

TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI “FEDERICO II”

DIPARTIMENTO DI INGEGNERIA ELETTRICA  
E DELLE TECNOLOGIE DELL'INFORMAZIONE

DOTTORATO DI RICERCA IN  
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

---

LOCAL IMAGE DESCRIPTORS FOR  
BIOMETRIC LIVENESS DETECTION

---

**Diego Gragnaniello**

Il Coordinatore del Corso di Dottorato    Il Tutore

Ch.mo Prof. Daniele RICCIO

Ch.ma Prof.ssa Luisa VERDOLIVA

A. A. 2014-2015



*“Non semper ea sunt quae videntur,  
decipit frons prima multos.”*



# Contents

<b>List of Figures</b>	<b>vii</b>
<b>1 Liveness Detection</b>	<b>1</b>
<b>2 State of the art</b>	<b>5</b>
2.1 Dynamic features . . . . .	5
2.2 Global features . . . . .	6
2.3 Local features . . . . .	8
2.4 Independent Quantization of Features . . . . .	13
2.4.1 LBP . . . . .	13
2.4.2 CoA-LBP and Ric-LBP . . . . .	15
2.4.3 LPQ . . . . .	16
2.4.4 WLD . . . . .	17
2.4.5 BSIF . . . . .	19
2.5 Joint Quantization of Features . . . . .	19
2.5.1 SIFT, DAISY . . . . .	19
2.5.2 SID . . . . .	21
<b>3 Main contributions</b>	<b>25</b>
3.1 Wavelet-Markov features . . . . .	25
3.2 LPQ and WLD concatenation . . . . .	27
3.3 LCPD . . . . .	29
3.3.1 Spatial-domain component . . . . .	29
3.3.2 Transform-domain component . . . . .	31
3.3.3 Combination . . . . .	32
3.4 LBP from the residue image . . . . .	33
3.4.1 Local Binary Patterns for mobile devices . . . . .	34
3.4.2 Emphasizing local patterns . . . . .	36

---

3.4.3	Liveness detection algorithm . . . . .	38
3.5	SID and Bag-of-Features . . . . .	39
<b>4</b>	<b>Experimental results</b>	<b>43</b>
4.1	Fingerprint . . . . .	44
4.2	Finger Veins . . . . .	51
4.3	Iris . . . . .	55
4.4	Contact lens classification . . . . .	65
4.5	Face . . . . .	72
4.6	Cells classification . . . . .	75
4.7	Complexity . . . . .	82
	<b>Conclusion</b>	<b>85</b>

# List of Figures

1.1	A typical biometric authentication system equipped with a liveness detection module. . . . .	2
2.1	Feature extraction step. . . . .	10
2.2	Independent Features Quantization and Joint Features Quantization . . . . .	14
2.3	LPQ patch processing . . . . .	18
2.4	Different sampling grids. . . . .	20
2.5	Log-polar grid and spatially varying filtering kernels. . . . .	23
3.1	The pyramid of approximations and details subbands used in the algorithm . . . . .	27
3.2	Example of differential excitation and gradient orientation fields for live and fake fingerprint images . . . . .	28
3.3	WLD histograms for live and fake fingerprints . . . . .	29
3.4	Importance of contrast in LCPD . . . . .	30
3.5	Contrast-based field block-diagram . . . . .	31
3.6	Examples of LCPD contrast and phase fields. . . . .	32
3.7	LCPD block-diagram. . . . .	33
3.8	Resulting LCPD histograms for live and fake fingerprints. . . . .	34
3.9	Some neighborhood systems used by LBP. . . . .	35
3.10	Effectiveness of the residual image . . . . .	37
3.11	Block-diagram for the feature extraction step of the proposed method. . . . .	37
4.1	Live and fake fingerprint samples. . . . .	46
4.2	ROCs for the Italdata 2011 dataset. . . . .	50
4.3	Live and printed fake finger veins samples. . . . .	52
4.4	Live and printed fake iris samples. . . . .	59

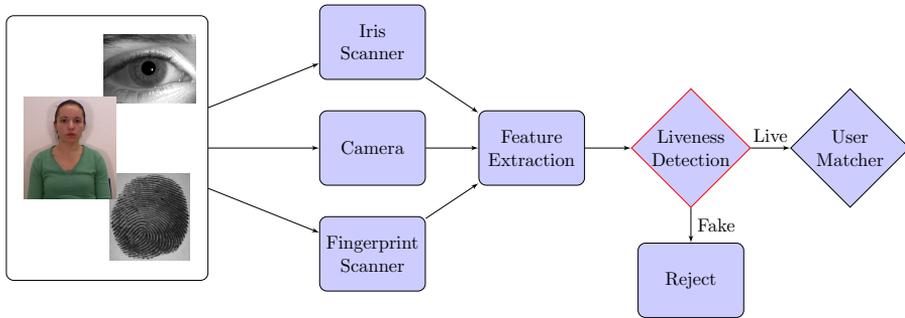
4.5	Live and fake printed iris examples from MobBIOfake database . . . . .	60
4.6	Live and fake screen iris images from MICHE database . .	61
4.7	Live and fake iris samples wearing contact lenses. . . . .	62
4.8	ROCs for Cogent dataset. . . . .	63
4.9	Iris segmentation algorithm . . . . .	67
4.10	Different segmentation regions. . . . .	68
4.11	Segmentation performances. . . . .	70
4.12	Feature extraction procedure. . . . .	71
4.13	Contact lens samples. . . . .	72
4.14	Live and fake face samples. . . . .	75
4.15	Examples of different staining patterns . . . . .	78
4.16	Examples of BoW histograms for cells classification . . . .	80

# Chapter 1

## Liveness Detection

**B** iometric systems are more and more often used for authentication in various security applications. By relying on physiological attributes of each individual they offer simplicity of use and reliability at the same time, avoiding typical problems of systems based on the use of passwords, which can be forgotten, transferred or stolen. Fingerprint, face, and iris are the biometric traits most frequently used in present authentication systems [134]. Of course, biometric systems have their own weaknesses, in particular they are relatively vulnerable to some sophisticated forms of spoofing. For example, fingerprint-based systems are among the most commonly used and, for this very same reason, more subject to attacks. Indeed, early systems could be easily fooled by fake fingerprints, reproduced on simple molds made of materials such as silicone, Play-Doh, clay or gelatin [88, 37]. Likewise, iris-based systems can be attacked with fake irises printed on paper or on wearable plastic lenses [87, 113], while face-based systems can be fooled with sophisticated 3D masks (easily bought online once a few photos of the subject are provided) or, again, with faces printed on paper [67] or, also, with video reproduced on mobile and tablet devices [16].

Clearly, these attacks have elicited a race towards some reliable anti-spoofing systems, and in particular towards liveness detection techniques, which use various physiological properties to distinguish between real and fake traits. Fig.1.1 shows the typical placement of a liveness detection module in the context of a biometric authentication system. In principle, besides being reliable, blocking attackers and allowing legitimate users in the systems, liveness detection methods should posses



**Figure 1.1:** A typical biometric authentication system equipped with a liveness detection module. Once the biometric trait has been acquired a feature extraction process followed by a classification step labels the image as fake or live. Only in this last case the image is considered for the recognition phase.

other important properties [82], being *non-invasive*, *user friendly*, *fast*, and *low cost*.

A large number of methods have been proposed in recent years to combat spoofing [83, 122]. Some of them rely on the detection of vitality signs at the acquisition stage. Hence they require additional hardware embedded in the sensor which verifies vitality by measuring particular intrinsic properties of a living trait, such as temperature, odor, sweat, blood pressure, or reflection properties of the eye [4, 105], sometimes also in response to specific stimuli [69]. By combining multiple sources of information, this approach turns out to be more resilient to specific attacks, providing a very good reliability. However, it is a relatively expensive and rigid solution, potentially vulnerable to attacks not considered at design time.

On the contrary, software-based methods, based on signal-processing techniques, are certainly more appealing, for their reduced cost and invasiveness, and their higher flexibility. They try to detect liveness by analyzing synthetic image features that are peculiar of vital biometric traits and not easily reproduced on fakes. In some cases, features are singled out based on a deep study of the physics of the problem and/or a careful analysis of the statistical behaviour of the captured images. A large number of such methods have been proposed in recent years [30, 92, 107, 125, 2, 59, 17, 93, 126, 96, 86, 127, 84, 36, 36], based on

---

clever and well-founded ideas, testifying on the relevance and difficulty of this problem. As an example, a few methods base their decision on some ridge-valley characteristics peculiar of live fingerprints. Other techniques are based on more generic features, like the energy observed in certain frequency bands, computed globally on the whole image.

In parallel with these approaches, based on global and high-level image descriptors, techniques based on local descriptors (LD) have been taking hold recently [94, 40, 47, 124, 143, 31]. LDs, as the name suggest, describe the statistical behavior observed *locally* in very small patches of the image by means of histograms (frequencies of occurrence, empirical probability distributions) collected over the ensemble of all patches. These histograms are then used as features to classify the image by means of conventional classification tools. LD-based approaches have been used in a large number of image classification problems, from image mining and retrieval [154], to texture classification [102, 137], face recognition [1, 12], steganalysis [35], forgery detection [6, 18], image quality assessment [91], always with impressive results.

It is somewhat surprising that these alternative methods, showing often little or no data-specific clues, overcome so easily the *physics-based* approaches. On the other hand, liveness detection, as all tasks related with digital security, can be seen as a game with two players, and it is only reasonable to expect that methods based on macroscopic features will be sooner or later tricked by smart attackers, equipped with better materials and better knowledge of the specific biometric traits statistics. Local descriptors represent a natural evolution towards the discovery of microscopic features that are more discriminating and also harder to tamper with. Till now, however, only general-purpose descriptors have been usually considered for the liveness detection task, while image and video descriptors conceived specifically for these very specific biometrics traits can provide a still better performance.

The thesis is organized as follows

**Chapter 2** introduces a general framework which describes most of the local feature-based techniques for different biometric traits: fingerprint, iris and face. In particular, two possible approaches are presented depending on whether features associated with a pixel are quantized independently or jointly.

**Chapter 3** presents the main contributions of this work to the liveness detection research topic. In particular, it has been proposed a new local descriptor designed for the fingerprint liveness detection task, namely LCPD (Local Contrast Phase Descriptor). In addition, a method based on rich local features later compacted using the Bag-of-Words paradigm is presented. Particular attention is given also to the design of liveness detection tools for mobile devices implementation, for which suitable low-complexity features are adopted.

**Chapter 4** describes the experimental results obtained for various biometric traits. In this analysis, carried out on publicly available datasets, we have always considered homogeneous basic tools for classification (e.g. linear-kernel SVM) and adopted the same experimental protocol, so as to put all competing solutions on the same ground. To measure the robustness of the approach based on the Bag-of-Words paradigm, results for a cell image multi-class classification task are also presented.

**Chapter 5** draws some conclusions, summarizing the results obtained for each specific biometric traits, and proposes future work guidelines.

## Chapter 2

# State of the art

There is a large body of literature on biometric spoofing, especially thanks to the competitions held in recent years [85, 144, 40, 11] which provided publicly available databases to test performance. In the following we will review the literature on liveness detection for fingerprint, iris and face images. The majority of techniques follow the paradigm shown in Fig.2.1, differing mainly in the type of features extracted. We will first describe techniques based on dynamic features, which require multiple acquisitions. Then we will focus on techniques based on a single sample image, distinguishing between the approaches that analyze the image as a whole, extracting *global features*, and those that extract *local features* at each pixel to summarize them eventually in a single feature vector, typically by means of histograms.

### 2.1 Dynamic features

Assuming that multiple acquisitions are allowed, some peculiar characteristics of living bodies can be exploited to detect spoofing. For fingerprints, some intrinsic characteristics of the skin can be used, like perspiration, deformation or elasticity [107, 2, 86]. Likewise, for irises, controlled light reflection, pupil dynamics, and pupil constriction have been used as a basis for spoofing countermeasures [105, 56]. Eye blinking has been used also for face liveness detection, as well as the movement of lips [57, 106]. Other techniques based on motion analysis rely on the observation that a 3D face generates a 2D motion which is higher at the center (e.g., nose) than in peripheral regions, (e.g., ears), while a trans-

lated photograph generates a constant motion field. In [65], for example, face parts are detected by a model-based Gabor decomposition, and the trajectories of various parts are evaluated by optical flow estimation. A different approach is followed in [62] based on the fact that face and background follow different motion models, in a live video, contrary to what happens in the presence of a moving photo. Finally, in [108] the authors base liveness detection on a spatio-temporal *dynamic texture* information, which combines facial appearance and motion.

Although these methods may be highly reliable, they require additional hardware and/or an increased processing time. Moreover, the user may experience an increased waiting time before being granted access. In addition, in some cases, the attack is intrinsically resilient to such countermeasures, like when a human attacker wears cosmetic lenses with fake irises.

## 2.2 Global features

A common approach consists in detecting the (expected) lower quality of images relative to fakes w.r.t. those originated by live samples. A warning is due on the fundamental hypothesis, though, since technological improvements, for materials, printers, etc., may, in time, frustrate the whole approach [53].

In [92], among the first techniques based on this principle, fingerprint coarseness is used as a discriminative feature, exploiting the imperfections in materials used for the fakes. The analysis is performed in the wavelet domain. A denoised version of the image is subtracted from the original one, then, the variance of the noise residual is estimated from the wavelet subbands corresponding to fine details along the various orientations, and used as a measure of coarseness.

Fourier analysis has been used for iris and face images. In particular, with reference to fakes based on printed images, Daugman [22] suggested to analyze the image in the Fourier domain to reveal the periodic pattern of dots left by the printer, while the reduced high-frequency content in photographs of faces has been used in [72]. In [127], ridges and valleys are extracted using a mask, and separate related features are computed both in the spatial and in transform (Fourier, wavelet) domains.

Always following the idea that spoofing produces low quality images, [36] considers some fundamental properties of fingerprints, like ridge

---

strength, continuity and clarity, collecting a set of ten discriminative quality-based features for fake detection. The same principle guides the definitions of the measures proposed for iris images, where properties like focus, blur, contrast are considered [39]. The same idea can be applied to different kind of attacks, like for example the print attack for iris- or face-based systems, exploiting the roughness of the paper surface to spot fake images. In [38] generic image quality measures, such as image sharpness, structural distortions, are considered as features, with no reference to any specific biometric trait, and can be therefore used indifferently for fingerprint, iris and face images.

In the literature, much attention has been also devoted to the phase, which is able to better preserve the correlation between neighboring samples [104] and, when evaluated locally, can give a more precise localization of patterns. Short-time Fourier transform (STFT) and the more general Gabor filter banks both capture well this type of information. Especially the latter has been extensively used to characterize textural information in different areas, and in particular for biometric feature identification [26, 149, 148]. The differences arising in inter-ridge distances, ridge frequencies and widths between live and fake fingerprints seem to be well captured by textural features extracted using Gabor filters, as shown in [93]. However, in this last work only global features are considered using second order statistics.

Morphology-based and perspiration-based features have been taken into account in [84], where a proper feature selection approach is also adopted.

Another large class of methods is based on the analysis of textures. In fact, biometrical traits are often characterized by abundant and strong textural information. This is certainly true for fingerprints, with their regular ridge-valley patterns<sup>1</sup>, but also for irises, where there are rich patterns of furrows, ridges, and pigment spots. Therefore, methods based on statistical texture analysis can be reasonably expected to provide good results. For example, textural features related to inter-ridge distances, and ridge frequencies and widths, extracted using Gabor filters [93], seem to discriminate well between live and fake samples. Likewise, distinctive textural features of the iris are defined in [52], where

---

<sup>1</sup>The fingerprint pattern is comprised of ridges, touching the sensor plane, and valleys, which remain far from it, and this ridge-valley structure typically looks different for fake and live images.

four features extracted from the gray level co-occurrence matrix have been derived. Transform-based methods drawn from the texture classification literature can be easily adapted to deal with fingerprints which, due to their strong regular structure, can be assimilated to textures themselves. These methods work on the energy spectrum of textures, assuming that different classes of texture exhibit different distributions of energy in the transform domain [111]. In [59] and [17], in particular, it was observed that the different ridge-valley periodicity of live and fake fingerprints translates into easily captured differences in their spectral ring patterns. Live fingerprints have higher energy content in the ring patterns than their fake counterparts, a property readily used for reliable classification. Other features related to spectral energy distribution have been proposed with reference to different transforms, like wavelet [125] and ridgelet [96], with similar results.

### 2.3 Local features

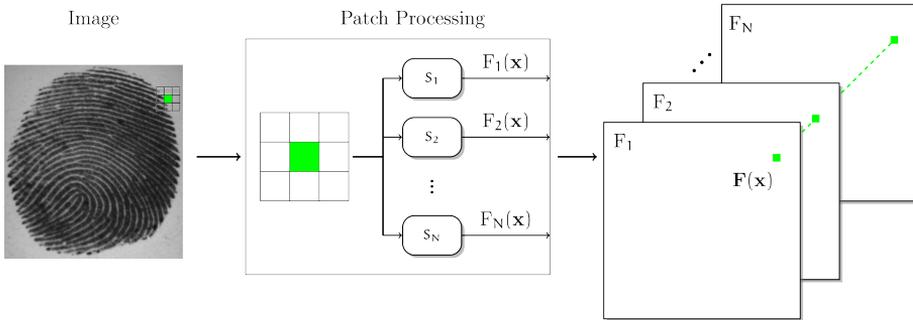
In the last few years the analysis of *micro*-textures has gained large popularity for a number of image processing tasks. To this end, the fine-scale statistical behavior is observed locally in small patches of the image, typically after some form of high-pass filtering meant to emphasize the high-frequency content. This form of textural analysis has proven very effective for liveness detection.

Some techniques for the detection of biometric spoofing are based on the Local Binary Patterns (LBP), a descriptor first proposed in 2002 [102] for texture classification and then gaining a long score of successes in very different and challenging tasks, such as face recognition [1], fingerprint recognition [110], camera identification [10], or image forgery detection [27]. These results testify on the potential of these tools and, more in general, on the effectiveness of the patch-processing that is a dominant paradigm in all branches of image processing and machine vision, from image denoising [21, 29], to no-reference measures of image quality [91], to texture classification [73]. In LBP the variations observed in a circular neighborhood of each image pixel are computed and encoded in a scalar, then a feature vector is built by computing the histogram of occurrences of such scalars over the whole image. LBP first appears in the fingerprint liveness detection literature in a 2008 paper [95] where wavelet-domain energy features are complemented by

---

the LBP descriptor. In 2009 [54] proposes the use of LBP to detect iris spoofing based on contact lenses. The iris is divided into six relevant subregions, these are rectified, and LBPs are extracted at various scales. Finally the Adaboost algorithm is used to learn the most discriminative regional LBP features. In [58], following [81], a multiscale version of LBP was proposed. The LBP operator was applied with large radii, after a regularizing low-pass Gaussian filtering, so as to take into account long-range dependencies. More recently, [143] divides the eye image in three regions: pupil, iris and sclera, using LBPs at multiple scales for each of them. The potential of LBP descriptor has been explored also for face liveness detection. In 2011 the method proposed in [80] divides the images in blocks, a combination of different LBP operators, obtained by varying sampling and scale, are then extracted from each block, and eventually concatenated. Further studies, conducted in [16], show that a similar performance can be obtained also by processing the whole image. Some extensions of LBP have been also investigated in [31] for the more challenging problem of detecting attacks by 3D face masks. Performance analysis highlights again the good performance obtained by the LBP descriptor, even by considering a single frame of a video. An interesting variant of LBP is proposed in [150], in the context of iris spoofing detection, where the SIFT descriptor is used to sort the LBP elementary features (two-pixel differences) in order of importance, defining a new rotation-invariant version. Instead, in [48] LBP is evaluated on the residual images, obtained through a proper high-pass filtering, for print-based and screen-based attacks in the context of a iris mobile biometric identification system.

In the literature, much attention has been also devoted to the local phase, which is able to better preserve the correlation between neighboring samples and can give a more precise characterization of patterns, especially when considering iris images [23]. Short-time Fourier transform (STFT), and the more general Gabor filter banks, both capture well this type of information. In [41] we find the first attempt to use the local phase for fingerprint liveness detection. It is based on the LPQ (Local Phase Quantization) descriptor, proposed in [103], which has a structure similar to LBP, but encodes some phase information extracted through a short-time Fourier transform of the local patch, rather than gradients. A new descriptor combining contrast-based and phase-based information, has been proposed in [47], obtaining very good results in



**Figure 2.1:** Feature extraction step. This process associates at the pixel of the image a feature vector. In general, the process can be sparse (only for some selected pixels) or dense (for each pixel). In our framework we consider dense approaches where the neighborhood of each pixel,  $\mathbf{x}$ , is analyzed to extract a discriminative feature vector:  $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_N(\mathbf{x})]$ .

fingerprint liveness detection. Gabor filters are used in [140], for the detection of fakes based on contact lenses. Iris-texton features are extracted using a Gabor filter bank and K-means clustering. A further improvement of this technique is proposed in [124] for the more general iris classification task. A novel texture representation method, called Hierarchical Visual Codebook (HVC), is used to encode the texture primitives of iris images, while dense SIFT descriptors are used to encode the gradient information of the local iris region. Eventually, K-means is used to extract compact features. Another approach [42], recently proposed for fingerprint liveness detection, considers a variant of SIFT and HOG called HIG (Histogram of Invariant Gradients) with the aim of preserving robustness to variations in gradient positions.

Other local descriptors, of course, do not fit in the above classes. In [45], for example, intra- and inter-band dependencies of wavelet coefficients are exploited to detect fake fingerprints. Another interesting evolution is represented by the BSIF (Binarized Statistical image Features) descriptor [60]. BSIF was used in [40] for fingerprint liveness detection with very competitive results. Rather than using a fixed set of filters, BSIF learns its own using the statistics of image patches and maximizing the statistical independence of the filter responses. Some very recent works have considered deep learning for biometric spoofing

---

detection. In particular, Convolutional Neural Networks (CNN) have been considered in [97] and [145] for the specific task of fingerprint and face liveness detection, respectively, while in [90] they have been applied to handle different biometrics traits. A CNN is a multi-layered representation, where each layer is obtained through a proper convolution and local pooling. This procedure allows to extract patterns and hence find deep representational connections. The filter weights can be learned directly from the data (although random weights are considered in both [97] and [90]) and a large number of hyper-parameters, like number of layers and filter sizes, must be optimized in the training phase, which makes this approach computation-intensive.

Despite the excellent results obtained, these proposals show little or no effort to adapt the descriptors to the specific fingerprint problem, and the experimental results do not point out any *killer* features, related to some physical or statistical trait of the data. A common effort, however, is to combine several descriptors [41, 43], as also done in other fields, using information coming from the space, frequency, and orientation domains [70], or concatenating the LBP and LPQ features [153]. In particular, in [43] we tested several popular descriptors for the liveness detection problem, i.e., LBP, LPQ and WLD, the Weber local descriptor [14] based on orientation and differential excitation. In that context, we also tested the descriptors obtained by concatenating the basic ones, that is, by forming a new feature of length  $\sum_i N_i$  by simply putting together several features of length  $N_i$ . Experiments were extremely informative. A first striking result, already mentioned in the Introduction, was the gap in performance gain provided by all classifiers based on local descriptors w.r.t. all other classifiers. Moreover, we observed a further significant gain when we tried some simple concatenations of descriptors. This was not obvious, a priori, because adding features increases dimensionality, a problem to be dealt with through some feature selection techniques. The performance improvement indicates that the new features are indeed discriminative or, more precisely, that the selected concatenation allows for a better exploration of the whole data space.

However, not all concatenations provided an equally good performance, and this shed some light on their respective importance. In particular, results showed that the phase-oriented features encoded by the LPQ descriptor hold a major discriminative power for this task. In hindsight, this is no surprise, considering the quasi-sinusoidal local be-

havior of fingerprint images, and the paramount importance of the phase in the Fourier-domain description of signals. Nonetheless, LPQ by itself scores worse than its concatenation with either LBP or WLD, indicating that some more information is required to complement the phase, very likely related with the local amplitude patterns encoded by LBP and by the first component of WLD.

This analysis of the literature, with the experimental results reported therein, confirmed that features based on a local analysis of textures are among the most promising for the liveness detection task. They can help telling apart different classes of images based on features that the human eye cannot catch. On the other hand, textures, and especially micro-textures, are everywhere, which suggest these features can be used with success in a wide range of situations.

A large number of computer vision tasks require some form of image classification. However, feeding a classifier with the images as they are is not only unwieldy, but also ineffective. In fact, studies on the mechanisms of early vision in mammals, proved that high-level tasks require, as a first fundamental step, the extraction of some local high-pass features, much more informative than sheer intensity.

These observations well explain the current quest towards compact and effective image descriptors based on expressive local features. A key step in this direction can be traced back to [71] where, to recognize surfaces made of different materials, a vector of features  $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_N(\mathbf{x})]$  was associated with each pixel  $\mathbf{x}$ , where each feature is obtained as the output of a specific linear filter. The potential of this paradigm, depicted in Fig.2.1, was soon recognized, and a large number of image descriptors based on local features were proposed in the following years, some of which are briefly reviewed in the reminder of this Section. As for compactness, however, we are now given  $N$  feature planes in place of a single (possibly multi-band) image. Therefore, further steps are necessary, see Fig.2.2, where the feature vector associated with a pixel is converted into a scalar by means of quantization, and a synthetic descriptor is eventually built by computing the histogram of these scalars (coded features) over the whole image.

This brief summary of concepts skips, obviously, countless important issues and details. In the following, however, we will focus on the crucial feature coding step, classifying descriptors in two big families, depending on whether the features associated with a pixel are quantized indepen-

dently or jointly. The pros and cons of the two solutions, depicted in Fig.2.2, are easily understood:

- with the *independent quantization of features*, the components of a feature vector are quantized separately, and the corresponding indexes are combined to obtain the scalar local descriptor; this solution is very simple, hence preferable for low-power contexts, or when a fast response is required;
- with the *joint quantization of features* the vector of features is quantized as a whole, through a suitable partition of the feature space (think of K-means clustering) and the partition index represents the local descriptor. Vector quantization (like more advanced solutions [13]) is relatively complex, and requires a careful training phase; on the other hand, exploiting joint dependencies can be expected to provide a better performance.

In the following, we explore some of the most well-known image descriptors fitting in these two classes. Our selection of methods, however, takes into account also on the availability of software code to carry out the reproducible experiments discussed in the Section.

## 2.4 Independent Quantization of Features

### 2.4.1 LBP

Let  $\mathbf{x}$  be a generic pixel, and  $\eta_i(\mathbf{x})$  the  $i$ -th of  $P$  neighbors sampled uniformly on a circle of radius  $R$  centered on  $\mathbf{x}$ , resorting to interpolation when the neighbor position does not coincide with a pixel site. The basic features used in LBP are simply the directional differences

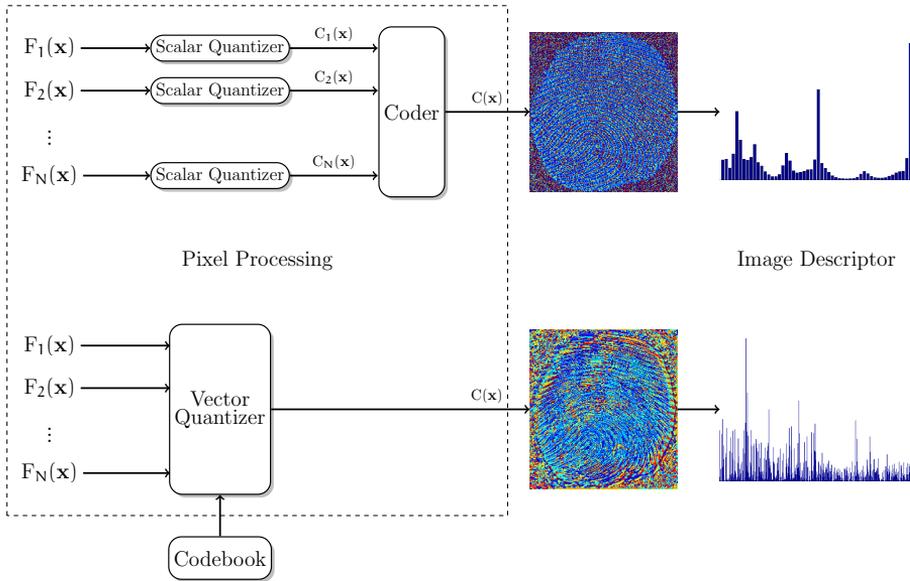
$$F_i(\mathbf{x}) = I(\mathbf{x}) - I(\eta_i(\mathbf{x})) \quad (2.1)$$

These features are quantized independently, with a fixed 2-level symmetric quantizer, obtaining the indexes

$$C_i(\mathbf{x}) = \begin{cases} 0 & F_i(\mathbf{x}) \leq 0 \\ 1 & \text{otherwise} \end{cases} \quad (2.2)$$

that is, a string of bits, represented synthetically by the integer

$$C(\mathbf{x}) = \text{LBP}(\mathbf{x}) = \sum_{i=0}^{P-1} C_i(\mathbf{x})2^i \quad (2.3)$$



**Figure 2.2:** This figure shows the two different paradigms: Independent Quantization of Features (up) and Joint Quantization of Features (down). In the first case each component of the feature vector is scalar quantized, typically with a two-level quantizer. Instead, in the second case a vector quantization is used. The resulting image descriptor is an histogram which represents number of occurrences of the code  $\mathbf{C}(\mathbf{x})$ , which can assume values in a broader range when the features are vector quantized.

In the basic version of LBP, these quantities are not subject to further processing, leading to a feature vector of length  $2^P$ ,  $\mathbf{h} = \text{hist}(C)$ , where

$$h(i) = \sum_{\mathbf{x}} \delta(C(\mathbf{x}) - i) \quad (2.4)$$

with  $\delta(i) = 1$  for  $i = 0$  and 0 otherwise. Typically,  $R \leq 3$ , corresponding to very small neighborhoods, and  $P$  goes from 8 to 24.

Despite its simplicity, LBP has proven very powerful in a number of applications, to begin with texture classification considered in the seminal work [102]. In the very same work, some variants of LBP have been proposed as well. In the *rotation-invariant* version, all strings presenting the same pattern after a circular shift are collected together, obtaining

not only some invariance to local rotations, but also shorter vectors. Note that, in general, reducing the length of the feature vector (without sacrificing descriptive power) can improve significantly both the training and the classification phases. In this same direction goes the *uniform* version, which encodes individually only strings with no more than two 0-1 or 1-0 transitions, corresponding to relatively regular local patterns, pooling all the others (rarely occurring) in a single codeword. By so doing, the feature length is drastically reduced and becomes manageable also for  $P = 24$ . As a consequence, a further *multi-resolution* variant can be considered, where multiple uniform LBP features, obtained for different values of  $R$  and  $P$ , are concatenated, enriching the description and covering a larger image patch.

### 2.4.2 CoA-LBP and Ric-LBP

Multi-resolution LBP testifies of a first effort towards the use of larger neighborhoods, so as to exploit richer and longer-range dependencies. The problem, of course, lies in the length of the feature vector which grows rapidly as more features are included.

A simple solution is found with the CoA-LBP (Co-occurrence of Adjacent LBPs [99]). The name is self-explaining: after extracting  $C$ ,  $K$  bi-dimensional histograms  $\mathbf{h}_k$  are computed, with

$$h_k(i, j) = \sum_{\mathbf{x}} \delta(C(\mathbf{x}) - i, C(\mathbf{x} + \Delta_k) - j) \quad (2.5)$$

These histograms are then vectorized, and concatenated to form the final feature  $\mathbf{h}$ . The  $k$ -th bi-dimensional histogram accounts for co-occurrences of couples of LBP's separated spatially by the vector  $\Delta_k$ , thereby exploiting dependencies at a somewhat extended range. In [99], four such histograms are computed, with LBP's taken at distance  $\|\Delta_k\| = 3$  along directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . To reduce the feature length,  $P = 4$  is considered for the basic LBP, obtaining eventually features of length 1024 (without considering symmetries).

Ric-LBP [100] is a further evolution, where couples of LBP's corresponding to rotated configurations are pooled together, obtaining some feature length reduction.

### 2.4.3 LPQ

Local Phase Quantization (LPQ) was proposed originally for texture classification in the presence of blurring [103]. Following the description in [103], the patch surrounding the target pixel  $\mathbf{x}$  is analyzed in the frequency domain by means of a short-time Fourier transform (STFT),

$$\hat{I}_{\mathbf{x}}(\mathbf{u}) = \sum_{\mathbf{y}} I(\mathbf{y})w(\mathbf{y} - \mathbf{x})e^{-j2\pi\mathbf{u}\cdot\mathbf{y}} \quad (2.6)$$

where  $\mathbf{x}, \mathbf{y}$  are bi-dimensional spatial coordinates,  $\mathbf{u}$  indicates bi-dimensional spatial frequencies,  $w(\cdot)$  is a compact window that enforces locality of the transform, and  $\hat{I}_{\mathbf{x}}(\cdot)$  is the output STFT around  $\mathbf{x}$ . Afterwards, only four frequencies are considered,  $\mathbf{u}_0 = (\alpha, 0)$ ,  $\mathbf{u}_1 = (\alpha, \alpha)$ ,  $\mathbf{u}_2 = (0, \alpha)$ , and  $\mathbf{u}_3 = (-\alpha, \alpha)$ , with  $\alpha \ll 1$ , corresponding to the directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . The basic features of LPQ are the phases of the selected coefficients, that is

$$F_i(\mathbf{x}) = \angle \hat{I}_{\mathbf{x}}(\mathbf{u}_i)$$

For each of these frequencies, the phase  $F_i(\mathbf{x})$  is computed and quantized with a 2-bit uniform quantizer in  $[-\pi, \pi]$ . Finally we obtain an 8-bit feature  $[q_1, \dots, q_8]$  that are represented as integer values in the range  $[0 - 255]$ :

$$C(\mathbf{x}) = \text{LPQ}(\mathbf{x}) = \sum_{i=1}^8 q_i 2^{i-1}$$

In all practical applications,  $w(\cdot)$  is a square window (with or without Gaussian smoothing) with odd side  $N$ , and  $a = 1/N$ . Therefore the LPQ feature encodes information on the phase of the lowest-frequency plane-waves oriented at 0, 45, 90, and 135 degrees w.r.t. the horizontal axis. With proper choice of the parameters, the features  $F_i(\mathbf{x})$  can be viewed alternatively as the output phases of four Gabor filters oriented along directions  $45^\circ$  apart. This is a sensible choice (as amply supported by experimental results) as is well-known that, in the Fourier transform, the phase spectrum is more informative than the amplitude spectrum. With the given dimensionality constraints, these quantized phases provide therefore a very informative local description of the image.

This interpretation suggests that LPQ can be effective in describing images characterized by a local wave-like behavior, like fingerprint and,

to some extent, iris images. A rotation-invariant version of LPQ has also been proposed in [103], where the patch is preliminary rotated along a characteristic direction  $\beta(x)$  computed in advance.

The LPQ descriptor can be further refined by considering its rotation-invariant version,  $\text{LPQ}^{ri}$ , described in [103]. The idea is straightforward: rather than using the original patch to compute the STFT, a rotated version of it is used, which is aligned along a characteristic direction  $\beta(x)$  computed from the patch itself. By doing so, the  $\text{LPQ}^{ri}$  provides a description of the patch which does not depend on its random local orientation but only on its intrinsic properties. This idea is particularly compelling for the case of fingerprint images given the plane-wave local appearance of patches.

In more detail, for each pixel  $x$ ,  $F_x(u)$  is computed on  $M$  points placed uniformly on a circle of radius  $r$ , namely, at the frequencies

$$v_i = (r \cos \phi_i, r \sin \phi_i), \quad \text{with } \phi_i = 2\pi i/M, \quad i = 0, \dots, M-1 \quad (2.7)$$

obtaining the vector  $[F_x(v_0), \dots, F_x(v_{M-1})]$ . Then, the sign of the imaginary part of this vector is retained, say  $[c_0, \dots, c_{M-1}]$ , and the characteristic orientation  $\beta(x)$  is eventually computed as the phase of the complex quantity

$$b(x) = \sum_{i=0}^{M-1} c_i e^{j\phi_i} \quad (2.8)$$

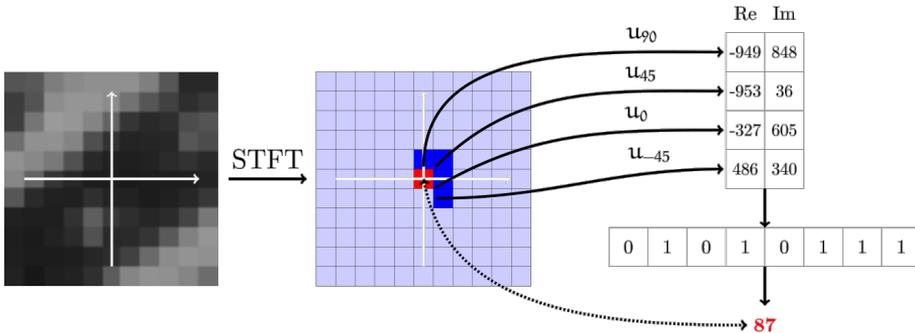
With this definition, due the symmetry properties of the Fourier transform, the characteristic orientation of a patch rotated by  $\theta$  can be approximated as  $\beta(x) + \theta$  [103], with a precision depending only on  $M$ . We obtained a satisfactory performance with  $M=36$ . Moreover we set  $N=9$ , as it guarantees for all datasets to include two nearby ridges in the analysis window  $w(\cdot)$ , and  $a = 1/N$ .

#### 2.4.4 WLD

The Weber LD is built starting from two dense fields of features, orientation and differential excitation. The orientation is simply the angle of the local gradient

$$F_1(\mathbf{x}) = \theta(\mathbf{x}) = \text{angle}(\nabla I(\mathbf{x})) \quad (2.9)$$

where the sign is chosen so as to obtain the correct angle formed by the gradient with the reference axis, while the differential excitation writes



**Figure 2.3:** Patch processing of the LPQ descriptor.

as:

$$F_2(\mathbf{x}) = \xi(\mathbf{x}) = \frac{\bar{I}_{3 \times 3}(\mathbf{x}) - I(\mathbf{x})}{I(\mathbf{x})} \quad (2.10)$$

where  $\bar{I}_{3 \times 3}(\mathbf{x})$  indicates the sample average of  $I$  over the  $3 \times 3$ -pixel square centered on  $\mathbf{x}$ . The numerator is the difference between the intensity of the target pixel and the average intensity of its neighbors, therefore the feature is zero in flat areas of the image and grows larger in the presence of discontinuities. However, the very same difference can have a quite different perceptual importance depending on where it occurs in the image: it can be barely distinguishable in a high intensity region, and quite significant instead in low-to-medium intensity regions. This observation is captured by the well-known Weber's law that states that the just-noticeable difference between two stimuli is proportional to the magnitude of the stimuli. In accordance with this principle, this difference is normalized to the pixel intensity itself.

The angle is quantized uniformly with  $N_1$  output levels in the range  $[-\pi, \pi]$ , while the excitation is quantized non-uniformly, to take into account the high-dynamics and the unbounded range induced by the ratio, using a uniform  $N_2$ -level quantizer in  $[-\pi/2, \pi/2]$  after an arctan nonlinearity.

The outputs are then concatenated in a single integer

$$C(\mathbf{x}) = \text{WLD}(\mathbf{x}) = C_1(\mathbf{x})N_2 + C_2(\mathbf{x}) \quad (2.11)$$

with values in  $[0, N_1N_2 - 1]$ . Typical values are  $N_1 = 8$  and  $N_2 = 120$ , for a feature vector with length of 960 features.

---

The rationale of WLD is to take into account jointly the local orientation of the gradient and the local activity of the neighborhood, invoking the Weber law on the psychophysics of vision which states that the visibility of features depends indeed on the local contrast. It may be argued that psychophysics is not relevant in machine learning, unless some model of the human visual system is incorporated in the classifier, but classifying local neighborhoods based on their activity seems nonetheless a good idea, which enhances the descriptive power of the gradient, as confirmed by experiments [14, 43].

### 2.4.5 BSIF

The path followed so far testifies of an intense effort towards the selection of the local features that more compactly and effectively catch the relevant micro-textural characteristics of the image. In [60] there is an interesting deviation w.r.t. this path, since the features are not manually designed anymore, but obtained automatically to satisfy some desired requirements.

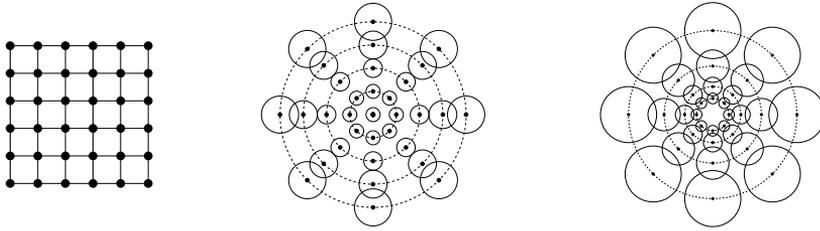
The proposed descriptor, called BSIF (Binarized Statistical Image Features) keeps the same structure of LBP, with  $N$  local features  $F_i(\mathbf{x})$ , obtained by filtering the input patch, binary quantized and finally combined into a single scalar  $C(\mathbf{x})$ . However, the filters themselves are not known in advance, but designed so as to maximize the statistical independence of the outputs  $F_i(\mathbf{x})$ .

Of course, independent features may be expected to be more informative than dependent ones. Moreover, they can be quantized independently with little performance loss w.r.t. the case of joint quantization.

## 2.5 Joint Quantization of Features

### 2.5.1 SIFT, DAISY

The Scale-invariant Feature Transform (SIFT) is one of the most famous and successful descriptors, used for a variety of computer vision tasks. As the name suggests, one of its main properties is the scale invariance, namely, the ability to provide a meaningful description of a region that, up to a certain extent, remains the same as the scale of observation changes. A detailed description of SIFT is out of the scope of this work, we will instead try to convey some very general concepts, fitting in the



**Figure 2.4:** Examples of different sampling grids: rectangular grid with bilinear weights (SIFT), polar grid with gaussian weights (DAISY) and log-polar grid with gaussian weights (SID).

scheme outlined in Fig.2.1, sacrificing precision for the sake of synthesis. The reader is referred to [78] for more details.

First of all, it is important to note that SIFT has been conceived originally as a *keypoint-based* descriptor. By this, we mean that it tries to characterize each pixel of interest (the keypoints) with a high-dimensionality feature vector, rich enough to enable the accurate and reliable matching of keypoints. Therefore, for each keypoint  $\mathbf{x}$ , SIFT characterizes a whole subimage centered on  $\mathbf{x}$ , not just a small neighborhood, by means of a suitable histogram-based feature vector. To make an example with simpler LDs, for each point  $\mathbf{x}$  one might extract a subimage centered on  $\mathbf{x}$ , and compute the corresponding 256-component LBP feature vector. This would provide a set of *local* features  $F_1(\mathbf{x}), \dots, F_{256}(\mathbf{x})$  associated with the point.

In SIFT, this local feature vector is the estimated histogram of the gradient direction, quantized uniformly to 8 bins in the  $[0, 2\pi]$  range. This histogram is computed on a very small patch of  $4 \times 4$  pixels, but various expedients are taken to smooth it (using a suitable weighting scheme) and improve its significance [78]. Moreover, to provide also a *spatial* characterization of dominant directions around the keypoint, 16 such histograms are actually computed on neighboring patches covering a total area of  $16 \times 16$  pixels. All such histograms are concatenated to form a set of 128 features associated with  $\mathbf{x}$ .

As said before, armed with such a rich description, one can reliably match keypoints within a single image (for example, to detect copy-move forgeries) or across different images (for example to register mosaics of photos). However, if the goal is to characterize the whole image using *dense* SIFT descriptors, computed for all pixels or on a regular sampling

grid, these local features must be necessarily summarized, before an histogram can be computed. The typical solution is to resort to the Bag-of-Words / Bag-of-Features paradigm, that is, more technically, to the vector quantization of the local features, as shown in the bottom part of Fig.2.2. Given a training set of SIFT features, large enough to well represent the source, one can easily design a vector quantizer that operates jointly on the vector of local features, associating with each vector the index of the (128-dimensional) cell it belongs to. These indexes can be regarded as a summary of the rich feature-vector, but they are scalars, allowing for the computation of the usual histogram-based feature vector  $\mathbf{h}$  representing the whole image.

Given the huge success of SIFT a number of descriptors have been proposed to improve upon it under some respects. One of the most interesting is DAISY [130], characterized by an efficient computational scheme, of special interest for use in dense modality, and by a different organization of the spatial neighborhoods, placed on concentric rings (from which the name, see also Fig.2.4) rather than on a rectangular grid, obtaining also a better robustness to rotations.

### 2.5.2 SID

Just like scale invariance, also rotation invariance can be an important feature of a descriptor, useful for a number of applications. In copy-move forgeries, for example, objects are often rotated and resized before copying, but this should not prevent the descriptor from discovering similarities. Invariance properties, however, are important also in the context of image description, as they allow to obtain more compact and expressive feature vectors, to the benefit of classification accuracy. These properties are guaranteed to a good extent by the Shift-Invariant Descriptor (SID) proposed in [64] and refined in [63]. SID has already proven extremely powerful in such diverse applications as fingerprint liveness detection [46] and cell classification [51]. We provide here a brief description of SID, but the reader is referred to [63] for a more thorough treatment and analysis.

The basic idea is straightforward. Consider two images obtained from one another through a rigid shift  $I_2(\mathbf{x}) = I_1(\mathbf{x} + \Delta\mathbf{x})$  (let us neglect boundary issues). Their Fourier transforms  $\hat{I}_1$  and  $\hat{I}_2$  will be identical but for a phase term and, therefore, the modulus of the Fourier transform is a shift-invariant representation of an image. However, we are

not interested in shift-invariance, but rather in invariance to rotations and rescaling. All we have to do, therefore, is to resample the image on a log-polar grid around its center (see Fig.2.4), extracting samples at logarithmic increasing distance from the center along radii with uniform angular separation. The modulus of the Fourier transform of this resampled image will be rotation- and scale-invariant. It is worth underlining that, contrary to what happens with SIFT, where the scale must be estimated in advance, scale invariance in SID is intrinsic and does not require any further processing. Of course, there are a number of important issues to deal with in order to obtain a meaningful and expressive local descriptor  $C(\mathbf{x})$ . First of all, the Fourier transform is taken not on the original image but on directional derivatives. Then, a suitable adaptive smoothing scheme is necessary to avoid aliasing. Therefore, the SID descriptor is actually computed through the following steps:

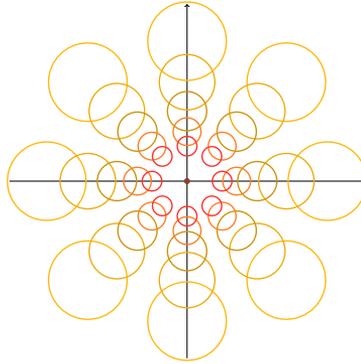
1. log-polar transformation;
2. multiscale smoothing;
3. computation of directional derivatives;
4. 2d discrete Fourier transform.

In the following we will analyze each of these steps.

### Log-polar transformation

The log-polar transformation is a non linear and non uniform sampling of the spatial domain. This mapping is extensively used in image registration and turns out to be similar to the sampling pattern of the human visual front-end [115]. Note that this spatial grid has been already used in [89] to increase robustness and distinctiveness of the SIFT descriptor and has been shown to give good performance in [141].

The log-polar mapping has the interesting property that rotations and rescalings relative to the center correspond to mere translations [75], a feature that will be exploited to build the rotation and scale invariant descriptor. Samples will be taken around the target point  $\mathbf{x} = (x_1, x_2)$  of the image  $f$  along  $K$  directions, at angles  $\theta_1, \theta_2, \dots, \theta_K$  with  $\theta_k = 2\pi k/K$  and at  $N$  distances from the center  $r_1, r_2, \dots, r_N$  with  $r_i/r_{i-1} = \rho > 1$ . The measures on the sampling grid can be rearranged



**Figure 2.5:** Log-polar grid and spatially varying filtering kernels.

using a  $N \times K$  matrix:

$$h_{\mathbf{x}}(n, k) = f(x_1 + r_n \cos \theta_k, x_2 + r_n \sin \theta_k)$$

### Multiscale smoothing

In order to avoid aliasing caused by the irregular sampling, it is necessary to remove high frequency components by band-pass filtering the image before extracting features from it. This is done by using gaussian filters with spatially varying standard deviations [64, 130], proportional to the distance from the center of the log-polar sampling grid,  $\sigma_n = \alpha r_n$ , as shown in figure 2.5. Note that in [74] it is shown that a scale-space representation can be obtained through convolution with gaussian kernels of different size. Moreover this guarantees that scaling the image only scales the features and does not distort them in any other way. Hence:

$$h_{\mathbf{x}}(n, k) = f_{\sigma_n}(x_1 + r_n \cos \theta_k, x_2 + r_n \sin \theta_k)$$

where  $f_{\sigma_n}$  is the sampled field  $f$  smoothed with a gaussian kernel with standard deviation  $\sigma_n$ .

### Computation of directional derivatives

To compute the descriptor, filtering and sampling operate on directional gradients, rather than on the original values, as already proposed in [130] to construct the DAISY descriptor. This is a common practice when building descriptors to provide invariance to intensity changes.

Interestingly, this approach is in qualitative agreement with the results of biological evolution. Neurophysiological studies by have shown that there are receptive fields in the mammalian retina and visual cortex, whose measured response profiles can be well modelled by Gaussian derivatives up to order four [75].

In particular, in [63] directional derivatives are computed with polarization, i.e. separating the negative and positive parts. In order to ensure that image rotation amounts to feature translation directional derivatives need to be aligned with the ray directions. Hence the directional derivatives on ray  $k$  and scale  $n$  are computed for  $D$  orientations:  $\theta_{k,d} = 2\pi k/K + 2\pi d/D$ . In addition, since the amplitude of spatial derivatives decreases with scale, due to the scale-space smoothing, the output is rescaled by the kernel's standard deviation [75]:

$$h_{\mathbf{x}}^{d,\pm}(n,k) = \sigma_n f_{\sigma_n}^{\theta_{d,k,\pm}}(x_1 + r_n \cos \theta_k, x_2 + r_n \sin \theta_k)$$

where  $(\cdot)^\pm$  are the operators such that  $(a)^+ = \max(a, 0)$  and  $(a)^- = \min(a, 0)$ , This way, we have now a total of  $2D K \times N$  matrices for each point.

## 2d discrete Fourier transform

A common problem that emerges in the computation of local descriptors is the variability of the signal scale. The standard approach to cope with this is scale selection, which consists in estimating a characteristic scale and rotation around the few image or shape points where scale estimation can be performed reliably. The log-polar sampling strategy allows to construct a descriptor which is invariant w.r.t. rotation and scaling. This is already done in the context of image registration, but on a global rather than local scale. Here, due to the time-shifting property of the Fourier transform, and to the log-polar sampling, by considering only the amplitude of the Fourier transform we obtain a rotation and scale invariant descriptor [64, 63]. Actually, thanks to the symmetry properties of the Fourier transform we can keep only  $N \times (K/2 - 1)$  points for each matrix. Concatenating all components in a single vector will form eventually a local descriptor of dimension  $DN(K - 2)$ .

# Chapter 3

## Main contributions

This chapter describes the main research contributions given in this thesis work to the liveness detection task. In Sec.3.1 we explore the use of a Wavelet-Markov descriptor for fingerprint liveness detection, while in Sec.3.2 show that the simple concatenation of multiple independent descriptors can significantly improve upon each of them. This issue is studied in more depth in Sec.3.3, where a new local descriptor is proposed based on the joint analysis of suitable elementary features. Instead in Sec. 3.4 we propose an approach to deal with the iris liveness detection on mobile devices. Finally in Sec.3.5 we present a robust technique based on the SID descriptor and the Bag of Words paradigm, to deal with different biometric authentication systems spoofing.

### 3.1 Wavelet-Markov features

Inspired by the work of [55], we compute a local descriptor based on co-occurrences in the wavelet transform domain. We carry out a three-level wavelet transform of the original image  $A^0$ , obtaining a total of 13 subbands, that is, the image  $A^0$  itself and, for each level  $l=1,2,3$ , the approximation  $A^l$ , and the detail subbands,  $H^l$ ,  $V^l$  and  $D^l$ , along the horizontal, vertical and diagonal directions, respectively, as shown in Fig. 3.1.

Then, being  $X$  any one of the above mentioned subbands, and  $(i, j)$  generic coordinates, we compute a feature vector  $F$  for classification by performing the following processing steps:

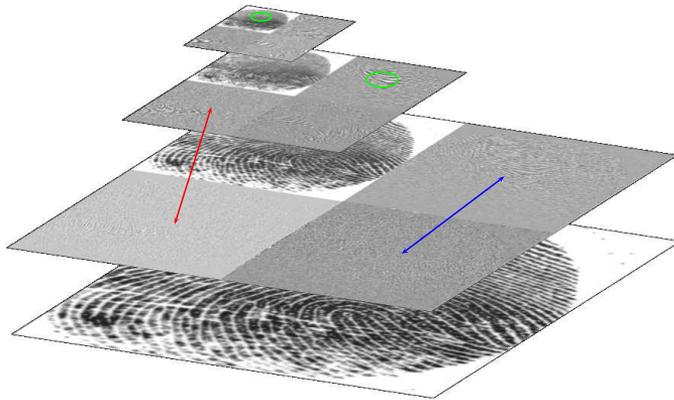
1.  $X_{ij} \leftarrow \text{round}(\text{abs}(X_{ij}));$

- 
2.  $R_{ij} = X_{ij} - \widehat{X}_{ij}$ ;
  3.  $R_{ij} \leftarrow \text{trunc}_T(R_{ij})$ ;
  4.  $F_{rs} = \sum_X \mathbf{1}(R_{ij} = r, R_{i+di, j+dj} = s)$ .

As step 4 shows, our local descriptor is based on second-order probabilities (by which the *Markov* label), estimated by means of co-occurrence matrices. However, rather than the actual value  $X_{ij}$ , we use its residual (step 2) w.r.t. a suitable prediction  $\widehat{X}_{ij}$ . Residuals account for local deviations from the expected behavior, a much more valuable information than the original values to detect anomalies. Likewise, we use  $\text{abs}(X_{ij})$  rather than  $X_{ij}$  because wavelet coefficients are only weakly correlated, while their moduli exhibit stronger dependencies. The rounding and truncation operations serve to limit the number of possible pairs  $(r, s)$ , and hence the feature vector length, equal now to  $(2T + 1)^2$ .

The feature vector depends on three main quantities, the subband  $X$ , the predictor  $P$ , and the displacement  $(di, dj)$ . We consider co-occurrences only along adjacent pixels in the horizontal ( $di = 1, dj = 0$ ) and vertical ( $di = 0, dj = 1$ ) directions, since pixels farther apart or on the diagonal show weaker dependencies. Then, we consider three classes of predictors, symbolically shown in Fig. 3.1, that account for dependencies across position (green), scale (red), and orientation (blue). As an example, the orientation predictor depicted in blue in the figure, predicts  $D_{ij}^1$  from  $H_{ij}^1$ , that is, from the same-position pixel taken from a same-level detail subband having different orientation. By considering only the simplest combinations of subband, predictor, and displacement, we end up with 82 feature vectors, each one with  $(2T + 1)^2$  Markov features. Preliminary experiments proved that  $T=4$  guarantees the best trade-off between performance and complexity, leading to a grandtotal of over 6000 features.

In this work we use a Support Vector Machine, with a radial basis functions kernel, as classifier. Given the limited size of the available training sets we need to reduce the feature vector dimensionality before performing classification. To this end, we resort to PCA feature reduction and choose the number of features that guarantees the smallest overall classification error by using a cross-validation procedure. In all cases, less than 100 PCA components were selected.

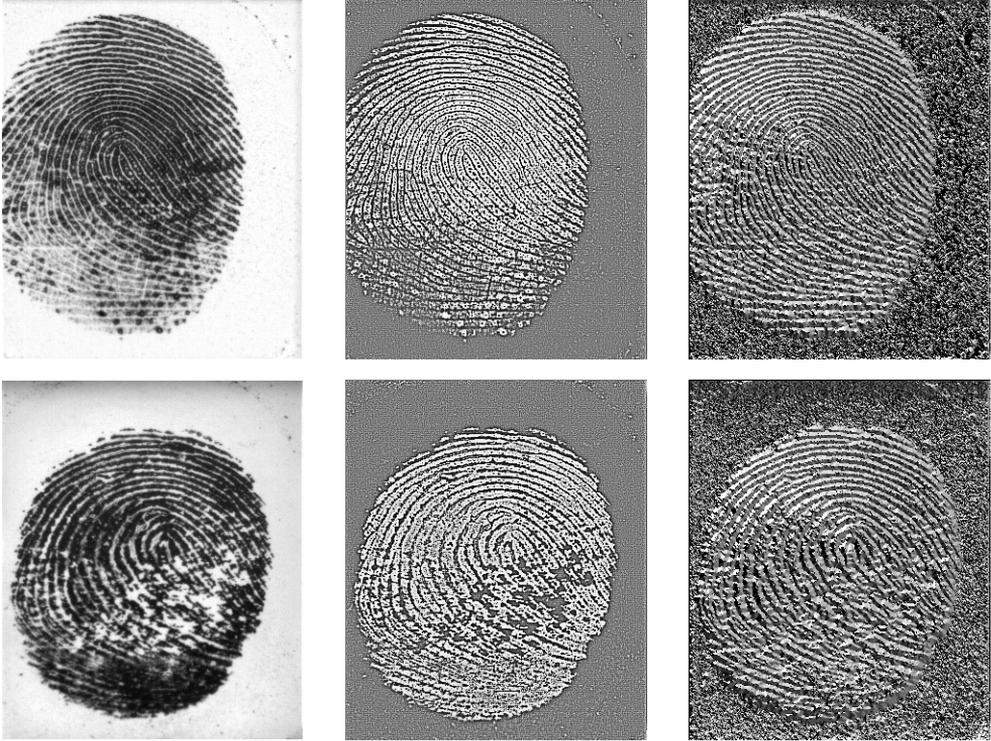


**Figure 3.1:** The pyramid of approximations and details subbands used in the algorithm. The colored arrows and circles indicate the various types of dependencies exploited for prediction: across position (blue), scale (red), and orientation (green).

## 3.2 LPQ and WLD concatenation

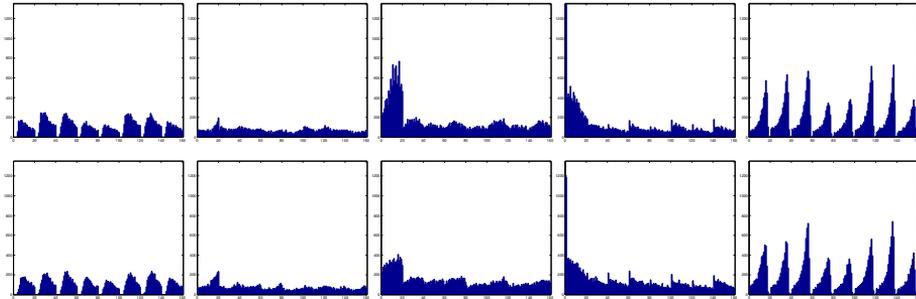
To perform the fingerprint liveness detection task, we simply extract from each training fingerprint image its WLD, and then use all such descriptors to train a linear kernel SVM classifier. In Fig.3.2 the two fields of the WLD are depicted for a live and a fake samples, while in Fig.3.3 are shown the respective histograms, spotting a noticeable difference between the two classes samples in some bins values. This approach is versatile, indeed we can expand the features used simply concatenating the corresponding histograms. The performance is then assessed on some publicly available datasets provided in the context of the Fingerprint Liveness Detection Competitions LivDet 2009 [85] and LivDet 2011 [144]. Each dataset is divided into a training set and a test set in which the evaluation results are obtained. In order to generate unbiased results, no overlap is present between training and test sets (i.e., samples corresponding to each user are included in only one of the sets).

First of all, techniques based on local descriptors outperform most of the times those based on global features. Among the local descriptors, WLD appears to be much better than LBP and somewhat better than LPQ on the LiveDet09 datasets, see also Tab.4.4 in Section 4, but not so



**Figure 3.2:** Live (top) and fake (bottom) fingerprints with the corresponding differential excitation and gradient orientation fields.

on the (more challenging) LiveDet11 datasets, where it performs mostly on par with the others except for the ItalData set, when it is much worse. In *all* cases, the best performance is obtained by using some combination of local descriptors. On the LiveDet09 datasets it is always the concatenation WLD+LPQ while on the LiveDet11 datasets two combinations are almost equally good, LBP+LPQ and, again, WLD+LPQ. However, combining multiple local descriptors is not always a good idea, and in fact the LBP+LPQ combination does not always improve the performance of the best of the two, and only slightly when this happens. On the contrary, the WLD+LPQ combination works always much better than both WLD and LPQ, indicating that these descriptors must somehow complement one another, thus greatly improving their discriminating ability.



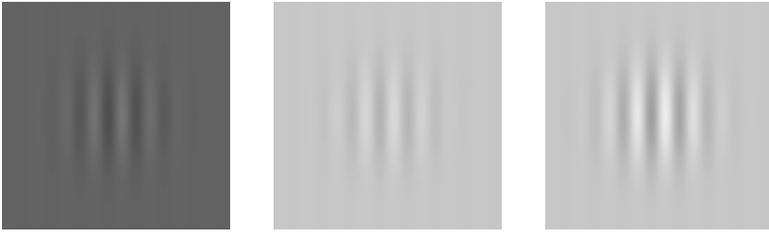
**Figure 3.3:** Histograms corresponding to the live (top) and fake (bottom) fingerprints of Fig.3.2

### 3.3 LCPD

In the previous subsection we analyzed the effect of concatenating several descriptors summing up their dimensionality, observing a significant performance gain [43]. Here, following also [14, 76], we set to combine individual features so as to exploit their dependencies. Eventually, our proposed descriptor is based on a spatial-domain component, inspired to the homologous component of WLD, and a phase component, which is the rotation-invariant version of LPQ, ending up with the Local Contrast-Phase Descriptor (LCPD) illustrated below. In the next three subsections we provide details on the two components and on their combination.

#### 3.3.1 Spatial-domain component

Local descriptors in the spatial domain work typically on the residual obtained through some high-pass filtering of the image. In fact, information on the local image behavior is contained mostly in the *variations* of the amplitude, while low-pass trends are of little or no interest and can be neglected. However, variations with the very same amplitude may have a very different importance depending on where they occur in the image: they can be regarded as little more than structured noise in a high intensity region, and be instead quite significant in low-to-medium intensity regions. In the context of vision, this observation is captured by the well-known Weber’s law which states that the just-noticeable difference between two stimuli is proportional to the magnitude of the stimuli themselves. This is clarified by the example of Fig.3.4: the local



**Figure 3.4:** Importance of contrast. The images show smoothed sinusoidal patterns with peak amplitude  $A$ , over a flat background with amplitude  $B$ . (a)  $A=25$ ,  $B=100$ ; (b)  $A=25$ ,  $B=200$ ; (c)  $A=50$ ,  $B=200$ .

sinusoidal pattern is clearly visible in image (a) and barely distinguishable in image (b) although variations have the very same amplitude in both images, with a peak of 25 gray-levels. In (a), however, the contrast is twice as large as in (b) -  $25/100$  as opposed to  $25/200$  - and equal to the contrast in (c),  $50/200$ . Although Weber's law has to do with psychophysics, it is generally true that the *contrast*, more than the amplitude variation itself, conveys valuable information on the local image behavior. Of course, this is very relevant for the fingerprint case, because the amplitude properties of the acquired images can change significantly with the hardware and with the specific acquisition.

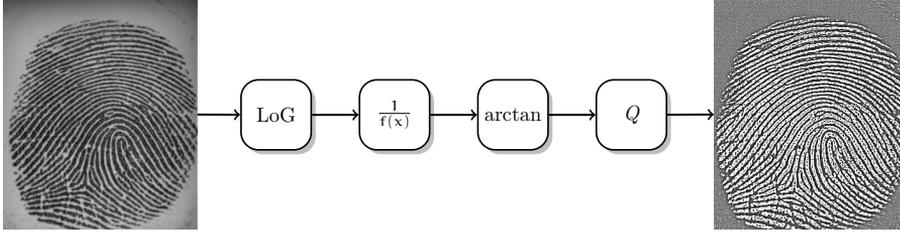
This principle is taken into account in the Weber Local Descriptor (WLD) proposed in [14]. In WLD two complementary pieces of information are combined, local contrast  $\xi(x)$  and local orientation  $\theta(x)$ . Since our phase component, described in next subsection, conveys the same type of information as  $\theta(x)$ , and in a more effective way, we will neglect this latter in the following.

The local contrast, called differential excitation in [14] writes as:

$$\xi(x) = \arctan \left[ \frac{1}{f(x)} \sum_{i=0}^7 (f_i - f(x)) \right] \quad (3.1)$$

where  $f(x)$  is the value of the image in the target pixel  $x$ , while the  $f_i$ 's are the values of the neighbors in a  $3 \times 3$  patch surrounding  $x$ , as shown below

$$\begin{bmatrix} f_0 & f_1 & f_2 \\ f_7 & f(x) & f_3 \\ f_6 & f_5 & f_4 \end{bmatrix}$$



**Figure 3.5:** Contrast-based field block-diagram. The image is LoG filtered and normalized by the value of the central pixel. Finally the non-linearity is applied and the output is finally quantized.

The summation implements the high-pass filtering, in particular by means of a  $3 \times 3$  Laplacian filter. Then the result is normalized to the intensity of the target pixel to compute the contrast. Due to this normalization, the resulting feature has unbounded range. Therefore to obtain an integer-valued feature with a finite range, a non-uniform quantization is needed. In alternative, uniform quantization can still be used after a suitable non-linear transformation, which is why the arctan operator is considered. A mid-thread quantizer is used, with an odd number of levels  $S$ , in order to count 0 among the possible output values.

Among the many candidate amplitude components for our descriptor,  $\xi(x)$  provided consistently the best performance. The only variation that provided a significant improvement was the use of a smoothing filter before the Laplacian, so in the end we used the well-known LoG filter (Laplacian of Gaussian) as also proposed in [76] with a  $5 \times 5$  window and  $\sigma = 0.5$ . The optimal value of  $S$  changes with the dataset, and is set through cross-validation on the training set.

### 3.3.2 Transform-domain component

A fingerprint image can be thought of as a system of oriented textures, where the gray levels along ridges and valleys can be modeled locally as a sinusoidal-shaped wave along a direction normal to the ridge orientation, while the ridge frequency varies slowly throughout the image. The frequency domain is therefore perfectly suited to extract a compact and informative description of a signal that exhibits such a strong quasi-periodic behavior.

Here, we will resort to the rotation invariant Local Phase Quantiza-

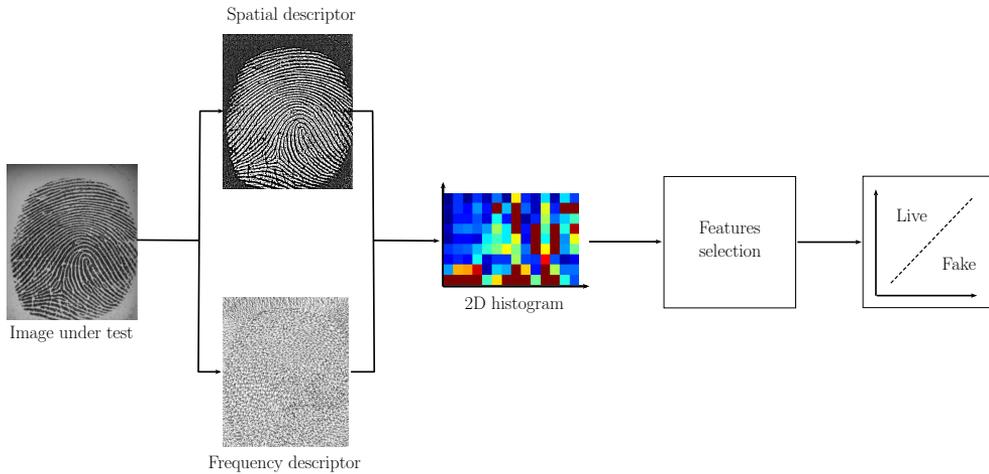


**Figure 3.6:** Live (top) and fake (bottom) fingerprints with the corresponding contrast and quantized phase fields.

tion ( $\text{LPQ}^{r_i}$ ) described in section 2.4.3.

### 3.3.3 Combination

In Fig.3.7 we show the main processing steps of the proposed liveness detection method. The image under test is analyzed, in parallel, in the spatial and in the transform domain. For each pixel  $x$  two features are computed, the local contrast  $\xi(x)$  and the  $\text{LPQ}^{r_i}(x)$  feature, called here  $\phi(x)$ . These couples, collected over the whole image, populate a two-dimensional histogram  $\text{LCPD}(\xi(x), \phi(x))$ , of dimension  $S \times 256$ . As a consequence, the intensity of each cell corresponds to the frequency of a certain (quantized) contrast-phase couple. All the rows of the histogram are finally concatenated together to form a 1D-histogram. Fig.3.8 shows some of the rows of the bi-dimensional histogram, before concatenation. Since the overall dimensionality is typically too large to obtain good results, a feature selection block follows. The usefulness of the feature



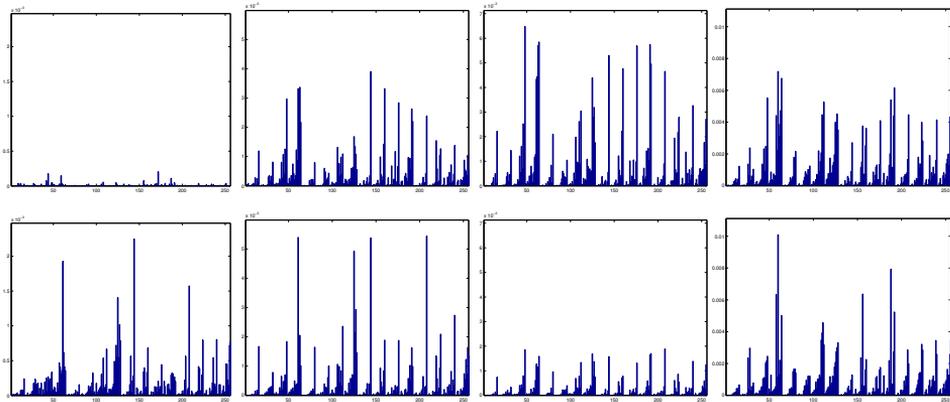
**Figure 3.7:** LCPD block-diagram.

selection will be experimentally evaluated in the next Section. Eventually, the output feature vector is fed into a SVM classifier with a linear kernel which makes the decision.

In Fig.3.6 we show  $\xi(x)$  and  $\phi(x)$  for a live (up) and a fake (bottom) fingerprint. Despite the low quality of these images, corrupted also by a significant noise, the fields of features appear to capture quite accurately the relevant features of these images.

### 3.4 LBP from the residue image

Nowadays mobile devices are equipped with biometric authentication systems which should protect sensible data, like for example contacts and private photo or, more important, personal credentials for bank account. In particular a reliable and low-cost solution could be to exploit the front camera sensor for iris recognition, avoiding additional hardware needed for fingerprint scanning. However a robust liveness detection system is necessary to protect such data from any person who can gain access to someone else's mobile device.



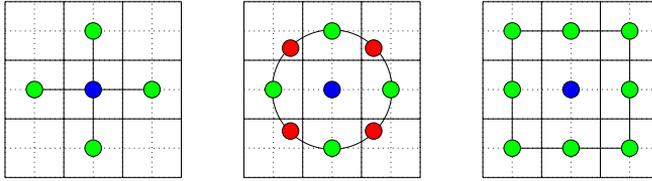
**Figure 3.8:** Histograms corresponding to a live (top) and a fake (bottom) fingerprints.

### 3.4.1 Local Binary Patterns for mobile devices

In order to build a very-low complexity iris liveness detector, suitable for implementation on a mobile device, we rely on the LBP descriptor. LBP encodes the information on the variations of intensity occurring between the image pixels and their neighbors. In more details, for each target pixel,  $x$ ,  $P$  neighbors are sampled uniformly on a circle of radius  $R$  centered on  $x$ . These neighbors are then compared with  $x$ , taking only the sign of the difference, and forming thus a vector of  $P$  binary values, which is eventually converted to an integer. In formulas,

$$\text{LBP}_{P,R}(x) = \sum_{i=0}^{P-1} \text{u}(x_i - x) 2^i$$

where  $x_i$  is the  $i$ -th neighbor of  $x$ , and  $\text{u}(t) = 1$  when  $t \geq 0$  and 0 otherwise. The resulting feature provides information on the level of activity in the target-pixel area. By scanning the whole image, a field of integers is obtained, which is then summarized by a histogram, accounting for the frequency of occurrence of each feature. The parameter  $P$  controls the quantization of the angular space, while the radius  $R$  determines the spatial resolution of the operator. With  $R = 1$ , considering  $P = 4$  and  $P = 8$ , we obtain the first two configurations shown in Fig.3.9. Note that, since the neighbors are located on a circle surrounding the target, they do not always correspond to actual pixel values (see the diagonal



**Figure 3.9:** Some neighborhood systems used by LBP. The first two configurations are circular symmetric, with  $R=1$  and  $P=4$  (left) and with  $R=1$  and  $P=8$  (center). In the second case some neighbors (red) do not correspond to actual pixels and must be computed by interpolation. The last configuration does not exhibit circular symmetry but does not require interpolation.

neighbors in the central configuration of Fig.3.9), in which case they must be computed by interpolation.

When  $P$  is large (fine angular resolution) the resulting feature vector has a considerable length. Besides the increased complexity, this fact implies that many patterns occur very rarely in the image, and their frequency of occurrence is an unreliable feature. Therefore, to improve performance, some more compact versions of LBP have been proposed. The circular symmetry allows for the definition of a rotation invariant version of the descriptor [102], recalled by the superscript  $ri$

$$\text{LBP}_{P,R}^{ri}(x) = \min_{n \in \{0, \dots, P-1\}} \sum_{i=0}^{P-1} u(x_i - x) 2^{[(i+n) \bmod P]}$$

Since each rotation-invariant pattern corresponds to several original patterns, this new descriptor is more compact than the original one, and can be more accurately characterized. Moreover, in [102] it was also observed that the so called *uniform* patterns,  $\text{LBP}_{P,R}^{u2}(x)$ , characterized by just two transitions, from 0 to 1 and back, account for the vast majority of occurrences in natural textures. For a given value of  $P$ , only  $P + 1$  uniform and rotation-invariant patterns exist, as opposed to a total of  $2^P$  generic patterns. Keeping only such patterns, and pooling all the remaining ones in a “left-over” pattern, leads to a very compact and reliable image descriptor.

Besides these basic descriptors, with fixed  $R$  and  $P$ , in [102] a multiscale version of LBP was also proposed, using various  $(R, P)$  combinations,  $(1, 8)$ ,  $(2, 16)$ , and  $(3, 24)$ , and uniform and rotation-invariant

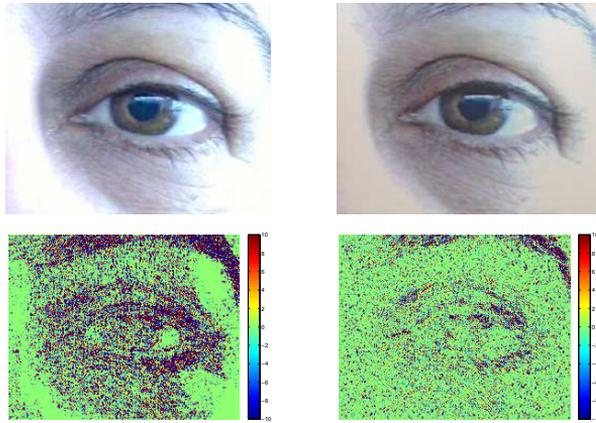
patterns, for a feature vector of length 54. Likewise, in the vast literature on LBP, a number of more complex variants have been proposed, not considered here for brevity.

### 3.4.2 Emphasizing local patterns

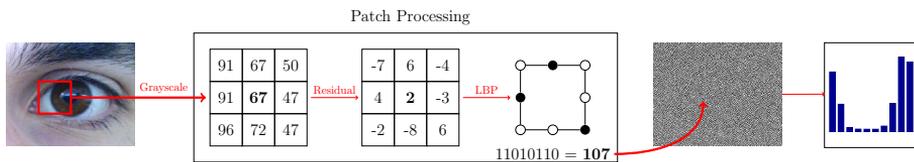
When trying to design highly discriminative features for a given application, it is very important to have a good statistical model of the images of interest and, especially, to work in the most appropriate domain. Once in these favourable conditions, one can effectively look for the distinctive features that enable a reliable classification. This general rule certainly applies to the iris liveness detection case. To gain insight about this, Fig. 4.5 shows three live iris images and the correspondent fakes taken from the MobBIOFake Database [116]. It should be clear that telling apart lives from fakes in this original domain is not a trivial task.

Working on the original intensity images might actually hide discriminative local alterations, preventing correct classification. Ways to improve the discrimination ability of LBP have been proposed in [151] and [148], where the local descriptor is computed on the magnitude and phase, respectively, of some Gabor filter outputs. In [152] LBP is computed after applying a Sobel filter along the vertical and horizontal directions in order to better enhance the edge information. The conjecture at the basis of such variants is that high-order derivative features contain more detailed and more discriminative information. This concept is well formalized in [147], through the use of local derivative patterns (LDP). LDP can capture the change of derivative directions among local neighbors, obtaining a better performance w.r.t. LBP, at the price of a higher complexity. In particular, the  $n^{\text{th}}$ -order LDP is a binary string describing, through co-occurrences, the local behavior of  $(n - 1)^{\text{th}}$ -order directional derivatives. Since there are several possible directions to take into account, the resulting code is 32 bit long for each pixel.

Similar ideas are used in steganalysis where a message hidden in the image data must be detected. In this context, features are extracted after some form of high-pass filtering, since fine details hold most of the information of interest. Typically, the low-pass content is suppressed by means of a predictor [32, 155, 109], obtaining the so-called prediction-error image, or also residual image, which can be more easily modeled



**Figure 3.10:** Live (left) and fake (right) iris images from the Mob-BIOfake database (top), with the corresponding residual images (bottom).



**Figure 3.11:** Block-diagram for the feature extraction step of the proposed method.

and analyzed. In [35] a large number of filters, of different order and dimension, are tested for this task. The same approach has been also used for camera identification [142] and forgery detection [18], confirming that working on fine details of images is extremely promising.

It is fair underlining that the use of some high-pass versions of the image has been already considered in biometric spoofing detection. In [92], for example, in the context of fingerprint liveness detection, a denoised version of the image is subtracted from the original one, and the variance of the noise residual in detail wavelet subbands is computed as a measure of coarseness, and used for classification. Since a very simple denoising filter is used, this noise residual does not differ much from the prediction residuals considered in the previous cases, containing high-frequency signal components useful for detection. Similarly, in [38] the majority of features proposed for iris liveness detection are evaluated on

the difference between the input image and a smoothed version. Local descriptors, however, hold a much stronger description power with respect to features computed globally, as proven by the technique recently proposed in [45], working on wavelet-domain residuals like [92] but providing a much better performance. It is worth pointing out that the use of residuals to help in distinguishing microstructures has been also used in other research fields like in [118] for material recognition.

Fig.3.10 provides some immediate insight on the value of the residual image. A live and a fake iris are shown together with their image residuals. While the original images do not exhibit striking differences, the live residual is much richer of details than the fake residual, suggesting that classification must be viable in this domain. One must therefore rely of fine-grain details, seemingly invisible alterations of the natural characteristics of the image, in order to detect the attack.

### 3.4.3 Liveness detection algorithm

Following the above discussion, the proposed algorithm comprises three steps:

1. computation of the high-pass image residual;
2. feature extraction based on a suitable LBP descriptor;
3. classification through SVM with a linear kernel.

We selected LBP as our basic tool because of its high descriptive power (confirmed by preliminary experiments) and, not least, its limited hardware requirements, in terms of both CPU and memory. LBP is computed on the residual obtained by means of a very simple high-pass filter already successfully used in steganalysis [61, 35]. In particular, w.r.t. the  $3 \times 3$  neighborhood of the target pixel  $x$  shown below

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ x_7 & x & x_3 \\ x_6 & x_5 & x_4 \end{bmatrix}$$

the residual  $r$  is computed as

$$r = x - \frac{1}{2} \sum_{i \text{ odd}} x_i + \frac{1}{4} \sum_{i \text{ even}} x_i$$

To avoid fractional coefficients, all quantities are actually multiplied by 4.

Note that, while LBP encodes first-order spatial variations computed on two-pixel supports, the use of a preliminary high-pass filters amounts to considering higher-order statistics computed on a larger support, as shown in the example below w.r.t. the main diagonal difference

$$\begin{bmatrix} 0 & -1 & 2 & -1 \\ 1 & 0 & -3 & 2 \\ -2 & 3 & 0 & -1 \\ 1 & -2 & 1 & 0 \end{bmatrix}$$

Turning to LBP, in order to meet the complexity constraints imposed by mobile devices, we use two simplified configurations which avoid interpolation, the only step requiring floating point operations, and hence the most onerous of the whole feature computation. In particular, we consider either only the 4-connected  $\{x_1, x_3, x_5, x_7\}$  or all the 8-connected  $\{x_0, \dots, x_7\}$  neighbors. Therefore, in the first case we sample the angular space quite roughly, and in the second case renounce rigorous rotation invariance. Despite the seemingly crude simplifications, the experimental analysis presented in next Section will fully support these choices.

Finally, we used a properly trained support vector machine (SVM) with a linear kernel for the actual classification. The overall procedure is outlined pictorially in Fig.3.11.

### 3.5 SID and Bag-of-Features

Bag of Words model (BoW) has shown its superiority over many conventional global features for image classification. The idea underlying BoW was originated by text classification [129] where the frequency of words was used as a feature for training a classifier. In this context the BoW is a set of words belonging to different topics. Given a text one can count the occurrence of each word in the BoW, obtaining the histogram of the BoW which features the text under examination. Texts that deal with different topics result in very different distributions that can tell them apart.

This approach has been then successfully used in image processing, since images can be treated as texts and thus words are represented by

image patches. To define words it is necessary to extract a vector of features from a large set of patches. These local descriptors are then clustered into groups, for example by using k-means, and each group is then represented by a centroid, the *codeword*, that now acts as a word. Finally, the image is represented by the histogram of these codewords. Each patch of the image under test is compared with those of the BoW to find the most similar one, in terms of Euclidean norm, thus incrementing the respective bin of the histogram of the BoW. This is equivalent to count the amount of patches belonging to each of the clusters defined from the BoW.

A first technique of this kind was proposed for the task of image retrieval [154]. Following this approach, however the same object may not be recognized in different images due to some changes in visualization like scaling or rotation. For this reason later works exploit local descriptors which are robust with respect to such visualization changes. The approach, sometimes named Bag of Features, works on local descriptor of the patch instead of the patch itself to build up the BoW and to compute the distribution that characterizes each image. This approach was first exploited for texture classification [136] in which the local descriptors are the response of a filter bank. For each texture class the author compute 10 centroids via k-means algorithm to build up the BoW and then the classification is obtained comparing the BoW histogram of the image under test with the ones obtained from the training images. More recently, numerous [techniques/variants] have been proposed with the aiming of improving performance with respect to the basic approach. Main contributions of the scientific community are in two different ways: to improve local features, so as to achieve better local description of the image, or to improve global image features, overcoming the limits of the vector quantization.

Among recent local descriptors, those extracted on keypoints (e.g. SIFT [78]) are gaining more relevance for various computer vision application, such as object recognition, image retrieval or 3D reconstruction. This due to the fact that keypoints are able to catch the global appearance of the images. This allows us, for example, to recognize an object in different images taken from various points of view. However, in our application we aim to classify very similar images (e.g. live and fake access attempts) based on fine details that make the difference. Therefore we compute the local descriptors in a *dense* fashion for all the pixels of

---

the image. Then we resort to a global descriptor of the image using the BoW histogram with a learned vector quantization, in fact the BoW is computed each time from the training data, so that the classifier can better adapt to the specific data.

Among the dense descriptors recently proposed, SID [63] exhibits invariance to rotation and scale change and is computed in a very efficient way. For this reasons we adopted the SID features in the BoW environment. We have adopted this approach for the liveness detection application [44, 46], achieving both effectiveness and robustness. However the success of this approach is strictly related to the data. To correctly compare the histograms one should be sure that each sub-histogram contains information about the same part of the images. This means that same part of the images should appear in the same sub-regions of the partition, thus the images should be co-registered. Moreover, in our application the discriminant information between live and fake class is in the relationship between adjacent pixels, that is the same in all the regions of the image.

For these reasons, in [68] the authors propose to partition the image in regions with finer resolution and to compute sub-histograms of BoW for each of this region, thus keeping spatial information in the global image descriptor. A similar approach has been already proposed in [120]. In [135] it is shown that a soft-assignment of the vector to the more similar elements of the BoW can improve performances with respect to the classical hard-assignment (vector quantization). For each feature vector of the image under test we obtain a vector of weights representing the similarity with each element of the BoW. We have adopted this approach in [51] for image classification task. More recently [146, 139] new techniques have been proposed to compute this weights in the sparse representation domain, further increasing the descriptive power.



## Chapter 4

# Experimental results

In this Chapter we present the results of several experiments carried out to assess and compare the methods discussed so far on several liveness detection tasks. To show the generality of the proposed approaches based on local image descriptors, we also present the results for a cell images classification application. Software code is available online for all considered descriptors (see Tab.4.23), except for the comparisons with state-of-the-art approaches (whose performance results are drawn by the referred papers) present in the top part of each table. Experiments are carried out on publicly available datasets so as to guarantee fully reproducible research. Since we aim at comparing the discriminative power of the various descriptors, we fix all other experimental conditions, accepting also a possible (slight) impairment of performance w.r.t. the implementation used in the original paper. So, we use always a linear SVM classifier, which does not require parameter tuning, and avoid any feature selection procedure. No preliminary segmentation step is applied to fingerprint, finger veins and iris images for the two-class classification task of the liveness detection, where the background is very specific and constant, and can be assumed not to modify the relevant statistics. On the contrary, we propose a segmentation technique to exploit sclera information for the three-class lens classification task which significantly improves performances w.r.t. the state of the art. Segmentation is explicitly required also for one of the face datasets (Replay-Attack) to avoid biases, and a mask of size going from  $64 \times 64$  to  $80 \times 80$  pixels, including the face, is provided with the images. Therefore, we used segmentation also for the other available dataset (3DMAD)

building a square mask including the face, based on the position of the eyes, provided with the images. Finally, we use K-means clustering with Euclidean distance for the joint quantization of features (300 clusters for each class when not specified), although the use of more advanced techniques [13] appears as one of the most promising topics for future research.

All results for two-class classification problems are in terms of Half Total Error Rate,  $HTER = (FGR+FFR) / 2$ , where FGR[FFR] is the False Genuine[Fake] Rate, that is the percentage of fake[genuine] samples mis-classified as genuine[fake]. For multi-class classification problems we report per class Accuracy and the mean accuracy among classes. For some challenging cases we also plot the Receiver Operating Curves (ROC) in order to gain better insights into the relative performance and robustness of descriptors. Finally, we provide some hints on computational complexity through CPU-times, instead a deeper complexity analysis is presented for application on mobile devices for which this aspect is crucial.

## 4.1 Fingerprint

### Datasets description

For fingerprint images, performance has been tested on several datasets made available in the context of the LivDet 2009<sup>1</sup> [85], LivDet 2011<sup>2</sup> [144] and LivDet 2013<sup>3</sup> [40] competitions. For this last dataset we avoid using the Crossmatch dataset, since as explicitly said on the web-site it was affected by an acquisition problem and cannot be used for assessing the performance of a liveness detection algorithm.

All datasets, whose characteristics are summarized in Tab.4.1 - 4.3 respectively, comprise disjoint training and test sets. They differ from one another under many respects, number of subjects and samples, sensor (hence, image size and resolution), number and type of materials used for spoofing and, most important, modality of fingerprint reproduction, consensual (with the cooperation of the user) or not. Clearly, the consensual modality allows for better fakes, and hence is more challenging for the liveness detection module. This can be also appreciated

---

<sup>1</sup><http://prag.diee.unica.it/LivDet09/>

<sup>2</sup><http://people.clarkson.edu/projects/biosal/fingerprint/index.php>

<sup>3</sup><http://prag.diee.unica.it/fldc/>

DATASET	<i>LivDet2009</i>		
Scanner	Biometrika	CrossMatch	Identix
Model No.	FX2000	V300LC	DFR2100
Res. (dpi)	569	500	686
Image size	312x372	480x640	720x720
Live Samples	1986	4000	3000
Fake Samples	1986	4000	3000
Total subjects	50	254	160
Materials	1	3	3
Co-operative	Yes	Yes	Yes

**Table 4.1:** Characteristics of the datasets used in *LivDet2009*.

DATASET	<i>LivDet2011</i>			
Scanner	Biometrika	Italdata	Sagem	Digital Persona
Model No.	FX2000	ET10	MSO300	400B
Res. (dpi)	500	500	500	500
Image size	312x372	640x480	352x384	355x391
Live Samples	2000	2000	2000	2000
Fake Samples	2000	2000	2000	2000
Total subjects	200	92	200	82
Materials	5	5	5	5
Co-operative	Yes	Yes	Yes	Yes

**Table 4.2:** Characteristics of the datasets used in *LivDet2011*.

in Fig.4.1, which shows live and fake samples drawn from two consensual (Biometrika-2009 and Biometrika-2011) and a non consensual (Biometrika-2013) datasets. In the first two cases, live and fake samples are much more similar to one another. Another important aspect which impact the quality of the fake images is the type of material used, due to this we will present a per-material performance analysis.

DATASET	<i>LivDet2013</i>		
Scanner	Biometrika	Italdata	Swipe
Model No.	FX2000	ET10	-
Res. (dpi)	569	500	96
Image size	312x372	480x640	1500x208
Live Samples	2000	2000	2500
Fake Samples	2000	2000	2000
Total subjects	50	50	100
Materials	5	5	4
Co-operative	No	No	No

**Table 4.3:** Characteristics of the datasets used in *LivDet2013*.



**Figure 4.1:** Some examples of live (up) and fake (down) fingerprints coming from Biometrika for the three databases used.

Ref.	2009					2011					2013				
	Biom.	Xmatch	Identix	Biom.	D.P.	Ital.	Sagem	Biom.	Ital.	Swipe	avg.				
Winner	18.2 (13)	15.2 (16)	10.6 (16)	14.7 (14)	20.0 (17)	36.1 (16)	21.8 (17)	*1.7 (7)	*0.8 (1)	*3.5 (2)	(-)				
Pers./morph.	[84]	12.6 (11)	15.2 (16)	9.7 (15)	40.0 (17)	8.9 (12)	13.4 (16)	(-)	(-)	(-)	(-)				
IQA-based	[38]	12.8 (-)	10.7 (-)	1.2 (-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)				
HIG	[42]	(-)	(-)	(-)	(-)	(-)	(-)	3.9 (-)	*1.7 (-)	14.4 (-)	(-)				
CNN	[97]	9.5 (8)	*3.8 (5)	2.8 (13)	9.9 (10)	*1.9 (1)	7.9 (11)	4.6 (13)	47.7 (16)	6.0 (8)	9.9 (8.6)				
LBP	[102]	28.6 (17)	8.6 (14)	32.6 (17)	10.2 (11)	12.8 (15)	9.1 (14)	*1.6 (6)	3.0 (9)	10.1 (14)	13.4 (12.6)				
CoA-LBP	[99]	23.6 (14)	6.8 (12)	1.2 (5)	*6.8 (4)	*2.4 (3)	13.3 (5)	*1.8 (8)	2.0 (3)	5.0 (5)	6.8 (6.4)				
Ric-LBP	[100]	24.4 (16)	*3.5 (4)	1.4 (9)	*6.5 (3)	6.4 (10)	17.5 (9)	*1.2 (3)	2.6 (6)	6.0 (8)	7.5 (7.5)				
WLD	[43]	*1.2 (3)	5.7 (8)	2.0 (10)	13.3 (13)	13.8 (16)	27.7 (13)	6.7 (10)	5.2 (14)	7.1 (13)	8.4 (12)				
LPQ	[41]	5.1 (5)	6.0 (10)	2.9 (14)	12.8 (12)	9.7 (14)	15.6 (8)	6.3 (9)	2.5 (9)	3.6 (11)	7.1 (11)				
LBP+LPQ	[43]	10.5 (10)	5.1 (7)	2.3 (12)	*6.9 (6)	9.6 (13)	*13.8 (6)	6.1 (8)	*1.3 (5)	4.3 (12)	5.9 (7)				
WLD+LPQ	[43]	*0.3 (1)	*3.1 (1)	1.2 (5)	7.2 (7)	8.0 (11)	*12.7 (4)	*3.7 (2)	*0.8 (1)	2.6 (6)	4.6 (4.8)				
wt-Markov	[45]	23.7 (15)	*4.0 (6)	*0.7 (2)	16.5 (15)	*2.9 (4)	35.4 (15)	8.5 (12)	3.0 (11)	11.4 (15)	11.8 (11.0)				
BSIF	[40]	9.0 (6)	5.8 (9)	*0.7 (2)	*6.8 (4)	4.1 (6)	13.9 (7)	5.6 (6)	*1.1 (2)	3.0 (9)	5.2 (6)				
LCPD	[47]	*1.0 (2)	*3.4 (3)	1.3 (7)	*4.9 (1)	4.7 (7)	12.3 (3)	*3.2 (1)	*1.2 (3)	*1.3 (2)	4.7 (4)				
Keyp. SIFT	[78]	12.6 (11)	6.1 (11)	*0.5 (1)	20.9 (16)	5.0 (8)	28.0 (14)	*4.2 (3)	8.8 (15)	2.2 (5)	12.3 (16)				
Dense SIFT	[78]	9.5 (8)	11.7 (15)	1.3 (7)	8.5 (9)	*3.0 (5)	27.2 (12)	8.7 (13)	3.6 (12)	2.0 (3)	*3.8 (3)				
DAISY	[130]	9.0 (6)	6.8 (12)	2.0 (10)	7.3 (8)	5.0 (8)	25.7 (11)	11.5 (15)	9.4 (16)	10.9 (14)	*2.8 (1)				
SID	[63]	3.8 (4)	*3.3 (2)	*0.7 (2)	*5.8 (2)	*2.0 (2)	11.2 (2)	*4.2 (3)	2.5 (9)	2.7 (8)	9.3 (13)				
											4.5 (4.7)				

**Table 4.4:** Performance comparison for fingerprint images. The best result on average is reported in bold. \* denotes a statistically significant difference ( $p < 0.05$ ) with respect to the runner-up.

	EcoFlex	Gelatine	Latex	Playdoh	Silgum	Silicone	Modasil	WoodGlue	Bodydouble	Other	TOTAL
Biometrika 2011	116	169	129	-	154	-	-	256	-	-	824
Dig.Pers. 2011	-	184	246	100	-	41	-	127	-	-	698
Italdata 2011	367	308	159	-	203	-	-	668	-	42	1747
Sagem 2011	-	361	16	64	-	73	-	68	-	-	582
Biometrika 2013	0	43	28	-	-	-	3	28	-	-	102
Italdata 2013	22	239	74	-	-	-	32	169	-	-	536
Swipe 2013	-	-	95	792	-	-	-	71	48	-	1006

**Table 4.5:** Total number of errors , aggregated over all descriptors, for each of the material used in *LivDet2011* and *LivDet2013*.

---



---

Reference	Biometrika	Crossmatch	Identix
position	10.33	3.58	0.36
scale	14.94	11.80	1.78
orientation	5.42	6.65	0.76
position+scale	9.86	2.92	0.33
position+orientation	5.12	3.43	0.29
scale+orientation	7.98	5.48	0.62
all (proposed)	5.36	2.82	0.33

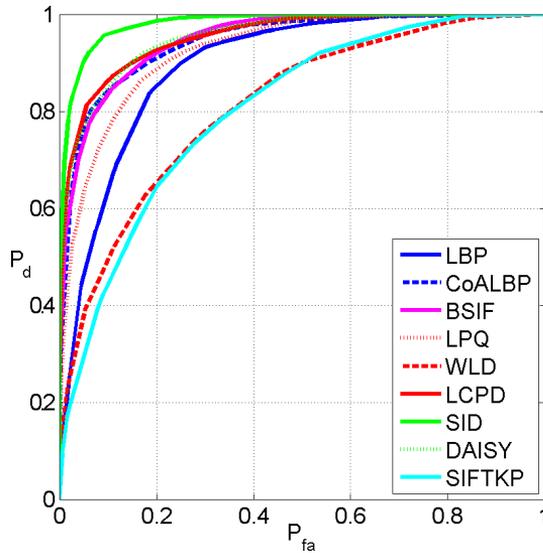
---

**Table 4.6:** Results on *LivDet2009* for the wt-Markov feature using PCA and RBF kernel SVM. Different contributions of the features are reported.

## Experimental results

Numerical results are reported in Tab.4.4. For each descriptor we provide a reference, either the paper where it was first used for liveness detection or, when this never happened, the paper where it was originally proposed. We also report two examples of features concatenation, namely LBP+LPQ and WLD+LPQ, combining complementary information of local texture analysis and local phase of the image in a straightforward manner. Together with the error rate, we show the rank observed for each dataset. This is important when we consider averages, in the last columns, because the average error rate depends very much on the most challenging datasets, while the average rank is more robust to outliers. In addition, we mark with a star the best result and all results with a statistically insignificant ( $p = 0.05$ ) difference with respect to it. External performance references are given by the winner of the competition (first row), by two techniques not based on local descriptors [84, 38], by one using a variant of SIFT and HOG [42] and by the CNN technique proposed in [97], and based, like [90], on the implementation of [15], but tested on all datasets. Note that some of these algorithms provide results only for one of the available datasets, hence some data are missing. Of course, in some cases the experimental protocol differs from ours. For example, to further reduce the error rate, [97] uses PCA for feature reduction and non-linear SVM for classification.

A first remark concerns the very good performance of LD-based techniques in general. Most descriptors guarantee an average error rate be-



**Figure 4.2:** ROCs for the Italdata 2011 dataset.

low 10%, and three of them below 5%. The large gap w.r.t. the winner of the 2009 and 2011 competitions testifies on the great advances observed in this field in the last few years, mostly due to the use of local descriptors. LCPD provides the best average performance and the best average rank, although results are slightly worse than in the original paper [47] where feature selection is also used. This is even more relevant for wt-Markov features, in fact in the original paper [45] PCA is adopted to dramatically reduce the features dimensionality from 6000 to less than 100, then an RBF kernel SVM is trained. For comparison, we report in Tab.4.6 the original paper results for each features configuration and the final proposed approach. LCPD was developed explicitly for this biometric traits, trying to capture its peculiarities through dedicated expressive features. The good discriminant power of those local features, based on local contrast and phase, is confirmed by the runner-up that is the concatenation of WLD and LPQ. Nonetheless, the unspecific SID follows very closely, showing the importance of invariance properties and the adaptivity of the BoW approach. Fourth comes BSIF, suggesting that adaptive filters can provide a further boost. Putting together the best peculiarities of these approaches is not easy, but certainly an interesting avenue for next research. In Fig.4.2 we show the ROCs of the

---

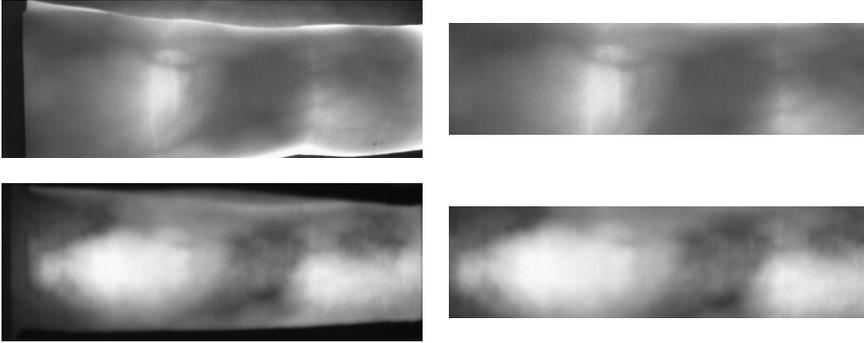
descriptors for the very challenging Italdata-2011 dataset, which shows SID to provide not only a good but also robust performance.

In order to gain a deeper insight on the influence of each material used to create the spoofed fingerprint, in Table 4.5 we present the total number of errors, aggregated over all descriptors, for each of the material used in Livedet2011 and Livedet2013 datasets. Note that each dataset includes 200 fakes for each type of material except Swipe where the number is 250. It appears that fakes made of silicone seem easier to detect, while those made of gelatin pose a more severe threat to the system. Note, however, that there is also a strong dependence on the specific sensor as well as on the type of spoofing (consensual or not). For example, with the same sensor and the same material, errors increase dramatically if the procedure is consensual.

## 4.2 Finger Veins

The vulnerability of finger vein recognition to spoofing attacks has emerged as a crucial security problem in the recent years mainly due to the high security applications where this technology is used.

Among all biometric technologies, finger vein recognition is a fairly new topic, which utilizes the vein patterns inside a person's finger. This technology is widely used in the financial sector for its intrinsic robustness to spoofing attack. The fact that the vein pattern used for identification is embodied inside the finger prevents the data to be easily stolen, contrary to face that can be captured with a camera or fingerprints that can be collected from latent prints. However, acquiring images of finger vein patterns is not impossible. In fact recent works shown that finger vein biometrics is vulnerable to spoofing attacks, pointing out the importance to investigate counter-measures against this type of fraudulent actions. The goal of the 1st Competition on Counter Measures to Finger Vein Spoofing Attacks is to challenge researchers to create counter-measures effectively detecting printed attacks. The submitted approaches have been evaluated on the Spoofing-AttackFinger Vein Database and the results are presented in [131].



**Figure 4.3:** Some examples of live (up) and corresponding printed fake (down) finger veins images. The leftmost are extracted from the *full* subset, while the rightmost came from the *cropped* subset.

## Dataset description

The Spoofing-Attack finger vein database<sup>4</sup> [132] consists of 440 index finger vein images of both real-access and attacks attempt to 110 different identities. The total number of images in the database is 880 (240 in the training set, 240 in the development set and 400 in the test set.) It is important to highlight that clients that appear in one of the data sets (train, dev or test) do not appear in any other set. The established protocol assert to use the training set in order to train the anti-spoof classifier and the development set for threshold estimation. Finally, the test set was used to report the final performance. The competition were split into 2 different sub-tasks according to the visual information available: *full* printed images and *cropped* images. Two live samples and their correspondent fakes are shown in Fig.4.3. This classification scheme makes-up a total of two protocols that can be used for studying the performance of counter-measures to finger vein attacks. Both full and cropped protocols are designed to use prior information based on trained classifiers or non trained approaches where the decision is taken based just on the input.

<sup>4</sup><https://www.idiap.ch/dataset/fvspoofingattack>

## Experimental results

We propose two different approaches to handle cropped and full images, both based on the use of local descriptors.

For cropped images we extract the LBP [102] over residual (high-pass version) of the image with  $3 \times 3$  integer kernel as described in [48], in order to improve the discrimination ability of LBP and better explore the image statistics. In particular, for a  $3 \times 3$  neighbourhood of the target pixel  $x$  shown below:

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ x_7 & x & x_3 \\ x_6 & x_5 & x_4 \end{bmatrix} \quad (4.1)$$

the residual  $r$  is computed as:

$$r = x - \frac{1}{2} \sum_{i \text{ odd}} x_i + \frac{1}{4} \sum_{i \text{ even}} x_i \quad (4.2)$$

To avoid fractional coefficients, all quantities were multiplied by 4. Note that, while LBP encodes first-order spatial variations computed on two-pixel supports, the use of a preliminary high-pass filters amounts to considering higher-order statistics computed on a larger support. LBP is then evaluated on the residual image  $r$  by considering 8 neighbours sampled uniformly on a circle of radius 1. The resulting vector is formed by 256 features. For full images indeed we used the concatenation of Local Phase Quantization (LPQ) [41] and Weber Local Descriptor (WLD) [14]. Since these two descriptors extract information on the patch in different domains we complement one another and can give better results in terms of discrimination ability [43]. The resulting combined vector is formed by 1,216 features.

Finally, for both the approaches we train an SVM with linear kernel as classifier.

Results of the contest are shown in Tab.4.7, where is indicated with *GRIP-PRIAMUS* the team name for the proposed approach. Performances are evaluated in term of HTER and, in case of a tie, *decidability*  $d'$ , a measure of the classes separability, that is defined as:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2} (\sigma_1^2 + \sigma_2^2)}} \quad (4.3)$$

The results are compared with a baseline approach, a texture-based algorithm implemented in Python<sup>5</sup> that exploits subtle changes in the finger vein images due to printed effects computing a global feature in the frequency domain. To recognize the static texture, the Fourier Transformation is extracted from the raw image after applying an histogram equalization. Then the percentage of energy in a vertical subband, that is weakly manifested in fake images, is computed obtaining a score in  $[0, 1]$ .

For full protocol, 3 over 4 approaches (including the proposed one) reach perfect performances, however the baseline one achieves best decidability among the participants. Meanwhile the crop protocol seems to be a harder task to address, indeed only the proposed approach still achieves perfect performances. In particular the baseline achieves worst performances in this case, showing also the weakness of global features in certain cases. Please refer to [131] for a more detailed performances analysis.

Team	Protocol	Development				Test			
		FAR	FFR	HTER	$d'$	FAR	FFR	HTER	$d'$
Baseline	<i>Full</i>	0.00	0.00	0.00	9.75	0.00	0.00	<b>0.00</b>	11.17
GUC		0.00	0.00	0.00	5.46	0.00	8.00	4.00	4.47
CVSSP		0.00	0.00	0.00	9.05	0.00	0.00	0.00	8.06
GRIP-PRIAMUS		0.00	0.00	<b>0.00</b>	11.10	0.00	0.00	0.00	8.03
Baseline	<i>Cropped</i>	4.17	40.83	22.50	1.58	11.00	30.00	20.50	1.82
GUC		0.00	0.00	0.00	5.09	1.50	4.00	2.75	3.81
CVSSP		0.00	0.00	0.00	6.68	0.00	2.50	1.25	5.54
GRIP-PRIAMUS		0.00	0.00	<b>0.00</b>	6.28	0.00	0.00	<b>0.00</b>	5.20

**Table 4.7:** Results of the 1st Competition on Counter Measures to Finger Vein Spoofing Attacks for the development set and the test set. The proposed approach, namely GRIP-PRIAMUS, reach perfect classification performances on both databases. For the full, best performances in terms of decidability are obtained from the baseline, while for the cropped database the proposed approach ranks first.

<sup>5</sup>[https://pypi.python.org/pypi/antispoofing.fvcompetition\\_icb2015](https://pypi.python.org/pypi/antispoofing.fvcompetition_icb2015)

---

## 4.3 Iris

Biometric authentication systems, based on fingerprints, face, iris, or other distinctive traits, provide security with little involvement on part of the user [134]. Iris-based systems [79, 25, 8], in particular, are very popular thanks to their high reliability in both identification and verification tasks. The iris pattern is unique for each individual, even for identical twins, and so rich of distinctive features that a casual wrong identification is rarely observed. Moreover, as an internal organ of the eye, the iris is well protected from the environment and stable with age [24].

However iris scanning authentication systems are not free from attack. Two different spoofing approaches are possible: those based on a printed version of the iris image or using contact lens to modify the iris pattern of the subject.

### Datasets description

For iris liveness detection we follow the same experimental protocol used for fingerprints, so we point out, here, only significant differences. the most important being the spoofing modality, which can be based on printed iris images or cosmetic contact lenses.

There are two datasets available for the first type of attack, ATVS<sup>6</sup> [113] and Warsaw<sup>7</sup> [20]. In the same category we find two datasets acquired using mobile devices: MObBioFAKE<sup>8</sup> [117] and MICHE [28] dataset. Even very simple descriptors, suitable for an implementation on mobile devices, provide a perfect accuracy on these datasets [48]. Due to this here we do not report the performances of all the descriptors for these latter datasets.

ATVS database has been created using iris images from 50 users of the BioSec baseline database. After a preprocessing, the images are printed on a piece of paper and then presented at the iris sensor, obtaining the fake image. The authors accurately control the quality of the fake samples exploring a wide range of settings including two different printers, various paper types and a specific enhancement of the image before the print. Therefore, data for the experiments consists of

---

<sup>6</sup><http://atvs.ii.uam.es/>

<sup>7</sup><http://people.clarkson.edu/projects/biosal/iris/registration.php>

<sup>8</sup><http://mobilive2014.inescporto.pt/>

---

100 distinct eyes acquired in two different sessions for a total of 800 iris images, and its corresponding printed fakes.

Similar acquisition procedure has been carried out for the Warsaw database. The data was collected for 237 volunteers for a total of 426 authentic eyes, ending up with 1274 images. Based on all authentic images, the authors prepared the printouts and then checked their fraudulent power in a commercial ET-100 camera, thus discarding those fake images exhibiting poor quality. The printouts that passed the verification step successfully, were then photographed by the AD100 camera. The total number of printout images is 729 acquired from 243 distinct eyes and partitioned into “low resolution” printouts prepared with the HP LaserJet 1320, and “high resolution” printouts, prepared with the Lexmark c534dn.

The MobBIO Multimodal Database [116] comprises biometric data from 100 volunteers, each one contributing 16 images (8 of each eye). The samples were acquired by an Asus Transformer Pad TF 300T, using the back camera, version TF300T-000128, with 8 MP of resolution and autofocus. The iris images, of size  $250 \times 200$  pixels, were captured in very different conditions in order to consider a large variability of scenarios. The MobBIOfake dataset [117] is composed by 800 iris images from MobBIO and their corresponding fake copies, and is already divided into a training and a test set, each composed by 400 live images and 400 fake ones. The fakes were obtained from printed images of the original ones. A contrast enhancement was applied to improve their quality, then the images were printed using a professional printer and high quality photographic paper and eventually acquired using the same portable device and in similar lighting conditions as the original ones. Examples of iris images belonging to this dataset and the relative fake images are shown in Fig.4.5.

In the MICHE dataset described in detail in [28] we considered two different type of attacks: printed-iris images and screen-iris ones. In particular, by referring to the first scenario 75 subjects participated to the test for a total of 338 live iris images of dimension  $4128 \times 2322$  pixels acquired by a Samsung Galaxy S4, forty of these images were then printed on a high-quality device and used as fakes to test the system. Besides the printed-iris attack, explicit object of this investigation, we considered also another scenario to gather some information on robustness, where the attack is carried out by presenting the iris image reproduced

on the screen of an iPad-mini with a  $1024 \times 768$  pixel resolution. For this scenario we collected 640 fake and 640 live images of dimension  $3264 \times 2448$  pixels by means of an iPhone 5. Examples of iris images belonging to this dataset and the relative fake screen images are shown in Fig.4.6.

Four datasets are instead relevant to the second type of attack, Notre Dame<sup>9</sup> I and II, and IIIT<sup>10</sup> Cogent and Vista. Images of irises with soft contact lenses are considered as genuine, because the iris pattern is still visible through the lenses, enabling correct identification. Those datasets will be used also in the next Section with the aims of detect the specific contact lens in a three-class classification problem.

The Notre Dame Contact Lens Detection database is divided into two datasets. Dataset I consists of a training set of 3000 images and a verification set of 1200 images, all acquired with an LG 4000 iris camera. Instead dataset II consists of a training set of 600 images and a verification set of 300 images, all acquired with an IrisGuard AD100 iris camera. Both datasets are divided equally into three classes:

1. no contact lenses,
2. soft, non-textured contact lenses,
3. textured contact lenses.

In particular, textured contact lenses of various colors and coming from three different suppliers are represented. Moreover for each image a manual segmentation of the iris region is provided. Samples images are shown in Fig.4.13.

The IIIT-D Contact Lens Iris database is divided into two dataset captured using two iris sensors: Cogent dual iris sensor (CIS 202) and VistaFA2E single iris sensor. For each subject, images without lens, with soft lens, and with textured lens are captured. The database is comprised of 6570 iris images pertaining to 101 subjects (both left and right eyes). Like Notre Dame dataset, contact lenses of various colors and different manufacture are represented. The database contains a minimum of three images for each iris class in each of the above mentioned lens categories for both the iris sensors.

---

<sup>9</sup>[http://www3.nd.edu/~cvrl/CVRL/Data\\_Sets.html](http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html)

<sup>10</sup><https://research.iiitd.edu.in/groups/iab/irisdatabases.html>

DATASET	Warsaw	ATVS
Printers (dpi)	HP LaserJet 1320 (600)	HP Deskjet 970cxi (600)
	Lexmark c534dn (1200)	LaserJet 4200L (1200)
Camera	IrisGuard AD100	LG Iris Access EOU3000
Image size	480x640	480x640
Live Samples	1274	800
Fake Samples	729	800
Total subjects	237	50
Total Eyes	243	100

**Table 4.8:** Characteristics of printed iris datasets.

DATASET	MobBIOFake	MICHE	MICHE
Attack type	Print	Print	iPad mini screen
Camera	Asus Transformer Pad TF 300T	Samsung S4 iPhone 5	iPhone 5
Image size	250x200	4128x2322	3264x2448
Live Samples	800	338	640
Fake Samples	800	40	640
Total subjects	100	75	75
Total Eyes	200	150	150

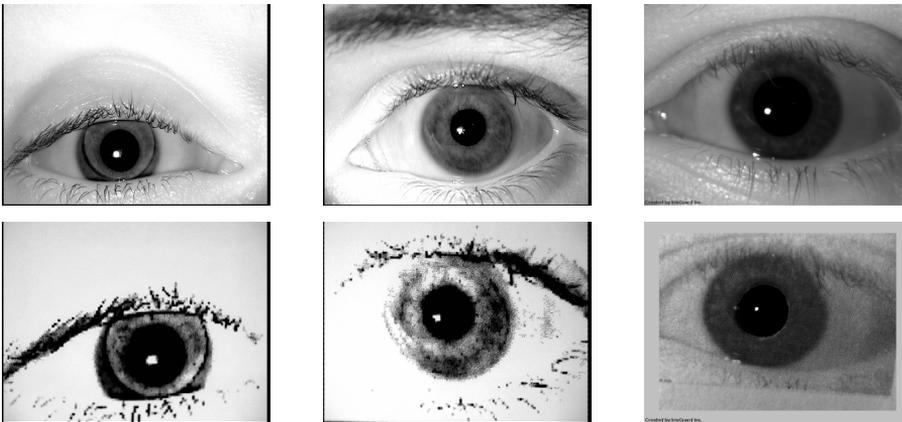
**Table 4.9:** Characteristics of printed iris datasets.

The reader can refer to [143] for a detailed description of Notredame and IIIT-D datasets. Information on all these datasets is provided in Tab.4.8 - 4.10, while Fig.4.4 - 4.7 show examples of live and fake images obtained from both printed irises or cosmetic contact lenses.

Notredame, Warsaw and MobBIOfake datasets are already divided in training and test sets, hence we made a two-fold cross validation both on Cogent and Vista and on ATVS, so as to compare our results with [38]. Instead for MICHE dataset, due to the low amount of fake printed samples, we performed a leave-one-out per subject cross-validation procedure.

DATASET	ND-I	ND-II	Cogent	Vista
Camera	LG 4000	IG AD100	CIS 202	VistaFA2E
Image size	480x640	480x640	480x640	480x640
Live Samples	2800	600	2306	2010
Fake Samples	1400	300	1160	1005
Total subjects	213	69	101	101
Total Eyes	287	89	202	202
Lens Manufactur.	3	3	2	2
Lens Colors	4	4	4	4

**Table 4.10:** Characteristics of contact lens iris datasets.



**Figure 4.4:** Some examples of live (up) and printed fake (down) iris images coming from *ATVS* and *Warsaw*.

## Experimental results

Numerical results are reported in Tab.4.11. In this case, SID is by far the best descriptor, ranking often first, and proving hence to work well with different biometric traits (note that it has been also used for contact-lens iris classification giving again very good performance [44]). The fingerprint-specific LCPD, however, keeps providing a good performance, probably because of deep structural similarities between the discriminative micro-textures found in fingerprint and iris images. In general, it seems that printed irises can be reliably recognized as fakes by SID and some other descriptors. Cosmetic contact lenses, as expected,



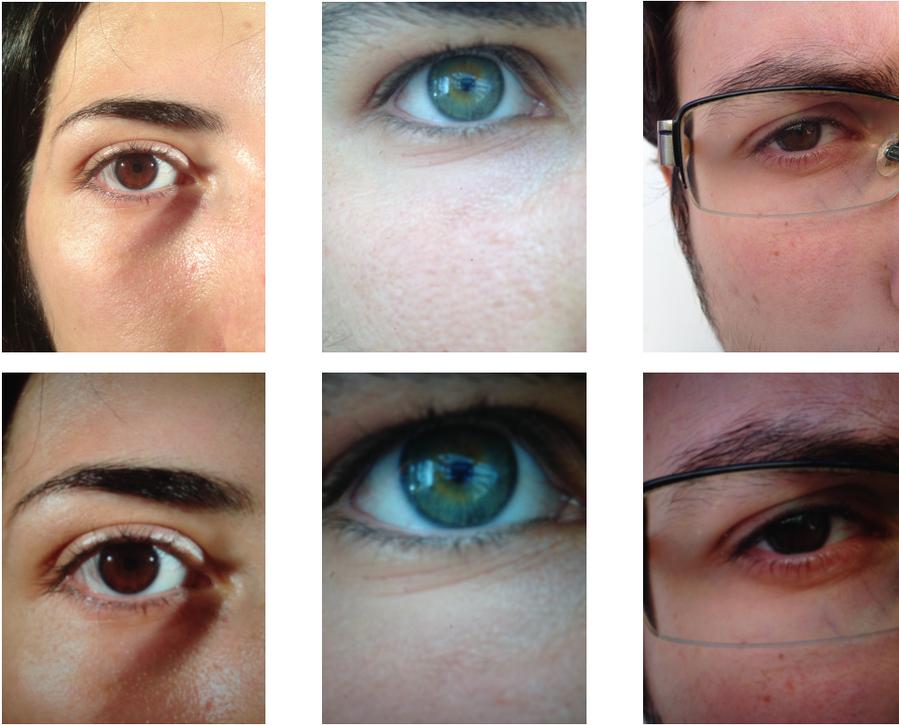
**Figure 4.5:** Some examples of live (top) and fake (bottom) printed iris images coming from MobBIOfake database.

pose more serious problems, especially for the IIT datasets. In Fig.4.8 we show the ROC curves for the Cogent dataset, with SID consistently superior to all other descriptors.

### Experimental results for mobile devices application

For the application over mobile devices, the proposed approach [48] performances has been assessed using ad-hoc datasets previously presented. In this Section we analyze the performance of a number of LBP-based image descriptors over several image databases used in the iris liveness detection field. We consider the three neighborhood systems of Fig.3.9, that is circular (identified with symbol  $\circ$ ), cross (+), and square ( $\square$ ), always with  $R = 1$ . For each one, we consider the basic (basic), rotation-invariant (ri), uniform (u2) and rotation-invariant uniform (riu2) features. As said before, we neglect all complex LBP variants, with the only exception of the multiresolution (MR) LBP on circular neighborhoods, with rotation-invariant uniform (riu2) features, included for comparison.

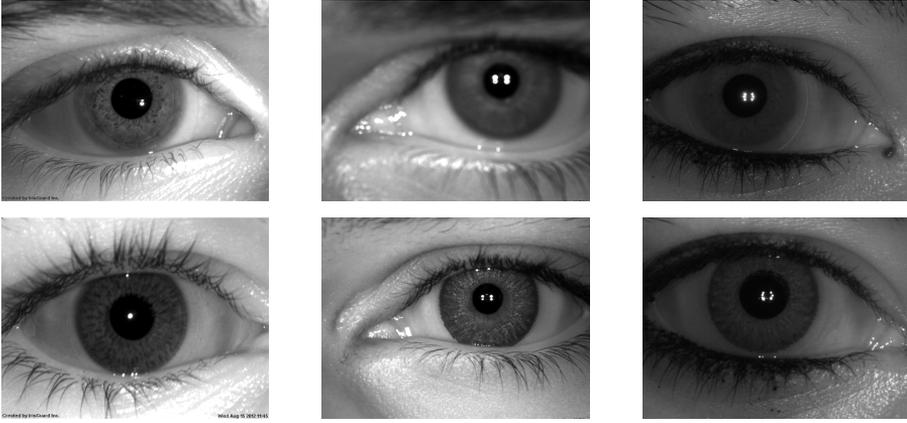
Results for MobBIOfake database for all considered LBP-based descriptors are reported in Tab.4.12 in terms of Half Total Error Rate (HTER). For each descriptor, we consider two versions, with features computed on the original, and on the residual images. From the anal-



**Figure 4.6:** Examples of live (top) and fake (bottom) screen iris images from MICHE database.

ysis of results, two facts emerge clearly: *i*) descriptors computed on the original images do not provide a satisfactory performance, and *ii*) almost *all* descriptors computed on residuals provide a near-perfect performance. However, it seems advisable not to use the rotation-invariant uniform features which, though very short, lead to some performance impairment. Looking for the best descriptor, in these conditions, is meaningless, since all those with an average error below 0.48 (marked with a star) are statistically indistinguishable ( $p < 0.05$ ). Several of these descriptors compute very fast, and are therefore very appealing for implementation on a mobile platform.

To gain insight about the absolute significance of these results, Tab.4.13 reports the performance obtained by the methods participating in the 1st Mobile Iris Liveness Detection Competition on the same test set (MobBIOfake) used to obtain the results shown in Table 4.12. Only



**Figure 4.7:** Some examples of live (up) and contact lens fake (down) iris images coming from Notredame, Cogent and Vista.

%	N.dame I	N.dame II	Cogent	Vista	Warsaw	ATVS	average
IQA-based [38]	(-)	(-)	(-)	(-)	(-)	2.2 (-)	(-)
CNN [90]	(-)	(-)	(-)	(-)	*0.2 (-)	(-)	(-)
LBP	3.6 (11)	2.8 (6)	20.6 (10)	8.9 (10)	1.9 (5)	*0.0 (1)	6.3 (7.2)
CoA-LBP	0.8 (8)	3.3 (8)	9.9 (2)	*1.9 (1)	26.9 (11)	*0.0 (1)	7.1 (5.2)
Ric-LBP	2.6 (10)	9.3 (11)	14.5 (5)	3.6 (5)	4.6 (7)	*0.0 (1)	5.8 (6.5)
WLD	*0.0 (1)	3.0 (7)	32.9 (11)	10.0 (11)	10.1 (9)	0.5 (11)	9.4 (8.3)
LPQ	0.4 (7)	2.0 (4)	16.4 (7)	7.0 (8)	1.4 (4)	*0.0 (1)	4.5 (5.2)
BSIF	*0.0 (1)	2.5 (5)	17.9 (9)	4.5 (6)	23.1 (10)	*0.0 (1)	8.0 (5.3)
LCPD	*0.1 (4)	*0.8 (3)	11.0 (3)	*3.1 (3)	7.1 (8)	*0.0 (1)	3.7 (3.7)
Keypoint SIFT	1.8 (9)	4.8 (10)	15.1 (6)	5.6 (7)	2.0 (6)	*0.1 (1)	4.9 (6.5)
Dense SIFT	*0.2 (6)	3.5 (9)	13.9 (4)	*2.5 (2)	0.5 (2)	*0.0 (1)	3.4 (4.0)
DAISY	*0.0 (1)	*0.5 (2)	17.2 (8)	8.8 (9)	0.9 (3)	*0.1 (1)	4.6 (4.0)
SID	*0.1 (4)	*0.0 (1)	*6.2 (1)	3.5 (4)	*0.0 (1)	*0.0 (1)	<b>1.6 (2.0)</b>

**Table 4.11:** Performance comparison for iris images. The best result on average is reported in bold. \* denotes a statistically significant difference ( $p < 0.05$ ) with respect to the runner-up.

the two best ranking methods are comparable with the residual-based LBPs considered here. We also include, as a very recent literature ref-

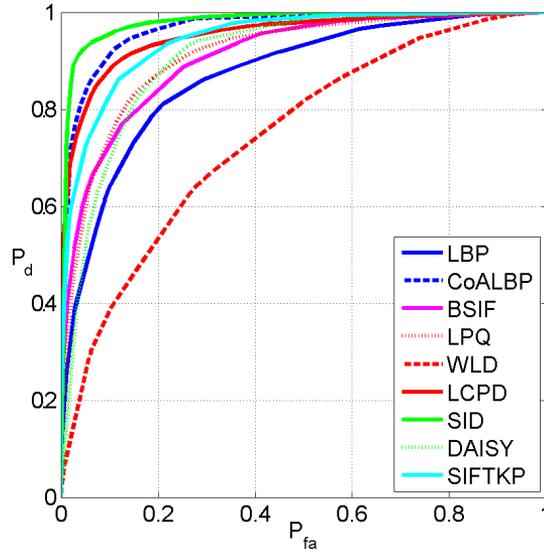


Figure 4.8: ROCs for Cogent dataset.

LBP desc.	on image			on residual		
	FGR	FFR	HTER	FGR	FFR	HTER
o-basic	9.25	0.25	4.75	0.00	0.00	*0.00
o-ri	8.75	0.75	4.25	0.00	0.00	*0.00
o-u2	11.25	0.75	6.00	0.25	0.25	*0.25
o-riu2	18.50	1.00	9.75	1.50	1.75	1.63
□-basic	3.50	0.25	1.88	0.00	0.00	*0.00
□-ri	17.00	1.75	9.38	0.00	0.00	*0.00
□-u2	5.75	0.25	3.00	0.25	0.25	*0.25
□-riu2	29.00	3.75	16.38	3.00	1.00	2.00
+ -basic	26.25	2.00	14.13	0.00	0.25	*0.13
+ -ri	30.25	3.50	16.88	1.00	0.50	0.75
+ -u2	27.00	3.25	15.13	0.00	0.00	*0.00
+ -riu2	30.25	3.50	16.88	1.00	0.50	0.75
MR-riu2	8.50	0.25	4.38	0.50	0.00	*0.25

Table 4.12: Performance of LBP descriptors on MobBIOfake.

erence, the technique proposed by [38], which does not provide better results.

These data, although encouraging, should be taken with caution. Of course, results may depend on the type of attack and on the experimental conditions, so we proceeded to experiments with two more datasets to

	FGR	FFR	HTER
1 <sup>st</sup> (IIT Indore)	0.00	0.50	*0.25
2 <sup>nd</sup> (GUC)	0.00	0.75	*0.38
3 <sup>rd</sup> (Federico II)	0.00	1.25	0.63
4 <sup>th</sup> (LIV-IC-UNICAMP)	2.00	0.50	1.25
5 <sup>th</sup> (IrisKent)	3.75	0.25	2.00
6 <sup>th</sup> (HH)	7.00	29.25	18.13
Galbally 2014	2.25	0.25	1.25

**Table 4.13:** Performance of methods participating in the 1st Mobile Iris Liveness Detection Competition.

LBP desc.	HQ print			screen		
	FGR	FFR	HTER	FGR	FFR	HTER
o-basic	0.00	0.00	*0.00	0.16	0.00	0.08
o-ri	0.00	0.00	*0.00	0.31	0.16	*0.23
o-u2	0.00	0.00	*0.00	0.31	0.00	*0.16
o-riu2	0.00	0.00	*0.00	0.94	0.00	*0.47
□-basic	0.00	0.00	*0.00	0.16	0.00	*0.08
□-ri	0.00	0.00	*0.00	0.31	0.00	*0.16
□-u2	0.00	0.00	*0.00	0.31	0.00	*0.16
□-riu2	0.00	0.00	*0.00	4.22	0.47	2.34
+ -basic	0.00	2.50	1.25	3.28	0.31	1.80
+ -ri	0.00	2.50	1.25	8.13	3.44	5.78
+ -u2	0.00	0.00	*0.00	3.44	0.31	1.88
+ -riu2	0.00	2.50	1.25	8.13	3.44	5.78
MR-riu2	0.00	0.00	*0.00	0.31	0.16	*0.23

**Table 4.14:** Performance of residual-based LBP descriptors on MICHE.

study robustness to changing conditions.

Results for MICHE dataset are reported in Tab.4.14 considering now only the residual-based descriptors. They mostly confirm what was already observed for the MobBIOfake database, except for the descriptors with cross configuration, which now performs a little bit worse than the others. In any case, several simple residual-based LBP descriptors guarantee a near-perfect detection, even for the case of screen-based attack which, however, is a more challenging task.

---

## 4.4 Contact lens classification

A weak point of iris-based system is the use of contact lenses, widespread among users, which affects the performance of recognition systems. The impairment is quite significant in the case of *cosmetic* lenses, characterized by a colored iris pattern, however it may be not negligible also for *transparent* lenses, as recently shown in [3]. To guarantee the best possible performance, therefore, it is important to detect whether the user wears contact lenses, and of which type, colored or transparent.

We propose a new machine-learning technique for detecting the presence and type of contact lenses in iris images. Following the usual paradigm, we extract the regions of interest for classification, compute a feature vector based on local descriptors, and feed it to a properly trained SVM classifier. Major improvements w.r.t. current state of the art concern the design of a more reliable segmentation procedure and the use of a recently proposed dense scale-invariant image descriptor. Experiments on publicly available datasets show the proposed method to outperform significantly all reference techniques.

### Introduction

Most of the research on contact lens classification topic is focused on colored lenses, starting from the pioneering work of Daugman in 2003 [24] where printed irises are detected based on the periodicities left by the dot-matrix printers used to fabricate them. However, approaches based on the low quality of fakes become soon ineffective as technology advances. Recent successful approaches are based therefore on local descriptors, which are extremely effective for the analysis of microtextures, because they capture the fine-scale statistical behavior observed locally in small patches of the image. Indeed, the iris, with its complex structure, provides abundant textural information to exploit for classification purposes.

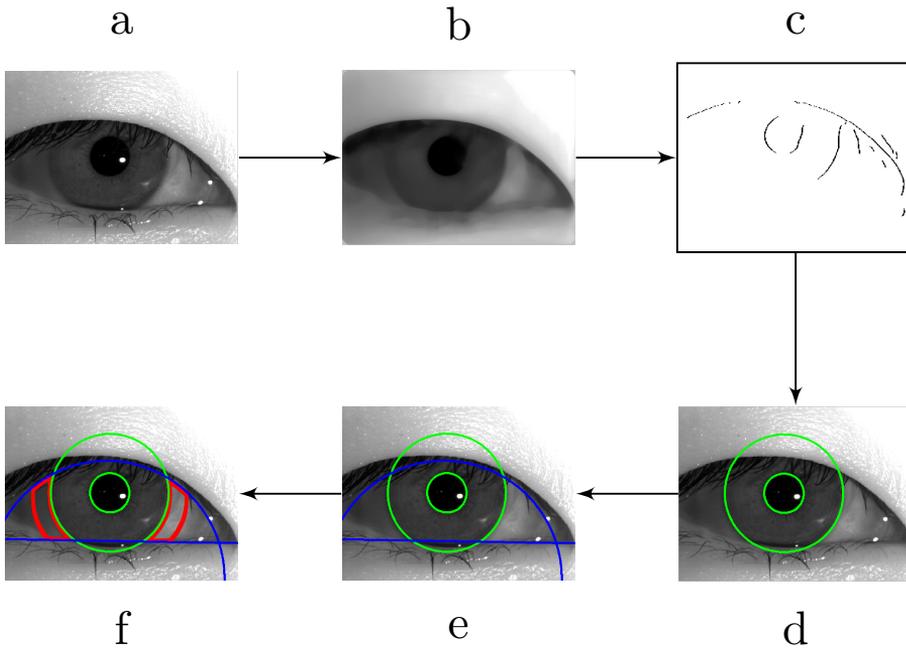
In [140] different measures are proposed: iris edge sharpness, iris-texton features based on Gabor filters and features based on co-occurrence matrix. Local binary patterns (LBP), extracted at multiple scales, are instead used in [54], while in [150] the SIFT descriptor is used to guide the LBP encoding procedure after a preliminary denoising. In [124] dense SIFT descriptors are used starting from the gradient image. Variants of LBP are adopted again in [143] for the full 3-class problem:

no-lenses, colored lenses, transparent lenses.

Besides the specific descriptor used, these techniques differ significantly in the pre-processing phase, typically involving a segmentation step, designed to select one or more regions of the eye image where relevant information can be extracted, and to discard data of no diagnostic value. In [140] and [124] the features are extracted only from the iris region, resampled in polar coordinates (a process called *normalization* in this field) so as to deal with more manageable rectangular images. In [54] the normalized iris is divided into six-subregions, corresponding to different scales of observation and orientation, computing and concatenating features for each region. In [150], instead, no normalization is applied, in order not to distort the original patterns, and features are computed in a square region bounding the iris, thereby including part of the sclera. The importance of collecting information also from the sclera and the pupil regions is explicitly recognized in [143] where, after normalization, LBP features are extracted independently from all regions and concatenated. [143] is especially valuable also because the Authors describe some iris image databases available online to assess the performance of competing techniques, establishing thus a solid experimental protocol.

In this section, therefore, we consider [143] as our main reference, dealing with the same 3-class problem, and keeping the same general framework based on segmentation and local descriptors. With respect to [143], however, our algorithm differs under several important respects: we use a real segmentation algorithm, that excludes eyelids and avoids normalization, taking into account only information belonging to the iris and part of the sclera; then, we replace LBP with the more sophisticated scale-invariant descriptor (SID), recently proposed in [63], and already applied with success to several classification problems [46, 44, 51], finally, we use the more effective Bag of Words (BoW) paradigm. Experimental results on the same datasets used in [143] confirm the superior performance of the proposed method.

In the rest of the section we describe the proposed segmentation, local descriptor, and classification procedure. Then we provide experimental evidence of the method's performance, and draw conclusions.

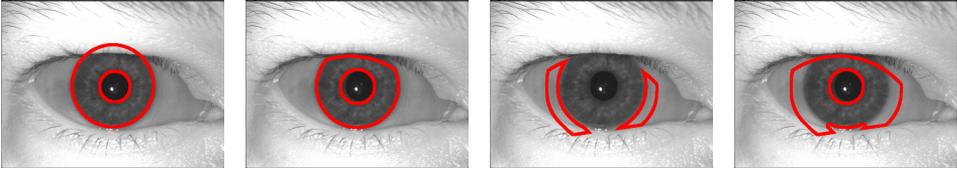


**Figure 4.9:** Steps of the segmentation algorithm: (a) original image, (b) median filtering, (c) edge detection, (d) iris boundaries detection, (e) eyelids detection, (f) sclera regions identification.

## Segmentation

The goal of our segmentation algorithm is to extract the iris and the sclera regions, which both convey precious information for the classification task. To this end, we aim at identifying the iris-sclera and iris-pupil circular boundaries, and then the boundaries with the upper and lower eyelids.

Our algorithm is inspired in the recent work of [114], based on the Circular Hough Transform (CHT) [5]. Given the well-defined structure of the regions of interest, [114] follows a parametric approach, where a circle of suitable center and radius is fit to the each boundary of interest, partially detected by means of an edge detection step. In our implemen-



**Figure 4.10:** Regions used for classifications: (a) ideal iris, (b) detected iris, (c) detected sclera, (d) detected iris+sclera.

tation, as in [114], we adopt the well-known Canny edge detector [9]. Preliminarily, however, we apply a large-window median filter on the image, so as to flatten out the iris pattern while preserving the strong iris-sclera and iris-pupil edges, thereby preventing the detection of useless edges in the region of interest. When looking for iris boundaries, we assign larger weights to vertical edges than to horizontal ones, in order to de-emphasize the edges related to upper and lower eyelid boundaries.

The core of the algorithm is in the Circular Hough Transform, which outputs a 3d matrix of coefficients,  $C(x, y, r)$ , where  $x$  and  $y$  are image coordinates and  $r$  a distance. The value of  $C(x, y, r)$  says how well the detected image edges are matched by a circle with center  $(x, y)$  and radius  $r$ . Therefore, the largest values in this matrix should correspond to the desired circular structures.

Needless to say, due to noise and occlusions, false alarms may occur. To reduce them, one can make a judicious use of the available prior information. First of all, assuming that the images are acquired in controlled conditions, with the pupil pretty well centered (which holds for our datasets), or that they are centered afterwards through some simple pre-processing, the CHT coefficients can be safely multiplied by Gaussian weights  $w(x, y)$  which penalize circles centered in the peripheral areas of the image. A more important point, however, and major difference of our algorithm w.r.t. [114], is that we look *jointly* for the two iris boundaries, that is, we try to identify two concentric circles, the iris-pupil and the iris-sclera boundaries, with significantly different radii. Although this condition might not fully hold, because pupil and iris can happen not to be perfectly concentric, this approach allows us to avoid most false alarms because we look for a very particular structure which can hardly arise due to spurious detected edges. For each candidate center  $(x, y)$ , we look for the two largest maxima of  $C(x, y, r)$  along the radius coordinate  $r$ , with the further constraint that the selected

radii,  $r_1 < r_2$ , are at least  $\Delta r$  pixels apart. Eventually, we appoint as center of iris and pupil the pixel for which the sum of these two values is maximum.

The localization of upper and lower eyelid boundaries is performed following closely the approach in [114], to which the reader is referred for details. In particular, in this case, the horizontal edges are emphasized, and the results obtained by the iris boundary detection are used to help localizing the upper and lower eyelids. Finally, to single out a significant portion of the sclera, a further ring concentric with the first two is built, with radius  $r_3 = \alpha r_2$ , and  $\alpha > 1$ . The region delimited by the circles of radius  $r_2$  (outer iris boundary) and  $r_3$  is then intersected with the region between the two eyelids. All these processing steps are summarized in Fig.4.9, while Fig.4.10 shows, for a typical test image, the various regions of interest and the ideal segmentation accompanying the image in the annotated database. It can be easily appreciated that the detected iris region is very close to the ideal segmentation.

Segmentation accuracy can be assessed by computing the F-measure between the detected region and the ideal circular ring provided with the dataset

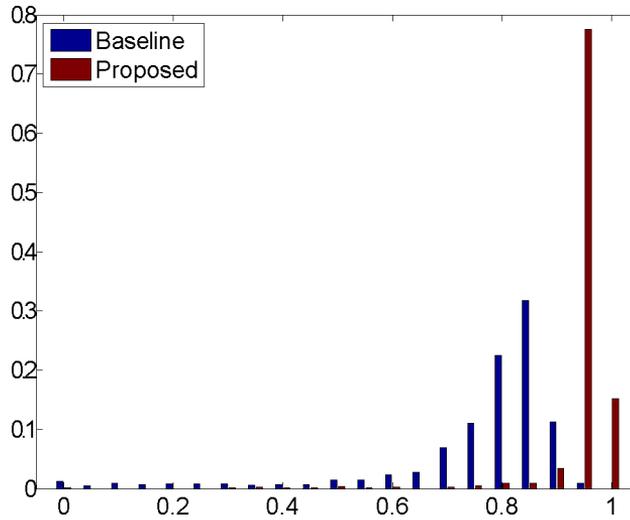
$$F = \frac{2 TP}{2 TP + FP + FN} \quad (4.4)$$

where TP, FP and FN are, respectively, the true positive, false positive and false negative pixels in the detected mask. We set segmentation parameters as  $r_1 \geq 30$ ,  $r_2 \leq 210$ ,  $\Delta r \geq 25$  and  $\alpha = 4/3$ . In Fig.4.11 we show the histogram of the F-measures obtained for the NotreDame II dataset with the proposed algorithm (red bars). For almost all images the F-measure is very close to 1 (perfect segmentation), while errors rarely occur. On the contrary, the algorithm of [114] is affected by frequent errors, and the corresponding histogram (blue bars) is generally shifted towards smaller values of F-measure.

In this work we will adopt the Scale-invariant descriptor (SID) [63] previously described in Sec.2.5.2. In Fig.4.12 a block-scheme of the patch processing that yields to the global image descriptor is depicted.

## Experimental results

To perform the classification task we considered the Bag of Words (BoW) model. The idea underlying BoW was originated by text classification where the frequency of words was used as a feature. This approach

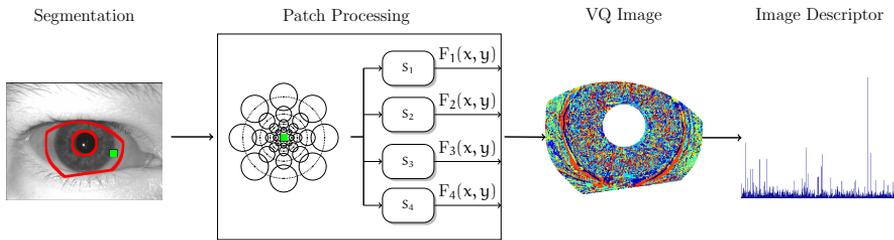


**Figure 4.11:** Histograms of F-measure values obtained with the proposed and reference segmentation algorithm on the NotreDame II database.

has been then successfully applied to the image processing field. To define words in this context it is necessary to extract suitable features associated with each pixel or patch. These local descriptors are then clustered, for example by using vector quantization, and each cluster is then represented by its centroid, the *codeword*, that now acts as a visual word. Finally, the image is represented by the histogram of these codewords. In our implementation, for each of the three classes, 200 codewords are computed for a total of a 600 dictionary elements.

We extracted the feature vectors for each pixel in the previously segmented regions, i.e. iris or sclera, where information about the lens may be present. These features are then vector quantized using the partitions defined by the dictionary, obtaining the histogram used as input of the SVM linear classifier.

The performance of the proposed method has been assessed on the publicly available datasets described in [143], we refer to that paper for their thorough description, reporting here only the characteristics of interest for the experiments. In particular, the NotreDame-I and II datasets are provided with an ideal segmentation of the iris, which is not available for the IIIT-D Cogent, and IIIT-D Vista datasets. Moreover,

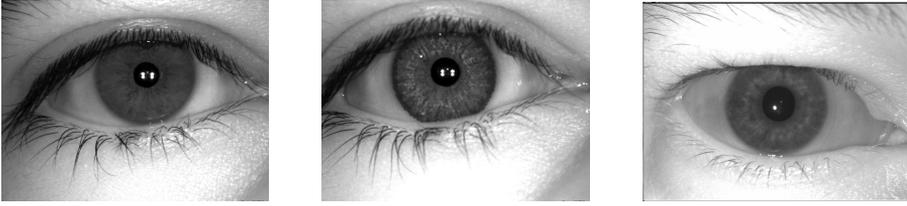


**Figure 4.12:** Feature extraction procedure. For each patch, four directional derivatives are computed on each point of the log-polar sampling grid, then four 2d Fourier transforms are computed, whose absolute values represent the local feature vector; this is vector quantized, and the histogram of the quantization indexes over the whole image forms the final feature vector.

the NotreDame datasets are divided explicitly in training and test sets, while the other do not present such separation, in which case we use two-fold cross validation in the experiments.

We consider three classes: no-lens, colored-lens and transparent-lens (some examples are shown in Fig.4.13), however, for uniformity with [143], in the tables we use the symbols N, for no-lens, S, for soft (that is transparent) lens, and T, for textured (that is, colored) lens. For each experiment, we compute the confusion matrix with generic entry  $\Pr(i|j)$ , the probability to decide for class  $i$  given that the sample belongs to class  $j$ . However, for brevity, we report in the tables only the diagonal entries, that is, the probability of correct decision for each class, together with their average.

Tab.4.15 reports the results obtained with the proposed algorithm and various choices of the region where the SID feature is computed: whole image with no segmentation, iris with the ideal segmentation provided with the datasets (only for the ND datasets), iris with real segmentation, sclera with real segmentation, iris and sclera with real segmentation. The textured lenses are detected in all cases with very high accuracy, with very limited dependence on the region used for SID. In the other cases, instead, the performance varies significantly with the region, and the sclera seems to be extremely important for a correct detection. Nonetheless, the joint use of iris and sclera provides a further gain in performance, up to 3 percent points. Using only the iris, instead,



**Figure 4.13:** Examples of iris with no lense (left), cosmetic lens (middle) and transparent lens (right).

leads to much worse performance. Notice that the ideal and real segmentation of the iris lead to very similar results, confirming the quality of our segmentation algorithm.

In Tab.4.16 we show again the performance of the proposed algorithm in its best configuration (iris+sclera) in comparison with the state-of-the-art methods proposed in [143], using plain LBP descriptor [101] and two *ad hoc* variations of it, LBP+PHOG (Pyramid of Histograms of Orientation Gradients) [7] and mLBP, which is the multiscale version of LBP. The proposed method guarantees always the best average accuracy, with a gain with respect to the best competitor which exceeds 10 percent points on the average.

## 4.5 Face

### Datasets description

Research on face liveness detection is definitely less mature than in the two preceding fields. Among the available databases we used the 3DMAD dataset<sup>11</sup> [31], based on wearable 3d masks, and the Replay-Attack dataset<sup>12</sup> [16], that considers printed and screen based fakes (made by mobile phones or tablets). Moreover, each attack subset can be sub-classified in two different groups based on the strategy used to hold the attack replay device: hand-based or fixed-support. Details on the datasets are provided in Tab.4.17, and examples of live and fake samples are shown in Fig.4.14. The Replay-Attack database comprises three separate subsets for training, development and testing, which we used following the protocol suggested in [16]. For 3DMAD, instead, we

<sup>11</sup><http://www.idiap.ch/dataset/3dmad>

<sup>12</sup><http://www.idiap.ch/dataset/replayattack>

Dataset	Class	Image	Ideal	Iris	Sclera	Iris + Sclera
NotreDame I	N-N	78.00	86.00	85.50	93.50	95.75
	T-T	100.00	99.75	100.00	99.00	99.75
	S-S	61.75	71.75	67.25	77.50	84.00
	Avg.	79.92	85.83	84.25	90.00	93.17
NotreDame II	N-N	71.00	66.00	60.00	84.00	79.00
	T-T	98.00	98.00	99.00	99.00	99.00
	S-S	53.00	49.00	63.00	71.00	78.00
	Avg.	74.00	71.00	74.00	84.67	85.33
Cogent	N-N	62.43	-	66.72	81.85	79.80
	T-T	93.13	-	95.63	87.06	95.54
	S-S	58.02	-	59.57	76.46	76.29
	Avg.	71.19		73.97	81.79	83.88
Vista	N-N	68.65	-	64.88	88.98	87.77
	T-T	95.29	-	99.19	96.19	98.59
	S-S	65.56	-	58.95	85.91	82.99
	Avg.	76.50		74.34	90.36	89.78

**Table 4.15:** Classification accuracy of the proposed algorithm as a function of the region where the feature vector is computed.

performed a leave-one-out per subject cross-validation as in [31]. Recall that the protocol for Replay-Attack requires to consider a small region cropped from the face (of dimension  $64 \times 64$  to  $80 \times 80$  pixels), and we used this procedure for 3DMAD as well, to avoid biases (note that otherwise almost all the descriptors give a perfect detection).

## Experimental results

Results are reported in Tab.4.18. In this case we do not report average values, which have a limited significance for just two datasets. However, CoA-LBP appears as the best descriptor, in this case, followed by SID which can be therefore considered a robust and discriminative descriptor for biometrics spoofing. It is also clear that the Replay-Attack database is more challenging, calling for the use of temporal information, as already done in [66] and in [108], where the average error reduces to 5.1 and 7.6, respectively. In Tab.4.19 we show also the results obtained by considering individually each type of attack (printed, and screen based

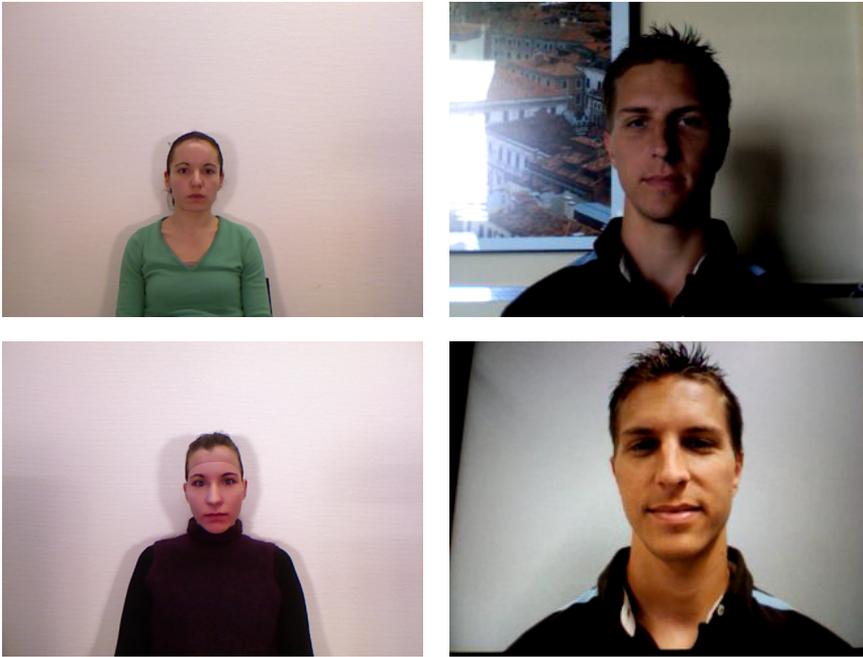
Dataset	Class	LBP	LBP+PHOG	mLBP	Proposed
NotreDame I	N-N	70.00	81.25	85.50	95.75
	T-T	97.00	96.25	96.50	99.75
	S-S	60.15	65.41	45.25	84.00
	Average	75.73	80.98	75.58	<b>93.17</b>
NotreDame II	N-N	42.00	42.00	81.00	79.00
	T-T	100.00	96.00	100.00	99.00
	S-S	54.00	60.00	52.00	78.00
	Average	65.33	66.00	77.67	<b>85.33</b>
IIITD Cogent	N-N	65.53	59.73	66.83	79.80
	T-T	89.39	91.87	94.91	95.54
	S-S	42.73	52.84	56.66	76.29
	Average	66.40	68.57	73.01	<b>83.88</b>
IIITD Vista	N-N	53.37	49.49	76.21	87.77
	T-T	98.64	99.42	91.62	98.59
	S-S	50.90	59.32	67.52	82.99
	Average	68.04	69.84	80.04	<b>89.78</b>

**Table 4.16:** Classification accuracy of proposed and reference techniques on the available databases. Best results in bold.

DATASET	3D-MAD	Replay-Attack
Camera	Microsoft Kinect	Apple Macbook
Image size	480x640	240x320
Live Samples	1700	3787
Fake Samples	850	11869
Total subjects	17	50
Frames per video	10	15

**Table 4.17:** Characteristics of 3DMAD and Replay-Attack datasets.

on mobile and high-definition tablets). Besides CoA-LBP and SID, also basic LBP performs well on these more homogeneous data. The high-definition images captured with the tablets represent clearly the most difficult dataset to deal with.



**Figure 4.14:** Some examples of live (up) and fake (down) images coming from *3D-MAD* database (left) and *Replay-Attack* database (right).

## 4.6 Cells classification

This work deals with the design of a classification method for cells extracted from Indirect Immunofluorescence images. In particular, we propose to use a dense local descriptor invariant both to scale changes and to rotations in order to classify the six categories of staining patterns of the cells. The descriptor is able to give a compact and discriminative representation and combines a log-polar sampling with spatially-varying gaussian smoothing applied on the gradients images in specific directions. Bag of Words is finally used to perform classification and experimental results show very good performance.

### Introduction

Indirect Immunofluorescence (IIF) images are generated by the interaction of biological tissue with special sources of light in order to generate

%	3DMAD	Replay-Attack
IQA-based [38]		15.2 (7)
LBP	1.4 (8)	12.8 (4)
CoA-LBP	<b>*0.0</b> (1)	<b>*9.4</b> (1)
Ric-LBP	*0.2 (4)	14.7 (6)
WLD	5.2 (11)	17.5 (10)
LPQ	4.7 (10)	21.7 (11)
BSIF	<b>*0.0</b> (1)	12.6 (3)
LCPD	2.1 (9)	14.0 (5)
Keypoint SIFT	0.3 (5)	25.2 (12)
Dense SIFT	<b>*0.0</b> (1)	17.0 (8)
DAISY	0.4 (6)	17.2 (9)
SID	1.4 (7)	*10.5 (2)

**Table 4.18:** Performance comparison for face images. The best result is reported in bold. \* denotes a statistically significant difference ( $p < 0.05$ ) with respect to the runner-up.

a fluorescent image responses. It is commonly used to diagnose autoimmune diseases by identifying specific patterns created by Anti-Nuclear Antibodies (ANAs) in the patient serum. Due to its effectiveness in the recent years the demand of diagnostic tests for systemic autoimmune diseases has rapidly increased. Currently the classification of the staining patterns is based on visual inspection by physician, however it is time-consuming and highly dependent on the analyst experience. It would be desirable to automate the process, but a comprehensive computer-aided diagnosis system for IIF is not yet available [34].

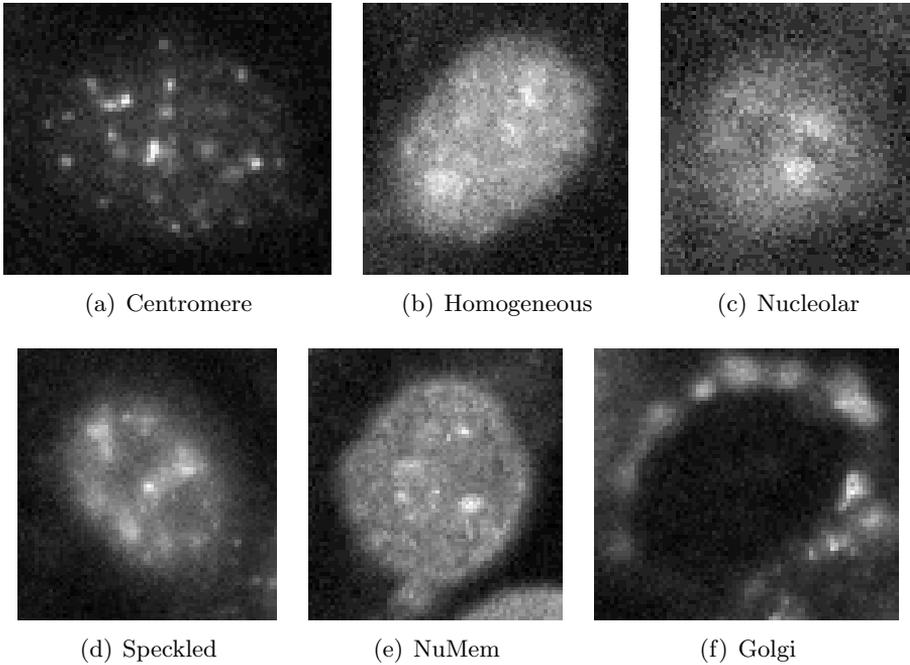
The recent contest on cells classification has given a strong impulse in the development of effective and specific algorithms for the recognition task of Human Epithelial type 2 (HEp-2) cell patterns [34, 33]. In particular, several works rely on the use of local descriptors, like the well-known local binary pattern (LBP) [102] (or its variants) and the SIFT descriptor [78]. This is the case for example of [123] where these descriptors are both used and combined with other ad hoc ones, or of [128] where the Co-occurrence of Adjacent LBP (CoALBP) has been considered in combination with SIFT. It is worth noting that CoALBP

%	Replay-Attack		
	Printed	Mobile	HighDef
IQA-based [38]	7.9	3.2	<b>*12.1</b>
LBP	<b>*2.8</b>	<b>*0.6</b>	*13.4
CoA-LBP	*3.5	1.6	*13.9
Ric-LBP	4.3	2.5	16.3
WLD	5.5	4.6	20.2
LPQ	8.5	9.3	23.1
BSIF	*2.9	1.6	17.2
LCPD	4.3	4.5	16.0
Keypoint SIFT	14.6	23.9	26.8
Dense SIFT	*3.9	4.9	18.9
DAISY	5.9	15.1	33.2
SID	*3.5	5.4	*14.1

**Table 4.19:** Results on Replay-Attack database considering individually the different types of attack. The best result is reported in bold. \* denotes a statistically significant difference ( $p < 0.05$ ) with respect to the runner-up.

was used by the top competitor of the contest [34]. Indeed, this approach was improved in [98] where a rotation invariant co-occurrence among adjacent LBPs (Ric-LBP) is proposed. In this way it is possible to deal with local image rotations. Indeed, constructing a rotation invariant descriptor seems to be desirable, as also noted in [119], where a rotationally invariant feature is proposed.

In this paper we follow this same path by resorting to a dense scale and rotation invariant descriptor [63]. This descriptor shares some common characteristics with the DAISY descriptor [130], like the use of an irregular grid and convolutions of the gradients in specific directions with spatially varying filters. The main difference lies essentially in the use of the Fourier transform amplitude to obtain a scale and rotation invariant descriptor. This approach was already used by the same Authors in [64], where the SID descriptor was proposed. It is worth nothing that there is an interesting similarity between the construction of the descriptor proposed in [63, 130] and the biological vision. In fact, as



**Figure 4.15:** Examples different staining patterns.

stated in [74] the scale-space representation, which is built by convolution with a family of Gaussian kernels and derivatives of increasing width, closely resemble receptive field profiles registered in neurophysiological studies of the mammalian retina and visual cortex. Note that, even if this descriptor is not specifically designed for the problem of IIF image classification, performance measured in terms of the mean-class accuracy are very good.

In the following section we will describe the dense local descriptor while in section 3 we will show the experimental results and make some comparison with recent state-of-the-art approaches. Finally, section 4 draws conclusions.

## Experimental results

Bag of Words model (BoW) has shown its superiority over many conventional global features for image classification, hence we used it in this context.

For the descriptor used in this work we set  $\alpha = 0.14$ ,  $N = 16$  rays,  $K = 10$  rings and  $D = 4$ . This setting turns out to be a good compromise between performance and computational complexity (note that the feature extraction can be computed very efficiently [130]). In this way we obtain eventually a local descriptor of length 560. We computed the euclidean distance between this feature vector (truly we extracted the feature vectors inside the segmentation map) and the 600 codewords of the dictionary to build the histogram that will be the input of the SVM linear classifier.

The proposed approach has been tested on the dataset provided by the ICPR 2014 contest<sup>13</sup> both for task1 and for task2. Task 1 is devoted to cell level classification. In particular, it consists of 13596 images for the training phase coming from 419 patient positive sera and categorized by trained physicians into six patterns: centromere, homogeneous, nucleolar, speckled (coarse and fine), nuclear membrane and golgi. An example of staining pattern for each class is shown in Fig.4.15. Task 2, instead, considers the problem of specimen level classification. It was collected from 1001 patient sera with positive ANA test and each specimen image belongs to one of the following pattern classes: homogeneous, speckled, nucleolar, centromere, golgi, nuclear membrane and mitotic spindle. All experimental results are reported in terms of Mean Class Accuracy (MCA):

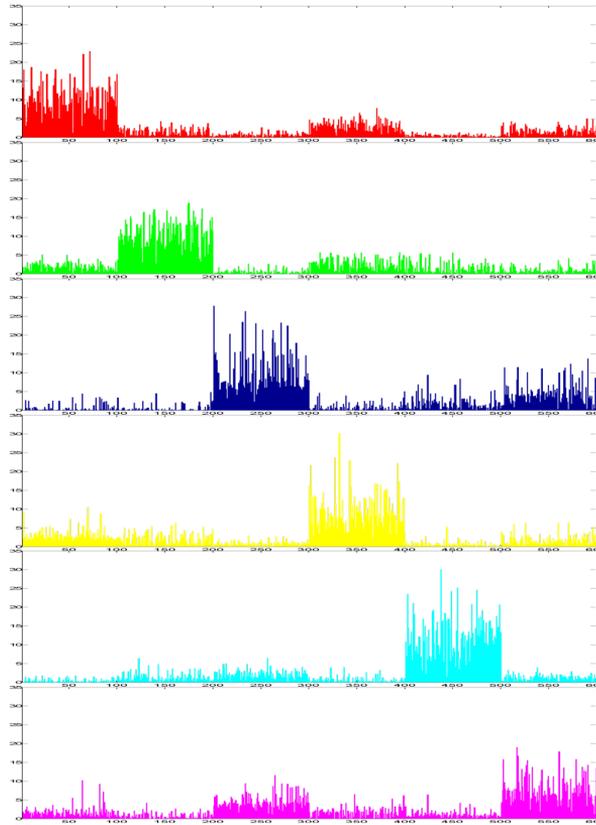
$$\text{MCA} = \frac{1}{K} \sum_{k=1}^K \text{CCR}_k = \frac{1}{K} \sum_{k=1}^K \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}, \quad (4.5)$$

where  $K = 6$  is the number of classes,  $\text{CCR}_k$  is the correct classification rate for the class  $k$  that is defined as the ratio between True Positive samples of the class  $k$  ( $\text{TP}_k$ ) and the total amount of samples of that class computed here as the sum of True Positive and False Negative sample ( $\text{TP}_k + \text{FN}_k$ ). In Fig.4.16 the average features vectors (i.e. the histograms of BoW) for each staining pattern are shown, proving the good discriminant power of the proposed features.

In Table 4.21 we report the classification results for phase 1 obtained with the technique based on the rotation and scale invariant descriptor on the training set by performing a leave-one-out per subject cross-validation procedure. The confusion matrixes for phase 1 and 2

---

<sup>13</sup><http://i3a2014.unisa.it/>



**Figure 4.16:** Examples of Bag of Words average histogram per cells staining patterns (from up to down): centromere, golgi, homogeneous, nucleolar, nuclear membrane, speckled. The BoW is composed using 100 centroid for each class. The average histograms show a good discriminant power of the features, indeed for each class it is possible to note an high value of the histograms for those bins corresponding to the same class and low values for the others. However we note a partial overlap between classes Homogeneous (blue) and Speckled (magenta).

are shown respectively in Tab.4.20 and Tab.4.22 and shows a very high correct classification rate (CCR) for all classes. Only for the Golgi cells, belonging to the less represented staining pattern, there a significant error, nearly 35%, mostly due to wrong classification as Nucleolar cells. A certain amount of misclassification errors involve class Homogeneous and

%	Centr.	Homog.	Nucleolar	Speckled	Nuc. memb.	Golgi
Centr.	<b>82.96</b>	0.11	3.21	13.46	0.04	0.22
Homog.	0.04	<b>82.04</b>	1.40	14.39	1.80	0.32
Nucleolar	2.08	1.31	<b>92.42</b>	2.04	0.81	1.35
Speckled	12.33	12.36	3.60	<b>70.79</b>	0.60	0.32
Nuc. memb.	0.00	4.62	0.77	0.59	<b>92.93</b>	1.09
Golgi	0.41	4.97	20.72	3.45	4.97	<b>65.47</b>

**Table 4.20:** Confusion matrix for the proposed algorithm (cell level classification). The final result is the recall mean value and is equal to 81.10%.

%	Centr.	Homog.	Nucleolar	Speckled	Nuc. memb.	Golgi	avg
[102] LBP	73.73	48.32	49.58	37.69	59.78	1.93	45.17
[102] LBPri-MR	76.80	61.03	64.67	51.18	74.50	14.09	57.04
[98] RiCLBP	80.04	70.41	70.21	65.77	76.90	26.10	64.91
[78] SIFT	84.82	79.47	84.57	72.55	88.95	41.16	75.25
[130] DAISY	76.43	72.09	77.29	64.11	81.75	54.42	71.02
Proposed w/o invariance	84.13	76.02	88.99	68.24	88.32	62.98	78.11
Proposed	82.96	82.04	92.42	70.79	92.93	65.47	81.10

**Table 4.21:** Comparison of the proposed approach with other state-of-the-art descriptors.

Speckled, accordingly to the average features vectors shown in Fig.4.16. The reader can refer to [77] for a detailed performance analysis.

However, looking at the results provided by a number of other descriptors, reported in table 4.21, it is clear that recognizing the Golgi cells is an intrinsically difficult task, and the proposed descriptor works indeed much better than all the competitors gaining about 10% points over DAISY and more than 20% over SIFT. Also for several other classes of cells (e.g. Nucleolar, Nuclear membrane) a significant improvement is observed, leading to an average performance gain going from 6% over SIFT to about 35% over LBP. Note that also for SIFT (computed densely) and DAISY we use exactly the same strategy of classification by means of BoW, while for LBP and its variants a linear SVM.

%	Centr.	Homog.	Nucleolar	Speckled	Nuc. mem.	Golgi	Mit. sp.
Centr.	<b>98.04</b>	0.00	0.00	1.96	0.00	0.00	0.00
Homog.	0.00	<b>83.02</b>	0.00	9.43	0.00	0.00	7.55
Nucleolar	0.00	2.00	<b>96.00</b>	0.00	0.00	0.00	2.00
Speckled	1.92	9.62	0.00	<b>86.54</b>	0.00	0.00	1.92
Nuc. mem.	0.00	4.76	0.00	0.00	<b>90.48</b>	0.00	4.76
Golgi	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	0.00
Mit. sp.	0.00	26.67	0.00	6.67	13.33	0.00	<b>53.33</b>

**Table 4.22:** Confusion matrix for the proposed algorithm (specimen level classification). The final result is the recall mean value and is equal to 86.77%.

To study the importance of scale and rotation invariance, the table shows also the results obtained avoiding the last step which gives the invariance property. Results are very good also for this solution, but scale and rotation invariance properties guarantee an extra gain of 3% point. Note that also the rotation invariant and multiresolution version of LBP (LBPri-MR) [102] has a significant gain over the basic solution, and RiCLBP [98] further improves.

In addition, for the proposed descriptor we considered a different BoW cardinality for each class, and in particular 200 for Centromere and Nuclear membrane, 100 for Nucleolar and Speckled, and 50 for Golgi and Homogeneous. Since soft-assignment can improve performance, as shown in [135], we followed this path by considering gaussian weights. Using the described approach we were able to obtain a mean class accuracy of 81.76 % on the training set.

Finally, in Table 4.22 we report the results obtained for specimen level classification. Although they are very promising (MCA is equal to 86.77 %), we have not yet optimized the method for this second task, and we expect to obtain further improvements.

## 4.7 Complexity

In Tab.4.23 we report the average CPU-time observed on a desktop computer for the feature extraction and coding phases of the various descriptors considered. These numbers must be taken with more than a

descriptor	size	feature extraction time	coding time	implementation
LBP	54	0.001	0.39	Matlab <sup>a</sup>
CoA-LBP	3072	0.041	0.14	Matlab <sup>b</sup>
Ric-LBP	408	0.116	0.26	Matlab <sup>b</sup>
WLD	960	1.502	1.31	Matlab <sup>a</sup>
LPQ	256	1.507	0.03	Matlab <sup>a</sup>
BSIF	4096	0.027	0.02	Matlab <sup>c</sup>
LCPD	2304	1.621	2.19	Matlab <sup>d</sup>
Keypoint SIFT	600	0.735	0.07	C <sup>e</sup>
Dense SIFT	600	0.794	0.72	C <sup>e</sup>
DAISY	600	0.984	1.16	[Matlab / C] <sup>f</sup>
SID	600	0.918	3.26	[Matlab / C] <sup>g</sup>

<sup>a</sup><http://www.cse.oulu.fi>

<sup>b</sup><http://www.cvlab.cs.tsukuba.ac.jp/~nosaka>

<sup>c</sup><http://www.ee.oulu.fi/~jkannala/bsif/bsif.html>

<sup>d</sup><http://www.grip.unina.it>

<sup>e</sup>VLFeat toolbox, version 0.9.18 [138]

<sup>f</sup><http://cvlab.epfl.ch/software/daisy>

<sup>g</sup><http://vision.mas.ecp.fr/Personnel/iasonas/descriptors.html>

**Table 4.23:** Average CPU-times (in seconds) for  $480 \times 640$ -pixel images on a 3.40 GHz 64 bit desktop computer with 8 GB memory.

grain of salt, because they refer to implementations provided by different programmers with different programming languages and different care for efficiency. For example, SIFT is well-known to be computation intensive, but the implementation provided in [138] is efficient enough to put it on par with much simpler descriptors. Nonetheless, knowing that it can run in dense modality in less than 2 seconds is valuable information. In any case, even the slower descriptors like SID and LCPD, run in less than 4 seconds, while simpler ones, like LBP, take much less, down to the light-speed 50-ms BSIF, giving solid guarantees on their applicability in low-power context, such as mobile phone authentication.

A deeper analysis has been conducted for the approaches proposed for implementation on mobile devices. In Tab.4.24 we report, for each LBP descriptor, the length of the corresponding feature vector and the

LBP desc.	length	sums	mul.s	tests	CPU-time
o-basic	256				
o-ri	36	20	16	8	0.2464
o-u2	59				
o-riu2	10				
□-basic	256				
□-ri	36	8	0	8	0.1002
□-u2	59				
□-riu2	10				
+ -basic	16				
+ -ri	6	4	0	4	0.0548
+ -u2	15				
+ -riu2	6				
MR-riu2	54	156	144	48	1.8979

**Table 4.24:** Complexity and feature length for LBP-based descriptors. CPU times are evaluated in  $\mu\text{s}/\text{pixel}$ .

theoretical complexity in terms of number of operations per pixel. In the computation of complexity, we count 4 multiplications for each interpolated sample. With some approximations, these could be replaced by shifts, but the square configuration can be seen itself as an approximation of the circular one, so we just neglect this option. Using look-up tables, no further operation is needed to compute rotation-invariant and uniform features. The table shows clearly that only MR-LBP requires a significant computational load, while LBP is already relatively simple with the circular configuration, and much simpler with square and cross configurations. These data are confirmed by the average CPU-times observed on a 2.20 GHz 64 bit desktop computer with 6 GB memory.

# Conclusion

Biometric authentication systems are quite vulnerable to sophisticated spoofing attacks. To keep a good level of security, reliable spoofing detection tools are necessary, preferably implemented as software modules. The research in this field is very active, with *local descriptors*, based on the analysis of micro-textural features, gaining more and more popularity, thanks to their excellent performance and flexibility.

In this thesis work, after carrying out a review of state of the art regarding image descriptors applied in different contexts, we have proposed new local descriptors and LD-based techniques to address the liveness detection task for fingerprint, iris, and face images. We have assessed the potential of these descriptors comparing their performances on public available datasets with a number of recently proposed techniques. The analysis shed some light on the relative pros and cons of the various proposed solutions, pointing out the most interesting lines of development, and laying a solid ground for further investigations. In this analysis, special attention was devoted to generality, the robustness with respect to uses in different domains, which is a very appealing property in liveness detection [38] considering the large array of biometric traits that have been proposed for authentication purposes. To this aims, the proposed approach based on SID and the Bag of Words paradigm has proven especially valuable. Indeed, it has been used with equally good results also in a very different context, concerning cell image classification.

The main lesson learnt through the experimental analysis is that they work very well. For example, in case of fingerprint, *all* tested descriptors improve w.r.t. the winner of the LivDet competition of just 4 years ago, and some reduce the average error by as much as 75%.

Future research in this field must necessarily start from this point. On the other hand, we could not single out a descriptor performing uni-

formly better than the others. Although SID looks very promising and stable over all case studies, the fingerprint-specific LCPD performs indeed better on fingerprints, showing that some space remains for clever design. A significant example is the proposed technique for the classification of contact lens, for which information coming from the sclera allowed to distinguish between transparent-lens and no-lens access, outperforming the state of the art. Likewise, for application on mobile devices, where complexity is a key issue, variants of the basic LBP worked best. Of course, new descriptors appear by the day in the literature, some of which look especially well-suited to the liveness detection task like, for example, those proposed in [121] and in [133].

However, although designing better and better descriptors is certainly of interest, it is very likely that more significant improvements come from a sensible decision-level fusion of existing ones. This will be a main topic of our future research. Another issue deserving deeper investigation is robustness to imprecise training. All descriptor-based classifiers, in fact, require training on a large set of images describing the source. While learning the sensor characteristics may make sense, overfitting to certain types of spoofing is very dangerous, as new attacks may go totally undetected. An interesting approach tackling this problem has been recently proposed in [112], where the detector automatically adapts to spoofs fabricated using novel materials. This work and others following this same path can be precious for scientists and engineers alike.

# Bibliography

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [2] A. Antonelli, R. Cappelli, D. Maio, and D. Maltoni, “Fake finger detection by skin distortion analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 3, pp. 360–373, 2006.
- [3] S. E. Baker, A. Hentz, K. W. Bowyer, and P. J. Flynn, “Degradation of iris recognition performance due to non-cosmetic prescription contact lenses,” *Computer Vision and Image Understanding*, vol. 114, no. 9, pp. 1030–1044, 2010.
- [4] D. Baldissera, A. Franco, D. Maio, and D. Maltoni, “Fake fingerprint detection by odor analysis,” in *International Conference on Biometrics*, vol. 3832, january 2006, pp. 265–272.
- [5] D. H. Ballard, “Generalizing the hough transform to detect arbitrary shapes,” *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.
- [6] T. Bianchi and A. Piva, “Image forgery localization via block-grained analysis of jpeg artifacts,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, Jun. 2012.
- [7] A. Bosch, A. Zisserman, and X. Munoz, “Representing shape with a spatial pyramid kernel,” in *ACM international conference on Image and video retrieval*, 2007, pp. 401–408.

- 
- [8] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” *Computer vision and image understanding*, vol. 110, no. 2, pp. 281–307, 2008.
- [9] J. Canny, “A computational approach to edge detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986.
- [10] O. Çeliktutan, B. Sankur, and I. Avcibas, “Blind Identification of Source Cell-Phone Model,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 553–566, 2008.
- [11] M. Chakka *et al.*, “Competition on counter measures to 2-d facial spoofing attacks,” in *IEEE International Joint Conference on Biometrics*, 2011, pp. 1–6.
- [12] C. Chan, M. Tahir, J. Kittler, and M. Pietikäinen, “Multiscale Local Phase Quantization for Robust Component-Based Face Recognition Using Kernel Fusion of Multiple Descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1164–1177, 2013.
- [13] K. Chatfield, V. Lempitsky, A. Vealdi, and A. Zisserman, “The devil is in the details: an evaluation of recent feature encoding methods,” in *British Machine Vision Conference*, 2011.
- [14] J. Chen *et al.*, “WLD: a robust local image descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, July 2010.
- [15] G. Chiachia, “<https://github.com/giovanichiachia/convnet-rfw>,” 2014.
- [16] I. Chingovska, A. Anjos, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in *International Conference of the Biometrics Special Interest Group*, 2012, pp. 1–7.
- [17] P. Coli, G. Marcialis, and F. Roli, “Power spectrum-based fingerprint vitality detection,” in *Proc. of IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 169–173.

- 
- [18] D. Cozzolino, D. Gagnaniello, and L. Verdoliva, "Image forgery detection through residual-based local descriptors and block-matching," in *IEEE International Conference on Image Processing (ICIP)*, October 2014, pp. 5232–5236.
- [19] D. Cozzolino, D. Gagnaniello, and L. Verdoliva, "Image forgery localization through the fusion of camera-based, feature-based and pixel-based techniques," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 5302–5306.
- [20] A. Czajka, "Database of iris printouts and its application: development of liveness detection method for iris recognition," in *Proc. of the 18th International Conference on Methods and Models in Automation and Control*, 2013.
- [21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–95, 2007.
- [22] J. Daugman, "Recognizing persons by their iris patterns: Countermeasures against subterfuge," in *Biometrics. Personal Identification in a Networked Society*, J. et al., Ed., 1999, pp. 103–121.
- [23] J. Daugman, "Statistical richness of visual phase information: update on recognizing persons by iris patterns," *International Journal of Computer Vision*, vol. 45, no. 1, pp. 25–38, 2001.
- [24] J. Daugman, "The importance of being random: Statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279–291, 2003.
- [25] J. Daugman, "New methods in Iris Recognition," *IEEE Transactions on Systems, Man and Cybernetics-Part B*, vol. 37, no. 5, pp. 1167–1175, october 2007.
- [26] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.

- [27] R. Davarzani, K. Yaghmaie, S. Mozaffari, and M. Tapak, "Copy-move forgery detection using multiresolution local binary patterns," *Forensic Science International*, vol. 231, pp. 61–72, 2013.
- [28] M. De Marsico, C. Galdi, M. Nappi, and D. Riccio, "FIRME: Face and Iris Recognition for Mobile Engagement," *Image and Vision Computing*, vol. in press, 2014.
- [29] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, "Exploiting patch similarity for SAR image processing: the non-local paradigm," *IEEE Signal Processing Magazine*, july 2014.
- [30] R. Derakhshani, S. Schuckers, L. Hornak, and L. Gorman, "Determination of vitality from a non-invasive biomedical measurement for use in fingerprint scanners," *Pattern Recognition*, vol. 36, no. 2, pp. 383–396, 2003.
- [31] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084–1097, july 2014.
- [32] H. Farid and L. Siwei, "Detecting hidden messages using higher order statistics and support vector machines," in *5th International Workshop on Information Hiding*, 2002, pp. 340–354.
- [33] P. Foggia, G. Percannella, A. Saggese, and M. Vento, "Pattern recognition in stained hep-2 cells: Where are we now?" *Pattern Recognition*, vol. 47, no. 7, pp. 2305–2314, July 2014.
- [34] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking hep-2 cells classification methods," *IEEE Transactions on Medical Imaging*, vol. 32, no. 10, pp. 1878–1889, October 2013.
- [35] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics*, vol. 7, no. 3, pp. 868–882, June 2012.
- [36] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 311–321, 2012.

- 
- [37] J. Galbally *et al.*, “An evaluation of direct attacks using fake fingers generated from ISO templates,” *Pattern Recognition Letters*, vol. 31, pp. 725–732, 2010.
- [38] J. Galbally, S. Marcel, and J. Fierrez, “Image quality assessment for fake biometric detection: application to iris, fingerprint and face recognition,” *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, February 2014.
- [39] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia, “Iris liveness detection based on quality related features,” in *Proc. of 5th IAPR International Conference on Biometrics (ICB)*, April 2012, pp. 271–276.
- [40] L. Ghiani, A. Hadid, G. Marcialis, and F. Roli, “Fingerprint liveness detection using binarized statistical image features,” in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2013.
- [41] L. Ghiani, G. Marcialis, and F. Roli, “Fingerprint liveness detection by local phase quantization,” in *International Conference on Pattern Recognition*, 2012, pp. 537–540.
- [42] C. Gottschlich, E. Marasco, A. Yang, and B. Cukic, “Fingerprint liveness detection based on histograms of invariant gradients,” in *International Joint Conference on Biometrics*, 2014, pp. 1–7.
- [43] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “Fingerprint liveness detection based on weber local image descriptor,” in *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, 2013, pp. 46–50.
- [44] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “Contact lens detection and classification in iris images through scale invariant descriptor,” in *Workshop on Insight on Eye Biometrics*, november 2014, pp. 560–565.
- [45] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “A wavelet-markov local descriptor for detecting fake fingerprints,” *Electronics Letters*, vol. 50, no. 6, pp. 439–441, March 2014.

- [46] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 849–863, April 2015.
- [47] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "Local contrast phase descriptor for fingerprint liveness detection," *Pattern Recognition*, vol. 48, no. 4, pp. 1050–1058, april 2015.
- [48] D. Gragnaniello, C. Sansone, and L. Verdoliva, "Iris liveness detection for mobile devices based on local descriptors," *Pattern Recognition Letters*, in press 2015.
- [49] D. Gragnaniello, C. Chaux, J.-C. Pesquet, and L. Duval, "A convex variational approach for multiple removal in seismic data," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 215–219.
- [50] D. Gragnaniello, G. Poggi, and L. Verdoliva, "Classification-based nonlocal sar despeckling," in *Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS)*, 2012, pp. 121–125.
- [51] D. Gragnaniello, C. Sansone, and L. Verdoliva, "Biologically-inspired dense local descriptor for indirect immunofluorescence image classification," in *1st Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis (I3A)*, 2014, pp. 1–5.
- [52] X. He, S. An, and P. Shi, "Statistical textural analysis based approach for fake iris detection using support vector machine," in *International Conference on Biometrics*, 2007, pp. 540–546.
- [53] X. He, Y. Lu, and P. Shi, "A new fake iris detection method," in *in proc. of the Third International Conference on Advances in Biometrics*, 2009, pp. 1132–1139.
- [54] Z. He, Z. Sun, T. Tan, and Z. Wei, "Efficient iris spoof detection via boosted local binary patterns," in *Advances in Biometrics*, vol. 5558, 2009, pp. 1080–1090.
- [55] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on markov features in dct and dwt domain," *Pattern Recognition*, vol. 45, no. 12, pp. 4292–4299, 2012.

- 
- [56] X. Huang, C. Ti, Q. Hou, A. Tokuta, and R. Yang, "An experimental study of pupil constriction for liveness detection," in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2013, pp. 252–258.
- [57] H. Jee, S. Jung, and J. Yoo, "Liveness detection for embedded face recognition system," in *Proc. of World Academy of Science, Engineering and Technology*, vol. 18, 2006.
- [58] X. Jia *et al.*, "Multi-scale local binary pattern with filters for spoof fingerprint detection," *Information Sciences*, vol. 268, pp. 91–102, june 2014.
- [59] C. Jin, H. Kim, and S. Elliott, "Liveness detection of fingerprint based on band-selective fourier spectrum," in *Proc. Int. Conf. on Information Security and Cryptology (ICISC)*, vol. 4817, 2007, pp. 168–179.
- [60] J. Kannala and E. Rahtu, "BSIF: binarized statistical image features," in *International Conference on Pattern Recognition*, 2012, pp. 1363–1366.
- [61] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 681 905–681 905.
- [62] Y. Kim, J.-H. Yoo, and K. Choi, "A motion and similarity-based fake detection method for biometric face recognition systems," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 2, pp. 756–762, 2011.
- [63] I. Kokkinos, M. Bronstein, and A. Yuille, "Dense scale invariant descriptors for images and surface," INRIA, Research Report RR-7914, 2012.
- [64] I. Kokkinos and A. Yuille, "Scale invariance without scale selection," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.
- [65] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *IEEE Workshop on Automatic Identification Advanced Technologies*, 2005, pp. 75–80.

- [66] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, “Complementary countermeasures for detecting scenic face spoofing attacks,” in *International Conference on Biometrics*, 2013, pp. 1–7.
- [67] N. Kose and J.-L. Dugelay, “On the vulnerability of face recognition systems to spoofing mask attacks,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 2357 – 2361.
- [68] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [69] E. Lee and K. Park, “Fake iris detection based on 3d structure of iris pattern,” *International Journal of Imaging Systems and Technology*, vol. 20, no. 2, pp. 162–166, 2010.
- [70] Z. Lei, S. Liao, M. Pietikäinen, and S. Li, “Face Recognition by Exploring Information Jointly in Space, Scale and Orientation,” *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 247–256, 2011.
- [71] T. Leung and J. Malik, “Representing and recognizing the visual appearance of materials using three-dimensional textons,” *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.
- [72] J. Li, Y. Wang, T. Tan, and A. Jain, “Live face detection based on the analysis of fourier spectra,” in *SPIE*, vol. 5404, 2004, pp. 296 – 303.
- [73] S. Liao, M. W. K. Law, and A. C. S. Chung, “Dominant Local Binary Patterns for Texture Classification,” *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 1107–1118, 2009.
- [74] T. Lindeberg, “Scale-space theory: A basic tool for analysing structures at different scales,” *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224–270, 1994.

- 
- [75] T. Lindeberg and L. Florack, “Foveal scale-space and the linear increase of receptive field size as a function of eccentricity,” *Computer Vision and Image Understanding*, 1996.
- [76] F. Liu, Z. Tang, and J. Tang, “WLBP: Weber local binary pattern for local image description,” *Neurocomputing*, vol. 120, pp. 325–335, 2013.
- [77] B. C. Lovell, G. Percannella, M. Vento, and A. Willem, “Performance evaluation of indirect immunofluorescence image analysis systems,” UNISA, Research Report, 2014. [Online]. Available: <http://i3a2014.unisa.it/>
- [78] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [79] L. Ma, T. Tan, Y. Wang, and D. Zhang, “Efficient Iris Recognition by Characterizing Key Local Variations,” *IEEE Transactions on Image Processing*, vol. 13, no. 6, pp. 739–750, june 2004.
- [80] J. Määttä, A. Hadid, and M. Pietikäinen, “Face spoofing detection from single images using micro-texture analysis,” in *IEEE International Joint Conference on Biometrics*, 2011, pp. 1–7.
- [81] T. Mäenpää, “The local binary pattern approach to texture analysis - extensions and applications.” Ph.D. dissertation, 2003, dissertation. Acta Univ Oul C 187, 78 p + App.
- [82] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. New York, NY, USA: Springer-Verlag, 2009.
- [83] E. Marasco and A. Ross, “A survey on anti-spoofing schemes for fingerprint recognition systems,” *ACM Computing Surveys*, vol. 47, no. 2, p. 28, 2014.
- [84] E. Marasco and C. Sansone, “Combining perspiration- and morphology-based static features for fingerprint liveness detection,” *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1148–1156, 2012.

- [85] G. Marcialis *et al.*, “First international fingerprint liveness detection competition - livdet 2009,” in *Lect. Notes Comput. Sci.*, 2009, pp. 12–23.
- [86] G. Marcialis, F. Roli, and A. Tidu, “Analysis of fingerprint pores for vitality detection,” in *International Conference on Pattern Recognition*, 2010, pp. 1289–1292.
- [87] T. Matsumoto, “Artificial irises: importance of vulnerability analysis,” in *2nd Asian Biometrics Workshop*, vol. 45, 2004.
- [88] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, “Impact of artificial gummy fingers on fingerprint systems,” in *Proc. of SPIE*, vol. 4677, 2002, pp. 275–289.
- [89] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, october 2005.
- [90] D. Minotti *et al.*, “Deep representations for iris, face, and fingerprint spoofing attack detection,” *arXiv preprint arXiv:1410.1980*, october 2014.
- [91] A. Mittal, A. Moorthy, and A. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [92] Y. Moon, J. Chen, K. Chan, K. So, and K. Woo, “Wavelet based fingerprint liveness detection,” *Electronic Letters*, vol. 41, no. 20, pp. 1112–1113, 2005.
- [93] S. Nikam and S. Agarwal, “Gabor filter-based fingerprint anti-spoofing,” in *LNCS 5259*, 2008, pp. 1103 – 1114.
- [94] S. Nikam and S. Agarwal, “Local binary pattern and wavelet-based spoof fingerprint detection,” *International Journal of Biometrics*, vol. 1, no. 2, pp. 141–159, 2008.
- [95] S. Nikam and S. Agarwal, “Texture and wavelet-based spoof fingerprint detection for fingerprint biometric systems,” in *First International Conference on Emerging Trends in Engineering and Technology*, 2008, pp. 675–680.

- 
- [96] S. Nikam and S. Agarwal, "Ridgelet-based fake fingerprint detection," *Neurocomputing*, vol. 72, pp. 2491–2506, 2009.
- [97] R. Nogueira, R. de Alencar Lotufo, and R. Machado, "Evaluating software-based fingerprint liveness detection using convolutional networks and local binary patterns," in *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, october 2014.
- [98] R. Nosaka and K. Fukui, "Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns," *Pattern Recognition*, vol. 47, no. 7, pp. 2428–2436, July 2014.
- [99] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," *5th Pacific-Rim Symposium on Image and Video Technology*, vol. 7088, pp. 82–91, 2011.
- [100] R. Nosaka, C. Suryanto, and K. Fukui, "Rotation invariant co-occurrence among adjacent lbps," in *International Workshop on Computer Vision With Local Binary Pattern Variants*, vol. 7728, 2012, pp. 15–25.
- [101] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [102] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, july 2002.
- [103] V. Ojansivu, E. Rahtu, and J. Heikkilä, "Rotation invariant blur insensitive texture analysis using local phase quantization," in *International Conference on Pattern Recognition*, 2008.
- [104] A. Oppenheim and J. Lim, "The Importance of Phase in Signals," *Proc. IEEE*, vol. 69, no. 5, pp. 529–541, 1981.
- [105] A. Pacut and A. Czajka, "Aliveness detection for iris biometrics," in *Proc. of the 40th Annual IEEE International Carnahan Conferences Security Technology*, 2006, pp. 122–129.

- [106] G. Pan, L. Sun, Z. Wu, and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic webcam,” in *International Conference on Computer Vision (ICCV)*, 2007, pp. 1080–1090.
- [107] S. Parthasaradhi, R. Derakhshani, L. Hornak, and S. Schuckers, “Time-series detection of perspiration as a liveness test in fingerprint devices,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applic. and Rev.*, vol. 35, no. 3, pp. 335–343, aug. 2005.
- [108] T. Pereira *et al.*, “Face liveness detection using dynamic texture,” *EURASIP Journal on Image and Video Processing*, pp. 1–15, 2014.
- [109] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215–224, june 2010.
- [110] J. Qian, J. Yang, and G. Gao, “Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction,” *Pattern Recognition*, vol. 46, pp. 2724–2729, 2013.
- [111] T. Randen and J. H. Husoy, “Filtering for Texture Classification: A Comparative Study,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.
- [112] A. Rattani and A. Ross, “Automatic adaptation of fingerprint liveness detector to new spoof materials,” in *International Joint Conference on Biometrics*, october 2014, pp. 1–8.
- [113] V. Ruiz-Albacete *et al.*, “Direct attacks using fake images in iris verification,” in *BIOID 2008, LNCS 5372*, B. S. *et al.*, Ed. Berlin Heidelberg: Springer-Verlag, 2008, pp. 181–190.
- [114] S. A. Sahmoud and I. S. Abuhaiba, “Efficient iris segmentation method in unconstrained environments,” *Pattern Recognition*, vol. 46, no. 12, pp. 3174–3185, 2013.
- [115] E. L. Schwartz, “Spatial mapping in the primate sensory projection: analytic structure and relevance to perception,” *Biological Cybernetics*, vol. 25, no. 4, pp. 181–194, 1977.

- 
- [116] A. Sequeira, J. Monteiro, A. Rebelo, and H. Oliveira, “Mobio: a multimodal database captured with a portable handheld device,” in *9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2014, pp. 133–139.
- [117] A. Sequeira, J. Murari, and J. Cardoso, “Iris liveness detection methods in mobile applications,” in *9th International Conference on Computer Vision Theory and Applications*, 2014, pp. 22–33.
- [118] L. Sharan, C. Liu, R. Rosenholtz, and E. Adelson, “Recognizing Materials Using Perceptually Inspired Features,” *International Journal of Computer Vision*, vol. 103, pp. 348–371, 2013.
- [119] L. Shen, J. Lin, S. Wu, and S. Yu, “Hep-2 image classification using intensity order pooling based features and bag of words,” *Pattern Recognition*, vol. 47, no. 7, pp. 2419–2427, July 2014.
- [120] L. Shen, J. Lin, S. Wu, and S. Yu, “Hep-2 image classification using intensity order pooling based features and bag of words,” *Pattern Recognition*, 2013.
- [121] K. Simonyan, A. Vedaldi, and A. Zisserman, “Learning local feature descriptors using convex optimisation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1573–1585, august 2014.
- [122] Y. Singh and S. Singh, “Vitality detection from biometrics: State-of-the-art,” in *Proc. of the World Congress on Information and Communication Technologies*, 2011, pp. 106–111.
- [123] R. Stoklasa, T. Majtner, and D. Svoboda, “Efficient k-nn based hep-2 cells classifier,” *Pattern Recognition*, vol. 47, no. 7, pp. 2409–2418, July 2014.
- [124] Z. Sun, H. Zhang, T. Tan, and J. Wang, “Iris image classification based on hierarchical visual codebook,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1120–1133, june 2014.
- [125] B. Tan and S. Schuckers, “Liveness detection for fingerprint scanners based on the statistics of wavelet signal processing,” in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2006.

- [126] B. Tan and S. Schuckers, “New approach for liveness detection in fingerprint scanners based on valley noise analysis,” *Journal of Electronic Imaging*, vol. 17, 2008.
- [127] B. Tan and S. Schuckers, “Spoofing protection for fingerprint scanner by fusing ridge signal and valley noise,” *Pattern Recognition*, vol. 43, pp. 2845–2857, 2010.
- [128] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, “Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space,” *Pattern Recognition*, vol. 47, no. 7, pp. 2367–2378, July 2014.
- [129] J. Thorsten, *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.
- [130] E. Tola, V. Lepetit, and P. Fua, “Daisy: An efficient dense descriptor applied to wide-baseline stereo,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, 2010.
- [131] P. Tome *et al.*, “The 1st competition on counter measures to finger vein spoofing attacks,” in *The 8th IAPR International Conference on Biometrics (ICB)*, May 2015.
- [132] P. Tome, M. Vanoni, and S. Marcel, “On the vulnerability of finger vein recognition to spoofing,” in *IEEE International Conference of the Biometrics Special Interest Group*, 2014.
- [133] T. Trzcinski and V. Lepetit, “Efficient discriminative projections for compact binary descriptors,” in *European Conference on Computer Vision*, vol. 7572, 2012, pp. 228–242.
- [134] J. Unar, W. Senga, and A. Abbasia, “A review of biometric technology along with trends and prospects,” *Pattern recognition*, vol. 47, no. 8, pp. 2673–2688, august 2014.
- [135] J. C. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. Smeulders, “Kernel codebooks for scene categorization,” in *Computer Vision—ECCV 2008*. Springer, 2008, pp. 696–709.

- 
- [136] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Computer Vision-ECCV 2002*. Springer, 2002, pp. 255–271.
- [137] M. Varma and A. Zisserman, "Texture classification: Are filter banks necessary?" in *Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on*, vol. 2, 2003, pp. II–691.
- [138] A. Vedaldi and B. Fulkerson, "Vlfeat - an open and portable library of computer vision algorithms," in *Proc. ACM Int. Conf. on Multimedia*, 2010.
- [139] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3360–3367.
- [140] Z. Wei, X. Qiu, Z. Sun, and T. Tan, "Counterfeit iris detection based on texture analysis," in *19th International Conference on Pattern Recognition*, 2008, pp. 1–4.
- [141] S. Winder and M. Brown, "Learning local image descriptors," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.
- [142] G. Xu and Y. Shi, "Camera model identification using local binary patterns," in *IEEE International Conference on Multimedia and Expo*, 2012, pp. 392–397.
- [143] D. Yadav *et al.*, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 851–862, may 2014.
- [144] D. Yambay *et al.*, "Livdet 2011 - fingerprint liveness detection competition 2011," in *IAPR/IEEE Int. Conf. on Biometrics*, 2012, pp. 208–215.
- [145] J. Yang, Z. Lei, and S. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408.5601*, august 2014.
- [146] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding." in *NIPS*, vol. 9, 2009, p. 1.

- [147] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, February 2010.
- [148] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, January 2007.
- [149] D. Zhang, W.-K. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1041–1050, 2003.
- [150] H. Zhang, Z. Sun, and T. Tan, "Contact lens detection based on weighted LBP," in *20th International Conference on Pattern Recognition*, 2010, pp. 4279–4282.
- [151] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," in *International Conference on Computer Vision*, 2005, pp. 786–791.
- [152] S. Zhao, Y. Gao, and B. Zhang, "Sobel lbp," in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 2144–2147.
- [153] S.-R. Zhou, J.-P. Yin, and J.-M. Zhang, "Local binary pattern (LBP) and local phase quantization (LPQ) based on Gabor filter for face representation," *Neurocomputing*, vol. 116, pp. 260–264, 2013.
- [154] L. Zhu, A. B. Rao, and A. Zhang, "Theory of keyblock-based image retrieval," *ACM Transactions on Information Systems (TOIS)*, vol. 20, no. 2, pp. 224–257, 2002.
- [155] D. Zou, Y. Shi, W. Su, and G. Xuan, "Steganalysis based on markov model of thresholded prediction-error image," in *International Conference on Multimedia and Expo*, 2006, pp. 1365–1368.