TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI INGEGNERIA ELETTRICA
E TECNOLOGIE DELL' INFORMAZIONE

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

IMAGE SEGMENTATION IN A REMOTE SENSING PERSPECTIVE

GIUSEPPE MASI

Il Coordinatore del Corso di Dottorato	Il Tutore
Ch.mo Prof. Daniele RICCIO	Ch.mo Prof Giuseppe SCARPA

A. A. 2015–2016

Contents

List of Figures

Introduction

Image segmentation in general is defined as a process of partitioning an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous [99]. Efficient image segmentation is one of the most critical tasks in automatic image processing [102, 69, 99, 150, 31] and image segmentation has been interpreted differently for different applications. For example, in machine vision applications, it is viewed as a bridge between low level and high level vision subsystems, in medical imaging as a tool to delineate anatomical structure and other regions of interest whose a priori knowledge is generally available [103] and in statistical analysis, it is posed as a stochastic estimation problem, with assumed prior distributions on image structure, which is widely used in remote sensing [78, 41]. In remote sensing, it is often viewed as an aid to landscape change detection and land use/cover classification. Aforementioned examples state that image segmentation is present in every kind of image analysis. This constitutes a plethora of literature on the image segmentation.

This thesis deals with the problem of image segmentation in the context of remote sensing, where, to give a leading example, one of the most common goals is a pixel-wise labeling in predefined thematic classes which can differ from one application to another. Obviously remote sensing is much more than classification and there are many other applications where segmentation can play a relevant role. From the methodolocical perspective it has also to be remarked that segmentation is a data-dependent problem. Solutions which perform well on data of a given sensor may be unsuited for other kind of data. In particular, in this thesis, several cases are investigated: very high spatial resolution optical data provided by the most advanced sensors like GeoEye or WorldView, hyperspectral data like AVIRIS, and Cosmo-SkyMed multitemporal SAR sequences. As will be later shown, different solutions have been proposed depending on the involved data.

Optical remote sensing imagery has been to a paradigm shift in the decade

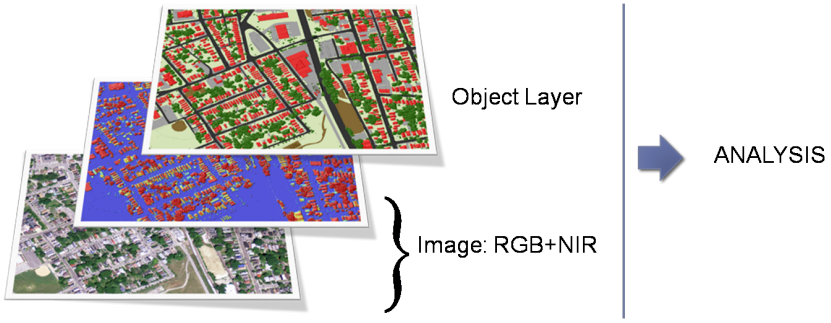


Figure 1: Obia Paradigm: An object layer is associated with the data before the analysis.

after year 1999. Landsat 7 launched in 1999 (with Multispectral (MS), 30m spatial resolution; Panchromatic (Pan), 15m spatial resolution), IKONOS launched in 1999 (MS, 4.0m; Pan, 1.0m), Quickbird launched in 2001 (MS, 2.44m; Pan, 0.61m), WorldView-1 launched in 2007 (Pan, 0.5m), GeoEye-1 launched in 2008 (MS, 1.65m; Pan, 0.42m), and WorldView-2 launched in 2009 (MS, 1.8m; Pan, 0.46m) are evidence of this shift. The spatial resolution has been changed so considerably that pixel size has become smaller than a size of car which was earlier bigger than two or three buildings. This led to research on new classification algorithms for high and very high resolution remote sensing images because traditional pixel based analysis was proved to be insufficient due to its incapability to handle the internal variability of complex scenes [119, 22, 30]. These also propelled object based approach or Object Based Image Analysis (OBIA) for very high resolution image segmentation [70]. Detailed applications and discussion on the development trends of OBIA can be found in [21]. The OBIA concept applied to a multispectral image is depicted in figure Figure 1

According to the aforementioned definition of segmentation, the major thrust is on determining the suitable homogeneity measure which can discriminate the objects from each other. However generating an object layer from HR remote sensing images is a complex, and often ill-defined, task. In fact, the same data hold different meanings depending on the *scale* of observation – individual buildings, roads, parking lots, etc. become just an “urban area” at a lower scale – and the user has often little clues to solve these ambiguities. A popular approach consists in carrying out a *multi-scale* image segmentation and let the user select the scales where the objects of interest are better de-

lineated. However in this thesis more attention is given to the segmentation techniques for the low-level segmentation, where "low level" means that the technique outputs the basic object-level image description, which is composed of homogeneous regions that can be easily processed to generate application oriented products with a higher semantic value. In particular two different approaches are considered. The first one resorts to a watershed transform based on a distance-from-edge topographic surface, so guaranting a faithful delineation of detailed contours. Such a characteristic is of fundamental importance in order to preserve all valuable information at an object-level description. In the second approach, a graph-based modeling is explored in order to generate super-pixel segmentation maps which are highly consistent with the edges of the image objects. Although these two approaches sound very similar as they both tend to achieve high local accuracy, the latter approach can be generalized to different, more complex, situations and is also less computationally demanding than the former. Image segmentation plays a very important role even in the interpretation and understanding of SAR images. It has received an increasing amount of attention and therefore hundreds of approaches have been proposed over the last few decades [149]. Different from optical images, SAR images are inherently contaminated by speckle noise, which inevitably deteriorates the performance of segmentation. Approaches with good performance are often involved in complex computation which may lead the whole process to be more expensive in terms of time [62, 61]. So it is still an urgent task to devise simple and efficient methods. In this thesis a new approach for SAR remote sensing data exploration, based on a tight humanmachine interaction, is explored. The analyst uses a number of powerful and user-friendly image classification/segmentation tools to obtain a satisfactory thematic map, based only on visual assessment and expertise. All processing tools are in the framework of the tree-structured MRF model, which allows for a flexible and spatially adaptive description of the data. The proposed approach is tested in the exploration of multitemporal COSMO-SkyMed data, appropriately registered,calibrated, and filtered, obtaining good performances in both subjective and objective terms, to that of comparable non interactive methods.

Chapter 1

Overview on segmentation

The goal of the present chapter is to provide an overview of the segmentation problem focusing, in particular, on the domains of application of the proposed methods, discussing the related state-of-the-art and highlighting critical aspects. In particular in the section 1.1 the case of the optical VHR multispectral data is concerned, with a discussion on the critical problem of scale-dependency of the segmentation. In section 1.2 the problem of super-pixel segmentation is focused together with a perspective on the use of graph-based models as a flexible mean to constraint a segmentation. Finally, in section 1.3, the challenging problem of segmentation of multitemporal Synthetic Aperture Radar (SAR) images is discussed.

1.1 Segmentation of multispectral Images

Extracting knowledge from image data is probably the main research challenge of these decades in remote sensing. Several important end-user applications fit in this class of problems, like land cover classification, urban planning, or change detection [46, 65, 124, 26, 92], which motivates the strong and unrelenting interest for this field of study. These high-level applications rest, in turn, on fundamental data processing methodologies like image enhancement, fusion, and segmentation [74, 7, 147, 146, 64], which are themselves object of intense research.

Segmentation, in particular, has always played a fundamental role for remote sensing applications and many powerful approaches have been proposed over the years [41], providing often very good results in challenging real-world problems. Nonetheless, the rapid progress in sensor technology, with images

characterized by an ever growing spatial (but also spectral and temporal) resolution, calls for new methods which take into account the different nature (not just higher volume) of the information by now available. In fact, classical pixel-based methods exhibit clear limitations with high-resolution (HR) images, due to the complexity that characterizes the regions of interest even at the finer scales [65, 117]. For this reason, most recent techniques for remote-sensing image analysis adopt an object-based approach, relying on a preliminary process that extracts the elementary image regions, thus forming a new description level which simplifies the emergence of salient information. As a matter of facts, *object-based image analysis* (OBIA - GEOBIA for geospatial data) has been pointed out [21] as the more suitable paradigm for the interpretation of HR images, while object-based methods have already been proposed for various processing tasks [27].

However, generating an object layer from HR remote sensing images is a complex, and often ill-defined, task. In fact, the same data hold different meanings depending on the *scale* of observation – individual buildings, roads, parking lots, etc. become just an "urban area" at a lower scale – and the user has often little clues to solve these ambiguities.

A popular approach consists in carrying out a *multi-scale* image segmentation and let the user select the scales where the objects of interest are better delineated. Multi-scale segmentation can be obtained, for example, through iterative region merging applied to a fine level segmentation [83], or by varying a scale-related parameter which defines the amount of information to be gradually "filtered away" in the growing process [16]. Both approaches are indeed used in well-known commercial softwares, like eCognition [12, 1] and ENVI [75]. In both cases, however, the user must undertake a painstaking trial-and-error process to identify the scales which provide a satisfactory trade off between contour precision and under/over-segmentation of objects. Techniques for the automated selection of the scale parameter have also been proposed [52, 47], but a single-scale description is often inadequate [11] in applications involving complex landscapes. On the other hand, the fusion of multiple-scale segmentations, considered for example in [88] for a landslide detection application, requires an intense *ad hoc* post-processing to provide an acceptable contour accuracy.

An interesting solution for multiple-scale object description is proposed in [4], based on a hierarchical segmentation, obtained by means of morphological object extraction with varying-size structuring elements. In each band of interest, a hierarchical segmentation is provided, and a meaningful pruning

of the relative tree structure is automatically selected based on the analysis of spectral homogeneity and neighborhood connectivity. Despite its flexibility and independence from fixed scales, this technique cannot directly manipulate multi-band (or multi-resolution) sources, and in the general case high-level analysis needs to "browse" objects among multiple segmentations at the feature level, hence in an application-dependent fashion.

Another approach relies on binary partition trees (BPTs). A BPT-based representation strategy has already been used in the context of hyperspectral image analysis and classification, see [136, 137, 116]. In [80], the BPT-based approach has been applied to multiresolution images: a hierarchical segmentation stack is first computed and then a binary partition tree is associated with it and properly pruned, providing a single but multi-scale segmentation. However, pruning is only semi-automatic, since it requires an expert user to train the algorithm on a sample image acquired by the same sensor and portraying a scene with similar content as the target image. This approach has been also generalized [79] to deal with multiple images of the same scene coming from sensors with different resolutions. In particular, the segmentation is first computed based on the coarsest resolution image and progressively refined by using the available data at higher and higher resolution. At each step, the boundaries previously determined are fixed, and the current data are used to reveal nested sub-structures.

The above method is just an example of the growing interest for multi-resolution (MR) images. MR sensors, such as Ikonos, GeoEye, WorldView or Pleiades, are becoming a standard in remote sensing as they meet the demand for ever higher spatial and spectral resolutions, unachievable with current technology. To this end, MR sensors acquire a single panchromatic (PAN) image at high resolution, typically, below 1 meter, complemented by a low-resolution multispectral (MS) image composed of 4-8 bands, relying then on signal processing techniques, like pansharpening [7], to recover a full resolution multispectral datacube.

With MR images the issue of scale is intrinsic in the original data. While some small objects may be appreciable only at the higher resolution, complex textures (forests, agricultural fields, etc.) are recognized as objects only through their spectral coherence at lower resolution. Although conventional segmentation techniques could be certainly applied to pansharpened data, better results (or at least not worse) can be achieved by working on the original data. In fact, pansharpening is well-known to introduce spectral distortion and/or spatial blur on the fused datacube, especially in high-resolution regions

of the image. Therefore, segmentation and, in general, image processing techniques devised expressly for multi-resolution images are likely to become hot research topics in the near future. Some contribution have already appeared in the recent literature. In [140] a clustering-based classification is performed by first segmenting separately the images at different resolutions and then jointly characterizing the connected regions. A more theoretical approach is followed in [97], where a multi-resolution image model is proposed for supervised land cover classification. Both techniques, however, appear to provide limited accuracy on fine details, as they focus mainly on land-cover classification. The region-based data fusion approach proposed in [28], instead, guarantees fine detail preservation, but carries out only a partial segmentation of the image, aimed at some particular land covers.

In this thesis, as will be explained in section 2.1.1, a novel technique based on marker-controlled watershed transform is discussed, for the low-level segmentation of single and multi-resolution images. By "low-level" is meant that the technique outputs the basic object-level image description mentioned before, composed of homogeneous regions that can be easily processed to generate application-oriented products with higher semantic value. By resorting to watershed, based on edge detection, a faithful delineation of high-resolution contours is ensured, a fundamental requirement in order to preserve all valuable information in the object-level description. The use of markers, on the other hand, allows us to reduce the over-segmentation typical of watershed, providing thus a *compact* object layer. More remarkably, objects with very different sizes are extracted with the same contour accuracy. That is, the proposed technique promotes the coexistence of very small localized objects and large structured objects covering large parts of the scene. For Multi Resolution (MR) inputs, the processing takes place, as far as possible, on the original data, avoiding the information losses caused by pansharpening. Markers of two types are defined, based on *i)* spectral, and *ii)* morphological properties. The former are used prevalently in low-detail areas of the image, while the latter are more important in high-detail regions, where the spectral information is less reliable. Marker extraction is fully automatic, with no supervision by the user, nor is the user required to define any scale parameter.

A first version of the technique has been published in a conference paper [64] and subsequently its multiresolution version [93]. In [63], with respect to that preliminary work, some fundamental processes are modified, like spectral marker generation and edge fusion, and improved a number of technical details, producing eventually a much more effective and efficient algorithm,

whose source code is available online for testing. In addition, a totally new functionality has been included for the segmentation of single resolution multispectral images. In Section 2.1.1 after recalling some necessary background notions on watershed-based segmentation and marker generation, a detailed explanation of the proposed method is given and validate its performance (in section 3) by both directly quantifying the meaningfulness of extracted objects and using the map for a simple application of classical supervised classification.

1.2 Graph based segmentation

Remote sensing images are characterized by an ever higher spatial resolution, improving the quality of existing products and allowing for the design of completely new ones. As a side effect, however, one has to manage a huge bulk of data. Algorithms well established for much smaller images turn out to be exceedingly cumbersome and slow in this case, calling for new approaches to image processing. Among the most popular solutions to this problem is the use of higher-level representations, from *superpixels* to semantic object layers, subsequently used in object-based image analysis (OBIA or GEOBIA) [21]. In particular, there has been great interest on *superpixels*, which partition the image in small homogeneous patches of approximately the same size, *e.g.*, SLIC [3].

Despite their simplicity and popularity, however, SLIC superpixels do not always provide a faithful representation of remote sensing images. With its mostly agglomerative nature, in fact, SLIC produces superpixels that are only approximately aligned with region boundaries, elements of great importance for a number of high-level tasks. Alternatively, one can generate superpixels starting from the edges themselves, so as to preserve their accuracy throughout the process [64, 93, 63].

The superpixel-level representation is an efficient way to deal with the segmentation of high-resolution images. However, it is only as an intermediate step towards higher-level segmentation, where elementary object are associated with expressive compact features, and with a suitable semantics. To move from one level (superpixels) to the other (meaningful objects) one can resort to efficient graph-based methods. The image is represented as a region-adjacency graph (RAG), with regions associated with vertices and boundaries with edges. Then, image segmentation is regarded as a graph partitioning problem [43], which can be solved through suitable optimization tools.

In recent years, several graph-based image segmentation methods have been proposed, e.g., [121, 139, 138, 44, 57, 23]. Unfortunately, their complexity grows very fast with the graph size, making them unsuited to large remote sensing images.

Here, we focus on the graph partitioning model first proposed in [13] in the field of document analysis and known as *Correlation Clustering*. Although optimal correlation clustering (CC) is itself a NP-hard problem, good approximate solutions can be obtained in limited time by means of suitable greedy heuristics. The main aspects discussed in this thesis, in particular in the chapter 2.2, can be therefore summarized as follows: i) use of an edge-oriented superpixel representation of the image; ii) solution of the graph partitioning problem under a correlation-clustering formulation; iii)use of a greedy heuristic which provides fast near-optimal solutions.

In the previous works [64, 93, 63] the object layer is reached by means of image processing tools, namely, superpixel agglomeration driven by both morphological and color/spectral markers. As long as pixel sizes remained typically coarser than, or at the best, similar in size to the objects of interest, emphasis was placed on per-pixel analysis, but with increasing spatial resolutions alternative paths have been followed, aimed at deriving objects that are made up of several pixels. Graph-based methods, however, developed mostly in the computer vision community, represent a promising a often more general alternative . Generally in this class of techniques the image is represented as an undirected, weighted graph $G(V, E)$ where each pixel (or super-pixel) in the image is considered as a vertex of the graph and an edge is formed between a pair of a adjacent nodes. In this thesis, I discuss a new image segmentation technique based on correlation clustering, presented in [91].

Superpixels are first obtained through edge detection and edge-based watershed, and associated with the vertices of an undirected graph. Then a simple method to characterize the relationships between couples of superpixels is introduced, and a greedy procedure to obtain a fast and accurate CC solution is discussed. Experiments on real-world remote sensing images prove the effectiveness of the method. In particular the necessary background on correlation clustering and proposed method is given in section 2.2. Experimental results, that show the effectiveness of the method, are shown in section 3.3.

1.3 Interactive segmentation of SAR images

As already stated in the previous sections, the constellations of sensors available nowadays provide data with unprecedented spatial resolution and revisit time. Therefore, the bulk of available data reached a level that can be hardly managed by human operators, leading to an extreme automation of algorithms, with techniques that alienate more and more the users from direct data management and analysis [126, 141, 15]. This paradigm, certainly necessary and advantageous in the data acquisition and storage steps, has been also extended to the data processing realm, leaving the user with the role of mere interpreter of results obtained through "black-boxes" implemented on the basis of some necessarily simplified models [68]. In remote sensing, this can lead to the misclassification of objects and the misidentification of phenomena, and eventually to a wrong interpretation of the data.

These problems can be mitigated by restoring the central role of the user as the key actor in a number of high-level decisional tasks. As brilliantly argued in [85], human beings and computer algorithms are good at solving different and mostly complementary tasks. As the interpreter cannot be asked to compute important data statistics and synthetic features, essential for all decision making processes, algorithms cannot be expected to make correct decisions in a wide range of unpredictable situations, which arise quite often in remote sensing and cannot be taken into account by compact mathematical models. To obtain the most from the available data, the user must be given the opportunity to *interact* with the computer, in a simple and easily understood way, to drive the decision process. In such a way, thanks to real-time actions and reactions, the processing can be transformed from an "objective coding of the image information content" [33] into a machine learning process guided by the user's knowledge and judgement.

In this paradigm, the computer is only a flexible (yet powerful) number-crunching tool driven by the expert towards a solution that is *context-aware*, as opposed to the *context-independent* solutions offered by totally automatic processes, where by "context" we mean the many possible data peculiarities and specific application needs which call for dedicated work-flows. The user-machine interaction is fundamental for recognizing the inconsistencies between the technique/model and the context in which it is applied, and the user becomes the central actor of this task, participating actively in the processing chain, based on the accumulated expertise [33].

Among the many image processing tasks relevant for remote sensing, segmentation is probably the one which could benefit most from the interactive

paradigm described above. Although it is not obvious, in general, how to manage the huge amount of data provided by remote sensing, and what workflow is best suited to extract the information of interest from them [120], segmentation is very likely to be part of it. In [85], a model for remote-sensing data exploration is proposed. Not surprisingly, segmentation is taken as a running example to prove its potential, adopting a number of tools, under the user supervision, to extract a meaningful thematic map based on high resolution optical data and a digital elevation model of the scene. Following this seminal paper, recent work has focused on a more formal definition of the human-machine interaction frameworks and their applications to remote-sensing data analysis [33, 24]. The tendency to leverage user interaction in this domain is further confirmed by very recent works both in the SAR [17] and optical [45] context.

In this thesis, inspired by [85], an innovative user-driven approach for the unsupervised land-cover classification of multitemporal Cosmo-SkyMed SAR images is proposed. SAR image processing, especially in the multitemporal case, is a perfect example of the added value represented by the user intervention in the processing chain. Interpreting SAR images requires, in general, a deep understanding of many relevant physical processes and models. Moreover, with multitemporal data, the physical parameters of the scene vary not only in space but also in time, often in an unpredictable way, affected by human activities which can induce both temporal patterns and local anomalies in the electromagnetic response. These circumstances can be controlled and mitigated interactively by the user who can modify the processing flow using available prior knowledge or information drawn from different sources.

In the proposed work-flow (explained in details in the sections 2.3 and 3.4.1), after the suitable preparation of data, the multitemporal stack of intensities and the coherence map extracted from it are combined to obtain an accurate thematic map. Unlike in [85], most of the segmentation tools belong to a single flexible algorithmic suite, the Tree-Structured Markov Random Field (TS-MRF) algorithm, based on the model of the same name, originally proposed for the unsupervised [104, 37] and supervised [105] land-cover classification of multispectral images. Indeed, with its hierarchical nature, TS-MRF represents a natural basis for interactive segmentation, allowing the user to check and then validate, or further process, results that are confined to a single class or region of the image, without interfering with satisfactory results observed elsewhere. Moreover, the opportunity to work in a single algorithmic framework (without precluding the use of others, of course) reduces the train-

ing required of the users to take full advantage of the image processing and data analysis tools available. TS-MRF does not need training data (in unsupervised mode), hence is especially suited to data exploration. Its effectiveness has been largely proven in the segmentation of optical remote sensing data, while its application to SAR imagery has been long prevented only by the lack of data reliable enough for a detailed segmentation, a problem now overcome thanks to the wealth of multitemporal SAR data provided by the COSMO-SkyMed constellation. Human-machine interaction represents the correct modality to find a meeting point between challenging SAR-related tasks and well-founded and long-proven methods developed in the optical domain.

In general SAR image segmentation is an extremely challenging task. Besides the mentioned need for a knowledge-based interpretation of phenomena, meaningful information must be extracted from data that exhibit a very high dynamics, and are characterized by speckle which severely corrupts region boundaries and fine details, preventing the use of classical image processing tools. In order to deal with speckled images most of the segmentation techniques proposed in the literature adopt a Markov Random Field (MRF) approach [77, 110, 128, 125]. By defining explicitly the spatial interaction between neighboring pixels, one can enforce suitable regularity constraints, avoiding so the highly fragmented segmentation output maps typical of pixel-level approaches. Many variation to the classical MRF based data-flow have been proposed as, for example, embedding the MRF model in the clustering space and using graph cuts to search the optimal data clusters [141], or using a hierarchical MRF for multiresolution segmentation, with suitable expedients to avoid block artifacts [144]. Irrespective of the detail, all these methods are based on the assumption of a multiplicative noise with circular Gaussian statistics, which makes full sense for fully developed speckle. For high resolution SAR images, however, this model is not always appropriate, because it cannot be assumed that a large number of scatterers fall in any resolution cell, especially in urban areas. This observation has spawned a number of recent papers. In [134] a new model for the statistics of the scattering process is proposed, and used to improve the classification of urban areas. Another way to deal with high resolution images goes through the use of statistical learning methods with appropriate local descriptors. In [39] texture features are included in the MRF model to identify distinct ice types. In [143] a hierarchical Markov "aspect" model is proposed to generate dense and efficient terrain-class labeling by exploiting both high-level context and multiscale features. In [145] conditional random fields are used to incorporate context interactions in the

extracted features. Clearly, for all these methods, a significant training phase is required. All the methods outlined above have been proposed for a single intensity or amplitude SAR image. However, co-registered multitemporal SAR images provide a much richer information, with new opportunities for land-cover classification. First of all, they allow one to use effective despeckling filters, improving significantly the quality of data. Despeckling has never been a popular option for the segmentation of single-look images, because of a possible resolution loss (mostly absent, though, in modern despeckling techniques [100, 42]). With multitemporal data available, however, undue smoothing effects can be avoided altogether. Needless to say, despeckling modifies data statistics, and models for unfiltered data do not apply anymore. Of course, besides the improved data quality, the mere fact that a *vector* of data is associated with each pixel opens the way to major improvements in segmentation and classification. Indeed, with their very high spatial resolution combined with a short revisit time these data represent a powerful tool for accurate interpretation of the ground scene. Therefore, in recent years, research has focused mostly on application-oriented tasks rather then methodological developments: supervised land-cover classification [25], urban-area segmentation [98, 40], flood mapping [89], flood monitoring [38, 109], wet snow cover in a mountainous area [118]. This short review of the literature is concluded by mentioning the method recently proposed in [6], where the binary partition trees [114] are used to perform the hierarchical segmentation of multidimensional SAR data. A multiple-resolution description of the image is obtained, with a structured representation which supports easy access and processing of subsets of regions. Although not based on MRF models, this method performs SAR image segmentation through a hierarchical approach, similar in principle to the one discussed in this thesis. This is not surprising, though, since the hierarchical approach fits very well the scale-dependent and non-stationary nature of SAR images.

Chapter 2

Main proposals

This chapter describes the three main proposals resulting from the thesis research activity. The first main outcome, object of Section 2.1, is an innovative watershed-based technique for multispectral and/or multiresolution optical data particularly suited for data acquired with sensors like Ikonos, GeoEye, WorldView, Pleiades al likes, where a submetric resolution panchromatic band is coulped with a lower resolution multiband component. Next, in Section 2.2, a new graph-based segmentation approach relying on the concept of correlation clustering [13] is presented. The proposed method allows one to better control and/or constraint a generic edge-based segmentation process in a very flexible manner thanks to a general framework where local cues can be easily injected. Such a framework is rather general as multisource (data fusion) image segmentation can be easily enclosed. Finally, in Section 2.3, the case of multitemporal SAR image sequence segmentation is faced. In particular, an interactive tool for segmentation is proposed, which is based on the Tree-Structured Markov Random Field (TS-MRF) framework, originally conceived for multispectral data [37].

2.1 Edge, mark and fill algorithm

Due to its local consistency properties, the watershed transform [127] represents a precious segmentation tool when a fine-scale decomposition of the scene is required, notably in the remote sensing domain [142, 83]. In watershed segmentation a suitable topographic surface is associated with the image, and is progressively filled with water until it is flooded. Whenever two basins meet, a virtual dam is built between them, and when the process stops, each

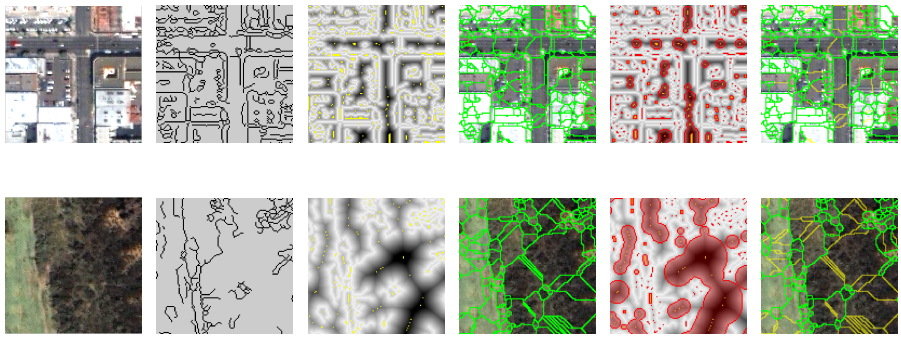


Figure 2.1: From edge-based watershed segmentation to *Edge, Mark and Fill*: from left to right, source image, edges, corresponding topographic surface with highlighted minima, watershed segmentation, morphological marker generation and final EMF segmentation with suppressed watershed boundaries in yellow.

basin represents an image segment.

Of course, the final output depends strongly on how the surface is defined. For low-level segmentation purposes, a popular choice consists in using the image gradient as surface, with the gradient crest lines assuming the role of region boundaries. However, the image gradient is typically quite noisy, leading to a deep over-segmentation, with objects that represent little more than superpixels. Smoothing the gradient, on the other hand, causes the loss of many fine details.

A better alternative consists in first performing edge detection, thereby exploiting the sophisticated filtering strategies adopted in this process, and then using the distance-from-edge map as topographic surface. This solution, adopted in the present work, allows to better control under-/over-segmentation by suitably tuning the parameters of the edge detection process. Moreover, many detected edges will be part of the final region boundaries. Finally, the watershed applied to the distance transform has the interesting property of decomposing the image into elementary shaped objects [127, 112, 64], thus allowing for the use of morphological properties to detect missing region boundaries from the edge map.

Unfortunately, when fine details need to be preserved, a high degree of over-segmentation is easily observed. In fact, tuning the edge detector to guarantee high-recall implies a relatively low-precision, with a large number of dubious edges. To better appreciate the extent of this problem, consider the

examples shown in Fig. 2.1. In the left column we show two small 128×128 sections of a pansharpened Ikonos image, a urban area (top) and a vegetation area (bottom), and in the second column the corresponding edge[1] maps. These maps, without further processing, do not provide an acceptable image segmentation. In the third column we show the surface associated with each edge map, with all local minima, seeds of image segments, shown in yellow. In the fourth column, the final segmentation boundaries are shown as green lines superimposed on the RGB original image. All relevant contours are kept but there is a clear over-segmentation.

In such circumstances, results can be much improved by exploiting some prior knowledge on the shape and extension of image objects, which can be taken into account by means of *markers*. In marker-controlled watershed transform [127], the user is allowed to guide the segmentation process by drawing one or more markers on the image: all segments touched by a marker will be eventually merged together, reducing over-segmentation. Of course, setting the markers manually , *e.g.*, [19], can be a low-precision and tedious task, if not plain impossible for object layer extraction from a large remote-sensing image. More interestingly, markers can be defined through an automatic procedure, like in [142, 131, 64, 18, 132], which takes into account suitable object models or statistical information on the image.

This latter approach is followed in the Edge, Mark and Fill (EMF) algorithm, originally proposed in [64], which performs a watershed segmentation with fully automatic markers, obtained from an arbitrary edge map through a purely morphological process. The rationale of EMF comes from observing that, in typical high-resolution images, most of the region seeds are due to minor irregularities in the edge profile, or even just to the discrete geometry of the problem. These seeds are often very close to one another and are not separated by detected edges (see again Fig. 2.1, third column) although they will be separated by new boundaries at the end of the process. Therefore they should be merged from the beginning, by means of suitable markers, to avoid the generation of useless segments. In EMF these "interacting" seeds, closer to one another than to any of the original edges, are detected and linked before applying the watershed transform.

EMF is described formally through the pseudo-code of Alg. 1. After com-

[1]Here, and throughout this description, we use the Canny edge detector [29] for its good performance (in high-recall regime) and wide availability, but this choice is immaterial under a conceptual point of view, and any other detector could be used. For multispectral data, edge detection is performed separately on each band, and the final edge map is obtained by overlapping the band-wise edge maps and performing a morphological thinning.

Algorithm 1 EMF

Require: I ▷ Input image
Ensure: Q ▷ Segmentation
1: $E = \text{EdgeDetection}(I)$
2: $D = \text{DistanceFromEdge}(E)$
3: $M = \text{MorphMarker}(D)$
4: $Q = \text{MarkedWatershed}(-D, M)$

puting the edge map E, and the corresponding distance transform D, the morphological marker map M is generated based excusively on this latter information, and used to drive the watershed transform[2].

Algorithm 2 Morphological Markers

1: **procedure** $M = \text{MORPHMARKER}(D)$
2: $seeds = \text{LocalMinima}(-D)$ ▷ list of local minima positions
3: **for** $k = 1 : |seeds|$ **do**
4: $s = seeds(k)$
5: $SE = \text{circle}(D(s) - \epsilon)$ ▷ structuring element
6: $M(k) = \text{Dilate}(1_s, SE)$ ▷ basic marker for s
7: **end for**
8: $M = \bigcup_k M(k)$ ▷ aggregated marker map
9: **end procedure**

The core of the algorithm is the marker generation procedure of Alg. 2. Each region seed, s, is dilated with a circular structuring element of radius $D(s) - \epsilon$, to generate a basic marker. Since $D(s)$ is the distance between s and the closest edge, such basic markers do not intersect edges. However, when corresponding to close seeds, they can overlap. The union of all basic markers generates thus a final map M comprising a smaller number of extended markers, used to reduce oversegmentation.

The effect of using EMF on our running examples is depicted in the last two columns of Fig. 2.1: first, we show how clusters of interacting seeds are covered by a unique morphological marker given by the union of the corresponding circles. These markers are coherent with the local image morphology, partly recovering linear structures like the roads (top) as well as isotropic regions like the wood (bottom). Then, on the rightmost column, we show the final segmentation boundaries, again as green lines superimposed to the im-

[2]In this thesis, we use the implementation proposed in [20], which does not generate a separate label for the region boundaries as in the classical version.

ages, together with the watershed boundaries (in yellow) suppressed through the morphological markers.

2.1.1 Including Spectral Marker

EMF is based on exclusively morphological criteria, assuming an implicit regular model for image segments. Therefore, it is a very general tool, which can be used on any kind of image, improving results w.r.t. plain watershed. However, it totally neglects the information coming from the spectral content of the image, especially valuable for high spectral resolution sources. Therefore, to exploit this information, we extend the basic algorithm by including a further process for the automatic generation of spectral markers.

The problem of automatic marker selection has been discussed in many papers, especially for gray-scale and color images, which are well summarized in [131]. A further approach for spectral marking, closer to the rationale of the Edge, Mark and Fill algorithm, is presented in [87] in the context of medical image analysis. Here, the interaction among the seeds of the watershed transform is studied looking at the photometric variations along *valley-lines* of the topographic surface. This solution, although inspiring, is hardly applicable to remote-sensing images, since the analysis of typically complex spectral variations cannot be reliably performed without taking into account regional characteristics.

More suitably for remote-sensing applications, it is worth mentioning some approaches, which will be also used as state-of-the-art references in the experimental analysis of Sec.3.1.2. A simple solution, proposed in [18] for land cover classification, considers as spectral markers single pixels randomly selected from a preliminary classification. The selection is repeated several times and, despite its simplicity, provides interesting results within the given context. A more sophisticated marker selection is performed in [131], where the target image is first classified by SVM, thanks to the availability of training samples, and a pixel-wise membership probability map is suitably computed. Then, each connected component is refined by eliminating all pixels whose membership probability is below a given threshold. A larger threshold is used for small-size components, introducing therefore a certain degree of adaptivity. Since the final application is classification, the reshaped connected components associated to the same class are kept together, forming a single global marker per class which is not necessarily connected. In [132] the process is slightly modified, since connected components are reshaped by means of morphological operations rather than by membership weighting. All the de-

scribed methods share a *global* approach to spectral marker generation, and rely on the availability of training samples which allow for a pre-classification process carried out on the whole image. Although some local processing is performed in [132] by means of the morphological erosion, in no cases the local (spectral, morphological) characteristics of the image are taken into account in the marker generation process. Moreover, object-level segmentation is subordinate to classification, limiting the applicability of the OBIA paradigm. We will instead consider a totally local approach, where spectral markers are generated in close domains singled out based on preliminary edge detection, and with no knowledge about image classes. The relative distance-based topographic surface will therefore provide a reference geometry for generation of locally accurate markers. Note that the idea of exploiting spectral information for marker generation is not novel: a notable example is [82], where data from a small number of manually selected markers are used to provide multiple topographic surfaces, and supervised segmentation is then carried out using a modified marker-based watershed framework. This solution has been successfully applied to remote sensing images in [108], in a context where manual marker selection was reasonable. However, this is simply not the case for object layer extraction, all the more when the aim is to provide a totally unsupervised segmentation technique.

We describe the proposed algorithm, named EMF+, by means of a simple toy example, were all phenomena of interest can be easily spotted. Fig. 2.2 shows a synthetic color image with the detected edges superimposed on it (a), the associated topographic surface (b), and the final unmarked watershed segmentation (c). A human interpreter would probably segment right-away this image in the two rectangular regions shown in part (h). Edge detection, however, does not provide, by itself, this nice result. A part of the "true" edge in the top of the image is lost due to vanishing contrast, while a "false" partial edge is detected in the bottom. Edge-based watershed completes both edges, leading to the observed over-segmentation. The problem is that edge detection, based only on local data, misses important contextual information at larger scales. To extract this information one can resort to a global process, like spectral-based segmentation. However, also in this case the result is often unsatisfactory due to the global/local scale mismatch, as shown in part (d): region boundaries are not accurate, and a small unwanted region appears in the bottom. Our idea, then, is to use the connected components singled out by this process only to generate additional *spectral* markers, which will be eventually merged with the *morphological* markers of the basic EMF.

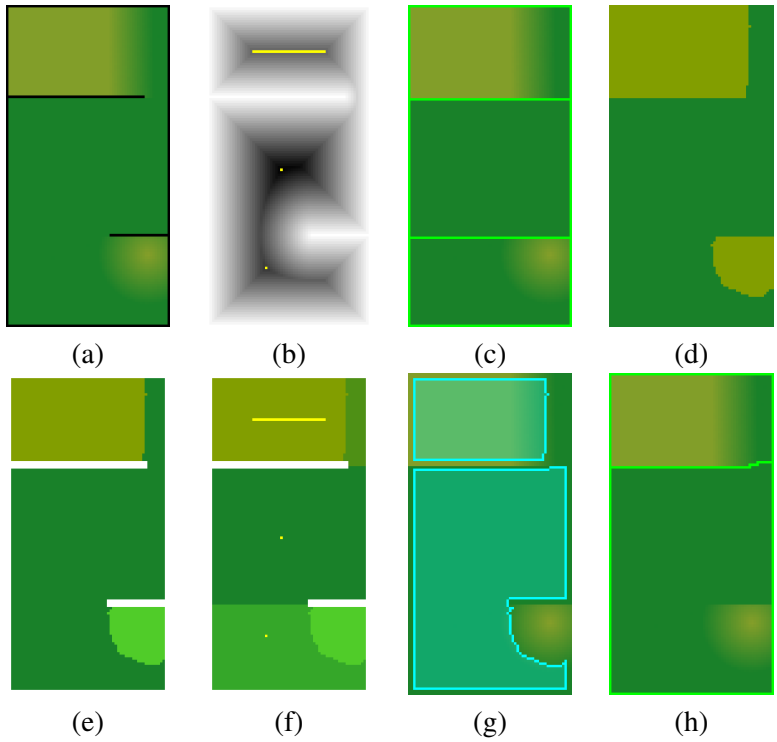


Figure 2.2: Toy example for spectral marker generation. Source image and edge map (a), topographic surface (b), watershed segmentation (c), spectral-based segmentation obtained using TS-MRF [37] (d), spectral marker domain (e), watershed-fit marker erosion (f), final spectral markers (g), marker-controlled watershed segmentation (h).

Before using such regions as additional markers, however, we force them to become consistent with the edge-based watershed segmentation, through a process called *Watershed-fit Marker Erosion* (WME). With reference to the example, we first restrict the spectral marker domain to the part of the image not covered by the dilated edges, and then perform spectral segmentation on this domain, labeling the resulting connected components as shown in (e). We then intersect the watershed segmentation (c) with the spectral segmentation (e), obtaining the five regions shown in (f). Two of such regions, however, do not cover any seed and cannot be associated with any of the watershed segments. Therefore they are regarded as artifacts and removed. The remaining regions are eventually grouped in connected components (in this process the

two regions in the bottom are merged) obtaining the final spectral markers (g), whose superposition to the topographic surface (b) leads to the final watershed segmentation (h).

Algorithm 3 EMF+

Require: I
Ensure: Q
1: $E = \mathrm{EdgeDetection}(I)$
2: $D = \mathrm{DistanceFromEdge}(E)$
3: $M = \mathrm{MorphMarker}(D)$
4: $S = 1 - \mathrm{Dilate}(E, \mathrm{square}(3))$
5: $S = \mathrm{SpectMarker}(S, I)$
6: $S = \mathrm{WME}(S, D)$
7: $Q = \mathrm{MarkedWatershed}(-D, M \oplus S)$

As for EMF, we provide a more formal description of EMF+ by means of the pseudo-code Alg. 3. The first part of the algorithm coincides with EMF, then the spectral marker map S is computed for the off-edge part of the image, and merged with the morphological marker map M to drive the watershed. Of course, this is only a very high-level description, a number of important details and relevant design choices are hidden in the SpectMarker procedure, described by means of the pseudo-code Alg. 4.

Algorithm 4 Spectral Markers

1: **procedure** $S = \textsc{SpectMarker}(M, I)$
2: $[\{R\}, labels] = \mathrm{ConnectedComponents}(M)$
3: **for** $k = 1 : |labels|$ **do**
4: $S(k) = R(k)$ ▷ basic spectral marker
5: $a = \mathrm{Activity}(I(R(k)))$
6: **if** $a > T_\eta$ **then**
7: $S(k) = \mathrm{TS\_MRF}(I(R(k)))$ ▷ update for active regions
8: **end if**
9: **end for**
10: $S = \bigcup_k S(k)$ ▷ aggregated marker map
11: **end procedure**

A first important observation is that spectral markers are generated independently for each closed region $R(k)$ singled out by the initial edge detection and dilation (the input mask M in Alg.4). These consistent boundaries are accepted with no further inquiry and hence the corresponding closed regions form spatial domains that do not interact anymore (for what segmentation is

concerned) with the rest of the image. Needless to say, this choice improves efficiency very much. For each closed region, a spectral activity measure is computed (the mean of the component-wise variances) and only high-activity regions are further segmented, while a single spectral marker is associated with the other regions. The local segmentation is performed by the unsupervised binary TS-MRF algorithm [37], followed by an independent labeling of the connected components of the two classes singled out. MRF regularization helps avoiding the unwanted fragmentation caused by noise with simpler clustering techniques. The spectral-based segmentation in the toy example of Fig. 2.2(d) has been indeed achieved using this technique. Although multi-class segmentation can be easily implemented with TS-MRF, experimental evidence shows that two classes are enough to deal effectively with these regions. The final step is the watershed-fit erosion, whose pseudo-code algorithm in reported in Alg. 5 without further comments.

Algorithm 5 *Watershed-fit Marker Erosion* (WME)

procedure $EM = \mathrm{WME}(M, D)$
  $W = \mathrm{Watershed}(-D)$
  $seeds = \mathrm{LocalMinima}(-D)$
  for $k = 1 : |seeds|$ **do**
    $s = seeds(k)$;
    $EM(k) = (W == W(s)) \odot (M == M(s))$
  end for
  $EM = \bigcup_k EM(k)$
end procedure

Fig. 2.3 illustrates the effects of using spectral markers on two real-world images. In both cases, edge detection isolates closed regions with spectrally heterogeneous areas (first column). Regarding these regions as segments would cause the undesirable merging of different objects like roads and grass (top) or sand and grass (bottom). On the other hand, watershed segmentation, even with morphological markers, would produce a clear over-segmentation (second column). By using the markers computed by spectral-based segmentation and WME (third column) properly conditioned and merged with morphological markers, a much better result is obtained (last column).

It is worth underlining that, despite some superficial similarity, EMF+ differs profoundly from region merging techniques, such as those implemented in eCognition and ENVI. These latter carry out iterative pairwise region merging based on some similarity criterion, e.g., merging neighbors with mean spectral values closer than a given threshold, thereby introducing the scale-dependency

Figure 2.3: Spectral markers for *Roads* and *Baseball* objects. From left to right: initial segments, EMF segmentation, spectral markers (after WME) and final EMF+ segmentation.

typical of these approaches.

To conclude this Section, Fig. 2.4 illustrates with two examples the impact of spectral markers on results. The final spectral markers are shown on the left. The final segmentation, shown on the right with region boundaries represented in green, is quite good. A number of boundaries originally singled out by the watershed have been removed thanks to the morphological or spectral markers (in orange and yellow, respectively), but fine-scale details are correctly preserved and, in general, all relevant scales are retained in the final map.

2.1.2 Multi Resolution Extension

Multi-resolution images are becoming widespread in remote sensing, and effective tools for their segmentation are definitely of interest. EMF+ can be also applied to multi-resolution images, without any further modification, provided a prior pansharpening of the data is carried out. This latter process, however, can impair segmentation accuracy, especially when the number of spectral bands increases, and certainly affects its computational complexity. Therefore, we developed a multi-resolution version of the algorithm, called *Multi-Resolution Edge, Mark and Fill* (MR-EMF), which uses only the orig-

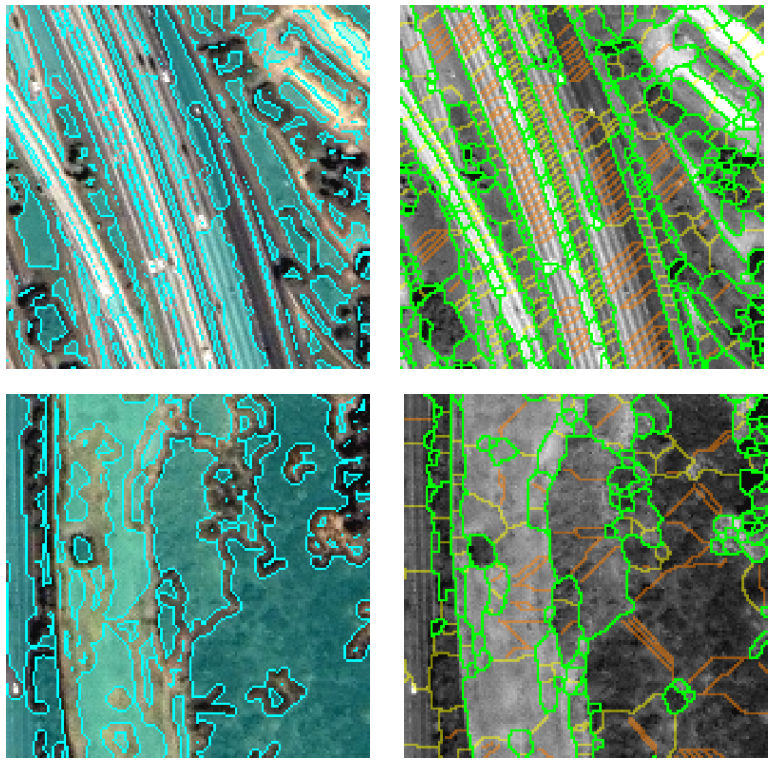


Figure 2.4: Impact of spectral markers on segmentation: spectral markers (left), and final segmentation (right) with superimposed watershed boundaries removed thanks to the morphological (orange) or spectral (yellow) markers.

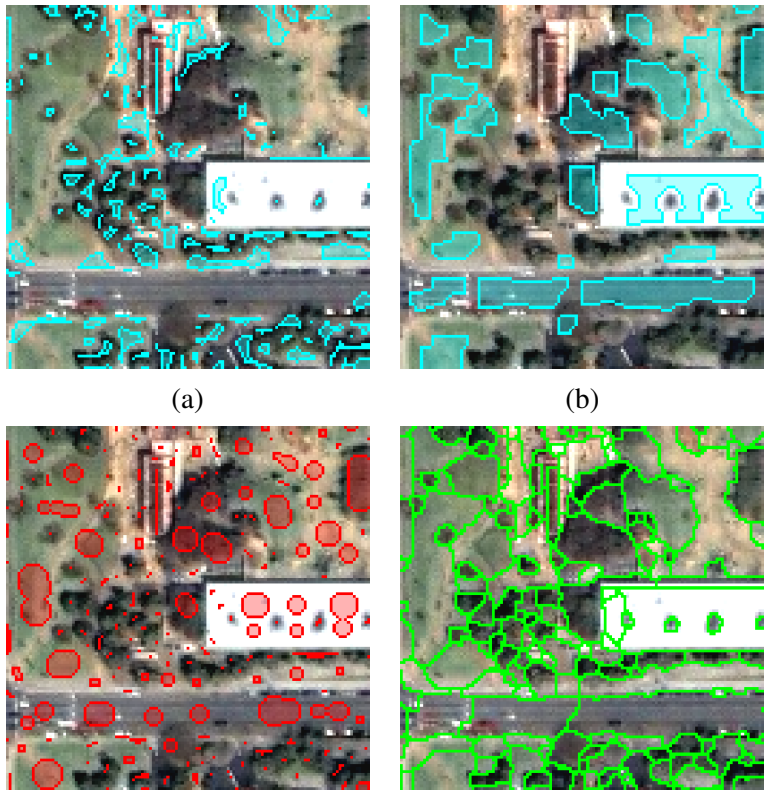
inal data, at multiple resolutions, without any intermediate, and potentially lossy, data processing steps. The spirit of MR-EMF is to selectively exploit the information at different resolutions, giving priority to the high-resolution panchromatic data to determine boundaries and fine details of the scene, and relying on the spectrally rich multispectral data to consistently detect objects at larger scales. More specifically, data processing at pixel level (edge detection and spectral-based segmentation) takes place independently on each component at its native resolution. Fusion of information extracted at multiple resolutions is then performed first at the edge level, then at the region level, by combining separate sets of spectral markers computed on two disjoint domains (PAN/MS).

Let us describe MR-EMF in more detail with the help of the pseudo-code

Algorithm 6 MR-EMF

Require: I_{pan}, I_{ms}
Ensure: Q
1: $E = \text{EDGEFUSION}(I_{pan}, I_{ms})$
2: $D = \text{DistanceFromEdge}(E)$
3: $M = \text{MorphMarker}(D)$
4: $O = (1 - \text{Dilate}(E, \text{square}(3)))$ ▷ O off-edge part of image
5: $SE = \text{square}(\rho)$
6: $S_{pan} = (D \leq \lambda) \odot O$
7: $S_{pan} = \text{WME}(S_{pan}, D)$
8: $S_{ms} = (1 - S_{pan}) \odot O$
9: $S_{ms} = \text{Open}(S_{ms}, SE)$
10: $S_{pan} = \text{SpectMarker}(S_{pan}, I_{pan})$
11: $S_{pan} = \text{WME}(S_{pan}, D)$
12: $S_{ms} = \text{SpectMarker}((S_{ms} \downarrow \rho), I_{ms}) \uparrow \rho$
13: $S_{ms} = \text{WME}(S_{ms}, D)$
14: $Q = \text{MarkedWatershed}(-D, M \oplus S_{pan} \oplus S_{ms})$

of Alg. 6. The EdgeFusion procedure, described in detail later on, provides the high-resolution edge map E by combining the edge maps independently extracted from both the panchromatic and multispectral components. Based on the overall edge map, the morphological marker map M is readily computed as in EMF (lines 2-3). As for spectral markers, they could be computed as in EMF+, after up-sampling the multispectral component. However, due to spectral mixing, the original MS data are not reliable near region boundaries, up to ρ (high-resolution) pixels away from them, with ρ the ratio between high and low resolutions. Therefore we first single out a suitable MS domain, S_{ms}, where reliable multispectral data are available, and compute the MS spectral markers only in this domain. In addition, we compute spectral markers also in the complementary PAN domain, S_{pan}, relying in this case only on panchromatic data. These operations are carried out in lines 4-9 of the pseudocode. More specifically, the PAN domain is defined as the set of pixels close to edges ($D \leq \lambda$) but disjoint with them (in the off-edge area O). This set is then refined by the WME procedure, which excludes points belonging to large regions, limiting the PAN domain to pixels locally "enclosed" by edge segments. The MS domain, instead, is defined as the complement of the PAN domain, again in the off-edge area. It is regularized as well by morphological opening, to separate regions connected by thin (less than $\rho \times \rho$) junctions. Finally, spectral markers are evaluated, based on the panchromatic image in the PAN domain (line 10)



(a) (b)

(c) (d)

Figure 2.5: Spectral markers in the PAN (a) and hi-res MS (b) domains; morphological markers (c); final segmentation (d).

and on the MS image in the MS domain (line 12). The last two calls to WME are needed to separate connected components singled out by the segmentation process and discard those which do not cover at least one minimum of the topographic surface. In this latter case, the map must be first downsampled, to fit the resolution of the MS data, and eventually upsampled again. The segmentation map Q is then obtained through Watershed controlled by the union of all selected markers.

In Fig. 2.5, parts (a)-(c), we show the three types of markers superimposed on an example image. Needless to say, MS and PAN spectral markers never overlap, but they can overlap morphological markers producing, after marker fusion, the basis for a fully satisfactory segmentation, part (d), obtained by

marker controlled Watershed.

Multi-Resolution Edge Fusion

As said before, we extract edges separately in the PAN and MS components, and selectively combine both pieces of information, giving priority to the former. In fact, most region boundaries are detected in both image components, in which case we want to keep only the more precise high-resolution PAN edges. Other boundaries, instead, are detected only in the PAN image (tiny regions under the MS resolution), or only in the MS bands (neighboring regions with very similar projection on the PAN), and all of them should be included in the final map.

Algorithm 7 EdgeFusion

1: **procedure** $E = \text{EDGEFUSION}(I_{pan}, I_{ms})$
2: $SE = \text{square}(\rho)$
3: $E_{pan} = \text{EdgeDetection}(I_{pan})$
4: $E_{ms} = \text{EdgeDetection}(I_{ms})$
5: $E^{+}_{ms} = \text{Thinning}(E_{ms} \uparrow \rho)$
6: $U_{pan} = \text{Dilate}(E_{pan}, SE)$
7: $E^{+}_{ms,only} = E^{+}_{ms} \odot (1 - U_{pan})$ ▷ remove double edges
8: $U_{ms,only} = \text{Dilate}(E^{+}_{ms,only}, SE)$
9: $E^{+}_{ms} = E^{+}_{ms} \odot U_{ms,only}$ ▷ restore edge terminals
10: $E = E_{pan} \oplus E^{+}_{ms}$ ▷ fusion
11: **end procedure**

We propose, therefore, a simple low-complexity edge fusion method based on morphological operations, summarized in the pseudo-code of Alg. 7. Edges from the panchromatic (E_{pan}) and multispectral (E_{ms}) components are extracted separately, and the latter are up-sampled and thinned (E^{+}_{ms}) to match the target higher resolution. Lines 6-7 remove MS edges close to PAN edges (within ρ high-resolution pixels), thus avoiding double edges. By so doing, however, we remove also the terminal parts of MS edges where they meet PAN edges, creating unwanted gaps. This problem is solved with lines 8-9. Eventually, the two types of edges are merged in a single high-resolution edge map. This simple solution, all based on morphological filtering, allows us to exploit edge information at both resolutions avoiding time-consuming processes. We will certainly improve it in further versions but, just like for edge detection, this will have no consequences on the overall work-flow, just on performance.

We illustrate this whole process in Fig. 2.6. In part (a) we show the PAN

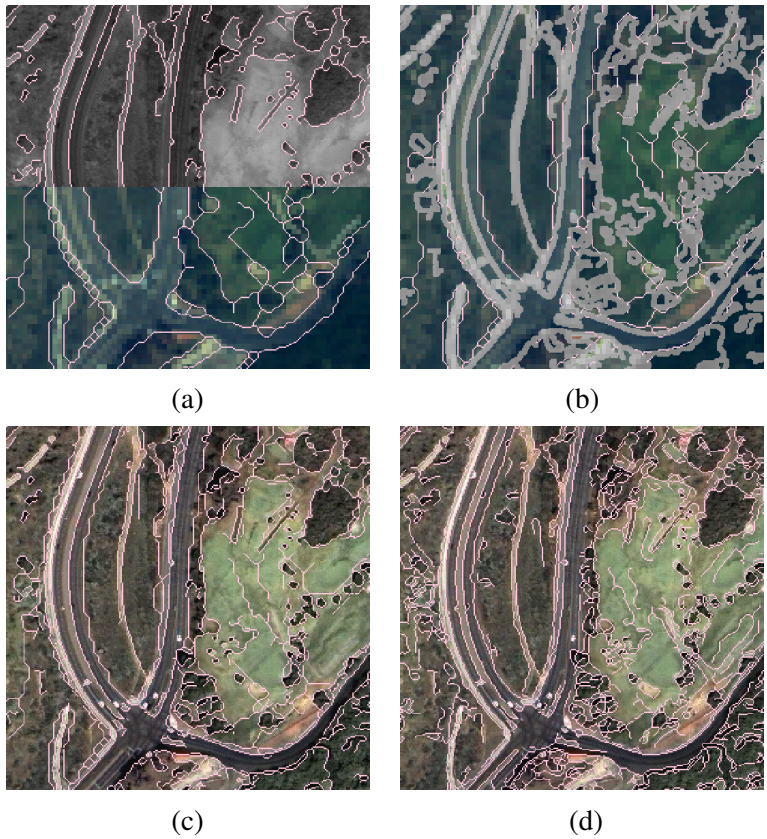


Figure 2.6: Multi-resolution edge fusion process. PAN/MS initial edges (a), final MS edges in pink (b), final edge map (c) and corresponding edges extracted on the pansharpened image (d).

with its edges in the upper half of the image, and the MS with its edges in the bottom half. Differences between the two types of edges can be appreciated especially in the transition area. In part (b) we show in light gray the U_{pan} mask obtained by dilating the PAN edges, together with the remaining off-mask MS edges. Part (c) shows the final edge map superimposed on the pansharpened version of the scene: the improvement w.r.t. both E_{pan} and E_{ms} is obvious. It is also interesting to compare this result with the edge map, shown in part (d), computed directly on the pansharpened image. Our multispectral edges are coarser than those drawn from the pansharpened image, since no post-processing is currently performed on them. Nonetheless, in many cases

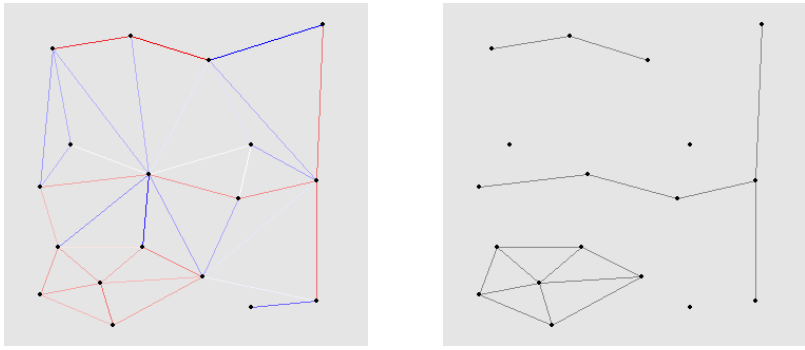
the MS edges separate more neatly objects which are spectrally close, as for example the road in the middle of the figure, where edges obtained after pansharpening appear to be more noisy and discontinuous. More in general, edges based on pansharpening appear to be more fuzzy, due to the "injection" of micro-textural information in the high-resolution product, which are unnecessary for object delineation and cause the edge detection process to linger over negligible details.

2.2 Segmenting with correlation clustering

The superpixel-level representation is an efficient way to deal with the segmentation of high-resolution images. However, it is only as an intermediate step towards higher-level segmentation, where elementary object are associated with expressive compact features, and with a suitable semantics. To move from one level (superpixels) to the other (meaningful objects) one can resort to efficient graph-based methods. The image is represented as a region-adjacency graph (RAG), with regions associated with vertices and boundaries with edges. Then, image segmentation is regarded as a graph partitioning problem [43], which can be solved through suitable optimization tools.

In recent years, several graph-based image segmentation methods have been proposed, e.g., [121, 139, 138, 44, 57, 23]. Unfortunately, their complexity grows very fast with the graph size, making them unsuited to large remote sensing images. Here, we focus on the graph partitioning model first proposed in [13] in the field of document analysis and known as *Correlation Clustering*. Although optimal correlation clustering (CC) is itself a NP-hard problem, good approximate solutions can be obtained in limited time by means of suitable greedy heuristics.

In this thesis, we propose a new image segmentation technique based on correlation clustering. Superpixels are first obtained through edge detection and edge-based watershed, and associated with the vertices of an undirected graph. We then introduce a simple method to characterize the relationships between couples of superpixels, and propose a greedy procedure to obtain a fast and accurate CC solution. Experiments on real-world remote sensing images prove the effectiveness of the proposed method.



(a) (b)

Figure 2.7: Example graph with color coded edges (a) and possible partitioning (b)

2.2.1 Correlation Clustering

Let $G = (V, E)$ be an undirected graph, where V is the set of nodes and E a set of edges, typically non complete, connecting couples of nodes. We want to partition the nodes in clusters, cutting all edges linking nodes belonging to different clusters and keeping the others. In correlation clustering [13] with each edge $e = (i, j)$ a weight is associated, w_{ij}, expressing the *correlation* between nodes i and j. Although the precise meaning of correlation depends on the specific problem, a positive[negative] correlation indicates, in general, a tendency of the linked nodes to belong to the same[different] cluster. Therefore, based on such weights, we aim at partitioning the graph so as to keep together nodes with positive correlation and separate the others. Fig.1 shows an example planar graph, with color-coded edges, red[blue] for positive[negative] correlation, with a possible partition, where only the red edges are eventually kept.

In general, with complex graphs, in higher-dimensional spaces, the solution is not so simple. Formally, the problem can be cast as a constrained energy minimization. Let x_e be the binary indicator variable specifying whether edge e is cut ($x_e = 1$) or retained ($x_e = 0$), and $\mathbf{x} \in \{0, 1\}^{|E|}$ a generic configuration of the edges. Of course, if nodes i and j are to belong to different clusters ($x_{ij} = 1$), any other node k cannot be grouped simultaneously with both i and j ($x_{ik} + x_{jk} \geq 1$). This implies a set of "transitivity" constraints that a valid partition must obey, expressed compactly as

$$x_{ij} - x_{jk} - x_{ik} \leq 0 \quad \forall\, i, j, k \tag{2.1}$$

and only the configurations respecting these constraints, $\mathbf{x} \in X_c$, correspond to acceptable solutions of the partitioning problem. If we now define the energy associated with a configuration, $\mathcal{E}(\mathbf{x})$, as the sum of the weights of all cut edges the correlation clustering problem can be eventually expressed as

$$x^{CC} = \arg\min_{x \in X_c} \mathcal{E}(\mathbf{x}) = \arg\min_{x \in X_c} \sum_{e \in E} w_e x_e \tag{2.2}$$

Note that, since the transitivity constraints are linear, the optimal graph partition can be found by resorting to Integer Linear Programming (ILP) [13]. However, for large graphs, computing the optimal solution is prohibitively complex.

2.2.2 Proposed Method

We illustrate the proposed technique with the help of the running example of Fig.2. The starting point (a) is a reliable but incomplete region boundary map, A, computed through the Canny edge detector. Due to open boundaries, this is not a partition of the image, so we perform a watershed transform applied to the distance function computed w.r.t. A, obtaining a valid superpixel representation (b). Note that this is an over-segmentation of the image, since many neighboring superpixels are homogeneous and could be reasonably merged. The superpixel map is then associated with a graph (c), where vertices correspond to superpixels, and edge to boundaries, B_{ij}, between neighboring superpixels.

Our goal, now, is to suitably merge some superpixels so as reduce oversegmentation and obtain a meaningful object layer for the image. This corresponds to clustering the vertices of V, a problem that can be solved by correlation clustering once we define reasonably the correlation, and hence the weights w_{ij}. With respect to our goal of preserving the reliable Canny edges, the closed boundary map B associated with our final segmentation will present two types of errors: missing edges (present in A but not in B) and filling edges (reverse). In general, we can attribute different weights to these errors, say $\alpha \in [0, 1]$ and $1 - \alpha$, respectively. Therefore, the decision of removing or keeping the boundary segment $B_e = B_{ij}$ has cost

$$c_e = \begin{cases} l_e o_e \alpha & (x_e = 0 \rightarrow B_e \text{ removed}) \\ l_e (1 - o_e)(1 - \alpha) & (x_e = 1 \rightarrow B_e \text{ retained}) \end{cases} \tag{2.3}$$

where l_e is the segment length, and $o_e \in [0, 1]$ its overlap with the Canny edge

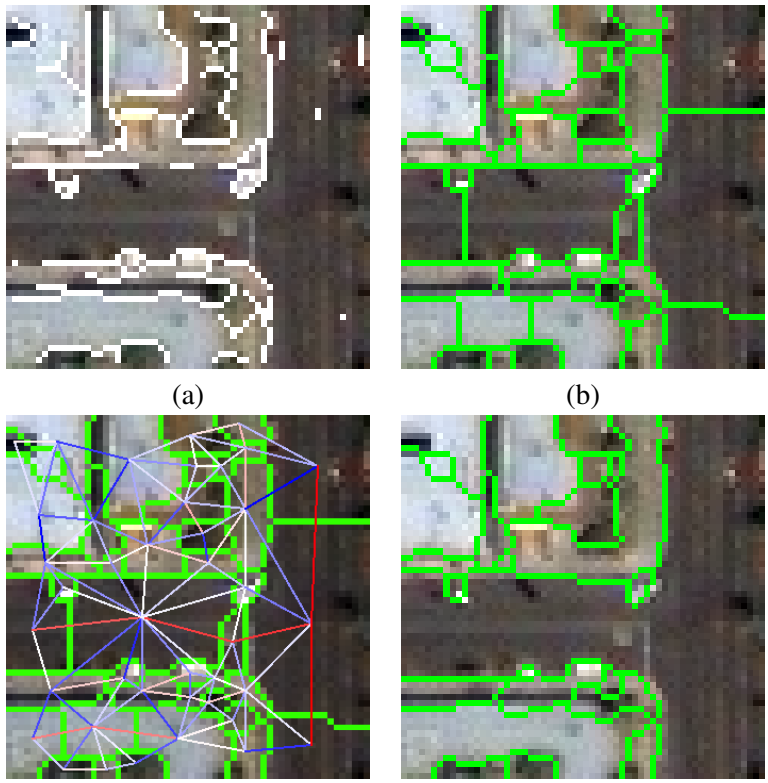


Figure 2.8: Correlation Clustering Example on a Real MS image: (a) initial Canny edge map; (b) Superpixel representation obtained through the watershed transorm; (c) Region Adjacency Graph with color-coded links, saturation indicated intensity; (d) Final CC segmentation.

map. More compactly,

$$\begin{aligned} c_e &= (1-x_e)l_e o_e \alpha + x_e l_e (1-o_e)(1-\alpha) \\ &= (1-\alpha-o_e)l_e x_e + \alpha o_e l_e \end{aligned} \tag{2.4}$$

By neglecting the last term, which does not depend on decisions, the overall cost associated with configuration $\mathbf{x}$ is therefore

$$C = \sum_{e \in E} (1-\alpha-o_e) l_e x_e \tag{2.5}$$

and therefore, the minimum-cost solution is obtained by solving the CC problem (2) with weights $w_e = (1-\alpha-o_e)l_e$. Fig.2(d) shows the optimal solution obtained through ILP.

Unfortunately, the CC problem is known to be NP hard [13], and in fact ILP becomes quickly unfeasible when the number of nodes grows, a case far too common with remote-sensing images. A simple approach [8] to reduce complexity is to delete at once all edges with negative weight and then look for connected components in the remaining graph. However, even a single edge with positive weight may cause two almost separate regions to be merged, leading to severe under-segmentation phenomena. A much better solution is to sequentially merge the pair of regions connected by the maximum-weight edge at each step, and re-compute all edges affected by this change, typically a few ones, recovering gradually the boundaries of the main regions of the image with near-optimal quality. We therefore consider this choice, also because, thanks to how the weights are defined, their updating is extremely simple.

Even the above suboptimal algorithm, however, can become very complex when the graph has high cardinality. We therefore propose a further variation, where we start from a more compact superpixel representation obtained by using a marker-controlled watershed. In particular, we use the markers introduced in [63], based on both spectral and morphological properties. This solution reduces significantly the initial number of superpixels, *i.e.* nodes, at the cost of a limited computational overhead. With large graphs, this cost is largely compensated when turning to the CC-based optimization.

2.3 Segmentation of multitemporal SAR images

In this Section, we recall briefly the principles of TS-MRF image modeling, and define the related elementary actions for unsupervised segmentation which

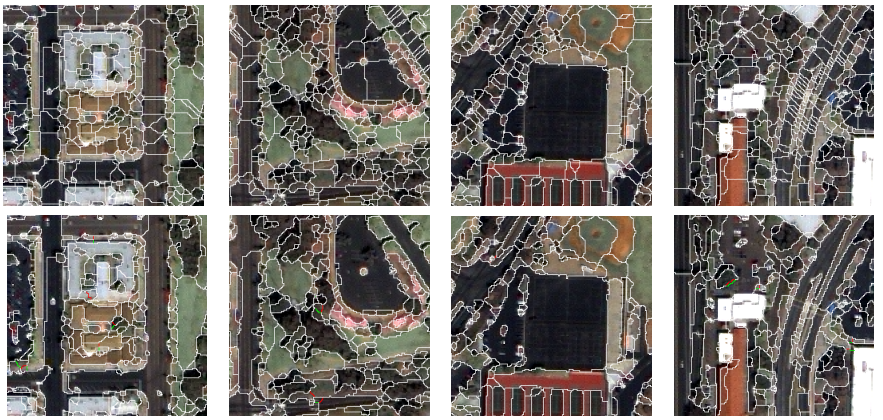


Figure 2.9: Superpixel contours (top) and CC final segmentation (bottom) for some relevant clips extracted from the test IKONOS image.

are combined to obtain the final segmentation map. This combination, and hence the detailed processing chain, was fully automated in the original TS-MRF algorithm, while here it is the user that selects at each time the actions most suitable for the data exploration goal. We refer the reader to [37, 105, 36] for a more detailed description. As will be better explained in section 3.4 we utilize the TS-MRF model in order to abtain, in an interactive way, a segmentation of a multitemporal SAR image stack.

2.3.1 MRF based modeling

In the Tree Structured Random Markov Field (TS-MRF) model, the image, defined on the set of sites $\mathcal{S}$, with observable data y, is associated with a binary unbalanced tree. Each node t of the tree is associated with a region (not necessarily connected) $\mathcal{S}^t$ of the whole image, and hence with the corresponding data y^t. To each *internal* node, a label map x^t is also associated which, for each pixel $s \in \mathcal{S}^t$ can assume only two values, $x_s^t \in \{t^{\text{left}}, t^{\text{right}}\}$, pointing at the two children nodes. Therefore, the label map of node t defines the regions associated with its children nodes $\mathcal{S}^{\text{left}} = \{s \in \mathcal{S}^t : x_s^t = t^{\text{left}}\}$ and $\mathcal{S}^{\text{right}} = \{s \in \mathcal{S}^t : x_s^t = t^{\text{right}}\}$. The root is associated with the whole image $(\mathcal{S}, y)$, and its binary label map divides the image in two non-overlapping regions. Proceeding recursively, each internal node/region is further partitioned, until the K leaves of the tree are reached which, collectively, partition the

whole image in K disjoint regions.

To this structural model, we now add a statistical model. Each label map is modeled as a binary Markov random field (MRF), with distribution $p(x^t) = \Pr(X^t = x^t)$, while the observable at the leaves are modeled as multivariate Gaussian variables, $y_s^t \sim N(\mu^t, \mathbf{C}^t)$, independent on one another given the label. As a consequence, the observables in the internal nodes are mixtures of Gaussian. However they can be also approximated as Gaussian if detailed information on the nodes is lacking (unsupervised case). In this tree-structured model, a dedicated binary MRF is associated locally with each node/region, which allows to adapt accurately to the non-stationary behavior typical of images. The non-stationarity is indeed *the* major issue in image modeling, and certainly the major limit of "flat" MRF models. TS-MRF modeling is a powerful method to address this problem.

2.3.2 Proposed interactive segmentation tool

Given the above model, the TS-MRF recursive segmentation is readily described. The fundamental action is the so called *node splitting*, while further actions, the *merge-split refinement*, and the *topological split* allow to improve overall accuracy.

Node splitting

For each node t, a binary MRF segmentation is carried out according to the *Maximum a Posteriori* (MAP) criterion

$$\widehat{x}^t = \arg\max_{x^t} p(x^t|y^t) = \arg\max_{x^t} p(y^t|x^t)p(x^t) \tag{2.6}$$

Therefore, $\widehat{x}^t$ is the most probable label map given the observables and the MRF prior at the node. Although any binary MRF prior can be adopted at the nodes, the classical Potts model is preferred for the sake of simplicity, and class parameters are estimated with a Maximum Likelihood (ML) approach. If no prior information is available (general unsupervised case), segmentation and class-wise parameters are jointly computed in a Estimation-Minimization (EM) fashion, by iteratively performing ML (given the class statistics) and MAP (given the model parameters) estimation. Refer to [37] for further details.

In the supervised case [105], a significant prior knowledge is supposed to be available, thanks to preliminary data exploration or other sources of information. In particular, the structure of the tree is known in advance, and hence

the number of leaves K, corresponding to the number of classes. Moreover the parameters $\mu^t, \mathbf{C}^t$ are supposed to be known for each class, and therefore all node likelihoods $p(y^t)$ are also perfectly known. In this setting, the only matter is the solution of the binary segmentation problems (2.6). In the unsupervised case [37], instead, all information must be estimated for each node. This includes the tree structure itself, and the number of leaves. In [37] this latter problem is solved by using an indicator, computed locally for each node, the split gain, which drives the growth of the tree by indicating at any time which leaf must be split and providing a stopping condition. Obviously, lacking any annotation of the source data, the meaning of each region singled out is not provided, and the task of associating regions to semantic classes is left to the user. It is worth underlining that, for a given number of classes, TS-MRF segmentation is computationally lighter than flat MRF segmentation.

Merge-split refinement

The exclusive use of *binary* splits represents a constraint which might impair the segmentation performance, because of the inability of the algorithm to deal with non-binary structures. In [37] a new action was added to address this problem, the merge-split refinement.

The example of Figure 2.10 illustrates a typical over-segmentation problem due to the binary constraint. Part (a) shows a synthetic image with three distinct regions, x, y and z. In some infortunate cases, due to region statistics, the first binary split may produce a segmentation, like in part (b), where region y is split between nodes 2 and 3. The further split of these nodes will produce the final 4-class segmentation of part (c), where two different nodes, 5 and 6, correspond to two adjacent parts of the same region y, a clear failure of the algorithm.

This over-segmentation problem is solved by introducing, after each split, a merge-split phase. Each newly created *child* node is tentatively merged with each of the other nodes, except the sibling, and then split again based on a local binary MRF. For each tested merge-split, a merging gain is computed. Eventually the merge-split with the largest gain (if positive) is validated. The overall effect of this action is a refinement of the boundary between the two *involved* nodes. In the bottom part of Figure 2.10 we illustrate the effect of one such merge-split action. After the splitting of node 2, we have nodes 4, 5, and 3, in part (d); the merging of nodes 5 and 3, in part (e), reassembles the over-segmented region y, while the subsequent split of the merged node (5+3), in part (f), provides the desired segmentation.

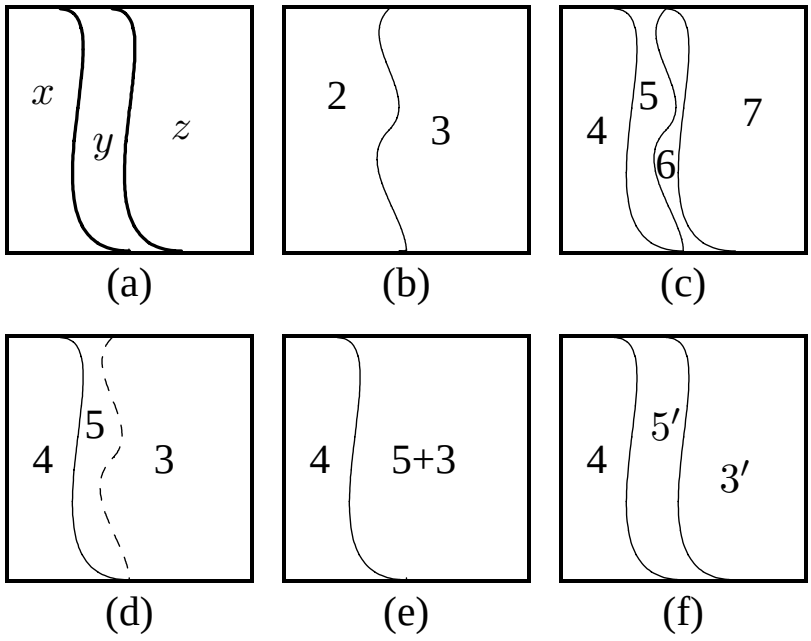


Figure 2.10: Merge-split refinement

Topological split

Another important action was introduced in [36]. The segmentation carried out by TS-MRF is clearly class-oriented. However, if the user is interested in segmentation in a more strict sense, rather than classification, there is no point in keeping a single class-wise data description, as it represents only a constraint which can impair local accuracy. In fact, separate segments belonging to the same class can have quite different statistics, especially with large and noisy images.

Therefore, when TS-MRF is used in the context of pure segmentation, aimed at building an object-level description of the image, after each binary split a topological split of the children classes follows, in which disjoint segments are assigned different labels. Each MRF split, which generates always two children nodes, is therefore followed by a topological split, which can be void, if the class is already connected, but more often generates a very large number of children. Of course, the huge increase in the number of nodes has a significant cost in terms of computational burden. On the other hand, a description of the data local to each segment cannot but improve the accuracy of

subsequent binary MRF splits.

The elementary actions described above were proposed originally for automatic segmentation, driven by suitable numerical indicators, like the split gain, the merge gain, and other node statistics. Here, they will be given to the user as basic tools, to be used interactively on the basis of a continuous inspection of results.

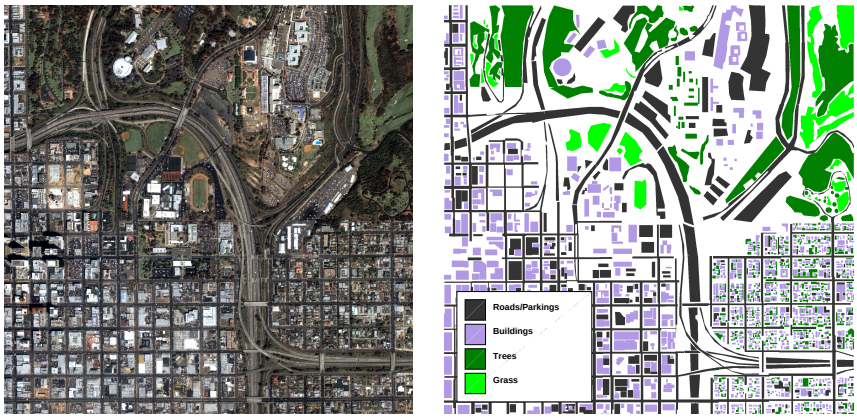
Chapter 3

Experimental results

This chapter gathers the results of several experiments carried out to numerically assess and compare the methods discussed thus far on several segmentation tasks. In particular, in Section 3.1, the MR-EMF method is compared both with all its intermediate variants above discussed and with the state-of-the-art methods for multi-scale segmentation of single resolution multispectral remote sensing images (pansharpened data, when needed). Moremore, the proposed technique is analyzed in the context of two different real-world applications. The first is the Ground Truth (GT) design problem [101], i.e. the semi-automatic generation of ground-truths for multispectral images. The second, described in Section 3.5, is the detection of environmental hazards, a work in collaboration with the Italian Aerospace Research Center (CIRA) and committed by the local authorities, carried out on a case study in southern Italy. In section 3.3 the experimental evaluation of the proposed method based on correlation clustering 2.2 is discussed. Finally, in Section 3.4, the performances of the proposed TS-MRF-based tool for interactive segmentation of multitemporal Cosmo-SkyMed SAR data are assesed and discussed.

3.1 Edge, mark and fill: evaluation

In this Section we assess the performance of the proposed EMF+ and MR-EMF algorithms on real-world single and multi-resolution images. In particular, we will compare MR-EMF with all its intermediate variants discussed in previous Sections, that is, simple watershed segmentation applied to the distance transform (WS), and EMF/EMF+ working on a single resolution image. Furthermore, we will compare performances with two state-of-the-art commercial



(a) (b)

Figure 3.1: Pansharpened RGB channels of the Ikonos image of San Diego, CA, USA, 2004×2004 pixels at 1m resolution, used in experiments (a) and the corresponding 4-class ground truth (b) obtained by photo-interpretation [65].

softwares which provide object layers for multispectral (or pansharpened) remote sensing imagery: the multi-scale segmentation technique provided by the *eCognition Developer* software [12], and the segmentation algorithm embedded in the ENVI suite [75]. Since both these techniques depend on a tunable scale parameter, we will consider multiple segmentation maps, dubbed eCognition-xx and ENVI-xx respectively, with xx representing the value of the scale parameter. Of course, the proposed techniques depend on some key parameters as well, their setting will be later discussed in detail. Finally, we will also compare results, when possible, with some state-of-the-art techniques recently published [18, 132] in the remote-sensing literature.

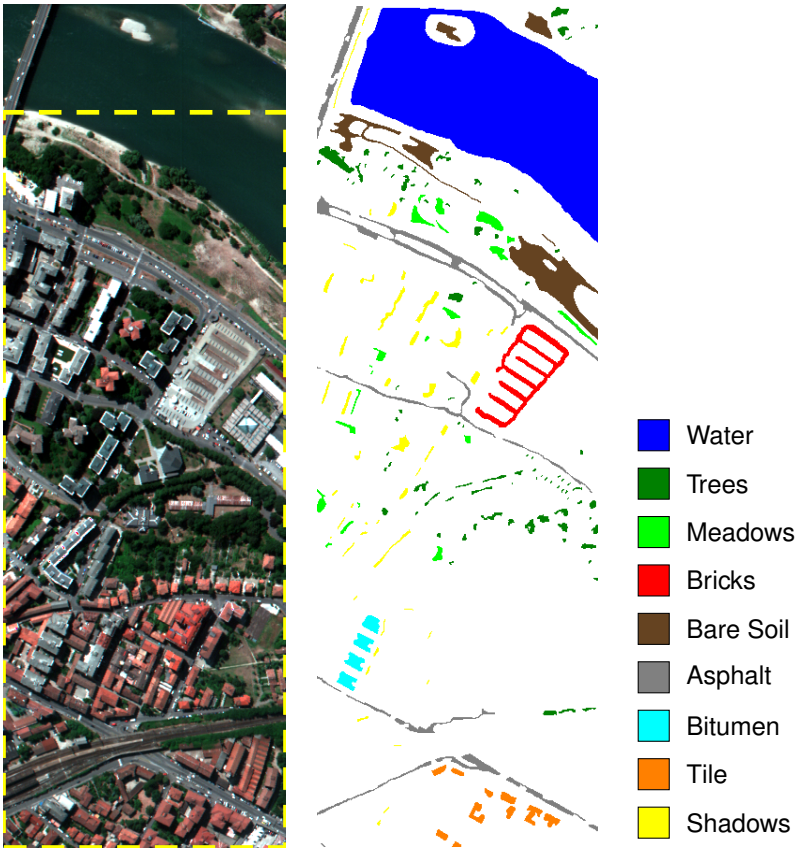
A first set of experiments is carried out on a Ikonos multi-resolution image of San Diego, CA, USA. This image is composed by a 2004×2004 panchromatic band at about $1m$ spatial resolution and four 501×501 multispectral bands in the blue, green, red and near-infrared spectrum. The RGB pansharpened version of the image is shown in Fig. 3.1(a). For this image, a handmade photo-interpreted ground truth (GT) is available, shown in Fig. 3.1(b). The original 7-class version, first used in [65], was conceived to identify objects homogeneous both in spectral behavior and spatial context, so as to assess the hierarchical texture-based segmentation algorithm there introduced.

Here, instead, we consider a more conventional spectral-based classification, and therefore generate a 4-class version of the original ground-truth by merging three couples of spectrally homogeneous classes, that is, respectively, large and small buildings, roads and parking lots, trees and green areas.

A second set of experiments is carried out on a hyperspectral image acquired by the ROSIS airborne sensor during a flight over the city of Pavia, Italy. It consists of 102 bands at 1.3m spatial resolution. The original 1096×1096 image has been cut to 300×900 (resp. 300×785) pixels to allow for a comparison of classification accuracy with the reference techniques presented in [18] (resp. [132]). A RGB representation of this image cut is shown in Fig. 3.2(a). All our algorithms, as well as the other reference segmenters, use a 4-band spectrally reduced version of the image, obtained by applying PCA to 4 contiguous sets of highly correlated spectral bands [56]. Moreover, two additional images have been generated from this 4-band version to allow the testing of our multi-resolution algorithm: a simulated multi-resolution dataset obtained by respectively averaging the four spectral bands (for the panchromatic) and downsampling them with factor 4 (for the low-resolution multispectral); and a pansharpened dataset obtained by applying PCA-based pansharpening to the simulated multi-resolution image. For the Pavia image, a 9-class ground-truth is available with the original dataset, depicted in Fig. 3.2(b).

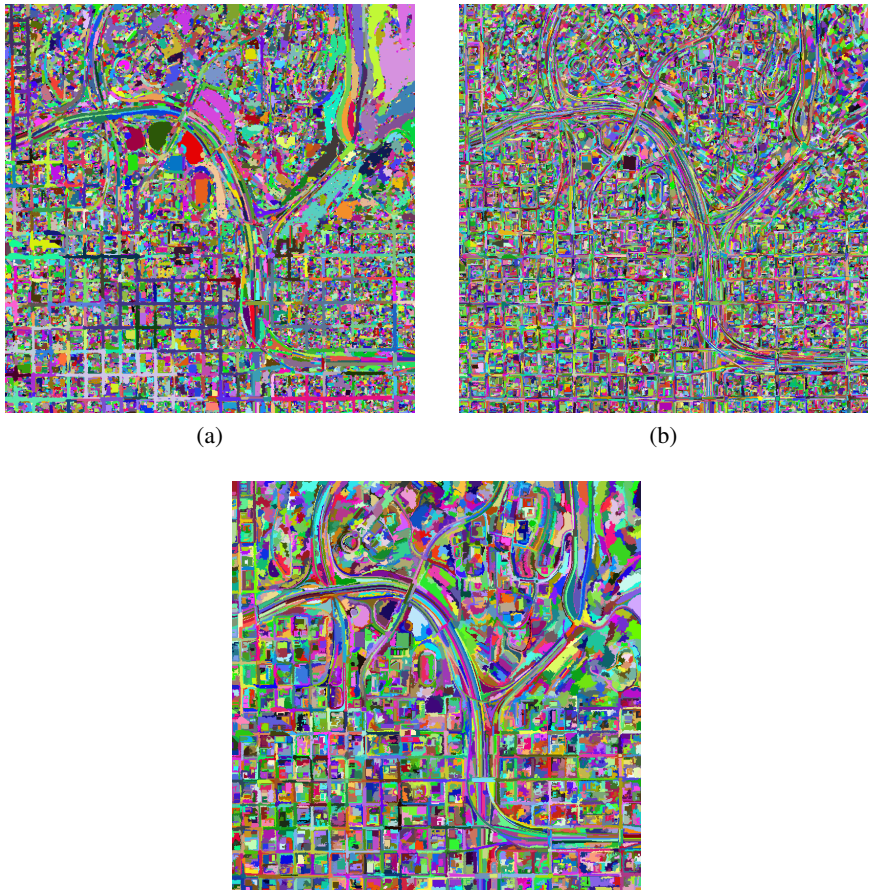
Both these ground truths are precious tools to evaluate object-layer quality. Nonetheless, they present obvious deficiencies. In order to avoid problems with mixed-signature pixels, segments with different labels almost never touch, preventing any study of boundary accuracy. Moreover, despite the fastidious work over high-detail regions, a large number of segments should be further subdivided in the Ikonos GT, depending also on the scale of the analysis. This cannot be accomplished, for such a large image, without some *ad hoc* tools for assisted GT design. On the contrary, segments in the ROSIS GT are more homogeneous but pretty sparse, especially in dense textured areas. Despite these problems, which must be taken into account in the analysis of results, these GTs will allow us to draw several interesting objective indications.

Before turning to numerical results, we show in Fig. 3.3 the segmentation maps provided for the Ikonos image by MR-EMF, and by eCognition with small (30) and large (80) scale parameter. The fixed-scale effect is striking in the eCognition maps, where the size of segments varies only within a small range, dictated by the scale parameter. Conversely, the MR-EMF map delineates objects at considerably different scales: big vegetation spots, and road segments covering most of the image coexist with a large number of smaller



(a) (b)

Figure 3.2: RGB representation of the ROSIS image of *Center of Pavia*, Italy, cut as in [18], 300×900 pixels at 1.3m resolution (a), and the corresponding 9-class ground truth (b) provided by the University of Pavia. The sub-image highlighted in (a) corresponds to the cut used for comparison with [132].



(a) (b)

(c)

Figure 3.3: Segmentation maps provided by the proposed MR-EMF technique (a), and by eCognition software with scale parameter 30 (b) and 80 (c).

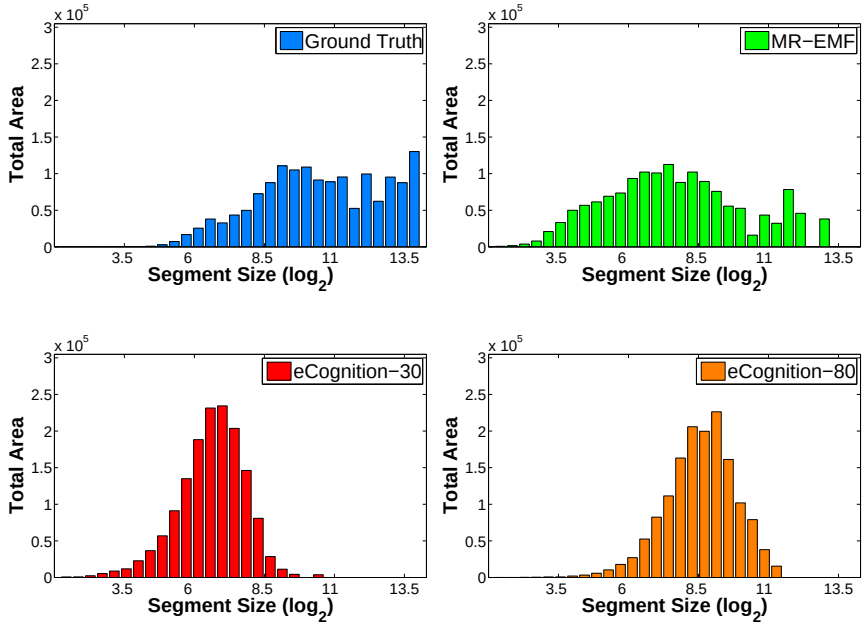


Figure 3.4: Distribution of image area by segment size.

building segments in the downtown area and, at a still smaller scale, in the fine-grain residential area.

This visual analysis is confirmed by the histograms of Fig. 3.4 showing the distribution of image area by segment size in the ground truth (limited to labeled objects) and in the three maps shown before. For the two eCognition maps the distribution peaks strongly around a specific size, while it is much more uniformly spread for the MR-EMF following more closely that of the ground truth, except for the much larger area covered by small fragments and for the huge road networks which is divided in several smaller objects.

Let us now turn to numerical assessment of performance. To this end we will consider two criteria, 1) the matching scores between GT and segmentation maps, and 2) the accuracy obtained in object-based land cover classification.

3.1.1 Object matching

Assuming the ground truth to be a faithful synthetic representation of the test image, it makes sense trying to match the segments singled out by segmentation with those outlined in the GT. Ideally, one should always observe a one-to-one perfect matching between couples of segments drawn from the two maps. In practice, due to both segmentation errors and the limited detail of the GT, matching will never be perfect, so we need some more sophisticated measures. Here, we apply the region-based metrics used in the Prague Texture Segmentation Dataset and Benchmark [95]. First of all, we mask the segmentation map to exclude the parts not labeled in the ground truth (white in Fig.3.1(b) and Fig.3.2(b)). Then, we try to match GT segments with map segments, with one of the following possible outcomes:

- *Correct Segmentation* (CS): a GT segment G is correctly segmented if and only if a map segment R exists such that the two overlap for a fraction larger than α of their respective areas (here α is set to 0.75), that is $|G \cap R| > \alpha|G|$, and $|G \cap R| > \alpha|R|$;

- *Over-segmentation* (OS): a GT segment G is over-segmented if a group of map segments $R_1, .., R_n$ exists such that $\sum_{i=1}^{n} |G \cap R_i| > \alpha|G|$, and $|G \cap R_i| > \alpha|R_i|, i = 1..n$;

- *Under-Segmentation* (US): dual to the previous case, the GT segments $G_1, .., G_m$ are under-segmented if a map segment R exists such that $\sum_{j=1}^{m} |G_j \cap R| > \alpha|R|$, and $|G_j \cap R| > \alpha|G_j|, j = 1..m$;

- *Missed Errors* (ME): a GT segment which does not fit any of the above cases is labeled as missed.

The synthetic matching scores are eventually computed as the fraction (percentage) of the cumulative *matched* area of segments with a given label (CS, OS, US) over the total labeled area of the GT. ME is instead computed as the fraction of cumulative area of missed GT segments over the total labeled area. Note that portions of the segmentation map which belong to segments labeled as CS/OS/US, but do not match the corresponding GT object(s), are excluded, which justifies the four indicators summing to less than 100%. The computation of these figures matches exactly that used in [95].

	#Obj	CS	OS	US	ME
WS	63640	2.98	94.78	0.03	0.71
EMF	60054	4.88	92.61	0.03	0.79
EMF+	46172	15.23	74.95	2.34	3.24
MR-EMF	35765	14.87	76.74	1.03	3.49
eCognition-20	57966	0.94	96.93	0.00	0.92
eCognition-30	28724	3.28	92.16	0.07	2.44
eCognition-50	11348	7.91	82.26	1.01	5.23
eCognition-80	4823	11.95	70.72	3.46	9.15
eCognition-120	2307	18.45	54.31	9.04	11.23
ENVI-30	317988	5.47	91.56	0.86	0.62
ENVI-45	212556	10.52	76.04	3.03	5.93

Table 3.1: Object matching performance for the Ikonos dataset.

Ikonos dataset

Results are reported in Tab. 3.1. A first due observation concerns the OS figure which is uniformly very high, above 70% in all cases. This is not surprising considering that our ground truth comprises just 1630 objects, while most segmentation maps count more than ten times as many segments. In practice, figures this large are mostly due to the limited level of detail of the ground truth, and should not be considered too alarming, but properly analyzed. However, they further stress the importance of visual analysis besides indicators.

Let us focus for the moment on the watershed-based techniques, namely the watershed transform based on the distance-from-edge topographic surface[1] (WS) and the EMF suite. In all cases, except MR-EMF, the pansharpened input is used.

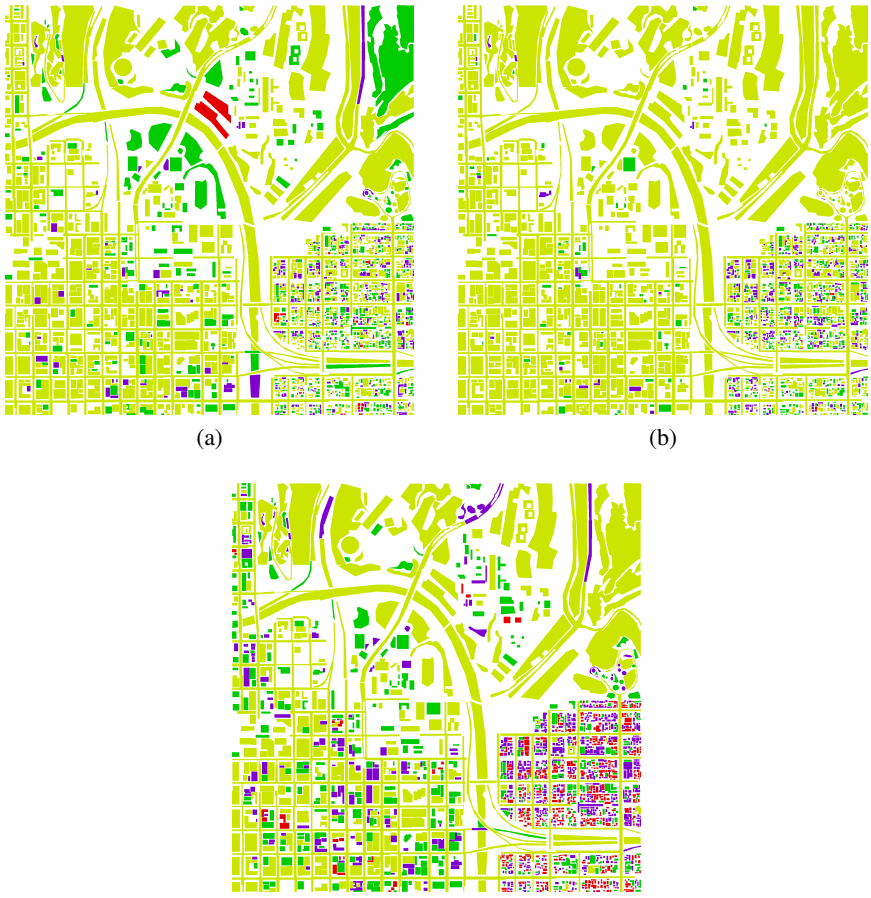
As expected, thanks to the morphological markers, EMF reduces somewhat the number of fragments w.r.t. WS, improving also slightly the CS figure. A much stronger improvement, however, is obtained in EMF+ and MR-EMF, when spectral markers are used as well. Compared to WS and EMF, techniques based on spectral markers reduce significantly the number of segments, and exhibit a much higher CS. They exhibit also larger errors, according to the

[1] In this thesis, the euclidean distance function implemented in Matlab (`bwdist`) is used, applied directly to the binary edge map, to compute the distance-from edge topographic surface.

US and ME figures, but only because of the strong over-segmentation of WS, by which US and ME tend necessarily to zero, while OS approaches 100%. Turning to the comparison between EMF+ and MR-EMF, the former, due to the presence of fuzzy contours and micro-textural details in the pansharpened data, outputs a much larger number of segments, with no benefit in terms of matching scores. By working only on the original data, MR-EMF provides much better results, despite some residual mis-alignments between PAN and MS contours.

The maps provided by eCognition with small scale parameters, 20 and 30, exhibit a clear over-segmentation, with the OS figure close to 100%. This is maybe reasonable for the first map, which has about as many objects as the WS map, but not for the second one, which has less objects than the MR-EMF map. By increasing the scale parameter to 50 and 80, the number of objects goes down rapidly, but US and ME figures grow just as fast, pointing to a still unsatisfactory segmentation. With scale parameter 120, the CS is highest, but US and ME also grow very much. In practice, only large objects are correctly recovered, while the vast majority of them, nearly 80% of the total, are undersegmented or missed altogether. The ENVI segmenter does not seem to be competitive at all for this image. Even in the best case, the output maps comprise a huge number of segments, with matching scores uniformly worse than that of MR-EMF.

The results reported in the table can be better understood by looking at the object matching maps of Fig. 3.5, relative to the same segmentation maps of Fig 3.3, where each connected component of the ground truth has been colored according with the corresponding matching label: green for CS, yellow for OS, red for US and violet for ME. All maps are characterized by widespread over-segmentation which, as explained before, depends mostly on the limits of the ground truth. However, the MR-EMF map provides also many correct matches, diffused uniformly over the whole image and at all scales. The much more critical under-segmentation is very rare, and in some instances, as for the large twin regions in the upper part of the image, is due again more to the limits of the GT than to an algorithm fault. Missed errors concentrate mostly on the residential area, due to the poor multispectral information available in this region which cannot help finding accurate contours. Conversely, both eCognition maps, as expected, seem able to match correctly only objects belonging to a given scale. With scale parameter 30, the image is almost completely over-segmented, with exceptions in the fine-grain residential area. On the other hand, scale parameter 80 allows to match some larger objects, to the benefit of



(a) (b)

(c)

Figure 3.5: Score maps associated with the segmentation maps of Fig. 3.3, obtained using MR-EMF (a), eCognition-30 (b) and eCognition-80 (c). Green = CS (correct segmentation), yellow = OS (over-segmentation), red = US (under-segmentation), violet = ME (missed error).

the CS indicator, but leads to a diffuse under-segmentation of the residential area, together with a significant number of errors due to inaccurate contours in the pansharpened image.

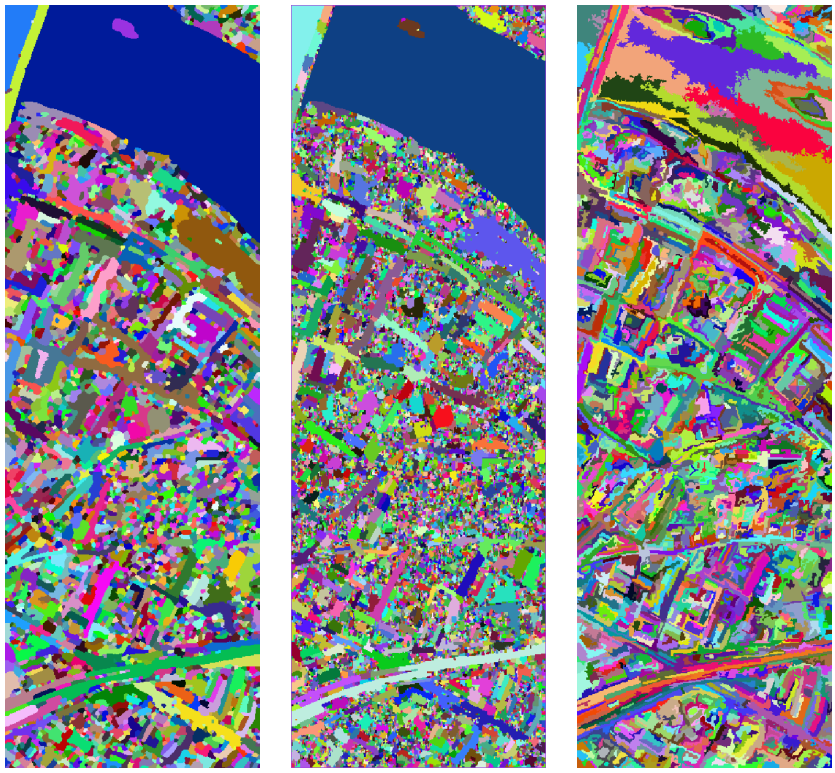
ROSIS dataset

	#Obj	CS	OS	US	ME
WS	4919	3.46	95.39	0.12	0.77
EMF	4706	66.72	32.07	0.12	0.79
EMF+/full	3642	74.54	17.91	3.27	2.69
MR-EMF	3177	73.26	19.90	1.22	3.78
EMF+/pansh	3815	74.79	20.66	1.54	1.47
eCognition-100	10871	1.86	96.79	0.25	0.66
eCognition-150	5224	4.75	92.50	0.74	1.08
eCognition-200	3184	6.14	89.44	0.90	2.55
eCognition-250	2184	7.20	87.38	1.07	3.60
eCognition-300	1637	7.18	83.52	1.51	6.93
eCognition-400	978	7.32	79.75	4.14	7.88
eCognition-500	659	7.49	77.34	5.86	8.38
eCognition-1200	132	59.20	9.20	11.42	12.38
ENVI-30	18612	65.04	33.89	0.08	0.34
ENVI-40	14949	72.31	23.92	0.51	1.09
ENVI-50	11350	76.83	13.75	3.49	4.18

Table 3.2: Object matching performance for the ROSIS dataset, full resolution 4-band version, except for MR-EMF (synthetic MR) and EMF+/pansh (synthetic MR + pansharpening).

Object matching scores for the Pavia dataset are reported in Tab. 3.2. The ground truth of Fig. 3.2(b) puts in even more evidence the existence of objects at strongly varying scales, from small trees and thin roads to the very large river segment. Evidently, this is the cause of the significant gaps in the CS/OS figures among techniques which correctly segment larger objects and those who over-segment them.

The automatic marker generation strategy proposed in this work confirms as a rewarding choice. Passing from the distance-based watershed (WS) to the

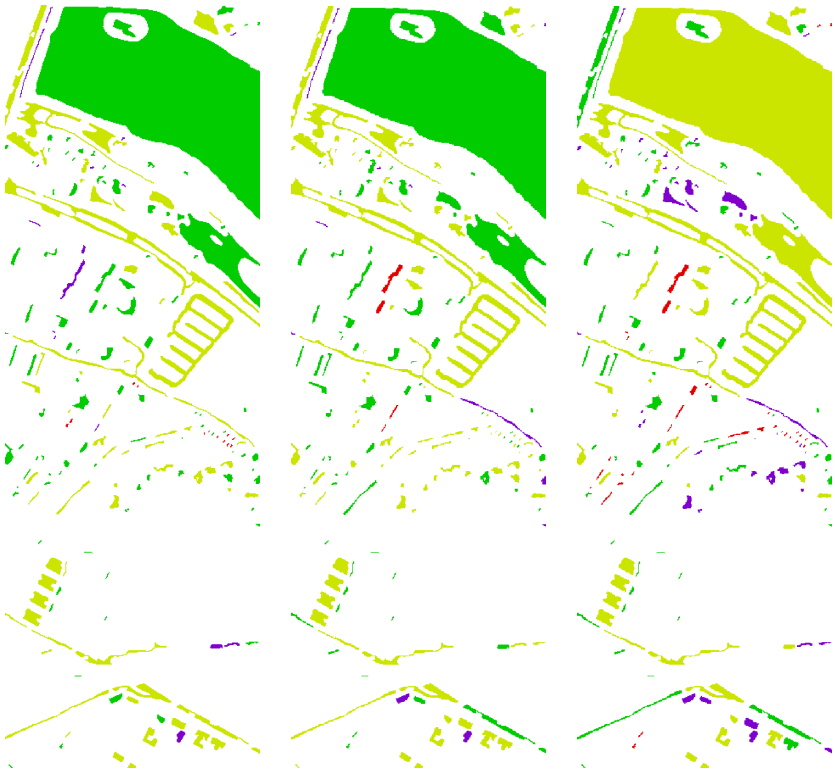


(a) (b) (c)

Figure 3.6: Segmentation maps provided by EMF+/full (a), ENVI-40 (b) and eCognition-250 (c).

basic EMF, there is a huge improvement in terms of CS, with no appreciable increase in US and ME. This is mostly due to the ability to recompose some large objects (notably the river) which had been oversegmented due to minor irregularities. Spectral marking on the full 4-band image further improves the overall quality of segmentation, with a gain of about 5 percent point in CS at the price of a slightly higher error. It is also worth underlining that MR-EMF provides results almost as good, though working on the lower-quality synthetic multiresolution dataset, and with a significantly shorter computational time.

Turning to the reference segmenters, it is interesting that, for this image, ENVI outperforms clearly eCognition. This is clearly related to the strong scale variability, a situation where a fixed-scale algorithm cannot work properly. At low scales, large objects are consistently oversegmented, while OS and



(a) (b) (c)

Figure 3.7: Score maps associated with the segmentation maps of Fig. 3.6, obtained using EMF+/full (a), ENVI-40 (b) and eCognition-250 (c).

ME are already significant. Only with a scale parameter of 1200 large objects are eventually captured, but then under-segmentation and errors are unacceptably large. The negative behavior of the fixed-scale approach is confirmed by the inspection of both the segmentation map, Fig. 3.6(c), where false contours appear in larger areas, and the object score map, Fig. 3.7(c), with almost all small objects mis-segmented. The gradient-driven watershed based ENVI segmenter provides better numerical results, taking advantage of the significant presence of classes with highly homogeneous objects, like shadows, or water. The best score is obtained with a scale parameter of 40, and is very close to the score as EMF+, as made obvious by comparing the two score maps of Fig. 3.7(a) and (b). However, since the ENVI segmenter inhibits region growing in highly textured areas, it singles out many more segments than EMF+,

15000 vs 3600. Therefore, large parts of the image are indeed oversegmented by ENVI (mainly in areas not covered by the ground truth), as clear by comparing the segmentation maps in Fig. 3.6(a) and (b).

3.1.2 Classification

Land cover classification is one of the main tasks performed on remote-sensing imagery. Traditionally, it is carried out independently on each pixel, but one can instead take advantage of the available object layer assigning a unique class to each object. This approach has a number of advantages: by working on extended objects, one reduces the impact of noise on accuracy, especially relevant for some imaging modalities, like SAR [66]; geometrical features can be taken into account to improve classification; interactions among objects can be studied to characterize complex scenes, as done in [65]; last but not least, working on objects rather than pixels reduces the number of processing atoms, thus limiting computational complexity.

For our test images, since a ground truth is available, we can compute class-wise features on a training subset and resort to *supervised* classification. A simple and robust way to exploit objects consists in classifying first each pixel of the source image independently, and then labeling each object based on a majority vote. Note that this approach only exploits the object layer *a posteriori* w.r.t. the classification, hence we refer to this process as *object-based regularization* of supervised pixel-wise classification.

A direct object-based classification scheme has also been put in place, which relies on a simple unsupervised feature extraction strategy recalling the one used in [4]: pixels from the multispectral input are first clustered using k-means ($k = 25$), then each object is characterized by the spectral mean computed over the pixels belonging to the largest cluster. This solution, though robust to the presence of a few outliers, is still more error prone than the previous scheme based on pre-classification. A more complex region-based modeling could overcome this problem, but would require a more accurate object-based ground truth and, in any case, this issue lies outside the scope of this thesis.

Classification accuracies have been assessed by analyzing the confusion matrix A, where entry a_{ij} is the number of ground-truth pixels of class j that have been classified as belonging to class i. Sums along rows a_{i+} and columns a_{+j} give the number of pixels belonging to each class, in the map and in the ground truth, respectively.

Based on confusion matrices, several global quality indicators are usually computed. The *overall accuracy* (OA), defined as $\tau = \sum_i a_{ii}/N$, is the per-

centage of sample pixels that are correctly classified. The *Kappa parameter*, defined as $\kappa = (N \sum_i a_{ii} - \sum_i a_{i+}a_{+i})/(N^2 - \sum_i a_{i+}a_{+i})$, discounts successes obtained by chance, and is therefore more conservative (it can be also negative). The average accuracy (AA), also frequently used, is defined as the mean of per-class *producer's accuracies* a_{ii}/a_{+i}. Finally, the normalized accuracy τ^{norm} [32] is computed on a confusion matrix modified in order to give equal importance to all classes, irrespective of the number of samples in each one.

Ikonos dataset

For this test image, about 15% of ground truth objects have been used to train a pixel-wise classifier based on the Maximum Likelihood (ML) estimator. Class-wise probability distributions are assumed to be Gaussian.

Tab. 3.3 shows the confusion matrix for pixel-wise classification. A significant mis-classification rate is observed between the *Roads* and *Buildings* classes and between the *Trees* and *Grass* classes.

	Roads	Buildings	Trees	Grass	Total
Roads	**649849**	129193	2727		781769
Buildings	87782	**327053**	7586	116	422537
Trees	34609	18605	**339025**	7380	399619
Grass	14	25	7809	**81196**	89044
Total	772254	474876	357147	88692	

Table 3.3: Confusion matrix for pixel-wise classification of the San Diego (IKONOS) image.

Tab. 3.4 shows, instead, the confusion matrix obtained with object-based regularization using the MR-EMF segmentation map. By comparison with the ML confusion matrix, it is clear that mis-classifications is much reduced, especially for the *Roads* class, which the object layer generally disentangles from the detail-rich urban area.

Synthetic indicators are reported in Tab. 3.5. First of all, it is clear that object-based regularization improves performance significantly. Overall accuracy τ, for example, grows by almost three percent points, from 82.53% for pixel-wise ML to 85.30% for MR-EMF. Among all object layers, MR-EMF scores uniformly best under all measures, followed closely by EMF+ which

	Roads	Buildings	Trees	Grass	Total
Roads	**690419**	129181	4070	3	823673
Buildings	62254	**326832**	5124	43	394253
Trees	19515	18857	**342997**	4707	386076
Grass	66	6	4956	**83939**	88967
Total	772254	474876	357147	88692	

Table 3.4: Confusion matrix for the classification of the San Diego (IKONOS) image with object-based regularization (MR-EMF map).

loses some accuracy due to prior pansharpening. The last column reports the overall accuracy, τ', obtained with direct object classification. As expected, results are uniformly worse than those obtained with majority-vote regularization, but the loss is contained within 1-2 percent points.

	τ (OA)	κ	AA	τ^{norm}	τ' (OA)
ML	82.53	73.79	84.87	86.05	82.53
WS	84.76	76.97	86.39	87.98	84.34
EMF	84.83	77.07	86.68	88.16	84.41
EMF+	85.19	77.58	86.91	88.49	**85.01**
MR-EMF	**85.30**	**77.80**	**87.23**	**88.72**	84.89
eCognition-20	84.16	76.16	86.01	87.60	83.92
eCognition-30	84.28	76.35	86.16	87.85	83.99
eCognition-40	84.42	76.54	86.16	87.89	84.21
eCognition-50	84.34	76.42	86.10	87.79	84.09
eCognition-60	84.39	76.50	86.10	87.76	84.08
eCognition-80	84.75	77.02	86.42	88.10	84.30
eCognition-120	83.88	75.68	85.50	87.09	84.78
ENVI-30	83.84	75.67	86.09	87.40	83.63
ENVI-45	84.28	76.32	86.99	88.07	83.42
ENVI-50	79.90	69.39	76.21	82.96	76.78

Table 3.5: Classification accuracy indicators (percent) for the San Diego image.

ROSIS dataset

For the *Center of Pavia* test image, pixel-wise classification has been obtained using a support vector machine (SVM), since it is the pixel-wise classifier used in [18] and [132], where some classification techniques are proposed, used as further reference for performance comparison. In these papers, slightly different cuts of the image have been used, hence we designed two different SVM classifiers, choosing the training sets in order to both match the number of per-class samples and approximate the pixel-wise accuracies reported in the papers. In Fig. 3.2(a), the smaller cut used in [132] is highlighted in the yellow dashed box. Note that pixel-wise SVM classification has been performed on the original 102-band dataset.

Like in the previous case, and despite the differences in source characteristics and classification engine, object-based regularization generally improves the overall classification accuracy, τ, with a 2-3% gain w.r.t. basic pixel-wise processing, already quite good. Besides this, the most remarkable result is the flatness of performance across all methods, except for scale-dependent techniques when scales too large are used. This is arguably due to the available ground truth, quite sparse over the image except for a few very large objects. Indeed, plain watershed seems to be the best solution in this case. The proposed techniques, insensitive to the scale issue, perform always among the best. Reference techniques proposed in the literature also provide similar results. The marker-controlled RD-MSF [18] is somewhat poorer because it does not use edge information to preserve fine details. A better performance is obtained by MSF [132], where pre-classified markers are improved through a morphological erosion (i.e., a local process). In fact, using the same markers to select a segmentation map in the HSEG stack [133] (M-HSEG^{op}) works equally well.

With direct object classification (last two columns) there is a sharper decline in overall accuracy, τ', w.r.t. object-based regularization, around 2-3 percent points, making it comparable to pixel-wise classification. Besides the intrinsic vulnerability to outliers, already observed for the San Diego image, in this case we also have a relatively small training set (see [18]), with a limited number of samples per class, which probably does not allow for a correct design of the object-based classifier.

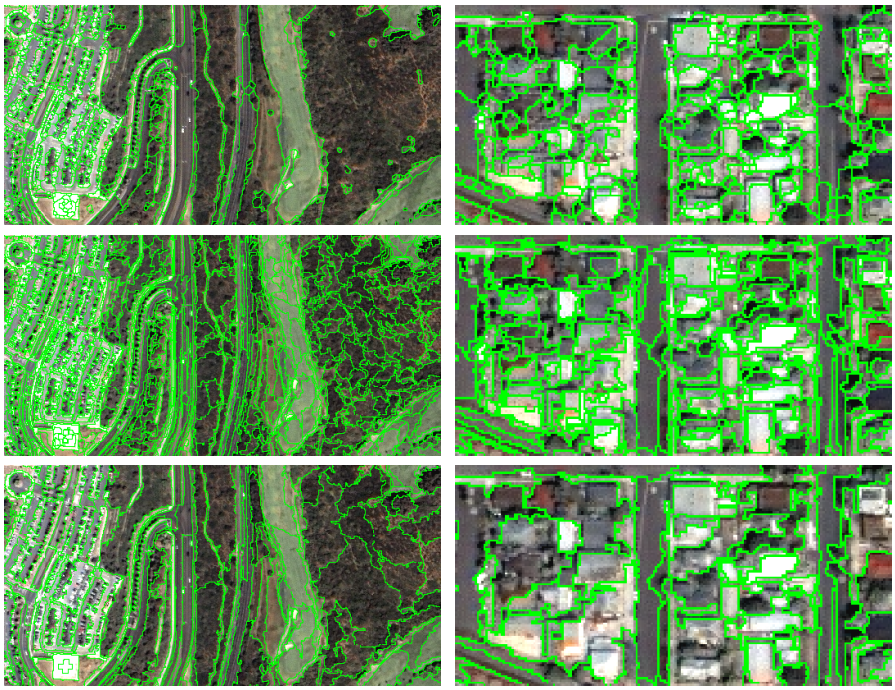
	τ(OA)		τ'(OA)	
	Cut #1	Cut #2	Cut #1	Cut #2
SVM (pixel-wise)	95.63	95.14	95.63	95.14
WS	97.87	98.05	96.02	**96.76**
EMF	97.87	97.69	**96.07**	96.39
EMF+/full	**98.16**	97.81	95.29	95.69
MR-EMF	97.24	97.30	94.17	95.70
EMF+/pansh	97.59	98.01	95.57	96.17
eCognition-100	97.24	97.35	95.32	95.35
eCognition-150	97.68	97.70	95.52	95.17
eCognition-200	97.79	97.63	95.49	94.78
eCognition-250	97.79	97.18	94.32	94.49
eCognition-300	97.56	97.02	94.40	94.16
eCognition-400	97.39	97.00	94.52	93.48
eCognition-500	97.56	96.18	94.84	92.84
ENVI-30	97.44	97.17	95.68	95.29
ENVI-40	97.75	97.01	95.91	94.61
ENVI-50	97.06	91.29	94.18	83.07
RD-MSF (L1) [18]	97.17	–	–	–
MSF (SAM) [132]	–	**98.13**	–	–
M-HSEG^{op} [132]	–	98.00	–	–

Table 3.6: Classification accuracy results (percent) for the Center of Pavia image.

3.1.3 Visual inspection

We complete our analysis with an accurate visual inspection of some sample results, which allows us to gain insight into aspects, like contour accuracy, that cannot be measured with the help of the ground truth, and more in general to appreciate phenomena hardly captured by numbers.

In Fig. 3.8 we show two areas of the Ikonos image with superimposed green contours corresponding, respectively, to the MR-EMF (top), eCognition-30 (middle), and eCognition-80 (bottom) maps. The superior ability of MR-EMF to adapt to the local image scale is immediately clear. The eCognition-30



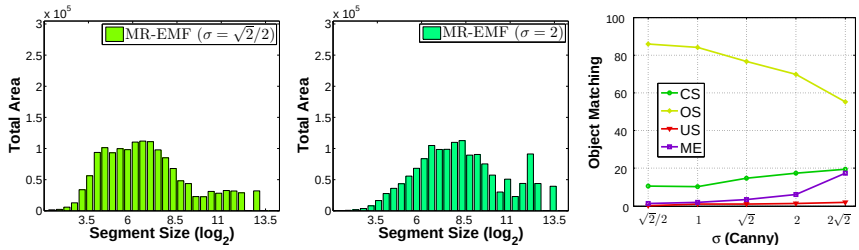
(a) (b)

Figure 3.8: Results for MR-EMF, eCognition-30, and eCognition-80 (top to bottom) over a rural area (a) and a dense urban area (b) of the Ikonos image.

map exhibits always a strong over-segmentation, except for the small buildings within the blocks in the residential area (b). The eCognition-80 map reduces over-segmentation, but only to a limited extent, since the fixed scale prevents the extraction of large objects. In addition, it loses a large number of fine details, especially in the parking lots and on small buildings. As for MR-EMF, the large-scale forest spot in (a) is correctly delineated, and quite good results are obtained also on the roads and parking lots (b). At the same time, most of the details are correctly preserved in both transition areas and regions with sparse vegetation, and especially in the smaller-scale dense urban scene.

Further experiments have been carried out to validate the proposed approach on images acquired with different sensors. In Fig. 3.9, we show two details of a multi-resolution Worldview image of a dense urban area in the town of Maddaloni (Italy), composed by eight 450×450 multispectral bands (2.4-m resolution) and a 1800×1800 panchromatic band (0.6-m). Segmen-

Figure 3.9: Results for MR-EMF, eCognition-20, and eCognition-50 (top to bottom) over some areas of a WorldView image.



(a) (b) (c)

Figure 3.10: Robustness of MR-EMF vs. σ (Ikonos dataset). Distribution of area by segment size for $\sigma = \frac{1}{2}\sqrt{2}$ (a), and $\sigma = 2\sqrt{2}$ (b). Matching scores as a function of σ (c).

tation results are represented as usual by green contours superimposed on the pansharpened image, both for the proposed MR-EMF technique (left) and for eCognition with scale parameters 20 (middle) and 50 (right). As with the Ikonos image, MR-EMF works pretty well at all scales, isolating correctly most of the buildings and small roads, only slightly over-segmented, as well as large-scale objects like the parking lot in the left image or the tree garden in the right image. Scale adaptivity is again a major problem with eCognition, with over-segmentation of large objects when tuned on buildings (middle) and diffuse under-segmentation when a larger parameter is used. Moreover, in this case the image comprises 8 low-resolution bands, instead of 4, which makes pansharpening more prone to generating artifacts over object boundaries. As a consequence, all maps provided by eCognition show an evident "staircase" effect on the boundaries, contrarily to MR-EMF maps where edge accuracy is mostly preserved by giving priority to panchromatic edges. Robustness with respect to the number of multispectral bands is a further qualifying point for the proposed technique.

3.1.4 Parameters setting and analysis of robustness

Implementation of the EMF algorithms calls for several design choices, and the setting of a few parameters. Sometimes they can be chosen in advance in a sensible way, in other cases, a preliminary analysis of robustness has been carried out.

The margin for seed dilation in EMF must be a small integer $\epsilon > 1$ to avoid edge crossing with 8-connectivity. Performance depends very weakly on this

parameter, which is eventually set to 3. The threshold for PAN/MS domain discrimination is naturally set to $\lambda = \rho$, always 4 in our images. Therefore, objects thinner than 2ρ PAN pixels are associated with the PAN domain, as they might happen not to include any "pure" MS pixel. As for spectral markers, they are generated only for regions with activity index larger than a given threshold T_η, with activity computed as the sum of the band-wise data variances. Too large a threshold might lead to under-segmentation. On the other hand, by lowering it too much, TS-MRF ends up working on homogeneous regions, producing many small segments that are removed anyway by WME. Based on preliminary experiments, a reasonable rule is to set T_η so that only a fraction $\eta = 0.01$ of all the regions are classified as active. This is not a small value, considering that the vast majority of off-edge connected domains on which spectral markers are extracted are small and homogeneous, while the few active ones account for a large fraction of the image (e.g., around 30% for the Ikonos dataset). Pansharpening, when necessary, is carried out by the PCA-based technique implemented in ENVI.

A key step of our algorithms is edge detection, which can be expected to impact significantly on performance. As already said, we use the Canny algorithm, in the implementation of MATLAB version R2012b, since it guarantees a good performance especially when fine details must be preserved. We use the default parameters, $\sigma = \sqrt{2}$, and low and high thresholds set automatically based on the percentage of non-edge gradient points, 70% by default. However, they were selected only after extensive preliminary experiments. As an example, we discuss the effect of varying σ, but a similar behavior, with obvious differences, is observed by varying the thresholds. The analysis is carried out for σ going from $\frac{1}{2}\sqrt{2}$ to $2\sqrt{2}$, with multiplicative step $\sqrt{2}$. Smaller and larger values make clearly no sense for our problem. By modifying the intensity of smoothing, σ acts as a scale parameter, favoring the detection of small (large) objects when it is itself small (large). This appears clearly in the histograms of Fig. 3.10(b) and (c), where the distribution of segment size changes as expected with the parameter. In all cases, however, it follows the distribution computed on the ground truth much better than fixed-scale eCognition (see Fig.3.4). More conclusive results are shown in Fig. 3.10(c), reporting the object matching scores as a function of σ. The performance is pretty stable in a relatively large range around $\sqrt{2}$, where the critical US/ME figures remain quite low, and one only must cope with the compromise between correct segmentation and over-segmentation. At larger values of σ, errors increase, confirming that for remote sensing images, characterized by neat edges and rich

micro-textures, an intense smoothing is not advisable, as it may filter away fine-detail information.

3.1.5 Computational complexity

All the proposed algorithmic suite (EMF/EMF+/MR-EMF) has been developed in Matlab, with the exception of few modules implemented in C++ and ported in Matlab (watershed transform [20] and TS-MRF based unsupervised segmentation [37]). By relying mostly on morphological operations, the algorithm turns out to be quite fast, also compared with the other reference techniques: the MR-EMF segmentation of the multi-resolution image of Fig. 3.1 completes in about 40 seconds on a Intel® Core® I7-3537U CPU 2.00 Ghz, a satisfying performance compared to the 15 seconds necessary (after pansharpening) with the commercial softwares eCognition and ENVI, which benefit from an optimized implementation. A more detailed analysis shows that a large fraction of the overall CPU-time is spent on processing steps that are not specific of the proposed method, mostly TS-MRF segmentation[2] (about 85%), and Canny edge detection (about 10%). Replacing these tools with faster ones would impact significantly on speed.

Note also that, when multi-resolution data is processed, EMF+ on the pansharpened image is much slower than MR-EMF, taking about 120 seconds on the Ikonos image, mostly because of the marker segmentation step, carried out on 4 high-resolution pansharpened bands. This further supports the choice of multi-resolution processing adopted in MR-EMF, which provides better results in a much shorter time. The Matlab code for our algorithms, along with the pre-compiled *mex*-files for C++ modules, is available on the GRIP website, at `http://www.grip.unina.it`.

3.2 Ground truth design via EMF

3.2.1 Problem overview

An unprecedented wealth of remote-sensing images is available nowadays, opening the door to a large number of valuable applications, like urban planning [2], land use management [113], environmental crime detection [51, 50], etc. To analyze this growing bulk of data, however, one is forced to use automatic tools, since expert photo-interpreters are both rare and expensive. On

[2] An analysis of the computational complexity for TS-MRF based segmentation has been conducted in [37].

the other hand, no automatic tool can replace the skill and expertise of trained humans, calling for some form of human-computer interaction [85, 63], and in particular for the use of supervision in the design phase of automatic and/or interactive tools.

Indeed, supervised design, when not strictly mandatory, keeps providing a significant benefit in terms of algorithm performance. Very often, the supervision includes the design of some suitable ground truth (GT) data, necessary to train the automatic algorithms. Of course, the availability of GT data is also of paramount importance when an algorithm is first designed and validated. Under this point of view, the remote sensing world suffers a tremendous gap w.r.t. other fields, where the widespread diffusion of rich datasets and benchmark sites, together with the good practice of reproducible research, allows for faster design and validation. The Prague remote sensing segmentation benchmark [95, 96] is a notable exception. However, it uses synthetic image mosaics, with simple GTs produced by an algorithm and revealed to the user. In practical cases one has just a single class of images to work on, without annotations. Being able to extract automatically and in a reasonable time a reliable GTs to use for training and validation becomes therefore a major issue.

The manual design of a detailed GT may be a painstaking task, taking many hours of precious man-power. Moreover, it is an error-prone process, since high-dimensionality data cannot be easily visualized. In the simplest cases, one may just select and classify a few spot regions of interest for the intended task, for example some homogeneous regions used to train a point-wise classifier. GTs of this kind, however, are very limited in scope and, in general, do not allow for a good validation of results. Indeed, one should be able to label a whole image, or a large part of it, such to be used as a reliable guide for subsequent design, rejecting just a few critical areas. In classification, for example, conventional GTs do not include areas near region boundaries (see for example the Pavia datasets) to avoid the ensuing uncertainties. However, these are exactly the situations where the performance of classifiers may differ more significantly, and experimental results on such areas would allow one to select the most reliable tool.

In this thesis, we propose a simple and effective interactive framework (already published in [90]) aimed at assisting a photo-interpreter in the design of large and detailed GTs for remote sensing images. Here, we consider classification as the final goal, but the process can be obviously tailored to other tasks. Following the framework first proposed in [101], the image is preliminarily segmented, then the interpreter is presented with spectrally homogeneous re-

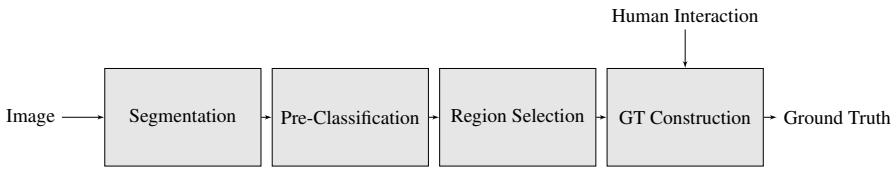


Figure 3.11: Proposed Ground Truth Design Framework

gions, with synthetic information attached, which can be labeled as they are, further refined, or just rejected.

Contrary to [101], we aim for GT that guarantee a large coverage of the image. For this reason, great emphasis is given to the initial segmentation tool, which must be fast, reliable both on inner regions and borders, and able to process various kinds of input data, including the multi-resolution data typically provided by most modern sensors. To this end, we resort to the recently proposed Multi-Resolution Edge Mark and Fill (MR-EMF) algorithm [93, 63]. which possesses all these properties and is freely available online[3]. A further desirable property of MR-EMF, especially in complex scenarios, is its ability to provide segments in a wide range of scales.

Turning to region selection, in order to reach quickly a good coverage of the image, larger segments should be considered first. However, since only a fraction of all segments are typically analyzed, this choice may lead to biases in favor of classes with large segments, and to the absence of representatives for some classes. To guarantee a more "fair" sampling, in [101], segments are clustered based on their spectral content, creating "buckets" that are sampled in round-robin modality. However, it may easily happen that two or more classes share similar spectral characteristics, but comprise segments of wildly different size, like for example building rooftops and street network. In this case the small-size segments would still be penalized. To deal with this problem, we discriminate also with respect to the size, defining and sampling a two-dimensional array of buckets.

3.2.2 Proposed solution

The easiest way to construct a GT by photointerpretation requires the iteration of two elementary steps: 1) carefully delimiting a homogeneous region of the image and 2) assigning a label to it. Drawing manually the region boundaries

[3] `http://www.grip.unina.it`

is all but trivial, especially when considering large images (think of last generation sensors like GeoEye or WorldView) with fine-detail regions, like in urban areas. A human operator may spend many hours on this task, with declining attention and hence a large probability of error. Moreover, results may depend strongly on the experience of the human operator.

Here we propose a semiautomatic tool for GT design where a human operator is only asked to inspect a candidate segment and decide whether to accept it as a GT sample or not. The general scheme is depicted in Figure 3.11. At the core of the proposed scheme is an automatic segmentation engine which extracts meaningful candidate segments. Next, an unsupervised classifier creates clusters of segments homogeneous in terms of both spectral characteristics and size. Subsequently, all cluster are visited, with a suitable schedule, and each segment is shown to the operator who first verifies its accuracy and then, in the positive case, associates a label with it. Clustering and region selection have the critical role of ensuring a balanced composition of the ground truth, which must be representative of all classes and region sizes.

Segmentation

In order to build quickly a representative GT, with as few iterations as possible, the segmenter should:

a) **avoid undersegmentation**: segments spread over multiple classes are harmful for subsequent applications, they should not be presented to the operator to avoid errors and to save analysis time;

b) **limit oversegmentation**: these errors cannot be avoided, especially given the previous constraint; however they reduce only the efficiency of the process and only marginally the quality of the GT;

c) **provide interscale segmentations**, where all types of object, small and large, are selected at their intrinsic scale (*e.g.,* forests and buildings).

Motivated by the above considerations, we have chosen the recently proposed MR-EMF algorithm [93, 63]. It meets very well the above requirements and, in addition, allows one the deal equally well with single-resolution and multiresolution data (like Ikonos, GeoEye, WorldView), without the need of pansharpening, and exploiting all data at their full spatial and spectral resolutions.

Pre-classification and selection

To achieve the target image coverage with the smaller possible number of iterations one may be tempted to give priority to the largest segments. However, this may penalize classes composed mostly of small objects (for example buildings) which could be eventually under represented in the GT, when covering a significant part of the image. Moreover, some classes may occur more frequently than others. In order to balance class coverage and speed we create subsets of fragments which are homogeneous in terms of both spectral signature and size. Then, the selection scheme will draw candidates from all subsets at the same rate, in a predefined order. For each subset, however, the largest remaining segment is chosen first. The subsets are created with a simple k-means over the spectral features (average spectral signature of the segment) followed by a k-means over the segment size for each of the previous sets. As a rule of thumb, we used in both clustering processes a value of k equal to the number of expected classes.

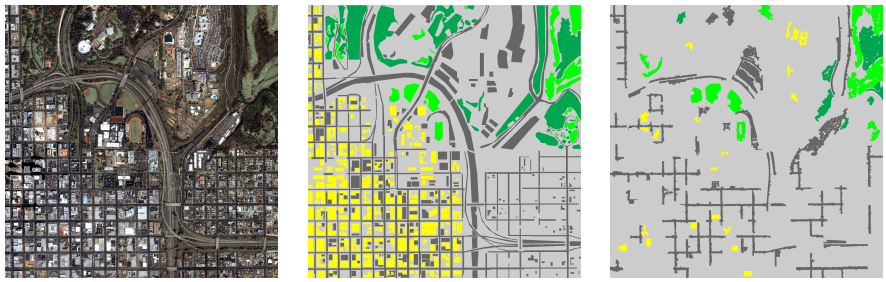
Interaction

In the last step of each iteration the operator is asked to check whether the candidate segment is acceptable or not, rejecting those showing undersegmentation or not belonging to any class of interest, and labeling the accepted ones.

3.2.3 Experimental results and discussion

In Figure 3.12 we show the test image (a), collected by the Ikonos sensor over San Diego (USA), the reference hand-drawn GT (b), whose construction took many hours of work, and the GT built with the proposed method, obtained in about 2 minutes (including segmentation) with 100 iterations. Visual inspection reveals that the semi-supervised GT is much more accurate than the hand-drawn one. This should not be too surprising: in order to obtain a high coverage of such a large image, about 2000×2000 pixels, the photointerpreter used only regular shapes, mostly rectangular, incurring often in undersegmentation, especially in the urban areas characterized by a myriad of small segments. Such problems can be easily avoided with the proposed tool by tuning the segmentation parameters to achieve the smallest undersegmentation degree.

To obtain some numerical evidence on the effectiveness of the proposed tool we focus on a subsequent task, the supervised ML classification of the



(a) (b) (c)

Figure 3.12: Ikonos MR test image. (a) RGB composite, (b) hand-drawn GT, (c) GT built with the proposed method.

	Roads	Buildings	Trees	Grass
Roads	**84.15**	27.21	0.76	0.00
Buildings	11.37	**68.87**	2.12	0.13
Trees	4.48	3.92	**94.93**	8.32
Grass	0.00	0.01	2.19	**91.55**

Table 3.7: Confusion Matrix using the hand-made GT: overall accuracy 82.52%, Kappa coefficient 0.738

image, using for training the semi-supervised GT designed by the proposed method, and the hand-drawn GT. In both cases, 800 pixels per class were picked at random for training the ML classifier, while accuracy is evaluated on the hand-drawn GT, excluding all sites involved in the training of both classifiers. The results, reported in Table 1 and 2, are very similar for the two ground truths.

Finally, to gain insight into the importance of the segmentation tool for the proposed method, the results obtained using MR-EMF were compared with those obtained with other standard segmentation algorithms, widely used in remote sensing, eCognition [12] and ENVI [75]. Performance is assessed here in terms of speed in the GT formation. In Table 3.9 we report for each segmenter, and at various stages of the process (30, 60 , 100 iterations) the coverage and rejection rate indicators. Coverage is simply the fraction of the image covered by the ground truth under construction, providing information into how fast a sufficiently complete GT can be obtained. The Rejection rate, instead, measures the segmenter precision, as it accounts for the area associated to seg-

	Roads	Buildings	Trees	Grass
Roads	**89.24**	30.86	4.83	0.13
Buildings	7.93	**66.27**	0.85	0.05
Trees	2.80	2.62	**83.02**	2.61
Grass	0.04	0.25	11.29	**97.21**

Table 3.8: Confusion Matrix using the semi-supervised GT: overall accuracy 81.90%, Kappa coefficient 0.727

ments that have been rejected during the process as unsuitable w.r.t. the total available area. The scale parameters of ENVI and eCognition were tuned to achieve the best performance.

The table clearly shows that MR-EMF overcomes the reference segmenters under both points of view. Comparative methods, particularly eCognition, suffer of a high rejection rate, eventually slowing down the labeling process. The gap between MR-EMF and the comparative solutions, is mainly due to the single scale nature (controlled by a parameter) of the latter. In fact, for images containing objects of different scales, a compromise scale parameter has to be fixed, hence balancing somehow over- and under-segmentation. For example, smaller scale parameters than those which provided the results of Table 3.9, would reduce the rejection rate (smaller under-segmentation), but also the average segment size, and hence area accepted in a given number of iterations.

In conclusion a simple and effective method for fast semi-automatic ground truth design which can be easily applied to very high resolution, or multiresolution, images acquired by last generation sensors is presented. Experiments carried out on a typical multi-resolution image prove the proposed framework to allow for a simple GT design in a fraction of the time necessary with conventional techniques, without impairing the performance on the the final application, point-wise spectral classification. Moreover, MR-EMF is able to provide for this task better results than the most widespread commercial software like eCognition [12] and ENVI [75].

3.3 Segmentation with correlation clustering

In order to test the performance of the proposed techniques some experiments have been carried out on regions of various sizes cropped from a large IKONOS multispectral image od San Diego. In table 3.10 we compare ILP

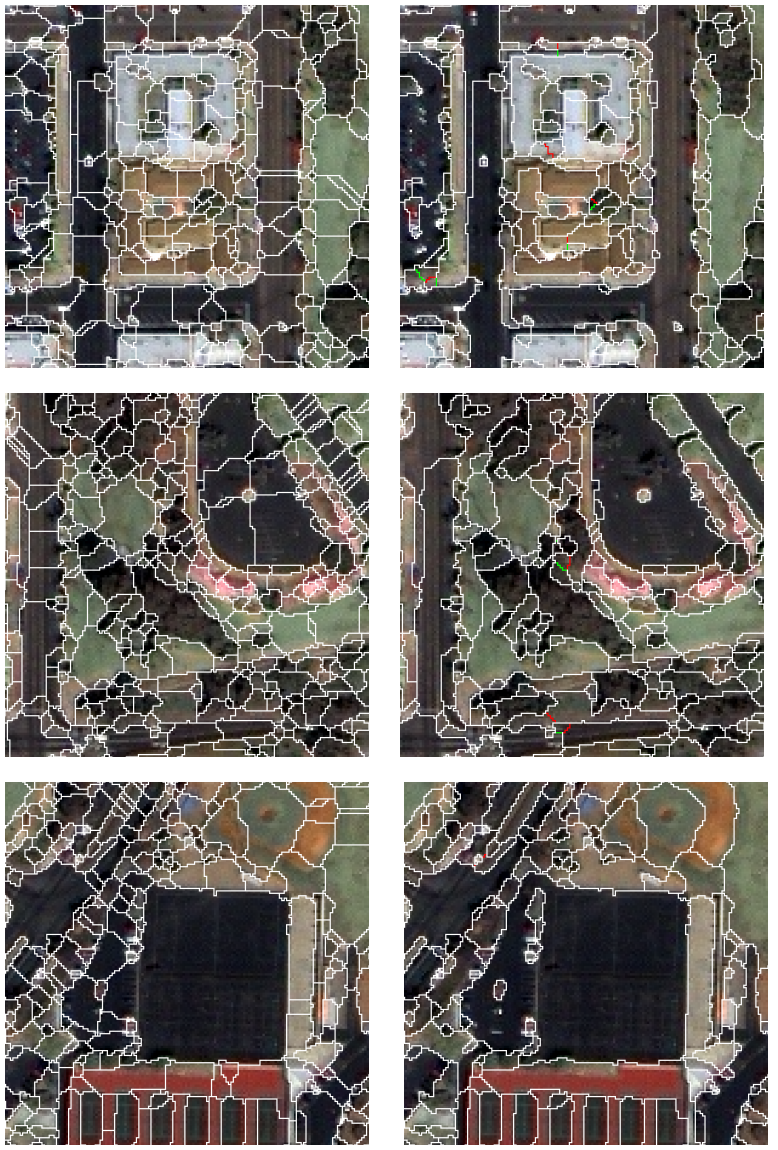


Figure 3.13: Superpixel contours (right) and Correlation Clustering segmentation (bottom) for some relevant clips extracted from the test IKONOS image.

	30 Iterations		60 Iterations		100 Iterations	
	Coverage	Rejection rate	Coverage	Rejection rate	Coverage	Rejection rate
MR-EMF	8.0%	4.2%	12,0%	6,1%	16,0%	5,9%
eCognition	8,2%	30,0%	11,2%	35,0%	14,3%	34,3%
ENVI	3,5%	17,0%	5,0%	18,0%	8,1%	17,2%

Table 3.9: Coverage and rejection rates for MR-EMF, eCognition and ENVI.

clip size	400x400		600x600		800x800		1200x1200	
# of nodes	2251		4857		8666		18751	
	Energy	Time	Energy	Time	Energy	Time	Energy	Time
ILP	-11081	1.7	-23889	9.7	-42491	29.0	-92789	158.4
greedy	-11078	1.2	-23884	6.5	-42486	19.8	-92766	100.2
greedy w/markers	-10850	5.7 (4.3)	-23396	7.7 (5.8)	-41710	14.7 (8.4)	-90838	45.5 (17.5)

Table 3.10: Energy and CPU-time for the various CC-based algorithms considered.

with the greedy algorithm, with and without markers, for the solution of the CC problem. Results are given in terms of Energy and CPU-time, the latter always on the same desktop PC. The greedy algorithm reaches always an energy level extremely close to that of ILP, with a consistent time saving of about 40%. Using markers, instead, some nodes are merged in advance and a slightly higher energy level is reached. However, it is worth pointing out that the resulting segmentation is not necessarily worse than that provided by ILP: the marker-based merging criterion, even if not energy-minimizing, may be more meaningful than that pursued with correlation clustering. Indeed, defining the energy so as to satisfy high-level requirements is still an open question, and our current solution is only a reasonable proposal. On the positive side, markers induce a time saving that grows significantly with the image size, and may become decisive for large images. For small images, instead, the marker generation time (shown in parentheses) dominates the overall cost. It should be also pointed out that the relatively low complexity exhibited by ILP in these experiments stems also from the good quality of the image, with many long edge segments (see Figure 2.8). With lower quality images (think of SAR) the situation would change dramatically. As an example, if half of the initial edges are removed at random from the 1200×1200 clip, ILP is not able to provide a result in acceptable times.

To conclude, Figures 3.13 and 3.14 shows a few sample results for the

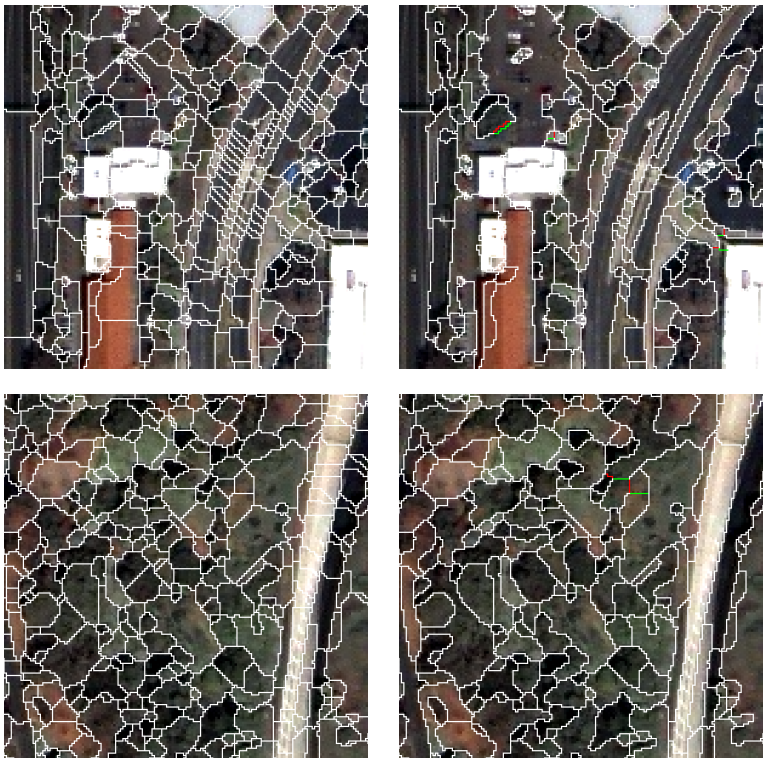


Figure 3.14: Superpixel contours (right) and Correlation Clustering segmentation (bottom) for some relevant clips extracted from the test IKONOS image.

IKONOS image. The proposed CC technique improves very much the segmentation quality w.r.t. the initial superpixel representation, preserving all relevant details. In green and red are shown the few differences between the solution provided by ILP and the greedy algorithm.

3.4 Segmentation of multitemporal SAR images

3.4.1 Case study and data

Our case study concerns the province of Caserta, in southern Italy, between the Volturno river and the Regi Lagni artificial channel (reference coordinates are $41°01'50''N, 13°59'04''E$). The area, whose Google Earth view is shown in Figure 3.15, is prevalently rural, with densely inhabited coastal settlements,

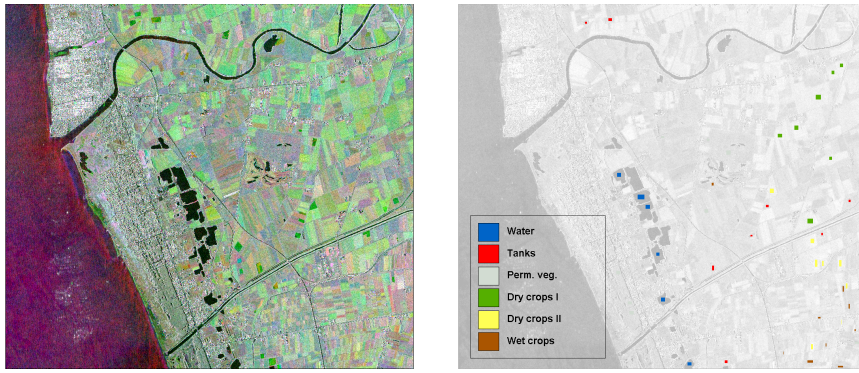
and a flat topography. It includes cultivated fields, human settlements, large tanks and small water harvesting facilities. Most of the fields are managed by family farms, so that the agricultural production units are very small (less than 4 hectares on average) and the terrain is cultivated with different plantations, each providing a different temporal feature in the radar reflectivity.

Figure 3.15: Google Earth view of the study area. The covered area is approximately 15 km $\times$ 13 km.

Fifteen COSMO-SkyMed stripmap SAR images, of size 5200$\times$4600 pixels, are available, spanning a temporal interval of two years, between December 14, 2009, and October 17, 2011. The data are HH polarized, acquired with ascending orbit and a look angle of approximately 33°. Spatial resolution is 3 meters, for an overall coverage of about 195 km^2.

Our aim is to recover the best possible range of land-cover information, in the absence of a ground truth, through the interactive segmentation of the whole area, driven by joint visual inspection of the SAR data and the corresponding optical view of the scene.

Figure 3.16(a) shows a false-color representation of the scene, obtained using the intensities of three SAR images acquired in different seasons, April (red), August (green) and December (blue), 2010. In Figure 3.17 we show some selected sections of this image and highlight some classes that might reasonably be found in a good thematic map. A "water" class (up-left) is clearly distinguishable from its low response at any season. Another "tanks" class (up-middle) comprises small agricultural tanks which are empty in the summer and filled in the other seasons, and appear in full green in our RGB



(a) (b)

Figure 3.16: False-color representation of the data (a), and selected ground truth (b).

composition. Pine groves and uncultivated crops are included in a "permanent vegetation" class (up-right) which has a fairly stable response throughout the year and hence appears close to gray in the image. Finally, we could identify three main types of crops, characterized by different seasonal behaviors, called "dry crops I", "dry crops II", and "wet crops", in the absence of more specific information, and shown in the bottom part of the figure.

Based on this set of classes, a ground truth was generated, shown in Figure 3.16(b), by manually annotating several areas of the image. It is worth underlining that this ground truth was *not* used in any way during the interactive segmentation phase, but only at a later stage for testing purposes, allowing us to compare the proposed method with suitable references in terms of classification accuracy.

A further "man-made" class is eventually considered, comprising the urban areas and other artificial structures. The urban areas, in particular, characterized by tiny details and a high response heterogeneity, cannot be easily identified based on the multitemporal vector of intensities. An expert could fairly easily find them based on textural properties which, however, are hardly captured by statistical models [115]. Here, we will resort to a further piece of information the average coherence, which is typically very large for stable artificial structures and much smaller otherwise.

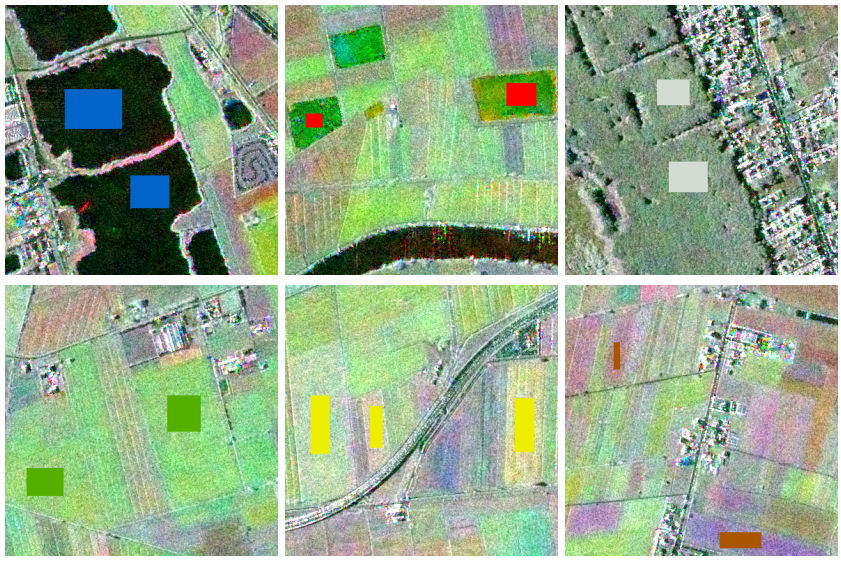


Figure 3.17: Ground truth samples for the homogeneous classes (*Water*, *Tanks*, *Permanent vegetation*, *Dry crops I* and *II*, *Wet Crops*).

3.4.2 Data preparation

In order to fully exploit the wealth of information provided by the COSMO-SkyMed data, using the TS-MRF suite with the smallest possible variations w.r.t. the optical image case, a number of preliminary processing steps are necessary:

- spatial registration;
- radiometric calibration;
- despeckling;
- homomorphic transform.

The data registration has been carried out via the three-step procedure proposed in [10]. After a coarse registration based on orbital data, a refinement based on correlation of amplitude data, and a final step based on coherence evaluation, the images are aligned with sub-pixel precision.

With multitemporal images, a meaningful comparison of data acquired in different dates requires a reliable calibration procedure. This step is of fundamental importance for a better visual inspection by the user. As explained in

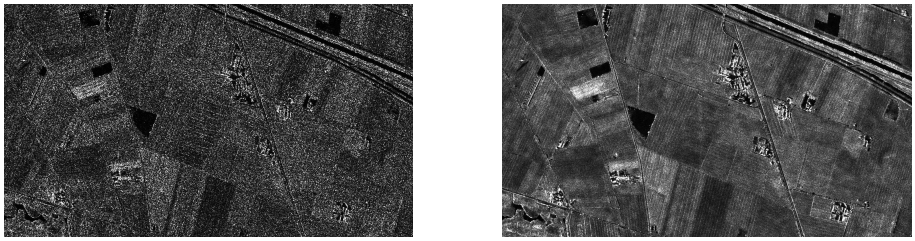
Table 3.11: Dataset summary

Acquisition date	Calib. coeff. [dB]
2009-12-14	-36.3176
2009-12-30	-36.9188
2010-01-15	-39.0413
2010-03-20	-40.7196
2010-04-05	-38.4177
2010-04-21	-40.1378
2010-08-11	-46.5234
2010-08-27	-41.1693
2010-09-12	-40.5112
2010-09-28	-39.4239
2010-12-17	-39.9272
2011-01-18	-40.0279
2011-02-03	-40.8033
2011-06-27	-46.9393
2011-10-17	-46.5963

[49], COSMO-SkyMed Single Look Complex Balanced products are already corrected for effects related to the sensor and the acquisition geometry. Hence, the sigma naught can be evaluated by applying a calibration factor which can be computed from ancillary data. The list of the available product dates and the corresponding calibration factors is shown in Table 3.11.

As mentioned in Section 1.3, the availability of a number of co-registered images of the same scene, gives us the opportunity to significantly improve the data quality by a suitable despeckling. In the proposed processing chain, we apply an optimal weighting De Grandi filter [34] which allows a speckle reduction in the order of 12 equivalent number of looks, without any loss in spatial resolution. In Figure 3.18 we show a subset of the 2010-04-05 image together with its despeckled version: the quality improvements is obvious, as well as the preservation of spatial resolution.

At this processing stage, data intensities do not follow anymore an exponential statistic (if they ever did before despeckling) but they are certainly not Gaussian, not even approximately. On the other hand, TS-MRF relies on the hypothesis that data are Gaussian, conditionally on the class they belong to. Therefore, in order to apply the TS-MRF suite without any structural modification, we perform a point-wise homomorphic transformation of the data,



(a) (b)

Figure 3.18: Comparison between SLC products before (a) and after (b) the application of the De Grandi filter.

which provides class-wise statistics of the scene much closer to the Gaussian distributions. It goes by itself that, even after this processing, this is still only a convenient approximation, but this holds just as well for optical images. More important is the fact that the class-wise distributions show a negligible skewness, allowing us to proceed as usual with second-order statistics.

To provide some more insight into the effects of these operations, Figure 3.19 shows a number of data distributions observed at various stages of the processing chain, for the 2010-04-05 image. In all cases, we report the frequency of occurrence of observations as a function of the intensity, using always the same scale on both x- and y-axis to enable an easy comparison. In particular, the distributions are computed after internal calibration (first column), after despeckling (middle column), and after logarithmic rescaling (last column). The first three rows show class-wise statistics for the water, permanent vegetation, and a crop class, respectively while the last row concerns the whole image.

As expected, despeckling modifies significantly the statistics observed in homogeneous areas, which pass from the characteristic exponential distribution of the SLC intensity product to a more symmetric one. The homomorphic transform, eventually, leads to distributions that are reasonably well fit by Gaussians.

Note that, while the water class is clearly separable from the other two, these latter have distributions that overlap significantly. However, the permanent vegetation has a fairly stable response during all the year, contrary to the crops, where the response is significantly influenced by state of the cultivations. By exploiting information on the whole time series, these two classes can be easily separated as well.

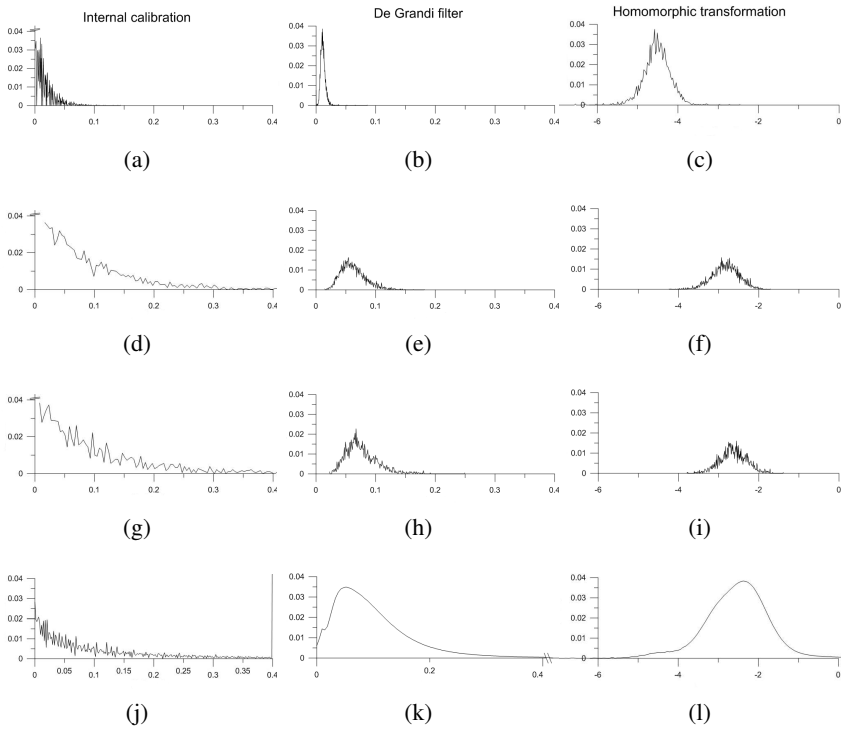


Figure 3.19: Variation of the image statistics along the processing chain for different classes (water, permanent vegetation, crops) and for the whole scene.

3.4.3 Interactive TS-MRF based segmentation

We now describe the interactive segmentation of our multitemporal SAR stack carried out with the tools provided by the TS-MRF suite. As explained in the section on TS-MRF, the user can select at any moment one of the following three actions

- split;
- split-and-merge refinement;
- topological split;

Since our aim is land-cover classification, we consider only the first two actions, initially, leaving the third one for the final stage when we build an object layer used to recover the man-made class.

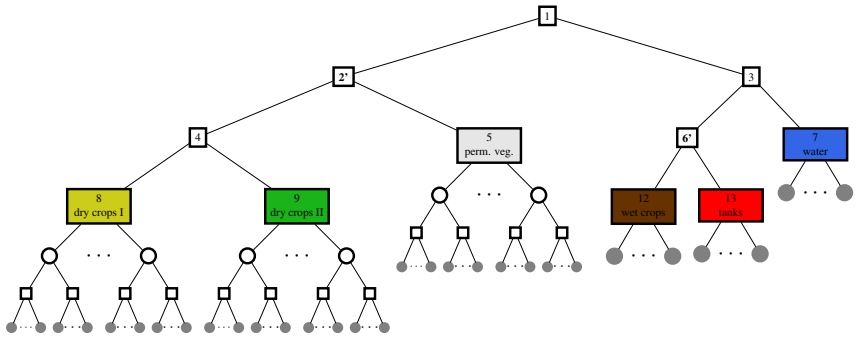


Figure 3.20: TS-MRF tree evolution: squared nodes come from binary MRF splits; circular nodes come from topological splits; prime superscript indicate a class obtained by merge-split refinement; filled circles represent elements of the object layer.

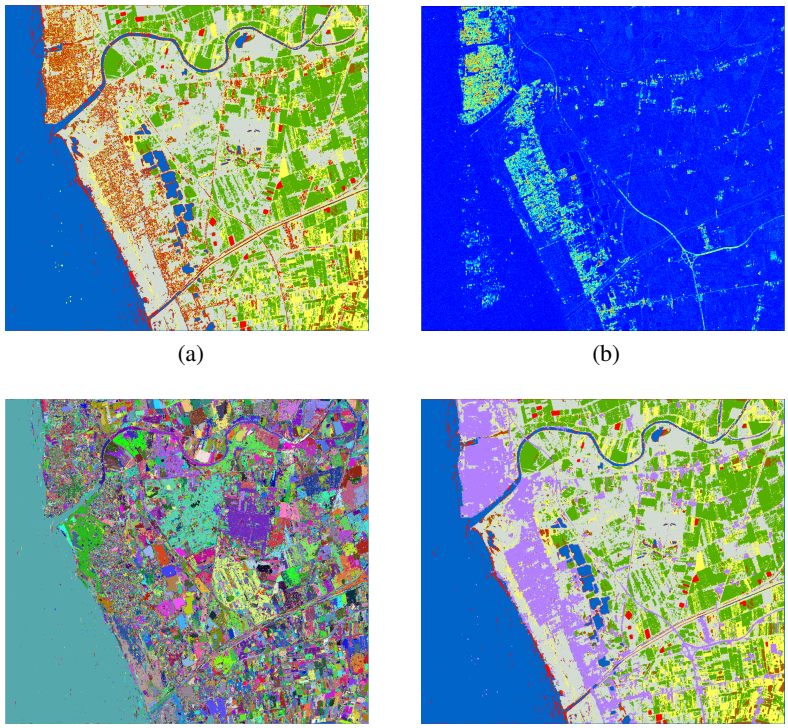
In automatic TS-MRF algorithm, the choice on whether to split some nodes, and in which order, is based on a locally computed split gain parameter. In interactive mode, instead the user assesses by visual inspection the meaningfulness of any split to decide whether to proceed, validating the split, or to stop. After each class split, the newly created classes can be compared with the other ones to check whether a merging is needed. Again, in the automatic version of the algorithm, a specific parameter, the merging gain, is used to drive this process. In the interactive mode, this decision is left to the user responsibility. In general, merge-split refinement should not be abused, resorting to it only when one of the children classes is clearly over-segmented, with complementary parts dropped into another class.

In our experiment, we obtained fairly naturally the six-class segmentation tree shown in Figure 3.20 (stopping at the colored nodes) using only visual information on class homogeneity and region compactness. The colors have been set only afterwards, by optimizing the matching of the selected classes with the ground-truth classes. Notice that only one merging was actually required to prevent over-segmentation. Such refinement affected the nodes labeled as 2' and 6', obtained by means of a suitable reshaping of the original classes labeled 2 and 6. More specifically, such reshape eventually helped recovering the integrity of class 12 ("wet crops"), eventually anchored to 6'. A full summary of the user actions is reported in Tab. 3.12, together with the classes emerging at each step of the process.

Figure 3.21(a) shows the segmentation map associated with our 6-class

	Action	Node	Emerging Classes
Classification	NODE SPLITTING	ROOT	2 *dry* 3 *dry + wet*
	NODE SPLITTING	3	6 *semi-wet + dry* 7 ***water***
	MERGE-SPLIT REF.	2,6	2' *dry* 6' *semi-wet*
	NODE SPLITTING	6'	12 ***wet crops*** 13 ***tanks***
	NODE SPLITTING	2'	4 *dry crops* 5 ***perm. vegetation***
	NODE SPLITTING	4	8 ***dry crops I*** 9 ***dry crops II***
Obj. L.	TOPOLOGICAL SPLIT	*dry* LEAVES (8,9,5)	-

Table 3.12: Summary of user actions.



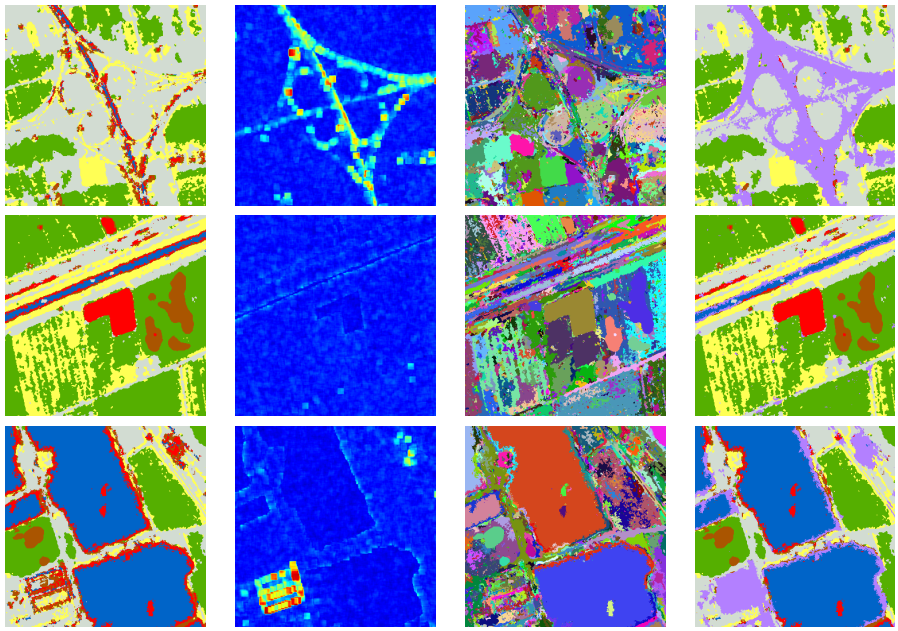
(a) (b)

(c) (d)

Figure 3.21: Segmentation products: (a) 6-class TS-MRF classification; (b) low-resolution coherence map; (c) full object-layer; (d) final 7-class segmentation.

tree. This result, however interesting, is obviously incomplete, as it lacks a dedicated man-made class. Indeed, the presence of some areas characterized by a significant class variability can be noticed in the map. These areas correspond mainly to urban settlements, characterized by very fine details, therefore they cannot fall into any of the above classes, but parts of them are retrieved in all six of them.

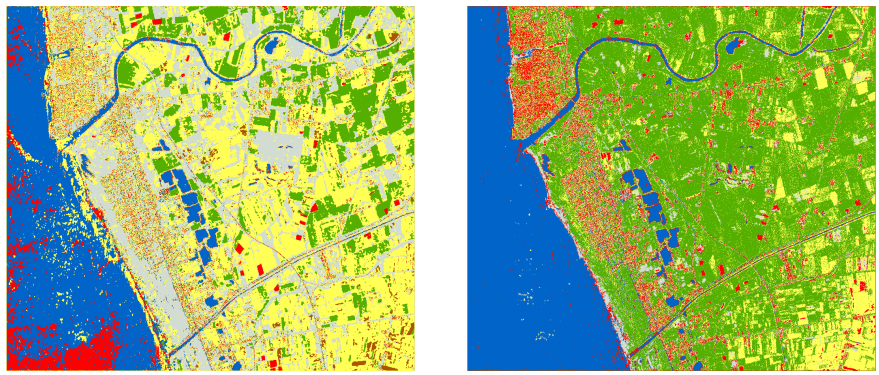
To single out man-made regions we resort to the coherence map, shown in Figure 3.21(b), obtained by averaging the pair-wise coherence of the oldest with all the others images. To single out man-made regions we resort to the coherence map, shown in Figure 3.21(b), obtained by averaging the pair-wise coherence of the oldest with all the others images. In fact, built-up areas



(a) (b) (c) (d)

Figure 3.22: Close-ups from Fig. 3.21. From left to right: 6-class map, coherence, object layer, 7-class map

exhibit typically a high coherence, which seems to be confirmed in our experiment. A simple way to obtain a reasonable man-made class is by thresholding the coherence map. By so doing, however, very irregular areas would be eventually extracted, and many inland artificial structures would be lost, especially the thin roads, due to the lower resolution of the coherence map. Such inconveniences can be avoided by resorting again of the TS-MRF suite. By applying a topological split to all classes, and then a further MRF split of each new segment, followed by a final topological split, a new tree is obtained, with terminal nodes (filled circles in Figure 3.20) which correspond to elementary connected components of the map. The set of all these components forms an "object layer", that is a higher-level representation of the image opposed to the pixel-level, shown in Figure 3.21(c) for our case study. We then perform thresholding at object-level rather than at pixel-level, featuring each object with its average coherence value. This very simple solution, which strongly relies on the available segmentation and hence constitutes a byproduct of the interactive classification, provides a much more consistent and accurate man-made class,



(a) (b)

Figure 3.23: Thematic maps obtained using the supervised (b) and unsupervised (c) flat MRF classification.

shown in lilac in the final output map of Figure 3.21(d) together with the other six original classes, properly reshaped. Note that, since the coherence information is projected on the original high-resolution objects, no loss of details is observed, and tiny structures are faithfully preserved. Note also that segments generated by the topological split of classes in the right part of the tree ("water", "tanks" and "wet crops") do not need a further MRF split since they are already quite homogeneous, hence we skip this last step for such segments.

3.4.4 Performance assessment

Visual inspection of the 7-class map of Figure 3.21(d) seems to confirm the potential of the proposed TS-MRF based technique for interactive segmentation. The most relevant regions of interest have been clearly extracted, with a good level of detail up to the finest scales, as confirmed by the close-ups shown in Figure 3.23. Of course, the quality of the SAR data used in this experiment, in terms of both spatial resolution and number of observations, plays a fundamental role in such good results. Nonetheless, the obtained performance seems much superior to that of conventional techniques working on the very same data.

In Figure 3.23(b) and Figure 3.23(a) we show the 6-class segmentation maps obtained by using a "flat" (non tree-structured, non interactive) MRF-based segmentation, both in unsupervised and supervised modality, to be compared with the analogous 6-class map of Figure 3.21(a) (obviously, the man-

made class, based on external features, is not considered in this comparison). To train the supervised segmenter, we used a fraction of the ground truth of Figure 3.16(b) as training set (around 35% of the area for each class), leaving the rest as test set for numerical evaluation. In Tables 3.13, 3.14, and 3.15, we report the confusion matrices computed on the test set for unsupervised MRF, supervised MRF and interactive TS-MRF, respectively.

The unsupervised classifier scores very poorly in terms of overall accuracy, τ=56.18%. With its balanced approach, given only the number of classes as prior information, it tends to the refinement of the classes with the higher data variability, missing altogether several others. The supervised classifier, as expected, performs much better, achieving an overall accuracy τ=81.82%. Interactive TS-MRF significantly outperforms both, with 87.22% of correctly classified pixels. These results were not at all obvious in advance. Remember that the proposed classifier is, in the usual sense, unsupervised, that is, it makes no use of prior information available on the classes of interest. The user can only decide which nodes to split, and possibly merge again, but has no influence on the binary local segmentations. Under this point of view, the most correct reference for the proposed approach is indeed the unsupervised MRF, and the huge performance gain speaks volumes about the importance of human/computer interaction.

Nonetheless, we observe a considerable gain also w.r.t. the supervised classifier. Looking in detail at the two confusion matrices, we notice that in both cases the "water" and "tanks" classes have been almost perfectly recovered thanks to their distinctive features. Significant differences arise instead on several vegetation classes. The poor performance of the supervised classifier on the permanent vegetation class is likely due to its strong inner variability, which can be hardly captured by a single multivariate Gaussian distribution. In the interactive approach, this class is identified by exclusion, after several other classes have been already well defined, hence it suffer less from over-segmentation. The "wet crops" class, instead, is a smaller class very local to the image, which is well recovered in the interactive case mainly thanks to the merge-split refinement performed in early stages. In summary, the observed gain is probably due to the better class-adaptivity of the tree-structured model, together with the opportunity of exploiting it through the user intervention.

For the man-made class, we limit the assessment to a suitable visual inspection of the result. In Figure 3.24 we compare, for various thresholds, a section of the man-made class obtained working at pixel-level (up) and at object-level (bottom) using the object layer provided by interactive TS-MRF.

	Water	Tanks	Perm. Veg.	Dry Crops I	Dry Crops II	Wet Crops	User acc.
Water	**8516**	0	16	0	0	0	99.81%
Tanks	54	**3444**	11	0	0	1402	70.13%
Perm. Veg.	1	1231	**888**	253	33	338	32.36%
Dry Crops I	0	0	6215	**4334**	649	833	36.02%
Dry Crops II	0	0	1527	4721	**7614**	1977	48.07%
Wet Crops	51	29	0	0	0	**0**	0%
Prod. acc.	98.77%	73.21%	10.26%%	46.56%	91.78%	0%	$\tau =$ **56.18%**

Table 3.13: Confusion matrix for the unsupervised flat MRF-based classifier.

	Water	Tanks	Perm. Veg.	Dry Crops I	Dry Crops II	Wet Crops	User acc.
Water	**8541**	0	0	0	0	0	100%
Tanks	55	**4636**	0	23	0	0	98.35%
Perm. Veg.	1	0	**4465**	0	0	639	87.46%
Dry Crops I	5	0	3	**9187**	740	0	92.47%
Dry Crops II	20	68	3893	98	**7495**	2124	54.72%
Wet Crops	0	0	296	0	61	**1787**	83.35%
Prod. acc.	99.06%	98.55%	51.58%%	98.70%	90.34%	39.27%	$\tau =$ **81.82%**

Table 3.14: Confusion matrix for the supervised flat MRF-based classifier.

	Water	Tanks	Perm. Veg.	Dry Crops I	Dry Crops II	Wet Crops	User acc.
Water	**8515**	0	0	0	0	0	100%
Tanks	107	**4704**	0	0	0	0	97.78%
Perm. Veg.	0	0	**6260**	0	0	654	90.54%
Dry Crops I	0	0	1980	**8721**	918	538	71.74%
Dry Crops II	0	0	101	2	**7378**	440	93.14%
Wet Crops	0	0	316	585	0	**2918**	76.41%
Prod. acc.	98.76%	100%	72.31%	93.69%	88.93%	64.13%	$\tau =$ **87.22%**

Table 3.15: Confusion matrix for the interactive TS-MRF-based classifier.

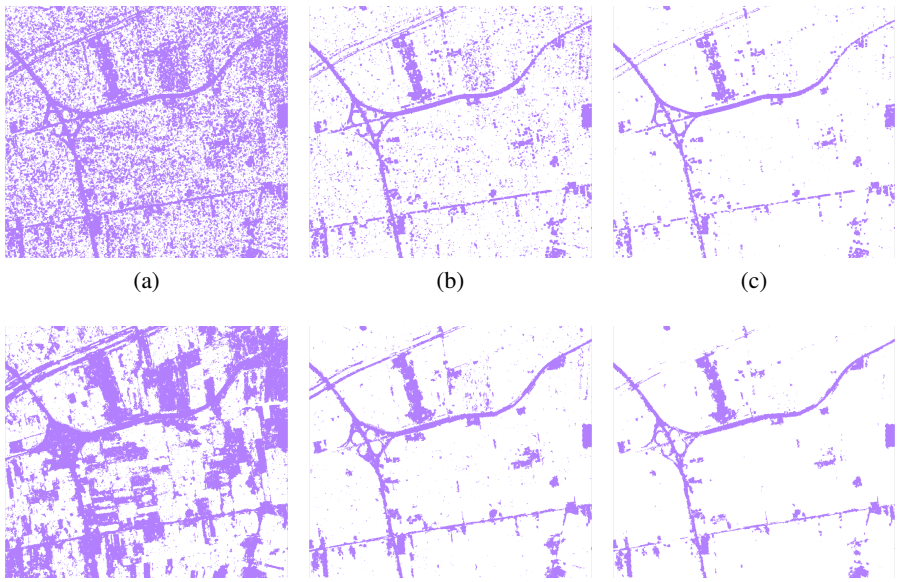


Figure 3.24: Pixel layer vs. object layer man-made class extraction. Top: maps obtained through pixel-level thresholding of the coherence map, with thresholds 0.15, 0.18, 0.21. Bottom: maps obtained through object-level thresholding at the same levels.

At all thresholds, the object-based class appears to be less noisy, and to better preserve the shape of the component regions. Moreover, in the pixel based solution, the variation of the threshold changes rather gracefully the level of noise in the map, providing little clues on which threshold best trades-off noise against the preservation of important details. With the object based solution, instead, by varying the threshold, entire objects of the scene appear/disappear, allowing for an easier selection of the "correct" level according to the highlighted content.

One might argue that a better man-made class could be obtained through a direct contextual segmentation of the coherence map. However, Figure 3.25 clearly shows that this is not the case. Next to a detail of the original continuous-valued coherence map (a), we show its binary segmentation obtained with the MRF model of 2.3.2 (b), and the result obtained by object-based thresholding (c). Direct MRF segmentation suffers the obvious problems related to the low-resolution original data: most thin details are lost due

(a) (b) (c)

Figure 3.25: Direct contextual segmentation of the coherence map vs. object-based thresholding: (a) detail of the coherence map; (b) MRF segmentation; (c) object-based thresholding.

to excessive regularization, and even large objects exhibit less continuity where coherence values are less dense. This last example underlines the potential offered by the object-layer, and hence by the low-level segmentation, for SAR data exploration.

3.5 Detection of environmental hazards

In this section a particular application based on segmentation is considered. In particular we propose a new workflow for the detection of potentially hazardous cattle-breeding facilities, exploiting both synthetic aperture radar and optical multitemporal data together with geospatial analyses in the geographic information system environment.

3.5.1 The case study

In this subsection, we present the case study: the detection of hazards with reference to BBFs. Also, we describe the available data sources, both optical and SAR, acquired in the province of Caserta.

Environmental hazards related to buffalo breeding

Pollutants from manure, litter, and process wastewater can seriously affect human health and the environment [71, 55, 94]. Whether from poultry, cattle, or swine, these contain substantial amounts of nutrients (nitrogen, phosphorus, and potassium), pathogens, heavy metals, and smaller amounts of other elements and pharmaceuticals [67]. This material is commonly applied to crops

associated with concentrated animal feeding operations (CAFOs) or transferred off site. Whether over-applied or applied before precipitation events, excess nutrients can flow from agricultural fields, causing harmful aquatic plant growth, commonly referred to as algal bloom, which can cause fish death and contribute to dead zones. In addition, algal bloom often releases toxins that are harmful to human health.

More than 40 diseases found in manure can be transferred to humans, including the causative agents of salmonellosis, tuberculosis, and leptospirosis. Exposure to waterborne pathogen contaminants can result from both recreational use of affected surface water (accidental ingestion of contaminated water and dermal contact during swimming) and ingestion of drinking water derived from either contaminated surface water or groundwater. Heavy metals such as arsenic, cadmium, iron, lead, manganese, and nickel are commonly found in CAFO manure, litter, and process wastewater [76]. Some heavy metals, such as copper and zinc, are essential nutrients for animal growth, especially for cattle, swine, and poultry. However, farm animals excrete excess heavy metals in their manure, which in turn is spread as fertilizer, causing potential run-off problems.

To promote growth and to control the spread of disease, antibiotics, growth hormones, and other pharmaceutical agents are often added to feed rations or water, directly injected into animals, or administered via ear implants or tags. Most antibiotics are not metabolized completely and are excreted from the treated animal shortly after medication. As much as 8090% of some administered antibiotics occur as parent compounds in animal wastes. Steroid hormones are of particular concern because there is laboratory evidence that very low concentrations of these chemicals can adversely affect the reproduction of fish and other aquatic species. The dosing of livestock animals with antimicrobial agents for growth promotion and prophylaxis may promote antimicrobial resistance in pathogens, increasing the severity of disease and limiting treatment options for diseased individuals (EPA 2011).

BBFs in the province of Caserta

This specific environmental problem is very relevant for the Caserta area in southern Italy, which therefore represents an interesting case study [72]. Caserta is the northernmost province of Campania, one of the most densely populated regions of Italy, and among the poorest. Campania is an agricultural region, very productive and highly specialized, with a model of extensive cultivation. Nearly 80% of farm work is carried out on family farms, so agri-

cultural production units are very small (3.6 ha on average). Mainly fruit and vegetables are produced, but buffalo breeding for mozzarella production is also important. In fact, the Caserta area is one of the main production sites of the "Mozzarella di Bufala Campana", the world-famous fresh cheese holding the status of a protected designation of origin under the European Union. In 2006 Campania produced 34,000 t of mozzarella, about 80% of national production.

The food production system in Italy, and especially in Campania, is relatively vulnerable to waste contamination [14]. Sometimes, this is due to the massive level of crime perpetrated by large-scale criminal organizations, but also it is the result of a culture of illegal practices and neglect widespread among small farm owners [53, 135].

Concerning BBF, in particular, besides many technologically advanced and lawabiding companies, many small factories exist which are not even on the productive activity register, and are not easily monitored or surveyed. Awareness of this problematic issue has been raised by many recent cases of pollution due to illicit spills involving BBFs. These cases have been reported in the course of inspections carried out by forestry personnel in collaboration with the local agencies in Campania. In particular, these investigations have made it clear that some holders did not properly accumulate and download all heaps of manure, with several cubic metres having been downloaded over a few square metres. This is in open violation of the established specific rules on how wastewater can be spread on soils. Manure cannot be accumulated in a small area as this represents a serious source of pollution. This is a bad habit that becomes a serious danger when BBFs are located in proximity to rivers, archaeological areas, or urban centres.

Available data

Two multi-resolution optical images were used in this work, acquired by the GeoEye sensor on 29 July 2010 and 12 August 2011, covering a region of about 20 km 16 km in the province of Caserta. Although other images were available, we considered only these two, acquired at about the same time of the year, in order to carry out a reliable multitemporal analysis. Each image comprises a panchromatic band with geometric resolution of 0.5 m/pixel, and a four-band multispectral image (Blue, Green, Red (RGB), Near-Infrared (NIR)) co-registered with the panchromatic band but with a geometric resolution of 2 m/pixel. Radiometric resolution is 8 bits for all data. Figure 3.26 shows the RGB composite of the 2010 image.

A set of 15 COSMO-SkyMed single-look complex balanced stripmap SAR

Figure 3.26: RGB composite of the first three bands of an optical GeoEye image. This image is originally composed of four bands: Blue (450520 nm), Green (520600 nm), Red (625695 nm), and NIR (760900 nm). The area is about 14 km 12 km with a spatial resolution of 2 m/pixel for the multispectral bands and 0.5 m/pixel for the panchromatic band.

images is available for the project, unevenly spanning a temporal interval of 2 years, between 14 December 2009 and 17 October 2011. All the data are Horizontal-Horizontal (HH) polarized, acquired with ascending orbit and look angle of approximately 33°. The data cover an area of about 40 km 40 km, with 3 m/pixel spatial resolution (in both range and azimuth). A calibration set for correcting the effects related to the sensor and the acquisition geometry can be extracted from the ancillary data provided by the Italian Space Agency (ASI). In such way, the achievable radiometric accuracy is about 1 dB. A subset of 5200 4600 pixels was used for the proposed project, covering an area of about 195 km2.

Figure 3.27 shows one of the available SAR images, geocoded and resampled on a map grid of 0.5 m/pixel (for comparison to the pan-sharpened optical image) after the application of the multitemporal De Grandi filter [34], fol-

Figure 3.27: One of the available SAR images in amplitude format. The area is about 14 km 12 km with spatial resolution of 3 m/pixel.

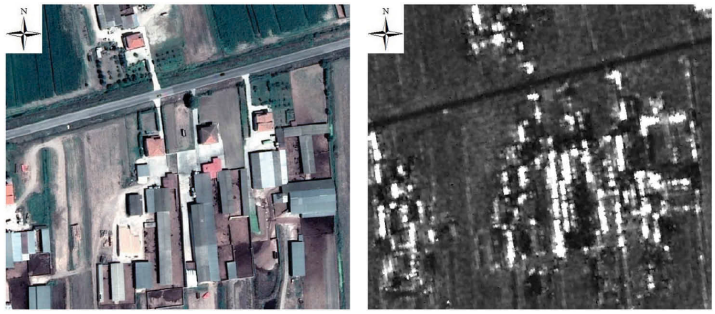


Figure 3.28: RGB composite of the first three bands (Blue, Green, and Red) of an optical GeoEye image and the amplitude SAR image enhanced by nonlinear processing for visualization purposes. The selected region is about 337 m 226 m and presents several BBFs.

lowed by a spatial non-local filter [100]. Multitemporal filtering, by exploiting time diversity, helps in reducing speckle and hence improves the performance of the successive segmentation step [63]. In particular, the De Grandi filter is relatively simple and has proved very effective in the context of several different applications [5, 59, 9]. The subsequent non-local filter exploits spatial dependencies to further reduce speckle, while preserving relevant image structures, as shown in [35].

3.5.2 Proposed approach

There are probably many ways to combine and exploit the available data to detect small BBFs.

In the following, we describe a simple processing chain, based on some preliminary observations on the characteristics of these facilities. As recorded from the satellite (see Figure 3.28), BBFs are mainly characterized by the adjacent sheds and fenced uncovered spaces used for both breeding the buffalo and accumulating animal waste. Sheds are clearly visible in both optical images, where they have a saturated response due to their high reflectivity, and the SAR images, where they contribute bright lines due to double-reflection mechanisms. However, these responses are not at all specific and can be confused with other highly reflective covers in the optical images (e.g. bitumenized roads) and, especially, with generic buildings in both sources. Moreover, as mentioned above, sheds are not always close to the fenced spaces where the

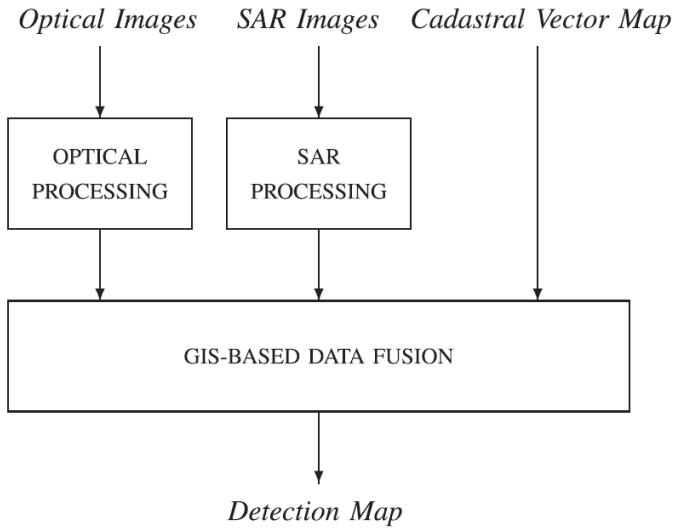


Figure 3.29: High-level processing chain.

buffalo live. Conversely, the spectral signature of the manure is highly characteristic, easily discriminated from bare ground in the NIR band, and stable to changes in solar illumination. Needless to say, manure is always present and abundant where the animals live, and indoor breeding is not an acceptable option in the highly standardized buffalobreeding protocol for the "Mozzarella di Bufala Campana" industry. Therefore, we decided to use GeoEye-1 optical images as main source of information, focusing on the manure signature.

In SAR images, manure does not exhibit a distinctive backscatter. However, we use the SAR stack to detect and mask built-up areas, thus reducing false alarms. This is especially valuable since most false alarms are related to shadows projected by buildings over bare soil, which abound in urban areas due to the high density of buildings.

In Figure 3.29 we show a high-level block diagram of the proposed workflow. Several optical images available at different dates are processed independently to generate maps of candidate BBFs. Instead, the entire multitemporal SAR stack is processed jointly to produce the built-up areas'mask. These outputs are then converted to vector format and processed in the GIS environment, together with the cadastral map including prior information on the location of facilities officially registered. The final product is a map of likely BBFs not on the official registry, which can be used in turn as input for on-site inspections

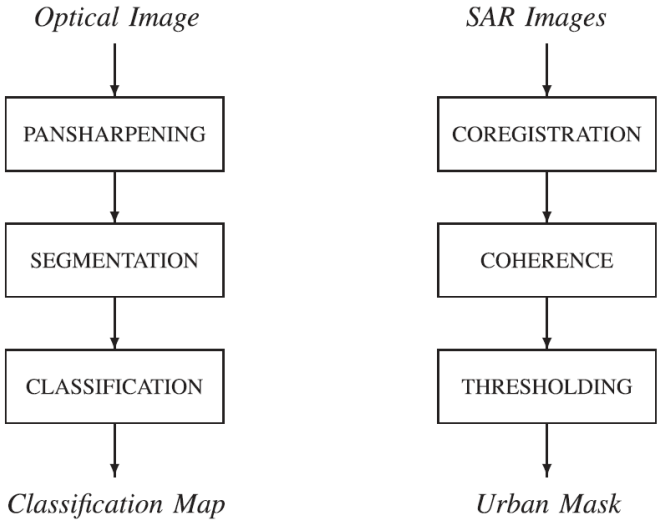


Figure 3.30: Detailed processing chain of optical and SAR domains.

by the environment protection and law enforcement agencies.

In Figure 3.30, we show finer details for the optical and SAR processing chains. The main goal of optical image processing is classification: based on their properties, pixels are labelled as either ”manure” or ”not manure”. To improve performance and reduce complexity, classification is carried out on homogeneous image segments rather than isolated pixels. Therefore, after a preliminary pan-sharpening, the image is segmented in order to identify its elementary homogeneous regions. Each high-resolution segment is then characterized by its spectral signature and classified.

The SAR image processing block, instead, provides a map of the urban areas in the scene. This is obtained by first co-registering the SAR images to a common master, and then computing the stack of corresponding coherence maps which is thresholded to provide the desired urban mask.

In the data fusion block (see Figure3.29), after georeferencing and co-registering all products, the classification maps corresponding to the various dates (two in our case) are combined through a logical AND, discarding in advance, however, regions that are too small and isolated. Detections occurring in urban areas, singled out by the SAR domain processing, are removed as well. The resulting map, converted from raster to vector format, is eventually compared to the cadastral map to determine suspect BBFs. Despite its

simplicity, this workflow turns out to be quite effective, as shown in Section 3.5.5. In addition, it is easily manageable by non-expert users, as the operators of governmental agencies may be expected to be. In fact, the output BBF map can be obtained and updated with a small number of simple operations, making the humanmachine interaction experience quick and trouble-free, thus orienting towards the end-user community [85, 63, 9].

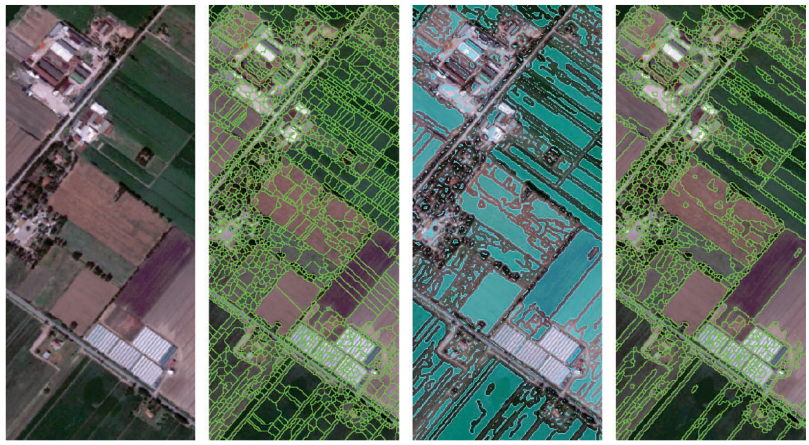
3.5.3 Processing in the optical domain

The first task carried out on the multiresolution images is pan-sharpening, which provides a data cube with full spatial and spectral resolution. We resort here to the GramSchmidt method, which has become very popular for pan-sharpening [81] due to its good performance over a wide variety of applications [48, 148]. Moreover, it is implemented in the ENVI package, one of the most commonly used commercial software packages for the processing of remote-sensing images. In the following, the segmentation and classification processing are discussed.

Segmentation

Remote-sensing images come in the form of arrays of pixels, hardly a good basis on which to make reliable decisions. Therefore, it is convenient to raise the description to a higher level, by identifying elementary regions, or segments, which are internally homogeneous, and hence characterized by means of a few compact features. These are, however, large enough to simplify all subsequent processing and enable the fast and reliable achievement of all application goals. This processing paradigm is also referred to as object-based image analysis (OBIA) or geospatial OBIA (GEOBIA), and is widely adopted as shown in the review proposed in [21]. In the present study, given the need to extract region contours as accurately as possible for subsequent vectorization, we resort to edge-oriented segmentation techniques based on the watershed transform. First, we compute the map of image edges on the high-resolution panchromatic component.

However, the application of watershed to real-world remote-sensing images (see Figure 3.31(a)) provides an exceedingly large number of regions (see Figure 3.31(b)), many of which are due to either minor imperfections of the edge map or simply the discrete geometry of the images and should obviously be merged. We therefore apply a more sophisticated segmentation algorithm, termed Edge Mark and Fill (EMF), proposed originally in [64] and



(a) (b) (c) (d)

Figure 3.31: Edge Mark and Fill (EMF) segmentation. (a) Original RGB clip (about 540 m 200 m); (b) watershed segmentation map; (c) morphological and spectral markers; (d) EMF segmentation map.

generalized in [63] for colour and multi-resolution images. Edge detection is here performed using the Canny edge detector [29], which is largely available and flexible and has been proven to perform well within the EMF framework.

The EMF performs a marker-controlled watershed segmentation. Markers are regions superimposed on the original image that reorganize all pixels covered by a given mark into the same segment. They can be inserted individually by an operator, a tedious and low-precision task, or, more interestingly, through some specific automatic procedure [142, 64]. In EMF, two types of marker are automatically generated and fused, based on the morphological properties of the Canny edge map and the spectral properties of the corresponding adjacent regions, respectively. (see Figure 3.31(c)), we show some of the markers generated by EMF, superimposed on the original image. By using such markers, the final segmentation map, shown in part (d), comprises a much smaller number of segments with the same accuracy, partially closing the gap with the ideal map that a human being might generate.

Spectral classification

Our aim is to classify each segment of the area of interest as either “manure” or “not manure”, based on the spectral response vectors of the component pixels, obtained through the pan-sharpening of the multi-resolution optical image.

Table 1. Comparison of training/classification combinations.

Training	Classification	Precision	Recall	*F*-measure
Pixel	Pixel	0.862	0.975	0.915
Pixel	Segment	0.902	0.995	0.946
Segment	Segment	0.854	0.994	0.919

Table 3.16: Comparison of training/classification combinations.

Table 2. Fifteen-class confusion matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	UA
1	9712	2502	89													0.79
2	42	4764														0.99
3	76	86	1450		4	71	174		541	982	16					0.43
4	4	8	3	593						2						0.97
5			62		4788	3		6		10					163	0.88
6			18			9779	311	3	311	161					35	0.92
7						79	3026									0.97
8						5	1	1991							402	0.83
9			609			15,191		9	848	1						0.05
10			1203		5	17			542	7203	2				11	0.80
11	265		14		41	57			3	18	1402		616	47	42	0.56
12	433											14,417	17			0.97
13												60	4214	574		0.87
14			18			165			6	159	130	109	153	11,199	67	0.93
15			5		48	144		49		2					10,939	0.98
PA	0.92	0.65	0.42	1.00	0.98	0.38	0.86	0.97	0.38	0.84	0.90	0.99	0.84	0.95	0.94	

Note: 1, shallow water; 2, deep water; 3, asphalt; 4, pools; 5, rock; 6, bare soil; 7, wet soil; 8, clay roofs; 9, asphalt roofs; 10, metal roofs; 11, green roofs; 12, trees; 13, grass; 14, sparse vegetation; 15, dry vegetation. UA, user's accuracy; PA, producer's accuracy.

Table 3.17: Fifteen-class confusion matrix.

To this end, given the wealth of information available, we resort to supervised classification. The spectral response of "manure" cannot be discriminated from that of other semantic classes, and hence a more general 'wet soil" class was used with regard to classification. As discussed in the following, "manure" can be discriminated with respect to other land covers of the "wet soil" spectral class only by going beyond spectral analysis. The proposed model eventually comprises 15 classes (see Table 3.17), including for example "green vegetation", "dry vegetation", and "bare soil". For our purposes, however, all segments not classified as "wet soil" are eventually collected in a single class and discarded from further analysis.

A relatively small fraction of the image was selected as the training set, taking care to include all the features of interest. More precisely, for each class of interest, we selected from 20 to 50 segments each comprising a few hundreds pixels, except for some classes of particularly small objects, such as clay, asphalt, green roofs, and trees. In fact for these classes smaller ground truth segments are needed (fewer than 100 pixels) for reliable annotation. Overall, about 200,000 pixels were used for the training set. In the same manner, we

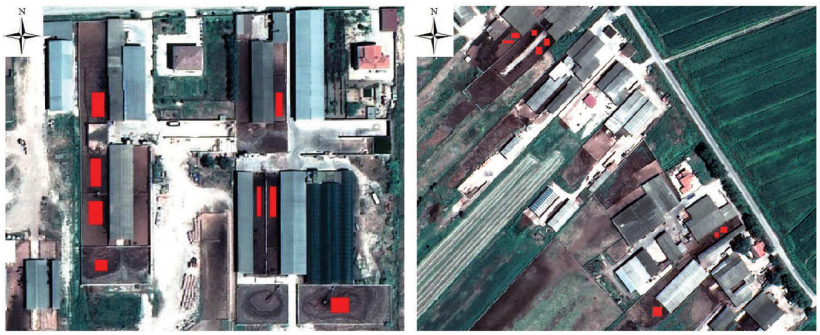


Figure 3.32: Training (left) and test (right) sets for classification. Red boxes correspond to "wet soil" areas.

formed a test set of approximately 110,000 pixels, taking care to avoid any intersection between training set and test set segments.

Figure 3.32(a) shows some "wet soil" training set segments (to avoid cluttering the figure, segments of other classes are not shown), while Figure 3.32(b) shows some test set segments of the same class.

Given the detailed 15-class model, it is reasonable to characterize each class $c = 1, .., C$ through a single-mode probability density function (PDF), and in particular, a multivariate Gaussian random vector with mean μ_c and covariance matrix $\mathbf{\Sigma}_c$. These synthetic statistics are maximum likelihood estimated based on the training data, with high reliability, given the low dimensionality (four) of the vector space, much lower than the number of available pixels per class. Given the spectral vector $\mathbf{X}(s) = \mathbf{x}$ associated with pixel s the label or class $\hat{c}(s) \in C$ is chosen according to the Maximum A-posteriori Probability (MAP) rule. However, lacking any prior information on the classes, this reduces to the maximum likelihood rule, and eventually, for the assumed Gaussian statistics, to

$$\hat{c}(s) = \arg\min_{c \in C} \left[ln \left| \mathbf{\Sigma}_c \right| + (\mathbf{x} - \mu_c)^T \mathbf{\Sigma}_c^{-1} (\mathbf{x} - \mu_c) \right] \tag{3.1}$$

To reduce the influence of noise, the decision is made on segments, rather than pixels. For each homogeneous region singled out by segmentation, the average spectral signature is computed and used for classification. Segmentation granularity is kept high in order to preserve the homogeneity of the spectral response in the same segment. As a consequence, the physical objects of the scene are often composed by several segments. Because of the use of seg-

ments rather than pixels, the classification appears to be fairly reliable, despite the simple multivariate Gaussian model adopted.

3.5.4 Processing in the SAR domain

The goal of SAR domain processing is to extract a pixel-based map of urban areas through the analysis of interferometric coherence. To this end, SAR images are preliminarily co-registered with one another by a three-step procedure [84] during which the alignment is progressively refined. First orbit information is used, then the cross-correlation between coupled windows, and finally optimization of results with Powells method [107]. Man-made areas can be separated from natural ones based on the interferometric coherence between successive acquisitions, because of their different scattering formation physical principles [111]. Indeed, in man-made areas, the backscattering is dominated by multiple reflections between building elements and the ground. Moreover, double and triple reflections due to the dihedrals and trihedrals are stable with respect to variations in the observation geometry [58]. Rural areas, typically composed of trees, cultivated fields, grasses, and crops, instead exhibit backscattering values that are strongly influenced by both the observation geometry and changes in the scene. This causes low values of interferometric coherence, especially if computed with a large temporal baseline. Of course, the typical time lapse necessary to cause an appreciable drop in coherence is different for each of the above-cited objects and depends also on the season (e.g. a single rainfall event can change the scene characteristics with a dramatic reduction in the coherence between the pre- and post-event images) [60]. In the available data set, the average interval between two acquisitions is in the order of one month, which is sufficient for assuming that only stable targets (e.g. buildings and roads) exhibit a high level of interferometric coherence. However, in order to minimize the probability of false alarms, a mean coherence map was generated by averaging all the coherence maps between an image assumed as reference and all others.

Figure 3.33 represents the mean coherence values (left) and map (right) of the whole SAR scene under analysis, projected onto the WGS84 geographic coordinate system (north at top), where the sea has been manually removed since it is irrelevant to the analysis conducted in this work. Note that to obtain the binary map, man-made and natural areas are separated by simple thresholding [63]. As a final step, we move from pixel- to region-level maps, based on prior information. Indeed, urban zones are areas of significant spatial extension with a high density of man-made structures. Following this definition,

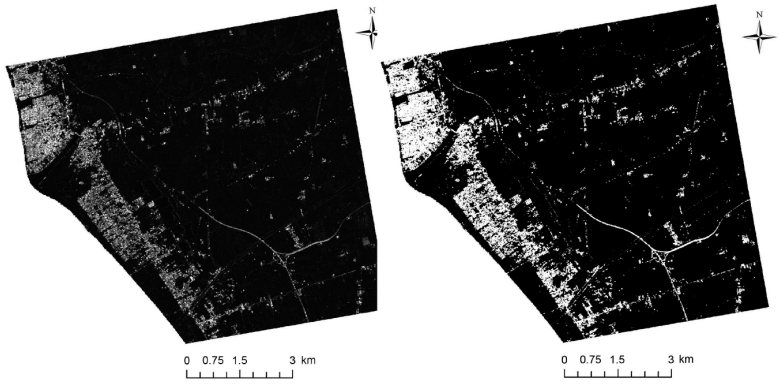


Figure 3.33: Geocoded mean coherence map (left) and the corresponding man-made mask (right).

we first compute local density by averaging the pixel-level map in a circular region of radius 200 pixels (600 m) centred on the target. Then the density map is thresholded, and small (¡10,000 pixels) isolated regions are removed. Eventually, only high-density large regions are classified as urban areas. Figure 3.34 depicts the full pan-sharpened data set available. The highlighted portion (in yellow) overlaps the co-registered SAR multitemporal data and is hence used in this work.

3.5.5 Data fusion

In the data fusion block, decisions are made based on all available pieces of information. Before that, however, co-registration and rectification are required, in order to provide coherent data.

Co-registration between SAR and optical images

SAR/optical image registration aims at correcting the misalignment of geocoded SAR with respect to the rectified optical images. The SAR geocoded images were obtained via a range-Doppler mode, with a 20 m-resolution digital elevation model (DEM) for compensation of terrain-induced distortion. Although many automatic registration techniques have been proposed in the literature (e.g. those in [73, 129, 54]), their robustness is still limited. Therefore, we refined the SAR/optical alignment with user-defined ground control

Figure 3.34: Dense urban areas (in lilac) extracted by refining the man-made map, superimposed on the RGB composite image.

points (GCPs), the selection of which is time-consuming and influenced by operator sensitivity. However, this needs be performed only once for the whole data set. In order to improve the accuracy of GCP identification in the SAR maps, a multitemporal De Grandi filter [34] was applied to the entire available data set. In order to further reduce the effects of speckle without loss of spatial resolution, the filter was followed by non-local spatial despeckling [100]. Then, the selection of GCPs was based on the identification of points that are easily recognizable and detectable in both the SAR and optical maps, despite their different geometries. Hence, the point candidate to be selected as GCPs should be relative to areas which are stable in amplitude (on the SAR map) and easily separable from surrounding regions [151], such as road crossings, buildings, boundaries between homogeneous areas, or other dominant features observable in both images. Thirty uniformly distributed GCPs were selected for building the warp polynomial needed to align the SAR to the optical reference image. This processing allowed an alignment in the order of pixel size, which is consistent with the project objectives.

Rectification based on a DEM

Often georeferencing is not sufficient to guarantee a precise spatial correspondence among physical regions and objects in the various images, even when DEMs are used for either orthorectification or geocoding. To improve the geometric quality of the original images, rectification was applied using the rational function model (RFM) [130], adopted with success in many applications [86]. This approach requires a DEM of the whole area and at least 39 GCPs with known image coordinates and 3D (altimetric and planimetric) position in a geodeticcartographic reference system. A 5 m 5 m DEM of the region of interest was built by means of linear interpolation on vector maps at a scale of 1:5000, with 99 GCPs for the first image and 201 for the second. Accuracy was tested by considering the difference between the exact and estimated coordinates, by means of root mean squared error ($RMSE = \sqrt{MSE}$). RMSE turned out always to be below 1 m for GCPs, and slightly higher than that for a disjoint set of control points.

Decision

After co-registering all products, several simple decision rules can be enacted based on the segment-level classification of optical images and on the urban mask. In particular, we show results obtained using only one or both of the optical images and with and without urban masking. When two images are used, only segments detected in both are taken into account (hence, a more conservative choice). Moreover, the urban mask, when used, allows one to discard segments detected in urban areas. A more detailed description of the decision process is deferred to the next section, in the context of the experimental analysis.

Experimental results

Here, we report results of the experiments carried out to validate the proposed algorithm. We analyse separately the performance of classification and detection tasks. However, since in both cases we are eventually interested in detecting the presence of a given target class, “manure” in classification, and “buffalo breeding facility” in detection, we always consider the same measures used in two-class hypotheses tests: namely precision, P, and recall, R, and the synthetic F_1 measure F. These measures are defined, with reference to a generic target class T, as

$$P = \Pr(c = T \mid \hat{c} = T), \tag{3.2}$$

$$R = \Pr(\hat{c} = T \mid c = T), \tag{3.3}$$

and

$$F = \frac{2PR}{P + R}, \tag{3.4}$$

where Pr denotes probability, and c and $\hat{c}$ indicate the true and selected class/hypothesis, respectively. High precision indicates that when the target class is detected, the decision is very likely correct. High recall indicates that when the target class is present, it will very likely be detected. Therefore, both measures are desired to be large, and by tuning the classification parameters one may increase either one while decreasing the other. A synthetic measure of performance is the F-measure, a harmonic mean of precision and recall, which is large only when one or both indicators are rapidly reducing in value.

Classification

Our classifier is trained on pixels drawn from the training set, while the decision is made on segments, namely the average spectral response computed over all pixels belonging to a segment. This mixed solution was chosen after comparing its performance to the other relevant alternatives, where training and classification are performed on pixels, segments, or both. Results are reported in Table 3.16, with respect to the target class "wet soil", and are computed pixel-based irrespective of how the decision was made. Although the performance is definitely good in all cases, with *F*-measure above 0.9, the selected mixed solution guarantees an appreciable gain in precision, and therefore in the *F*-measure. Indeed, when decisions are made on individual pixels, the influence of noise is more relevant, causing a drop in both precision and recall. In the third case, the problem is likely the limited number of segments available for training, which reduces the ability of the classifier to deal with outliers of other classes. We underline also that the pixel-based solution must be excluded not only for its inferior performance, but also because we use segments as the basis for the detection of BBFs. For the selected solution, we computed the complete 15-class confusion matrix **A** over a total of N = 113,367 pixels, with entries a_{ij} counting the number of pixels of class j that were classified

as belonging to class i. Based on a confusion matrix, several global quality indicators are usually computed. The overall accuracy, τ, defined as

$$\tau = \sum_i a_{ii}/N \tag{3.5}$$

is the percentage of sample pixels correctly classified. The kappa parameter, defined as

$$\kappa = \frac{N \sum_i a_{ii} - \sum_i a_{i+} a_{i+i}}{N^2 - \sum_i a_{i+} a_{i+i}} \tag{3.6}$$

with $a_{i+} = \sum_j a_{ij}$ ans $a_{+i} = \sum_j a_{ji}$ discounts successes obtained by chance, and is therefore more conservative (it can also be negative). The average accuracy (AA), also frequently used, is defined as the mean of per-class producer's accuracy, a_{ii}/a_{+i}. Finally, the normalized accuracy, τ^{norm}, is computed on a confusion matrix modified as described in [32] in order to give equal importance to all classes, irrespective of the number of samples in each. These indices are all very high for our classifier: $\tau = 78,19\%$, $\kappa = 76,05\%$, $AA = 80,98\%$, $\tau^{norm} = 86,92\%$, especially considering the large number of classes considered, some of which are similar to one another. In Table 3.17, we report the confusion matrix. With perfect classification, only diagonal entries should be larger than 0, and indeed, most off-diagonal entries are 0 (blank) or very close to it. In any case, we are especially interested in the "wet soil" class, number 7, including "manure", for which both producer's and user's accuracies are clearly very high.

Detection

Detection performance is assessed on a large part of the available images, shown in the yellow box in Figure 3.35, while the small region in the red box is used only for visual inspection. Measuring performance is less obvious in this case. Our goal is to detect BBFs, when present, and to avoid declaring their presence otherwise. First, to measure success in the first task, we need a ground truth which identifies all such facilities in the test area. Therefore, an expert photo-interpreter thoroughly analysed the whole image. Then, based also on other complementary sources of information, 76 BBFs were detected, their approximate contours shown in GIS as regular polygons and in yellow (nine of them) in the example clip of Figure 3.36.

Figure 3.35: The image used in the experiments. Performance is computed on the large region in the yellow box. The small region in the red box is used for detailed visual inspection of results

Figure 3.36: Segment-level decisions on the same small area of the image at the two dates. Green, correct; red, false alarm.

Variant	Image(s)	Urban mask	Precision	Recall	F-measure
1	T_1	–	0.973	0.154	0.266
2	T_1	No	0.986	0.234	0.379
3	$T_1 + T_2$	–	0.934	0.862	0.897
4	T_1	–	0.973	0.272	0.425
5	T_1	Yes	0.986	0.426	0.596
6	$T_1 + T_2$	–	0.934	0.949	0.941

Table 3.18: Detection performance with different variants of the proposed procedure.

This figure also shows the segments classified as "wet soil", in green (correct decision) when more than 50% of the segment is inside a BBF, or in red (false alarm) otherwise. However, we are interested in detecting facilities, not segments. Therefore, we use these data to label the 76 BBFs as either detected (when comprising at least one green segment) or missed (when no green segment falls within its bounding polygon). In the example clip, all nine BBFs are detected at both dates. With this information, we can compute a meaningful recall indicator. In regard to precision, no similar conversion seems possible. So we are forced to operate at segment level, computing precision as the ratio between the number of segments (green) correctly declared "wet soil", possibly manure, and the number of all segments (green or red) declared "wet soil", irrespective of their real class, thus including errors. Although working at segment level, this latter indicator provides a good insight into the quality of the whole procedure. If precision is too low, the technique indicates many more targets than actually present, becoming basically useless. To reduce false alarms we resort to the urban mask of Figure 3.34, computed from the SAR coherence map.

In Table 3.18, we report the performance indicators obtained using all pieces of information available (last row) or just some of them. In the first two cases, only one of the optical images is used, either T_1 or T_2. In both cases almost all BBFs are detected (high precision), but also thousands of "manure" regions unrelated to BBFs (low recall), resulting in an acceptable overall performance, as testified by the very low F-measure value. This was to be expected from the analysis of Figure 3.36, where many red segments appear. However, while regions in BBFs are persistent, because they are continuously covered by manure, external regions are only occasionally classified as such possibly because they are periodically fertilized and can thus be eliminated through multitemporal analysis. By combining the maps relative to both time instances through a simple logical AND (case 3), much better recall is ob-

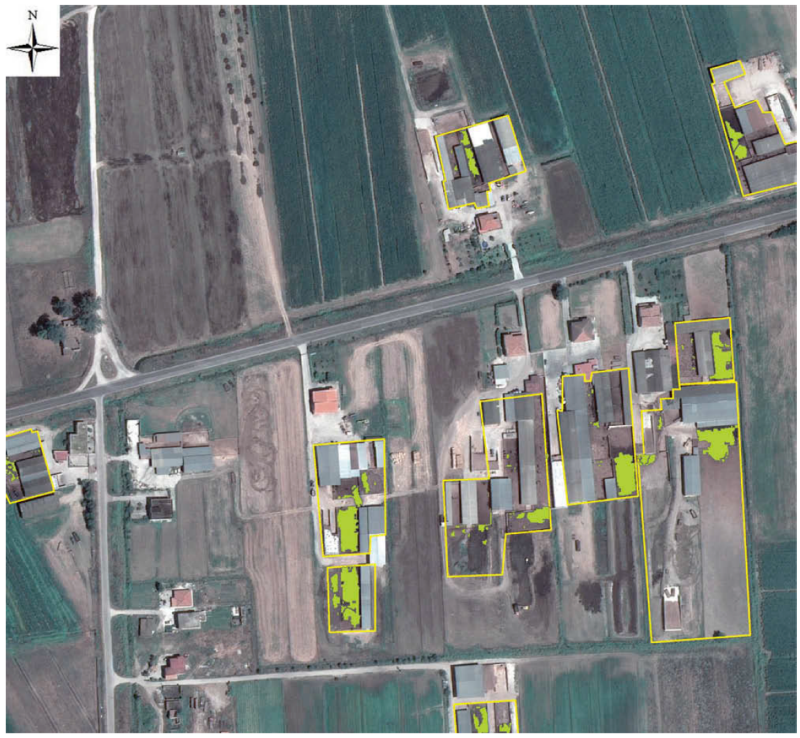


Figure 3.37: Segment-level decisions based on multitemporal data. No false alarm occurs in the clip.

tained. However, although not in the example clip, some BBFs are lost due to the logical AND, slightly reducing precision. Despite this loss, a much higher F-measure is observed. Figure 3.37 shows the effects of the logical AND on our example clip. In the last three rows of the table, we report the same data as above when the mask for dense urban areas, derived from SAR images, is also used. This mask allows us to reject a number of bare soil areas that, when shadowed by buildings, are spectrally indistinguishable from wet soil, generating a large number of false alarms. Therefore, recall increases significantly with respect to the corresponding cases without urban mask. Precision is obviously not affected by masking, because BBFs are always rather distant from large urban centres. The fully fledged technique (case 6) ultimately guarantees both high precision, with 71 facilities detected out of 76, and high recall, with only 30 false alarm segments out of 590.

Conclusions and perspectives

In this thesis a new technique for the segmentation of MR remote sensing images is proposed, whose general scheme is also applicable to single-resolution data. Segmentation is regarded here as a means to build an object-level representation of the image; based on which, more advanced analysis tasks can be easily carried out, with lower computational cost and better accuracy. In this perspective, it is essential that all relevant information is preserved in the segmentation map. Therefore, we resort to watershed transform based on a preliminary edge detection phase. The oversegmentation typical of watershed is controlled by means of suitable morphological and spectral markers, which are extracted automatically from the data. For MR data, the PAN and MS components are processed independently in their respective domains, avoiding any possible information loss induced by pansharpening. Processing products are then merged at the highest resolution, using for each image region the most appropriate pieces of information. Experiments on both hyperspectral and common MR imagery fully support our choices. The performance of object matching and land cover classication, computed using an available detailed GT, improve both significantly, whereas visual inspection confirms the superior ability of the proposed scheme to preserve all image scales. Currently, we are working on including topological information to further improve the quality of the object layer. Moreover, we are testing the potential of the object layer for road extraction applications based on geometric features.

Furthermore, the proposed segmentation scheme is tested on two real world applications.

The first application is a simple and effective method for fast semi-automatic ground truth design which can be easily applied to very high resolution, or multiresolution, images acquired by last generation sensors. Experiments carried out on a typical multi-resolution image prove the proposed framework to allow for a simple GT design in a fraction of the time necessary with conventional techniques, without impairing the performance on the the fi-

nal application, point-wise spectral classication. Moreover, we show MR-EMF to provide for this task better results than the most widespread commercial software like eCognition and ENVI.

The second application is a methodology for detecting small buffalo-breeding facilities based on multi-sensor and multitemporal remote-sensing data and GIS-based processing. The performance of the proposed system is quite satisfactory with an F-measure always above 0.9. Hence, it can be a valuable tool for monitoring environmental hazards, adaptable to different tasks by modifying the input data, and also in regard to various highly data-dependent processing tasks, such as denoising or segmentation. For example, work is under way to adapt the tool to the detect illegal landfill. Of course, there is room for further improvement in several aspects. First, with more images available, a better decision strategy could be implemented to detect all areas of interest, with limited false alarms. However, even with the data currently available, performance could be improved by better exploiting information available in the GIS, such as the position of candidate areas with respect to road networks and water ways, or other geographic layers from different sources. Work is currently under way to investigate these issues.

Another contribution of this thesis is a new graph-cutting segmentation algorithm for remote sensing images, based on a preliminary superpixel segmentation. Thanks to the use of edge-oriented superpixels, boundary accuracy is always preserved. Segmentation is then formulated as a correlation clustering problem on the RAG associated with the superpixels. Both visual quality and complexity appear as very promising, the latter being a major issue for high-resolution images. Current work is on the definition of alternative energies which better capture image quality.

With reference to the challenging case of multitemporal high-resolution SAR data, in this thesis we show that much better results can be obtained by following an interactive data exploration paradigm. Here, no prior analysis is required, but the user can perform simple actions, based on prior knowledge and experience, that guide the process toward the most satisfactory results. This approach, of course, rests up on the availability of powerful and easy-to-use basic tools. We show that the TS-MRF algorithmic suite fits well this approach, and allows one to obtain very valuable results. The hierarchical image model, which adapts to the local statistics of the classes, provides the required flexibility to deal with non-standard problems. In the considered case study, the interactive use of TS-MRF allowed us to obtain a better thematic map than with a conventional supervised approach. Moreover,by leveraging on

the object-layer made available by the suite, we could also accurately recover a high-definition man-made class. Despite such good results, there is much room for improvements under many respects. In this work, we used a very limited set of possible actions, but many more can be conceived. Keeping the very promising hierarchical framework, more sophisticated MRF models can be used, possibly varying from node to node, possibly K-ary instead of binary. Other classes of models, alternative to the MRFs, can also be considered to deal with specific problems, such as the recovery of macro-textures, road networks, etc. We are already investigating some of these topics.

Bibliography

[1] *eCognition user manual*. Munich, Germany: Definiens AG, 2003.

[2] G. Abbate, L. Fiumi, C. De Lorenzo, and R. Vintila, "Evaluation of remote sensing data for urban planning. Applicative examples by means of multispectral and hyperspectral data," in *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, May 2003, pp. 201–205.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, Nov 2012.

[4] H. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, July 2008.

[5] I. Ali, C. Schuster, M. Zebisch, M. Forster, B. Kleinschmit, and C. Notarnicola, "First Results of Monitoring Nature Conservation Sites in Alpine Region by Using Very High Resolution (VHR) X-Band SAR Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2265–2274, 2013.

[6] A. Alonso-Gonzàlez, S. Valero, J. Chanussot, C. Lòpez-Martìnez, and P. Salembier, "Processing multidimensional SAR and hyperspectral images with binary partition tree," *Proc. of the IEEE*, vol. 101, no. 3, pp. 723–747, March 2013.

[7] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Bruce, "Comparison of Pansharpening Algorithms: Outcome of the

2006 GRS-S Data-Fusion Contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[8] A. Alush and J. Goldberger, "Break and Conquer: Efficient Correlation Clustering for Image Segmentation," in *Proc. of SIMBAD*, 2013, pp. 134–147.

[9] 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 Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 117–133, 2015.

[10] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, G. Ruello, M. N. Papa, F. Ciervo, and Y. Koussoube, "Effectiveness of high-resolution SAR for water resource management in low-income semi-arid countries," *International Journal of Remote Sensing*, vol. 35, no. 1, pp. 70–88, 2014.

[11] R. Arbiol, Y. Zhang, and V. Palà, "Advanced classification techniques: a review," in *ISPRS Commission VII Mid-term Symposium "From Pixel to Processes"*, Enschede, NL, 2006.

[12] M. Baatz and A. Schäpe, "Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000*, H. W. Verlag, Ed., Karlsruhe (Germany), 2000, pp. 12 – 23.

[13] N. Bansal, A. Blum, and S. Chawla, "Correlation Clustering," *Machine Learning*, vol. 56, no. 1-3, pp. 89–113, Jul. 2004. [Online]. Available: http://link.springer.com/10.1023/B:MACH.0000033116.57574.95

[14] M. Barba, A. Mazza, C. Guerriero, M. D. Maio, F. Romeo, P. Maranta, I. Marino, M. G. Paggi, and A. Giordano, "Wasting Lives: The Effects of Toxic Waste Exposure on Health. the Case of Campania, Southern Italy," *Cancer Biology & Therapy*, vol. 12, no. 2, pp. 106–111, 2011.

[15] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, 2005.

[16] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3, pp. 239–258, 2004.

[17] G. P. Bernad, L. Denise, and P. Réfrégier, "Hierarchical feature-based classification approach for fast and user-interactive SAR image interpretation," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 1, pp. 117–121, January 2009.

[18] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. Benediktsson, "Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2008–2021, April 2012.

[19] S. Beucher, "Marker-controlled segmentation: an application to electrical borehole imaging," *Journal of Electronic Imaging*, vol. 1, no. 2, p. 136, Apr. 1992.

[20] A. Bieniek and A. Moga, "An efficient watershed algorithm based on connected components," *Pattern Recognition*, vol. 33, no. 6, pp. 907–916, 2000.

[21] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 2–16, 2010.

[22] T. Blaschke and J. Strobl, "Whats wrong with pixels? some recent developments interfacing remote sensing and gis," *GeoBIT/GIS*, vol. 6, no. 1, pp. 12–17, 2001.

[23] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *Int. J. Comput. Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006. [Online]. Available: http://dx.doi.org/10.1007/s11263-006-7934-5

[24] D. Bratasanu, I. Nedelcu, and M. Datcu, "Interactive spectral band discovery for exploratory visual analysis of satellite images," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 5, no. 1, pp. 207–224, February 2012.

[25] L. Bruzzone, M. Marconcini, U. Wegmller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1321–1334, June 2004.

[26] L. Bruzzone and F. Bovolo, "A Novel Framework for the Design of Change-Detection Systems for Very-High-Resolution Remote Sensing Images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[27] M. Cagnazzo, G. Poggi, and L. Verdoliva, "Region-Based Transform Coding of Multispectral Images," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2916–2926, Dec. 2007.

[28] F. Calderero, F. Eugenio, J. Marcello, and F. Marques, "Multispectral Cooperative Partition Sequence Fusion for Joint Classification and Hierarchical Segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1012–1016, Nov. 2012.

[29] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[30] A. Carleer, O. Debeir, and E. Wolff, "Assessment of very high spatial resolution satellite image segmentations," *Photogrammetric engineering and remote sensing*, vol. 71, no. 11, pp. 1285–1294, 2005.

[31] H. Cheng, X. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition*, vol. 34, no. 12, pp. 2259 – 2281, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320300001497

[32] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35 – 46, 1991.

[33] M. Datcu and K. Seidel, "Human centered concepts for exploration and understanding of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 52–59, 2005.

[34] G. 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 *IEEE International Geoscience and Remote Sensing Symposium*, 1997, pp. 1047–1050.

[35] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, , and L. Verdoliva, "Exploiting Patch Similarity for SAR Image Processing: The Nonlocal Paradigm," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 69–78, 2014.

[36] C. D'Elia, G. Poggi, and G. Scarpa, "Improved tree-structured segmentation of remote sensing images," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, no. 1, August 2003, pp. 1805–1807.

[37] C. D'Elia, G. Poggi, and G. Scarpa, "A tree-structured Markov random field model for Bayesian image segmentation." *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1259–73, Jan. 2003.

[38] S. Dellepiane, E. Angiati, and G. Vernazza, "Processing and segmentation of COSMO-SkyMed images for flood monitoring," in *IEEE International Geoscience and Remote Sensing Symposium*, 2010, pp. 4807–4810.

[39] H. Deng and D. Clausi, "Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 528–538, March 2005.

[40] J. Deng, Y. Ban, J. Liu, L. Li, X. Niu, and B. Zou, "Hierarchical segmentation of multitemporal radarsat-2 sar data using stationary wavelet transform and algebraic multigrid method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4353–4363, July 2014.

[41] V. Dey, Y. Zhang, and M. Zhong, "A review on image segmentation techniques with remote sensing perspective," in *ISPRS TC VII Symposium - 100 Years ISPRS*, W. Wagner and B. Szeékely, Eds., vol. XXXVIII. IAPRS, 2010, pp. 31–42.

[42] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1596–1615, March, 2014.

[43] S. Dickinson, M. Pelillo, and R. Zabih, "Introduction to the special section on graph algorithms in computer vision," *IEEE PAMI*, vol. 23, no. 10, pp. 1049–1052, Oct. 2001.

[44] C. H. Q. Ding, X. He, H. Zha, M. Gu, and H. D. Simon, "A min-max cut algorithm for graph partitioning and data clustering," in *Proc. IEEE ICDM*, 2001, pp. 107–114.

[45] J. dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. F. ao, "Interactive multiscale classification of high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 4, pp. 2020–2034, August 2013.

[46] J. A. dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. da S. Torres, and A. X. Falao, "Multiscale Classification of Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3764–3775, Oct. 2012.

[47] L. Drăguţ, D. Tiede, and S. R. Levick, "ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *International Journal of Geographical Information Science*, vol. 24, no. 6, pp. 859–871, Apr. 2010.

[48] Q. Du, N. H. Younan, R. King, , and V. P. Shah, "On the Performance Evaluation of Pan-Sharpening Techniques," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 518–522, 2007.

[49] e geos. COSMO-SkyMed Image Calibration. [Online]. Available: http://www.e-geos.it/products/pdf/COSMO-SkyMed-Image_Calibration.pdf

[50] A. Errico, C. V. Angelino, L. Cicala, G. Persechino, C. Ferrara, M. Lega, A. Vallario, C. Parente, G. Masi, R. Gaetano, G. Scarpa, D. Amitrano, G. Ruello, L. Verdoliva, and G. Poggi, "Detection of environmental hazards through the feature-based fusion of optical and sar data: a case study in southern italy," *International Journal of Remote Sensing*, vol. 36, no. 13, pp. 3345–3367, 2015. [Online]. Available: http://dx.doi.org/10.1080/01431161.2015.1054960

[51] A. Errico, C. V. Angelino, L. Cicala, D. P. Podobinski, G. Persechino, C. Ferrara, M. Lega, A. Vallario, C. Parente, G. Masi, R. Gaetano,

G. Scarpa, D. Amitrano, G. Ruello, L. Verdoliva, and G. Poggi, "Sar/multispectral image fusion for the detection of environmental hazards with a gis," pp. 924 503–924 503–8, 2014. [Online]. Available: http://dx.doi.org/10.1117/12.2066476

[52] 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, Jul. 2006.

[53] M. Esposito, F. P. Serpe, F. Neugebauer, S. Cavallo, P. Gallo, G. Colarusso, L. Baldi, G. Iovane, and L. Serpe, "Contamination Levels and Congener Distribution of PCDDs, PCDFs and Dioxin-Like PCBs in Buffalo's Milk from Caserta Province (Italy)," *Chemosphere*, vol. 79, no. 3, pp. 341–348, 2010.

[54] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and sar satellite images by exploring the spatial relationship of the improved sift," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 657–660, 2013.

[55] D. Fatta-Kassinosa, I. K. Kalavrouziotisb, P. H. Koukoulakisc, and M. I. Vasqueza, "The Risks Associated with Wastewater Reuse and Xenobiotics in the Agroecological Environment," *Environmental Health Perspectives*, vol. 409, no. 19, pp. 3555–3563, 2011.

[56] M. Fauvel, "Spectral and spatial methods for the classification of urban remote sensing data," Ph.D. dissertation, Grenoble Institute of Technology, University of Iceland, 11 2007.

[57] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77

[58] A. Ferretti, A. Monti-Guarnieri, C. Prati, F. Rocca, and D. Massonnet, *InSAR principles: guidelines for SAR interferometry processing and interpretation*. Postbus 229, 2200 AG Noordwijk: ESA Publications, ESTEC, 2007.

[59] G. Fontanelli, A. Crema, R. Azar, D. Stroppiana, P. Villa, and M. Boschetti, "Crop Mapping Using Optical and SAR Multi-Temporal

Seasonal Data: A Case Study in Lombardy Region, Italy," in *IEEE International Geoscience and Remote Sensing Symposium*, 2014, pp. 1489–1492.

[60] G. Franceschetti, A. Iodice, D. Riccio, G. Ruello, , and S. Cimmino, "SAR Raw Signal Simulation for Urban Structures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1986–1995, 2003.

[61] R. Gaetano, D. Amitrano, G. Masi, G. Poggi, G. Ruello, L. Verdoliva, and G. Scarpa, "Exploration of multitemporal cosmo-skymed data via interactive tree-structured mrf segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 7, pp. 2763–2775, July 2014.

[62] R. Gaetano, D. Amitrano, G. Masi, G. Poggi, G. Ruello, L. Verdoliva, and G. Scarpa, "Interactive segmentation of high resolution synthetic aperture radar data by tree-structured mrf," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, July 2014, pp. 3734–3737.

[63] R. Gaetano, G. Masi, G. Poggi, L. Verdoliva, and G. Scarpa, "Marker-controlled watershed-based segmentation of multiresolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 2987–3004, June 2015.

[64] R. Gaetano, G. Masi, G. Scarpa, and G. Poggi, "A marker-controlled watershed segmentation: Edge, mark and fill," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, July 2012, pp. 4315–4318.

[65] R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical texture-based segmentation of multiresolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2129–2141, 2009.

[66] R. Gaetano, G. Moser, G. Poggi, G. Scarpa, and S. B. Serpico, "Region-Based Classification of Multisensor Optical-SAR Images," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 4. IEEE, 2008, pp. IV – 81–IV – 84.

[67] C. P. Gerba and J. E. Smith, "Sources of Pathogenic Microorganisms and Their Fate during Land Application of Wastes," *Journal of Environmental Quality*, vol. 34, no. 1, pp. 42–48, 2005.

[68] R. N. Giere, "Using models to represent reality," in *Model-Based Reasoning in Scientific Discovery*, L. Magnani, N. J. Nersessian, and P. Thagard, Eds. Springer US, 1999, pp. 41–57.

[69] R. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics and Image Processing*, vol. 29, no. 1, pp. 100–132, 1985.

[70] G. J. Hay and G. Castilla, " Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 6, 2006.

[71] L. Horrigan, R. S. Lawrence, and P. Walker, "How Sustainable Agriculture Can Address the Environmental and Human Health Harms of Industrial Agriculture," *Environmental Health Perspectives*, vol. 10, no. 5, pp. 445–456, 2002.

[72] R. Infascelli, S. Faugno, S. Pindozzi, R. Pelorosso, and L. Boccia, "The Environmental Impact of Buffalo Manure in Areas Specialized in Mozzarella Production, Southern Italy," *Geospatial Health*, vol. 5, no. 1, pp. 131–137, 2010.

[73] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2104–2120, 2004.

[74] J. H. Jang, S. D. Kim, and J. B. Ra, "Enhancement of Optical Remote Sensing Images by Subband-Decomposed Multiscale Retinex With Hybrid Intensity Transfer Function," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 983–987, Sep. 2011.

[75] X. Jin, "Segmentation-based image processing system," Patent US 20 090 123 070 A1, 2007.

[76] A. W. Jongbloed and N. P. Lenis, "Environmental Concerns about Animal Manure," *Journal of Animal Science*, vol. 76, no. 10, pp. 2641–2648, 1998.

[77] P. Kelly, H. Derin, and K. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 10, pp. 1628–1641, October 1988.

[78] I. B. Kerfoot and Y. Bresler, "Theoretical analysis of multispectral image segmentation criteria," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 798–820, Jun 1999.

[79] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant, "Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology," *Pattern Recognition*, vol. 45, no. 2, pp. 685 – 706, 2012.

[80] C. Kurtz, N. Passat, A. Puissant, and P. Gançarski, "Hierarchical segmentation of multiresolution remote sensing images," in *Mathematical Morphology and Its Applications to Image and Signal Processing*, ser. Lecture Notes in Computer Science, P. Soille, M. Pesaresi, and G. Ouzounis, Eds. Springer Berlin Heidelberg, 2011, vol. 6671, pp. 343–354.

[81] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," US Patent 6,011,875, 200.

[82] S. Lefèvre, "Knowledge from markers in watershed segmentation," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, W. Kropatsch, M. Kampel, and A. Hanbury, Eds. Springer Berlin Heidelberg, 2007, vol. 4673, pp. 579–586.

[83] P. Li, J. Guo, B. Song, and X. Xiao, "A Multilevel Hierarchical Image Segmentation Method for Urban Impervious Surface Mapping Using Very High Resolution Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 103–116, Mar. 2011.

[84] Z. Li and J. Bethel, "Image coregistration in sar interferometry," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Services, XXXVII (B1)*, 2008, pp. 433–438.

[85] V. Madhok and D. A. Landgrebe, "A process model for remote sensing data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 680–686, 2002.

[86] P. Maglione, C. Parente, and A. Vallario, "Coastline Extraction Using High Resolution WorldView-2 Satellite Imagery," *European Journal of Remote Sensing*, vol. 47, pp. 685–699, 2014.

[87] K. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1153–1163, June 2006.

[88] T. R. Martha, N. Kerle, C. J. van Westen, V. Jetten, and K. V. Kumar, "Segment Optimization and Data-Driven Thresholding for Knowledge-Based Landslide Detection by Object-Based Image Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4928–4943, Dec. 2011.

[89] S. Martinis and A. Twele, "A hierarchical spatio-temporal Markov model for improved flood mapping using multi-temporal X-band SAR data," *Remote Sensing*, vol. 2, pp. 2240–2258, 2010.

[90] G. Masi, R. Gaetano, G. Poggi, and G. Scarpa, "A ground truth design tool for multiresolution images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, July 2015, pp. 4999–5002.

[91] G. Masi, R. Gaetano, G. Poggi, and G. Scarpa, "Superpixel-based segmentation of remote sensing images through correlation clustering," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, July 2015, pp. 1028–1031.

[92] G. Masi, R. Gaetano, G. Scarpa, and G. Poggi, "Dynamic segmentation for image information mining," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, July 2010, pp. 1992–1995.

[93] G. Masi, G. Scarpa, R. Gaetano, and G. Poggi, "A watershed-based segmentation technique for multiresolution data," in *Image Analysis and Processing–ICIAP 2013*. Springer Berlin Heidelberg, 2013, pp. 241–250.

[94] H. Menzi, O. Oenema, C. Burton, O. Shipin, P. Gerber, T. Robinson, and G. Franceschini, "Impacts of Intensive Livestock Production and Manure Management on the Environment," in *Livestock in a Changing Landscape: Drivers, Consequences and Responses*, H. Steinfeld, H. Mooney, L. E. Neville, and F. Schneider, Eds. Washington, DC: Island Press, 2010.

[95] S. Mikes, M. Haindl, and G. Scarpa, "Remote sensing segmentation benchmark," in *Pattern Recognition in Remote Sensing (PRRS), 2012 IAPR Workshop on*, Nov 2012, pp. 1–4.

[96] S. Mikes, M. Haindl, G. Scarpa, and R. Gaetano, "Benchmarking of remote sensing segmentation methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015.

[97] G. Moser and S. B. Serpico, "Joint classification of panchromatic and multispectral images by multiresolution fusion through Markov random fields and graph cuts," in *2011 17th International Conference on Digital Signal Processing (DSP)*. IEEE, Jul. 2011, pp. 1–8.

[98] X. Niu and Y. Ban, "An adaptive contextual SEM algorithm for urban land cover mapping using multitemporal high-resolution polarimetric SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1129–1139, August 2012.

[99] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1277 – 1294, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0031320393901351

[100] S. Parrilli, M.Poderico, C.V.Angelino, and L.Verdoliva, "A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 606–616, February 2012.

[101] E. Pasolli, F. Melgani, N. Alajlan, and N. Conci, "Optical image classification: A ground-truth design framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3580–3597, June 2013.

[102] T. Pavlidis, "Image analysis," *Annual review of Computer Science*, vol. 3, pp. 121–146, 1988.

[103] D. Pham, C. Xu, and P. J.L., "A survey of current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315–337, 2000.

[104] G. Poggi and R. Ragozini, "Image segmentation by tree-structured Markov random fields," vol. 6, no. 7, pp. 155–157, July 1999.

[105] G. Poggi, G. Scarpa, and J. Zerubia, "Supervised segmentation of remote sensing images based on a tree-structured MRF model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1901–1911, August 2005.

[106] G. Poggi, F. Sica, L. Verdoliva, G. Fornaro, D. Reale, and S. Verde, "Non-local methods for filtering interferometric sar datasets," in *Advances in Radar and Remote Sensing (TyWRRS), 2012 Tyrrhenian Workshop on*. IEEE, 2012, pp. 136–139.

[107] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Computer Journal*, vol. 7, no. 2, pp. 155–162, 1964.

[108] A. Puissant, S. Lefèvre, S. Rougier, and J.-P. Malet, "Automated mapping of coastline from high resolution satellite images using supervised segmentation," in *Proceedings of the 4th GEOBIA*, Rio de Janeiro, Brazil, 2012, pp. 515–517.

[109] L. Pulvirenti, N. Pierdicca, M. Chini, and L. Guerriero, "Monitoring flood evolution in vegetated areas using COSMO-SkyMed data: The Tuscany 2009 case study," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 4, pp. 1807–1816, August 2013.

[110] E. Rignot and R. Chellappa, "Segmentation of polarimetric Synthetic Aperture Radar data," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 281–300, July 1992.

[111] P. A. Rosen, S. Hensley, I. R. Joughin, F. K. Li, S. N. Madsen, E. Rodriguez, and R. M. Goldstein, "Synthetic Aperture Radar Interferometry," *Proceedings of the IEEE*, vol. 3, no. 88, pp. 333–382, 2000.

[112] J. C. Russ, *The Image Processing Handbook, Sixth Edition*. CRC Press, 2011.

[113] B. Sadoun and S. Al Rawashdeh, "Applications of GIS and remote sensing techniques to land use management," in *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA).*, May 2009, pp. 233–237.

[114] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 561–576, January 2000.

[115] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical multiple Markov chain model for unsupervised texture segmentation," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1830–1843, August 2009.

[116] G. Scarpa, G. Masi, L. Verdoliva, G. Poggi, and R. Gaetano, "Recursive-tfr algorithm for segmentation of remotely sensed images," in *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on*, Nov 2012, pp. 174–181.

[117] G. Scarpa, G. Masi, R. Gaetano, L. Verdoliva, and G. Poggi, "Dynamic hierarchical segmentation of remote sensing images," in *Image Analysis and Processing–ICIAP 2013*. Springer Berlin Heidelberg, 2013, pp. 371–380.

[118] T. Schellenberger, B. Ventura, M. Zebisch, and C. Notarnicola, "Wet snow cover mapping algorithm based on multitemporal COSMO-SkyMed X-band SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 1045–1053, June 2012.

[119] J. Schiewe, "Segmentation of high-resolution remotely sensed data concepts, applications and problems," in *Symposium on Geospatial theory, Processing and Applications*, 2002.

[120] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu, "Spatial information retrieval from remote-sensing images - Part II: Gibbs-Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1446–1455, 1998.

[121] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE PAMI*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[122] F. Sica, L. Alparone, F. Argenti, G. Fornaro, A. Lapini, and D. Reale, "Benefits of blind speckle decorrelation for insar processing," in *SPIE Remote Sensing*. International Society for Optics and Photonics, 2014, pp. 92 430D–92 430D.

[123] F. Sica, D. Reale, G. Poggi, L. Verdoliva, and G. Fornaro, "Nonlocal adaptive multilooking in sar multipass differential interferometry," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 8, no. 4, pp. 1727–1742, 2015.

[124] B. Sirmacek and C. Unsalan, "A Probabilistic Framework to Detect Buildings in Aerial and Satellite Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 211–221, Jan. 2011.

[125] P. Smits and S. Dellepiane, "Synthetic Aperture Radar image segmentation by a detail preserving Markov random field approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 844–857, July 1997.

[126] L.-K. Soh and C. Tsatsoulis, "Segmentation of satellite imagery of natural scenes using data mining," *IEEE Geoscience and Remote Sensing Letters*, vol. 37, no. 2, pp. 1086–1099, 1999.

[127] P. Soille, *Morphological Image Analysis: Principles and Applications*. Springer, 2004.

[128] A. S. Solberg, T. Taxt, and A. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 100–113, January 1996.

[129] S. Suri and P. Reinartz, "Mutual-information-based registration of terrasar-x and ikonos imagery in urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 939–949, 2010.

[130] C. V. Tao and Y. Hu, "Image rectification using a generic sensor model - rational function model," in *International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B3*, 2000, pp. 874–881.

[131] Y. Tarabalka, J. Chanussot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest

grown from automatically selected markers," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 5, pp. 1267–1279, Oct 2010.

[132] Y. Tarabalka, J. Tilton, J. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 1, pp. 262–272, Feb 2012.

[133] J. Tilton, "Image segmentation by region growing and spectral clustering with a natural convergence criterion," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium Proceedings, IGARSS'98*, vol. 4, Jul 1998, pp. 1766–1768 vol.4.

[134] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution sar images," *IEEE Transactions on Geoscience and Remote Senisng*, vol. 42, no. 10, pp. 2046–2057, October 2004.

[135] M. Triassi, R. Alfano, M. Illario, A. Nardone, O. Caporale, , and P. Montuori, "Environmental Pollution from Illegal Waste Disposal and Health Effects: A Review on the Triangle of Death," *International Journal of Environmental Research and Public Health*, vol. 12, no. 2, pp. 1216–1236, 2015.

[136] S. Valero, P. Salembier, and J. Chanussot, "New hyperspectral data representation using binary partition tree," in *2010 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Jul. 2010, pp. 80–83.

[137] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 4, pp. 1430–43, Apr. 2013.

[138] O. Veksler, "Image segmentation by nested cuts," in *IEEE CVPR*, 2000.

[139] S. Wang and J. M. Siskind, "Image segmentation with ratio cut," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 675–690, 2003.

[140] C. Wemmert, A. Puissant, G. Forestier, and P. Gancarski, "Multiresolution Remote Sensing Image Clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 533–537, Jul. 2009.

[141] G.-S. Xia, C. He, and H. Sun, "A rapid and automatic MRF-based clustering method for SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 596–600, October 2007.

[142] P. Xiao, X. Feng, S. Zhao, and J. She, "Multispectral IKONOS image segmentation based on texture marker-controlled watershed algorithm," in *International Symposium on Multispectral Image Processing and Pattern Recognition*, Y. Wang, J. Li, B. Lei, and J. Yang, Eds. International Society for Optics and Photonics, Nov. 2007, pp. 67 900U–67 900U–9.

[143] W. Yang, D. Dai, B. Triggs, and G. Xia, "SAR-based terrain classification using weakly supervised hierarchical Markov aspect models," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4232–4243, September 2012.

[144] Y. Yang, H. Sun, and C. He, "Supervised SAR image MPM segmentation based on region-based hierarchical model," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 517–521, October 2006.

[145] Y.Ding, Y.Li, and W.Yu, "SAR image classification based on CRFs with integration of local label context and pairwise label compatibility," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 300–306, Jan. 2014.

[146] L. Yi, G. Zhang, and Z. Wu, "A Scale-Synthesis Method for High Spatial Resolution Remote Sensing Image Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 4062–4070, Oct. 2012.

[147] J. Yuan, D. Wang, and R. Li, "Remote Sensing Image Segmentation by Combining Spectral and Texture Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–9, 2013.

[148] Y. Yusuf, I. Alimuddin, J. T. S. Sumantyo, , and H. Kuze, "Assessment of Pan-Sharpening Methods Applied to Image Fusion of Remotely Sensed Multi-Band Data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 165–175, 2012.

[149] A. E. Zaart, D. Ziou, S. Wang, and Q. Jiang, "Segmentation of {SAR} images," *Pattern Recognition*, vol. 35, no. 3, pp. 713 – 724, 2002, image/Video Communication. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132030100070X

[150] Y. Zhang, "Evaluation and comparison of different segmentation algorithms," *Pattern Recognition Letters*, vol. 18, no. 10, pp. 963 – 974, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865597000834

[151] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.