial resolution: 0.61 m. Default EN6 2% linear histogram stretching applied for visualization purposes. The two red rectangular outlines represent two zoomed areas, shown in Fig. 11-5 and Fig. 11-6 respectively.





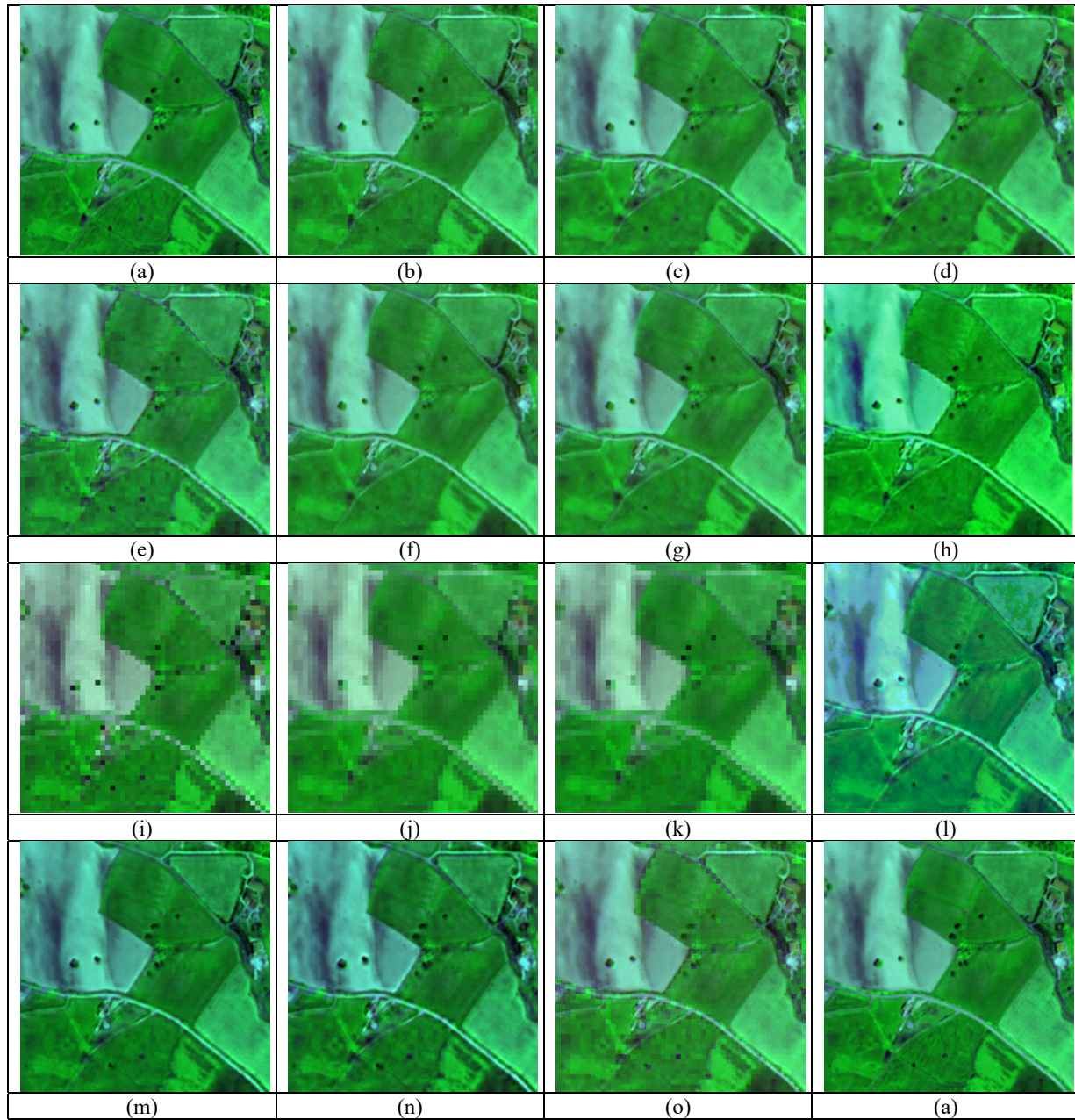


Fig. 11-6. Zoom of an agricultural area selected from the validation image shown in Fig. 11-4. Comparison of the reference sensory image,  $MS_i$ , with the fourteen fused image. All images are depicted in false colors (R: band Red, G: band Near Infrared, B: band Blue). Spatial resolution: 2.44 m. No histogram stretching is applied for visualization purposes. (a) Reference image (depicted top-left and bottom-right); (b) PC1\_NN\_PA; (c) PC2\_B; (d) PC3\_CC; (e) GS1\_NN; (f) GS2\_B\_PA; (g) GS3\_CC\_PA; (h) CN2\_PA; (i) DWT1; (j) DWT1\_PA; (k) ATW2\_PA; (l) EH1\_PA; (m) HCS3\_NN; (n) HCS7\_CC; (o) RM.

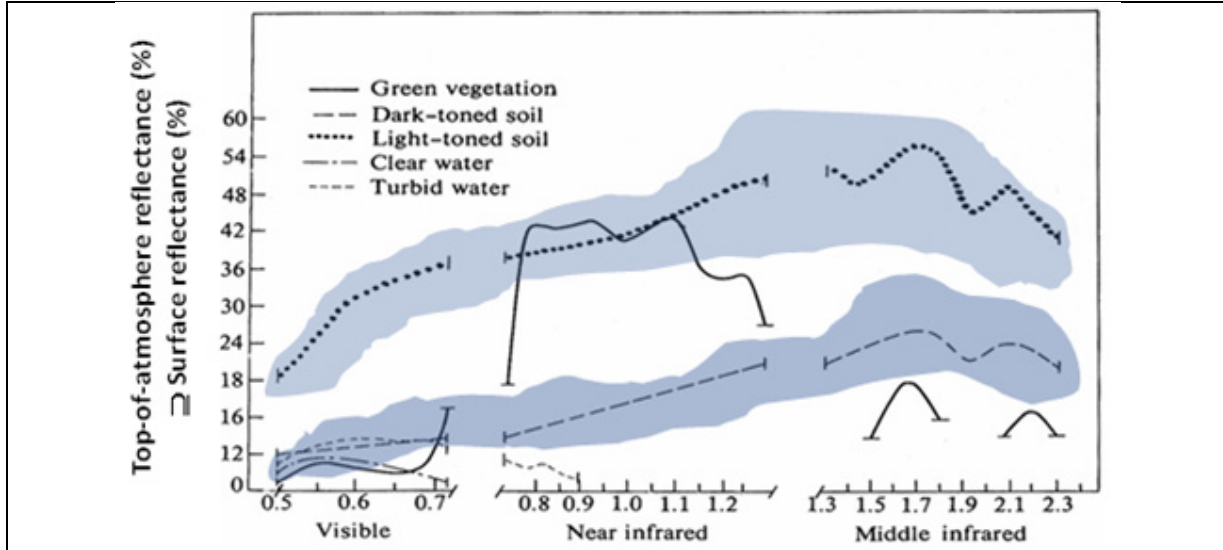


Fig. 11-7. Examples of land cover (LC)-class specific families of spectral signatures in top-of-atmosphere reflectance (TOARF) values. A within-class family of spectral signatures (e.g., dark-toned soil) in TOARF values forms a buffer zone (support area, envelope) which includes surface reflectance (SURF) values as a special case in clear sky and flat terrain conditions.



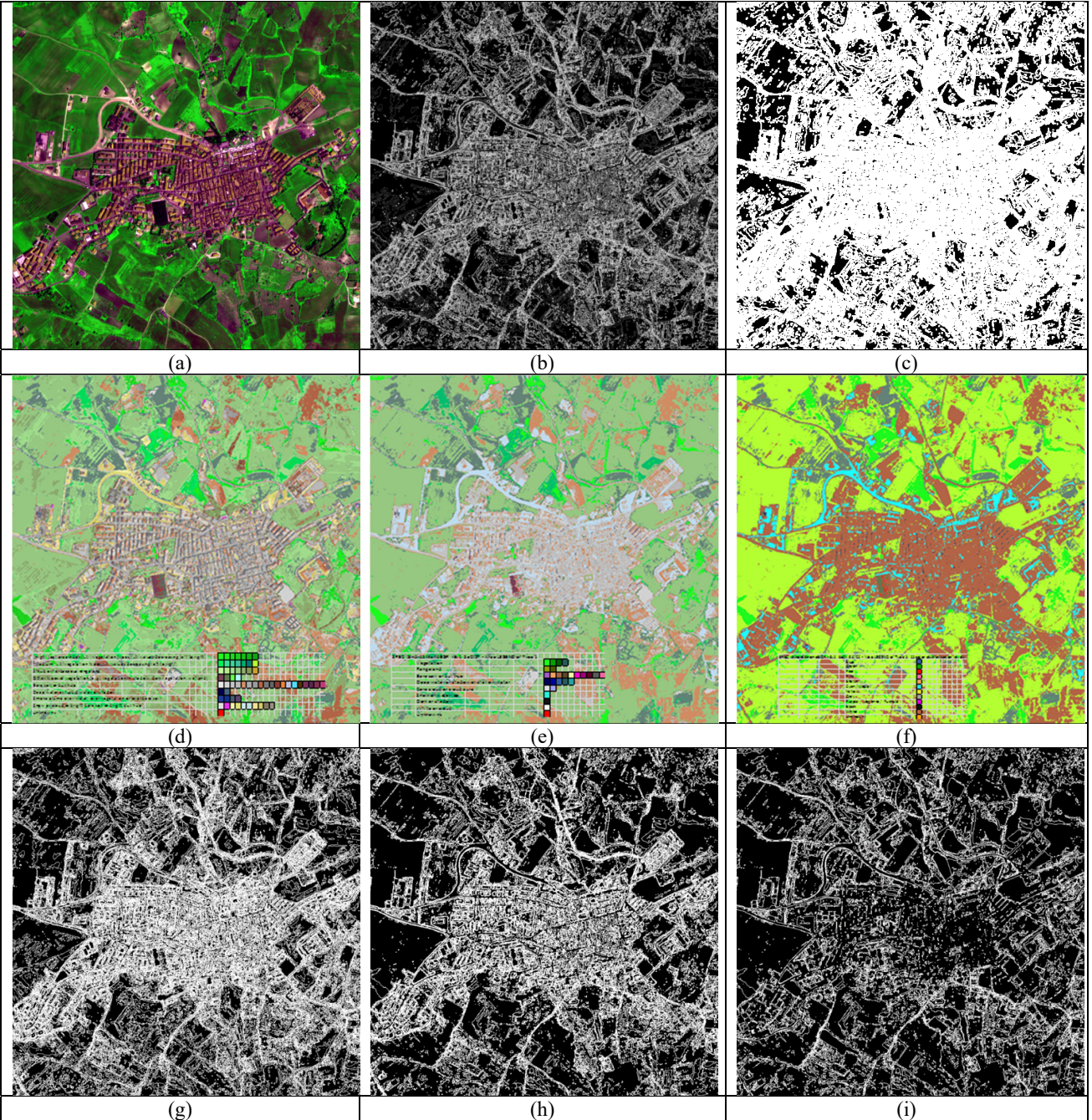
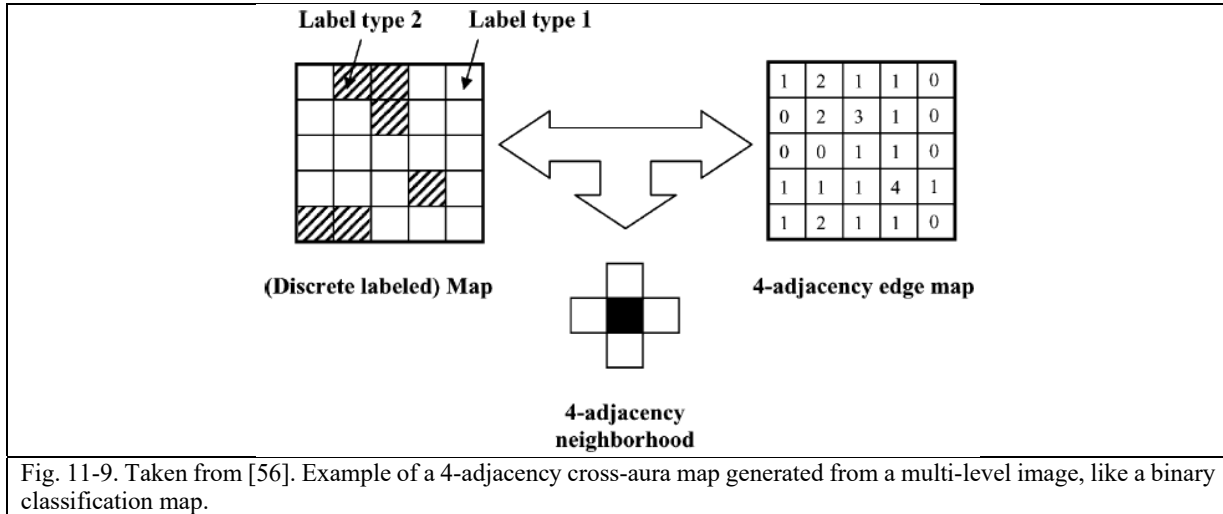


Fig. 11-8. (a) Zoomed area selected from the validation image shown in Fig. 11-4, depicted in false colors (R: band Red, G: band Near Infrared, B: band Blue), with histogram stretching for visualization purposes. (b) Three-level sum of 8-adjacency cross-aura measures shown in Fig. 11-8(g), (h) and (i). The per-pixel three-level cross-aura measure belongs to range  $\{0, 24 = 8 \times 3\}$ . (c) Binarization of the three-level sum of cross-aura measures, shown in Fig. 11-8(b). (d), (e), (f) Q-SIAM pre-classification map, at fine/ intermediate/ coarse discretization levels, corresponding to 61/28/12 spectral categories (see Table 11-3). (g), (h), (i) 8-adjacency cross-aura measure in range  $\{0, 8\}$  per pixel, generated from the Q-SIAM pre-classification map at fine/ intermediate/ coarse discretization levels shown in Fig. 11-8(d), (e) and (f) respectively.





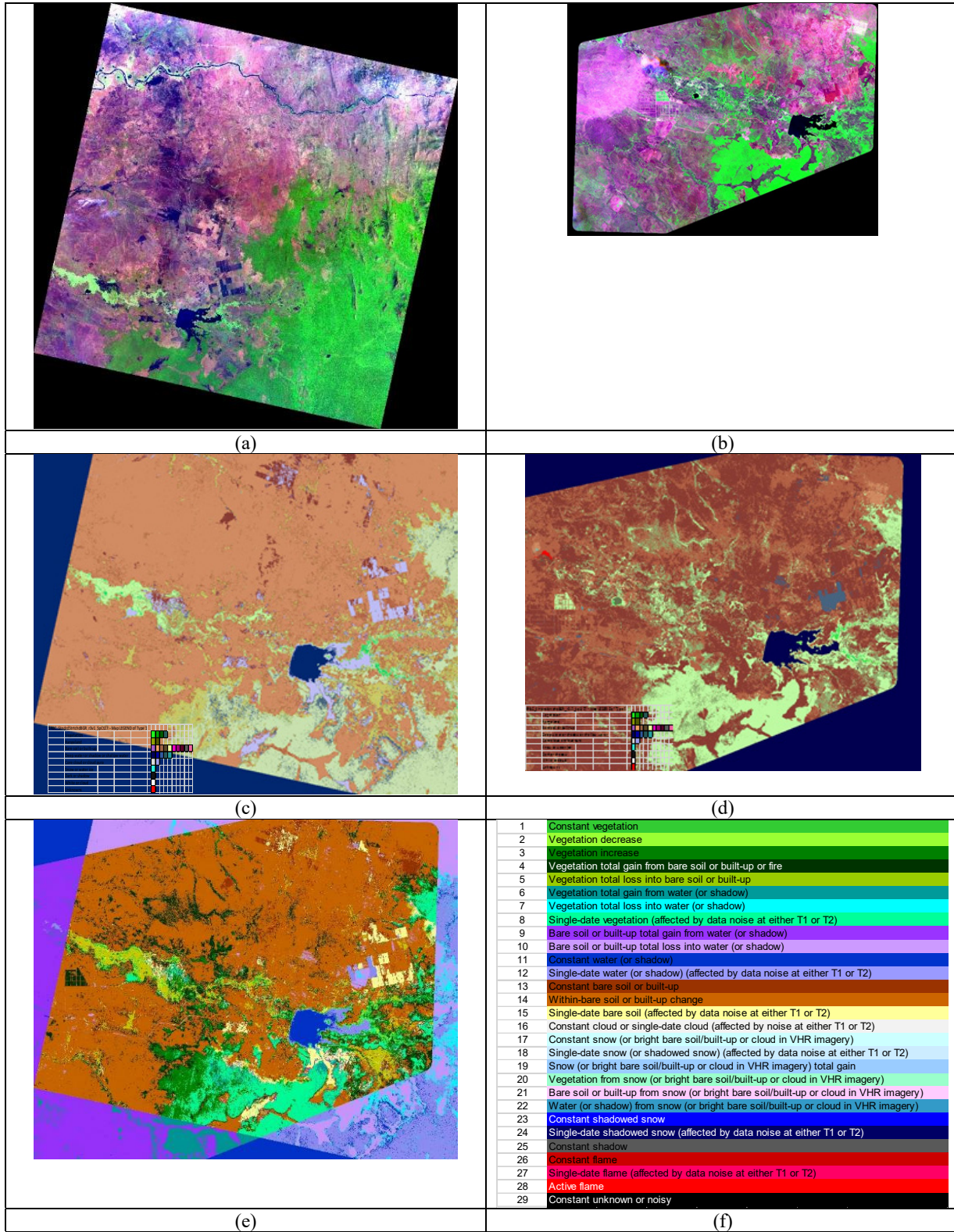


Fig. 11-10. Automatic SIAM-based post-classification change/no-change detection. (a) SPOT-5 image acquired on 2008-02-14, covering a surface area around Gambella, Ethiopia (DATASET\_NAME: SCENE 5 130-334), spatial resolution: 10 m, depicted in false colors (R: MIR, G: NIR, B: G). Radiometrically calibrated into TOARF values. Default EN6 histogram stretching, linear 2%. (b) Mosaic of 6 RapidEye images acquired on 2014-04-08, covering a surface area around Gambella, Ethiopia, spatial resolution: 5 m, depicted in false colors (R: R, G: NIR, B: B).

Radiometrically calibrated into TOARF values. Default EN6 histogram stretching, linear 2%. (c) S-SIAM pre-classification map depicted in pseudo colors, generated from the SPOT-5/RapidEye inter-image overlapping area, upscaled to 5 m. This inter-sensor SIAM's map legend consists of 33 "shared" spectral categories. (d) Q-SIAM pre-classification map depicted in pseudo colors, generated from the SPOT-5/RapidEye inter-image overlapping area, 5 m resolution. This inter-sensor SIAM's map legend consists of 33 "shared" spectral categories. (e) Bi-temporal inter-sensor post-classification land surface change/no-change detection, automatically generated from the two SIAM's pre-classification maps. (f) SIAM-based post-classification change/no-change map's legend, consisting of 29 spectral categories.

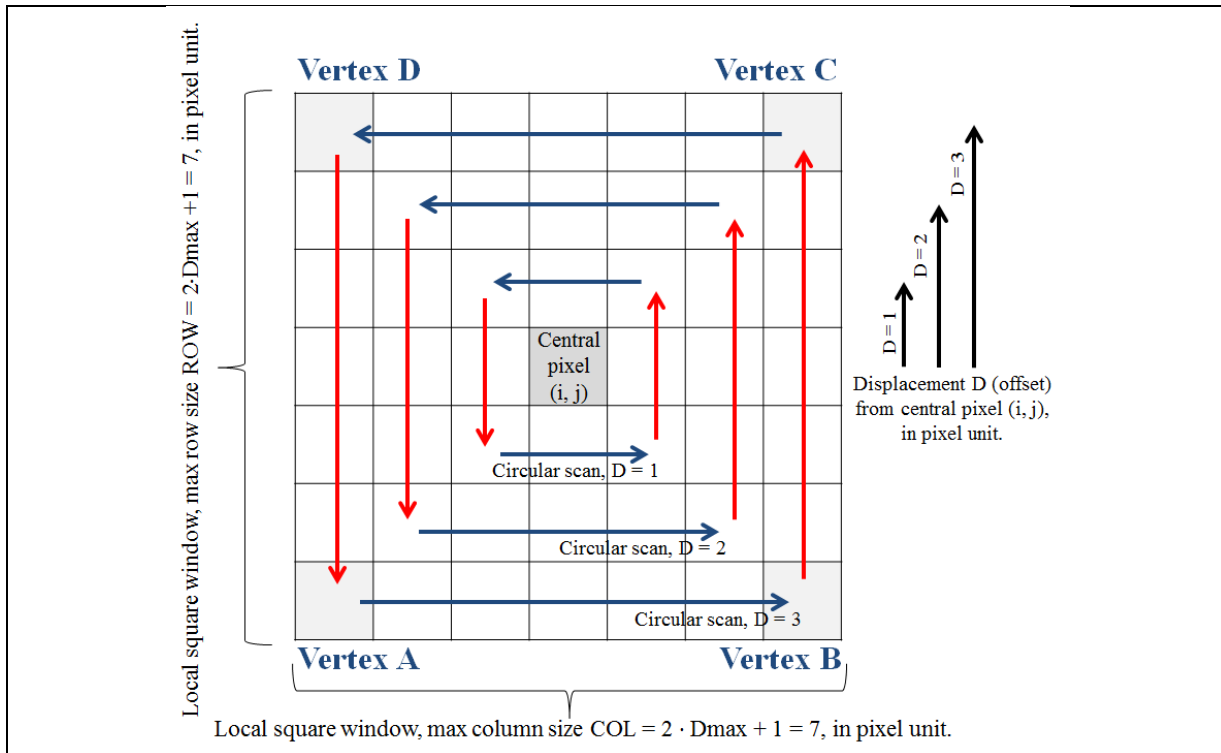


Fig. 11-11. Sketch of the TIMS-GLCM implementation, to collect 3<sup>rd</sup>-order statistics in the spatial domain. In this example, the scanning order of the local window centered on pixel (i,j) is counter-clockwise, for each spatial scale (radius, in this example,  $r = 1, 2, 3$  in pixel unit, where  $r$  is equivalent to a displacement  $D$  from the center pixel). The central pixel, whose gray level is identified as  $GL_1$ , provides the first  $GL$  value of each 3-tuple collected while scanning a complete circumference around the central pixel. The 2<sup>nd</sup> pixel is collected on the Right side – Bottom up (see arrow in red, pointing up), while the 3<sup>rd</sup> pixel is collected on the Left side – Top down (see arrow in red, pointing down). To close the circumference at a given radius  $r$ , the 2<sup>nd</sup> pixel is collected on the Top side – Right to left, while the 3<sup>rd</sup> pixel is collected on the Bottom side – Left to right. In each collected 3-tuple, the gray levels (GLs) of the 2<sup>nd</sup> and 3<sup>rd</sup> pixels are sorted, such that  $GL_2 < GL_3$ . The upper triangular TIMS-GLCM is input with values: row =  $GL_2$ , column =  $GL_3$ , depth =  $GL_1$ , where  $GL_2 < GL_3$ , also refer to Fig. 11-12.



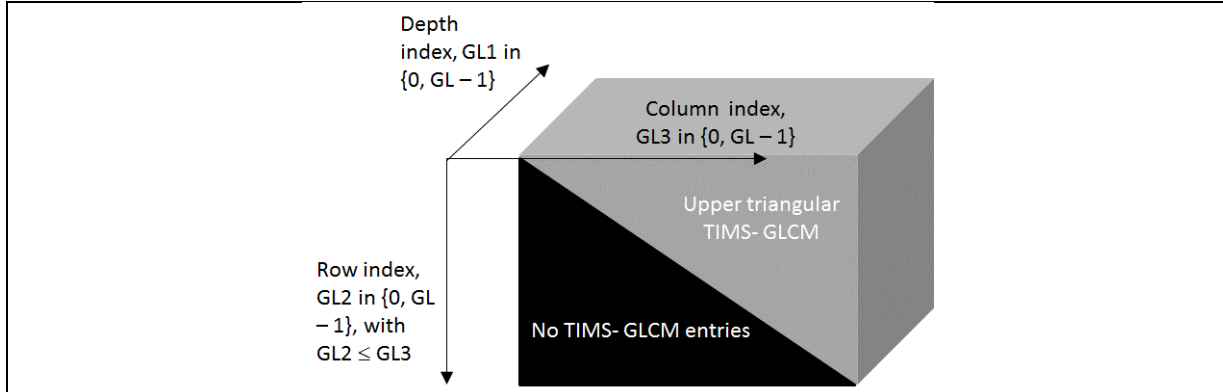


Fig. 11-12. Upper triangular TIMS-GLCM. An entry 3-tuple is: Row =  $GL2$ , Column =  $GL3$ , Depth =  $GL1$ , with  $GL2 \leq GL3$ .

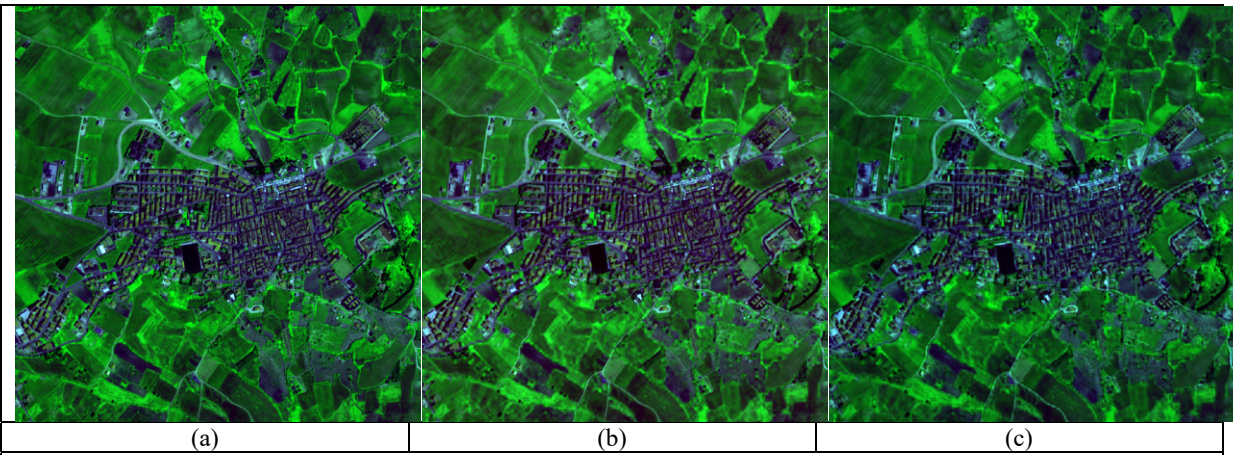


Fig. 11-13. Zoomed area selected from the validation image shown in Fig. 11-4, same as that shown in Fig. 11-8. Comparison of the reference image,  $MS_i$ , with the two fused  $MS^*_i$  images considered “best” by human subjects. All images are depicted in false colors (R: band Red, G: band Near Infrared, B: band Blue). Spatial resolution: 2.44 m. No histogram stretching is applied for visualization purposes. (a) Reference image; (b) PC3 CC; (c) HCS3 NN.



**Tables and table captions in Chapter 11**

Acronym	Tested MS image PAN-sharpening algorithm	Discriminative input parameter - Resampling algorithm
PC1_NN_PA	Principal Component	Nearest Neighbor (EN6)
PC2_B		Bilinear (EN6)
PC3_CC		Cubic Convolution (EN6)
GS1_NN	Gram - Schmidt	Nearest Neighbor (EN6)
GS2_B_PA		Bilinear (EN6)
GS3_CC_PA		Cubic Convolution (EN6)
CN2_PA	Color Normalized	Nearest Neighbor (EN6)
DWT1	Discrete Wavelet Transform	Nearest Neighbor (EN6)
DWT1_PA		Pixel Aggregate (IDL)
ATW2_PA	A Troust Wavelet Transform	Pixel Aggregate (IDL)
HCS3_NN	Hyperspectral color space	Nearest Neighbor (ERDAS)
HCS7_CC		Cubic Convolution (ERDAS)
EH1	Ehlers	Nearest Neighbor (ERDAS)
RM	Resolution Merge	Nearest Neighbor (ERDAS)

Table 11-1. Test set of MS image PAN-sharpening algorithms. For each algorithm, there were one or more system’s free-parameters to be user-defined. One input parameter was selected for discriminative purposes, i.e., its changes in value led to different runs of the same algorithm with different outcome. Other input parameters, if any, were kept fixed in the different runs by the same algorithm.

Either spatial or spectral quality	Human assessment
A	1 - Excellent
B	2 - Very good
C	3 - Good
D	4 - Fairly good
E	5 – Sufficient
F	6 - Insufficient (bad)
G	7 - Very bad

Table 11-2. Perceptual inter-image quality (similarity) assessment by human subjects: legend of the qualitative (categorical) variable.



SIAM, r88v5	Input bands	Preliminary classification map output products: Number of output spectral categories			
		Fine discretization levels	Intermediate discretization levels	Coarse discretization levels	Inter-sensor discretization levels (*)
L-SIAM	7 – B, G, R, NIR, MIR1, MIR2, TIR	96	48	18	33 * employed for inter-sensor post-classification change/no-change detection
S-SIAM	4 – G, R, NIR, MIR1	68	40	15	
AV-SIAM	4 – R, NIR, MIR1, TIR	83	43	17	
Q-SIAM	4 – B, G, R, NIR	61	28	12	

Table 11-3. The SIAM prior knowledge-based MS data quantizer is an EO system of systems, scalable to any past, existing or future MS imaging sensors, in compliance with the GEOSS implementation plan [4].

Product QI category	QI name and description	Metric	Acronym	Statistic analysis	Statistic order in the spatial domain
1. Context-insensitive (pixel-based) Position (row and column)-independent Spectral cost indexes (SPCTRL)	Mean scalar value, band-specific	MDB	MeanUnvrt	Unvrt	1 <sup>st</sup>
	Standard deviation scalar value, band-specific	MDB	StDvUnvrt	Unvrt	1 <sup>st</sup>
	Skewness scalar value, band-specific	MDB	SkwnsUnvrt	Unvrt	1 <sup>st</sup>
	Kurtosis scalar value, band-specific	MDB	KrtsUnvrt	Unvrt	1 <sup>st</sup>
	Entropy scalar value, band-specific	MDB	EntrpyUnvrt	Unvrt	1 <sup>st</sup>
2. Context-insensitive Position-dependent Spectral cost indexes (SPCTRL & SPTL1)	Cumulative (image-wide) per-pixel absolute difference in a pair of SIAM-based pre-classification maps (post-classification change detection)	-	PostClChngDtctnMvrt	Mvrt	1 <sup>st</sup>
	Inverse correlation coefficient (to be considered a cost value, to be minimized) computed inter-image band-specific	ABA	InvrCrlnBivrt	Bivrt	1 <sup>st</sup>
3. Context-sensitive Position-independent Spectral cost indexes (SPCTRL & SPTL2)	3 <sup>rd</sup> -order Contrast scalar value, band-specific	MDB	3odrCntrstUnvrt	Unvrt	3 <sup>rd</sup>
	3 <sup>rd</sup> -order Energy scalar value, band-specific	MDB	3odrEnrgyUnvrt	Unvrt	3 <sup>rd</sup>
	3 <sup>rd</sup> -order Large Number Emphasis scalar value, image band-specific	MDB	3odrLneUnvrt	Unvrt	3 <sup>rd</sup>
	Mean (image-wide) per-pixel SIAM-based multi-level 8-adjacency cross-aura contour measure (in range {0, 8} per pixel)	MD	CntourXauraMvrt	Mvrt	1 <sup>st</sup>
4. Context-sensitive Position-dependent Spectral cost indexes (SPCTRL & SPTL1 & SPTL2)	Cumulative (image-wide) per-pixel absolute difference in a pair of SIAM-based multi-level 8-adjacency cross-aura binary contour maps (in range {0, 1} per pixel)	MD	BinaryCntourMvrt	Mvrt	1 <sup>st</sup>

Table 11-4. Novel categorization of MS image PAN-sharpening product QIs and quality metrics, based on three nominal scales: (a) univariate (one-channel, Unvrt)/ bivariate (two-channel, Bivrt)/ multivariate (multi-channel, Mvrt), (b) 1st- to 3rd-order statistic in the spatial domain, (c) categories 1 to 4: SPCTRL, SPCTRL & SPTL1, SPCTRL & SPTL2, SPCTRL & SPTL1 & SPTL2. The Inter-QI metric functions (Metric) considered are: Minkowski distance of order 1 (MD) across bands (MDB), Across-band average (ABA).



Algorithm	Acronym	Subjective quality assessment	
		Spectral quality	Spatial quality
Principal Component	PC1 NN PA	C	A
	PC2 B	B	A
	PC3 CC	C	A
Gram - schmidt	GS1 NN	C	D
	GS2 B PA	D	D
	GS3 CC PA	D	D
Color Normalized	CN2 PA	E	E
Discrete Wavelet Transform	DWT1	F	G
	DWT1 PA	F	G
A Trouw Wavelet Transform	ATW2 PA	F	G
Hyperspherical Color Space	HCS3 NN	A	B
	HCS7 CC	B/C	D
Ehlers	EH1	G	E
Resolution Merge	RM	C	E

Table 11-5. Visual score of the tested MS image PAN-sharpened outcome. Green highlight: first- and second-best choice by human subjects.

Algorithm	SPCTRL					SPCTRL & SPTL1		SPCTRL & SPTL2				SPCTRL & SPTL1 & SPTL2
	MeanUnvrt	StDvUnvrt	SkwnsUnvrt	KrtsUnvrt	EntpryUnvrt	PostCIChn gDtctnMvrt	InvrnCrt nBivrt	3ordrC ntrstUnvrt	3ordrE nrgyUnvrt	3ordrLn eUnvrt	CntourX auraMvrt	BinaryCnto urMvrt
PC1_NN_PA	0.033	0.180	0.366	2.324	0.011	0.271	0.006	0.014	0.011	0.011	3.867	0.233
PC2_B	0.001	0.199	0.348	2.213	0.037	0.261	0.005	0.012	0.010	0.170	3.753	0.231
PC3_CC	0.002	0.029	0.356	2.270	0.058	0.264	0.006	0.066	0.013	0.030	3.699	0.232
GS1_NN	0.000	0.171	0.300	1.847	0.082	0.336	0.003	0.113	0.016	0.300	4.214	0.249
GS2_B_PA	0.153	0.178	0.299	1.855	0.081	0.299	0.007	0.019	0.011	0.510	4.093	0.254
GS3_CC_PA	0.218	0.176	0.302	1.901	0.106	0.299	0.008	0.009	0.012	0.724	4.088	0.243
CN1	3.702	0.411	0.312	2.117	0.149	0.323	0.005	0.289	0.019	15.762	4.221	0.301
DWT1	0.086	0.169	0.174	0.635	1.457	0.329	0.085	0.046	0.015	0.113	4.433	0.381
DWT1_PA	0.072	0.226	0.044	1.206	1.650	0.241	0.039	0.741	0.050	0.518	4.261	0.394
ATW2_PA	0.020	0.359	0.222	1.792	1.753	0.238	0.031	0.791	0.052	0.332	4.272	0.409
HCS3_NN	0.099	0.216	0.104	0.577	0.028	0.268	0.010	0.129	0.001	0.502	3.918	0.223
HCS7_CC	0.409	0.627	0.090	0.002	0.171	0.319	0.039	0.710	0.017	0.916	4.020	0.237
EH1	46.67	7.388	2.124	12.517	0.718	0.950	0.201	1.069	0.127	277.94	4.885	0.383
RM	4.349	0.961	0.338	2.229	0.059	0.462	0.007	0.156	0.009	16.923	4.916	0.330
MEAN	3.987	0.806	0.384	2.392	0.454	0.347	0.032	0.297	0.026	22.482	4.189	0.293
STDV	12.37	1.909	0.512	3.004	0.658	0.182	0.054	0.364	0.033	73.752	0.366	0.071
Standardized QI (z = (QI - MEAN) / STDV, such that E[z] = 0, STDV[z] = 1.												
PC1_NN_PA	-0.320	-0.328	-0.036	-0.023	-0.673	-0.419	-0.484	-0.777	-0.456	-0.305	-0.879	-0.837
PC2_B	-0.322	-0.318	-0.071	-0.059	-0.634	-0.472	-0.501	-0.783	-0.494	-0.303	-1.190	-0.865
PC3_CC	-0.322	-0.407	-0.055	-0.040	-0.603	-0.458	-0.497	-0.634	-0.412	-0.304	-1.338	-0.849
GS1_NN	-0.322	-0.333	-0.164	-0.181	-0.566	-0.062	-0.552	-0.506	-0.312	-0.301	0.070	-0.620
GS2_B_PA	-0.310	-0.329	-0.167	-0.179	-0.567	-0.263	-0.469	-0.764	-0.449	-0.298	-0.261	-0.544
GS3_CC_PA	-0.305	-0.330	-0.161	-0.163	-0.529	-0.265	-0.463	-0.791	-0.415	-0.295	-0.275	-0.705
CN1	-0.023	-0.207	-0.142	-0.091	-0.465	-0.131	-0.512	-0.022	-0.207	-0.091	0.089	0.113
DWT1	-0.315	-0.334	-0.410	-0.585	1.524	-0.102	0.986	-0.689	-0.345	-0.303	0.668	1.239
DWT1_PA	-0.317	-0.304	-0.664	-0.395	1.817	-0.583	0.133	1.218	0.749	-0.298	0.198	1.422
ATW2_PA	-0.321	-0.234	-0.317	-0.200	1.973	-0.597	-0.027	1.353	0.790	-0.300	0.228	1.625
HCS3_NN	-0.314	-0.309	-0.547	-0.604	-0.647	-0.431	-0.415	-0.463	-0.753	-0.298	-0.739	-0.976
HCS7_CC	-0.289	-0.094	-0.574	-0.795	-0.430	-0.155	0.130	1.131	-0.286	-0.292	-0.461	-0.789
EH1	3.451	3.447	3.398	3.370	0.401	3.304	3.141	2.117	3.106	3.464	1.902	1.264
RM	0.029	0.081	-0.090	-0.054	-0.601	0.632	-0.471	-0.389	-0.516	-0.075	1.987	0.522
MEAN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000





STDV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Table 11-6. Test QuickBird MS image PAN-sharpening: product QI (actually, cost index) values and standardized QI values. Product QI (actually, cost index) category 1: SPCTRL. Product QI (actually, cost index) category 2: SPCTRL & SPTL1. Product QI (actually, cost index) category 3: SPCTRL & SPTL2. Product QI (actually, cost index) category 4: SPCTRL & SPTL1 & SPTL2. Cost index values to be minimized (best when smaller).

Algorithm	SPCTRL		SPCTRL & SPTL1				SPCTRL & SPTL2		SPCTRL & SPTL1 & SPTL2	
	Sum	PDPR	Sum, case (i)	PDPR, case (i)	Stndrdzd PostClChngDtctnMvrt, case (ii)	PDPR, case (ii)	Sum	PDPR	Stndrdzd BinaryCntourMvrt	PDPR
PC1_NN_PA	-1.38	8	-0.90	3	-0.41859	6	-2.42	3	-0.837	4
PC2_B	-1.41	7	-0.97	1	-0.47177	3	-2.77	1	-0.865	2
PC3_CC	-1.43	6	-0.96	2	-0.45796	4	-2.69	2	-0.849	3
GS1_NN	-1.57	3	-0.61	9	-0.06234	12	-1.05	7	-0.620	7
GS2_B_PA	-1.55	4	-0.73	5	-0.26270	8	-1.77	6	-0.544	8
GS3_CC_PA	-1.49	5	-0.73	6	-0.26511	7	-1.78	5	-0.705	6
CN1	-0.93	9	-0.64	7	-0.13064	10	-0.23	9	0.113	9
DWT1	-0.12	11	0.88	13	-0.10176	11	-0.67	8	1.239	11
DWT1_PA	0.14	12	-0.45	10	-0.58338	2	1.87	12	1.422	13
ATW2_PA	0.90	13	-0.62	8	-0.59653	1	2.07	13	1.625	14
HCS3_NN	-2.42	1	-0.85	4	-0.43093	5	-2.25	4	-0.976	1
HCS7_CC	-2.18	2	-0.02	11	-0.15457	9	0.09	10	-0.789	5
EH1	14.07	14	6.44	14	3.30399	14	10.59	14	1.264	12
RM	-0.63	10	0.16	12	0.63228	13	1.01	11	0.522	10

Table 11-7. Sum of within-category standardized product QIs (actually, cost indexes), to be minimized (best when more negative) and partial ranking for each category of product QIs (product partial rank, PDPR). Product QI (actually, cost index) category 1: SPCTRL - Sum of five standardized cost indexes to be minimized: MeanUnvrt, StDvUnvrt, SkwnsUnvrt, KrtsUnvrt, En-tryUnvrt. Product QI (actually, cost index) category 2: SPCTRL & SPTL1 – Case (i) = Sum of two standardized cost indexes to be minimized: PostClChngDtctnMvrt, InvrCrlnBivrt; Case (ii) = InvrCrlnBivrt is omitted, i.e., only PostClChng-DtctnMvrt is considered. Product QI (actually, cost index) category 3: SPCTRL & SPTL2 - Sum of four standardized cost indexes to be minimized: 3ordrCntrstUnvrt, 3ordrEnrgyUnvrt, 3ordrLneUnvrt, CntourXauraMvrt. Product QI (actually, cost index) category 4: SPCTRL & SPTL1 & SPTL2 - Single standardized cost index to be minimized: BinaryCntourMvrt.

Algorithm	Processing time (cost index 1, in seconds)	PSPR1	No. of system's free-parameters (cost index 2)	PSPR2
PC1_NN_PA	00:02:56:541	8	1	1
PC2_B	00:03:10:061	9	1	1
PC3_CC	00:03:23:870	10	1	1
GS1_NN	00:01:24:735	6	2	2
GS2_B_PA	00:01:18:118	5	2	2
GS3_CC_PA	00:01:39:022	7	2	2
CN1	00:00:28:820	1	1	1
DWT1	00:22:10:514	14	3	3
DWT1_PA	00:21:44:012	13	3	3
ATW2_PA	00:16:45:224	12	1	1
HCS3_NN	00:01:12:256	4	4	4
HCS7_CC	00:01:10:852	3	4	4
EH1	00:04:42:274	11	5	5
RM	00:00:58:713	2	5	5

Table 11-8. Battery of process QIs (actually, cost indexes, to be minimized) adopted by the new evaluation procedure. The number of system's free-parameters is a cost index, inversely related to the degree of automation. Process QI (actually, cost index) partial rank: PSPR.



Algorithm	SPCTRL, PDPR	SPCTRL & SPTL1, case (i), PDPR	SPCTRL & SPTL1, case (ii), PDPR	SPCTRL & SPTL2, PDPR	SPCTRL & SPTL1 & SPTL2, PDPR	Sum of PDPRs, with (i) - Case A	PDFR, with (i) - Case A	Sum of PDPRs, with (ii) - Case C	PDFR, with (ii) - Case C
PC1_NN_PA	8	3	6	3	4	18	4	21	4
PC2_B	7	1	3	1	2	11	2	13	2
PC3_CC	6	2	4	2	3	13	3	15	3
GS1_NN	3	9	12	7	7	26	7	29	8
GS2_B_PA	4	5	8	6	8	23	6	26	6
GS3_CC_PA	5	6	7	5	6	22	5	23	5
CN2_PA	9	7	10	9	9	34	9	37	9
DWT1	11	13	11	8	11	43	10	41	11
DWT1_PA	12	10	2	12	13	47	12	39	10
ATW2_PA	13	8	1	13	14	48	13	41	11
HCS3_NN	1	4	5	4	1	10	1	11	1
HCS7_CC	2	11	9	10	5	28	8	26	6
EH1	14	14	14	14	12	54	14	54	14
RM	10	12	13	11	10	43	10	44	13

Table 11-9. Product final ranks (PDFR), computed from the sum of the individual product quality partial ranks (PDPRs) collected from the four categories of product QIs, refer to Table 11-7. Case A is alternative to Case C. The latter holds when the InvsCrlnBivrt cost index is removed from the product QI category SPCTRL & SPTL1. Noteworthy, in these experiments, the sole QI be-longing to category 4, SPCTRL & SPTL1 & SPTL2, is the individual indicator that best approximates (which is highly cor-related with) the final ranks A and C, although no single "universal" quality indicator can exist. Yellow highlight: first-best choice by quantitative quality estimation. Red highlight: second-best choice by quantitative quality estimation.

Algorithm	SPCTRL, PDPR	SPCTRL & SPTL1, case (i), PDPR	SPCTRL & SPTL1, case (ii), PDPR	SPCTRL & SPTL2, PDPR	SPCTRL & SPTL1 & SPTL2, PDPR	PSPR1, Processing time	PSPR2, No. of system's free-parameters	Sum of PDPRs and PSPRs, Case B	PPFR, Case B	Sum of PDPRs and PSPRs, Case D	PPFR, Case D
PC1_NN_PA	8	3	6	3	4	8	1	27	4	30	4
PC2_B	7	1	3	1	2	9	1	21	2	23	2
PC3_CC	6	2	4	2	3	10	1	24	3	26	3
GS1_NN	3	9	12	7	7	6	2	34	7	37	8
GS2_B_PA	4	5	8	6	8	5	2	30	5	33	6
GS3_CC_PA	5	6	7	5	6	7	2	31	6	32	5
CN2_PA	9	7	10	9	9	1	1	36	9	39	9
DWT1	11	13	11	8	11	14	3	60	11	58	13
DWT1_PA	12	10	2	12	13	13	3	63	13	55	12
ATW2_PA	13	8	1	13	14	12	1	61	12	54	11
HCS3_NN	1	4	5	4	1	4	4	18	1	19	1
HCS7_CC	2	11	9	10	5	3	4	35	8	33	6
EH1	14	14	14	14	12	11	5	70	14	70	14
RM	10	12	13	11	10	2	5	50	10	51	10

Table 11-10. Product & Process final ranks (PPFRs), computed from the sum of product quality partial ranks (PDPRs) collected from each of the four categories of product's QI, refer to Table 11-7, in addition to the two process quality partial ranks (PSPRs), reported in Table 11-8. Case B is alternative to Case D. The latter holds when the InvsCrlnBivrt cost index is removed from the product QI category SPCTRL & SPTL1, refer to case (ii). Yellow highlight: first-best choice by quantitative quality estimation. Red highlight: second-best choice by quantitative quality estimation.



Algorithm	SAM (cost index)	PDFR, SAM	ERGAS (cost index)	PDFR, ERGAS	Q4 (quality index)	PDFR, Q4
PC1 NN PA	2.1950	3	2.2207	3	0.996300	4
PC2 B	2.0378	2	2.13219	2	0.996791	2
PC3 CC	2.0369	1	2.12856	1	0.996972	1
GS1 NN	3.0045	6	2.92803	10	0.993670	6
GS2 B PA	2.6781	4	2.42699	5	0.996542	3
GS3 CC PA	3.8092	5	2.40221	4	0.995557	5
CN2 PA	3.9176	7	2.5552	7	0.991644	7
DWT1	5.9411	13	4.32709	13	0.918766	13
DWT1 PA	4.5145	10	2.80043	8	0.963695	11
ATW2 PA	4.1037	9	2.5721	6	0.972055	10
HCS3 NN	3.9481	8	2.92367	9	0.975420	9
HCS7 CC	5.6735	12	3.44937	11	0.960850	12
EH1	6.1258	14	14.0384	14	0.437443	14
RM	5.1103	11	3.95332	12	0.987463	8

Table 11-11. Product final ranks (PDFRs) obtained from three popular multivariate (multi-band) QIs: SAM (angle in range [0, 90 degrees]), it is a cost to be minimized; ERGAS  $\geq 0$ , it is a cost to be minimized; “universal” (heterogeneous) Q4 (in range [0, 1]), it is a quality index to be maximized. Yellow highlight: first-best choice by quantitative quality estimation. Red highlight: second-best choice by quantitative quality estimation.

Algorithm	PDFR, Case A	PPFR, Case B	PDFR, Case C	PPFR, Case D	Subjective quality assessment		PDFR		
					Spectral	Spatial	SAM	ERGA S	Q4
PC1 NN PA	4	4	4	4	C	A	3	3	4
PC2 B	2	2	2	2	B	A	2	2	2
PC3 CC	3	3	3	3	C	A	1	1	1
GS1 NN	7	7	8	8	C	D	6	10	6
GS2 B PA	6	5	6	6	D	D	4	5	3
GS3 CC PA	5	6	5	5	D	D	5	4	5
CN2 PA	9	9	9	9	F	E	7	7	7
DWT1	10	11	11	13	B	F	13	13	13
DWT1 PA	12	13	10	12	B	F	10	8	11
ATW2 PA	13	12	11	11	B	F	9	6	10
HCS3 NN	1	1	1	1	A	B	8	9	9
HCS7 CC	8	8	6	6	B/C	E	12	11	12
EH1	14	14	14	14	G	G	14	14	14
RM	10	10	13	10	C	E	11	12	8

Table 11-12. Comparison between the new protocol's final ranks of Product QIs (PDFRs in Cases A and C, refer to Table 11-9; Case C is alternative to Case A, because the former omits cost index *InvrCrlnBivrt*) and Product & Process QIs (PPFRs in Cases B and D, refer to Table 11-10; Case D is alternative to Case B, because the former omits cost index *InvrCrlnBivrt*), the subjective ranks collected from human subjects (refer to Table 11-5) and the product final ranks (PDFRs) provided by three popular QIs, specifically, SAM, ERGAS and Q4, refer to Table 11-11. Green highlight: first-best and second-best choice by human subjects. Yellow highlight: first-best choice by quantitative quality estimation. Red highlight: second-best choice by quantitative quality estimation.



	Spearman's rank correlation coefficient (SRCC)				
	ERGAS	SAM	Q4	PDFR, Case C	PPFR, Case D
ERGAS	x	0.9253	0.8593	0.6967	0.6725
SAM	x	x	0.9692	0.7495	0.7560
Q4	x	x	x	0.6659	0.7209
PDFR, Case C	x	x	x	x	0.9626
PPFR, Case D	x	x	x	x	x

Table 11-13. The Spearman's rank correlation coefficient (SRCC) values, in range  $[-1, 1]$ , generated from pairwise comparisons of ranked variables: ERGAS, average SAM, Q4, PDFR - Case C and PPFR - Case D, refer to Table 11-12. Unlike the Pearson's correlation coefficient (PCC), the SRCC assesses how well the relationship between two ranked variables can be described using a monotonically increasing or decreasing function, even if their relationship is not linear.



## 12 Technical report 2 (made available in the public archive arXiv: 1701.04256): Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+

### Motivation and Contributions to the Dissertation

Among the original pair of expert systems (prior knowledge-based decision trees) for color naming presented in Chapter 3 (Technical report 1) and adopted by an Earth observation (EO) Image Understanding for Semantic Querying (EO-IU4SQ) system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the off-the-shelf Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program was submitted to a Stage 4 validation for systematic ESA EO Level 2 product generation, refer to Chapter 7 (Manuscript 4) and Chapter 8 (Manuscript 5). In Chapter 9 (Manuscript 6) the off-the-shelf RGB Image Automatic Mapper™ (RGBIAM™) lightweight computer program for true- or false-color RGB cube partitioning into a dictionary of color names was discussed in detail. By definition of the European Space Agency (ESA), an EO Level 2 product comprises a multi-spectral (MS) image corrected for geometric, atmospheric, topographic and adjacency effects, stacked with its data-derived general-purpose, user- and application-independent scene classification map (SCM), whose legend includes quality layers such as cloud and cloud-shadow. The present Chapter 12 (Technical report 2) reviews existing cloud and cloud-shadow detectors and proposes a novel hybrid (combined deductive and inductive) EO image understanding system (EO-IUS) design (architecture) for automatic spatial context-sensitive cloud/cloud-shadow detection in multi-source MS imagery, where input information sources include the SIAM and RGBIAM color maps automatically generated from a single-date MS image.

In Chapter 1 (Doctoral Research Objectives), the modular design of a hybrid (combined deductive and inductive) feedback EO-IU subsystem, implemented in operating mode as necessary not sufficient pre-condition of a closed-loop EO-IU4SQ system, was sketched in Fig. 1-3. For the sake of clarity, Fig. 1-3 is reported hereafter, where processing blocks involved with and described by the present Chapter 12 (Technical report 2) are color filled.

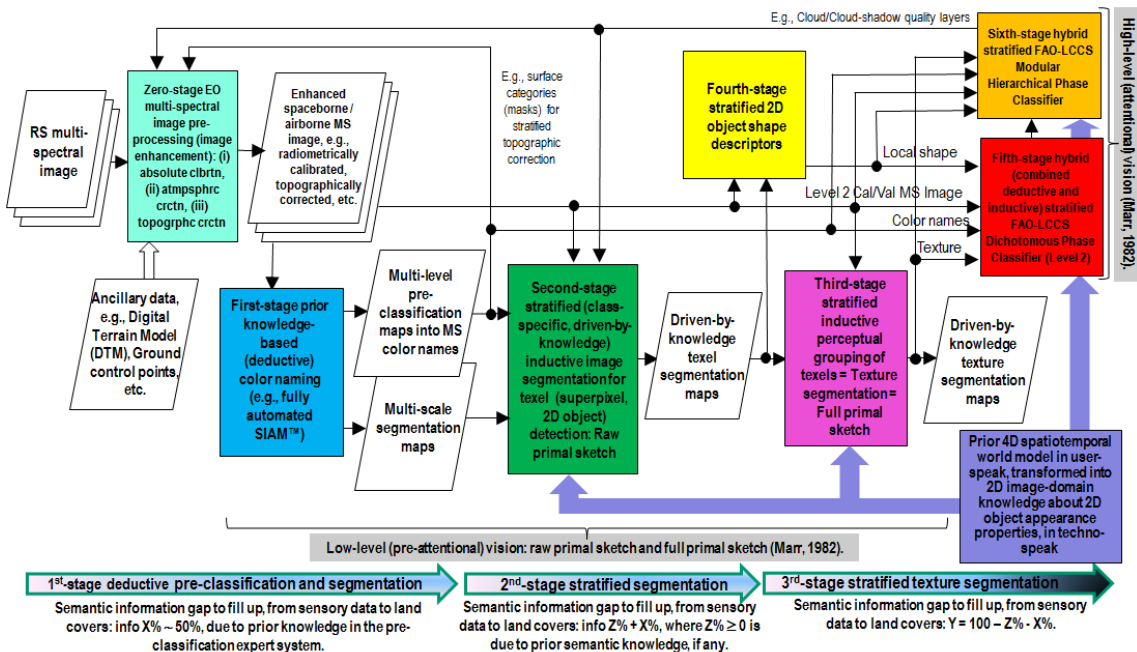


Fig. 1-3 revisited. Original six-stage hybrid (combined deductive/ top-down/ physical model-based and inductive/ bottom-up/ statistical model-based) EO image understanding system (EO-IUS) design provided with feedback loops. It pursues a convergence-of-evidence approach to computer vision (CV) whose goal is scene-from-image reconstruction and understanding. This CV system architecture is adopted by the EO-IU subsystem implemented by the proposed EO image understanding for semantic querying (EO-IU4SQ) system prototype {42}. This hybrid feedback inference system architecture is alternative to feedforward inductive learning-from-data inference adopted by a large majority of the CV and the remote sensing (RS) literature. Rectangles depicted with color fill identify processing blocks involved with and described by the present Chapter 12 (Technical report 2).



**Invitation to tender ESA/AO/1-8373/15/I-NB – “VAE: Next Generation EO-based Information Services”**

**Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images –**

**AutoCloud+**

Author: Andrea Baraldi, Senior scientist: .....

Dept. of Agricultural and Food Sciences, University of Naples Federico II, Portici (NA), Italy (e-mail: [andrea6311@gmail.com](mailto:andrea6311@gmail.com))  
 Paris Lodron University of Salzburg (PLUS), Austria / Department of Geoinformatics (Z\_GIS)

Date: Nov. 3, 2015.

Prof. Dr. Josef Strobl, Head of Z\_GIS-PLUS: .....

**12.1 Introduction**

The proposed Earth observation (EO)-based value adding system (EO-VAS), hereafter identified as AutoCloud+, consists of an innovative EO image understanding system (EO-IUS) design and implementation capable of automatic spatial context-sensitive cloud/cloud-shadow detection in multi-source multi-spectral (MS) EO imagery, whether or not radiometrically calibrated, acquired by multiple platforms, either spaceborne or airborne, including unmanned aerial vehicles (UAVs). It is worth mentioning that the same EO-IUS architecture is suitable for a large variety of EO-based value-adding products and services, including: (i) low-level image enhancement applications, such as automatic MS image topographic correction, co-registration, mosaicking and compositing, (ii) high-level MS image land cover (LC) and LC change (LCC) classification and (iii) content-based image storage/retrieval in massive multi-source EO image databases (“big data” mining).

**12.2 EO-VAS objectives, technical requirements and proposed approach**

Since the proposed EO-VAS for cloud/cloud-shadow detection in multi-source MS images has not been published yet, the present project proposal as well as any further documentation regarding the proposed activity shall be regarded as “Proprietary Sensitive Information”, subject to Articles 6.1.2 and 6.1.3 of the ESA Contract No. 4000xxxxx/15/I-NB, ARTICLE 6 –Information to be provided by the Contractor – Protection of information, whose quotes are the following.

- “6.1.2 For the purpose of this Contract Proprietary Sensitive Information... The Contractor shall not mark any (electronic) documentation as Proprietary Sensitive Information, unless agreed in advance with the Agency. Any request from the Contractor shall be submitted in writing accompanied by an appropriate justification.”
- “6.1.3 Neither Party shall disclose any documentation obtained from the other Party, and which both Parties recognise as being Proprietary Sensitive Information without the other Party’s previous written authorisation.”

**12.2.1 EO-VAS aims and degrees of innovation**

In the remote sensing (RS) community, a well-known prerequisite for clear-sky RS image compositing [1]-[5], suitable for further retrieval of land surface variables, either quantitative [6], such as biophysical variables, e.g., the leaf area index (LAI), or categorical (nominal) variables [6], such as LC/LCC classes, is accurate masking of clouds and cloud shadows, see Fig. 12-1 to Fig. 12-3. Intuitively, cloud contamination is a relevant problem in LCC analysis, because unflagged clouds may be mapped as false LCC occurrences.

In compliance with the Quality Assurance Framework for Earth Observation (QA4EO) guidelines, developed by the intergovernmental Group on EOs (GEO) [7], the ambitious goal of the present software project is to undertake the first research and technological development (RTD) of an EO-IUS software pipeline capable of cloud and cloud shadow detection in one input EO multi-source MS image subject to the following RTD project requirements specification. The proposed EO-IUS must be: (I) automatic, i.e., it requires no user’s interaction. (II) In operating mode, i.e., ready-for-use, by scoring high in a set of metrological/statistically-based quantitative quality indicators (Q<sup>2</sup>Is) of operativeness (Q<sup>2</sup>IOs), to be RS community-agreed upon. Proposed Q<sup>2</sup>IOs to be jointly maximized encompass [8], [9]: (i) degree of automation,



(ii) accuracy, (iii) efficiency, (iv) robustness to changes in input parameters, (v) robustness to changes in input data, (vi) scalability/transferability, (vii) timeliness, from data acquisition to data-derived product generation, and (viii) economy (vice versa, costs in manpower and computer power must be kept low). Noteworthy, this set of Q<sup>2</sup>IOs has never been adopted in the RS literature on a regular basis. (III) Sensor-independent, which means: (a) multi-scale, from regional to global spatial extents, (b) multi-resolution, from coarse ( $\approx 1$  km) to very high ( $< 1$  m), and (c) multi-platform, either spaceborne or airborne, including unmanned aerial vehicles (UAVs) [10]. (IV) Input with an MS image that is either: (a) radiometrically calibrated into top-of-atmosphere (TOA) reflectance (TOARF) or surface reflectance (SURF) values [11], or (b) uncalibrated. Although highly recommended in compliance with the QA4EO guidelines [7], radiometric calibration of digital numbers (DNs) is not considered mandatory by the present RTD project [11], to cope with consumer-level color cameras typically mounted on board light-weight UAVs or terrestrial photocameras [10].

Conceived to outperform existing state-of-the-art cloud/cloud-shadow detectors (see Fig. 12-4(a)), which are typically semi-automatic, site-specific and sensor-dependent (see Table 12-3), the aforementioned RTD software project's requirements, (I) to (IV), are ambitious, but realistic. They rely upon several EO-IUS software units already implemented, tested and validated by third-parties in recent years [8], [9], [12]-[14]. These software units can be combined according to an innovative EO-IUS architecture, see Fig. 12-4(b), featuring hybrid inference mechanisms (combined deductive/top-down/prior knowledge-based inference with inductive/bottom-up/learning-from-data inference) and provided with feedback loops. According to Marr [15], the linchpin of success of any information processing system is addressing the level of understanding of computational theory (system design), rather than algorithms or implementation. Noteworthy, in the EO-IUS design proposed in Fig. 12-4(b), first-stage prior knowledge-based inference (analogous to genotype) predates and conditions second-stage inductive data learning (equivalent to phenotype), because the latter is inherently ill-posed and requires *a priori* knowledge in addition to data to become better posed for numerical solution [21]. Feedback loops allow to back-project high-level categorical variables onto input quantitative variables, to accomplish either data enhancement, such as automatic EO image topographic correction [35] (see Fig. 12-5), or stratified (masked, conditioned) biophysical variable estimation, e.g., LAI estimation [54]. The hybrid feedback EO-IUS architecture shown in Fig. 12-4(b) is alternative to that adopted by the mainstream RS community, whose EO-IUSs adopt a feedforward inductive data learning strategy.

In addition to being considered realistic, the present RTD software project can be assessed as potentially relevant for the whole RS community. If successful, it would provide the first proof-of-concept that the proposed novel EO-IUS design and implementation strategies are capable of transforming multi-source EO image "big data" into operational, comprehensive and timely information products, e.g., cloud/cloud-shadow masks, in compliance with the QA4EO recommendations and with several ongoing RS international programs [16], [17]. This first proof-of-concept would open a wide spectrum of future research, educational and market opportunities, including the following.

(A) Accomplish automatic estimation of either continuous variables, such as topographically corrected TOARF values [35] (see Fig. 12-5) or biophysical variables, e.g., LAI [54], or categorical variables, such as LC/LCC classes, from existing massive multi-sensor EO image repositories [46], to better understand the systemic and interrelated nature of global LC/LCC dynamics.

(B) Integrate near real-time internet-based satellite mapping services on demand with virtual Earth geo-browsers, such as the popular Google Earth, see Fig. 12-6.

(C) Augment the scientific and commercial impact of European space infrastructures, including the Sentinel-2 MSI, Sentinel-3 OLCI and SLSTR imaging sensors, the Meteosat satellite 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> generation sensor series, etc., because the adopted expert system for MS image pre-classification and segmentation, the Satellite Image Automatic Mapper (SIAM) [9], [12]-[14], is capable of mapping any radiometrically calibrated MS image acquired by past, existing or future MS imaging sensors [18], e.g., Envisat AATSR, Meris, SPOT-1/2/3/4/5, SPOT-6/7, Pleiades-1A/B, IRS-1C/D, IRS-P6, AVHRR, Modis, Landsat-4/5/7/8, World View-2/3, Ikonos, QuickBird, GeoEye-1, RapidEye, Skybox, Planetlabs, Alos-1/2, etc. Noteworthy, the SIAM pre-classification maps available to date, see Fig. 12-3 and Fig. 12-7, are more informative than the future Sentinel-2A Level 2 products, consisting of a cloud mask and a land/water mask, to be generated on a non-systematic basis, and of the Landsat-8 quality bands, consisting of cloud masks, already available on a regular basis.

(D) Integrate the visual analysis of uncalibrated RGB images acquired by consumer-level terrestrial and aerial color cameras, such as those mounted onboard light-weight UAVs, onto the same EO-IUS pipeline adopted for the interpretation of MS images radiometrically calibrated into TOARF or SURF values in compliance with the QA4EO guidelines, see Fig. 12-8.

(E) The automatic extraction of content maps from EO imagery allows each EO image stored in a massive EO image database to be provided with one or more content maps. The solution of the problem of automatic image-derived content



map extraction guarantees the solution of its dual problem, specifically, content-based image storage/retrieval [46], which is an open problem to date. The latter would become a seamless navigation through content maps where image-objects (segments, patches, polygons) provided with semantic labels can be tracked through time [51].

### 12.2.2 EO-VAS architecture and implementation

To comply with the EO-IUS requirements specified in Chapter 12.2.1 and overcome operational limitations of the existing cloud/cloud-shadow detectors listed in Table 12-3, to be further discussed in Chapter 12.2.4, an original implementation of the four-stage EO-IUS architecture, shown in Fig. 12-4(b), is proposed in agreement with the software pipeline sketched in Fig. 12-9.

#### I. Spaceborne/airborne MS image pre-processing, identified as Stage 0 (zero) in Fig. 12-4(b).

In compliance with the QA4EO guidelines [7], the proposed RS image pre-processing Stage 0 must include the radiometric calibration of DNs into TOARF, SURF or surface albedo values, where  $TOARF \supseteq SURF$  (i.e., SURF is a special case of TOARF in flat terrain and very clear sky conditions [28]). On theory, the RS community regards as common knowledge the prerequisite that for physically based, quantitative analysis of airborne and satellite sensor measurements in the optical domain their calibration to spectral radiance or reflectance values is mandatory [7]. Unfortunately, in the RS common practice, scientists, practitioners and institutions tend to overlook radiometric calibration as a necessary not sufficient pre-processing requirement capable of harmonizing large-scale multi-temporal multi-sensor EO datasets. Physical model-based and hybrid (combined physical and statistical) EO-IUSs do require as input sensory data provided with a physical unit of radiometric measure. On the contrary, statistical model-based EO-IUSs do not require as input numerical variables subject to radiometric calibration, e.g., refer to Table 12-3. Nonetheless, statistical systems too can benefit from the harmonization of input data accomplished via radiometric calibration. It is worth mentioning that the present RTD software project's background relies, free of cost, on an existing battery of computer programs for sensor-specific MS image radiometric calibration of DNs into TOARF values.

#### II. Physical model-based per-pixel pre-attentive vision first stage for MS image multi-granule pre-classification and multi-scale segmentation, identified as Stage 1 in Fig. 12-4(b).

Although rarely acknowledged by the RS community, prior knowledge-based pre-classification for MS data space discretization has a long history. For example, it is part of the atmospheric correction implemented in the ATCOR commercial software product [25]. It is also part of the NASA and the Canadian Centre for Remote Sensing (CCRS) automatic processing chain for MODIS data composites [1], [29]. Finally, Shackelford and Davis adopted a statistical model-based (maximum likelihood) pre-classification first stage to stratify (partition) a second-stage battery of LC class-specific feature extractors and classification modules [30], [31]. Equivalent to color naming in a natural language [32], prior knowledge-based color space discretization is the deductive counterpart of popular unsupervised (unlabeled) data learning algorithms for vector quantization (VQ) [21], like the popular k-means VQ algorithm, not to be confused with unsupervised data clustering algorithms [33].

To be input with a MS image whether or not radiometrically calibrated, in compliance with Chapter 12.2.1, the proposed EO-IUS for cloud/cloud-shadow detection, sketched in Fig. 12-4(b), adopts an original implementation of the pre-attentive vision Stage 1. It consists of two existing prior knowledge-based color data discretization algorithms in operating mode, the SIAM [8], [9], [12]-[14] (see Table 12-1) and the new RGB-Image Automatic Mapper (RGBIAM) [50] (see Table 12-2). When the input MS image is radiometrically calibrated, then both SIAM and RGBIAM can be run in parallel to combine color strata compatible with the presence of clouds and cloud shadows, according to a convergence-of-evidence approach. Otherwise, if the input image is uncalibrated then the sole RGBIAM can be employed. The two SIAM and RGBIAM expert systems are described below.

(i) The multi-source prior knowledge-based SIAM decision tree for MS data space discretization (color naming [32]) has been proposed, tested and validated by the RS community in recent years [8], [9], [12]-[14]. It is capable of generating, alternately automatically and in near real-time, multi-level pre-classification maps and multi-scale segmentation maps (computed from pre-classification maps via a well-posed two-pass connected-component multi-level image labeling algorithm [52]) of a spaceborne/airborne MS image radiometrically calibrated into TOARF or SURF values. At the level of abstraction of knowledge/information representation, the legend of a SIAM's pre-classification map generated from a single-date MS imagery consists of a discrete and finite vocabulary of *color names* [32], also called *spectral-based semi-concepts* or *spectral categories*, such as green-as-“*vegetation*”, brown-as-“*bare soil or built-up*”, blue-as-“*water or shadow*”, etc. [8], [9], [12]-[14], see Table 12-1, Fig. 12-3 and Fig. 12-7. Each spectral-based semi-concept can match none, one or more target LC classes whose spectral properties overlap, irrespective of the other dominant spatio-temporal properties of these LC classes (e.g., “*deciduous forest*”) in the 4-D real world-through-time. In other words, discrete color types, just like endmember fractions, “cannot always be inverted to unique LC class names” [27] (p. 147).





At the level of understanding of system design, the SIAM software product is implemented as an integrated system of four subsystems, including one "master" 7-band Landsat-like (visible blue (B), visible green (G), visible red (R), near infra-red (NIR), medium infra-red 1 (MIR1), medium infra-red 2 (MIR2) and thermal infra-red (TIR)) subsystem plus three "slave" (downscale) subsystems, namely, a 4-band SPOT-like (G, R, NIR and MIR1), a 4-band AVHRR-like (R, NIR, MIR1 and TIR) and a 4-band VHR-like (B, G, R and NIR), whose spectral resolutions overlap with Landsat's, but are inferior to Landsat's. The expression "Landsat-like MS image" adopted in this paper means: "an MS image whose spectral resolution mimics the spectral domain of the 7 bands of the Landsat family of imaging sensors", i.e., a spectral resolution where bands B, G, R, NIR, MIR1, MIR2 and TIR overlap (which does not mean coincide) with Landsat's.

(ii) A recent extension of the SIAM software product, called RGBIAM [50], was developed for preliminary classification of an RGB (true color or false color) data space not subject to radiometric calibration, but submitted to an original inductive (data-driven) histogram stretching capable of accomplishing the so-called *color constancy effect* in human vision, i.e., color normalization under a "canonical" illumination source [34], refer to Table 12-2. Typically, an image band-specific histogram features up to three main modes: high (scene foreground, if any), central and low (scene background, if any). If the background and foreground histogram modes exist and are detected, then they are compressed at the opposite ends of the color domain, e.g., equal to 0 and 255 in byte data-coding, whereas the central mode is stretched linearly between these two end values. Hence, the per-band image contrast is enhanced and the adaptive color constancy algorithm accomplishes inter-image harmonization, equivalent to radiometric calibration. Finally, an RGBIAM expert system partitions an RGB data cube, subject to color constancy, onto a mutually exclusive and totally exhaustive finite and discrete set of fixed (non-adaptive to input data) color names (quantization levels), corresponding to 3-D polyhedral. An example of an RGBIAM color mapping applied to a non-calibrated Landsat-8 image is shown in Fig. 12-8.

### III. Feedback loops, from the pre-attentional Stage 1 and the attentional Stage 3 into the pre-processing Stage 0, refer to Fig. 12-4(b).

This is where a relevant degree of novelty of the innovative four-stage EO-IUS architecture is located, see Fig. 12-4(b). A first feedback loop feeds the pre-attentive vision Stage 1's categorical output back to the pre-processing Stage 0's input for stratified (driven-by-knowledge, symbolic mask-conditioned) automatic RS image enhancement. Existing examples are stratified atmospheric correction [25], stratified topographic correction [35] (see Fig. 12-5), stratified image-co-registration and stratified image mosaic enhancement [14]. A second feedback loop feeds the attentive vision Stage 3's categorical output back to the pre-processing Stage 0's input, e.g., for cloud/cloud-shadow masking, refer to Fig. 12-4(b). *The principle of stratification, well-known in statistics [36], states that any inherently ill-posed statistical system (e.g., adopted for third-stage cloud/cloud-shadow detection), will always achieve greater precision by incorporating the "stratified" or "layered" approach, provided that the input strata are as uniform as possible in respect of the characteristic of interest.* In general, input information strata are difficult or expensive to collect. Fortunately, discrete color strata (e.g., green-as-"vegetation") detected by the Stage 1's pre-classifiers in an input MS image are stable (data independent), informative and generated at no cost in terms of user's interactions and nearly no cost in computation time. It is worth noting that the statistical principle of stratification is equivalent to the popular divide-and-conquer (*dividi-et-impera*) problem solving approach.

### IV. Stratified (driven-by-knowledge, better-posed) image segmentation, identified as Stage 2 in Fig. 12-4(b).

It is well known, but often forgotten in the RS common practice, that traditional (driven-without-knowledge) image segmentation is an inherently ill-posed cognitive problem [19]. On the contrary, the segmentation (partitioning) of a multi-level image is a well posed (deterministic) and automatic task to be accomplished in linear-time by a two-pass connected-component multi-level image labelling algorithm [52]. The Stage 1's output segments are image-objects (polygons) featuring low within-segment variance because their connected pixels feature the same color name, i.e., belong to the same color quantization level. Traditionally called texture elements (*texels*) or *textons* [52], these color-uniform image-objects are equivalent to the so-called *tokens* in the Marr nomenclature of the raw primal sketch [15]. In recent years, these texels have been renamed *superpixels* [53]. In series with the pre-classification Stage 1, a stratified better-posed image segmentation Stage 2 is expected to pursue so-called perceptual grouping (full primal sketch, texture segmentation) [15], i.e., it is expected to investigate the spatial organization of texels (superpixels) [37], [38].

### V. Stratified attentive vision third stage, identified as Stage 3 in Fig. 12-4(b).

This is the application-specific attentive vision unit yet to be developed, where a great deal of the RTD work in Person-Month (PM) must be allocated in the project work plan (refer to further Chapter 12.2.8). An attentive vision third-stage battery of stratified (driven-by-knowledge), hierarchical, spatial context-sensitive sensor- and application-specific hybrid inference-based (combined physical and statistical) models for LC/LCC class-specific feature extraction and classification is identified as Stage 3, refer to Fig. 12-4(b). Like in [31], each sensor- and class-conditional component of this second-stage battery of stratified hybrid inference-based algorithms is expected to become better/well posed in the Hadamard sense (i.e., one solution exists and is unique) [20], which also means automatic, i.e., no system's free-parameter is expected to



be user-defined. It is well known that 4-D real-world objects-through-time (e.g., cars, trees, etc.) are dominated by their 4-D spatiotemporal information [27]. Hence, at the attentive vision second stage, effective exploitation of spatial information through spatial reasoning becomes mandatory. Spatial information encompasses: (a) texture (perceptual grouping) [15], (b) inter-object spatial relationships, either topological (e.g., adjacency, inclusion, etc.) or non-topological (directional, i.e., space distance or angle difference), together with (c) image-object's geometric (shape and size) features. With special emphasis on cloud detection, it is important to remember that the ATCOR-4 commercial software product partitions a scene into: (I) *clear view* (clear-sky [1], [28]), (II) hazy and (III) *cloud regions* [25]. In the words of Huang *et al.* [3], because in the troposphere (0.5 - 9 km in height) and lower stratosphere (9 - 16 km), where most clouds occur, air temperature decreases in general as altitude increases, most clouds are colder than the land or water surfaces underneath them. Such temperature differences are especially significant for mid- (6 - 9 km) to high-altitude (9 - 12 km) clouds and can be very effective in identifying such clouds. For example, because high-altitude thin cirrus clouds (see Fig. 12-1) are not very bright spectrally, they are often very difficult to detect using non-thermal bands, but they are generally very cold, hence they can be identified relatively easily using the thermal band [3]. In addition, when the thermal band is used, like in [3], the rule-based cloud detection algorithm become much simpler than those used in cloud algorithms where no thermal band is used, e.g., [1]. In accordance with the cloud/cirrus/haze sorted sequence of mapping activities adopted in the ATCOR's decision-tree classifier (capable of generating the pixel-based output pre-classification file “\_out\_hcw.bsq”, where acronym hcw means haze, cloud, water) [25], the proposed cloud/cloud-shadow decision-tree classifier can be structured according to the following sorted list of actions (i) to (vii).

(i) Provide the EO-IUS input variables (required) and data flags (optional), expected to be the following. (a) Spatial resolution of the input image. (b) Sun-sensor position parameters: (I) the sun zenith angle, (II) the viewing zenith angle, (III) the relative azimuth angle, see Fig. 12-2. (c) Pre-attentive vision Stage 1: select the SIAM subversion - 7-band Landsat-like (B, G, R, NIR, MIR1, MIR2 and TIR), 4-band SPOT-like (G, R, NIR and MIR1), 4-band AVHRR-like (R, NIR, MIR1 and TIR) or 4-band VHR-like (B, G, R and NIR). (d) Attentive vision Stage 3: (I) Thermal band availability - Yes/No, (II) Cirrus band availability at 1.38  $\mu\text{m}$ : Yes/No [41], e.g., band 9 in the Landsat-8 Operational Land Imager (OLI), band S4 in the future Sentinel-3 SLSTR, band M9 in NASA-NOAA Preparatory Project (NPP) - Visible Infrared Imaging Radiometer Suite (VIIRS), etc.

(ii) **Candidate core cloud detection** (see Fig. 12-1), namely: (a) Low level clouds (0.5 - 6 km): cumulus (*Cu*), cumulonimbus (*Cb*), stratus (*St*), stratocumulus (*Sc*), nimbus (*Ni*); (b) Mid level clouds (6 - 9 km): altocumulus (*Ac*), altostratus (*As*), nimbostratus (*Ns*); (c) High level clouds (9 - 12 km): cirrus (*Ci*), Cirrocumulus (*Cc*), cirrostratus (*Cs*). Planned activities: (I) Selection of the SIAM's spectral categories, e.g., “*snow water-ice*”, “*core cloud*”, “*thick cloud*”, as core cloud candidate image-objects, see Table 12-1. (II) Selection of the RGBIAM's spectral category “*white*” to detect core cloud candidate image-objects, see Table 12-2. (III) Select a fusion strategy to combine the SIAM's evidence with the RGBIAM's evidence to detect core cloud candidate pixels, see Fig. 12-7 to Fig. 12-9. (IV) Merge cloud-candidate image-objects based on spatial topological relationships: e.g., inclusion, adjacency [39], see Fig. 12-7 to Fig. 12-9. (V) Remove false cloud-candidate image-objects based on geometric properties, i.e., shape and size [40], see Fig. 12-7 to Fig. 12-9.

(iii) **Cloud annulus detection**. For example, in [3], a specific rule-based strategy is implemented in the 2-D Red - normalized TIR space to detect clouds that typically become less bright and less cold towards their edges. Planned activities: (I) Selection of the SIAM's spectral categories, e.g., “*thin cloud over water*”, “*thin cloud over vegetation*”, as cloud annulus candidate image-objects, refer to Table 12-1. (II) Merge cloud-candidate image-objects, detected in step (ii), with cloud annulus image-objects based on spatial topological relationships: e.g., inclusion, adjacency, see Fig. 12-7 to Fig. 12-9. (III) Remove false cloud-candidate image-objects based on geometric properties, i.e., shape and/or size, see Fig. 12-7 to Fig. 12-9.

(iv) **Thin cirrus detection** (see Fig. 12-1). Thin cirrus clouds are located in the upper troposphere (6 - 9 km) or lower stratosphere (9 - 16 km) [25], hence they are typically very cold due to rapid decreases in temperature as altitude increases in the atmosphere [3]. In non-thermal wavelengths, cirrus clouds are difficult to detect, especially over land, because they are partially transparent. In [3], thin cirrus clouds over forest are identified relatively easily using the thermal band of Landsat imagery. Thin cirrus and other high altitude thin clouds over forest, which are not necessarily much brighter than the forest beneath them, but are typically very cold, are detected with a specific rule-based strategy in the 2-D Red - normalized TIR space, whereas clouds that are bright but not cold, including some near surface clouds, low fogs, and smokes, are not likely to be detectable using this cloud decision boundary. Some sensors are provided with a so-called cirrus band at 1.38  $\mu\text{m}$ , e.g., Landsat-8 OLI band 9, in agreement with [41]. Water vapor dominates in the lower troposphere and usually 90% or more of the atmospheric water vapor column is located in the 0 - 5 km altitude layer. Therefore, if a narrow spectral band is selected in a spectral region of very strong water vapor absorption, e.g., around 1.38  $\mu\text{m}$  or 1.88  $\mu\text{m}$ , the ground reflected signal will be totally absorbed (although surface snow and water ice do appear visible in the cirrus band, based on experience, in addition to cirrus clouds), but the scattered cirrus signal will be collected by a high-altitude (> 20 km) airborne/spaceborne sensor [25]. For example, in [25], if a narrow cirrus channel around 1.38  $\mu\text{m}$  exists, then



two different cirrus removal strategies are adopted for water and land pixels by means of a hybrid inference system where the cirrus channel is investigated in combination with, respectively, a MIR band (water) and a Red (R) or a NIR band (land). Planned activities: (I) Investigate the SIAM's output product called Haze five-level mask. (II) If there is no thermal band or cirrus band, are there other convergence-of-evidence criteria? To be investigated.

(v) **Haze/fog detection** (see Fig. 12-1). In the visible bands (0.35 – 0.75  $\mu\text{m}$ ), images contaminated by haze appear similar to those contaminated by cirrus clouds. However, in longer wavelength channels, starting from the NIR band (around 0.85  $\mu\text{m}$ ), haze effects are rarely visible [25]. Haze is located in the lower troposphere (0 - 3 km) as opposed to high altitude cirrus clouds, located in the upper troposphere (6 - 9 km) or lower stratosphere (9 – 16 km). For haze detection, ATCOR features two hybrid (combined rule-based plus statistical) strategies specialized to detect: (a) haze over land and, based on the visible bands Blue and Red, while the NIR band is used to exclude water pixels, and (b) haze over water, based on evidence collected from the NIR band once water pixels are masked either using spectral criteria or taking an external water map [25]. Planned activities: (I) Investigate the SIAM's output product called Haze Five-level Mask. (II) Other convergence-of-evidence criteria based on color and spatial properties?

(vi) **Smoke plume detection**. Unlike clouds, smoke plumes are not bright, but dark. Planned activities: (I) Selection of the SIAM's spectral categories, e.g., "Thin Smoke Plume over Water", "Thick Smoke Plume over Water", "Smoke Plume over Vegetation", "Smoke Plume over Bare soil or Built-up" as smoke plume candidate image-objects, refer to Table 12-1. (II) Merge smoke plume-candidate image-objects with smoke plume annulus image-objects based on spatial topological relationships: e.g., inclusion, adjacency, see Fig. 12-9. (III) Remove false smoke plume-candidate image-objects based on geometric properties, namely, shape and/or size, see Fig. 12-9.

(vii) **Cloud shadow detection**, see Fig. 12-9. In [3], the cloud height is estimated based on a statistic model-based normalized temperature and a digital elevation model (DEM) is required as ancillary input. In [1], the cloud height is assumed between 0.5 and 12 km and no DEM is employed. Planned activities: (I) Apply a physical model-based cloud shadow projection algorithm, whose input parameters are [1]: (a) the spatial location of the cloud, (b) cloud top and bottom heights, (b) the sun zenith angle, (d) the viewing zenith angle, and (e) the relative azimuth angle, see Fig. 12-2. (II) Selection of the SIAM's spectral categories eligible as shadow candidate areas, e.g., "Water or shadow", "Vegetation in shadow", "Vegetation in water or shadow", etc., refer to Table 12-1. (III) Image-object matching between projected cloud shadow areas and shadow candidate image-objects can be inspired by those proposed in [3] or [26] and [49].

To recapitulate, the original implementation of an innovative four-stage EO-IUS, sketched in Fig. 12-4(b), suitable for cloud/cloud-shadow detection, in agreement with Fig. 12-9, consists of several software modules that already exist. In this EO-IUS instantiation, the aforementioned low-level vision Stage 0, Stage 1 (including the RGBIAM and SIAM low-level vision expert systems, identified as processing blocks 3 and 10 in Fig. 12-9) and at least part of Stage 2 are already available. Only the context-sensitive attentive vision Stage 2 (in part) and Stage 3 have to be implemented on an application-specific basis. The aforementioned Stage 2 (in part) and Stage 3, yet to be developed, can be identified as the two spatial modelling blocks 11 and 12 shown in Fig. 12-9. Noteworthy, the proposed Stage 3's cloud/cloud-shadow map legend, refer to the aforementioned points (ii) to (vii), is more informative than those proposed by alternative approaches, see Table 12-3.

### 12.2.3 System integration, quality assessment and comparison of alternative solutions

Since the proposed four-stage EO-IUS design, see Fig. 12-4(b), shall be implemented in agreement with Fig. 12-9, where only the two processing blocks involved with spatial reasoning, identified as spatial modellers 11 and 12, are yet to be developed, then the project's RTD software activity plan focuses on the application-specific high-level vision Stage 3. In recent years, the proposed four-stage EO-IUS architecture has been implemented, integrated and tested in a variety of application-specific domains [8], [9], [12]-[14], [18], [35], [40]-[43], [50], [51]. Hence, based on past experience, the integration of a new application-specific Stage 3 with pre-existing Stages 0 to 2 is expected to be straightforward. The core of the RTD software project will focus on the development of the new high-level vision Stage 3.

The test/validation phase of the novel high-level classification Stage 3 is designed as follows.

- (i). A set of independent metrological  $Q^2$ IOs, to be community-agreed upon, is selected from the existing literature [8], [9], refer to Chapter 12.2.1.
- (ii). The statistically valid and spatially consistent probability sampling protocol for accuracy assessment of a fine-resolution thematic map, proposed in [42], is adopted, in contrast with non-probability sampling strategies traditionally employed in the RS common practice. In [42], both pixel-based thematic  $Q^2$ Is (TQ<sup>2</sup>Is) and polygon-based Spatial  $Q^2$ Is (SQ<sup>2</sup>Is), provided with a degree of uncertainty in measurement, are estimated in agreement with the GEO's QA4EO guidelines [7]. Proposed TQ<sup>2</sup>Is include the popular overall accuracy (OA), user's accuracies and producer's accuracies [3], [4]. For example, in [3], reference pixels belonging to three target LC classes, namely, cloud, cloud-shadow and clear-sky surfaces, were selected via photointerpretation. Finally, the per-pixel OA value plus per-class omission and commission errors were assessed. The same pixel-based TQ<sup>2</sup>Is estimated in [3] were also adopted in [4], [26] and [49] (refer to Table 12-3), without any estimation of their degree of uncertainty in measurement as a function of the reference sample size. In the present project proposal, in addition to polygon-based SQ<sup>2</sup>Is, which are omitted in [3], [4], [26] and [49], pixel-based TQ<sup>2</sup>Is must be provided with a degree of uncertainty in measurement, to comply with the GEO's QA4EO guidelines [7].



For example, typical USGS classification project requirement specifications are:  $OA \in [0, 1] \pm \delta$  fixed equal to  $0.85 \pm 2\%$  [44], where  $\pm \delta$  is the degree of uncertainty in measurement. A typical USGS target per-class classification accuracy,  $OA_{c,c} \in [0, 1] \pm \delta c$ ,  $c = 1, \dots, C$ , where  $C$  is the total number of target LC classes, is fixed about equal to  $70\% \pm 5\%$  [7]. According to Lunetta and Elvidge [45], if the desired level of significance  $\alpha = 0.03$  and  $C = 3$ , say, cloud, cloud-shadow and clear-sky surfaces, then the level of confidence  $(1 - \alpha/C) = 0.99$  and  $\chi^2(1, 1 - \alpha/C) = 6.63$ . In this case, if  $OA_{c,c} = 85\%$ , and  $\delta c = \pm 2\%$  with  $c = 1, 2, 3$ , are the target accuracy values, then the required reference sample set size (SSS) per class  $c$  is  $SSS_c = 2113$  where  $c = 1, 2, 3$ . If  $OA_{c,c} = 85\%$ , with  $\delta c = \pm 5\%$ , then  $SSS_c = 338$  with  $c = 1, 2, 3$ , and so on, refer to [42]. (III) The required reference sample must consist of multi-source images, both calibrated and non-calibrated, provided with cloud/cloud-shadow/clear-sky “truth” masks [3].

In this testing/validation scenario, two reference data sets are available free-of-cost.

(A) The Landsat-7 sensor-specific Cloud Cover Assessment Validation Data (L7CCVD) set, available for download [47], has been adopted as the reference sample in related works, like [4], [26] and [49]. Consisting of 180 Landsat-7 images, provided with radiometric calibration parameters, it covers the full range of global environments and cloud conditions. Manually selected cloud masks per reference scene and cloud-shadow masks for few reference scenes are available.

(B) Landsat-8 standard Level 1 products cirrus and non-cirrus cloud masks, encoded at the pixel level as high/medium/low confidence, can be downloaded [48]. Unfortunately, beyond these two L7CCVD and Landsat-8 reference datasets, no additional multi-source reference sample set is available yet.

To summarize, according to the selected probability sampling protocol proposed in [42]: (i) target  $OA \pm \delta$  and  $OA_{c,c} \pm \delta c$ ,  $c = 1 \dots, C = 3 = \{\text{cloud, cloud-shadow, clear-sky}\}$ , must be specified in advance, to compute the required  $SSS_c$ ,  $c = 1, 2, 3$ , in agreement with [45]. (ii) Each reference sample unit features a spatial type, equal to either pixel or polygon, depending on the target LC class  $c = 1, 2, 3$ . (iii) Once randomly sampled and scrutinized by the domain experts and/or potential users, reference sample polygons, if any, must be manually edited by photointerpreters, like in [3] and [4]. To conclude, in addition to the available free-of-cost L7CCVD set and Landsat-8 imagery provided with cloud and cirrus masks, a probability sampling of multi-source reference images, featuring manually edited cloud/cloud-shadow/clear-sky “truth” objects, shall be conducted independently by potential users in the product validation phase.

Alternative to the hybrid EO-IUS proposed in this RTD project, the current state-of-the-art in cloud/cloud-shadow detection, called Fmask [26], [49] (refer to Table 12-3), is a purely deductive program executable, suitable for Landsat images exclusively. Fmask can be downloaded from the web page: <https://code.google.com/p/fmask/>. It will be adopted for direct comparison with the proposed methodology.

#### 12.2.4 Differences between the proposed solution and alternative existing solutions

In the machine learning literature it is well known that *inductive data learning problems*, like image segmentation (partitioning) and its dual problem, image-contour detection [19], *are inherently ill-posed in the Hadamard sense, i.e., their solution does not exist or is not unique* [20]. Therefore, they are very difficult to solve. *To become better posed (better conditioned) for numerical treatment, inductive data learning algorithms "require a priori knowledge in addition to data"* [21] (p. 39). By definition, prior knowledge is available *in addition to* sensory data, i.e., *a priori* knowledge is data independent, although it is typically application specific. It means that *a priori* knowledge is eligible for providing initial conditions to an inherently ill-posed adaptive learning-from-examples (statistical, inductive, bottom-up, driven-by-data, driven-without-knowledge) algorithm, such that the latter (equivalent to phenotype) is conditioned to explore a neighborhood of the former (equivalent to genotype) in a solution space.

In contrast with this common knowledge, a great majority of the existing EO-IUSs, including popular geographic object-based image analysis (GEOBIA) systems, employ a driven-without-knowledge inductive learning-from-data image segmentation first stage, which always starts its data analysis from scratch, see Fig. 12-4(a). As a consequence, the GEOBIA first stage is affected by structural drawbacks. First, image segmentation is inherently ill-posed [19], i.e., it is semi-automatic and site-specific [22]. Second, it is sub-symbolic; as such, it falls short in addressing one key principle of vision, formulated by David Marr as follows: "vision goes symbolic almost immediately, right at the level of zero-crossing (raw primal sketch in the pre-attentive vision first stage)... without loss of information" [15] (p. 343).

Due to their lack of operativeness, existing EO-IUSs are outpaced by the ever-increasing rate of collection of EO images of enhanced quality and quantity, hereafter identified as EO “big data”. For example, to date, the European Space Agency (ESA) estimates as 10% or less the percentage of EO images ever downloaded by stakeholders from its EO databases.

Supported by increasing portions of the RS literature [23], [24], the thesis that, in the RS common practice, Q2IOs of existing EO-IUSs, including GEOBIA systems, score low, can be considered part of an ongoing multi-disciplinary debate, encompassing scientific disciplines like computer vision, artificial intelligence (focused on deductive inference) and machine learning (centered on inductive inference), believed to be inadequate to provide operational solutions to their





ambitious cognitive goals. *This controversy may mean that, if they are not combined, inductive and deductive inference systems show intrinsic weaknesses in operational use, irrespective of their implementation.* Whereas inductive inference systems are semi-automatic and site-specific [8], [9], [20], [21], it is well known that expert systems lack flexibility and scalability to complex problems [8], [9]. To take advantage of the unique features of each and overcome their shortcomings, statistical and physical models are increasingly combined into *hybrid inference systems* [8], [9], see Fig. 12-4(b). For example, several existing EO-IUS implementations for cloud/cloud-shadow detection adopt a hybrid inference approach, where a prior knowledge-based decision tree is included, see Table 12-3 [1]-[5], [25], [26], [49]. Nevertheless, *the existing cloud/cloud-shadow detectors listed in Table 12-3 do not satisfy the EO-IUS requirements specified in Chapter 12.2.1.* First, none of these algorithms is capable of mapping an input MS image whether or not it is radiometrically calibrated. Second, all of these algorithms, but one, the popular Atmospheric / Topographic Correction for airborne/spaceborne image (ATCOR) commercial software product [25], are imaging sensor-specific, i.e., they are not scalable/transferable to different sensors. Third, with the sole exception of the pixel- and object-based methods proposed in [5], [26] and [49], the remaining cloud/cloud-shadow detectors are pixel-based, i.e., they are exclusively based on imaging spectrometry; in particular, algorithms proposed in [1], [2] and [25] employ no spatial (contextual) information whatsoever. This is in contrast with the conceptual foundation of the GEOBIA community, according to which spatial (contextual) information cannot be ignored in RS image understanding when the imaging sensor's spatial resolution is  $\leq 20$  m, because spatiotemporal information dominates spectral (color, context-insensitive) information in the real world, as correctly pointed out by Adams *et al.* [23], [24], [27]. This unquestionable fact is so true that, in human beings, panchromatic vision is almost as effective as chromatic vision. Last but not least, *each single cloud/cloud-shadow detector listed in Table 12-3 is affected by specific operational limitations at the levels of understanding of the EO-IUS design and/or implementation phase* [15]. For the sake of brevity, among the algorithms listed in Table 12-3, let us examine in more detail the so-called "automated" algorithm for cloud and cloud shadow detection in 30 m resolution Landsat images proposed in [3]. It consists of: (a) a physical model-based (prior knowledge-based decision-tree) detection (masking) of non-vegetation LC classes, namely, dark bare soil and water, (b) an inductive histogram analysis of a bi-modal vegetation pixel distribution collected from a moving image-window, to detect dark vegetation pixels as belonging to LC class forest, (c) a temperature normalization, where the forest temperature is subtracted from the pixel's temperature, (d) a physical rule-based detection of clouds in a 2-D red band-normalized temperature space, (e) a physical model-based estimation of the cloud height, based on temperature, (f) a physical model-based surface projection of a predicted cloud shadow from a detected cloud, (g) a contextual search of a cloud shadow in the neighborhood of a predicted cloud shadow, where dark pixels are detected in the near infra-red (NIR) or medium infra-red (MIR) bands. Operational limitations of this specific workflow are that: (A) it is not automatic, but depends on several heuristic parameters to be user-defined, (B) forest pixels are required to be detected with high confidence, to estimate a mean surface temperature expected to be higher than that of clouds, (C) there is spectral confusion between snow and cloud, and between cloud shadow and water, (D) a thermal channel is considered mandatory to accomplish the cloud shadow detection, i.e., this algorithm is thermal sensor-specific, and (E) the spatial type of information primitives is pixel and never polygon, i.e., this algorithm lacks spatial information and inter-object spatial relationships.

To summarize, in agreement with Table 12-3, main differences between the proposed EO-VAS in comparison with existing cloud/cloud-shadow detectors can be found: (1) at the design level of system understanding. The former employs prior knowledge-based color space partitioners, such as SIAM (for radiometrically calibrated MS images) and RGBIAM (for non-calibrated RGB images) to initialize inductive data analysis algorithms, which no longer require input parameters to be user-defined based on heuristics. (2) At the level of understanding of system implementation. Original automatic algorithms for inductive color constancy and for deductive color space quantization, such as SIAM and RGBIAM, are implemented in operating mode, which encompasses linear-time and tile-streaming implementation solutions, to be capable of processing massive images in near real-time. Consequences are that, among the competing systems revised in Table 12-3, the proposed EO-VAS is the only one capable of satisfying the RTD software project's requirements listed in Chapter 12.2.1. Last but not least, the proposed low-level vision Stage 0 and Stage 1 in Fig. 12-4(b) accomplish an automatic mapping of color images into a mutually exclusive and totally exhaustive dictionary of color names, which is user- and application-independent. Hence, this low-level CV subsystem can be employed in a large variety of EO image-derived value-adding products and services, such as those listed at the end of Chapter 12.2.1, e.g., content-based EO image storage/retrieval [46].



### 12.2.5 Target user communities

Accurate detection and masking of clouds and cloud shadows is a well-known low-level vision prerequisite for clear-sky RS image mosaicking/compositing [1]-[5], suitable for further retrieval of land surface variables, either quantitative or nominal [6]. For example, cloud contamination is a relevant problem in LCC analysis, because unflagged clouds may be mapped as false LCC occurrences. In practice, accurate automatic cloud/cloud-shadow detection is a necessary not sufficient condition to transform EO image big data into operational, timely and comprehensive information products and services, in compliance with the QA4EO guidelines. For example, to date a large majority of text-based EO image querying systems employs a per-image summary statistic of cloud coverage, which carries no geospatial information about the distribution of clouds. Only few spaceborne imaging sensors, such as Landsat-8 and Sentinel-2, provide a cloud quality mask. Unfortunately, the accuracy of the Landsat-8 cloud masks has been assessed to be low [48].

To recapitulate, any existing user of EO images demands for an operational cloud/cloud-shadow detector. Therefore, potential users of the proposed computer vision VAS encompass the whole academia involved with scientific applications of EO imagery, the EO image providers, ranging from space agencies (ESA, NASA, JAXA, ISRO, CNSA, DLR, CNES, etc.) to space industry, e.g., RapidEye and DigitalGlobe, which are required to augment the accessibility of their EO big data archives by increasing the quality and quantity of image quality bands (such as the Sentinel-2A and Landsat-8 Level 2 products), the EO service industry, such as Google (Earth Engine), the EO image processing commercial software toolbox developers, such as Harris (ENVI/EDL) and Trimble (eCognition), the EO service providers, encompassing both private companies or public sector institutions, such as UN-FAO, and EO image end-users, i.e., private companies or public sector organisations where EO image-derived geo-information is integrated into their operational business practices on a regular basis.

### 12.2.6 Expected benefits of the proposed EO-VAS solution

As reported in Chapter 12.2.3, none of the existing cloud/cloud-shadow detectors listed in Table 12-3 satisfies the EO-IUS requirements specified in Chapter 12.2.1. On the contrary, the proposed multi-source EO-VAS implementation for cloud/cloud-shadow detection, sketched in Fig. 12-9, is expected to satisfy the Q<sup>2</sup>IOs required in Chapter 12.2.1. In addition, the proposed computer vision VAS instantiation belongs to a novel hybrid feedback EO-IUS architecture, shown in Fig. 12-4(b), suitable for a variety of low- (pre-attentional) and high-level (attentional) vision applications, listed in Chapter 12.2.1, including content-based image storage/retrieval in big EO image databases.

### 12.2.7 Future opportunities of the proposed EO-VAS solution

As reported in Chapter 12.2.1, if successful, the proposed EO-VAS solution would provide the first proof-of-concept that a novel hybrid feedback EO-IUS design and implementation strategies are capable of transforming multi-source EO image “big data” into operational, comprehensive and timely information products, in compliance with the QA4EO recommendations and with several ongoing RS international programs [16], [17]. This proof-of-concept would open a huge variety of future research, educational and market opportunities, refer to points (A) to (E) listed in Chapter 12.2.1.

### 12.2.8 Proposed approach to the work and first iteration of tasks

An iterative project development style is adopted, where the 1-year project is broken down into four 3-months iterations, corresponding to four project milestones (MLs). In each time box a quarter of the project requirements would be addressed by completing the software life cycle for that quarter: analysis, design, code and test. At the end of each iteration predating the last iteration, the system is not expected to be put into production, but should be of production quality, to get value from the system earlier and to get better-quality feedback. In practice, the iterative project development style employing time boxing removes the critical path traditionally related to a waterfall project development style. The critical path is defined as the total time for activities on this path that is greater than that in any other path through the activity network, such that a delay in any task on the critical path leads to a delay in the project.

The project schedule (Gantt chart) is shown in Fig. 12-10. Table 12-4 provides an overview of the scheduled work packages (WPs), including deliverables (Ds) and a first breakdown of manpower per WP. Intuitively, a first breakdown of costs per task can assume that costs per WP are linearly related to the person-month (PM) estimates per WP. WP1 is partitioned into Task 1 to Task 3, in agreement with the Statement of Work. WP2 (Stage 0 – Data pre-processing) to WP5 (Stage 3 – Attentive vision) correspond to Stage 0 to Stage 3 of the four-stage EO-IUS architecture sketched in Fig. 12-4(b). Their detailed descriptions can be found in Chapter 12.2.2. A detailed description of WP6 (System integration, testing and validation) can be found in Chapter 12.2.3. An overview of the four project MLs is provided in Table 12-5.

## 12.3 Potential problem areas

### 12.3.1 Identification of the main problem areas likely to be encountered in performing the activity

The proposed study is realistic because instances of Stage 0 to Stage 3 of the planned four-stage EO-IUS design, see Fig. 12-4(b), were implemented, integrated and validated in recent years [8], [9], [12]-[14], [18], [35], [42], [43], [50], [51]. In



Fig. 12-9 only the two processing blocks identified as modules 11 and 12, equivalent to two spatial modellers belonging to the Stage 3 shown in Fig. 12-4(b), are yet to be developed. Hence, based on past experience, the integration of a new application-specific Stage 3 with pre-existing Stages 0 to 2 is expected to be straightforward. The core of the RTD software project efforts will focus on the implementation of the new Stage 3's processing blocks 11 and 12 shown in Fig. 12-9. Pre-existing system units, such as SIAM and RGBIAM, belonging to Stage 1 in Fig. 12-4(b), are expected to require only standard maintenance. Overall, the proposed divide-and-conquer problem solving approach sketched in Fig. 12-9, to be pursued by an iterative project development style (refer to Chapter 12.2.8), is capable of diluting the project technical risk. In practice, the technical risk of a major project breakdown is expected to be low or null.

### 12.3.2 Proposed solutions to the problems identified

An iterative project development style is adopted, where the 1-year project is broken down into four 3-months iterations. In each time box a quarter of the project requirements would be addressed by completing the software life cycle for that quarter: analysis, design, code and test. Fast prototyping of the two novel spatial modellers, identified as processing blocks 11 and 12 in Fig. 12-9, can employ the eCognition commercial software toolbox available at the tenderer's facility.

#### 13.3.3 Proposed trade-off analyses and identification of possible limitations or non-compliances

At the present stage of proposal, no EO-VAS methodological limitation, technical limitation or non-compliance can be reasonably foreseen.

## 12.4 Technical implementation / Programme of work

### 12.4.1 Proposed work logic

The work plan consists of WPs, including deliverables (Ds), summarized in Table 12-4. WP1 is partitioned into Task 1 (Service Verification and Service Trial Definition), Task 2 (Conduct EO service Trial) and Task 3 (Generate Action Plan & Promotional materials), in agreement with the Statement of Work. The work plan, scheduled according to the Gantt chart sketched in Fig. 12-10, is partitioned into four quarters of the software life cycle, encompassing analysis, design, code and test. In this iterative project development style, alternative to and more flexible than a traditional waterfall project development style, there is no project's single flowchart defined beforehand where a traditional critical path can be identified, refer to Chapter 12.2.8.

### 12.4.2 Detailed procurement plan for the EO data

According to Chapter 12.2.3, two reference cloud/cloud-shadow image sets are available free-of-cost. First, the L7CCVD set [47], consisting of 180 Landsat-7 images provided with radiometric calibration parameters, covers the full range of global environments and cloud conditions. Manually selected cloud masks per reference scene and cloud-shadow masks for few reference scenes are available. Second, the Landsat-8 standard Level 1 products cirrus and non-cirrus cloud masks, encoded at the pixel level as high/medium/low confidence, can be downloaded [48]. In addition, radiometrically calibrated Sentinel-2A images are already available for download free-of-cost, see Fig. 12-7. Additional test images acquired by different spaceborne sensors, e.g., RapidEye, are expected to be provided free-of-cost by potential users or are already available in the archives of this tenderer. To conclude, neither operational agreements for access to EO data nor costs for EO image purchase are envisaged.

## References in Chapter 12

- [1] Y. Luo, A. P. Trishchenko and K. V. Khlopenkov, "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sensing of Environment*, vol. 112, pp. 4167–4185, 2008.
- [2] K.V. Khlopenkov and A.P. Trishchenko, "SPARC: New cloud, snow, and cloud shadow detection scheme for historical 1-km AVHRR data over Canada," *Journal of Atmospheric and Oceanic Technology*, vol. 24, pp. 322–343, 2007.
- [3] C. Huang, N. Thomas, S. N. Goward, J. G. Masek, Z. Zhu, J. R. G. Townshend, and J. E. Vogelmann, "Automated masking of cloud and cloud shadow for forest change analysis using Landsat images," *International Journal of Remote Sensing*, vol. 31, no. 20, pp. 5449-5464, 2010.
- [4] R.R. Irish, J.L. Barker, S.N. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 10, pp. 1179–1188, October 2006.
- [5] S. Le Hégarat-Masclé and C. André, "Reduced false alarm automatic detection of clouds and shadows on SPOT images using simultaneous estimation," *Proc. SPIE*, 6748-46, vol. 1, p. 1-12, 2010.
- [6] R. Capurro and B. Hjørland, "The concept of information," *Annual Review of Information Science and Technology*, vol. 37, pp. 343-411, 2003.



- [7] GEO/CEOSS, A Quality Assurance Framework for Earth Observation (QA4EO), version 4.0 2010, Available online: [http://qa4eo.org/docs/QA4EO\\_Principles\\_v4.0.pdf](http://qa4eo.org/docs/QA4EO_Principles_v4.0.pdf) (accessed on 15 November 2012).
- [8] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 1: Introduction," *Remote Sens.*, vol. 4, no. 9, pp. 2694-2735, 2012. doi:10.3390/rs4092694.
- [9] A. Baraldi and L. Boschetti, "Operational automatic remote sensing image understanding systems: Beyond Geographic Object-Based and Object-Oriented Image Analysis (GEOBIA/GEOOIA) - Part 2: Novel system architecture, information/knowledge representation, algorithm design and implementation," *Remote Sens.*, vol. 4, no. 9, pp. 2768-2817, 2012. doi:10.3390/rs4092768.
- [10] A. S. Laliberte and A. Rango, "Image processing and classification procedures for analysis of sub-decimeter imagery acquired with an unmanned aircraft over arid rangelands," *GIScience & Remote Sensing*, vol. 48, no. 1, pp. 4–23, 2011.
- [11] G. Schaepman-Strub, M.E. Schaepman, T.H. Painter, S. Dangel, and J.V. Martonchik, "Reflectance quantities in optical remote sensing—definitions and case studies," *Remote Sensing of Environment*, vol. 103, pp. 27–42, 2006.
- [12] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 9, pp. 2563-2586, Sept. 2006.
- [13] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1299 - 1325, March 2010.
- [14] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1326 - 1354, March 2010.
- [15] D. Marr, Vision, W.H. Freeman and Company: San Francisco, CA, U.S.A. 1982.
- [16] G. Gutman, A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss, D. Skole, B. L. Turner, M. A. Cochrane, Eds. Land Change Science. Kluwer: Dordrecht, The Netherlands, 2004.
- [17] GEO. The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan 2005. Available online: <http://www.earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf> (accessed on 10 Jan. 2012).
- [18] A. Baraldi, L. Boschetti, D. Roy, and C. Justice, Automatic near real-time preliminary classification of Sentinel-2 (and Sentinel-3) images with the Satellite Image Automatic Mapper™ (SIAM™), Sentinel-2 Preparatory Symposium, ESA-ESRIN, Frascati, Italy, April 23-27, 2012.
- [19] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, pp. 869–889, 1988.
- [20] J. Hadamard, "Sur les problemes aux derivees partielles et leur signification physique," *Princet. Univ. Bull.*, vol. 13, pp. 49–52, 1902.
- [21] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*; Wiley: New York, NY, USA, 1998.
- [22] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: John Wiley and Sons, 2004.
- [23] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Feitosa, F. Van Der Meer, H. Van Der Weerf, F. Van Coillie, and D. Tiede, "Geographic Object-based Image Analysis: a new paradigm in Remote Sensing and Geographic Information Science," *ISPRS International Journal of Photogrammetry and Remote Sensing*, vol. 87, no. 1, pp. 180-191, 2014.
- [24] G. J. Hay and G. Castilla, "Geographic Object-Based Image Analysis (GEOBIA): A New Name for a New Discipline", In *Object-Based Image Analysis: Spatial Concepts for Knowledge-driven Remote Sensing Applications*; Blaschke, T., Lang, S., Hay, G.J., Eds.; Springer-Verlag: New York, NY, USA, 2008; Chapter 1.4, pp. 81–92.
- [25] R. Richter and D. Schlapfer, *Atmospheric / Topographic Correction for Airborne Imagery (ATCOR)-4 User Guide, Version 6.3.0*, Dec. 2013.
- [26] Zhe Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery", *Remote Sensing of Environment*, vol. 118, pp. 83–94, 2012.





- [27] J. B. Adams, E. S. Donald, V. Kapos, R. Almeida Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie, "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon," *Remote Sens. Environ.*, vol. 52, pp. 137-154, 1995.
- [28] P. Chavez, "An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data", *REMOTE SENSING OF ENVIRONMENT*, vol. 24, pp. 459-479, 1988.
- [29] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS", *J. Geophys. Res.*, vol. 103, D24, no. 32, pp. 141-157, 1998.
- [30] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 1920–1932, 2003.
- [31] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 2354–2363, 2003.
- [32] R. Benavente, M. Vanrell, and R. Baldrich, "Parametric fuzzy sets for automatic color naming," *Journal of the Optical Society of America A*, vol. 25, pp. 2582-2593, 2008.
- [33] B. Fritzke, Some competitive learning methods, 1997. [Online]. Available: Draft document, <http://www.demogng.de/>. Accessed on: 28 Oct. 2014.
- [34] A. Gijsenij, T. Gevers, J. van de Weijer, "Computational Color Constancy: Survey and Experiments," *IEEE Trans. Image Processing*, vol. X, no. X, 2010.
- [35] A. Baraldi, M. Gironda, and D. Simonetti, "Operational two-stage stratified topographic correction of spaceborne multi-spectral imagery employing an automatic spectral rule-based decision-tree preliminary classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 112-146, Jan. 2010.
- [36] N. Hunt and S. Tyrrell, Stratified Sampling. Available online: <http://nestor.coventry.ac.uk/~nhunt/meths/strati.html> (accessed on 12 June 2014).
- [37] G. M. Espindola, G. Camara, I. A. Reis, L. S. Bins, and A. M. Monteiro, "Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation", *International Journal of Remote Sensing*, vol. 27, no. 14, pp. 3035–3040, 2006.
- [38] A. Baraldi and F. Parmiggiani, "Single linkage region growing algorithms based on the vector degree of match", *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 137-148, Jan. 1996.
- [39] G. Christodoulou, E.G.M. Petrakis, and S Batsakis, "Qualitative Spatial Reasoning using Topological and Directional Information in OWL", In: 24th International Conference on Tools with Artificial Intelligence (ICTAI 2012). pp. 1–7. Athens (November 2012).
- [40] V. B. Soares, A. Baraldi, and D. W. Jacobs, "Multi-objective software suite of two-dimensional shape descriptors for object-based image analysis," submitted for consideration for publication, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2015.
- [41] B. C. Gao, A. F. H. Goetz, and W. J. Wiscombe, "Cirrus cloud detection from airborne imaging spectrometer data using the 1.38  $\mu\text{m}$  water vapor band," *Geophysical Research Letters*, vol. 20, pp. 301–304, 1993.
- [42] A. Baraldi, L. Boschetti, and M. Humber, "Probability sampling protocol for thematic and spatial quality assessments of classification maps generated from spaceborne/airborne very high resolution images," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 1, Part: 2, pp. 701-760, Jan. 2014.
- [43] D. Tiede, F. Lühje, and A. Baraldi. Automatic post-classification land cover change detection in Landsat images: Analysis of changes in agricultural areas during the Syrian crisis. In E. Seyfert, E. Gülch, C. Heipke, J. Schiewe, & M. Sester (Eds.), *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF) e.V. Band 23*, Potsdam. 2014.
- [44] T. Lillesand and R. Kiefer, *Remote Sensing and Image Interpretation*. New York, NY, USA: Wiley, 1979.
- [45] R. S. Lunetta and C. D. Elvidge, *Remote sensing and Change Detection: Environmental Monitoring Methods and Applications*. Chelsea, MI, USA: Ann Arbor Press, 1998.
- [46] S. Natali and A. Baraldi, "Semantic-geospatial query of remotely sensed image archives," *ESA-EUSC 2006: Image Information Mining For Security and Intelligence*, EUSC Torrejon Air Base, Madrid (Spain), Nov. 27-29, 2006.
- [47] USGS, Cloud Cover Assessment Validation Data. Available online: <http://landsat.usgs.gov/ccavds.php#Austral> (accessed on 2 Dec. 2014).
- [48] V. Kovalsky and D. P. Roy, "A one year Landsat 8 conterminous United States study of cirrus and non-cirrus clouds", *Remote Sens.*, vol. 7, pp. 564-578, 2015.



- [49] Z. Zhu C. E. and Woodcock, “Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4-7, 8, and Sentinel 2 Images,” Remote Sensing of Environment, in press, 2015 (paper for Fmask version 3.2.).
- [50] A. Baraldi, D. Tiede, and S. Lang, “Automatic Linear-Time Prior Knowledge-Based Multi-Level Color Analysis and Synthesis of RGB Imagery for Superpixel Detection and Quality Assessment,” submitted for consideration for publication, IEEE Trans. Pattern Anal. Machine Intell., TPAMI-2015-06-0436, 2015.
- [51] A. Baraldi, D. Tiede, M. Belgiu, and M. Sudmanns, Project winner of the T-Systems Big Data Challenge of the Copernicus Masters 2015: “Satellite Image Automatic Mapper™ (SIAM™)-Through-Time (SIAMT2) for spaceborne/airborne multi-spectral image time-sequence classification in operating mode and content-based image database retrieval” (Project ID 150688)”. Awards Ceremony on Oct. 20, 2015 at the Satellite Masters Conference, 20-22 Oct. 2015, German Federal Ministry of Transport and Digital Infrastructure, Invalidenstraße 444, 10115 Berlin, Germany.
- [52] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis and Machine Vision. London, U.K.: Chapman & Hall, 1994.
- [53] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Machine Intell., vol. 6, no. 1, pp. 1-8, vol. 6, no. 1, 2011.
- [54] J. Clevers, O. Vonder, R. Jongschaap, J. Desprats, C. King, L. Prevot, and N. Bruguier, “Using SPOT data for calibrating a wheat growth model under mediterranean conditions,” Agronomie, EDP Sciences, vol. 22, no. 6, pp.687-694, 2002.

Figures and figure captions in Chapter 12

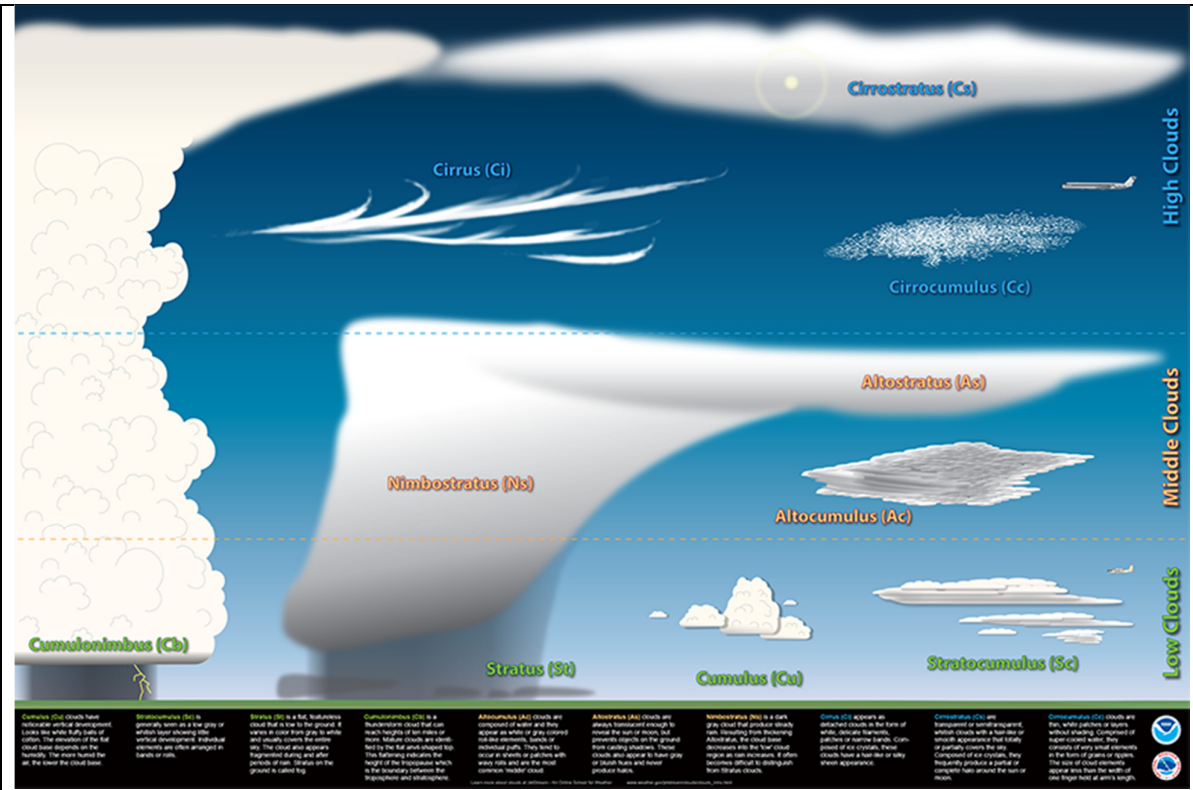


Fig. 12-1. Cloud classification according to the U.S. National Weather Service ([www.srh.noaa.gov/jetstream/clouds/corefour.htm](http://www.srh.noaa.gov/jetstream/clouds/corefour.htm)).

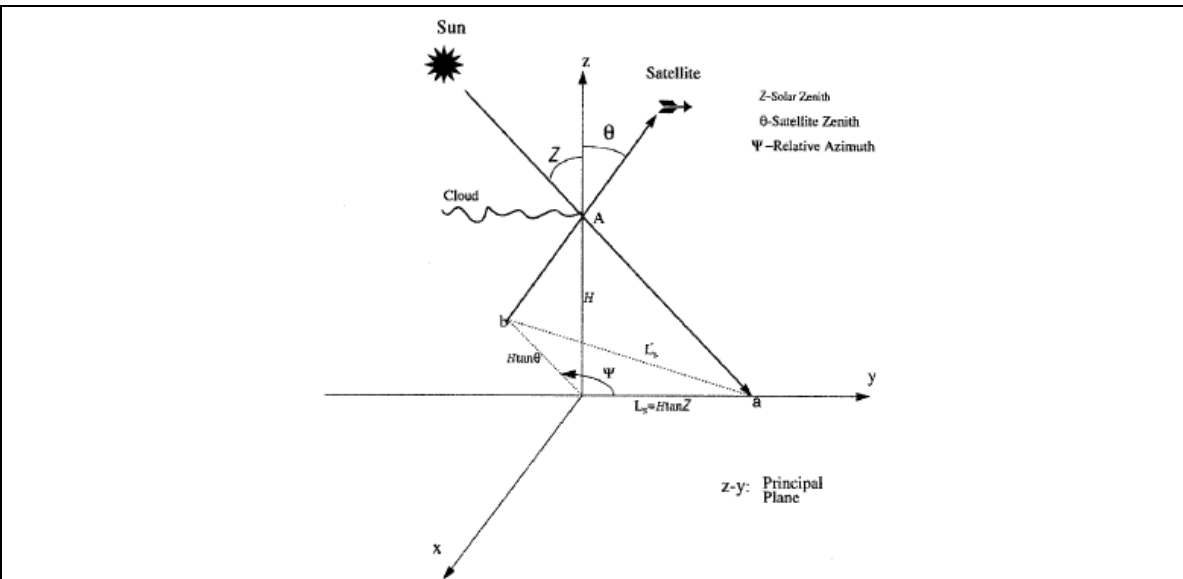


Fig. 12-2. Sun-cloud satellite geometry (leading cloud edge, point A) for arbitrary viewing and illumination conditions.  $H$  and  $LS$  are the actual cloud height and cloud shadow length.  $LS'$  is the apparent cloud shadow length.

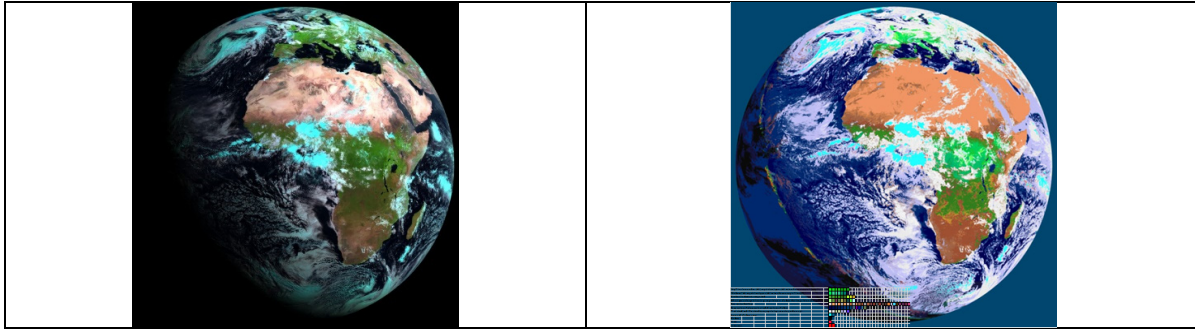


Fig. 12-3(a). Quick-look RGB image of a Meteosat 2<sup>nd</sup> Generation (MSG) SEVIRI image acquired on 2012-05-30, 08:45, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R: band 3, G: band 2, B: band 1), spatial resolution: 3 km. A default ENVI 2% linear histogram stretching was applied for visualization purposes.

Fig. 12-3(b). 4-band AVHRR-like Satellite Image Automatic Mapper (AV-SIAM) preliminary color map, featuring 83 color names depicted in pseudocolors, automatically generated from the MSG image shown in Fig. 12-3(a). Map legend: shown in the lower left corner. For greater details about the SIAM's map legends, refer to text.

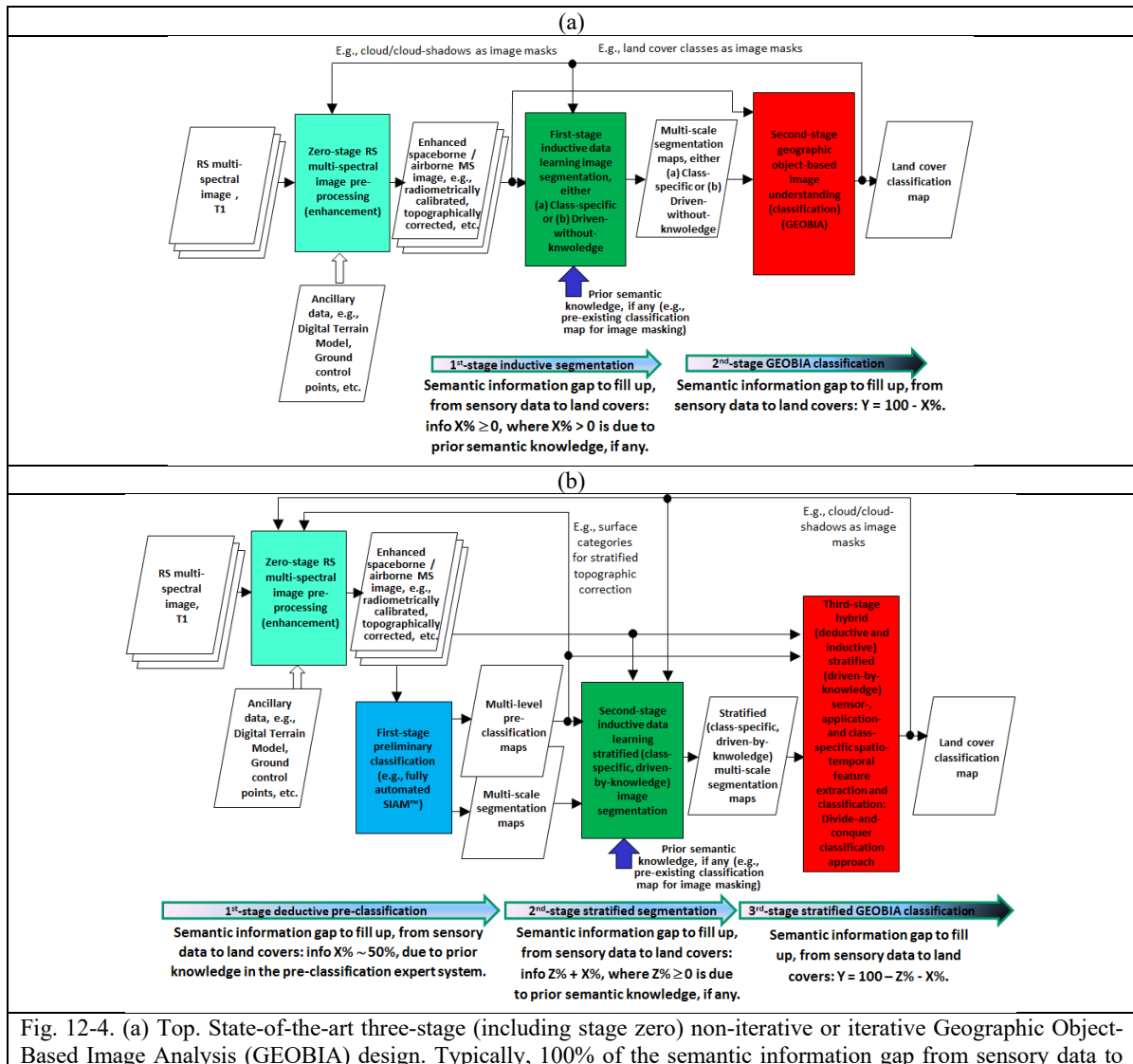


Fig. 12-4. (a) Top. State-of-the-art three-stage (including stage zero) non-iterative or iterative Geographic Object-Based Image Analysis (GEOBIA) design. Typically, 100% of the semantic information gap from sensory data to



land cover classes is filled up in the second step. (b) Bottom. Novel four-stage (including stage zero) hybrid inference-based feedback EO-IUS design. An estimated 50% of the semantic information gap from sensory data to land cover classes is filled up in the automatic deductive pre-classification first stage.

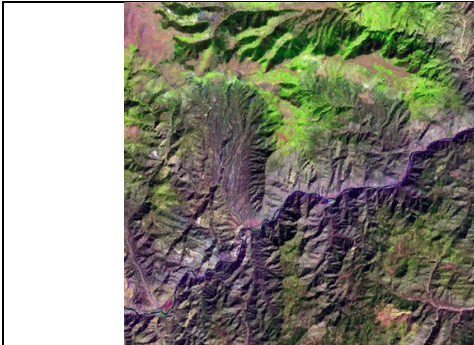


Fig. 12-5(a). Landsat 7 ETM+ image of Colorado, USA (path: 128, row: 021, acquisition date: 2000-08-09), depicted in false colors (R: band ETM5, G: band ETM4, B: band ETM1), 30m resolution, calibrated into TOARF values.

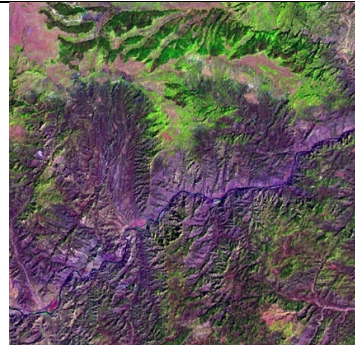


Fig. 12-5(b). Automatic SIAM-driven stratified topographic correction of the Landsat image shown in Fig. 12-5(a), based on a 16-class preliminary spectral map and the SRTM DEM, For further details, refer to [35].

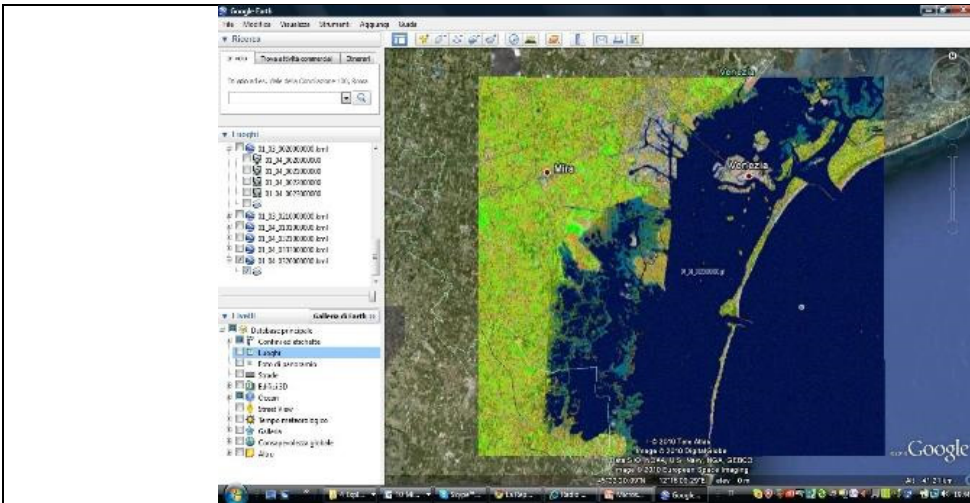


Fig. 12-6. Preliminary classification map, depicted in pseudocolors, generated by the SIAM expert system from a Landsat 7 ETM+ image of the Venice lagoon, Italy, radiometrically calibrated into TOARF values, spatial resolution: 30 m. The SIAM map was transformed into the .kml data format and uploaded as a thematic layer in a commercial 3-D Earth viewer (e.g., Google Earth).

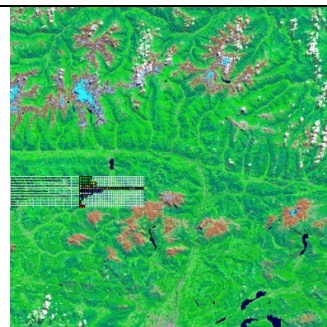
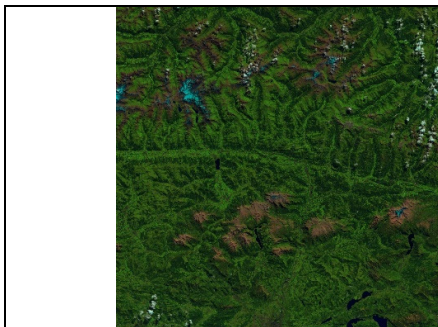




Fig. 12-7(a). False color quick-look RGB image of a Sentinel-2A image Level-1C, calibrated into TOARF values, depicting a surface area around Salzburg, Austria, acquired on 2015-XX-XX. Selected RGB bands are: R = MIR = Band 11, G = NIR = Band 8, B = Blue = Band 2. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes.

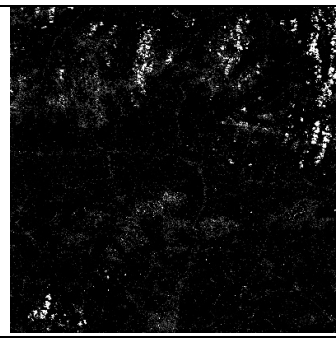


Fig. 12-7(c). Color map-derived cloud mask (spectral categories Core Cloud OR Thick Cloud), gray value = 150. Color map-derived cloud-aura mask (spectral categories Thin cloud over water OR Thin cloud over vegetation), gray value = 50. The cloud mask (gray value = 150) appears affected by false negatives. To recover from these false negatives, the cloud-aura mask (gray value = 50) is necessary. However, the latter appears affected by false positives, to be removed based on spatial analysis of segment shape properties and spatial relationships, refer to Fig. 12-9.

Fig. 12-7(b). First-stage 7-band L-SIAM pre-classification map of the Sentinel-2A image shown in Fig. 12-7(a). The SIAM pre-classification map at fine semantic granularity consists of 96 spectral categories, depicted with pseudocolors, refer to the map legend shown in Table 12-1.

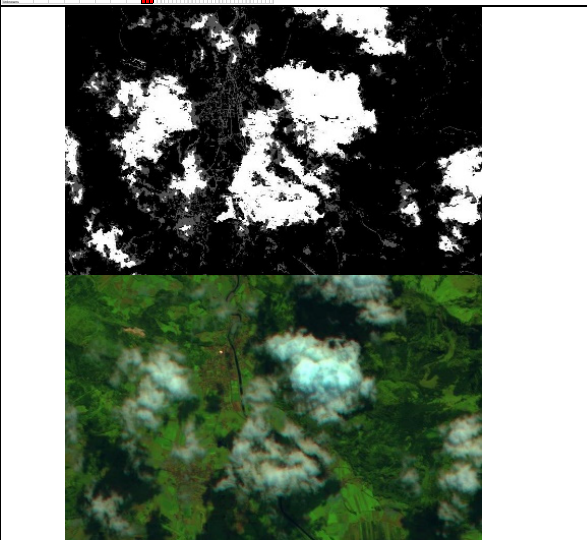


Fig. 12-7(d). Top. Zoomed area of Fig. 12-7(c). Bottom. Zoomed area of Fig. 12-7(a), shown for photointerpretation purposes.

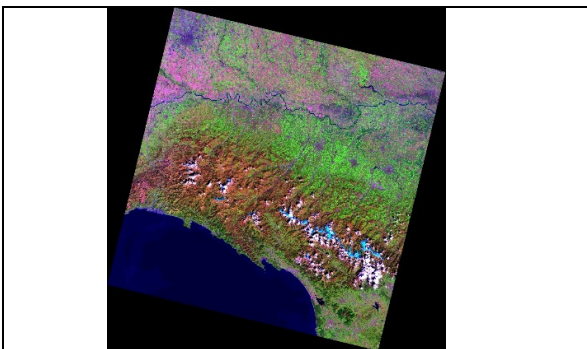


Fig. 12-8(a). False color quick-look RGB image of an uncalibrated Landsat-8 OLI+TIRS image of Northern Italy, acquired on 2014-04-07, after being subject to a color constancy algorithm. In particular: R = MIR = Band 6, G = NIR = Band 5, B = Blue = Band 2. Spatial resolution: 30 m. No histogram stretching is applied for visualization purposes.

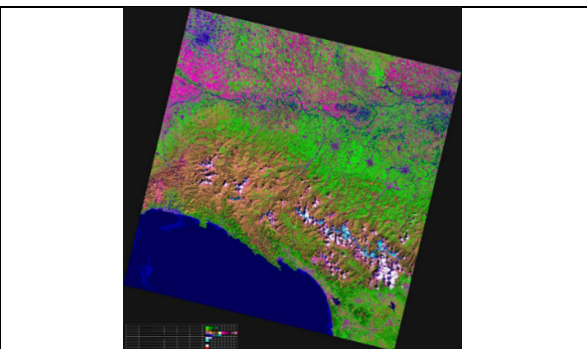


Fig. 12-8(b). First-stage 3-band RGBIAM pre-classification map of the Landsat-8 image shown in Fig. 12-8(a). The RGBIAM pre-classification map at fine semantic granularity consists of 27 spectral categories, depicted with pseudocolors, refer to the map legend shown in Table 12-2.



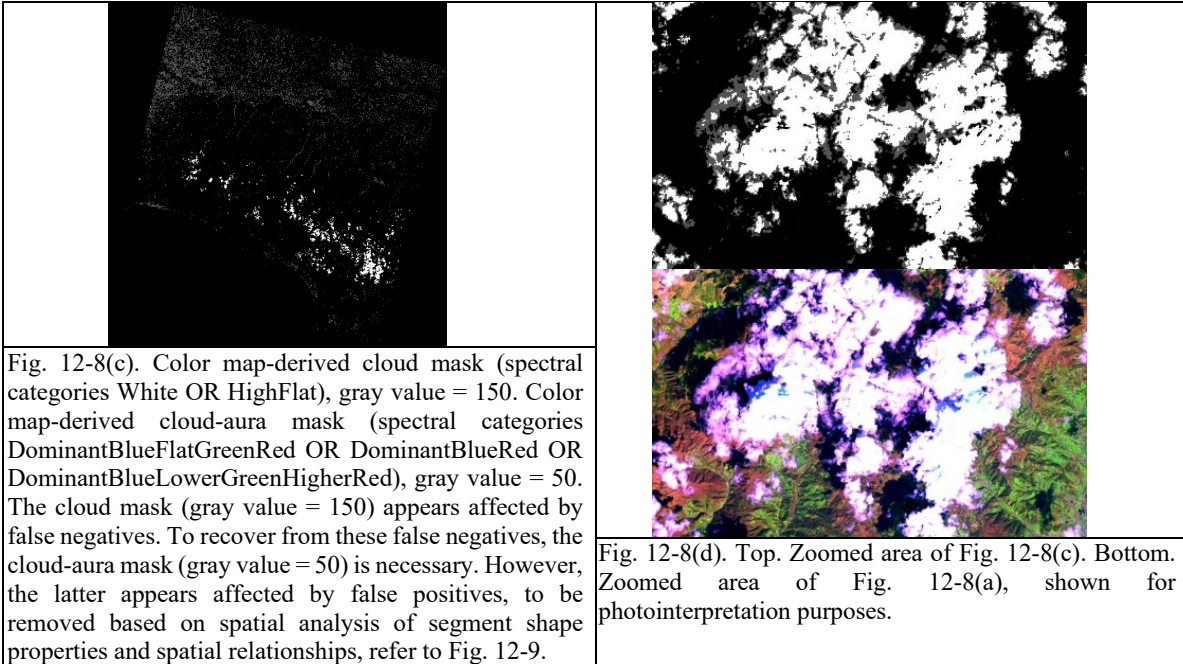


Fig. 12-8(c). Color map-derived cloud mask (spectral categories White OR HighFlat), gray value = 150. Color map-derived cloud-aura mask (spectral categories DominantBlueFlatGreenRed OR DominantBlueRed OR DominantBlueLowerGreenHigherRed), gray value = 50. The cloud mask (gray value = 150) appears affected by false negatives. To recover from these false negatives, the cloud-aura mask (gray value = 50) is necessary. However, the latter appears affected by false positives, to be removed based on spatial analysis of segment shape properties and spatial relationships, refer to Fig. 12-9.

Fig. 12-8(d). Top. Zoomed area of Fig. 12-8(c). Bottom. Zoomed area of Fig. 12-8(a), shown for photointerpretation purposes.

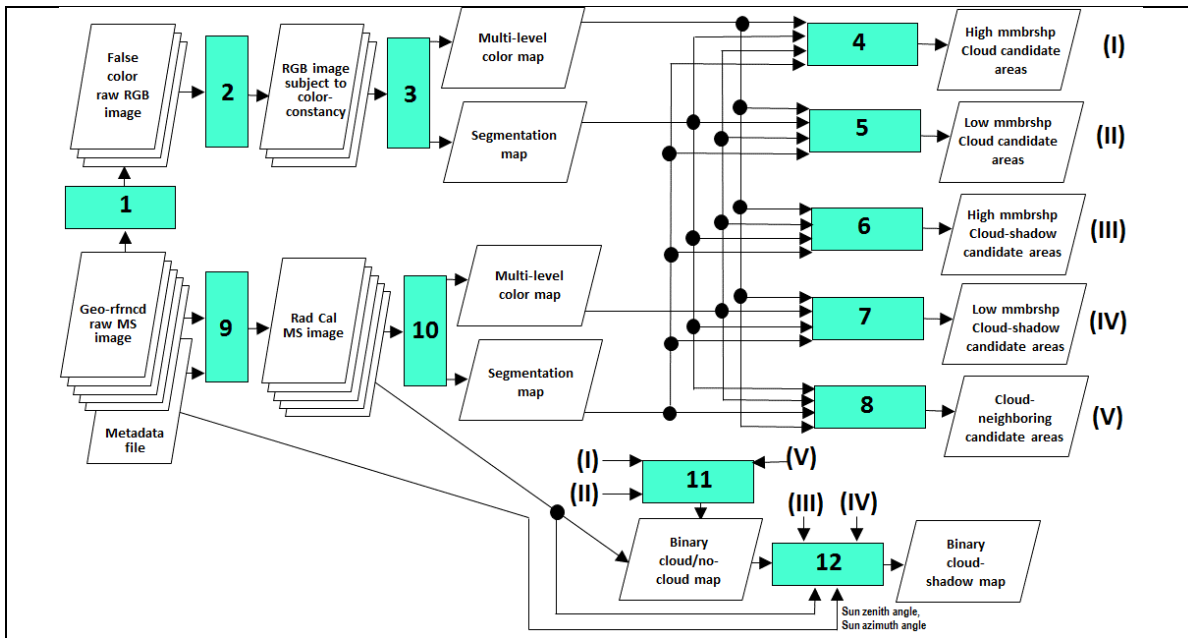


Fig. 12-9. In agreement with the general-purpose 4-stage EO-IUS design sketched in Fig. 12-4(b), the proposed EO-VAS architecture for automatic spatial context-sensitive cloud/cloud-shadow detection consists of a combination of deductive and inductive inference modules. (1) False color RGB channel selection, if possible (belongs to Stage 0 in Fig. 12-4(b)). (2) Statistical self-organizing color constancy algorithm possible (belongs to Stage 0 in Fig. 12-4(b)). (3) Prior knowledge-based RGBIAM color space discretizer (belongs to Stage 1 in Fig. 12-4(b)). (4) AND-Candidate cloud areas (belongs to Stage 2 in Fig. 12-4(b)). (5) NOTAND  $\times$  OR-Candidate cloud areas (belongs to Stage 2 in Fig. 12-4(b)). (6) AND-Candidate cloud-shadow areas (belongs to Stage 2 in Fig. 12-4(b)). (7) NOTAND  $\times$  OR-Candidate cloud-shadow areas (belongs to Stage 2 in Fig. 12-4(b)). (8) Candidate cloud neighboring areas (belongs to Stage 2 in Fig. 12-4(b)). (9) TOARF/SURF radiometric calibration (belongs to Stage 0 in Fig. 12-4(b)). (10) Prior knowledge-based SIAM color space discretizer (belongs to Stage 1 in Fig. 12-4(b)). (11) Spatial modeller (belongs to Stage 3 in Fig. 12-4(b)): Clouds detected from candidate core-cloud and cloud neighboring areas. (12) Spatial modeller (belongs to Stage 3 in Fig. 12-4(b)): Physical model-based cloud shadow detection.



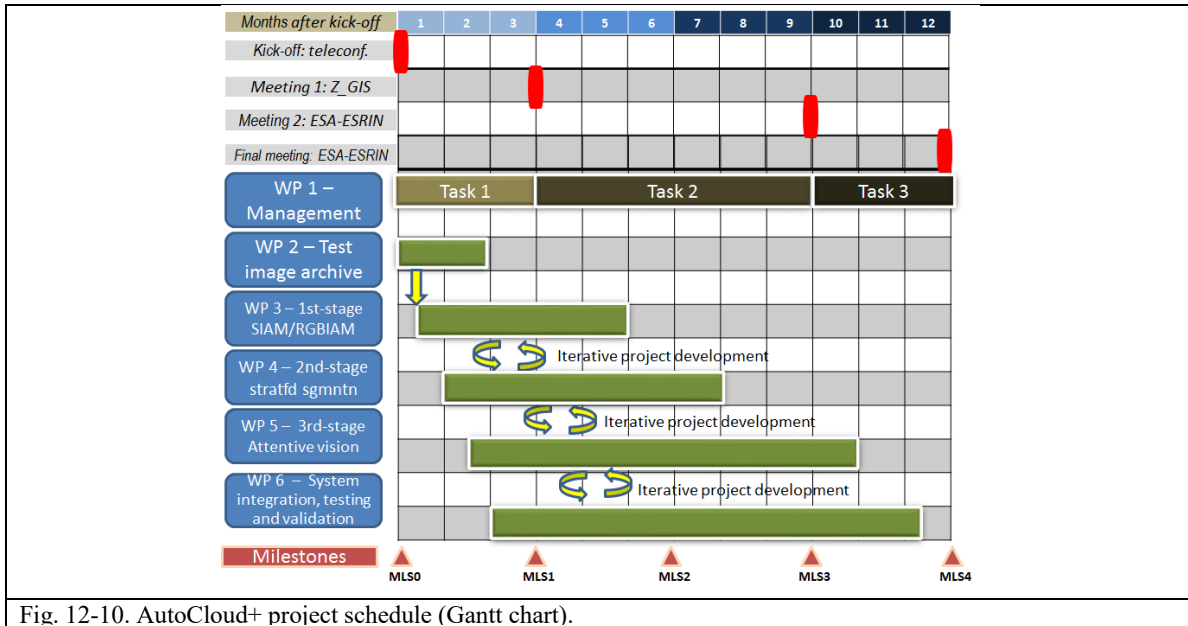


Fig. 12-10. AutoCloud+ project schedule (Gantt chart).



Tables and table captions in Chapter 12

"High" leaf area index (LAI) vegetation types (LAI values decreasing left to right)	
"Medium" LAI vegetation types (LAI values decreasing left to right)	
Shrub or herbaceous rangeland	
Other types of vegetation (e.g., vegetation in shadow, dark vegetation, wetland)	
Bare soil or built-up	
Deep water, shallow water, turbid water or shadow	
Thick cloud and thin cloud over vegetation, or water, or bare soil	
Thick smoke plume and thin smoke plume over vegetation, or water, or bare soil	
Snow and shadow snow	
Shadow	
Flame	
Unknowns	

Table 12-1. Example of a preliminary classification map's legend, adopted by the 7-band Landsat-like SIAM subversion (L-SIAM) at fine semantic granularity, consisting of 96 spectral categories (color names). Pseudocolors of the spectral categories are grouped on the basis of their spectral end member (e.g., brown-as-“bare soil or built-up”) or parent spectral category (e.g., “high” leaf area index (LAI) vegetation types). The pseudocolor of a spectral category is chosen so as to mimic natural colors of pixels belonging to that spectral category.

SRBC_StndrdStrtchdBGR_r1v1_SpCt26 - Map LEGEND of Type 1	
Vegetation	
Rangeland	
Bare soil or built-up	
Water or burned area or shallow water	
Core cloud or cloud aura	
Snow or water ice	
Dark or shadow	
White or cloud	
Unknowns	

Table 12-2. Example of a preliminary classification map's legend adopted by the 3-band RGB Image Automatic Mapper (RGBIAM), suitable for preliminary classification of non-calibrated color images, consisting of 26 spectral categories (color names). Pseudocolors of the spectral categories are grouped on the basis of their spectral end member (e.g., brown-as-“bare soil or built-up”). The pseudocolor of a spectral category is chosen so as to mimic natural colors of pixels belonging to that spectral category.

Paper	Sensor	Spatial resolution	MS bands	Thermal band	Rad. Cal.	Ancillary data	No-cloud/cloud-shadow LC classes		Cloud		Cloud-shadow			
							Inductive (statistical) / Deductive (physical model-based) / Hybrid - LC classes	Pixel/object-based	Inductive / Deductive / Hybrid (Deductive + Inductive)	Pixel/object-based	Pixel/object-based projection from shadow	Cloud eight estimation	Spatial search of cloud-shadow pixels	Inductive / Deductive / Hybrid (Deductive + Inductive)
[1] Luo et al.	MODIS	250 m	B1-B7, from Blue to MIR	N	TOARF or SURF	N	Deductive - LC classes: Bare soil, Vegetation, Snow/ice, Water	Pixel	Deductive	Pixel	Pixel	Assumed ranging from 0.5 km to 12 km	N	Deductive
[2] SPARC	AVHRR	1.1 km	B1-B4, from Red, NIR, MIR to TIR	Y	TOARF and °K (for the thermal band)	N	Deductive - LC classes: Water, Snow	Pixel	Deductive	Pixel	Pixel	Y (based on the TIR band)	Y (in a 4-neighborhood)	N
[3] Huang et al.	Landsat-4/5/7	30 m	TM1-TM7, from Blue to TIR	Y	TOARF and °K (for the thermal band)	DEM	Deductive - LC classes: Water, Dark soil	Pixel	Hybrid (Deductive + Inductive, e.g., tile-based forest)	Pixel + contextual (for cloud boundary dtctn)	Pixel	Y (based on the TIR band)	Y (in the vicinity of predicted shadow pixels)	Inductive



[4], ACCA	Landsat-4/5/7	30 m	TM1-TM7, from Blue to TIR	Y	TOARF and °K (for the thermal band)	N	Deductive – LC classes: Desert soil, Snow	Pixel	Hybrid (Deductive + Inductive, e.g., image-wide histogram-based)	Pixel + contextual (for filling cloud holes)	N	N	N	N	N
[5]	SPOT-4	20 m	B1-B4, from Green, RED, NIR to MIR	N	Green-MIR intercalibration	N	N	N	Inductive	Pixel + object-based	Object	N	Y		Inductive
[25], ATCOR-4	Airborne, spaceborne	Any	B, G, R, NIR, MIR1, MIR2 and Cirrus band (depending on the available spectral channels)	N	TOARF or SURF or Spectral albedo	N	Deductive – LC classes: Water, Land, Haze, Snow/ice (depending on the available spectral channels)	Pixel	Hybrid (Deductive + Inductive, e.g., image-wide histogram-based)	Pixel	N	N	N		Hybrid (Deductive + Inductive, e.g., image-wide histogram-based)
[26], [49] Zhu and Woodcock, Fmask	Landsat-7	30 m	TM1-TM7, from Blue to TIR	Y	TOARF and °K (for the thermal band)	N	Deductive – LC classes: clear land, clear water, snow.	Pixel	Deductive	Pixel + contextual (for isolated map pixel removal)	Pixel + object-based (derived by segmenting the cloud layer)	Y (based on the TIR band)	Y		Hybrid
EO-IUS proposed in the present proposal	Airborne, spaceborne	Any	B, G, R, NIR, MIR1, MIR2 and Cirrus band (depending on the available spectral channels)	Y	TOARF or SURF and °K (for the thermal band)	N	Hybrid – LC classes: Water, Shadow, Bare soil, Built-up, Vegetation, Snow or ice, Fire, Outliers, also refer to [30], [31].	Pixel and object-based, refer to Chapter 12.4	Hybrid	Pixel + object-based, refer to Chapter 12.4	Pixel + object-based, similar to [26], [49]	Y, inspired by [3] and [26], [49]	Y, inspired by [3] and [26], [49]		Hybrid, inspired by [3] and [26], [49]

Table 12-3. List of state-of-the-art cloud/cloud-shadow detectors in the spaceborne MS image domain.

WP No.	WP title (with % of the WP's PM covered by each of the 3 team members; acronyms: XX = ..., YY = ..., ZZ = ...).	Duration person-months (PM) – Total: 15 = 9 (XX, 75%) + 3 (YY, 25%) + 3 (ZZ, 25%), (hours/year @ PLUS: 1720)	Start-End, Kick-off months, Total: 12	Planned results, including Deliverables (Ds), in accordance with the Statement of Work, pp. 8 and 9.
1	Project Management, Quality Assessment/Validation Plan, Scientific results and software products dissemination and exploitation.	2	0-12	<b>Task 1</b> - D1.1 Service chain verification plan; D1.2 Service chain test report; D1.3 Service trial definition; <b>Task 2</b> - D1.4 Service trial report; D1.5 Service viability report; <b>Task 3</b> - D1.6 Action plan for service improvement and expansion; D1.7 Promotional material.
	XX = 50%, YY = 25%, ZZ = 25%; Individual PMs: 1.0 + 0.5 + 0.5 = 2			
2	Stage 0 (data pre-processing). Collection and radiometric calibration of multi-source EO test images and target land cover "truth", namely, cloud, cloud-shadow and clear-sky surfaces.	1.2	0-2	D2.1 Multi-source test image database provided with ground truth D2.2 Multi-source test image database radiometrically calibrated into TOARF values
	XX = 50%, YY = 25%, ZZ = 25%; Individual PMs: 0.6 + 0.3 + 0.3 = 1.2			



3	Stage 1 (pre-attentional vision - raw primal sketch). SIAM and RGBIAM installation, processing and maintenance. XX = 100%, YY = 0%, ZZ = 0%; Individual PMs: 0.5 + 0. + 0. = 0.5	0.5	0.5-5	D3.1 SIAM's products archive generated from multi-source test images
4	Stage 2. Stratified image segmentation (full primal sketch). XX = 70%, YY = 15%, ZZ = 15%; 1.4 + 0.3 + 0.3 = 2	2	2-7	D4.1 Stratified (driven-by-prior knowledge) image segmentation/texture segmentation.
5	Stage 3 (attentional color and spatial reasoning). Original stratified spatial context-sensitive software modules for second-stage attentive vision: development and testing. (I) Stratified geometric feature extraction from image-objects. (II) Estimation of inter-object spatial relationships: (i) topological (e.g., adjacency, inclusion) and (ii) non-topological (spatial distance, angle measures). (III) Cloud detection (to be accomplished before cirrus and haze, refer to procedural knowledge in ATCOR [25]). (IV) Cirrus detection. (V) Haze detection. (VI) Cloud-shadow detection. XX = 60%, YY = 20%, ZZ = 20%; Individual PMs: 4.02 + 1.34 + 1.34 = 6.7	6.7	1.5-11	D5.1 Refinement and integration of an existing software suite for stratified geometric feature extraction from image-objects. D5.2 Software suite to model inter-object spatial relationships: (i) topological and (ii) non-topological. D5.3 Software suite for stratified spatial context-sensitive cloud detection in multi-source test images: D5.3.1 Thermal channel available, D5.3.2 No thermal channel. D5.4 Software suite for stratified spatial context-sensitive cirrus detection in multi-source test images: D5.4.1 Thermal channel available, D5.4.2 No thermal channel. D5.5 Software suite for stratified spatial context-sensitive haze detection in multi-source test images: D5.5.1 Thermal channel available, D5.5.2 No thermal channel. D5.6 Software suite for stratified spatial context-sensitive cloud-shadow detection in multi-source test images: D5.6.1 Thermal channel available, D5.6.2 No thermal channel.
6	System integration, metrological/statistically-based quality assessment (testing) and third-party validation. XX = 60%, YY = 20%, ZZ = 20%; Individual PMs: 1.5 + 0.5 + 0.5 = 2.5	2.5	3-11.5	D6.1 Software suite for automatic cloud / cloud-shadow masking in multi-sensor multi-spectral EO images. D6.2 Test reference samples. D6.3 Validation reference samples.

Table 12-4. AutoCloud+ work plan, with an overview of work packages (WPs) scheduled for the proposed 12-month project, including deliverables (Ds), in agreement with Fig. 12-10. Team member acronyms: XX = ..., YY = ..., ZZ = ...

Milestone (MLS) No.	MLS title. MLS is achieved: at the end of every 1-of-4 iterations = 12 months / 4 = 3 months.	Relevant WPs involved	Expected date from (months kick-off)	Notes about WPs.
0	Kick-off		0	
1	Iteration 1, 1/4 of the software-life-cycle project requirements, from software analysis to design, code and test.	WP1 to WP6	3	WP2 is terminated. End of Task 1.
2	Iteration 2, 2/4 of the software-life-cycle project requirements, from software analysis to design, code and test.	WP1 to WP6	6	WP3 is terminated.
3	Iteration 3, 3/4 of the software-life-cycle project requirements, from software analysis to design, code and test.	WP1 to WP6	9	WP4 is terminated. End of Task 2.
4	Iteration 4, 4/4 of the software-life-cycle project requirements, from software analysis to design, code and test.	WP1 to WP6	12	WP5 and WP6 are terminated. End of Task 3.

Table 12-5. AutoCloud+ project overview of milestones (MLSs), in agreement with Fig. 12-10.

ID	Software tool to be delivered in protected format	Description	Software Language Implementation(s)	License Owner	Author(s)
1	RadCal	Library of sensor-specific TOARF calibrators, whose output is a MS image file in the SIAM's input file format	IDL	Andrea Baraldi	Andrea Baraldi
2	RGB cube color constancy	Automatic self-organizing RGB cube color constancy algorithm	IDL	Andrea Baraldi	Andrea Baraldi
3	SIAM	Expert system for MS calibrated image analysis (MS space partitioning into color names) and synthesis (reconstruction), equivalent to superpixel detection and quality assessment	C/C++, IDL	Andrea Baraldi	Andrea Baraldi
4	TPMLVIAS	Two-pass Multi-level Image Partition (analysis) and Piecewise Constant Continuous Image Reconstruction (synthesis, object-mean view)	C/C++	Andrea Baraldi	Andrea Baraldi
5	RGBIAM	Expert system for RGB image analysis (RGB cube partitioning into color names) and synthesis (reconstruction), equivalent to superpixel detection and quality assessment, where the RGB image must be subject to color constancy	IDL	Andrea Baraldi	Andrea Baraldi

Table 12-6. AutoCloud+ project background: Software tools to be delivered in protected format. Acronyms: Interfaculty Department of Geoinformatics – Z\_GIS of the Paris-Lodron University of Salzburg (PLUS; www.uni-salzburg.at).



### 13 R&D Summary and Future Works

To comply with guidelines for a successful Doctor of Philosophy (PhD) experience proposed in the engineering/computer science literature {1}, {2} (refer to Premise), this scientific dissertation does the following.

- In the interdisciplinary domain of cognitive science, it provides two original working hypotheses, refer to Chapter 1 (Doctoral Research Objectives and Definitions of Interest).
  - ✓ Vision is a cognitive (*information-as-data-interpretation*) task, synonym of scene-from-image reconstruction and understanding, acknowledged to be inherently ill-posed in the Hadamard sense and non-polynomial (NP)-hard in computational complexity. Vision encompasses both computer vision (CV) and human vision. The primary working hypothesis of this doctoral study was that no CV system in operating mode, capable of operational, comprehensive and timely Earth observation (EO) big image understanding (EO-IU) exists to date. In practice, systematic image understanding, including EO big data interpretation as a special case, remains unaccomplished to date. This is tantamount to saying that the visionary goal of the Group on Earth Observations (GEO)'s Global Earth Observation System of Systems (GEOSS), expected to transform large-scale multi-source EO image databases into timely, comprehensive and operational EO value-adding information products and services, is yet-unaccomplished by the remote sensing (RS) community. To be solved in operating mode, the cognitive problem of image understanding requires the CV and remote sensing (RS) communities to collaborate within the interdisciplinary domain of cognitive science. It is postulated that  $CV \supset EO\text{-}IU$  in operating mode, synonym of  $GEOSS \supset \text{European Space Agency (ESA) EO Level 2 product} \supset \text{human vision}$ , where human vision is considered a lower bound of CV, i.e., the inherently ill-posed cognitive problem of CV is conditioned by human visual perception to become better posed for numerical solution. This is tantamount to saying that a CV system in operating mode is required to include a computational model of human vision. Defined as a single-date multi-spectral (MS) image corrected for atmospheric, adjacency and topographic effects, stacked with its scene classification map (SCM), no ESA EO Level 2 product has ever been systematically generated at the ground segment.
  - ✓ The secondary working hypothesis of this doctoral study was that image understanding in operating mode is a necessary not sufficient pre-condition for semantic content-based image retrieval (SCBIR), where SCBIR is synonym of semantics-enabled information/knowledge discovery in massive image databases. In practice,  $SCBIR \supset CV$  in operating mode, i.e., the complexity of SCBIR is not inferior to the complexity of CV, acknowledged to be inherently ill-posed and NP-hard. No SCBIR system in operating mode exists to date.

Combined in series, these two working hypotheses postulate that  $SCBIR \supset CV \supset EO\text{-}IU$  in operating mode, synonym of  $GEOSS \supset \text{ESA Earth observation (EO) Level 2 product} \supset \text{human vision}$ . Hence, the solution of the primary (dominant, necessary not sufficient) CV problem can lead to a solution of the secondary (dependent) SCBIR problem, which are both cognitive problems open to solution in operating mode to date. In the RS common practice, systematic ESA EO Level 2 product generation is considered a GEOSS proof-of-concept necessary and sufficient for EO-SCBIR initialization.

- In Chapter 1 (Doctoral Research Objectives and Definitions of Interest), this dissertation recognizes CV and SCBIR as two cognitive problems open to solution in operating mode with respect to the state-of-the-art. It postulates a complexity relationship, where  $SCBIR \supset CV$  in operating mode  $\supset \text{human vision}$ , as an opportunity to better condition the inherently ill-posed CV problem with human visual perception in addition to sensory data. Next, to contribute toward filling an analytic and pragmatic information gap from EO big sensory data to timely, comprehensive and operational EO value-adding information products and services, it clearly states the ambitious, but realistic (well-conditioned) goals of this research & technical development (R&D) doctoral project in the multidisciplinary domain of cognitive science. First, to deliver an EO-IU subsystem in operating mode, capable of systematic ESA EO Level 2 product generation without human-machine interaction and in near real-time as a GEOSS proof-of-concept in support of SCBIR. Second, to develop a closed-loop EO image understanding for semantic querying (EO-IU4SQ) system prototype capable of incremental learning, where EO value-adding products generated as output increase their value-added with closed-loop iterations.
- It includes Chapter 2 (Introduction) to become accessible to everyone in engineering/computer science, not just to specialists.





- It is provided with a relevant survey value, e.g., refer to Chapter 3 (Computational models of human vision), Chapter 4 (Manuscript 1), Chapter 7 (Manuscript 4), Chapter 9 (Manuscript 6), Chapter 10 (Manuscript 7), Chapter 11 (Manuscript 8) and Chapter 12 (Technical report 2).
- It provides original, rigorous, experimental and/or formal arguments capable of convincing fellow scientists to substantiate or refute the proposed working hypotheses, e.g., refer to Chapter 3 (Computational models of human vision), Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2).
- It provides a new theory or concept at the levels of understanding of CV system requirements specification, knowledge/information representation and CV system architecture, refer to Chapter 3 (Computational models of human vision), Chapter 7 (Manuscript 4), Chapter 9 (Manuscript 6), Chapter 10 (Manuscript 7) and Chapter 12 (Technical report 2), together with a new closed-loop EO-IU4SQ system modular design, where the relationship between two known CV and SCBIR open problems is formalized as  $SCBIR \supset CV$  in operating mode. In the original EO-IU4SQ system architecture the two hybrid feedback EO-IU and EO-SQ subsystems are connected in closed-loop for incremental learning starting from an ESA EO Level 2 product generation as initial condition, refer to Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2).
- It presents and discusses new CV/ EO-IU/ EO-IU4SQ system solutions at the levels of understanding of algorithm and implementation. Sufficient information is provided for the implementation to be reproduced. Breaking points and failure modes of the proposed CV/ EO-IU/ EO-UE4SQ systems' algorithm and implementation are clearly acknowledged, if any, together with their advantages in terms of a minimally dependent and maximally informative (mDMI) set of outcome and process quantitative quality indicators (OP-Q<sup>2</sup>Is) provided with a degree of uncertainty in measurement in compliance with the Quality Assurance Framework for Earth Observation (QA4EO) *Val* requirements, e.g., refer to Chapter 3 (Computational models of human vision) and Chapter 4 (Manuscript 1). In agreement with project requirements, refer to Chapter 1 (Doctoral Research Objectives and Definitions of Interest) and Chapter 3 (Computational models of human vision), all proposed CV algorithms are fully automated, requiring neither user-defined parameters nor training data to run, and near real-time, with a computational complexity linear in image size.

A detailed summary of the original contributions in R&D of the present dissertation is proposed below.

### **Chapter 1 - Doctoral Research Objectives and Definitions of Interest**

In Chapter 1 (Doctoral Research Objectives and Definitions of Interest) the first working hypothesis was that no EO-IU system (EO-IUS) exists to date capable of transforming large-scale multi-source EO image databases into timely, comprehensive and operational EO value-adding information products and services, in compliance with the visionary goal of the intergovernmental GEO's GEOSS implementation plan for years 2005-2015, submitted to (constrained by) the Quality Assurance Framework for Earth Observation (QA4EO) calibration/validation (*Cal/Val*) requirements. In practice, no GEOSS exists to date, where GEOSS is synonym of EO-IUS in operating mode. To be considered in operating mode an EO-IUS has to score "high" in each quantitative quality indicator (Q<sup>2</sup>I) belonging to a minimally dependent and maximally informative (mDMI) set of outcome and process Q<sup>2</sup>Is (OP-Q<sup>2</sup>Is), proposed in Chapter 1 to be community-agreed upon. The conjecture that EO-IUSs in operating mode are lacking in the RS common practice is supported by several facts. For example, in 2002 the percentage of EO data ever downloaded from the European Space Agency (ESA) databases was estimated at about 10% or less. In addition, no ESA EO Level 2 product has ever been systematically generated at the ground segment. By definition an ESA EO Level 2 product comprises a single-date MS image radiometrically calibrated into surface reflectance (SURF) values corrected for geometric, atmospheric, topographic and adjacency effects, stacked with its data-derived general-purpose, user- and application-independent SCM, whose legend includes quality layers such as cloud and cloud-shadow. In greater detail, the first working hypothesis postulated that  $CV \supset EO-IU$  in operating mode  $\supset$  ESA EO Level 2 product  $\supset$  human vision, where human visual perception is considered a lower bound of CV within the multi-disciplinary domain of cognitive science, which is tantamount to saying that an inherently ill-posed CV system in operating mode is required to include a computational model of human vision to become better conditioned for numerical solution. Encompassing both human vision and CV in the interdisciplinary domain of cognitive science, vision is a cognitive problem, synonym of scene-from-image reconstruction and understanding, known as non-polynomial (NP)-hard in computational complexity and inherently ill-posed in the Hadamard sense. Hence, it requires a hybrid (combined deductive and inductive) inference approach to combine sensory data with *a priori* knowledge, available in addition to data, to become better posed for plausible solution(s). Vision is inherently ill-posed because affected by a 4D-to-2D data dimensionality reduction from the 4D spatiotemporal scene-domain to the (2D) image-domain, e.g., responsible of



occlusion phenomena, and by a semantic information gap from ever-varying sensory data to stable percepts belonging to a symbolic world ontology (world model, mental world) provided with semantics, equivalent to stable percepts. In the words of Iqbal and Aggarwal “biological vision is currently the only measure of the incompleteness of the current stage of CV, and illustrates that the CV problem is still open to solution” {19}. According to Pessoa, “if we require that a CV system should be able to predict perceptual effects, such as the well-known Mach bands illusion where bright and dark bands are seen at ramp edges (Figure 1-21), then the number of published vision models becomes surprisingly small” {20}. Hence, the doctoral project requirement that CV in operating mode  $\supset$  human vision has two goals. In general, it provides the CV system with R&D constraints, e.g., linear time complexity, required to make the inherently ill-posed vision problem better posed for numerical solution. In particular, it requires the CV system to agree with human visual perception.

The second working hypothesis of Chapter 1 was that no semantic content-based image retrieval (SCBIR) system is available to date, to process semantic queries in EO big data repositories such as “retrieve all EO images not necessarily cloud-free acquired by imaging sensor X where wetland areas are visible and located adjacent to a highway near a coast in the eastern part of country Y”, because existing EO content-based image retrieval (CBIR) systems lack EO-IU capabilities, i.e., SCBIR  $\supset$  CV  $\supset$  EO-IU in operating mode  $\supset$  human vision, where SCBIR is synonym of semantics-enabled information/knowledge discovery in big image databases. It means that SCBIR is not less difficult to solve than CV in operating mode, where vision, encompassing CV, is known to be very difficult to solve because NP-hard in computational complexity and inherently ill-posed in the Hadamard sense. Equivalent to two sides of the same cognitive problem, the SCBIR problem solution depends on the preliminary solution of the CV problem. In the RS common practice, semantic enrichment of large-scale multi-source EO image databases is considered a necessary not sufficient pre-condition for EO-SCBIR.

Combined in series, the two working hypotheses postulate that no EO-SCBIR can be realistically pursued if no EO-IU system in operating mode is available in advance. Unfortunately, no EO-IU in operating mode, synonym of GEOSS, exists to date. This explains why no EO-SCBIR system in operating exists to date either.

To overcome these lacks, the overarching goal of this R&D doctoral study was to develop a closed-loop EO-IU4SQ system prototype as a GEOSS proof-of-concept in support of SCBIR. This ambitious R&D project objective can be considered realistic if and only if a primary (dominant, necessary not sufficient) EO-IU subsystem in operating mode is delivered as initial pre-condition of a secondary (dependent) EO-SQ subsystem prototype. Chapter 3 (Computational models of human vision) provides a major contribution in making realistic the ambitious goal of this doctoral project.

## *Chapter 2 - Introduction*

Chapter 2 provides this doctoral dissertation with a relevant survey value to make it accessible to everyone in engineering/computer science, not just to specialists.

At the level of CV system understanding of visual information/knowledge representation, the obvious (based on common sense), but not trivial contribution of Chapter 2 is twofold. First, Chapter 2.6 highlights the inherent information loss of a spectral index, equivalent to an angular coefficient of a tangent in one point of a spectral signature, when it is employed as a compressed univariate (scalar variable) representation of a multivariate spectral variable equivalent to a spectral signature. Second, Chapter 2.7 stresses the undisputable fact that spatial topological and spatial non-topological information components dominate color information in vision. This true-fact is proved by observing that achromatic (panchromatic) human vision, familiar to everybody when wearing sunglasses, is nearly as effective as chromatic vision in scene-from-image representation and understanding. It means that the following necessary and sufficient condition holds for a CV system.

*A necessary and sufficient condition for a CV system to fully exploit spatial topological and spatial non-topological information components in addition to color is to perform nearly as well when input with either panchromatic or color imagery.*

Proofs that the aforementioned consideration, based on common sense in agreement with undergraduate knowledge of function analysis, are not at all trivial are that, first, spectral indexes dominate spectral pattern recognition in the RS literature. Second, 1D image analysis approaches, where a (2D) image is transformed into a 1D vector sequence of either pixel-based or spatial context-sensitive vector data, dominate the RS literature, including object-based image analysis (OBIA) algorithms. 1D image analysis is insensitive to permutations in the order of presentation of the input vector data sequence. As a consequence, it ignores spatial topological information which typically dominates color information in vision. These simple considerations explain why deep convolutional neural networks (DCNNs), capable of 2D image



analysis as synonym of retinotopic/topology-preserving (TP) feature mapping, typically outperform traditional 1D image analysis approaches.

**Chapter 3 - Technical report 1 (unpublished): Computational models of human vision: Developments and open challenges in automated detection of multi-scale image-contours, keypoints, texels and texture-boundaries in panchromatic and color images**

Provided with a relevant survey value, Chapter 3 presents and discusses an innovative pre-attentive (low-level, sub-symbolic) CV subsystem, designed and implemented to generate automatically (without human-machine interaction) in linear time (with computational complexity increasing linearly with image size) a raw primal sketch (sub-symbolic image segmentation into image-objects) and a full primal sketch (sub-symbolic texture segmentation of an image-plane into perceptually homogeneous textured areas), defined in line with the Marr terminology. Implemented as an innovative multi-scale spatial filter bank, this pre-attentive CV subsystem is a retinotopic/ TP convolutional neural network (TPCNN), capable of 2D image analysis, alternative to a large majority of existing image segmentation and texture segmentation approaches implemented as inductive learning-from-data algorithms, which are inherently ill-posed, semi-automatic (they require system-free parameters to be user-defined based on heuristics) and training data-specific, and/or 1D image analysis approaches, where a (2D) image is transformed into a 1D vector data sequence of either pixel-based or spatial context-sensitive vector data. Insensitive to permutations in the order of presentation of the input vector data series, 1D image analysis ignores spatial topological information, which typically dominates color information in vision (refer to Chapter 2, Introduction). The proposed multi-scale spatial filter bank is capable of automated near real-time image analysis (decomposition) and lossless synthesis (reconstruction), zero-crossing (ZX) image-contour detection, keypoint (endpoint, corner, T- and X-junction) detection, ZX image-object segmentation and texture segmentation, in agreement with a retinotopic/ TP/ 2D image analysis approach consistent with the Mach bands illusion in human visual perception. Its several degrees of novelty are summarized at the end of Chapter 3. Among these several degrees of novelty, the following are worth mentioning.

- Chapter 3.6 provides an original requirements specification for computational models of human vision.
- Chapter 3.9. Original automated statistical model-based color constancy algorithm, to be applied to panchromatic and color images provided with no metadata calibration file. In human vision, color constancy ensures that the perceived color of objects remains relatively constant under varying illumination conditions, but the biological mechanisms involved with the color constancy ability are not yet fully understood.
- Chapter 3.11. Lossless 1D (and 2D) signal reconstruction (synthesis) following signal analysis (near-orthogonal decomposition) consistent with the Mach bands illusion.

$$\text{Renstret}(x) = f(x) \circ G(x) + [f(x) \circ \partial^2 G / \partial x^2] / 2.$$

- Chapter 3.12. Operational unequivocal parameter-free definition of an image-contour pixel.

*A pixel  $I(n)$  with pixel coordinates  $n = (x, y)$  in a 2D array is an image-contour pixel if it is a zero-crossing (ZX) pixel, where the image local concavity, equal to  $[I(n) \cdot \partial^2 G / \partial n^2]$ , changes in sign, either from positive to non-positive, i.e., from positive to either zero or negative, or from negative to non-negative, i.e., from negative to either zero or positive, in comparison with the local concavity of any of its 8-adjacency neighboring pixel.*

Noteworthy, because the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  is a nonlinear operator, it neither commutes nor associates with the convolution. Therefore, the filtered image  $(\partial^2 G / \partial n^2 * I)$  is different from the 2<sup>nd</sup>-order derivative  $\partial^2 / \partial n^2$  applied to the low-pass image adopted by both Canny and Bertero, Torre and Poggio.

$$(\partial^2 G / \partial n^2 * I) \neq \partial^2 / \partial n^2 (G * I).$$

To the best of this author's knowledge, this analytic description of an image-contour, consistent with the Mach bands illusion, is original and alternative to those provided in the CV and the RS literature, such as those based on user-defined thresholding of local gradient, local contrast or local variance statistics in the image-domain.

- Chapter 3.13. Original perceptual image-pair (visual) dissimilarity metric (PVDm), consistent with human visual perception, including the Mach bands illusion. According to a relevant portion of the cognitive science and CV literature, the primary use of image quality metrics is to quantitatively measure an image quality that correlates with perceptual visual quality. So-called perceptual visual quality metrics, PVQMs, are objective models for predicting subjective visual quality scores, like the resultant mean opinion score (MOS) obtained by many observers through

repeated viewing sessions. No community-agreed PVQM exists to date. The proposed PVDM between a reference and test image-pair is

$$PVDM(I_R, I_T) = |I_R(x) \circ G(x) - I_T(x) \circ G(x)| + \left| \frac{I_R(x) \circ \frac{\partial^2 G}{\partial x^2}}{2} - \frac{I_T(x) \circ \frac{\partial^2 G}{\partial x^2}}{2} \right|,$$

where  $|I_R(x) \circ G(x) - I_T(x) \circ G(x)| \in [0, \text{MaxGrayValue}]$  and  $|\frac{I_R(x) \circ \partial^2 G / \partial x^2}{2} - \frac{I_T(x) \circ \partial^2 G / \partial x^2}{2}| \in [0, \text{MaxGrayValue}]$ . In mathematical terms, this is a Minkowski distance with degree  $d$  equal to 1. If appropriate,  $d$  can be set equal 2 (to apply to a Euclidean distance) or superior. Since it is a perceptual visual quality distance rather than a perceptual visual similarity indicator,  $PVDM(I_R, I_T) \in 2 * [0, \text{MaxGrayValue}]$  is best when minimized.

- Chapter 3.20. Original automated (parameter-free) implementation of a raw primal sketch consisting of discrete tokens (texels) as ZX segments. Unfortunately, in his seminal work Marr proposed no algorithm to extract ZX pixels, ZX segments and tokens from ZX segments. According to Li Zhaoping, "the computer vision community has tried to solve the problem of image segmentation for decades without a satisfactory solution."
- Chapter 3.21. Original conceptual unifying framework for spatial variance, spatial autocorrelation and the proposed 2D wavelet filter bank.
- Chapter 3.22. The mDMI set of OP-Q<sup>2</sup>Is characterizing the implemented low-level CV system for automated near real-time raw and full primal sketch generation is reported below in Table 13-1.

Legend of fuzzy sets of a quantitative variable.	
LOW	
MEDIUM	
HIGH	
<b>Computer vision (CV) Process (Pracs) and Outcome (Otcn) Q<sup>2</sup>Is ± δ ⊆ QA4EO Val</b>	Low-level vision: raw/full primal sketch
<b>Degree of automation (Pracs):</b> (a) inversely related to the number, physical meaning and range of variation of user-defined parameters, (b) inversely related to the collection of the required training data set, if any.	HIGH (unsurpassed, no free-paramtr)
<b>Effectiveness (Otcn), in agreement with human visual perception, i.e., CV ⊃ human vision, where human visual perception is a lower bound of CV.</b>	a) HIGH b) YES ➤ HIGH c) HIGH d) None
<b>Semantic information level (Pracs)</b>	LOW (sub-symbolic)
<b>Efficiency (Pracs):</b> (a) computation complexity in image size: Polynomial (P), P and linear (L), non-P (NP), and (b) run-time memory occupation.	(a) HIGH (L), (b) HIGH
<b>Robustness to changes in input image (Pracs),</b> e.g., large spatial extent data mapping (no toy problems).	HIGH
<b>Robustness to changes in input parameters (Pracs),</b> e.g., sensitivity analysis.	HIGH (unsurpassed, no free-paramtr)
<b>Scalability to changes in the sensor's specifications or user's needs (Pracs),</b> e.g., (a) pncchrmtc, (b) RGB, true- or false-color, (c) multi-spectral (MS), (d) super-spectral (SS), (e) hyper-spectral (HS).	HIGH, (a) YES, (b) YES, (c) YES, (d) YES, (e) YES.
<b>(Inverse of) Timeliness (Otcn),</b> from data acquisition to high-level product generation, increases with manpower and computing power.	HIGH
<b>(Inverse of) Costs (Otcn),</b> increasing with (a) manpower and (b) computing power.	(a) HIGH, (b) HIGH

Table 13-1, equivalent to Table 3.23-1. Outcome and process (OP) quantitative quality indicators (OP-Q<sup>2</sup>Is) of the proposed low-level CV system design and implementation.

*Chapter 4 - Manuscript 1 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal Remote Sensing of Environment): Systematic Earth Observation Level 2 product generation for semantic queryings*

*Chapter 5 - Manuscript 2 (unpublished, currently submitted for consideration for publication in the peer-reviewed journal European Journal of Remote Sensing): Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases*





In Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2) an innovative closed-loop EO-IU4SQ system architecture and prototypical implementation is proposed as proof-of-concept of a yet-unaccomplished GEOSS in support of a yet-unaccomplished SCBIR. This R&D project starts from the original postulate that SCBIR  $\supset$  CV  $\supset$  EO-IU in operating mode, synonym of GEOSS  $\supset$  ESA EO Level 2 product  $\supset$  human vision. In this postulate, human visual perception is considered a lower bound of CV within the multi-disciplinary domain of cognitive science, which is tantamount to saying that an inherently ill-posed CV system in operating mode is required to include a computational model of human vision to become better posed for numerical solution. Defined as single-date MS image corrected for atmospheric, adjacency and topographic effects, stacked with its SCM whose legend is general-purpose, user- and application-independent, no ESA EO Level 2 product has ever been systematically generated at the ground segment. In the closed-loop EO-IU4SQ system, systematic ESA EO Level 2 product generation is considered a necessary not sufficient initial condition for SCBIR and semantics-enabled information/knowledge extraction. EO value-adding information products generated as output by the closed-loop EO-IU4SQ system are expected to monotonically increase their value-added with closed-loop iterations.

The closed-loop EO-IU4SQ system architecture comprises a primary (necessary not sufficient) automated EO-IU subsystem in operating mode in closed-loop with a secondary (dependent) EO-SQ subsystem, provided with a graphic user interface (GUI) to streamline human-machine interaction for semantic querying and semantics-enabled information/knowledge discovery.

The implemented primary EO-IU subsystem adopts an innovative hybrid (combined deductive and inductive) feedback CV architecture, capable of TP image analysis based on a convergence-of-evidence approach, where dominant spatial topological and spatial non-topological information components in the image-domain are investigated in addition to color information, typically dominated by spatial information. This CV system design is alternative to the inductive feedforward CV architecture suitable for 1D image analysis typically adopted by the RS community, where spatial topological information in the image-domain is totally ignored. The proposed EO-IU subsystem implementation comprises the pre-attentional CV subsystem for automated sub-symbolic color naming, ZX image-object segmentation, planar shape analysis and texture segmentation discussed in Chapter 3 (Computational models of human vision), Chapter 7 (Manuscript 4), Chapter 8 (Manuscript 5), Chapter 9 (Manuscript 6) and Chapter 10 (Manuscript 7). It is followed by an attentional (high-level, symbolic) EO image classifier capable of systematic ESA EO Level 2 product generation, where the general-purpose ESA EO Level 2 SCM legend is selected as the standard 3-level 8-class Land Cover Classification System (LCCS) Dichotomous Phase (DP) taxonomy, originally proposed by the Food and Agriculture Organization of the United Nations (FAO), augmented to include quality layers cloud and cloud-shadow.

An *a priori* world model (world ontology, mental world) must be transferred from human-to-machine according to a deductive (top-down) inference approach, complementary not alternative to inductive (bottom-up) machine learning-from-data. The GUI of the EO-SQ subsystem supports the deductive knowledge transfer of a world model from users to the EO-IU4SQ system. The world model is an Entity-Relationship (ER) conceptual model of 4D geospatiotemporal information about entities as classes of spatiotemporal objects in the physical world and their relationships, either spatiotemporal (e.g., adjacency, inclusion, etc.) or semantics (e.g., part-of, subset-of). The ER conceptual model of the 4D physical world is graphically represented as a semantic network where entities are nodes and relationships are arcs between nodes. In addition, the GUI streamlines SCBIR operations and semantics-enabled information/knowledge discovery, constrained by the world model.

The latest EO-IU4SQ system implementation stores its fact base, consisting of multi-sensor EO images, where each individual image is always provided with information products, both numeric and categorical, either categorical sub-symbolic, e.g., quantized numeric variables such as quantized biophysical parameters, say, leaf area index, or categorical symbolic, e.g., SCMs, in an array database implemented as the Rasdaman. Considered a viable alternative to traditional flat files adopted in relational databases, an array database instantiates multiple spatiotemporal data cubes in compliance with the Open Geospatial Consortium (OGC) standards, to guarantee inter-system harmonization and compatibility. In a spatiotemporal data cube the third dimension is time.

Recent developments in the mandatory QA4EO *Val* by independent means of ESA EO Level 2 SCM product automatically generated by the EO-IU subsystem as initial not sufficient pre-condition for SCBIR are reported hereafter.

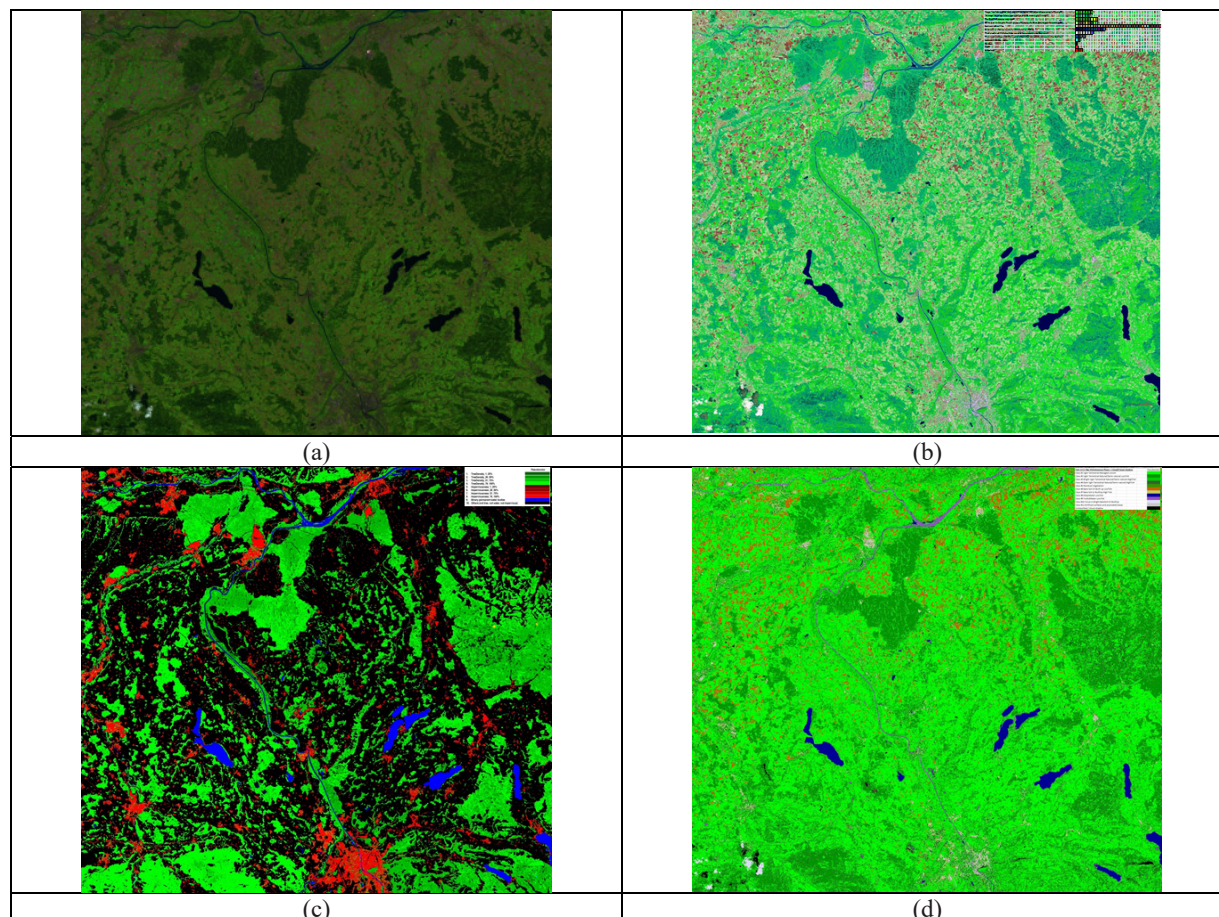


Fig. 13-1. (a) Subset of a 6-band (B, G, R, NIR, MIR1, MIR2) Sentinel-2A (S2A) image of Austria, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue). Acquired on 2015-08-13. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9. (c) European Environment Agency (EEA), GIO Land (GMES/Copernicus Initial Operations Land), Pan-European components: High Resolution Layers (HRLs), reference year 2012, 20 m spatial resolution, upscaled to 10 m resolution. Revisited map legend shown in Table 13-4. (d) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. To be compared with Table 13-2, showing an “ideal” standard 3-level 8-class LCCS-DP legend “augmented” with quality layers cloud and cloud-shadow + class Others (Unknowns), equal to  $8+2 = 10$  land cover (LC) classes + 1 non-LC class (cloud). In the implemented ESA EO Level 2 SCM generator, multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either.

To rely on a ground-truth reference map provided by an independent third party for thematic map *Val* purposes based on a wall-to-wall inter-map comparison without sampling (hence,  $\pm\delta_{OA} = 0\%$ ), where the test and reference maps feature the same spatial resolution and spatial extent, but whose legends are not the same and must be harmonized, the High Resolution Layers (HRLs), featuring 20 m spatial resolution and reference year 2012, were selected to overlap with a geographic area of interest (AOI) located on a Sentinel-2A image of Austria, 10 m resolution, acquired on 2015-08-13 and shown in Fig. 13-1. Delivered as Pan-European components by the European Environment Agency (EEA)’s GIO Land (GMES/Copernicus Initial Operations Land) initiative (EEA, 2017), these HRLs consist of: (I) three binary masks of permanent water bodies, wetlands (absent from the area of interest) and grassland, either natural or semi-natural (absent from the area of interest), (II) a numeric imperviousness index in range  $[0\%, 100\%]$ , and (III) a numeric tree percentage index in range  $[0\%, 100\%]$ . These binary and numeric reference variables, upsampled to 10 m resolution, were transformed into the ground-truth categorical map shown in Fig. 13-1(c), whose legend is depicted in Table 13.4. A dictionary-pair



binary relationship  $R: A \Rightarrow B \subseteq A \times B$  from set  $A = \text{DictionaryOfTestNames}$ , with cardinality  $|A| = a = \text{TestDictionaryCardinality} = 12$ , to set  $B = \text{DictionaryOfReferenceNames}$ , with cardinality  $|B| = b = \text{ReferenceDictionaryCardinality} = 11$ , where  $A \times B$  is the 2-fold Cartesian product (product set) between the two univariate categorical variables  $A$  and  $B$  estimated from the same geospatial population, was identified according to the 8-step guideline proposed for best practice in Chapter 7.4 (also refer to Table 7-5), based on a hybrid (combined deductive and inductive) combination of bottom-up frequentist inference with top-down prior beliefs, if any. This binary relationship  $R: A \Rightarrow B \subseteq A \times B$ , shown in Table 13-5 as color-filled cells, guides the interpretation process of the non-square BIVRTAB =  $\text{FrequencyCount}(A \times B)$ , whose minimally dependent maximally informative (mDMI) set of accuracy quantitative quality indicators ( $Q^2$ Is) consists of {overall accuracy  $OA = 99.38\% \pm 0\%$ , Categorical Variable Pair Association Index version 3 (CVPAI3)  $\in [0, 1] = 0.32$ } (refer to Chapter 7.5).












		<b>Pseudocolor</b>
<b>A11</b>	<b>1. Cultivated and Managed Terrestrial (non-aquatic) Non-vegetated Areas</b>	
<b>A12</b>	<b>2. Natural and Semi-Natural Terrestrial Vegetation</b>	
<b>A23</b>	<b>3. Cultivated Aquatic or Regularly Flooded Vegetated Areas</b>	
<b>A24</b>	<b>4. Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation</b>	
<b>B35</b>	<b>5. Artificial Surfaces and Associated Areas</b>	
<b>B36</b>	<b>6. Bare Areas</b>	
<b>B47</b>	<b>7. Artificial Waterbodies, Snow and Ice</b>	
<b>B48</b>	<b>8. Natural Waterbodies, Snow and Ice.</b>	
	<b>9. Quality layer: Cloud</b>	
	<b>10. Quality layer: Cloud-shadow</b>	
	<b>11. Others</b>	

Table 13-2, equivalent to Fig. 1-5. “Ideal” standard FAO LCCS Dichotomous Phase (DP) legend “augmented” with quality layers cloud and cloud-shadow + class Others (Unknowns), equal to  $8+2 = 10$  land cover (LC) classes + 1 non-LC class (cloud).

	<b>Pseudocolor</b>
<b>Class #1 Vgtn Terrestrial Managed Lowxtr</b>	
<b>Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxtr</b>	
<b>Class #3 Bright Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxtr</b>	
<b>Class #5 Residual Vegetation</b>	
<b>Class #6 Bare Soil Or Built-up LowTxtr</b>	
<b>Class #7 Bare Soil or BuiltUp High Txtr</b>	
<b>Class #8 DeepWater LowTxtr</b>	
<b>Class #9 TurbidWater LowTxtr</b>	
<b>Class #10 Cloud or Bright BareSoil Or BuiltUp</b>	
<b>Class #11 Artificial surfaces and associated areas</b>	
<b>Unclassified / cloud-shadow</b>	

Table 13-3, equivalent to Fig. 1-7. Implemented ESA EO Level 2 scene classification map (SCM) legend, consisting of 12 classes, approximately equivalent to a standard 2-level 4-class LCCS Dichotomous Phase (DP)-like 1st (veg/non-veg) and 2nd level (water/terrestrial). To be compared with Table 13-2, showing an “ideal” standard 3-level 8-class LCCS-DP legend “augmented” with quality layers cloud and cloud-shadow + class Others (Unknowns), equal to  $8+2 = 10$  land cover (LC) classes + 1 non-LC class (cloud).



1. TreeDensity\_1\_25%
2. TreeDensity\_26\_50%
3. TreeDensity\_51\_75%
4. TreeDensity\_76\_100%
5. Imperviousness\_1\_25%
6. Imperviousness\_26\_50%
7. Imperviousness\_51\_75%
8. Imperviousness\_76\_100%
9. Binary permanent water bodies
10. Others (not tree, not water, not impervious)

Pseudocolor

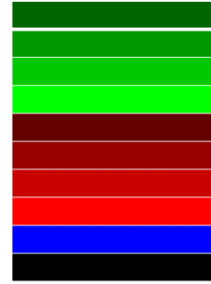


Table 13-4. European Environment Agency (EEA), GIO Land (GMES/Copernicus Initial Operations Land), Pan-European components: High Resolution Layers (HRLs), reference year 2012, 20 m spatial resolution. Revisited map legend: 4 fuzzy sets {1-25, 26-50, 51-75, 76-100} in Tree Density [0%, 100%], 4 fuzzy sets {1-25, 26-50, 51-75, 76-100} in Imperviousness Density [0%, 100%], Binary permanent water bodies.

	1. not_tree_not_water_not_impervious_of_unclassified	2. Binary permanent water bodies	3. Imperviousness_1_25%	4. Imperviousness_26_50%	5. Imperviousness_51_75%	6. Imperviousness_76_100%	7. TreeDensity_1_25%	8. TreeDensity_26_50%	9. TreeDensity_51_75%	10. TreeDensity_76_100%	11. Unclassifiable (clouds, shadow or snow)	Row sum = Test class in [0%, 100%] with respect to total samples	Tot entries per row	f(TC)
Hybrid (combined deductive and inductive) inference														
Unclassified	1.19	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.26	0.13	0.01	2.59	11	0.02425801
Class #1 Vgtn Terrestrial Managed LowTxr	35.39	0.02	0.10	0.43	0.76	0.66	0.17	0.54	1.33	2.78	0.00	43.93	10	0.04917368
Class #2 Vgtn Terrestrial Natural/Semi-natural LowTxr	5.13	0.00	0.00	0.00	0.01	0.00	0.21	0.11	0.78	4.52	0.00	11.89	6	0.39465155
Class #3 Bright Vgtn Terrestrial Natural/Semi-natural HighTxr	13.55	0.02	0.18	0.45	0.85	0.24	0.21	1.01	5.06	3.45	0.00	25.03	10	0.04917368
Class #4 Dark Vgtn Terrestrial Natural/Semi-natural HighTxr	0.38	0.03	0.00	0.02	0.01	0.00	0.02	0.19	1.84	1.27	0.00	3.67	6	0.39465155
Class #5 Residual Vegetation	1.26	0.01	0.02	0.10	0.12	0.05	0.03	0.16	0.85	0.54	0.00	3.08	10	0.04917368
Class #6 Bare Soil Or Built-up LowTxr	5.37	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	5.48	5	0.55153977
Class #7 Bare Soil Or Built-up HighTxr	0.91	0.01	0.01	0.09	0.30	0.75	0.02	0.02	0.02	0.01	0.00	2.14	9	0.09253528
Class #8 DeepWater LowTxr	0.00	0.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.93	2	0.9649297
Class #9 TurbidWater LowTxr	0.04	0.48	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.55	3	0.86377653
Class #10 Cloud or Bright BareSoil Or BuiltUp	0.09	0.00	0.00	0.00	0.01	0.09	0.00	0.00	0.01	0.05	0.00	0.25	11	0.02425801
Class #11 Artificial surfaces and associated areas	0.05	0.05	0.00	0.01	0.04	0.28	0.00	0.01	0.00	0.00	0.00	0.45	6	0.39465155
Column sum = Reference class in [0%, 100%] with respect to total samples	57.9692	1.6121	0.3552	1.4944	2.2521	2.4015	0.5089	2.1890	16.4592	14.7364	0.0219	100.00	Tot test sample	
Tot entries per column	11	5	6	8	8	8	8	8	8	7	12	100.00	Tot reference sample	
f(RC)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000		Tot RC	11
Overall accuracy =	99.3817												Tot TC	12
New Categorical Variable Pair Association Index (CVPAI) <sub>1</sub> = min(CVPAI <sub>1</sub> , CVPAI <sub>2</sub> ) =		1.0000	CVPAI <sub>2</sub> =	0.3208	min(CVPAI <sub>1</sub> , CVPAI <sub>2</sub> ) =	0.3208								

Table 13-5. Two-way contingency table (bivariate table or frequency table, BIVRTAB) between a test 12-class classification map automatically generated from a Sentinel-2A image of Austria, shown in Fig. 13-1(d), whose 12-class legend = A is shown in Table 13-3, which is input as test rows, overlapped wall-to-wall without sampling with the reference EEA HRL map shown in Fig. 13-1(c), whose 11-class legend = B is depicted in Table 13-4, which is input as reference columns. Since the two input categorical variables do not coincide and do not share the same cardinality, the two-way BIVRTAB = FrequencyCount(A × B) is a non-square frequency table featuring no main diagonal and where a binary relationship R: A ⇒ B ⊆ A × B guides the interpretation process, with cells belonging to binary relationship R depicted in color-fill. The binary relationship R: A ⇒ B was identified according to the hybrid 8-step guideline proposed for best practice in Chapter 7.4 (also refer to Table 7-5). The BIVRTAB's minimally dependent maximally informative (mDMI) set of accuracy Q<sup>2</sup>I values comprises {overall accuracy OA ∈ [0, 1] = 0.9938 ± 0%, Categorical Variable Pair Association Index version 3 (CVPSAI3) ∈ [0, 1] = 0.32}.

In agreement with the QA4EO Val requirements, a third-party reference EEA HRL map, shown in Fig. 13-1(c) and whose 11-class legend = B is depicted in Table 13-4, was overlapped wall-to-wall for Val purposes with the test ESA EO Level 2 SCM product, shown in Fig. 13-1(d) and whose 12-class legend = A is depicted in Table 13-3. The resulting overlapping area matrix (OAMTRX) is shown in Table 13-5, whose mDMI set of accuracy Q<sup>2</sup>I values is {OA ∈ [0, 1] = 0.9938 ± 0%, CVPSAI3 ∈ [0, 1] = 0.32}, where OA and CVPSAI3 are statistically independent (refer to Chapter 7.4). This OA value was computed on “correct” entry-pairs in a binary relationship, R: A ⇒ B ⊆ A × B, identified by the hybrid 8-step guideline proposed for best practice in Chapter 7.4 (also refer to Table 7-5). Color-filled in Table 13-5, these entry-pairs belonging to R: A ⇒ B are many, which contributes to obtain a “high” OA value. On the other hand, this semantic ambiguity in the identification of a single reference class starting from test classes causes the CVPSAI3 value to score “low”. It means the two codebooks A and B are very difficult to untangle, i.e., it is impossible to identify single words in reference codebook B as OR-combinations of codewords of test codebook A without replacement.

To provide a qualitative support to the aforementioned claims the following Fig. 13-2 to Fig. 13-4 are shown for photointerpretation purposes.



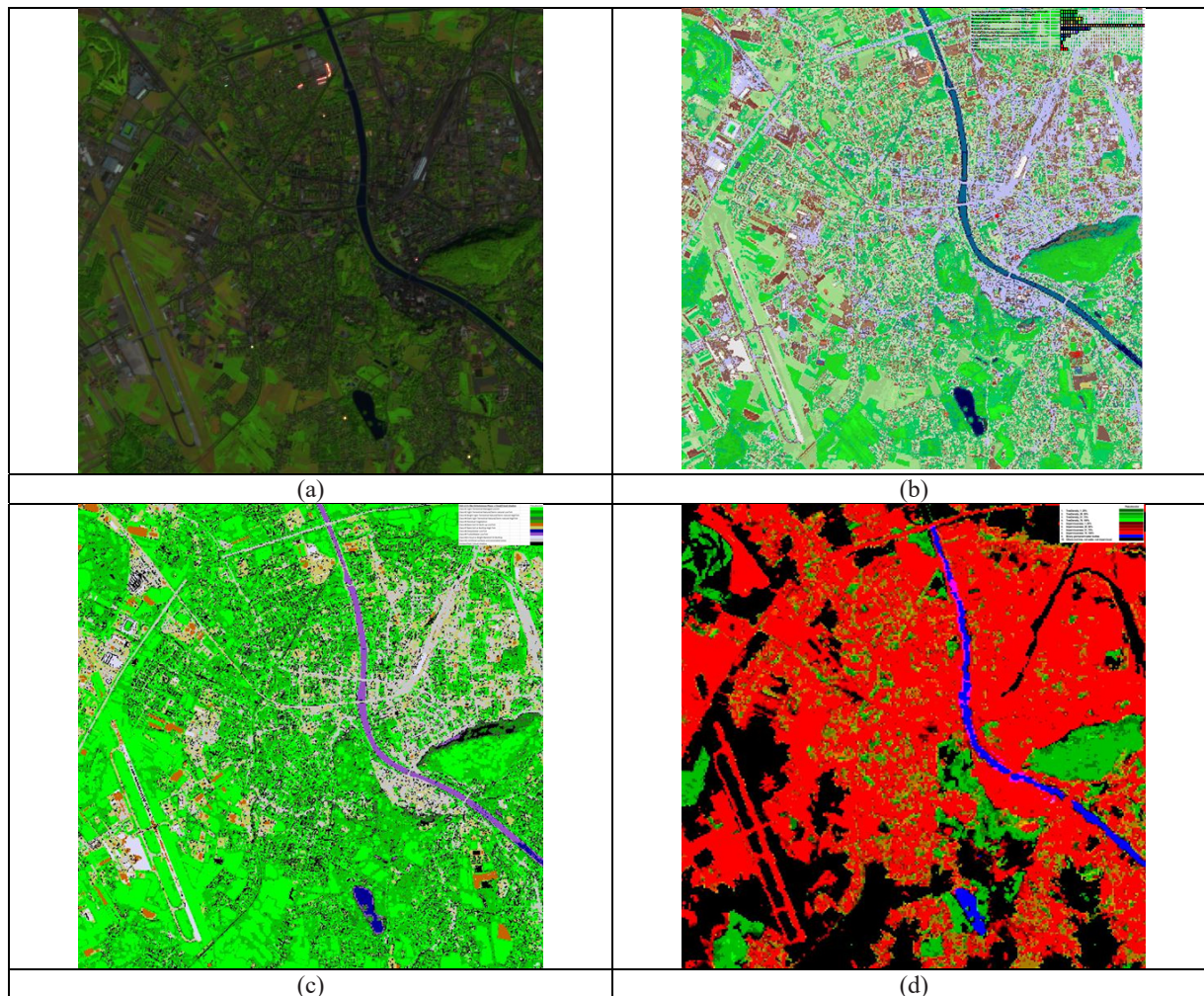


Fig. 13-2. (a) Zoom of a urban area in a 6-band (B, G, R, NIR, MIR1, MIR2) Sentinel-2A (S2A) image of Austria, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue). Acquired on 2015-08-13. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) European Environment Agency (EEA), GIO Land (GMES/Copernicus Initial Operations Land), Pan-European components: High Resolution Layers (HRLs), reference year 2012, 20 m spatial resolution, upscaled to 10 m resolution. Revisited map legend shown in Table 13-4.

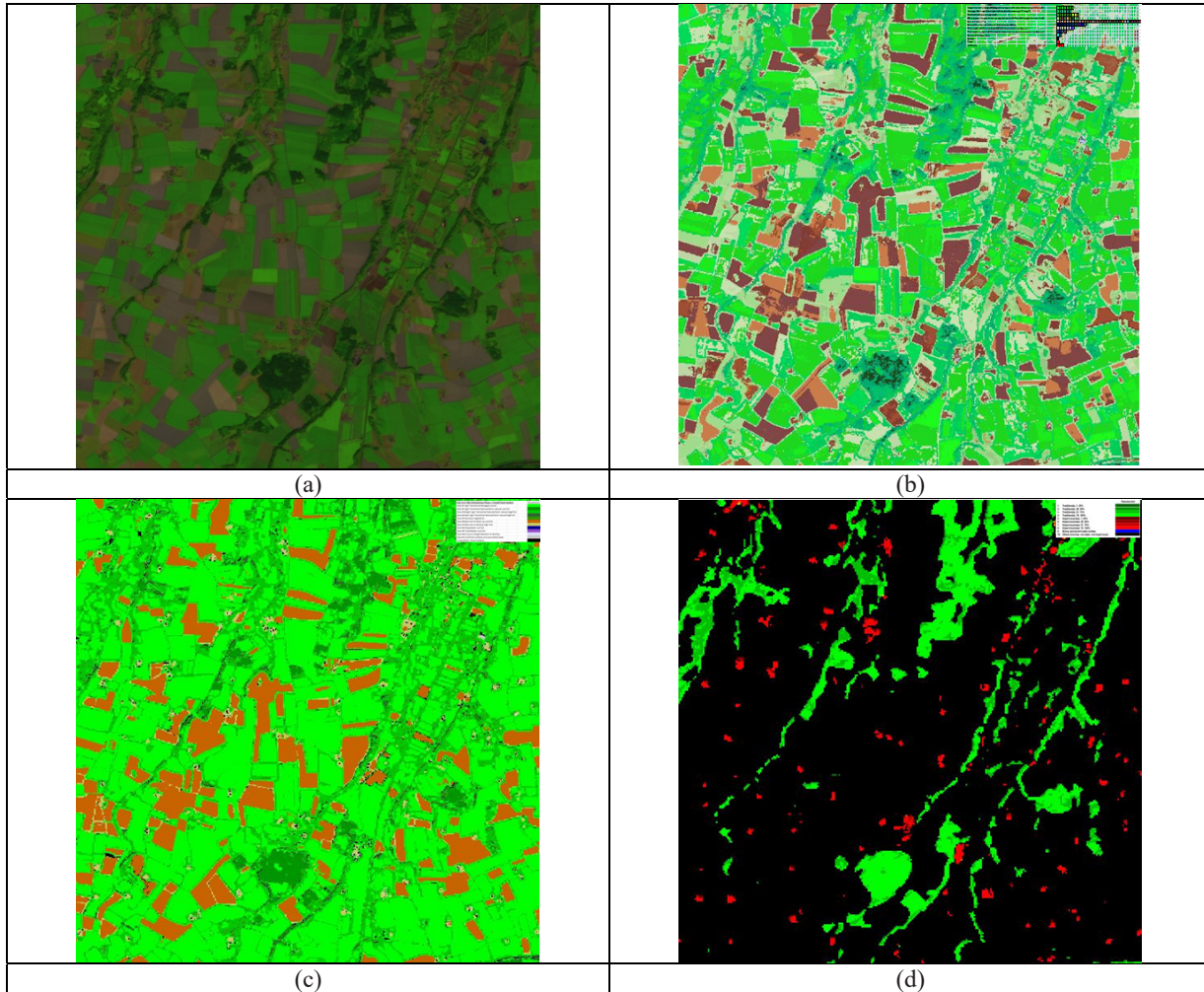


Fig. 13-3. (a) Zoom of an agricultural area in a 6-band (B, G, R, NIR, MIR1, MIR2) Sentinel-2A (S2A) image of Austria, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue). Acquired on 2015-08-13. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) European Environment Agency (EEA), GIO Land (GMES/Copernicus Initial Operations Land), Pan-European components: High Resolution Layers (HRLs), reference year 2012, 20 m spatial resolution, upscaled to 10 m resolution. Revisited map legend shown in Table 13-4.



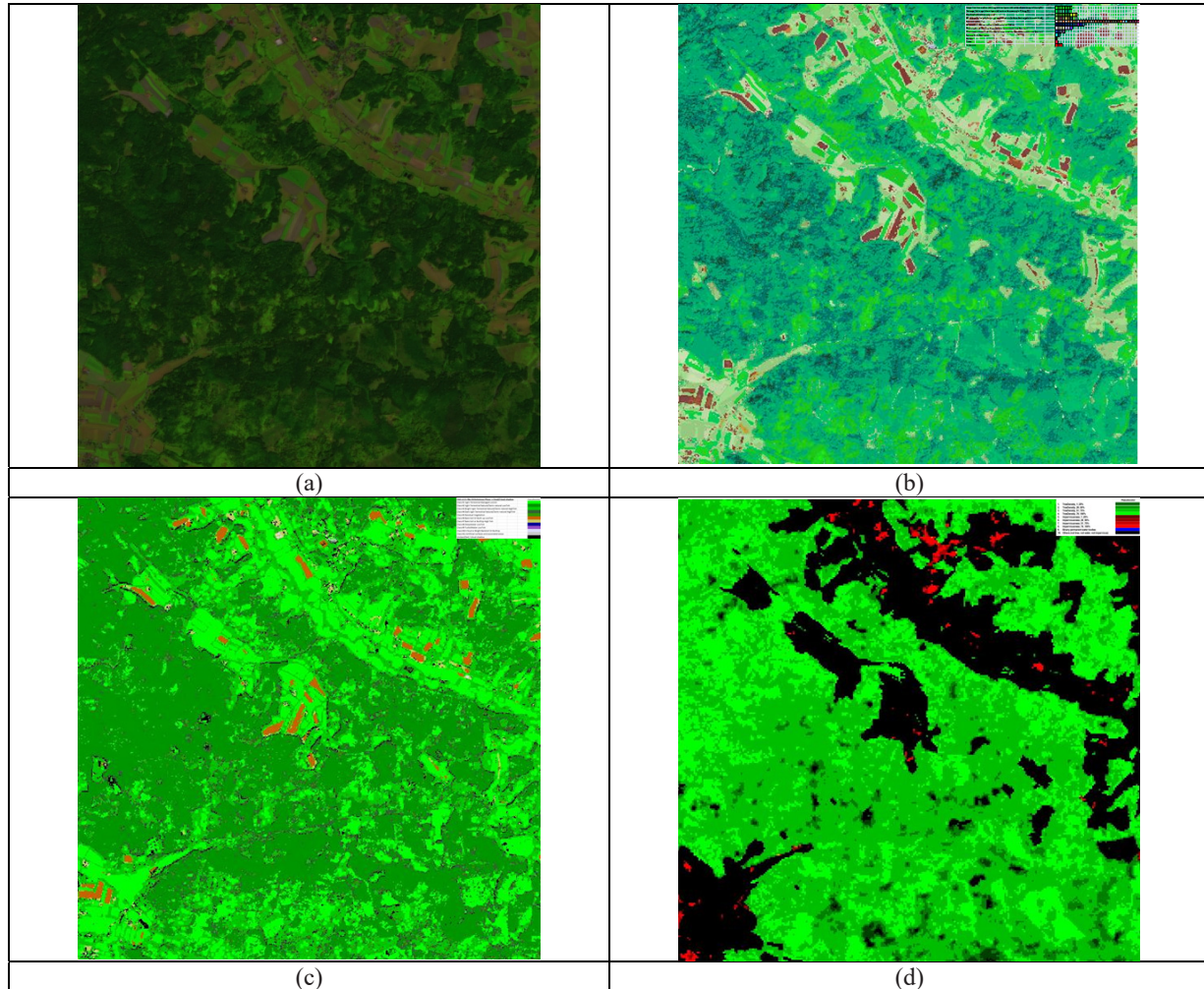


Fig. 13-4. (a) Zoom of a forest area in a 6-band (B, G, R, NIR, MIR1, MIR2) Sentinel-2A (S2A) image of Austria, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = MIR, G = NIR, B = Visible Blue). Acquired on 2015-08-13. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) L-SIAM color map at fine color granularity, consisting of 96 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) European Environment Agency (EEA), GIO Land (GMES/Copernicus Initial Operations Land), Pan-European components: High Resolution Layers (HRLs), reference year 2012, 20 m spatial resolution, upscaled to 10 m resolution. Revisited map legend shown in Table 13-4.

To test its robustness to changes in input data and its scalability to different imaging sensor specifications in spectral and/or spatial resolution, the same ESA EO Level 2 SCM generator implementation was applied to multi-source MS images, such as the 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values, 10 m resolution, acquired on 2004-13-06, shown in Fig. 13-5. For an intuitive qualitative assessment of this ESA EO Level 2 SCM product accuracy via photointerpretation, zoom-in areas are shown in Fig. 13-6 to Fig. 13-8.



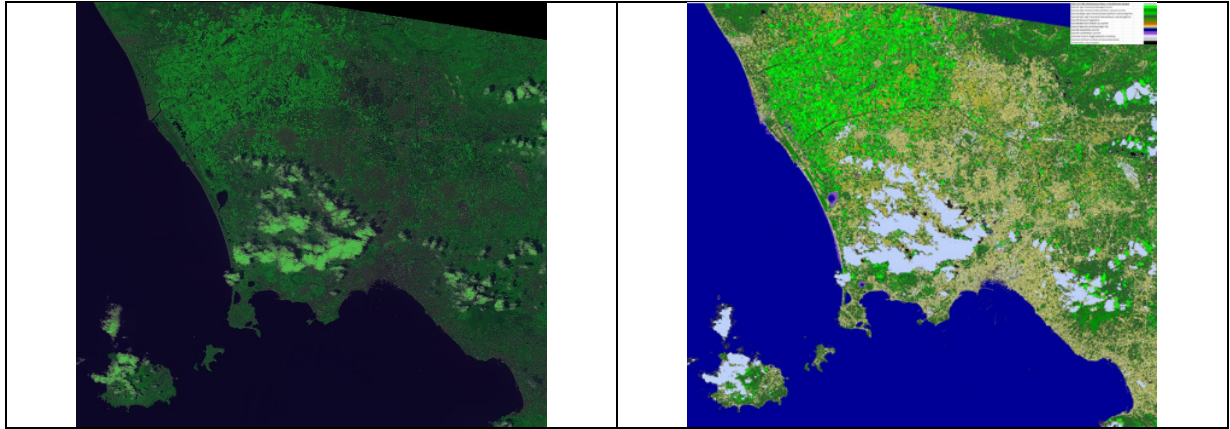


Fig. 13-5. Left: 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = Visible Red, G = NIR, B = Visible Blue), 10 m resolution. Acquired on 2004-13-06. No histogram stretching for visualization purposes. Right: Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. In the implemented ESA EO Level 2 SCM generator, multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either.

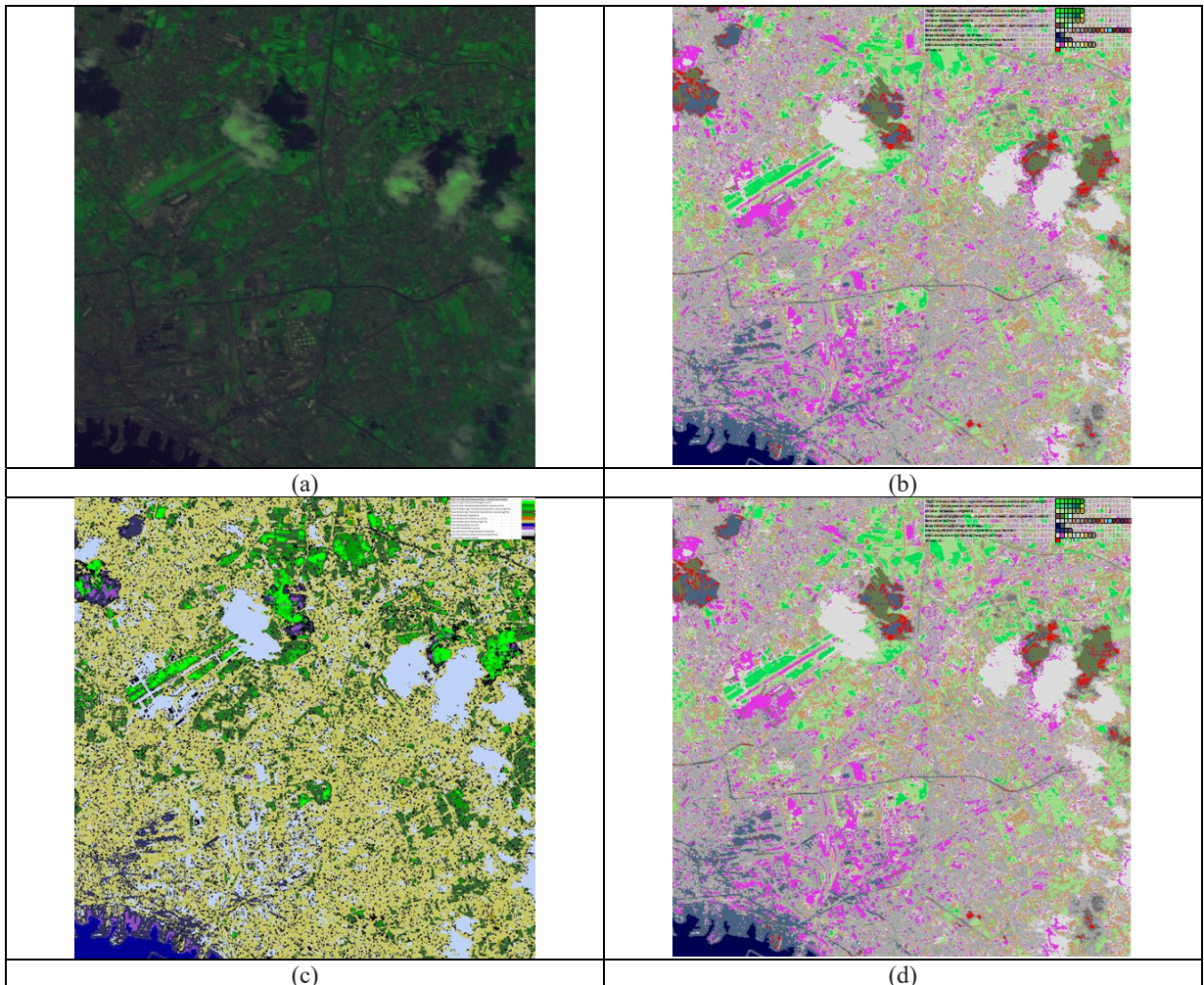






Fig. 13-6. (a) Zoom of a urban area in a 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = Visible Red, G = NIR, B = Visible Blue). Acquired on 2004-13-06. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) Q-SIAM color map at fine color granularity, consisting of 61 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) Same as Fig. (b).

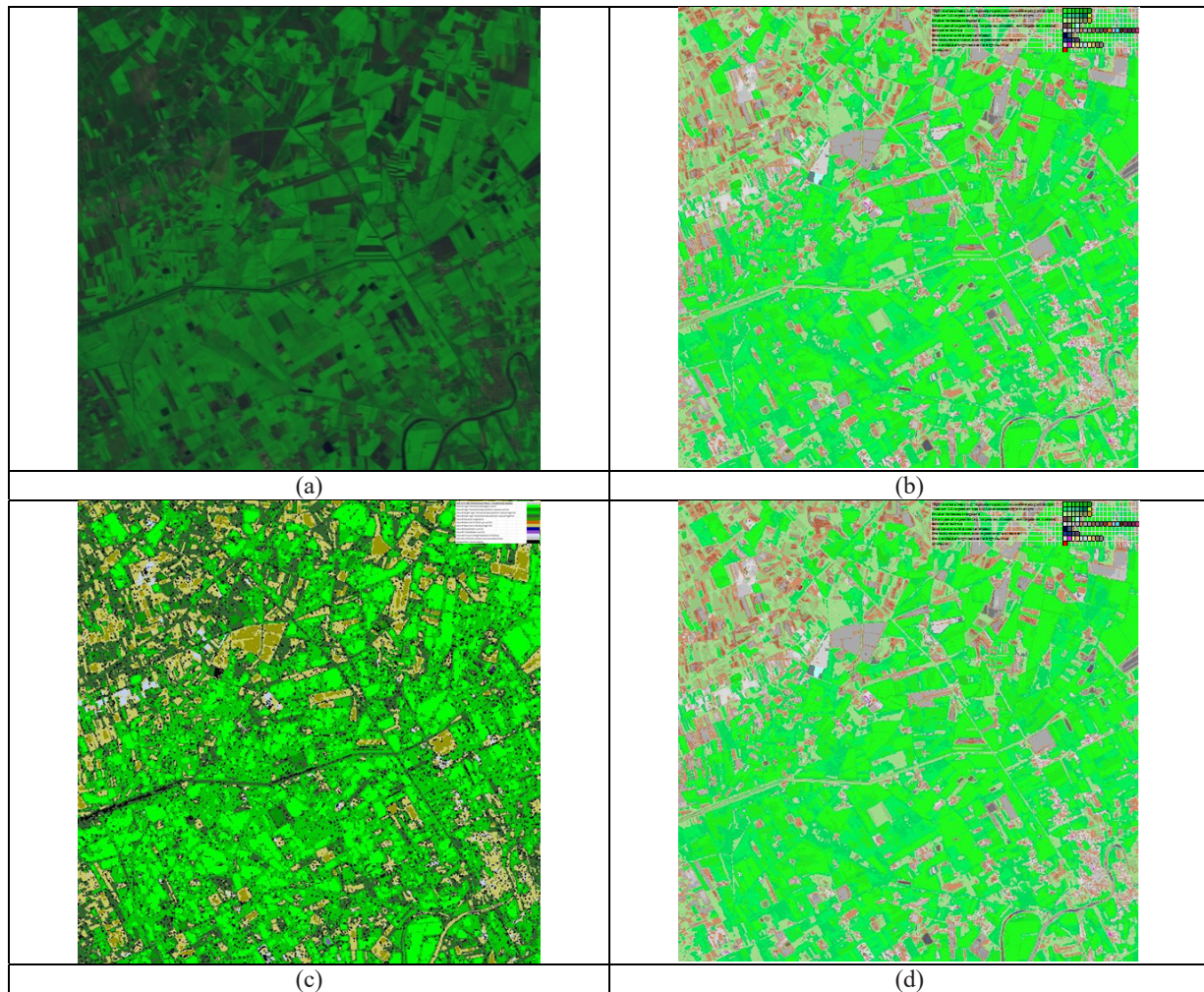


Fig. 13-7. (a) Zoom of an agricultural area in a 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = Visible Red, G = NIR, B = Visible Blue). Acquired on 2004-13-06. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) Q-SIAM color map at fine color granularity, consisting of 61 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) Same as Fig. (b).



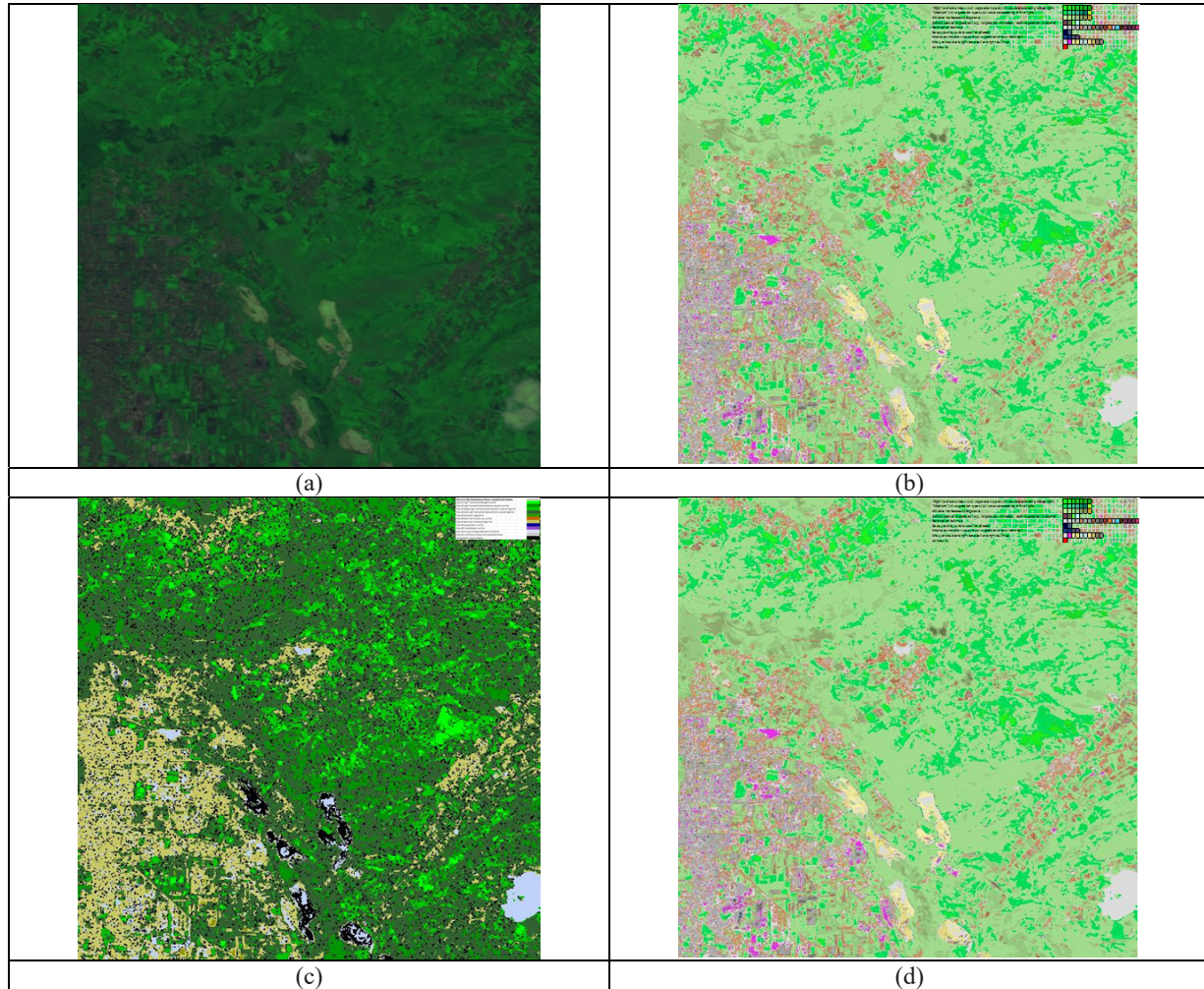


Fig. 13-8. (a) Zoom of a forest area in a 4-band (B, G, R, NIR) ALOS AVNIR-2 image of Campania, Italy, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values and depicted in false colors (R = Visible Red, G = NIR, B = Visible Blue). Acquired on 2004-13-06. Spatial resolution: 10 m. No histogram stretching is applied for visualization purposes. (b) Q-SIAM color map at fine color granularity, consisting of 61 spectral categories depicted in pseudo colors, shown in Fig. 3.10-9 of Chapter 3.10. (c) Automatically generated ESA EO Level 2 SCM whose legend is shown in Table 13-3. Multiple sources of visual evidence are the SIAM and RGBIAM color names (refer to Chapter 3.10) and a 3-scale texture binary profile in range  $\{0, 7\}$  automatically estimated from image-contours (refer to Chapter 3.20.4). Visual features not yet employed as input by the implemented ESA EO Level 2 SCM classifier were per-segment shape and size properties and inter-segment spatial relationships, either topological or non-topological. In addition, no cloud and cloud-shadow masking was applied either. (d) Same as Fig. (b).

In summary, the mDMI set of OP-Q<sup>2</sup>I values reported for the automated near real-time ESA EO Level 2 product generator, still implemented in a prototypical version where local shape, spatial topological and spatial non-topological information components, together with cloud and cloud-shadow quality layers, are not employed as input information components yet, can be summarized as shown in Table 13-6. These OP-Q<sup>2</sup>Is comply with the primary EO-IU subsystem requirements specification in user-speak proposed in Chapter 4.1, with the sole exception of the semantic information level, inferior to the target 3-level 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow, see Table 13-2. This semantic information gap is expected to be filled in as soon as the missing visual information primitives, already made available as described in Chapter 3 (Technical report 1), will be provided as input to the EO-IU subsystem prototype discussed in Chapter 4 (Manuscript 1).



Legend of fuzzy sets of a quantitative variable.

LOW
MEDIUM
HIGH

Computer vision (CV) Process (Pracs) and Outcome (Otcn) $Q^2Is \pm \delta \subseteq QA4EO VaI$	ESA EO Level 2 product generator
<b>Degree of automation (Pracs):</b> (a) inversely related to the number, physical meaning and range of variation of user-defined parameters, (b) inversely related to the collection of the required training data set, if any.	HIGH
<b>Effectiveness (Otcn), in agreement with human visual perception, i.e., <math>CV \supset human\ vision</math>, where human visual perception is a lower bound of CV.</b> a) Color constancy or radiometric calibration (when radiometric calibration metadata are available) b) 2D image analysis/ Retinotopic visual information representation/ Topology-preserving visual feature mapping / Spatial topological information extraction. > Necessary not sufficient condition: panchromatic vision performs nearly as well as chromatic vision. c) Pre-attentive image contour detection/ image segmentation quality, consistent with the Mach bands illusion in ramp-edge detection: spatial Qis (SQIs), provided with a degree of uncertainty in measurement $\pm\delta$ . d) High-level vision (classification). (a) thematic Qis (TQIs) and (b) spatial Qis (SQIs), provided with a degree of uncertainty in measurement $\pm\delta$ .	a) HIGH b) YES > HIGH c) HIGH d) HIGH
<b>Semantic information level (Pracs)</b>	MEDIUM/LOW (LEVEL 2 SCM legend)
<b>Efficiency (Pracs):</b> (a) computation complexity in image size: Polynomial (P), P and linear (L), non-P (NP), and (b) run-time memory occupation.	(a) HIGH (L), (b) HIGH
<b>Robustness to changes in input image (Pracs)</b> , e.g., large spatial extent data mapping (no toy problems).	HIGH
<b>Robustness to changes in input parameters (Pracs)</b> , e.g., sensitivity analysis.	HIGH (unsurpassed, no free-paramtr)
<b>Scalability to changes in the sensor's specifications or user's needs (Pracs)</b> , e.g., (a) pncchrmtc, (b) RGB, true- or false-color, (c) multi-spectral (MS), (d) super-spectral (SS), (e) hyper-spectral (HS).	HIGH, (a) YES, (b) YES, (c) YES, (d) YES, (e) YES.
<b>(Inverse of) Timeliness (Otcn)</b> , from data acquisition to high-level product generation, increases with manpower and computing power.	HIGH
<b>(Inverse of) Costs (Otcn)</b> , increasing with (a) manpower and (b) computing power.	(a) HIGH, (b) HIGH

Table 13-6, equivalent to Table 4-4. Outcome and process (OP) quantitative quality indicators (OP-Q<sup>2</sup>Is) of the proposed EO-IU subsystem design and implementation for systematic ESA EO Level 2 product generation.

Summarized in Table 13-6, OP-Q<sup>2</sup>I values collected from the implemented ESA EO Level 2 product generator appear superior to those of traditional EO-IU systems and SCBIR system prototypes built upon an inductive feedforward inference system for 1D image classification, such as a support vector machine, known to be inherently semi-automatic, training data-dependent and insensitive to spatial topological information which typically dominates color information in vision. For example, the implemented ESA EO Level 2 product generator is completely different from the Sentinel-2 software Toolbox developed and distributed by ESA, known as Sentinel 2 (atmospheric) Correction (SEN2COR) Prototype Processor to be run on user side. The latter is summarized as follows.

- (i) Sentinel-2 imaging sensor-specific. Input data sets: one Sentinel-2 image radiometrically calibrated into TOARF values and provided with its radiometric calibration metadata file (featuring acquisition time, sun azimuth and zenith position, sensor azimuth and zenith position, etc.), one digital terrain model (required for topographic correction).
- (ii) SEN2COR flow chart equal to that of the Atmospheric/Topographic Correction for Satellite Imagery (ATCOR) commercial software product, refer to Chapter 7.3 and see Fig. 7-11.
- (iii) ESA EO Level 2 SCM legend shown in Fig. 2-8 of Chapter 2 (Introduction).
- (iv) ESA EO Level 2 SCM generator is a static pixel-based spectral decision tree, i.e., it is a context-insensitive 1D image analysis approach, where spatial topological and non-topological information components are totally ignored, although they typically dominate color information in vision, refer to Chapter 2 (Introduction). It is automated, requiring no human-machine interaction, with computational complexity increasing linearly with image size.
- (v) Hybrid (combined physical model-based and statistical model-based) context-sensitive inference system for cloud and cloud-shadow detection.

On the contrary, the proposed ESA EO Level 2 product generator is summarized as follows.

- (i) Multi-sensor MS imagery. Input data sets: any MS image, either radiometrically calibrated into TOARF reflectance values and provided with its radiometric calibration metadata file or uncalibrated, but provided with radiometric calibration metadata file (featuring acquisition time, sun azimuth and zenith position, sensor azimuth and zenith position, etc.), one digital terrain model (required for topographic correction).
- (ii) ESA EO Level 2 product generator flow chart shown in Fig. 1-3 of Chapter 1 (Doctoral research objectives), where the MS image calibration into surface reflectance (SURF) values corrected for atmospheric, topographic and



- adjacency effects alternates with a SIAM-based classification step.
- (iii) ESA EO Level 2 SCM legend equal to the standard 3-level 8-class LCCS-DP taxonomy augmented with quality layers cloud and cloud-shadow, see Table 13-2.
  - (iv) ESA EO Level 2 SCM generator is a topology-preserving hybrid feedback CV system based on a convergence of evidence approach, where visual information primitives are color, local shape, texture and inter-object spatial topological and non-topological relationships, refer to Chapter 3 (Technical report 1). It is automated, requiring no human-machine interaction, with computational complexity increasing linearly with image size.
  - (v) Automated hybrid combined physical model-based and statistical model-based context-sensitive inference system for cloud and cloud-shadow detection based on a convergence of evidence approach, refer to Chapter 12 (Technical report 2).

***Chapter 6 - Manuscript 3 (published): Automated Hierarchical 2D and 3D Object-Based Recognition and Reconstruction of ISO Containers in a Harbor Scene***

Chapter 6 (Manuscript 3) consists of a published manuscript derived from a conference paper, titled “Geospatial 2D AND 3D object-based classification and 3D reconstruction of ISO-containers depicted in a LiDAR dataset and aerial imagery of a harbor”, presented in the IGARSS 2015 conference, Milan, Italy, 27-31 July 2015, which ranked 2nd in the IEEE GRSS Data Fusion Contest 2015.

Chapter 6 (Manuscript 3) presents and discusses an application example where the original CV algorithms presented in Chapter 3 (Computational models of human vision) and adopted by the closed-loop EO-IU4SQ system, proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), are applied to the experimental framework of the IEEE GRSS Data Fusion Contest 2015.

***Chapter 7 - Manuscript 4 (made available in the public archive arXiv:1701.01930): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 1: Theory***

***Chapter 8 - Manuscript 5 (made available in the public archive arXiv:1701.01932): Stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for Earth observation Level 2 product generation– Part 2: Validation***

Among the original CV algorithms proposed in Chapter 3 (Technical report 1) and adopted by an innovative EO-IU4SQ system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), there is an original pair of expert systems (prior knowledge-based decision trees) for color naming in a calibrated multi-spectral (MS) reflectance space or in an uncalibrated RGB color space, either true- or false-color. Color naming transforms a numeric variable (color value, colorimetric sensation) into a categorical variable, specifically, into color names belonging to a pre-defined dictionary of color names, equivalent to a latent/hidden variable and eligible for use in symbolic human reasoning to link sub-symbolic sensations (observables) in the physical world to symbolic and stable percepts in the modeled world. Chapter 7 (Manuscript 4) presents and discusses the non-trivial multi-disciplinary background of color naming, ranging from cognitive science to artificial intelligence (AI) and machine learning-from-data.

Provided with a relevant survey value, Chapter 7 (Manuscript 4) features several degrees of novelty. First, to cope with dictionaries of MS color names and land cover class names that do not coincide and must be harmonized, an original hybrid (combined deductive and inductive) guideline is proposed to identify a categorical variable-pair (binary) relationship. Second, an original quantitative measure of categorical variable-pair association index, CVPAI, is proposed, given a categorical variable-pair relationship.

Among the original expert systems for color naming presented in Chapter 3 (Technical report 1) and adopted by an EO-IU4SQ system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the Satellite Image Automatic Mapper™ (SIAM™) lightweight computer program was designed and implemented to provide multi-spectral (MS) reflectance space hyperpolyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time (with a computational complexity increasing linearly with image size). While the Part 1 of this paper (Chapter 7, Manuscript 4) provided the multidisciplinary background of color naming, in the Part 2 of this paper (Chapter 8, Manuscript 5) an off-the-shelf SIAM lightweight computer program was submitted to an intergovernmental Group on Earth Observations (GEO)’s Stage 4 validation, by independent means on a large-scale EO image time-series, for systematic EO Level 2 product generation.





Chapter 8 (Manuscript 5) features several degrees of novelty. First, it presents a novel protocol suitable for wall-to-wall thematic map quality assessment without sampling, where the test and reference thematic maps share the same spatial extent and spatial resolution, but whose map legends can differ in agreement with Chapter 8 (Manuscript 5). Second, it provides several instantiations of the categorical variable-pair relationship and CVP AI values proposed in Chapter 7 (Manuscript 4). Last but not least, it provides a non-trivial solution to the important question of general interest: if the “ultimate” accuracy of reference map A is validated in comparison with “absolute” ground-truth while the accuracy of test map B is assessed in relative terms of agreement in comparison with reference map A, what is the inferred “ultimate” accuracy of test map B in comparison with “absolute” ground-truth?

***Chapter 9 - Manuscript 6 (made available in the public archive arXiv:1701.01940): Automated Near Real-Time Detection and Quality Assessment of Superpixels in Uncalibrated True- or False-Color RGB Images***

Among the original pair of expert systems for color naming presented in Chapter 3 (Technical report 1) and adopted by an EO-IU4SQ system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the RGB Image Automatic Mapper™ (RGBIAM™) lightweight computer program was designed and implemented to accomplish true- or false-color RGB cube polyhedralization, superpixel detection and vector quantization (VQ) quality assurance in operating mode, specifically, automatically (without human-machine interaction) and in near real-time, i.e., in linear time complexity monotonically increasing with image size. In Chapter 9 (Manuscript 6) the RGBIAM lightweight computer program pipeline, including an original statistical model-based self-organizing color constancy algorithm required when an RGB image is not radiometrically calibrated, is presented and discussed in detail.

***Chapter 10 - Manuscript 7 (made available in the public archive arXiv:1701.01941): Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis***

In the convergence-of-evidence approach to CV proposed in Chapter 3 (Technical report 1), planar shape indexes are a source of spatial non-topological information in the (2D) image-domain, specifically, they are a source of spatial unit  $x$ -specific geometric information, where spatial unit  $x$  is either (0D) pixel, (1D) line or (2D) polygon. In the present Chapter 10 (Manuscript 7) an original minimally dependent and maximally informative (mDMI) set of planar shape indexes is presented and discussed to be considered eligible for use in the EO-IU4SQ system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2). The original contribution of Chapter 10 is twofold.

First, the proposed general-purpose dictionary of seven off-the-shelf 2D geometric descriptors in the spatial domain (plus area and orientation), specifically, *CnvxtyAndNoHole*, *FuzzyRuleBsdRctnglry*, *RndnssAndNoHole*, *MltSclStrghtnsOfBndrs*, *DMPmltSclChrtrstc*, *ElngrdnssAndNoHole* and *CombndSmplCnctvty*, provided with an intuitive physical meaning and submitted to a known *Val* policy for quantitative quality assurance ( $Q^2A$ ), can be integrated into software libraries of efficient and reliable computational geometry algorithms, such as CGAL, where “simple” 2D shape descriptors are absent. This compact feature set of certified quality is alternative to the large number and variety of 2D shape functions implemented in existing commercial or open source software libraries, such as eCognition’s, OpenCV and ENVI’s, whose multi-objective geometric feature representation and description criteria remain unknown for *Val* purposes. At the levels of understanding of system design and knowledge/information representation, three of the seven proposed geometric features, specifically, *MltSclStrghtnsOfBndrs*, *DMPmltSclChrtrstc* and *CombndSmplCnctvty* are totally new, i.e., they are absent from the three aforementioned commercial or open source software libraries and from the reference works by Nagao & Matsuyama and Shackelford & Davis. At the levels of understanding of algorithms and implementation, all the proposed geometric descriptors feature several degrees of novelty, introduced to optimize their quantitative quality index ( $Q^2I$ ) values in comparison with alternative solutions.

The second original contribution of Chapter 10 of potential interest to a wide scientific audience is the proposed hierarchical *Val* strategy for  $Q^2A$  of a set of quantitative random variables whose dependence (causality) must be minimized. To guarantee that numeric variables are not dependent, i.e., to avoid cause-effect relationships between numeric variable pairs, the traditional Pearson’s cross-correlation test is omitted because it is well known that, first, cross-correlation is sensitive to linear relationships exclusively and, second, that correlation does not imply causation. A Pearson’s chi-square test of statistical independence is applied instead, whose inputs are two categorical variables. Hence, each numeric variable must be transformed into a categorical variable whose bins (quantization levels) are equiprobable and whose cardinality (number of bins) is appropriate according to a heuristic statistical criterion. In search for inter-variable causality, if a feature-pair passes the Pearson’s chi-square test of statistical independence, it can be hierarchically submitted to a



Spearman's rank cross-correlation coefficient, showing whether two ranked random variables are monotonically increasing or decreasing, independent of linear relationships. Only if both tests are passed in comparison with any other feature a numeric variable is eligible for consideration in the mDMI set of features.

***Chapter 11 - Manuscript 7 (made available in the public archive arXiv:1701.01941): Multi-Objective Software Suite of Two-Dimensional Shape Descriptors for Object-Based Image Analysis***

To date no “universal” perceptual visual quality metric exists between a reference image and a test image. A special case of this perceptual image-pair comparison problem is when the test image is a MS image generated by panchromatic sharpening from a coarse-spatial resolution MS image and a fine-spatial resolution panchromatic image. In the present Chapter 11 (Manuscript 8) an original outcome and process quality assessment protocol for MS image panchromatic sharpening is proposed to comply with the QA4EO *Cal/Val* requirements. In this application framework, reference images are a coarse-spatial resolution MS image and a fine-spatial resolution panchromatic image, while the test image is a fused panchromatic-sharpened MS image at fine-spatial resolution. Typically, no process quality assessment is involved with the comparison of MS pan-sharpened images as outcome. Alternative to a traditional normalization of quality indexes adopted before index comparison and combination, one important strategy adopted by the proposed quantitative quality assessment protocol is that, before comparing and combining multiple quality indexes estimated from the same population, each individual quality index is standardized through the population to feature zero mean and unit variance, while its range of change remains unbounded above and below.

As a future development of the present Chapter 11 (Manuscript 8), an innovative perceptual image-pair quality/similarity/ dissimilarity index/ metric was proposed in Chapter 3.13 (Technical report 1).

***Chapter 12 - Technical report 2 (made available in the public archive arXiv: 1701.04256): Automatic Spatial Context-Sensitive Cloud/Cloud-Shadow Detection in Multi-Source Multi-Spectral Earth Observation Images – AutoCloud+***

Between the original pair of expert systems for color naming presented in Chapter 3 (Technical report 1) and adopted by an EO-IU4SQ system proposed in Chapter 4 (Manuscript 1) and Chapter 5 (Manuscript 2), the off-the-shelf SIAM™ lightweight computer program was submitted to a Stage 4 validation for systematic EO Level 2 product generation, refer to Chapter 7 (Manuscript 4) and Chapter 8 (Manuscript 5). In Chapter 9 (Manuscript 6) the off-the-shelf RGBIAM™ lightweight computer program for true- or false-color RGB cube partitioning into a dictionary of color names was discussed in detail. By definition, the ESA EO Level 2 product comprises a MS image corrected for geometric, atmospheric, topographic and adjacency effects, stacked with its data-derived general-purpose, user- and application-independent SCM, whose legend includes quality layers such as cloud and cloud-shadow.

Provided with a relevant survey value, Chapter 12 (Technical report 2) reviews existing cloud and cloud-shadow detectors and proposes a novel hybrid (combined deductive and inductive) EO-IU system design (architecture) suitable for automatic spatial context-sensitive cloud/cloud-shadow detection in multi-source MS imagery, where input information sources include the SIAM and RGBIAM color maps automatically generated from a single-date MS image.

Collected by D. Tiede (Z-GIS, Univ. of Salzburg) by the end of March 2017, recent results provided by the first prototypical implementation of the proposed hybrid EO-IU system architecture for automatic spatial context-sensitive cloud/cloud-shadow detection in multi-source MS imagery, where input information sources include the SIAM and RGBIAM color maps, appear extremely encouraging as shown in Fig. 13-9.

***Future developments***

In accordance with a “kaizen” paradigm of continuous improvement focused on identification, reduction and elimination of suboptimal processes, future works aiming at the development of an EO-IU4SQ system in operating mode, whose mDMI set of OP-Q<sup>2</sup>Is is required to score “high” in EO big data analytics and SCBIR tasks, will consider the EO-IU and EO-SQ subsystem prototypes, implemented in this doctoral work as a proof-of-concept, eligible for continuous optimization, modification or replacement at the four hierarchical levels of system understanding, specifically, system architecture, knowledge/information representation, algorithm and implementation.

In this doctoral dissertation, Chapter 3.22 provided a detailed list of future developments regarding the low-level CV algorithms whose output products, specifically, a raw primal sketch for image-contour detection, keypoint detection and

image segmentation and a full primal sketch for texture segmentation, are adopted by the proposed EO-IU subsystem prototype.

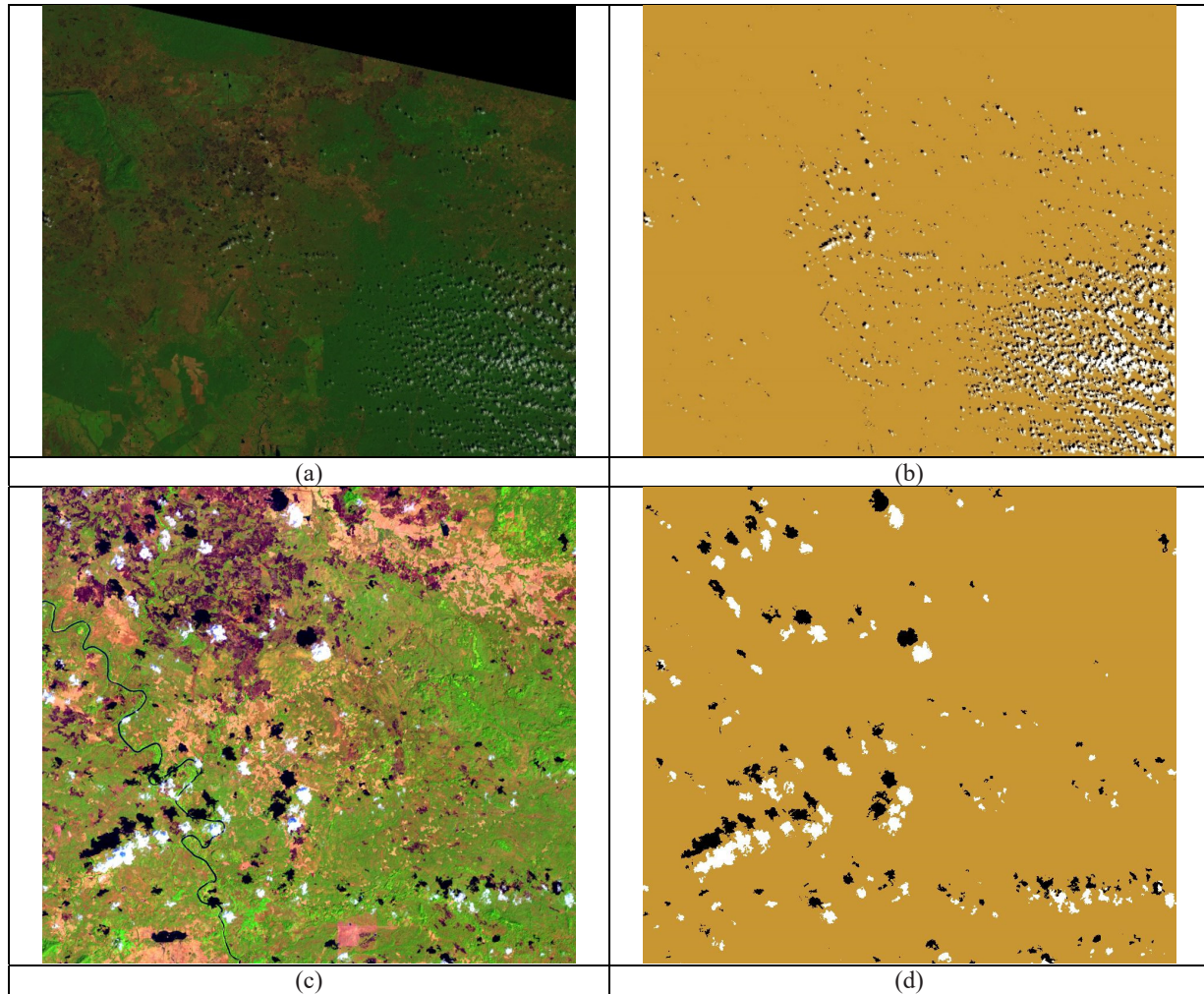


Fig. 13-9. Automatic hybrid (combined physical model-based and statistical model-based) spatial context-sensitive cloud/cloud-shadow detection (courtesy of D. Tiede, Z-GIS, Univ. of Salzburg). (a) Subset of a Landsat-8 OLI image of Cambodia (LC81260512017036LGN00, radiometrically calibrated into top-of-atmosphere reflectance (TOARF) values, depicted in false colors (R = MIR, G = NIR, B = Visible Blue), 30 m resolution, acquisition date: 03-13-2017. No histogram stretching is applied for visualization purposes. (b) Cloud/cloud-shadow thematic map. Legend in pseudolocors: White = cloud, Black = cloud-shadow, Brown = Otherwise. (c) Zoom-in of image (a), with ENVI standard histogram stretching applied for visualization purposes. (d) Zoom-in of thematic map (b).

High-level CV algorithms adopted by the EO-IU subsystem for single-date and multi-temporal EO image classification tasks, starting from systematic ESA EO Level 2 SCM product generation, will: (i) include local shape indexes and inter-object spatial relationships, either topological or non-topological, in addition to color names and texture categories adopted as input information sources to date, and (ii) benefit from future developments of and integration with an automated hybrid feedback cloud/cloud-shadow detector.

Planned future developments of the EO-SQ subsystem prototype will regard the semantic network formalism required to graphically represent the world model, to be augmented and integrated with an algebra capable of describing spatiotemporal data types and operations in a language-independent and formal way.

Finally, in agreement with the QA4EO *Val* requirements, an EO-IU4SQ system Stage 4 validation will be scheduled and pursued by independent means over multiple locations and time periods representing global conditions.