



UNIVERSIDADE DA BEIRA INTERIOR
Faculdade de Engenharia

***QUIS-CAMPI*: Biometric Recognition in Surveillance Environments**

Seminar



Hugo Pedro Martins Carriço Proença

In the scope of the application for the degree of *Agregado*, submitted at the University of Beira Interior, Covilhã, Portugal, in accordance to the Decree-Law n. 239/2007.

May, 2015

Preliminary Remark

This report was prepared in the scope of the application for the degree of *Agregado*, submitted at the University of Beira Interior, Covilhã, Portugal, in accordance to the Decree-Law n. 239/2007.

Contents

1	Introduction	7
1.1	Problem Statement	7
1.2	Major Contributions	8
1.3	<i>QUIS-CAMPI</i> Research Team	9
1.4	Document Structure and Notation	10
2	Biometric Recognition in Surveillance	13
2.1	State-of-the-Art	13
2.2	Recognition Chain	15
3	Low-level Vision Problems	19
3.1	Background Estimation	19
3.2	Human Detection	20
3.3	Pose Estimation	21
3.4	Tracking	22
4	High-level Vision Problems	25
4.1	Identities Pruning: Database Indexing / Retrieval	25
4.1.1	Indexing	25
4.1.2	Retrieval	27
4.2	Identities Pruning: Soft Biometrics	29
4.3	Identities Pruning: Gait Recognition	31
4.4	Iris Segmentation	32
4.4.1	Feature Extraction Stages	33
4.4.2	Sclera Stage	33
4.4.3	Iris Stage	34
4.4.4	Supervised Machine Learning and Classification	35
4.4.5	Shape Parameterization	36
4.4.6	Computational Complexity	38
4.5	Periocular Segmentation	38
4.5.1	Feature Extraction	40
4.5.2	Unary Potentials	41
4.5.3	Pairwise Potentials	42
4.6	Iris Recognition: Quality Assessment	43
4.6.1	Estimation of the Iris Center	44
4.6.2	Parameterization of the Biological Iris Boundaries	45
4.6.3	Focus Assessment	47
4.6.4	Motion Assessment	49
4.6.5	Off-Angle Assessment	50

4.6.6	Occlusions Assessment	51
4.6.7	Levels of Iris Pigmentation	52
4.7	Iris Recognition	53
4.7.1	Retinex	53
4.7.2	Noise-Free Iris Segmentation	54
4.7.3	Parameterization of Iris Boundaries	55
4.7.4	Partitioning the Iris into Regions	56
4.7.5	Color Descriptors	57
4.7.6	Order Statistics of Dominant Colors	59
4.7.7	Linear Assignment Problem	59
4.7.8	Histogram Matching	60
4.7.9	Shape Context Descriptor	60
4.7.10	Robustness to Data Variation Factors	61
4.8	Periocular Recognition	65
4.8.1	State-of-the-Art	65
4.8.2	Fusion of Iris and Periocular Recognizers	65
4.8.3	Strong Biometric Trait: Iris Texture	66
4.8.4	Iris Segmentation and Parameterisation	67
4.8.5	Preliminary Selection of Color Spaces	68
4.8.6	Feature Encoding	69
4.8.7	Learning Phase	70
4.8.8	Weak Biometric Trait: Eyelids, Eyelashes and Skin	71
4.8.9	Sclera Detection	71
4.8.10	Eyelids Parameterisation	72
4.8.11	Feature Encoding and Matching	72
4.8.12	Shape Descriptors	72
4.8.13	Texture Descriptors	73
4.9	Analysis of <i>Iriscodes</i> : Bit Discriminability	73
4.9.1	Study of Iris Codes	74
4.9.2	Bit Discriminability	75
4.9.3	Datasets and Preprocessing	77
4.9.4	Amount of Information in Iris Patches	79
4.9.5	Filters Parameterizations	81
4.9.6	Bit Discriminability	82
4.9.7	Discriminability vs. Fragility	84
4.9.8	Codes Quantization: How Much Discriminating Information Is Lost?	84
5	Ethics / Privacy in Non-Cooperative Biometric Recognition	87
6	Conclusions and Further Work	91
6.1	Concluding Remarks	91
6.2	Further Work	91

Chapter 1

Introduction

1.1 Problem Statement

Recent attacks in crowded urban environments reduced the perception of safety in modern societies, while the citizens' tolerance to *reasonable* risks has also been decreasing. Currently, there is a growing need of assure the safety of people, specially in places / events that concentrate large crowds, which are naturally perceived as those with the highest risk (due to e.g., 2001 New York 9/11, 2004 Madrid train bombing and 2013 Boston marathon attacks).

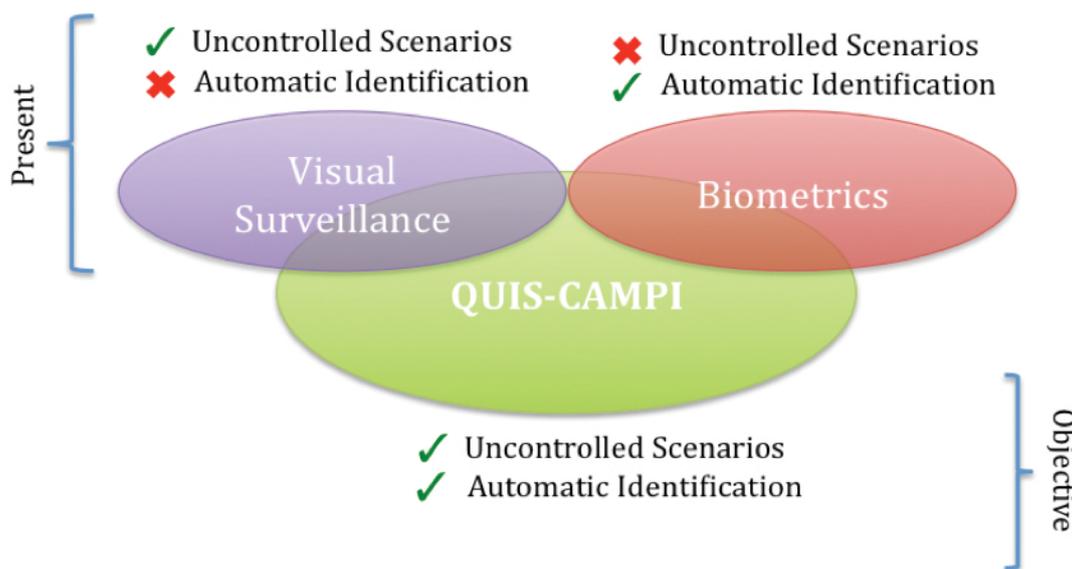


Figure 1.1: Main ambition in *QUIS-CAMPI* research.

To counterbalance this fear, visual surveillance is currently used worldwide. The deployment of surveillance cameras has grown astonishingly in the recent years, with more than 5.9 million CCTV cameras reported only in the United Kingdom [9]. However, contrary to popular belief, there are **still no fully automatic techniques to identify subjects without requiring their participation in data acquisition**, and the automated understanding of data is most times reduced to action recognition. For every identification attempt, it is required some kind of human intervention in the process. Even though national / international authorities have lists of potentially harmful individuals, it is particularly difficult for humans to confirm whether such elements are among a crowd. As an example, the *TIDE: Terrorist Identities Datamart Environment* from the U.S. National Counterterrorism Center has over 745,000

people listed in the database which authorities are willing to arrest, but only a small proportion of these was actually detected in visual surveillance systems.

Simultaneously, **biometrics is considered an especially successful case in the domain of pattern recognition**: biometric recognition systems have been deployed for many different applications (e.g., security assess or refugee control), but **performance is still strongly conditioned by the levels of cooperation demanded to subjects** and by the **environmental conditions** required to obtain data with minimal levels of quality.

Hence, as Fig. 1.1 illustrates, there is a complementarity between biometrics and visual surveillance in the environmental conditions they work on and the tasks performed: surveillance systems work in uncontrolled conditions but do not automatically identify suspects in a crowd, whereas the biometric systems are effective in automatic identification, but work exclusively in environments that produce good quality data.

The acronym *QUIS-CAMPI* is a composition of two Latin terms and summarizes the goals of this research line: *Quis* stands for *who is* and *campi* refers to *delimited spaces* (plural of campus). Hence, we focus in **the research and development of biometric recognition systems able to work in conditions currently associated to visual surveillance**. The ultimate goal is that whenever a subject enters a *QUIS-CAMPI* region, it is automatically identified using multiple biometric sources, without requiring any active participation from the subject side.

1.2 Major Contributions

Since 2004, our research efforts were concentrated in the development of less constrained biometric recognition systems. Based on the gained worldwide popularity of the *noisy iris* dataset created at the University of Beira Interior (the UBIRIS¹ [126]), several funded research projects have been running, which resulted in the proposal of new techniques and algorithms, described in one edited book, 23 publications in international journals, 4 special issues of international journals, 3 chapters in books and 27 papers published in proceedings of international conferences.

Having in mind the non-cooperative recognition problem, a **firm choice toward applied research** was made, even though some of the techniques and algorithms proposed can be easily extended to other types of problems. This choice had three basic motivations: 1) the intrinsic preferences of the author of this document, closely related with *engineering* and *problem-solving* paradigms; 2) the extremely ambitious goal of *QUIS-CAMPI*, requiring to concentrate as much efforts as possible on it; and 3) the University of Beira Interior (UBI) context, where most of the researchers that cooperated in *QUIS-CAMPI* have predominant skills in Informatics Engineering and were also biased to applied research.

Among the most relevant outputs of the research work being conducted, we highlight:

Identities Pruning: Database Indexing / Retrieval Proposal of a new strategy to index / retrieve *iris codes* at the bit level, particularly suitable for working in low-quality data, where substantial intra-class variability between codes is expected;

Iris Segmentation Proposal of a new iris segmentation strategy for visible-light images, robust to the typical factors that degrade the quality of such data when acquired in less constrained conditions: blur, occlusions, changes in scale and perspective. Also, this strategy works in time approximately linear with respect to the size of the input, which turns it suitable for real-time processing;

Periocular Segmentation Proposal of a method to perform the segmentation of all components in the periocular region (iris, sclera, eyelashes, eyebrows, eyelids, skin and glasses) according to

¹<http://iris.di.ubi.pt>

a *single-shot* model, that uses a unique set of feature descriptors (shape / texture) for all components and a Markov Random Field to guarantee the smoothness and biological plausibility of the solutions found.

Iris Quality Assessment Proposal of a method to assess the quality of visible-light iris images, in terms of their potential usefulness for biometric recognition.

Iris Recognition Proposal of a method to recognize visible light iris images, based on MPEG.7 descriptors and on relative measures, extracted from regions of the iris delimited in a non-supervised way. This strategy is particularly robust to changes in scale, 3D angle, global changes in lighting.

Periocular Recognition Proposal of an atomistic periocular recognition method, in the sense that uses specialized feature encoding / matching strategies with respect to the different biological components in the periocular region. This is in opposition with the previously published approaches, that are fully holistic.

Iris Codes Observation and publication of the iris bit discriminability phenomenon. Complementing previously published works that reported the existence of fragile bits, we observed that there are also certain bits in iris codes that have a predominating value even for different individuals, probably due to biological properties of the human iris tissue. This observation applies both to near-infrared and (more notoriously) to visible light data, which led to the extension of the concept of bit fragility to bit discriminability,.

1.3 *QUIS-CAMPI* Research Team

During the last decade (2004-present), several researchers actively collaborated in *QUIS-CAMPI* research, either faculty or research staff in the scope of two types of research grants: 1) Starting research grant, for undergraduates and recent graduated researchers; 2) Research grant, for graduated and master researchers. Most of these grants were funded by Portuguese *FCT: Fundação para a Ciência e Tecnologia* and the *IT: Instituto de Telecomunicações*, in the scope of six research projects:



NECOVID: Covert Negative Biometric Recognition, with reference PTDC/EIA-EIA/103945/2008. Host institution: Universidade da Beira Interior. Participating institution: IT-Instituto de Telecomunicações. Begin: 03-05-2010. End: 30-04-2013. Budget: 86.685 €. Funding institution: FCT-Fundação para a Ciência e Tecnologia (COMPETE, FEDER). Project team: **Hugo Proença** (coordination), Luís A. Alexandre (UBI), Paulo Fazendeiro (UBI), Pedro Almeida (UBI), Gil Melfe Mateus Santos e Sílvio Filipe.



QUIS-CAMPI: Biometric recognition in Surveillance Environments funded by the *IT-Instituto de Telecomunicações*. Begin: 01-09-2014. End: 31-08-2015. Budget: 40.000 €. Project team: **Hugo Proença** (coordination), Paulo Lobato Correia (*Instituto Superior Técnico*), Luís Ducla Soares (*ISCTE-Instituto Universitário de Lisboa*), Jaime Cardoso (*Faculdade de Engenharia da Universidade do Porto*) e Paulo Fiadeiro (Department of Physics, UBI).



NOISYRIS - Synthesis of Noisy Iris Images for Biometric Recognition Purposes funded by IT-Instituto de Telecomunicações. Begin: 01-09-2011. End: 31-08-2013. Budget: 40.000 €. Equipa de projecto: **Hugo Proença** (coordination), Frutuoso Silva (UBI).



MBIR-Multispectral Biometric Iris Recognition funded by IT-Instituto de Telecomunicações. Begin: 01-06-2010. End: 31-05-2012. Budget: 31.498 euro. Project team: **Hugo Proença** (coordination), Paulo Fiadeiro, (Department of Physics, UBI), Gil Santos Msc. (Department of Informatics, UBI) e Luís Lucas Bsc. (Department of Physics, UBI).



BioRec - Non-Cooperative Multimodal Biometric Recognition with reference PTDC/EIA/69106/2006. Host institution: Universidade da Beira Interior. Begin: 15-10-2007. End: 31-03-2011. Budget: 130.000 €. Funding institution: FCT-Fundação para a Ciência e Tecnologia (COMPETE, FEDER). Project team: Luís A. Alexandre (UBI, coordination), Pedro Domingues de Almeida (UBI), **Hugo Proença** (UBI), Francisco Catarino, Ricardo Santos, Gil Santos, João Oliveira, Sílvio Brás Filipe, Tiago Velho, Rui Silva, Chandrashekar Padole, Emundo Hoyle, Rui Raposo, Diogo Correia (UBI).



PAIRUE: Fusion of Palmprint and Iris Recognition in Uncontrolled Environments funded by IT-Instituto de Telecomunicações. Begin: 01-06-2011. End: 31-05-2012. Budget: 39.920 €. Project team: Luís Ducla Soares (ISCTE-Instituto Universitário de Lisboa, coordination), Paulo Lobato Correia (Instituto Superior Técnico) e **Hugo Proença** (UBI).

1.4 Document Structure and Notation

This remaining of this document is organised as follows: Chapter 2 summarises the state-of-the-art in the scope of *QUIS-CAMPI* and overviews our approach for the problem, also supplying a cohesive perspective of its sub-problems and of the major difficulties behind each one. The subsequent chapters are devoted to specific families of problems in the recognition chain: Chapter 3 discusses the so-called *Low-level Vision Problems*, while Chapter 4 (*High-level Vision Problems*) describes the most relevant research conducted in biometric recognition of degraded data. In Chapter 5 we briefly discuss some ethics and privacy concerns behind the development of biometric recognition systems that work covertly in public environments. Finally, Chapter 6 concludes this document, by summarising the most important points of the research being conducted and pointing for further directions in that context.

For comprehensibility, the **particularly important** parts of the document include a **Summary Table**, giving a cohesive / condensed perspective of the information in that part:

	Key statement 1 Key statement 2
---	--

Summary tables are also used to highlight the publications made in the scope of each topic covered in the document and to itemize the major challenges and open-problems behind each phase of the recognition chain proposed for *QUIS-CAMPI*, which we believe can be important as basis for the subsequent research.

Chapter 2

Biometric Recognition in Surveillance

2.1 State-of-the-Art

As above stated, there is no doubt that concerns about the *security* and *safety* in crowded urban areas have been increasing in recent times. These concerns raised the interest on biometrics and turned this topic among the most popular topics in the pattern recognition / computer vision domains, which can be objectively verified by the number of recent patents filled, the number of papers published and the amount of economic and human resources devoted to the topic. However, the successful deployment of biometrics technologies is still considered a grand-challenge[76], mainly due to three factors: performance, scale and usability. At this moment, there are not biometric recognition systems that work effectively using data acquired without subjects cooperation and in large-scale scenarios.

Currently, a huge number of visual surveillance systems is deployed worldwide. However, perhaps contrary to common belief, their automated analysis is most times constrained to action recognition (e.g., detect fights, suspicious behaviour, unattended luggage...). When it comes to human recognition, the error rates obtained in real-world data are still far from the demanded, which is specially concerning for large-scale scenarios, where even low error rates result in too many observations of failure recognition attempts. Hence, the current focus is putted in reducing the levels of cooperation that are demanded to subjects to obtain input data of minimal quality. In this context, the *VeriLook Surveillance* (VLS) system from *Neurotechnology*, constitutes a significant effort toward fusing biometrics and visual surveillance, but its performance is not considered satisfactory for large scale scenarios, and it still runs in relatively controlled environments: it uses high resolution cameras and requires a complex enrolment phase to obtain a large set of enrolment data per subject representing the most typical variations in a personalised way (e.g., differences in pose, lighting conditions and distances).

Fig. 2.1 gives the ambition of *QUIS-CAMPI* research with respect to the state-of-the-art in automated surveillance (VA, (also representing the VLS system) and biometric recognition (B) detecting two major challenges:

- **Challenge 1:** with respect to the state-of-the-art in biometrics, research for novel algorithms able to work effectively in data of much lower quality than the used currently, in result of the “transparent-to-user” data acquisition feature.
- **Challenge 2:** with respect to the state-of-the-art in visual surveillance (VA and VLS), improve substantially the ability to perform automated recognition, which currently still demands some kind of human intervention.

The Iris-on-the-Move [105] and the Human-Identification-at-a-distance (HumanID) [31] are good examples of biometric recognition systems that work in relatively unconstrained conditions. However, even though they are major examples of the engineering behind the new generation of biometric systems,

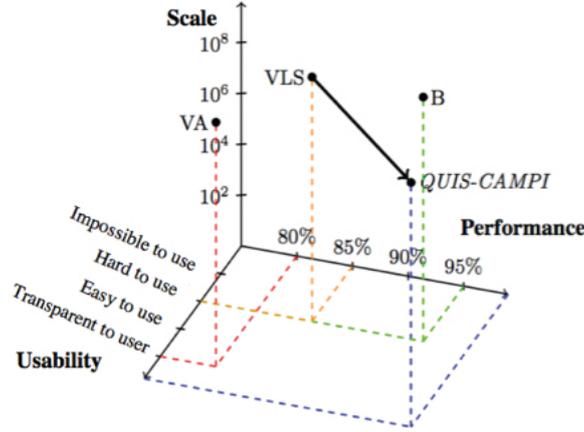


Figure 2.1: Schematic representation of the three issues behind the currently developed biometric recognition systems. “B”, and “VA” denote the current state-of-the-art biometric recognition and automated surveillance systems. VLS represents the *VeriLook Surveillance* system (adapted from[76]).

the concept of “user” of the system is still indubitably associated to them, which constrains their use for forensic purposes.

For most of the existing systems, the processing chain starts by the detection and tracking of human silhouettes, which is far from a trivial problem. Unconstrained environments are considered too intricate to be covered by a single camera, and often require multiple cameras for monitoring. While a single-camera tracker searches for correspondences between consecutive frames, a multi-camera tracker must also establish correspondences between observations of objects across cameras. The goal is to tag all instances of the same object at different locations and moments. A classical strategy for detecting objects across multiple cameras is based on camera calibration [23]. Alternatively, alignment-based approaches recover the geometric transformation between cameras automatically [94] using overlapping fields-of-view (FOV). However, proper alignment requires overlapping FOVs, which is not always feasible. To avoid that requirement, cameras are located in non-overlapping locations that nonetheless allow to establish path dependencies using probabilistic models [77].

Methods that rely on contextual information of the surveillance system may not be easily generalized. In these cases, object detection is based on feature matching, and the main challenge lies in the feature variability caused by different lighting conditions, poses and scales variations. Recently, these issues have been starting to be addressed by members of this project, using novel object representations (persons) and learning frameworks to mine streams with various lengths and starting points [132]. This type of algorithms handles the concept drift, accommodates new classes, deals with partially labeled and unlabelled data, and is of limited complexity. Still in this scope, the work of Kamgar-Parsi et al. [83] is considered the most important reference: it uses data acquired from surveillance cameras, and matches the unknown identities against a shortlist of potential suspects, using multiple classifiers (one per subject) followed by a fusion scheme at the decision level.

Among the strong biometric traits that are more frequently used in unconstrained conditions, the face, the iris, and the periocular region should be highlighted. According to the concept of *eigenface*, Wright *et al.* [167] used sparse coding techniques to augment the robustness against occlusions and changes in facial expressions. Using 3D data, Drira *et al.* [44] handled changes in facial expressions, occlusions and poses, using as reference the region around the nose, and extracting radial topologic curves of the face, subsequently matched by Riemann analysis techniques. Regarding the iris, the works of Tan *et al.* [153], Proença [128], [129] and Pillai *et al.* [123] are good references, as they all focus the robustness of the recognition process. Finally, the periocular region is regarded as a trade-off among the

face (too sensitive to facial expressions and occlusions) and the iris (a too small a moving target). Park *et al.* [115] fused global and local encoding techniques, obtaining results that encouraged further works. Woodard *et al.* [166] achieved performance levels similar to the attained by face recognition in visible wavelength data, using local appearance descriptors.

Due to the extreme conditions in the environments, the idea behind soft biometric traits [76] is to prune the searching spaces. This kind of techniques is becoming increasingly popular, since the pioneering work of Alphonse Bertillon in the XIXth century [13], which identified criminals based on anthropometric measurements (e.g., height and length of the arms, the morphological description of the appearance and body shape, eye color or finger anomalies). We also highlight the work of Cao *et al.* [24], which classified gender by means of anthropomorphic features. Zhang *et al.* [175] developed an algorithm based on super-resolution to recognize the type of gait in low-resolution images. Niinuma *et al.* [110] used soft biometrics for continuous subject authentication after the initial authentication has been done by strong biometrics.

In terms of automated surveillance, the research has been focused mostly in action recognition and not that much in the recognition process itself. As relevant examples, we highlight the Knight [144] system, which detects and classifies objects in a scene. Objects are modeled by combining color, shape and motion descriptors. In a similar way, the W⁴ [62] system analyzes the activities of subjects in a scene, modeling each region of the human body by circles and ellipses (strong restrictions for weak data), and inferred their actions based on the relative position and motion of such shapes.

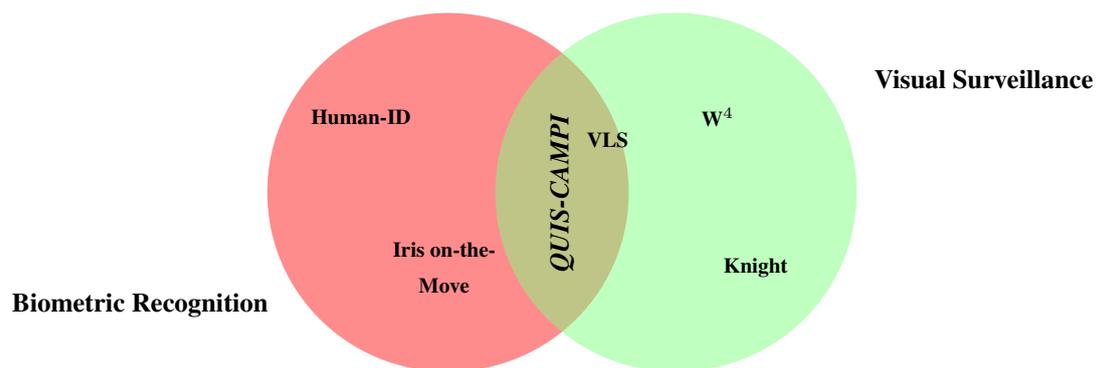


Figure 2.2: Positioning of *QUIIS-CAMPI* research with respect to the state-of-the-art.

In short, Fig. 2.2 summarizes the positioning of *QUIIS-CAMPI* with regard to the state-of-the-art in the biometrics and visual surveillance domains. Our goal is bridging both areas, i.e., to perform reliable biometric recognition from low quality data of limited resolution, such as the typically acquired from automated surveillance sensory.

2.2 Recognition Chain

To realize the main ambition behind *QUIIS-CAMPI*, several preliminary choices were made:

1. Use coupled visual surveillance and Pan-Tilt-Zoom (PTZ) imaging devices (Fig. 2.3). Following a sustainable economical perspective, we decided to use hardware similar to the already deployed in

many cities worldwide: a wide-view visual surveillance camera covers the whole scene, whereas the synchronised PTZ device collects high-resolution data from portions of the scene;

2. Use soft biometric information for indexing / retrieval in database identities. To compensate for the adversity of the environments and the low quality of data, it is important that soft biometric information (e.g., body metrology and gait) prune the set of potential enrolled identities for each query;
3. Use strong biometric traits for watch list detection. Receiving the set of identities pruned by soft biometric information, we explore innovative ways to fuse information from the face and periocular recognition modules, which were selected exactly due to their complementar features;
4. Faithfully balance ethics / privacy and safety / security issues. Considering that a *QUIS-CAMPI* system raises privacy concerns, we decided to assure the anonymity of the large majority of subjects in a scene. By attempting identification from the negative perspective, we actually ignore the identity of common citizens, whilst assuring that the potentially harmful subjects (elements of the watch list) are not among the crowd.



Figure 2.3: Example of a synchronised pair of wide-view (W) and PTZ (P) cameras, from a *QUIS-CAMPI* laboratorial prototype mounted in the *SOCIA: Soft Computing and Image Analysis lab.* outdoor wall.

The idealized *QUIS-CAMPI* recognition system is composed of nine modules / phases, from the moment the wide-view surveillance camera captures information from a moving subject, up to the decision about the corresponding identity. In Fig. 2.4 we illustrate a work flow of the most important phases, which are detailed in the subsequent parts of this document, and were divided into two main groups: 1) **low-level vision problems**; and 2) **high-level vision problems**. The first group (shaded in gray) contains the phases that are required to run in real-time and involve the perception of the scene where the system is running, up to the moment the PTZ device is pointed out for a particular region in the scene and acquires a high-resolution image of (ideally) a subject's head. Once that information is acquired, the subsequent phases belong to the second group and don't have strict requirements in terms

of the computational burden, as they can be forked to separate processes, each one performing for one recognition attempt per query sample.

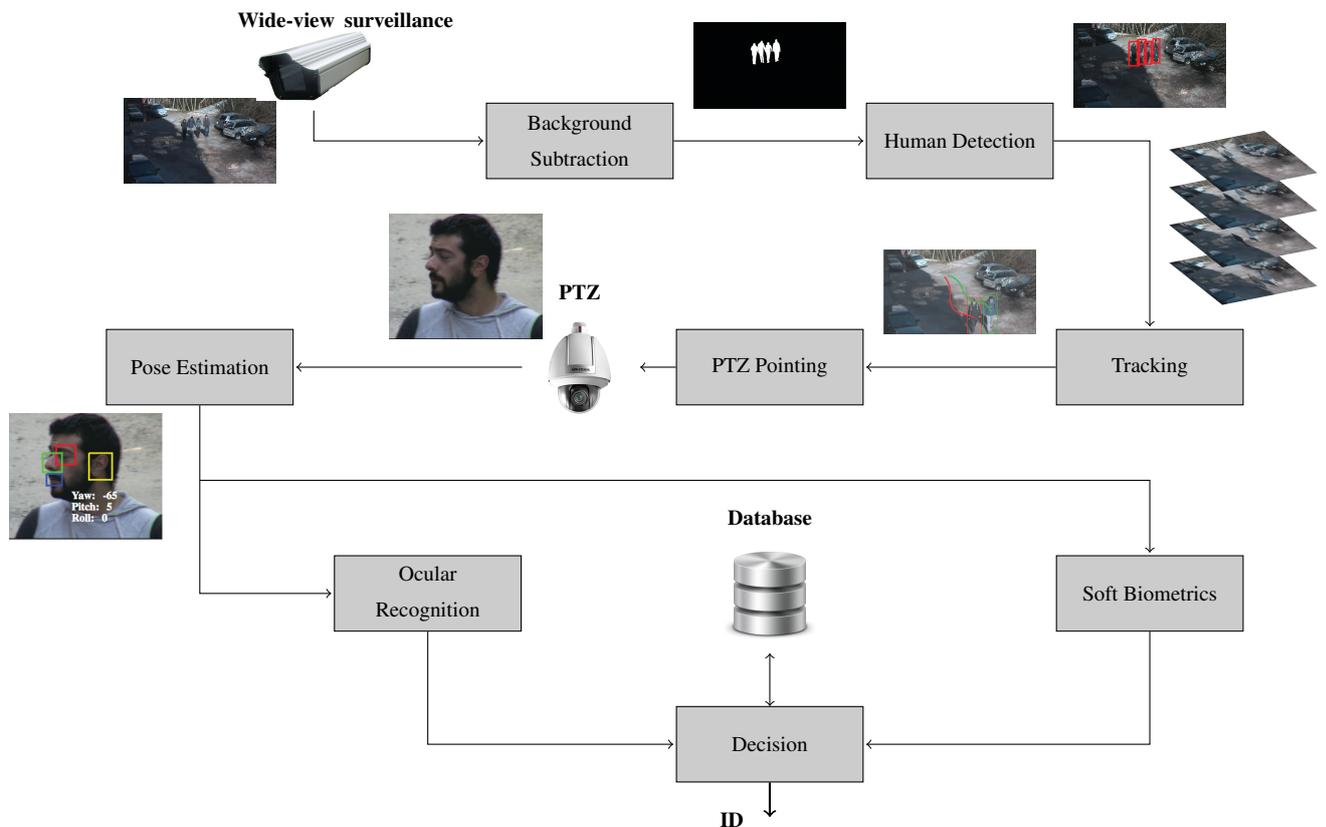


Figure 2.4: Major phases of the recognition process in a non-cooperative ocular recognition system.

The process starts by a wide view surveillance camera feeding a background subtraction module, returning as output the regions of interest that contain moving objects. This information enters a human detection module that is responsible for the discrimination between regions that contain a human shape and any other type of information. The sequences of positions of humans in the scene are the input of a tracking module, where objects are grouped and labeled into *subject A, B, ...*. Next, the PTZ pointing module is responsible for selecting the best candidates for the acquisition of high resolution images, pointing a PTZ device to a specific position in the scene. Both the wide-view and PTZ devices are synchronised, in order to appropriately exchange scene coordinates.

Once a high resolution image of a subject's head is acquired, the remaining modules perform the biometric recognition of that subject, either using holistic information (e.g., the face), or piece-wise components such as the ocular region, the nose, lips and the ear. Simultaneously, a soft biometrics module and a gait recogniser prune the space of possible identities for that query sample. Finally, the decision module gets information about the set of plausible identities for the sample and outputs the corresponding most probable identities in the watch list.

Chapter 3

Low-level Vision Problems

3.1 Background Estimation

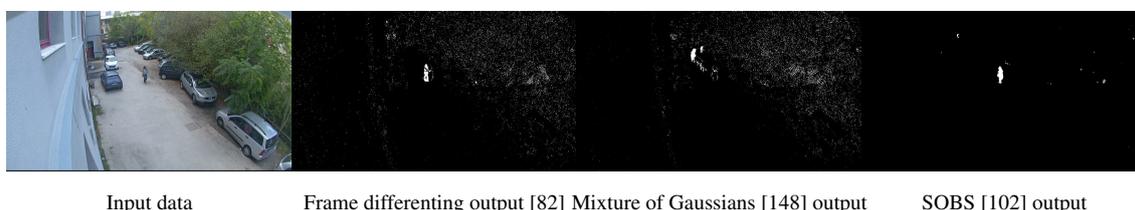


Figure 3.1: Examples of the output of three background subtraction algorithms, when running in outdoor scenes corresponding to typical visual surveillance conditions.

As illustrated in Fig. 3.1, background subtraction (BS) methods aim to divide the scene into two disjoint parts: 1) background, that contains the static regions in the input data, which usually should be disregarded from subsequent processing; and 2) foreground, that contains the regions-of-interest (ROIs) of all objects that the system should care about. This step is usually the basis of the processing chain, which augments its relevance as any error compromises the success of the whole process. Also, by pruning the scene and reducing the amount of information to be handled, the computational burden of the whole system is substantially reduced.

The previously published methods for background subtraction can be divided into three families: 1) basic; 2) Gaussian-based; and 3) machine-learning based, ordered by their level of complexity. The first family regards the most simple strategies and the pioneering approach dates back to 1979, when Jain and Nagel [75] analysed the differences in pixel intensity with respect to time to discriminate the useless regions in a scene. Similar approaches were more recently published (e.g., [146] [82]) and have as main advantage their reduced computational cost, although their lack in robustness. Another type of methods derives a coarse estimate of background from the earliest frames (e.g., [52]) or from the last frames in a scene (e.g. [106]), using simple statistics with respect to time (median filtering).

A family of methods of intermediate complexity models the density of the intensities of each pixel with respect to changes in time, either assuming single Gaussian (e.g., [168]), mixture of Gaussians (e.g., [148]) or non-parametric models (e.g., [48]). The underlying strategy is to obtain an estimate of the typical changes in each position in the image, and then report a pixel as foreground every time an outlier is observed. Simultaneously, the models are continuously updated, so to adapt to slight changes in the scene. This kind of methods is a trade-off between effectiveness and computational cost, increasing errors substantially when portions of the image are affected by multiple sources of noise at the same

time.

The most complex family of methods relies in well known machine-learning algorithms to obtain local representations of the background. Clustering-based approaches estimate the background by grouping pixels in different clusters, each one corresponding to a different source of background. The Codebook model [86] uses a set of code words to represent each cluster, using color and brightness information to define the distance function. Nearest neighbour techniques are also used for background estimation, fed by different features such as luminance [170] and chrominance [22]. More recently, unsupervised neural models have been tested to enhance robustness in real-world conditions (e.g., [102]). In this kind of methods, each pixel is modelled by a neural map, where each element stores typical RGB values at that position and acts as cluster centroid, also conditioning the values of its neighbours. The idea behind competitive neural networks [100] is highly similar to this strategy, in this case additionally using reinforcement learning techniques with respect to the winner neuron.

	Background Estimation Main Challenges:
	Long standing objects
	Low-frequency background changes
	Background initialization

3.2 Human Detection

The detection of humans is a particular instance of object detection, with some specificities that increase the challenge of the task, particularly in case of crowded scenes and unconstrained lighting environments (e.g., outdoor). Such challenges include the deformations in shape of the human silhouette with respect to movements of the legs, arms and head, partial occlusions due to other humans or objects, clothing and changes in perspective that might change dramatically the information sensed.

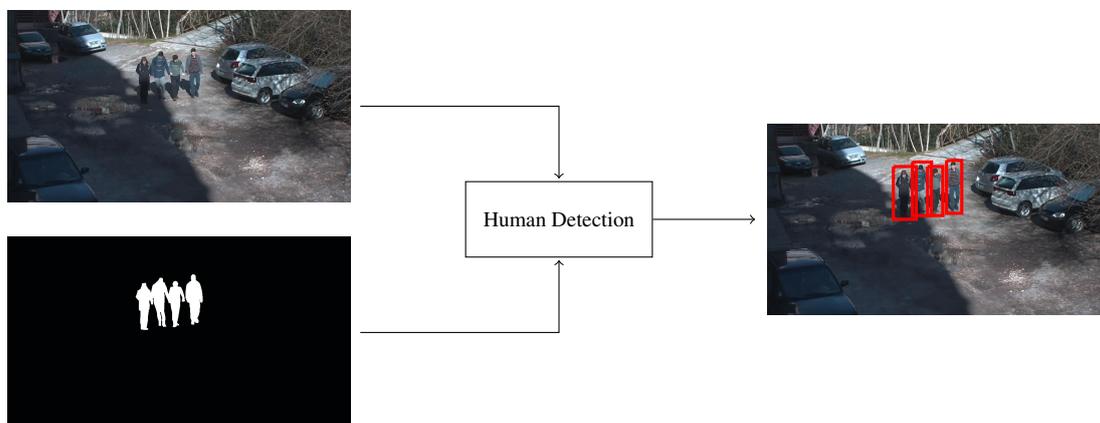


Figure 3.2: Human detection in surveillance.

Fig. 3.2 illustrates the human detection task: using the raw input data and the output of the background subtraction module, the goal is to define a set of regions-of-interest (ROIs), such that each one corresponds to one human in the scene. From the computer vision perspective, working in outdoor environments, with both global and local variations in lighting conditions, and imposing no constraints about the number of subjects represents a high challenging problem, even harder due to the limited data resolution and the required ability to work in real-time [171].

There are two major families of methods for human detection: 1) holistic methods, where the whole body is searched in the image; and 2) part-based methods, where each part of the body is detected inde-

pendently and information is further fused for consistency purposes. Most of the holistic methods learn a discriminative model, being well represented by the popular Viola and Jones' [160] method, adapted to detect humans using motion patterns [161] [162]. Similarly, Dalal and Triggs [32] use histograms of oriented gradients (HOGs) to feed a classifier, such as support vector machines (SVMs), in a way similar to [107] and [143]. Along with HOGs, local binary patterns (LBPs) features [113] have also been widely used for human detection [174] [164].

Regarding part-based methods, Mikolajczyk *et al.* [109] use a probabilistic model to assemble all parts of the body, each one detected in a coarse-to-fine strategy. Lin *et al.* [96] considered the head the most reliable part to be detected and to estimate the number of persons in a large crowd, similarly to Subburaman *et al.* [149], in this case using exclusively head features to attain state-of-the-art results. Zhao and Nevatia [177] analysed the silhouette boundaries from the background estimation mask and detect the head by searching for vertical peaks on these contours. Detections are subsequently filtered by cross-checking silhouette information with human anthropometric constraints. Wu and Nevatia [169] use four body parts (full-body, head-shoulder, torso, and legs), each one learned by boosting a set of weak classifiers based on edgelet features (short segments of edge pixels). The responses given by all detectors are fused to provide robustness to occlusions. This work was extended not only to improve detection performance but also to obtain a fine estimative of the human-body boundaries (segmentation).

	Human Detection Main Challenges:
	Superimposed / occluded regions
	Deformable shapes
	Local lighting changes (shadows)

3.3 Pose Estimation

As illustrated in Fig. 3.3, the task of estimation phase is to recover the 2D or 3D configuration of the human body, based on a set of observations. This problem is far from trivial, as one has to deal with occluded regions, deformable shapes and foreshortening. This phase actually requires that the whole scene is understood, i.e., all portions of the image should be linked to biological body components, which in our case determines the traits used to attempt biometric recognition.

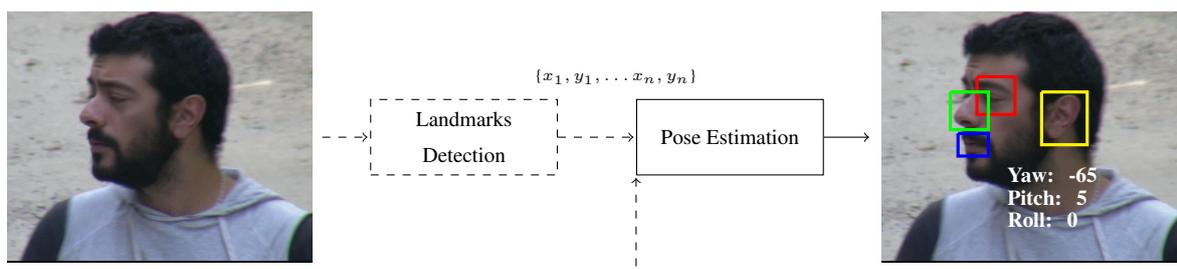


Figure 3.3: Pose estimation phase

The previously published approaches for pose estimation can be classified into two families: 1) generative; and 2) discriminative, whether they fit a model to the available data or attempt to establish correspondences between features and pose configurations.

Generative approaches estimate pose by analyzing prior information about human kinematics and appearance and are usually more computationally expensive than discriminative methods. This kind of methods is usually divided into top-down or bottom-up groups, whether they explicitly project a whole model onto the image and evaluate the corresponding likelihood or detect a set of candidate body parts

and assemble them to determine the subset that best explains the available data. Pictorial structure models [51] are widely popular in top-down algorithms. These models consist in a set of connected parts resembling the shape of the human body, where each part is represented by an appearance model. The geometric interactions between parts are modelled by deformation costs. Considering the high complexity of this kind of models, cascade classification /regression strategies were also proposed to alleviate this burden (e.g., Sapp *et al.* [142]). Contrary to top-down approaches, bottom-up approaches do not need to explicitly project the model onto the image. Instead, they detect a set of candidate body parts and assemble them in order to determine the subset that best explains the data and respects kinematics constraints. In general, bottom-up methods learn a set of detectors for each body part, which provides a set of possible locations. Then, as an exhaustive search over the space of all possible configurations is most times unpractical, tree-structured models, Markov networks and loopy graphs are used (e.g., Jiang and Martin [78]) to alleviate the computational burden of obtaining an acceptable solution.

Discriminative methods do not explicitly fit a model to the data and - instead - learn an appropriate mapping between model and image features, turning particularly important to have robust image descriptors and pose informative features. A learning set of pre-annotated images is required to learn a transfer function between the feature and pose spaces. Such function is usually inferred by means of machine-learning algorithms, fed by different kinds of descriptors, such as silhouette (e.g., Howe [71]), histograms of oriented gradients (e.g., Poppe [121]), shape-context (e.g., Lv and Nevatia [101]). However, in this kind of approaches, shape-based features are the most frequent logical choice, motivated by the work of Agarwal and Triggs [1], who noted that shape features are more informative than any other for this specific problem. Silhouette-based features are also popular (e.g., [137], [2] and [57]) due to three major reasons: 1) they provide useful information about the 3D pose; 2) they tend to be invariant to lightning conditions; 3) are easily extracted from the background subtraction mask.

When comparing generative to discriminative approaches, the latter are usually less accurate, mainly due to the restricted set of poses available in the learning data. By relying on pre-trained part detectors, generative approaches usually attain better performance, but are also more sensitive to data quality. This is particularly important for visual surveillance applications, which justifies that the most recent works in this scope focused only in the estimation of discrete pose states and belong to the discriminative family (e.g., Chen and Odobez [26] and Huang *et al.* [72]).

	Pose Estimation Main Challenges:
	Occluded landmarks
	Deformable shapes
	Computational cost of energy minimization

3.4 Tracking

Fig. 3.4 illustrates the task typically associated with tracking: given an initial estimate of the location of one object, the tracking phase determines the positions of that object in the subsequent frames, having in our case two major goals: 1) by perceiving the object path, accurate predictions of the location of the object in forthcoming frames can be made, allowing to timely point the PTZ device for a specific position; and 2) once the high-resolution data of a subject is acquired, that element can be ignored of any subsequent processing.

The existing approaches for object tracking can be divided with respect to the feature space they work in: 1) motion-based algorithms exploit the object dynamics, based on cues such as velocity, articulation and periodic constraints. Motion models are typically related to Bayesian tracking approaches, where dynamics is used to update the target state over time (e.g., Breitenstein *et al.* [21]) or shape information (e.g., Zhou *et al.* [180]). Tracking based on optical flow estimation is also a relevant example of this family, namely the KLT tracker [145], that assumes small movements between frames with brightness

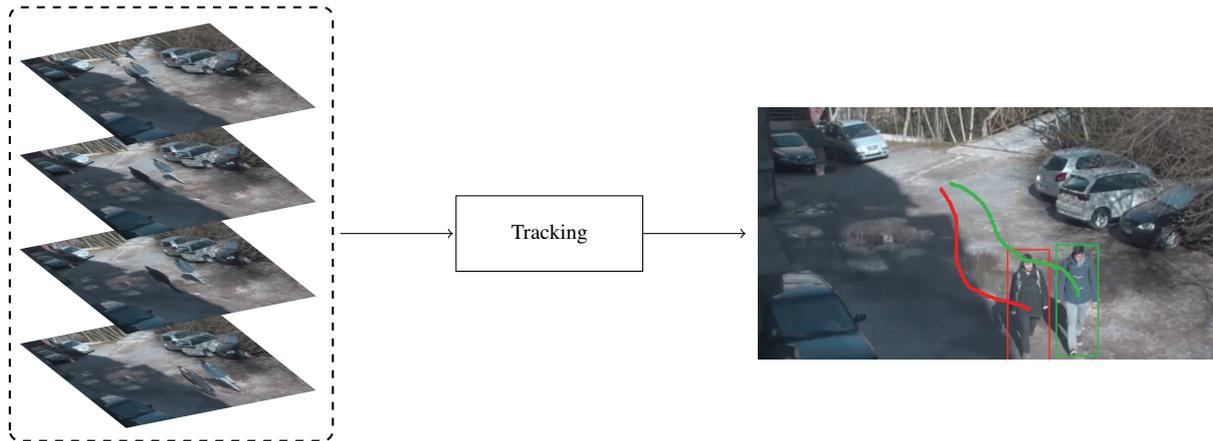


Figure 3.4: Main task in object tracking: given a sequence of images, determine the paths of each object-of-interest in the scene.

constancy, to follow a set of keypoints; 2) appearance-based algorithms are frequently associated to kernel-based methods that represent the target as a point in a high dimensional space, characterized in terms of histograms of intensities (e.g., Comaniciu *et al.* [29]), LBPs (e.g., Kalal *et al.* [80]) or sparse representations (e.g., Zhong *et al.* [178]) from channels of different color spaces; 3) shape-based algorithms eliminate the need to consider varying illumination and changes in appearance, yet turn more difficult to obtain a reliable estimate of the object boundary. However, shape information is most times used together with other families of cues (e.g., texture), which is particularly useful for low quality data (e.g., Liu *et al.* [97]).

Complementary, tracking methods can be classified with respect to the properties of their main algorithm. The earliest approaches attempted to track objects by searching for specific patterns in the neighbourhood of the previous known location (kernel / model tracking) or by evolving the state of the target according to a motion and appearance model (Bayesian tracking). More recently, a new strategy has been gaining popularity (tracking-by-detection), which is particularly suitable for arbitrary object tracking in unconstrained scenarios. A typical example of this family of algorithms is the proposal of Zhou and Aggarwal [179], using the Kalman filter with a constant velocity model to estimate the state of humans. Aiming at improving the robustness of tracking to dynamic environments, Zhang *et al.* [176] use a kernel-based Bayesian framework, where the feature space combines appearance and shape information. Mixture of Gaussians are also used to model the appearance model and the Chamfer matching provides a similarity measure between shapes. In case of shape cues, it is particular hard to be match shapes subject to severe occlusions and deformed shapes. For this reason, Saber *et al.* [140] use the concept of partial shape matching, as Husain *et al.* [73] did, to track objects in surveillance scenarios.

As stated above, the iterative use of detectors has been gaining popularity, mainly due to the high flexibility of this kind of algorithms and the hardware advances that have been reducing the amount of time required for execution. These algorithms estimate the target position by searching the position in the image that maximizes a similarity function between an image point and the feature vector of the target state. Contrary to other tracking families, no a priori target representation is required, inferring the corresponding model by online learning algorithms, allowing the resulting model to adapt to any kind of object and its variations in appearance. Regarding the used machine-learning classifier, online boosting classifiers were a typical strategy in the earliest approaches (e.g., Grabner *et al.* [55]), but state-of-the-art techniques, as Babenko *et al.* [8], now exploit multiple instance learning techniques to reduce the sensitiveness to slight changes in appearance.

Recently, substantial attention has been paid to multiple object tracking. Despite multiple instances

of a tracking algorithm can be used to address multiple targets, there is an exponential growth of computational complexity that restrains their use when the number of targets is high. Greedy strategies have been used to handle such complexity, where correspondences are regarded as an assignment problem based on spatial distance (e.g., Wu and Nevatia [169]). Offline or batch techniques methods comprise another solution for multiple target tracking, using the complete set of detections before estimate the trajectory. This phase is regarded as an optimization problem, where a function describes the cost of each solution (e.g., Andriyenko and Schindler [5]). Linear programming techniques are used in several works to cope with the computational burden of this optimization step. A continuous formulation of this problem was recently introduced by Andriyenko and Schindler [6], which yet high effective, has as main drawback the high latency required to analyse a video, turning it incompatible with real-time requirements.

**Tracking Main Challenges:**

Real-time multiple object tracking

Objects cross-paths

Local changes in typical image features (brightness, entropy,...)

Chapter 4

High-level Vision Problems

4.1 Identities Pruning: Database Indexing / Retrieval



Hugo Proença; Iris Biometrics: Indexing and Retrieving Heavily Degraded Data, IEEE Transactions on Information Forensics and Security, volume 8, issue 12, pag. 1975-1985, ISSN 1556-6013, Digital Object Identifier 10.1109/TIFS.2013.2283458, 2013.

According to the most acknowledged iris recognition method [37], matching *IrisCodes* primarily involves the accumulation of bitwise XOR operations. However, despite the extreme computational effectiveness of this matching scheme, the time required for exhaustive searches grows linearly with the number of enrolled identities. Also, the time required for de-duplication searches grows quadratically with respect to the size of the database, which is specially concerning in nationwide scales. As noted by Hao *et al.* [61], indexing is a specific case of the general *nearest neighbour search* problem, and motivated several proposals in the last years. However, most of these methods were designed for decision environments of good quality, with a clear separation between the genuine and impostor matching scores.

We were mainly interested in decision environments with a significant overlap between the genuine and impostors matching scores, corresponding to systems that operate in less controlled data acquisition protocols. We proposed an indexing / retrieval method that runs at the code level, i.e., after the feature encoding process. Codes are decomposed in a coarse-to-fine scheme, and their position in an n-ary tree determined. In retrieval, the probe is decomposed in a similar way, and the distances to multi-scale centroids are obtained, penalizing most of the paths in the tree and traversing only a subset of nodes down to the leaves. When compared to related works, the main contributions of the proposed method are three-fold: 1) it is compatible with different signature encoding methods; 2) it outperforms the state-of-the-art approaches in poor quality data, particularly in the performance range that is meaningful for biometrics (hit rates above 0.95); and 3) it has a reduced computational cost: compared to exhaustive searches, indexing becomes advantageous when more than a few thousands identities are enrolled.

The suitability of the proposed method to handle degraded data has its roots in the concepts of *coarse-to-fine analysis* (in indexing) and *non-excluding branches* (in retrieval): 1) in indexing, *IrisCodes* are grouped in branches of the tree according to their multi-scale features, being *tree-level* and *analyzed-scale* in direct relationship. This means that at the root (maximum) level, *IrisCodes* are grouped according to their lowest frequency components. 2) in retrieval, a parallel searching scheme was devised: starting with a residual value, the tree is traversed along different branches until the sum of residual penalizations for each branch guarantees that the identity of interest will not be found there.

4.1.1 Indexing

Let s_i denote an *IrisCode* from the i^{th} subject. As illustrated in Fig. 4.1, the rationale is to obtain coarse-to-fine representations of s as a function of the level l in the tree ($s^{(l)}$). These representations are grouped

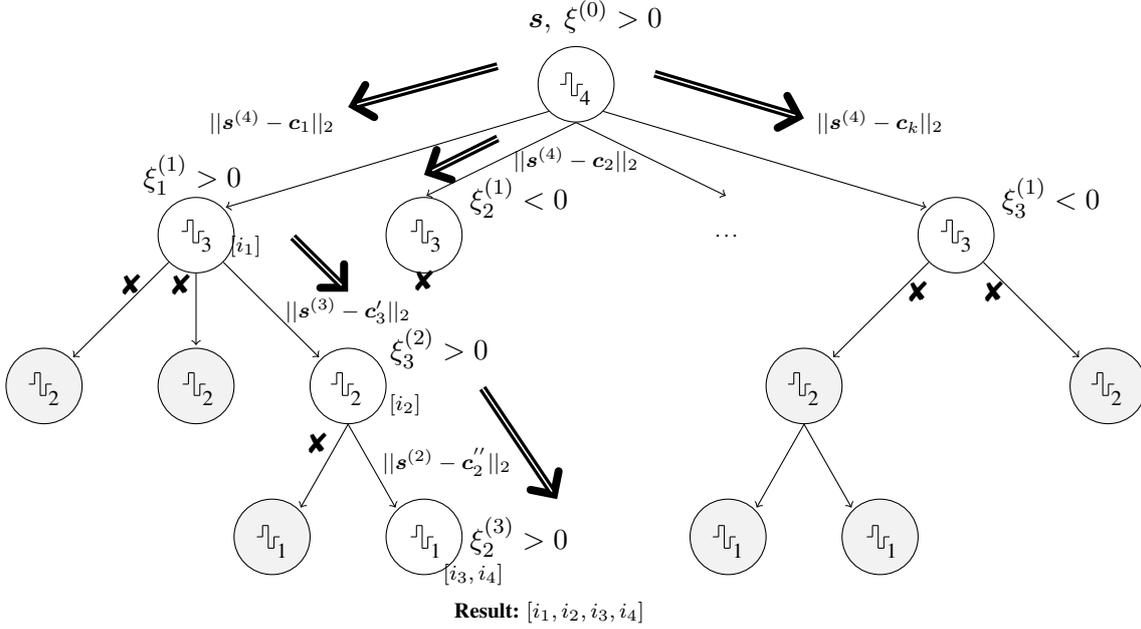


Figure 4.1: Cohesive perspective of the indexing structure and of the retrieval algorithm. For a query s with residual $\xi^{(0)}$, the distance between the decomposition of s at top level ($s^{(4)}$) to the centroids c_i is used to generate the new generation of residuals ($\xi^{(1)}$). For any branch with negative values, the search is stopped, meaning that subsequent levels in the tree are not traversed (illustrated by gray nodes). When traversing the tree, every identity found at any node while $\xi^{(\cdot)} > 0$ is included in the retrieved set.

according to their similarity in the L_2 metric space, and stored in each node. A node is considered a leaf when its centroid c is at a sufficiently small distance from all elements, i.e., $\|s_i^{(l)} - c\|_2 \leq \nu, \forall i$.

Let $\phi(x) = \sum_{k \in \mathbb{Z}} h(k) \sqrt{2} \phi(2x - k)$ and $\psi(x) = \sum_{k \in \mathbb{Z}} g(k) \sqrt{2} \phi(2x - k)$, where $h(\cdot)$ and $g(\cdot)$ are low-pass and high-pass filters. According to Mallat's multiresolution analysis [103], the operator representation of these filters is:

$$\begin{aligned} \mathbf{H}_a^{(k)} &= \sum_n h(n - 2k) a_n \\ \mathbf{G}_a^{(k)} &= \sum_n g(n - 2k) a_n, \end{aligned} \quad (4.1)$$

where $\mathbf{H}_a^{(k)}$ and $\mathbf{G}_a^{(k)}$ are one-step wavelet decompositions. Let $\text{len}(s^{(n)}) = N = 2^n$ be the length (in our experiments, $n = 11$) of the signal s represented at scale n by a linear combination of ϕ filters:

$$s^{(n)} = \sum_n a_k^{(n)} \phi_{nk}. \quad (4.2)$$

At each iteration, a coarser approximation $s^{(j-1)} = \mathbf{H} s^{(j)}, j \in \{1, \dots, n\}$, is obtained: $d^{(j-1)} =$

$\mathbf{G} \mathbf{s}^{(j)}$ are the residuals of the transformation $\mathbf{s}^{(j)} \rightarrow \mathbf{s}^{(j-1)}$. The discrete wavelet transformation of $\mathbf{s}^{(n)}$ is:

$$\mathbf{s}^{(n)} \equiv [\mathbf{d}^{(n-1)}, \mathbf{d}^{(n-2)}, \dots, \mathbf{d}^{(0)}, \mathbf{s}^{(0)}], \quad (4.3)$$

where $(\sum_{i=0}^{n-1} \text{len}(\mathbf{d}^{(i)})) + \text{len}(\mathbf{s}^{(0)}) = \text{len}(\mathbf{s}^{(n)}) = 2^n$.

In reconstruction, $\mathbf{s}^{(n)}$ can be approximated at different levels using \mathbf{H}^* and \mathbf{G}^* filters:

$$\begin{aligned} (\mathbf{H}_a^*)^{(n)} &= \sum_k h(n-2k)a_k \\ (\mathbf{G}_a^*)^{(n)} &= \sum_k g(n-2k)a_k, \end{aligned} \quad (4.4)$$

where $\mathbf{s}^{(n)} = \sum_{j=0}^{n-1} (\mathbf{H}_a^*)^{(j)} (\mathbf{G}_a^*)^{(j)} \mathbf{d}^{(j)} + (\mathbf{H}^*)^{(n)} (\mathbf{G}_a^*)^{(n)} \mathbf{s}^{(0)}$. Considering that *IrisCodes* are binary, the Haar wavelet maximally correlates them and its filter coefficients are $\mathbf{h} = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, $\mathbf{g} = [\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$, with similar reconstruction coefficients $\mathbf{h}^* = \mathbf{h}$ and $\mathbf{g}^* = -\mathbf{g}$. Under this strategy, \mathbf{H} acts as a smoothing filter and \mathbf{G} as a detail filter.

When reconstructing a signal at a given level, the detail coefficients of small magnitude can be disregarded, as they intuitively do not have a major role in the signal. This is possible because wavelets provide an unconditional basis, i.e., one can determine whether an element is important by analyzing the magnitudes of the coefficients used in the linear combination of the basis vectors.

The threshold (λ) for the minimal magnitude of the coefficients considered was found according to the idea of *universal threshold*, due to Donoho and Johnstone [41]. Here, detail coefficients with a magnitude smaller than the expected maximum for an independent and identically distributed (Normal dist.) noise sequence were ignored:

$$\lambda = \sqrt{2 \log(n)} \hat{\sigma}, \quad (4.5)$$

where 2^n is the length of the original signal and σ is given by:

$$\sigma = \sqrt{\frac{1}{N/2-1} \sum_{i=1}^{N/2} (d_i^{(l)} - \bar{d})^2}, \quad (4.6)$$

where $d_i^{(l)}$ denotes the i^{th} wavelet coefficient at level l and \bar{d} is the mean of coefficients. Fig. 4.2 illustrates representations at different levels l , ($l \in \{0, 1, \dots, 10\}$) of an *IrisCode* \mathbf{s} . The coarsest representation $\mathbf{s}^{(10)}$ retains the lowest frequency components of the signature (intensities are stretched for visualization purposes) and is used in the root of the tree. The finest representation $\mathbf{s}^{(0)}$ is used in the leaves.

As Fig. 4.3 turns evident, $\mathbf{s}^{(l)}$ are increasingly smoothed versions of \mathbf{s} . The leftmost plot shows the average residuals between \mathbf{s} and its reconstructions at level l (horizontal axis), being evident that residuals increase directly with respect to the decomposition level. The center and rightmost plots give histograms of the residuals for the coarsest (center) and finest scales (right), enabling to perceive that the reconstruction at the coarsest scale is essentially a mean of the original signal.

4.1.2 Retrieval

As below described, one of the most relevant properties of the proposed method is its *non-exclusiveness* in terms of the paths in the tree traversed for a query. This *parallel* searching scheme is particularly important for our purposes, as it contributes for the robustness against degraded data.

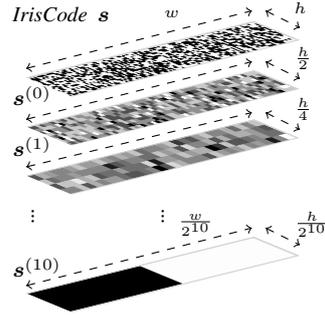


Figure 4.2: Representation of an *IrisCode* (upper image) at different levels, retaining coarse (bottom image) to fine information from the input code. The $s^{(10)}$ representation is used in the root of the tree and the remaining representations at the deeper levels. Intensities and sizes are stretched for visualization purposes.

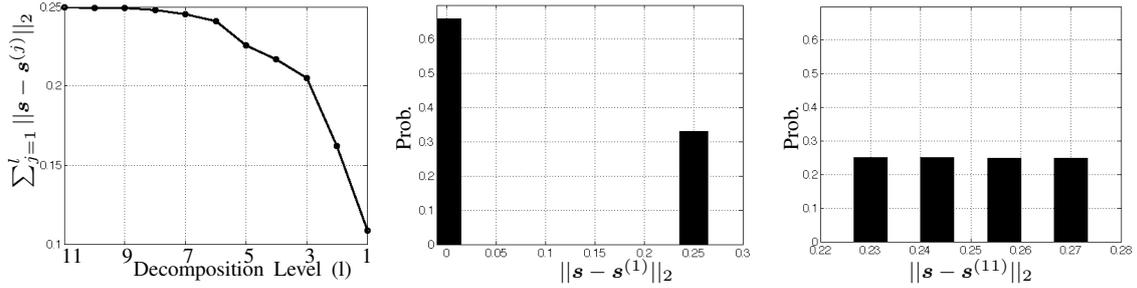


Figure 4.3: Average sum of residuals between an *IrisCode* s and its representations at different levels ($s^{(l)}$) (leftmost image). The images at the center and far right give the histograms of the residuals for decompositions/reconstructions at the finest (center) and coarsest (right) levels.

Retrieval requires a query signature s and a residual value $\xi > 0$. The idea is to iteratively decrease the residuals for each branch (ξ_j is the residual for the j^{th} branch of a node) and stop when $\xi_j < 0$. At each node, the ℓ_2 distance between the reconstructed signal $s^{(l)}$ and a cluster centroid c_i is subtracted from ξ_j , considering the maximum distance between c_i and all the identities in that branch. Formally, let $q(s, \xi^{(0)})$ be the query parameters. Let $s^{(l)}$ be the reconstruction of s at level l . The next generation of residual values $\xi^{(l+1)}$ is given by:

$$\xi^{(l+1)} = \xi^{(l)} - \max \left(0, \|s^{(l)} - c_i^{(l)}\|_2 - \max (\|s_b^{(l)} - c_i^{(l)}\|_2, \forall b \in \{1, \dots, t_i\}) \right), \quad (4.7)$$

being $c_i^{(l)}$ the i^{th} cluster at level l and $s_b^{(l)}$ the remaining signatures (total t_i) in that branch of the tree. The set of identities retrieved is given by:

$$q(\mathbf{s}, \xi^{(l)}) = \begin{cases} [\{i.\}, q(\mathbf{s}, \xi_j^{(l+1)})], \forall j, & \text{if } \xi^{(l)} > 0 \wedge l > 1 \\ \{i.\} & \text{if } \xi^{(l)} > 0 \wedge l = 1 \\ \emptyset & \text{if } \xi^{(l)} \leq 0 \end{cases} \quad (4.8)$$

where $[\cdot]$ denotes vector concatenation, $\xi_j^{(l)}$ denotes the residual value for the j^{th} branch at level l and $\{i.\}$ is the set of identities in a node.

Due to the intrinsic properties of wavelet decomposition, the distance values at the higher scales should be weighted by $w()$, as they represent more signal components:

$$w(l) = \frac{1 + \operatorname{erf}(\alpha(l - n))}{2}, \quad (4.9)$$

being α a parameter that controls the shape of the sigmoid. Fig. 4.4 shows one example of the histograms of the cuts in residuals ($\xi^{(l)} - \xi^{(l+1)}$, horizontal axis) with respect to the level in the tree. The dashed vertical lines indicate the cuts in the path that contained the identity of interest. Note that, with exception of the leaf level ($l = 1$), no cuts in the residuals were performed for the *interesting* path. This is in opposition to the remaining paths, where cuts occurred at all levels.

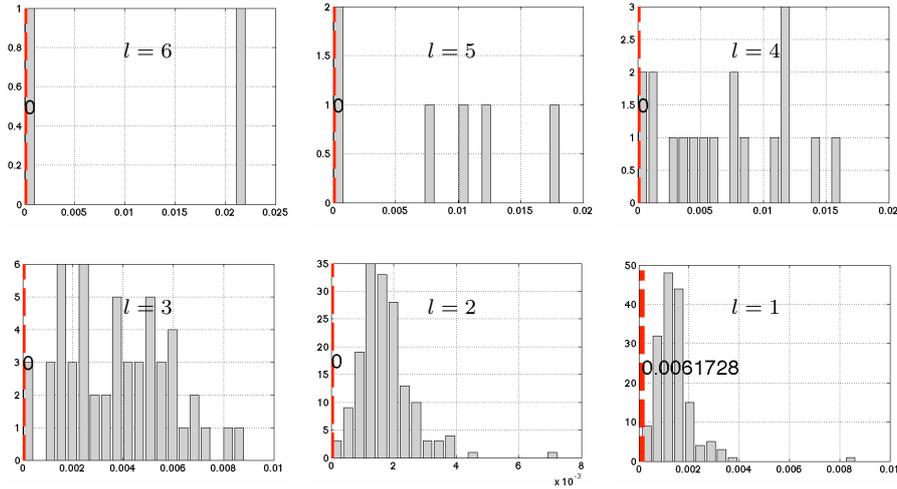


Figure 4.4: Histograms of the cuts in residuals $\xi^{(l)} - \xi^{(l+1)}$ per level during retrieval. The vertical dashed lines give the cumulative distribution values of the cuts in the path that contains the identity of interest. Gray bars express frequencies of the cuts in the remaining paths of the tree.

4.2 Identities Pruning: Soft Biometrics

The concept of soft biometrics has been recently regarded as an interesting possibility to compensate for the low quality of data acquired in less constrained protocols (e.g., visual surveillance). Soft traits are physical or behavioural human characteristics that have a limited discriminability between individuals, i.e., they are not enough to establish an identity, but can be combined or fused with strong biometric traits, in order to improve recognition accuracy or robustness.

At the bottom level, the idea behind soft biometrics can be regarded as a particular case of the concept of *multi biometrics*, which consists in using more than one source of information (trait) in the automatic recognition of a human being, and is largely discussed in the literature. As relevant examples,

Ross and Jain [138] reported a significant improvement in performance when summing the scores given by two different strong biometric experts, while Duca *et al.* [46] framed the problem according to the Bayes theory and estimated the biases of individual expert opinions, used to calibrate and fuse scores into a final decision. These works reported significant improvements in performance due to evidence fusion and did not pointed out any particular constraint about the individual performance of each fused expert. On the other way, as stated by Daugman [35], fusing different experts may not be good for all situations. Although the combination of tests enables the decision based in more information, on the other hand, if a stronger test is combined with a weaker one, the resulting decision environment is in some sense averaged, and the combined performance will lie somewhere between that of the two tests conducted individually. Accordingly, Poh and Bengio [124] analyzed four typical scenarios encountered in biometric recognition, mainly concerned about the issues of multi biometrics, having concluded that fusion is not always beneficial.



Figure 4.5: Key insight soft biometrics: in low quality data using soft biometric information might prune the space of plausible identities for a sample (image taken from [136]).

Moreover, one of the major advantages of soft biometric information is the intuitive relationship with the way humans recognise each other, based on a set of human compliant categories (Fig. 4.5). This is important not only to enable identification / retrieval based exclusively on a description given by an eyewitness, but also to provide intuitive justification for an identification, i.e., why did the system considered that a sample belongs to a particular subject. As pointed out by Reid *et al.* [135], there is a semantic gap between how humans and machines perform recognition and, when using most of the classical strong biometric traits, it is hard to provide a human understandable justification for a match / non-match reported by a machine.

There is nowadays a broad range of soft biometric traits, that can be divided into three groups, with respect to the region of the human body from where each trait is extracted [155]: 1) global traits refer to demographic information such as age, gender and ethnicity; 2) body traits are extracted from the whole body and describe the subject somatotype (e.g., height, figure, or proportions); and 3) head traits that are inferred exclusively based on images of the upper part of the subject, containing descriptions of the hair, beard and lips.

As in the case of other classical biometric traits (e.g., the iris or the face), it is important to perceive the degree of correlation between the labels of each soft biometric trait and scores that regard strong biometric traits, which at the ultimate level will justify the decision of using / not using the soft trait for a particular scenario. Also, the discriminating ability and the permanence of labels are other key factors that determine the usefulness of a soft trait.

**Identities Pruning Challenges:**

Retrieval based on disjoint excerpts of biometric signatures
High confidence pruning from soft biometric traits

4.3 Identities Pruning: Gait Recognition

The study of the human gait is an active research area, motivated by the difficulties faced by classical biometric traits in obtaining acceptable effectiveness in scenarios where the quality of the data is too low. For such conditions, it is considered that gait analysis can be one unobtrusive technology for detecting individuals that are possible threats to the community or that behave suspiciously [18].

Gait has several properties that turn it attractive, when compared to other biometric traits: even though it does not provide enough discriminating information to univocally establish an identity, it is obviously unobtrusive, i.e., unlike most strong biometric traits, it can be acquired from large distances and without requiring any active participation of the subjects, that can even be unaware of the recognition attempt. Moreover, it is extremely difficult to hide or fake, specially over large periods / distances. On the other side, while most biometric traits work in still images, most of the gait recognition algorithms rely on a video sequences, which significantly improves the computational burden of a recognition attempt. Moreover, previous studies report that gait changes over time and is particularly affected by clothes, footwear, walking surface, speed and emotional condition of the subject [147].

According to the above discussion, gait can be regarded from the *soft biometrics* perspective and use it mainly to prune the space of possible identities of a sample, narrowing down the database of possible matches, and letting the remaining identities be subject to the analysis of strong traits.

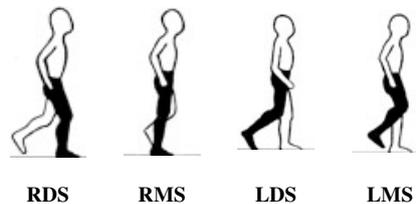


Figure 4.6: Four main phases of the human gait cycle, according to [28].

As illustrated in Fig. 4.6, the human walk is a cycled process, with four main phases [28]: 1) the right double support (RDS) phase, when both legs touch the ground, having the right leg on front; 2) the right mid stance (RMS), when legs are closest to each other, and the right leg touches the ground; 3) the left double support (LDS) and the 4) left mid stance (LMS). A *gait cycle* is composed by these four phases, while the corresponding time interval is called the *gait period*.

Published gait recognition algorithms can be coarsely divided into two strategies: 1) model-based approaches, which recover the human structure to provide information about walking dynamics (e.g., Lee and Grimson [93] and Gu *et al.* [58]); and 2) model-free approaches which directly analyse motion features from the image sequence (e.g., Han and Bhanu [60] and Iwama *et al.* [74]). Being more accurate, model-based methods are computationally more expensive and sensitive to appearance and occlusion issues than model-free methods, which turns them less adequate to work in degraded data. Several algorithms particularly suitable for surveillance of environments were recently proposed, either aiming at walking speed invariance (e.g., Priydarshi *et al.* [125]), clothes invariance (e.g., Hossain *et al.* [70]), view invariance (e.g., Goffredo *et al.* [54]) and robustness to limited data resolution (e.g., Zhang *et al.* [175]).

	<p>Gait Recognition Main Challenges:</p> <ul style="list-style-type: none"> Cross-view recognition Detection of counterfeit gaits Real-time gait recognition
---	--

4.4 Iris Segmentation

	<p>Hugo Proença; Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength, IEEE Transactions on Pattern Analysis and Machine Intelligence, August, 2010, volume 32, number 8, pag. 1502-1516, ISSN: 0162-8828, Digital Object Identifier 10.1109/TPAMI.2009.140</p>
---	---

We note that a significant majority of the listed methods operate on NIR images that typically offer high contrast between the pupil and the iris regions, which justifies the order in which the borders are segmented. Also, various innovations have recently been proposed, such as the use of active contour models, either geodesic ([139]), based on Fourier series ([38]) or based on the snakes model ([7]). These techniques require previous detection of the iris to properly initialize contours, and are associated with heavy computational requirements. Modifications to known form fitting methods have also been proposed, essentially to handle off-angle images (e.g., [181] and [158]) and to improve performance (e.g., [98] and [40]). Finally, the detection of non-iris data that occludes portions of the iris ring has motivated the use of parabolic, elliptical and circular models (e.g., [11], and [40]) and the modal analysis of histograms [38]. Even so, in noisy conditions, several authors have suggested that the success of their methods is limited to cases of image orthogonality, to the non-existence of significant iris occlusions, or to the appearance of corneal reflections in specific image regions.

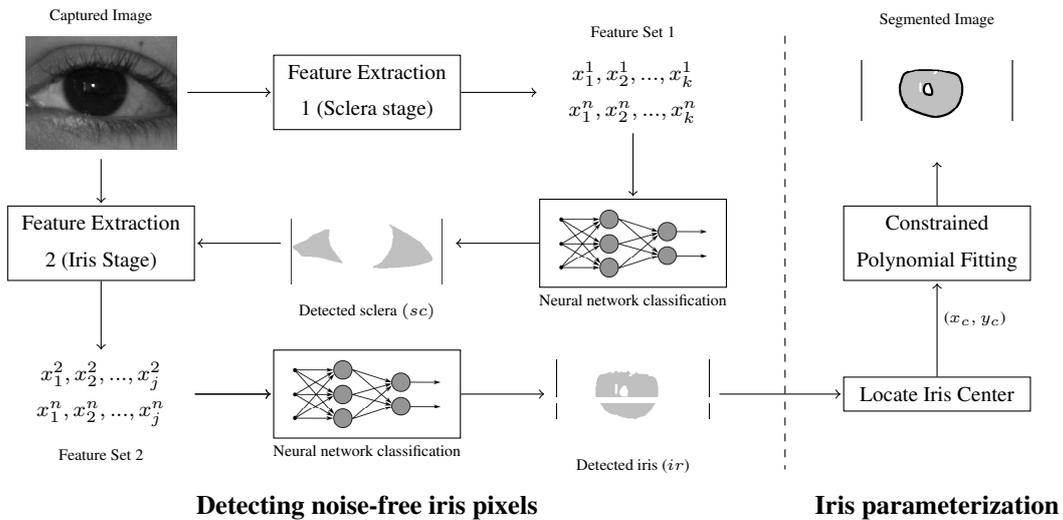


Figure 4.7: Block diagram of our iris segmentation method.

Fig. 4.7 shows a block diagram of our segmentation method, which can be divided into two parts: detecting noise-free iris regions and parameterizing the iris shape.

The initial phase is further sub-divided into two processes: detecting the sclera, and detecting the iris. The key insight is that the sclera is the most easily distinguishable region in non-ideal images. Next, we exploit the mandatory adjacency of the sclera and the iris to detect noise-free iris regions. We stress that the whole process comprises three tasks that are typically separated in the literature: iris detection, segmentation and detection of noisy (occluded) regions. The final part of the method is to parameterize the detected iris region. In our tests, we often observed small classification inaccuracies near iris borders.

We found it convenient to use a constrained polynomial fitting method that is both fast and able to adjust shapes with an arbitrary degree of freedom, which naturally compensates for these inaccuracies.

4.4.1 Feature Extraction Stages

We used local features to detect the sclera and noise-free iris pixels. Due to performance concerns, we decided to evaluate only those features that a single image scan can capture. Viola and others [159] proposed a set of simple features (reminiscent of Haar basis functions) and computed them over a single image scan with an intermediate image representation. For a given image I , they defined an *integral image*:

$$II(x, y) = \sum_{x'=1}^x \sum_{y'=1}^y I(x', y'), \quad (4.10)$$

where x denotes the image column and y denotes the row. They also proposed a pair of recurrences to compute the integral image in a single image scan:

$$s(x, y) = s(x, y - 1) + I(x, y) \quad (4.11)$$

$$II(x, y) = II(x - 1, y) + s(x, y), \quad (4.12)$$

with $s(x, 0) = II(0, y) = 0$. According to this concept, the average intensity (μ) within any rectangular region R_i , delimited by its upper-left (x_1, y_1) and bottom-right (x_2, y_2) corner coordinates, is determined by accessing just four array references. Let $T_i = (x_2 - x_1 + 1) \times (y_2 - y_1 + 1)$ be the number of pixels within R_i . Then

$$\mu(R_i) = \frac{1}{T_i} \left(II(x_2, y_2) + II(x_1, y_1) - II(x_2, y_1) - II(x_1, y_2) \right). \quad (4.13)$$

Similarly, the standard deviation (σ) of the intensities within R_i is given by

$$\sigma(R_i) = \sqrt{\mu(R_i^2) - \mu(R_i)^2}, \quad (4.14)$$

where $\mu(R_i)$ is given by (4.13) and $\mu(R_i^2)$ is obtained similarly, starting from an image with squared intensity values. According to (4.13) and (4.14), the feature sets used in the detection of the sclera and the noise-free iris regions are central moments computed locally within regions of varying dimension of different color spaces.

4.4.2 Sclera Stage

When examining degraded eye images, the iris region can be hard to discriminate, even for humans. Also, the sclera is much more naturally distinguishable than any other part of the eye, which is a key insight: our process detects pixels that belong to the sclera and, later, we exploit their mandatory adjacency with the iris in order to find the iris.

Our empirical analysis of different color spaces led to the selection of the hue (h), blue (cb) and red chroma (cr) color components. These serve to maximize the contrast between the sclera and the remaining parts of the eye, as illustrated in Fig. 4.8. Using the previously described average (4.13) and standard deviation (4.14) values, we extracted a 20-dimensional feature set for each image pixel:

$\{x, y, h_{0,3,7}^{\mu,\sigma}(x, y), cb_{0,3,7}^{\mu,\sigma}(x, y), cr_{0,3,7}^{\mu,\sigma}(x, y)\}$, where x and y denote the position of the pixel and $h(\cdot)$,

$cb()$, and $cr()$ denote regions (centered at the given pixel) of the hue, blue and red chroma color components. The subscripts denote the radii used (e.g., $h_{0,3,7}^{\mu,\sigma}(x, y)$ means that six features were extracted from regions of the hue color component: three averages and three standard deviations computed locally within regions of radii 0, 3 and 7).

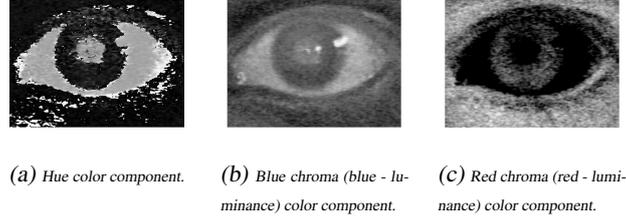


Figure 4.8: Discriminating between the regions that belong to the sclera and all the remaining types of information given by the hue (figure 4.8a), blue chroma (figure 4.8b) and red chroma (figure 4.8c) color components.

4.4.3 Iris Stage

The human eye's morphology dictates that any pixel inside the iris should either have an approximately equal amount of sclera to its left and right if the iris is frontally imaged, or have a much higher value at one of its sides if the iris was imaged off-axis. In any case, the number of sclera pixels in the upper and lower directions should be minimal if the image was acquired from standing subjects without major head rotations.

We used data obtained in the sclera detection stage ("Detected sclera" of figure 4.7) to extract a new type of feature, called "proportion of sclera" $p(x, y)$, for each image pixel. This feature measures the proportion of pixels that belong to the sclera in direction d with respect to the reference pixel (x, y) (in the experiments, the four main directions north \uparrow , south \downarrow , east \rightarrow , west \leftarrow were used). From (4.13), the result is given by:

$$p_{\leftarrow}(x, y) = \mu\left(sc\left((1, y - 1), (x, y)\right)\right) \quad (4.15)$$

$$p_{\rightarrow}(x, y) = \mu\left(sc\left((x, y - 1), (w, y)\right)\right) \quad (4.16)$$

$$p_{\uparrow}(x, y) = \mu\left(sc\left((x - 1, 1), (x, y)\right)\right) \quad (4.17)$$

$$p_{\downarrow}(x, y) = \mu\left(sc\left((x - 1, y), (x, h)\right)\right), \quad (4.18)$$

where $sc((., .), (., .))$ denotes regions of the image that feature the detected sclera (Figs. 4.9a and 4.9d), delimited by their top-left and bottom-right corner coordinates. w and h are the image width and height. By definition, the value of $p()$ was set to 0 for all the sclera pixels. Fig. 4.9 illustrates the $p_{\leftarrow}(x, y)$ and $p_{\rightarrow}(x, y)$ feature values for a frontal image in the upper row and an off-angle image in the lower row. You can see that in both cases, the simple overlap of the feature values almost optimally delimits the iris region.

These "proportion of sclera" values, the pixel position, the local image saturation and blue chrominance (obtained similarly to the previous feature extraction stage) are computed to yield a 18-dimension feature set: $\{x, y, s_{0,3,7}^{\mu,\sigma}(x, y), cb_{0,3,7}^{\mu,\sigma}(x, y), p_{\leftarrow, \rightarrow, \uparrow, \downarrow}(x, y)\}$. Again, we selected the color spaces empirically, according to the contrast between the sclera and the iris, as illustrated in Fig. 4.10. $s()$, and $cb()$ denote regions of the saturation and blue chrominance color components. As in the previously described

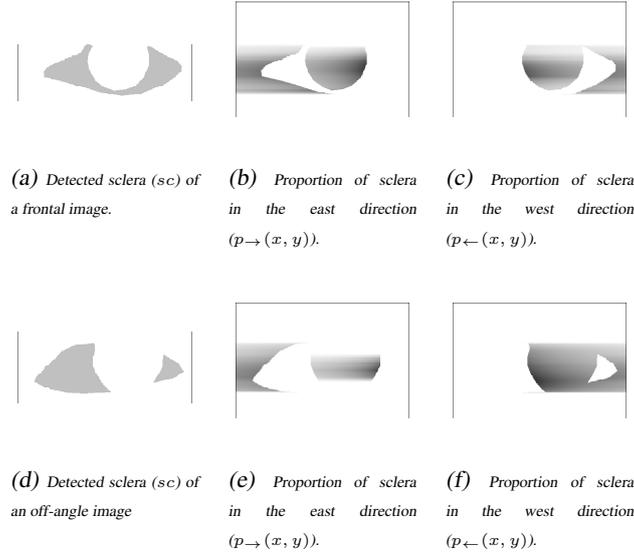


Figure 4.9: "Proportion of sclera" values towards the west ($p_{\leftarrow}(x, y)$) and east ($p_{\rightarrow}(x, y)$), obtained from the detected sclera of a frontal (upper row) and an off-angle (lower row) Image. For visualization purposes, darker pixels represent higher values.

feature extraction stage (sclera detection), the subscripts give the radii we used, centered at the given pixel.

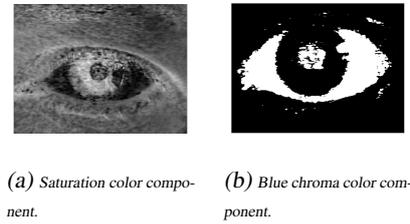


Figure 4.10: Color components used in iris detection.

4.4.4 Supervised Machine Learning and Classification

Both classifiers in our method operate at the pixel level and perform binary classification. For these, we evaluated several alternatives according to three fundamental learning theory issues: model capacity, computational and sample complexity. We were mindful of heterogeneity and the amount of data available for learning purposes, which justified the use of neural networks. We know that these types of classifiers can form arbitrarily complex decision boundaries. Thus, the model capacity is good. Also, the back-propagation learning algorithm propitiates good generalization capabilities using a relatively small amount of learning data.

As shown in Fig. 4.11, we used multi-layered perceptron feed-forward neural networks with one hidden layer for both classification stages, not considering the input nodes as a layer. All the networks feature as many neurons in the input layer (k_1) as the feature space dimension, (k_2) neurons in the hidden layer and a single neuron in the output layer. As transfer functions, we used the sigmoid hyperbolic tangent on the first two layers and pure linear on the output. Several parameters affect the networks'

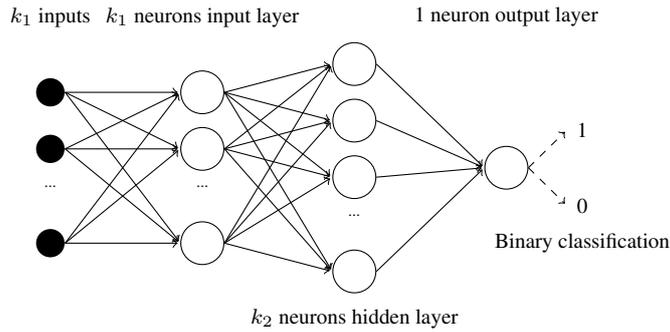


Figure 4.11: Schema for the multi-layered feed-forward neural networks used in both classification stages of our segmentation method.

results, such as the number of neurons used in the hidden layer, the amount of data used for learning and the learning algorithm. During the experimental period, we varied most of these parameters, to arrive at the optimal values as reported in the experiments section.

4.4.5 Shape Parameterization

Efficient shape parameterization is a key issue for post-segmentation recognition stages. With a set of image pixels that are classified as *noise-free iris*, the goal is to parametrically approximate the contour of the pupillary and scleric iris borders. Recently, researchers have proposed using active contour and spline techniques for this type of task, although they were not considered the most convenient for the purposes of our work, essentially due to performance concerns. Instead, we performed a polynomial regression on a polar coordinate system, which runs naturally fast and compensates for inaccuracies from the previous classification stage, as illustrated in Fig. 4.12. The process starts by roughly localizing the iris center. The center serves as a reference point in the translation into a polar coordinate system, where we perform the polynomial regression. Remapping the obtained polynomials into the original Cartesian space gives the parameterization of the pupillary and scleric iris borders.

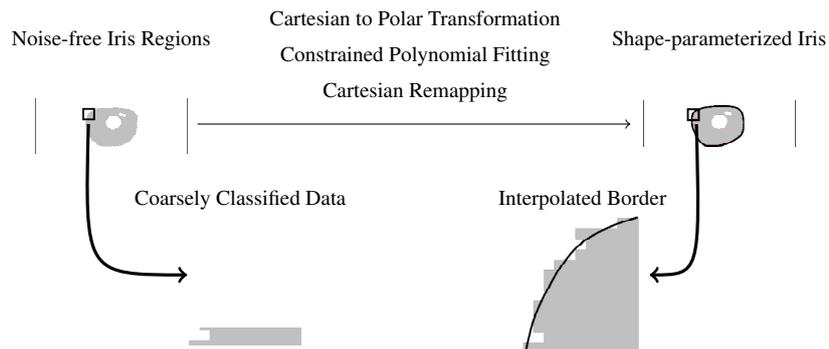


Figure 4.12: Parameterizing segmented noise-free iris regions through constrained polynomial fitting techniques.

The iris and pupil are not concentric, although their centers are not distant from one another. We identify a pixel (x_c, y_c) that roughly approximates these centers and use it as a reference point. Let B be a binary image that distinguishes between the noise-free iris regions and the remaining types of data

(figure 4.9d). Let $C = \{c_1, \dots, c_w\}$ be the cumulative vertical projection of B and $R = \{r_1, \dots, r_h\}$ be the horizontal projection: that is, $c_i = \sum_{j=1}^h B(i, j)$ and $r_i = \sum_{j=1}^w B(j, i)$. Since the iris regions are darker, the values of c_i and r_i decrease in the rows and columns that contain the iris, as illustrated in Fig. 4.13

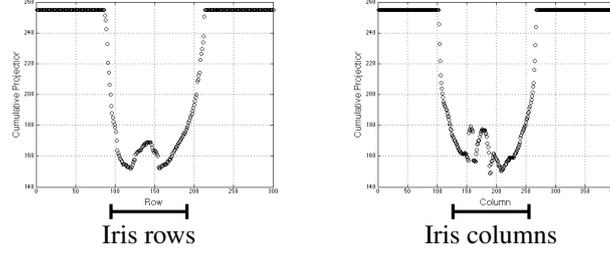


Figure 4.13: Horizontal and vertical cumulative projections of the iris image (*ir*) illustrated in Fig. 4.12.

Let $C^* = \{c_{1^*}, \dots, c_{m^*}\}$ be a subset containing the first-quartile elements of C_i , and $R^* = \{r_{1^*}, \dots, r_{n^*}\}$ be a subset containing the first-quartile elements of R_i , which correspond to the darkest columns and lines of the binary image. An approximation to the iris center (x_c, y_c) is given by the median values of C^* and R^* : that is, $x_c = c_{\frac{m^*}{2}}$ and $y_c = r_{\frac{n^*}{2}}$. We measure the distance between (x_c, y_c) and the pixels classified as *iris* along θ_i directions, such that $\theta_i = \frac{i2\pi}{t}$, $i = 1, \dots, t-1$. The highest value in each direction approximates the distance between the contour of the iris and the reference pixel (x_c, y_c) , as illustrated in Figs. 4.14a and 4.14b (Cartesian and polar coordinate systems). A set of simple semantic rules keeps incompletely closed pupil or iris shapes from degrading the process. The simplest rule is that contour points should be within the interval $[l_1, l_2]$. The regression procedure discards values outside this interval.

Hereafter, we regard the problem as a polynomial regression. We could use other shape-fitting techniques at this stage with similar results, but we chose this approach for its lower computational requirements. Given a set of t data points (x_i, y_i) , the goal is to optimize the parameters of a k^{th} degree polynomial $p(x) = a_0 + a_1x + \dots + a_kx^k$ so as to minimize the sum of the squares of the deviations S^2 :

$$S^2 = \sum_{i=1}^t (y_i - p(x_i))^2, \quad (4.19)$$

where y_i is the desired value at x_i and $p(x_i)$ is the response value at x_i . To guarantee a closed contour of the iris border in the Cartesian coordinate system, we must ensure that $p(x_1) = p(x_t)$, which gives rise to an equality constrained least squares problem [63]. The goal is to find a vector $x \in R^k$ that minimizes $\|Ax - b\|_2$, subject to the constraint $Bx = d$, assuming that $A \in R^{m \times k}$, $B \in R^{p \times k}$, $b \in R^m$, $d \in R^p$ and $rank(B) = p$. Here, A refers to the iris boundary points that are to be fitted and B is the constraint that guarantees a closed contour. Considering that the null spaces of A and B intersect only trivially, this problem has a unique solution x^* . As Loan describes [99], a possible solution is obtained through the elimination method, which uses the constraint equation to solve for m elements of b in terms of the remaining ones. The first step to the solution is to find an orthogonal matrix Q such that $Q^T B^T$ is upper triangular:

$$Q^T B^T = \begin{bmatrix} R_B \\ 0 \end{bmatrix}. \quad (4.20)$$

Next, we solve the system $R_B^T y_1 = d$ and set x_1 to $Q_1 y_1$, where $Q = [Q_1 Q_2]$, $Q_1 \in R^p$ and

$Q_2 \in R^{k-p}$. Again, we find an orthogonal matrix U such that $U^T(AQ_2)$ is upper triangular:

$$U^T(AQ_2) = \begin{bmatrix} R_A \\ 0 \end{bmatrix}. \quad (4.21)$$

We set $R_{Ay_2} = U_1^T(b - Ax_1)$ and $x_2 = Q_2y_2$, where $U = [U_1U_2]$, $U_1 \in R^{k-p}$ and $U_2 \in R^{m-k+p}$. Finally, the solution is given by

$$x^* = x_1 + x_2. \quad (4.22)$$

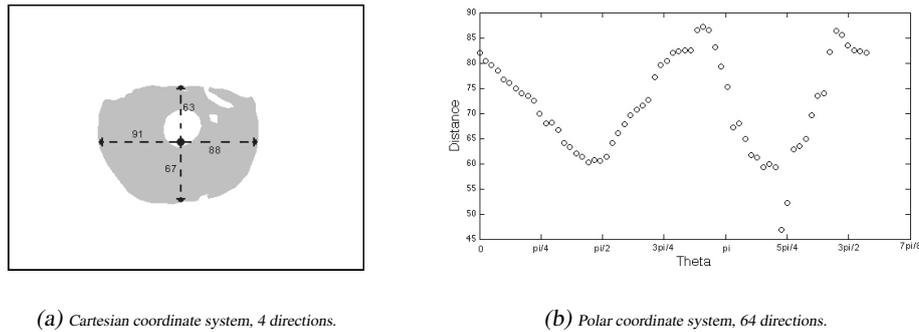


Figure 4.14: Largest distances between the iris center and the pixels classified as *iris* along θ directions, in the Cartesian (figure 4.14a) and in the polar (figure 4.14b) coordinate systems. The continuous line gives the 10th degree constrained polynomial for the purposes of data regression.

4.4.6 Computational Complexity

As noted previously, the computational complexity of the given segmentation method is a major concern for real-time data handling. The first part of the method operates at the pixel level, and all the corresponding operations receive as input all the image pixels: either their RGB, intensity or feature vectors. Let I be a RGB image with $n = c \times r$ pixels (typically $120\,000 = 400 \times 300$ in the experiments). Given this relatively large value, we must maintain an asymptotic upper bound on execution time that is linear in the size of the input, ensuring that the first stage of the method (and the most time-consuming) runs quickly. Thereafter, the parameterization of the iris borders depends on the number of directions from which reference points are picked, and on the polynomial degree. As these values are relatively low (in our experiments, the number of directions is 64 and the degree is 10), increased computational complexity is not a concern since it will not significantly lower the method's performance. Also, we emphasize that our method offers roughly deterministic performance, that its performance is linear in image size, and that it is significantly faster than other segmentation methods for similar scenarios.

	<p>Iris Segmentation Main Challenges:</p> <ul style="list-style-type: none"> Glossy reflections due to the cornea Low contrast pupillary boundary Real-time iris segmentation
---	---

4.5 Periocular Segmentation

	<p>Hugo Proença, João C. Neves and Gil Santos; "Segmenting the Periocular Region using a Hierarchical Graphical Model Fed by Texture / Shape Information and Geometrical Constraints" in Proceedings of the International Joint Conference on Biometrics - IJCB 2014, pag. 1-8, Clearwater, Florida, U.S.A., September 29- October 2, Digital Object Identifier: ?, 2014</p>
---	--

Most of the relevant periocular recognition algorithms work in a *holistic* way, i.e., they define a region-of-interest (ROI) around the eye and apply a feature encoding strategy independently of the biological component at each position. The exceptions (e.g., [152] and [30]) regard the iris and the sclera components, for which specific feature encoding / matching algorithms are used. This observation leads that some components (e.g., hair or glasses) might be erroneously taken into account and bias the recognition process.

The automatic labelling (segmentation) of the components in the periocular region has - at least - two obvious advantages: it enables to define better ROIs and conducts to more accurate estimates of subjects' pose and gaze. Hence, an image labelling algorithm for the periocular region was proposed, in order to discriminate between seven components in the periocular region: iris, sclera, eyelashes, eyebrows, hair, skin and glasses. The model is generated on two phases:

1. seven non-linear classifiers running at the pixel level are inferred from a training set, and provide the posterior probabilities for each image position and class of interest. Each classifier (neural network) is specialized in detecting one component and receives local statistics (texture and shape descriptors) from the input data;
2. the posteriors based on data local *appearance* are combined with geometric constraints and components' adjacency priors, to feed a hierarchical Markov Random Field (MRF), composed of a *pixel* and a *component* layer. MRFs are a classical tool for various computer vision problems, from image segmentation (e.g., [85]), image registration (e.g., [53]) to object recognition (e.g., [25]). Among other advantages, they provide non-causal models with isotropic behavior and faithfully model a broad range of local dependencies. The model proposed inherits some insights from previous works that used shape priors to constraint the final model (e.g., [14]) and multiple layered MRFs (e.g., [163]).

As Fig. 4.15 illustrates, the proposed MRF is composed of two layers: one works at the *pixel* level, with a bijection between each image pixel and a vertex in the MRF. The second layer regards the major *components* in the periocular vicinity, with six vertices representing the eyebrows, irises and corneas from both sides of the face. The insight behind this structure is that the pixels layer mainly regards the data appearance, while the components layer represents the geometrical constraints in the problem and assures that the generated solutions are biologically plausible.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph representing a MRF, composed of a set of t_v vertices \mathcal{V} , linked by t_e edges \mathcal{E} . Let t_p be the number of vertices in the *pixels* layer and let t_c be the number of vertices in the *components* layer, such that $t_v = t_p + t_c$. Let $\mathcal{C}(x, y)$ denote the biological component at position (x, y) of an image and \mathcal{T}_j be the component's *type* of the j^{th} component node: either 'iris', 'cornea' or 'eyebrow'.

The MRF is a representation of a discrete latent random variable $\mathbf{L} = \{L_i\}, \forall i \in \mathcal{V}$, where each element L_i takes one value l_i from a set of labels. Let $\mathbf{l} = \{l_1, \dots, l_{t_p}, l_{t_p+1}, \dots, l_{t_p+t_c}\}$ be one configuration of the MRF. In our model, every component node is directly connected to each pixel node and the pixel nodes are connected to their horizontal / vertical neighbors (4-connections). Also, the edges between component nodes correspond to geometrical / biological constraints in the periocular region: the nodes representing both irises, corneas and eyebrows are connected, as do the iris, cornea and eyebrow nodes of the same side of the face. Note that the proposed model does not use high-order potentials. Even though there is a point in Fig. 4.15 that joins multiple edges, it actually represents overlapped pairwise connections between one component and one pixel vertex.

The energy of a configuration \mathbf{l} of the MRF is the sum of the unary $\theta_i(l_i)$ and pairwise $\theta_{i,j}(l_i, l_j)$ potentials:

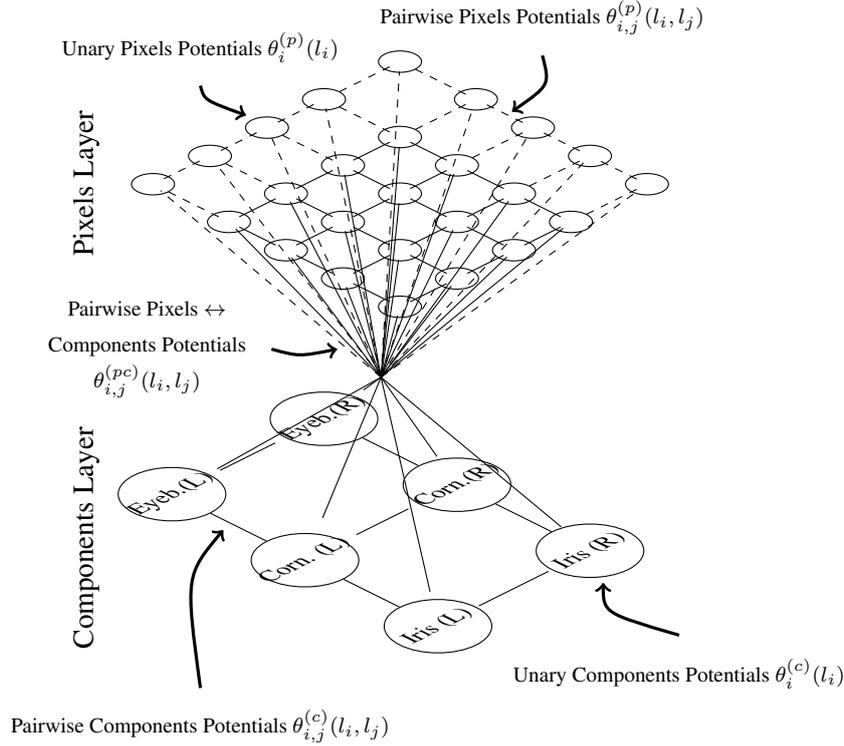


Figure 4.15: Structure of the MRF that segments the periocular region.

$$E(\mathbf{l}) = \sum_{i \in \mathcal{V}} \theta_i(l_i) + \sum_{(i,j) \in \mathcal{E}} \theta_{i,j}(l_i, l_j). \quad (4.23)$$

According to this formulation, labelling an image is equivalent to infer the random variables in the MRF by minimizing its energy:

$$\hat{\mathbf{l}} = \arg \min_{\mathbf{l}} E(\mathbf{l}), \quad (4.24)$$

where $\{\hat{l}_1, \dots, \hat{l}_{t_p}\}$ are the labels of the pixels and $\{\hat{l}_{t_p+1}, \dots, \hat{l}_{t_p+t_c}\}$ specify the components' parameterizations. In this work, the MRF was optimized according to the Loopy Belief Propagation [50] algorithm. Even though it is not guaranteed to converge to global minimums on loopy non submodular graphs (such as our MRF), we concluded that the algorithm provides visually pleasant solutions most of the times. As future work, we plan to evaluate the effectiveness of our model according to more sophisticated energy minimization algorithms (e.g., sequential tree-reweighted message passing [88]).

4.5.1 Feature Extraction

Previous works reported that the hue and saturation channels of the HSV color space are particularly powerful to detect the sclera [129], whereas the red / blue chroma values provide good separability between the skin and non-skin pixels [4]. Also, the iris color triplets are typically distant from the remaining periocular components and there is a higher amount of information in patches of the eyebrows

and hair regions than in the remaining components. Accordingly, a feature set at the pixel level is extracted, composed of 34 elements (Fig. 4.16): {red, green and blue channels (RGB); hue, saturation and value channels (HSV); red and blue chroma (yCbCr); LBP and entropy in the value channel}, all averaged in square patches of side $\{3, 5, 7\}$ around the central pixel. Also, the convolution between the value channel and a set of Gabor kernels \mathbf{G} complements the feature set:

$$\mathbf{G}[x, y, \omega, \varphi, \sigma] = \exp\left[-\frac{x^2 - y^2}{\sigma^2}\right] \exp[2\pi\omega i\Phi], \quad (4.25)$$

being $\Phi = x \cos(\varphi) + y \sin(\varphi)$, ω the spatial frequency, φ the orientation and σ the standard deviation of an isotropic Gaussian kernel ($\omega \in \{\frac{3}{2}, \frac{5}{2}\}$, $\varphi \in \{0, \frac{\pi}{2}\}$, $\sigma = 0.65\omega$).

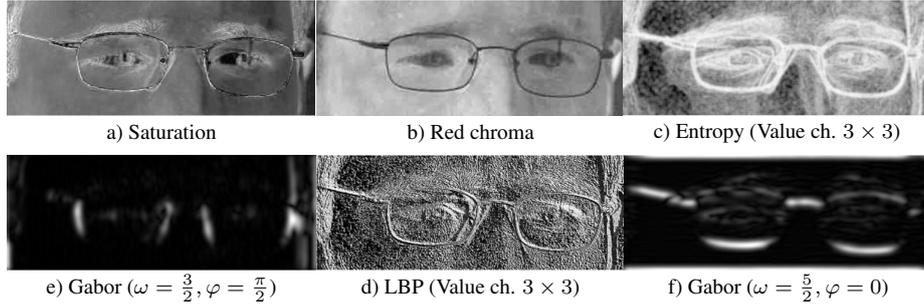


Figure 4.16: Illustration of the discriminating power of the features extracted, for the seven classes considered in the periocular region.

4.5.2 Unary Potentials

Let $\gamma : \mathbb{N}^2 \rightarrow \mathbb{R}^{34}$ be the feature extraction function, that for each image pixel (x, y) returns a feature vector $\gamma(x, y) \in \mathbb{R}^{34}$. Let $\Gamma = [\gamma(x_1, y_1), \dots, \gamma(x_n, y_n)]^T$ be a $n \times 34$ matrix extracted from a training set, that is used to learn seven non-linear binary classification models, each one specialized in detecting a component (class) $\omega_i \in \{\text{Iris, Sclera, Eyebrows, Eyelashes, Hair, Skin, Glasses}\}$. Let $\eta_i : \mathbb{R}^{34} \rightarrow [0, 1]$ be the response of the i^{th} non-linear model, used to obtain the likelihood of class ω_i : $p(\eta_i(\gamma(x, y)) | \omega_i)$. According to the Bayes rule, assuming equal priors, the posterior probability functions are given by:

$$P(\omega_i | \eta_i(\gamma(x, y))) = \frac{P(\eta_i(\gamma(x, y)) | \omega_i)}{\sum_{j=1}^7 P(\eta_j(\gamma(x, y)) | \omega_j)}. \quad (4.26)$$

The unary potentials of each vertex in the pixels layer are defined as $\theta_i^{(p)}(l_i) = 1 - p(\omega_i | \eta_i(\gamma(x, y)))$.

Each label in the components layer represents a parameterisation of an ellipse (found by the Random Elliptical Hough Transform (REHT)) [10] that roughly models the eyebrows, corneal or iris regions. Starting from images labelled by the index of the maximum posterior probability $I_m(x, y) = \arg \max_j p(\omega_j | \eta_j(\gamma(x, y)))$ (upper image in Fig. 4.17), a binary version per component can be obtained (bottom images in Fig. 4.17):

$$I_{m_i}(x, y) = \begin{cases} 1 & , \text{if } I_m(x, y) = i \\ 0 & , \text{otherwise} \end{cases} \quad (4.27)$$

The output of the REHT algorithm in $I_{m_i}(x, y)$ gives the unary potential of the component vertices: $\theta_i^{(c)}(l_i) = -\log(\kappa(i)), \forall i \in t_{p+1}, \dots, t_{p+c}$, being $\kappa(i)$ the votes returned by the REHT for the i^{th} ellipse parameterisation.

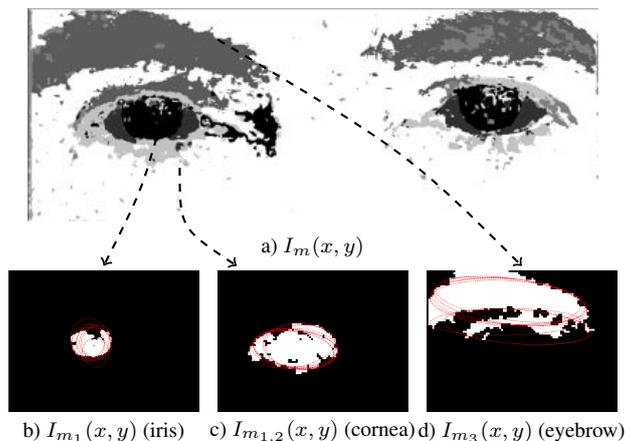


Figure 4.17: (Upper row) Example of an image labelled by the maximum of the posteriors given by the classification models $\eta_i(\gamma(x, y))$. The red ellipses in the bottom images represent the parameterisations returned by the REHT algorithm for the left iris, cornea and eyebrow.

4.5.3 Pairwise Potentials

There are three types of pairwise potentials in our model: 1) between two pixel nodes; 2) between two component nodes; and 3) between a pixel and a component. The pairwise potential between pixel nodes spatially adjacent $\theta_{i,j}^{(p)}(l_i, l_j)$ is defined as the prior probability of observing labels l_i, l_j in adjacent positions of a training set (e.g., it is much more probable that an "eyebrow" pixel is adjacent to a "skin" pixel than to an "iris" one):

$$\theta_{i,j}^{(p)}(l_i, l_j) = \frac{1}{\alpha_0 + P(\mathfrak{C}(x', y') = \omega_i, \mathfrak{C}(x, y) = \omega_j)}, \quad (4.28)$$

where $P(., .)$ is the joint probability, (x', y') and (x, y) are 4adjacent positions and $\alpha_0 \in \mathbb{R}^+$ avoids infinite costs (likewise, all α_i terms below are regularization terms).

The pairwise potentials between component nodes consider the geometrical constraints in the periorcular area, i.e., enforce that the irises are inside the cornea, and below the eyebrows. Also, both irises, corneas and eyebrows should have similar vertical coordinate and similar size. Let $(x_i, y_i, a_i, b_i, \varphi_i)$ be the i^{th} parameterisation of an ellipse, being (x_i, y_i) the ellipse centre, (a_i, b_i) its major / minor axes and φ_i the rotation. For pairs of nodes of the same type ($\mathfrak{X}_i = \mathfrak{X}_j$), similar vertical coordinates and similar sizes are privileged:

$$\theta_{i,j}^{(c1)}(l_i, l_j) = \alpha_1 |y_i - y_j| + \alpha_2 |a_i + b_i - a_j - b_j|. \quad (4.29)$$

For edges connecting the cornea (i^{th} node) and the eyebrow (j^{th} node) we privilege similar horizontal coordinates and locations having the eyebrow above the cornea:

$$\theta_{i,j}^{(c2)}(l_i, l_j) = \alpha_3 |x_i - x_j| + \alpha_4 \max\{0, y_i - y_j\}. \quad (4.30)$$

Regarding the iris / cornea pairwise potentials, we penalize parameterizations with portions of the iris outside the cornea:

$$\theta_{i,j}^{(c3)}(l_i, l_j) = \alpha_5 \left(1 - \frac{\sum_{x_i} \sum_{y_i} \psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j)}{\sum_{x_i} \sum_{y_i} 1} \right), \quad (4.31)$$

being (x_i, y_i) a pixel labelled as iris and $\psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j)$ an indicator function that verifies if that position is inside the ellipse defined by the j^{th} parameterisation (4.32).

$$\psi(x, y, x_i, y_i, a_i, b_i, \varphi_i) = \begin{cases} 1 & , \text{ if } \frac{(\cos(\varphi_i)(x-x_i) + \sin(\varphi_i)(y-y_i))^2}{a_i^2} + \frac{(\sin(\varphi_i)(x-x_i) + \cos(\varphi_i)(y-y_i))^2}{b_i^2} \leq 1 \\ 0 & , \text{ otherwise} \end{cases} \quad (4.32)$$

Overall, the pairwise potentials in the components layer are defined as:

$$\theta_{i,j}^{(c)}(l_i, l_j) = \sum_{k=1}^3 \theta_{i,j}^{(ck)}(l_i, l_j). \quad (4.33)$$

Lastly, the pairwise potentials between pixels and components enforce that pixels inside a component parameterisation are predominantly labelled by the value that corresponds to that type of node, whereas pixels outside that parameterisation should have label different of the component's type. Let (x_{jk}, y_{jk}) be the coordinates of the ellipse defined by the j^{th} parameterization. The pairwise cost between the i^{th} pixel node and the j^{th} component node is given by:

$$\theta_{i,j}^{(pc)}(l_i, l_j) = \begin{cases} \min_k \|(x_i, y_i) - (x_{jk}, y_{jk})\|_2, & \text{if } l_i \in \mathfrak{T}_j \\ & \text{and } \psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j) = 0 \\ 0, & \text{if } l_i \notin \mathfrak{T}_j \\ & \text{and } \psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j) = 0 \\ 0, & \text{if } l_i \in \mathfrak{T}_j \\ & \text{and } \psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j) = 1 \\ \max_k \|(x_i, y_i) - (x_{jk}, y_{jk})\|_2, & \text{if } l_i \notin \mathfrak{T}_j \\ & \text{and } \psi(x_i, y_i, x_j, y_j, a_j, b_j, \varphi_j) = 1 \end{cases}, \quad (4.34)$$

where $\|\cdot\|$ is the Euclidean distance.

	<p>Periocular Segmentation Main Challenges:</p> <ul style="list-style-type: none"> Samples with sun glasses Changes in samples roll Reduce the computational cost of segmentation
---	---

4.6 Iris Recognition: Quality Assessment

	<p>Hugo Proença; Quality Assessment of Degraded Iris Images Acquired in the Visible Wavelength, IEEE Transactions on Information Forensics and Security, March, 2011, volume 6, issue 1, pag. 82-95, ISSN 1556-6013, Digital Object Identifier http://doi.ieeecomputersociety.org/10.1109/TIFS.2010.2086446</p>
---	--

The concept of *good* metric for an iris image is not trivial, although the best one should maximally correlate with recognition effectiveness. Previous studies reported significant decays in effectiveness when data is degraded. There are several previously published methods, with differences in the used spectrum of light, the type of analyzed data (*raw* image, *segmented* or *normalized* iris region) and their output (*local* or *global*), as they operate at the pixel or image level. We note that most of the methods operate in NIR images and assess quality in the segmented data (either in the cartesian or polar coordinate systems). Exceptions are usually related with focus measurement, obtained by one of two approaches: (1) measuring the high frequency power in the 2D Fourier spectrum through a high-pass convolution kernel or wavelet-based decomposition (2) analyzing the sharpness of the iris borders through the magnitude of the first and second order derivatives. Another key characteristic is the level of analysis: some methods operate globally (at the image level), usually to determine focus, gaze or motion blur. As image quality varies across the iris, others operate at the pixel level to determine local obstructions. Motion is estimated by detecting interlaced raster shear that might be due to significant movements during the acquisition of a frame. Other approaches rely in the response of the convolution between the image and directional filters, being observed that linear motion blurred images have higher central peak responses than sharp ones. Gaze is estimated by 3D projection techniques that maximize the response of the Daugman's integro-differential operator or by the length of the axes of a bounding ellipse. Eyelids are detected by means of line and parabolic Hough transforms, active contours and machine learning frameworks. The modal analysis of the intensities histogram enables the detection of eyelashes, as do spectral analysis and edge-based methods. As they usually are the brightest regions of images, specular reflections are detected by thresholds while diffuse reflections are exclusive of VW data and more difficult to discriminate, being reported a method based in texture descriptors and machine learning techniques.

A global perspective of the proposed method is given in Fig. 4.18. The input is a VW iris sample and its corresponding segmentation mask, obtained according to the method of [65]. The center of the iris (x_c, y_c) is roughly found and used to parameterize the deemed biologic pupillary and limbic iris boundaries ($\{x_{pi}, y_{pi}\}$ and $\{x_{si}, y_{si}\}$). These boundaries permit to assess the quality of severely degraded samples that will be extremely difficult to handle by traditional approaches. For comprehensibility, the generated quality scores are denoted everywhere by " α " and a subscript: α_f for focus, α_m for motion blur, α_a for off-angle, α_o for occlusions, α_p for iris pigmentation, α_c for pixel count (area) and α_d for pupillary dilation. These quality factors are known to be strong correlated with recognition effectiveness and its assessment is made in a specific way for iris data. Other factors (such as Gaussian or salt and pepper noise) that are assessed in iris data as in any other type of images were not the scope of this work.

4.6.1 Estimation of the Iris Center

The convergence property of convex surfaces is the key insight this phase. The external boundary of the iris is predominantly convex, which was used to trace rays that pass through it and bend perpendicularly, converging into a *focal point* that gives an estimate of the iris center. As shown in figure 4.19, this gives a rough estimate of the true center, even in cases where the amount of unoccluded iris is very small. This approximation is acceptable because it is used as reference point by the shape descriptor that parameterizes the iris boundaries (normalized cumulative angular), which is invariant to the reference point as long as it is inside the desired contour.

Let I be an iris image (h rows \times w columns) and M the binary mask that gives the non-parameterized segmentation of its noise-free iris regions. Let $e = \{e_1, \dots, e_n\}$ be the set of edge pixels of M , each e_i with coordinates (e_x, e_y) . Let \vec{v}_i be a vector tangent to the iris boundary at e_i and $\perp \vec{v}_i$ its perpendicular. A line l_i through point e_i and perpendicular to \vec{v}_i is given by:

$$l_i = (b_x, b_y) + t \perp \vec{v}_i, \quad (4.35)$$

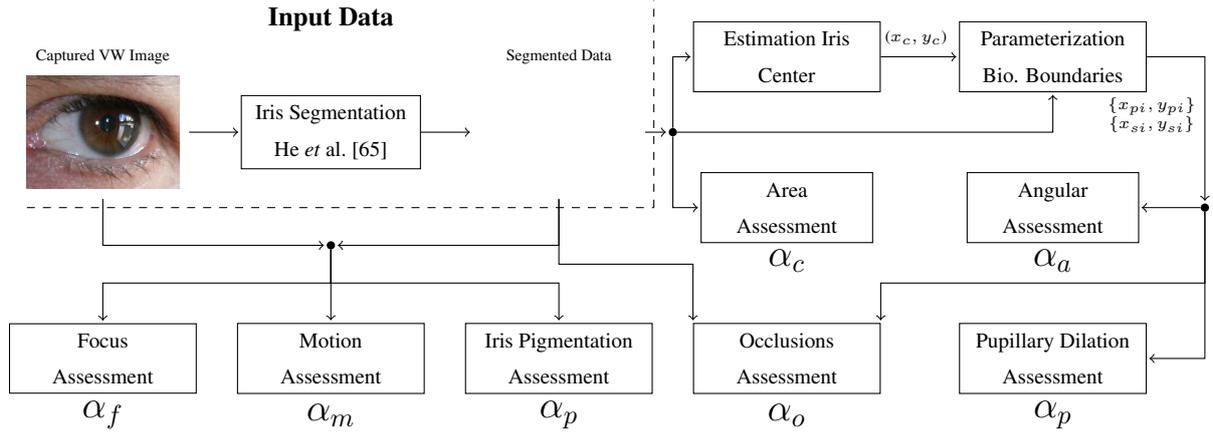


Figure 4.18: Cohesive perspective of the proposed quality assessment method. The iris segmentation mask is used to roughly estimate the center of the iris. This value is used as reference point to reconstruct the deemed biological iris boundary and to assess each of the seven quality measures.

with $t \in \mathbb{R}$. Let $l = \{l_1, \dots, l_n\}$ be the set of lines traced from e . We define an indicator function $\chi_1(x, y, l_i) \rightarrow \{0, 1\}$ that verifies whether the pixel (x, y) belongs to l_i :

$$\chi_1(x, y, l_i) = \begin{cases} 1, & \exists t \in]0, \sqrt{h^2 + w^2}] : (x, y) \in l_i \\ 0, & \text{otherwise.} \end{cases}$$

The center of the iris (x_c, y_c) is estimated as the point where the maximal number of l_i intersect, i.e., where the accumulated value of χ is maximal.

$$(x_c, y_c) = \arg \max_{x, y} \sum_x \sum_y \sum_{i=0}^n \chi_1(x, y, l_i), \quad (4.36)$$

where (x_c, y_c) give the coordinates of the iris center. Figure 4.19 gives two examples of this procedure: images in the left column have notoriously different proportion of noise-free iris and images in the right column give the accumulated $\chi_1(x, y, l_i)$ values. In both cases, the estimated center of the iris (denoted by the interception of the dashed lines) is acceptably close to the actual center.

4.6.2 Parameterization of the Biological Iris Boundaries

Efficient parameterization of the iris boundaries that are *behind* occlusions is a key issue regarding iris image quality assessment. Such boundaries not only permit to infer gaze, but also to estimate the proportion of occluded iris. This phase can be divided into two steps: (1) discriminate between the boundaries that correspond to iris biological borders and those that delimitate noisy regions; (2) reconstruct the biological iris boundary, according to the former boundary segments. The key insight is that biological boundaries can be faithfully described by periodic signals, which justifies the use of Fourier series for such purpose. Let (x_c, y_c) be the center of the iris and $f(x_c, y_c, \theta) = ((x_c - x_{\theta_j})^2 + (y_c - y_{\theta_j})^2)^{0.5} \cdot M(x_{\theta_j}, y_{\theta_j})$ the Euclidean distance between (x_c, y_c) and the noise-free iris pixels in a given direction θ_j , where $(x_{\theta_j}, y_{\theta_j})$ are given by:

$$(x_{\theta_j}, y_{\theta_j}) = (x_c, y_c) + t [0, 1] \begin{bmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{bmatrix}$$

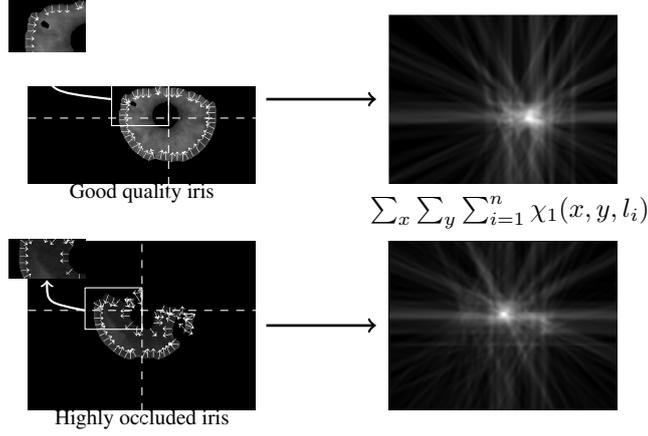


Figure 4.19: Estimated iris center (interception of the dashed lines) in a relatively unoccluded iris (upper row) and a heavily occluded one (bottom row). Images of the right column give the accumulated $\chi_1(x, y, l_i)$ values.

with $t \in]0, \sqrt{h^2 + w^2}]$. Let $b^s = \{b_1^s, \dots, b_n^s\}$ and $b^p = \{b_1^p, \dots, b_n^p\}$, b_i with coordinates (x_i, y_i) , be the noise-free iris pixels at respectively the farthest and closest distances from (x_c, y_c) in regularly spaced directions $\theta_i = \frac{2\pi i}{n}$, $i = 1, \dots, n$.

$$(x_i^s, y_i^s) = \arg \max_{x_{\theta_i}, y_{\theta_i}} f(x_c, y_c, \theta_i), \quad (x_i^p, y_i^p) = \arg \min_{x_{\theta_i}, y_{\theta_i}} f(x_c, y_c, \theta_i). \quad (4.37)$$

b^s and b^p give the outer and inner iris boundaries, and the former is illustrated in figure 4.20a. Special attention should be paid to extremely occluded irises, where some of the b_i might not exist and should be discarded of further processing by a simple semantic rule. Next, the cumulative angular function [173] is used as shape descriptor of each type of boundary, defined as the amount of angular change from an arbitrary starting point:

$$\gamma(t) = \int_0^{\frac{Lt}{2\pi}} k(r) dr - k(0) + t, \quad (4.38)$$

where $t \in [0, 2\pi]$ and $k(r)$ describes changes in direction at point t with respect to changes in arc length L . Here, if the boundary is a perfect circle, the corresponding angular description $\gamma(t)$ will be 0, $\forall t \in [0, 2\pi]$. An illustration of the $\gamma(t)$ values is given in figure 4.20b. Segments of the boundary that correspond to biological borders are discriminated by statistical estimation of the $\gamma(t)$, $\frac{\partial \gamma(t)}{\partial t}$ and $\frac{\partial^2 \gamma(t)}{\partial t^2}$ values. As illustrated in figures 4.20b-4.20d, biological boundaries have evident smoother values and lower energy, which gives raise to different values of the objective function:

$$O(t) = \beta_0 \gamma(t) + \sum_{i=1}^2 \beta_i \frac{\partial^i \gamma(t)}{\partial t^i}, \quad (4.39)$$

where β_i are regularization constants empirically obtained. Arguments of the first quartile of $O(t^*)$ — t^* regularly spaced in $[0, 2\pi]$ — are deemed to belong to the biological border and their coordinates (column and row) are illustrated by the dot and cross data points of figure 4.20e. The reconstruction of the complete biological border starts from these coordinates and is regarded as a nonlinear regression of a Fourier series of order r , given by:

$$c(x) = \frac{a_0}{2} + \sum_{k=1}^r (a_k \cos(x\omega k) + a_{r+k} \sin(x\omega k)). \quad (4.40)$$

Constraining the fundamental frequency to $\omega = 1$ assures closure and completeness of the obtained contour. For each boundary, both the column and row coordinates are fitted similarly. Given a set of $t^*/4$ data points x_i , the goal is to optimize the parameter vector $\vec{a} = (a_1, \dots, a_{2r+1})$ so as to minimize the sum of squares of the deviations:

$$S^2(\vec{a}) = \sum_{i=1}^{t^*/4} (y_i - c(x_i))^2, \quad (4.41)$$

where y_i is the desired value at x_i and $c(x_i)$ the actual response. Requiring the zero-gradient condition to hold at the minimum:

$$\nabla S(\vec{a}) = J^T (y_i - c(x_i)) = 0, \quad (4.42)$$

where J is the Jacobian. Let $b(\vec{a}) = -\nabla S(\vec{a}) = -J^T (y_i - c(x_i))$, the zero gradient is given by $b(\vec{a}) = 0$, which is solved iteratively using the Newton-Raphson iterations and starting from an initial guess of \vec{a} , corrected to $\vec{a} + \Delta \vec{a}$. The increment $\Delta \vec{a}$ satisfies the linear equation $J^*(\vec{a}) \Delta \vec{a} = -b(\vec{a})$, where J^* is the Jacobian of b :

$$J_{ij}^* = \frac{\partial b_i}{\partial a_j} = \frac{-\partial^2 S}{\partial a_i \partial a_j} = -H_{ij}, \quad (4.43)$$

As $J^* = -H$, $\Delta \vec{a}$ satisfies the system of linear equations:

$$H \Delta \vec{a} = b \quad (4.44)$$

$$\Delta \vec{a} = b H^{-1} \quad (4.45)$$

$$\Delta \vec{a} = H^{-1} J^T (y_i - c(x_i)). \quad (4.46)$$

Assuming moderate non-linearity, a reasonable approximation of H is given by $J^T J$, obtaining the next generation of the parameter vector \vec{a} . This is repeated until convergence is reached, i.e., $\|\Delta \vec{a}\| < \alpha$, being α a very small positive value. An illustration of the reconstructed biological iris boundary is given in figure 4.20f.

4.6.3 Focus Assessment

The Fourier domain is the most frequently used in focus assessment. The rationale is that focused data contains more high frequency components than blurred one. Daugman pointed out that defocus is equivalent to multiplying the 2D Fourier transform of a perfectly focused image by the 2D Fourier transform of a Gaussian kernel [36]. Thus, focus of an image I can be assessed by measuring its amount of high frequencies, i.e., the accumulated power of the convolution between I and a high-pass kernel H :

$$\alpha_f = \int_x \int_y |I * H|^2 dx dy. \quad (4.47)$$

This measure performs well if images are dominated by the iris. Otherwise, if the iris is just a small part of the data — dominated by eyelashes or hair that are full of minutia — it tends to fail, which is particularly probable in uncontrolled image acquisition setups. Hence, focus is exclusively assessed in the region segmented as noise-free iris (using the segmentation mask M), avoiding the described

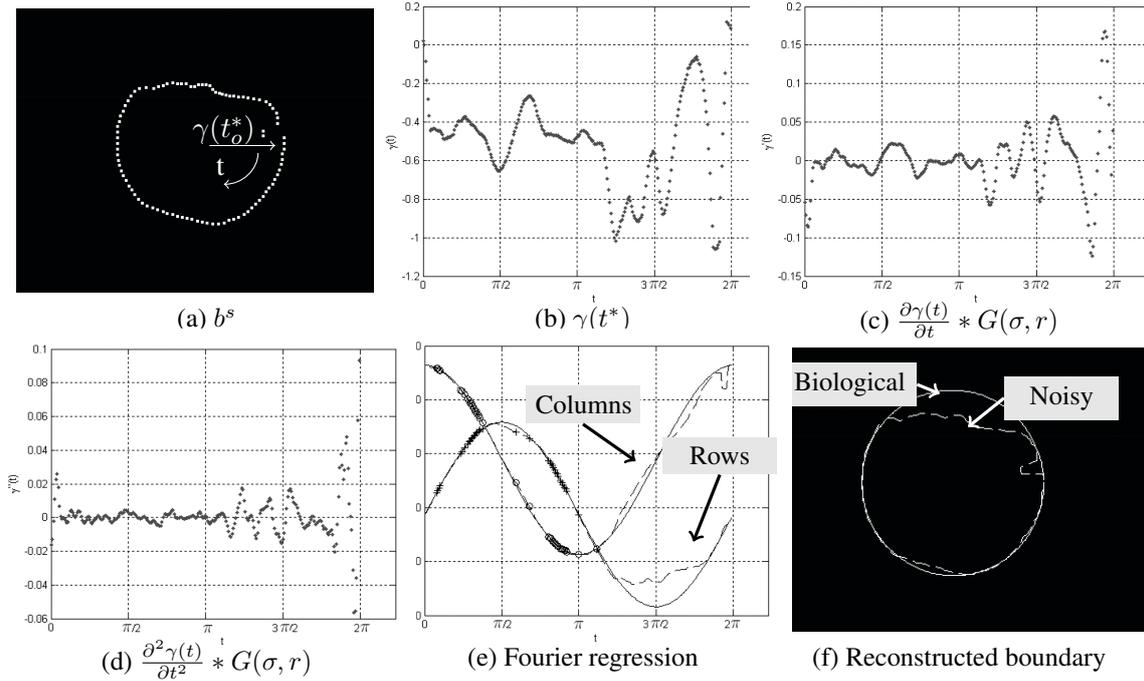


Figure 4.20: Estimation of the biological scleric border. The boundary points that are farthest from the iris center at each direction (figure a) are used to obtain the cumulative angular description $\gamma(t^*)$ of the contour (figure b) and its first and second order derivatives (figures c and d). The selection of the regions deemed to belong to the iris biological boundary (cross and circular data points of figure e) enables the reconstruction of the deemed biological border through a regression of a Fourier series (figure f).

problem. Also, assuming that segmentation inaccuracies predominate near the boundaries, we compared the results obtained when using exclusively the most interior regions of the noise-free irises (obtained by morphologic erosion of M and illustrated in the top images of figure 4.21), that have higher probability of actually being noise-free. Let M_p be an iris segmentation mask eroded by a circular structuring element of radius p . In order to keep the proportion of iris removed by erosion independent of the iris size, the radius of the structuring element is given by:

$$p = \arg \min_{p'} \sum_x \sum_y \frac{M_{p'}(x, y)}{M(x, y)} < T, \quad p' = 1, \dots, n, \quad (4.48)$$

where T is an empirically adjusted threshold. In the experiments, we used a data set of 1 000 focused images of the UBIRIS.v2 database D_1 and convolved them with Gaussian kernels $G(s, \sigma)$, $(s, \sigma) = (5i, 2i)$, $i = 2, \dots, 4$, obtaining three increasingly defocused versions of the data set: $\{D_2, D_3, D_4\}$. As it is given in the bottom row of figure 4.21, we compared the α_f values obtained by the Daugman [36] and Kang and Park [84] kernels when using: (1) the whole image; (2) the whole iris and (3) the most interior iris regions as input data. The plot at the bottom left corner gives the proportion between the average α_f values of consecutive data sets, i.e., $\sum_2^4 \bar{d}_{i-1} / \bar{d}_i$, where $\bar{d}_i = \frac{1}{n} \sum \alpha_f(i)$ is the average focus value in the D_i data set. Maximal separation was obtained around $T = 0.4$ (Daugman kernel) and $T = 0.3$ (Kang and Park kernel), which is justified by the smaller size of the latter kernel. Plots at the center and right contextualize the results, comparing the \bar{d}_i values obtained when using the whole image (dashed lines), the whole iris (continuous lines) and the interior iris regions (with $T = 0.35$, dotted lines) as source data. It can be confirmed that both spectral measures perform acceptably in VW data

and that results were consistently improved when exclusively the most interior iris regions were used as input (defocused versions of D_i obtained the lowest focus scores).

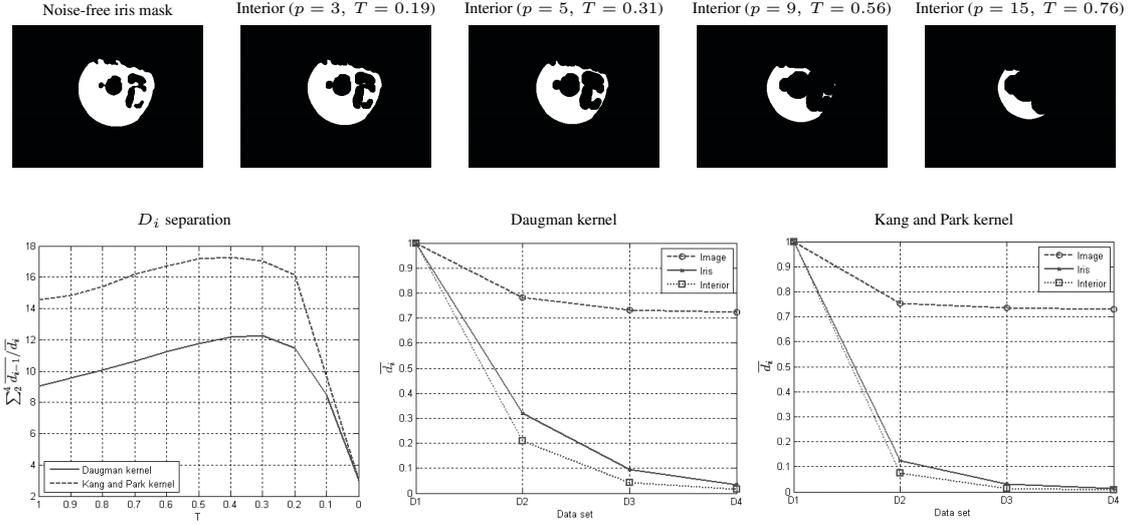


Figure 4.21: The upper row shows an iris mask and its increasingly most interior iris regions. The bottom left figure shows the separation between the focus scores obtained in defocused versions of the training data set, in respect to different values of T . The bottom center and right plots illustrate the improvement observed when just the most interior iris regions are used as source data (dotted lines), instead of the whole image (dashed lines) or whole iris (continuous lines).

4.6.4 Motion Assessment

There are various causes for motion blurred iris images, corresponding to the different types of movements in the scene: subjects, heads, eyes and eyelids, causing motion to be non-linear and particularly hard to determine and compensate for. In previous works, an oversimplification was made and assumed that motion is linear across the image, which cannot be guaranteed in unconstrained acquisition setups. Instead, we exclusively assumed that motion is linear in the iris ring and concerned about the detection of linear motion blur inside an iris bounding rectangle $B(x, y)$. The rationale is that linearly motion-blurred data has frequency responses with visible parallel strips of direction θ that correspond to the motion direction. As suggested by Kalka *et al.* [81], the width and power of these strips provide an estimate of the amount of motion. We used the concept of power cepstrum $C(B) = |F(\log(|F(B(x, y))|))|^2$, illustrated in three images of figure 4.22: a sharp and two motion-blurred. The primary direction of motion is deemed to be the one that minimizes the power of the derivatives of $C(B)$ along a direction $\theta_i \in [0, 2\pi]$.

$$\theta_m = \arg \min_{\theta_i} G(\sigma, r) * \sum_x \sum_y \left(\frac{\partial}{\partial \theta_i} C(x, y) \right)^2. \quad (4.49)$$

$\frac{\partial}{\partial \theta_i} C(x, y)$ gives the partial directional derivative of C along the θ_i direction ($\theta \in [0, \pi]$), obtained as described in [154]. The amplitude α_m of the partial directional derivatives in $[0, 2\pi]$ gives the motion quality score, as illustrated in figure 4.22d:

$$\alpha_m = \max_{\theta_i} \sum_x \sum_y \left(\frac{\partial}{\partial \theta_i} C(x, y) \right)^2 - \min_{\theta_i} \sum_x \sum_y \left(\frac{\partial}{\partial \theta_i} C(x, y) \right)^2. \quad (4.50)$$

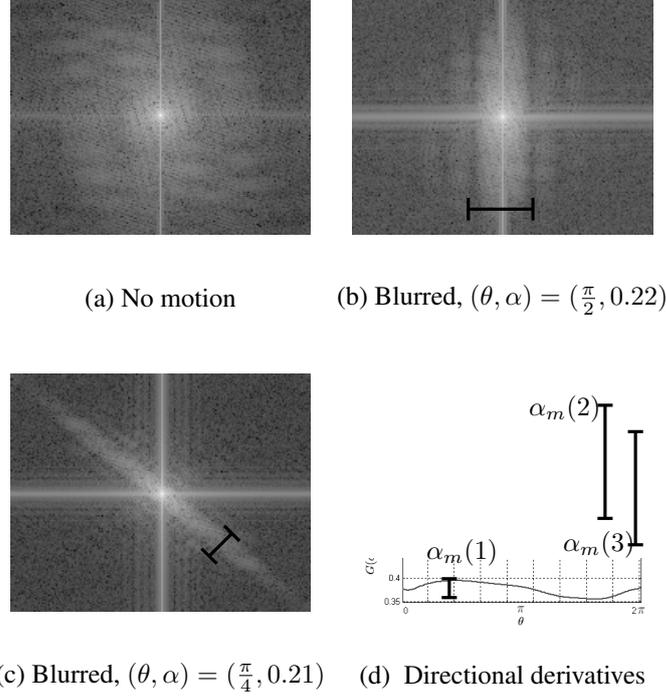


Figure 4.22: Power cepstrum of a sharp iris (figure a) and of two motion blurred irises (figures b and c). Fig. d gives the accumulated image derivatives along θ directions, where the minimum values indicate the primary blur direction. The amplitude of this signal is also used to discriminate between motion blurred and sharp irises, as $\alpha_m(2), \alpha_m(3) \gg \alpha_m(1)$.

4.6.5 Off-Angle Assessment

Assuming the circularity of the pupil as a measure of the off-angle, Kalka *et al.* [81] and Dorairaj *et al.* [42] projected an integro-differential operator at multiple yaw and pitch angles, yielding an optimization process in a 5D hyperspace with significant computational efforts. Daugman [38] used Fourier series expansions of the 2D coordinates of the pupillary boundary, which contain shape distortion information that is related to deviated gaze. Expressing the pupillary boundary as a set of coordinate vectors $X(t)$ and $Y(t)$, the direction and magnitude of the gaze deviation are contained in the form of Fourier coefficients on the harmonic functions $\cos(t)$ and $\sin(t)$ that — linearly combined — represent $X(t)$ and $Y(t)$. Algebraic manipulation estimates from four Fourier coefficients two gaze deviation parameters θ and γ . Our aim is to obtain a unique value that gives the magnitude of the deviation (used as quality score) and other one that corresponds to its major direction. The rationale is that the degree of circularity of the scleric boundary should be inversely correlated to off-angle acquisition. For such, we used the concept of minimal bounding rectangle and compared the length of its sides. Let $b^* = \{b_1^*, \dots, b_n^*\}$ be the coordinates of the biological scleric boundary, that delimitates a convex polygon P . As proposed in [156], (x_m, x_M, y_m, y_M) are the coordinates of the four extreme points of P , where m and M stand for the minimum and maximum values. Let $l_j : 1 \leq j \leq 4$, be four line segments through the extreme points and θ the minimal rotation such that one l_j coincides with a line segment defined by (b_i^*, b_{i+1}^*) . Let $l_j^\theta : 1 \leq j \leq 4$ be the versions of l_j rotated by θ . Minimizing the area of the enclosing rectangle is equivalent to keep the minimum of successive rotations, until the lines have been rotated by an angle greater than 90 degrees. Let $l_j^* : 1 \leq j \leq 4$ be the line segments that delimitate the minimum enclosing rectangle of extreme points $(x_j, y_j) : 1 \leq j \leq 4$. Let d_M and d_m be the lengths of these line segments, such that $d_m \leq d_M$. The magnitude of the deviation $\alpha_a : \mathbb{R}^{n \times 2} \rightarrow]0, 1]$ and its major direction

$\theta_a : \mathbb{R}^{n \times 2} \rightarrow [-\pi, \pi]$ are obtained as follows:

$$(\alpha_a, \theta_a) = \left(1 - \frac{d_m}{d_M}, \arctan(\vec{v})\right), \quad (4.51)$$

where $\vec{v} = \frac{(x_i, y_i) - (x_j, y_j)}{d_M}$, such that $y(j) \leq y(i)$ and $((x_i - x_j)^2 + (y_i - y_j)^2)^{0.5} = d_M$. Fig. 4.23 illustrates three iris images and the corresponding off-angle scores: image in the left column was frontally captured and the others resulted of moderated (central column) and severe deviations (right column) in image acquisition.

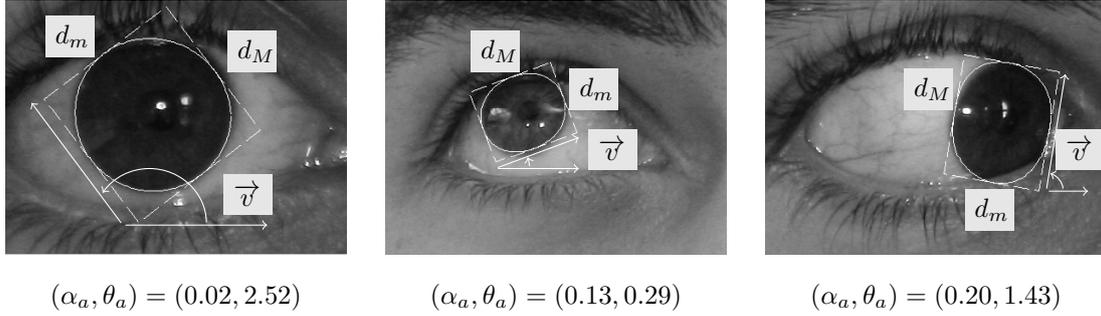


Figure 4.23: Examples of the proposed off-angle assessment, for aligned, moderately and severely deviated iris images.

4.6.6 Oclusions Assessment

An estimate of the amount of iris that is occluded by other types of data is obtained by comparing the area inside the deemed biologic iris ring (inside the scleric $\{s_1^*, \dots, s_n^*\}$ and outside the pupillary contour $\{p_1^*, \dots, p_m^*\}$), and the area of the noise-free iris segmentation mask. Assuming the convexity of the polygon defined by the set of vertices $\{b_1^*, \dots, b_n^*\}$, a pixel is on the interior of this shape if it is always on the same side of all line segments defined by (b_i^*, b_{i+1}^*) , $b_i^* = (x_i, y_i)$. Let $z_i = f(x, y, b_i^*, b_{i+1}^*) : \mathbb{N}^6 \rightarrow \mathbb{R}$, be a function that relates the position of (x, y) to the straight line defined by $z_i = (y - y_i)(x_{i+1} - x_i) - (x - x_i)(y_{i+1} - y_i)$.

An indicator function that discriminates between pixels inside and outside the iris ring is given by:

$$\chi_2(x, y) = \begin{cases} 1 & , \text{if } \forall j, j' : z_j^s, z_{j'}^s > 0 \wedge \exists k, k' : z_k^p, z_{k'}^p < 0 \\ 0 & , \text{otherwise,} \end{cases} \quad (4.52)$$

where $j, j' \in \{1, \dots, n\}$ and $k, k' \in \{1, \dots, m\}$. Function $o : \mathbb{N}^{w \times h} \times \mathbb{N}^{w \times h} \rightarrow [0, 1]$ has direct correspondence to the proportion of iris occluded by noise:

$$\alpha_o(I, M) = \frac{1}{n} \sum_x \sum_y \left(\chi_2(x, y) - (\chi_2(x, y) M(x, y)) \right), \quad (4.53)$$

where n is the area of the iris ring. Fig. 4.24 gives some examples of this procedure: white pixels denote regions segmented as noise-free that are inside the deemed biological boundaries; dark and light gray regions respectively denote occluded iris pixels and regions outside the deemed biological boundaries. It can be confirmed that assessments match the intuitive human perception, either for almost unoccluded, moderately or severely occluded iris images.

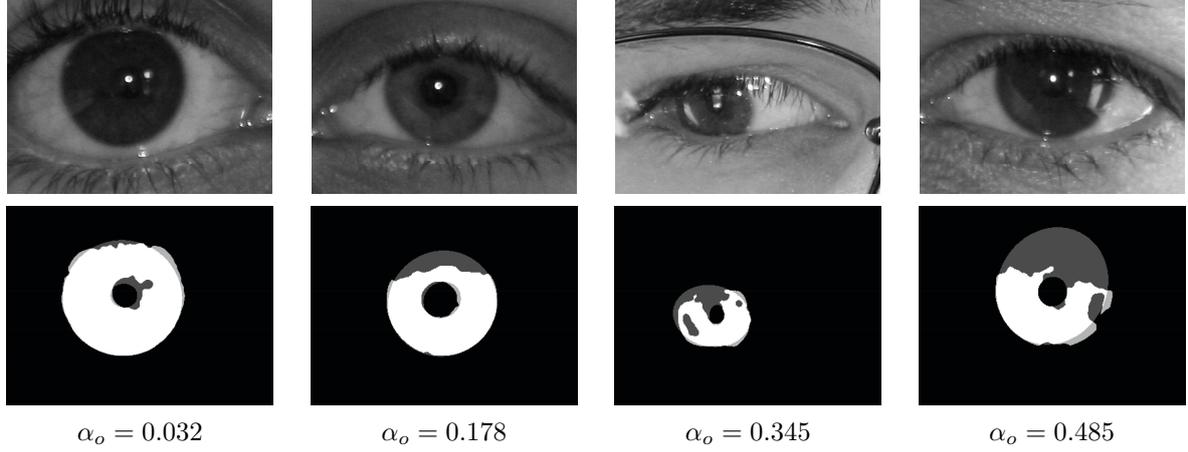


Figure 4.24: Examples of the proposed iris occlusions assessment, for practically unoccluded, moderately and severely occluded iris images. White regions were segmented as noise-free and are inside the deemed biological boundaries. Dark and light gray regions respectively denote noisy iris pixels and regions outside the biological boundaries.

4.6.7 Levels of Iris Pigmentation

The spectral radiance of the iris in respect of its level of pigmentation has greater variance in the VW than in the NIR, which increases the heterogeneity of the captured data: light pigmented irises provide more detail than heavy pigmented, although the number of visible crypts, arching ligaments, freckles and contraction furrows significantly varies between subjects. Also, the appearance of the striated anterior layer that covers the trabecular meshwork depends of the amount of light used in the acquisition process. Here, we propose a strategy to assess the levels of iris pigmentation based in the observation that local *hue* and *value* across the iris vary accordingly to its levels of pigmentation. Also, we consider the average brightness of a region deemed to contain the sclera, adjusting for different lighting conditions. Let I_{hsv} be the iris image represented in the *HSV* color space. Let $\{e_1, \dots, e_n\}$ be the set of edge pixels of M that are farthest from (x_c, y_c) in n regularly spaced directions ($n \in [0, 2\pi]$). Let $d = \{d_1, \dots, d_w\} : w = \sum_{i=1}^{n-1} i$, be the Euclidean distances between elements of e . The third quartile of d corresponds to the farthest distances between edges, which associated vectors tend to point to the direction of the sclera (avoiding eyelids and eyelashes occlusions). The direction of the sclera is given by the mean direction of the vectors drawn from these farthest points, i.e., $\vec{v} = \frac{4}{w} \sum_{i=1}^{w/4} (x_i - x_j, y_i - y_j)$. Thus, $(x_c, y_c) \pm k \vec{v}$ delimitate an image strip that spreads in direction \vec{v} and is illustrated in the left column of figure 4.25. Let B be a binary mask, such that $B(x, y) = 1$ inside the band and 0 otherwise. The levels of iris pigmentation are obtained by relating the hue and value of the iris and of the sclera pixels inside this band:

$$\alpha_p(I) = \frac{m_1 \sum_x \sum_y (I_{hv}(x, y) M(x, y) B(x, y))}{\beta m_2 \sum_x \sum_y (I_{hv}(x, y) B(x, y) (1 - M(x, y)))}, \quad (4.54)$$

where $m_1 = \sum \sum B(x, y)(1 - M(x, y))$ and $m_2 = \sum \sum B(x, y)(M(x, y))$. $I_{hv}()$ gives the pixel *hue* and *value* and $\beta \in [0, 1]$ is a regularization term. According to a set of manually classified 300 iris images of *light* and *heavy* pigmented irises, a k nearest neighbors strategy classifies the sample into one of k classes w_i : $w_i = k_l / (k_l + k_w)$, where k_l and k_h represent the number of light and heavy pigmented irises in the nearest neighbors. The plot at the right of figure 4.25 gives the α_p values of the

training set, where circular and cross data points represent light and heavily pigmented irises.

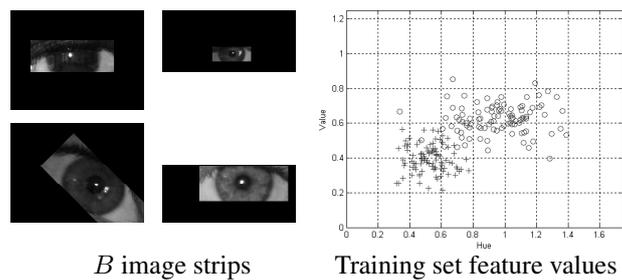


Figure 4.25: Images at the right column illustrate the image band deemed to contain the sclera. Plot at the center gives the α_p values of a training set of 300 images, where circular and cross points represent light and heavy pigmented irises.

	<p>Iris Image Quality Assessment Main Challenges:</p> <ul style="list-style-type: none"> Real-time quality assessment Quantification of the high-dimensional correlation between quality factors and recognition performance
---	---

4.7 Iris Recognition

	<p>Hugo Proença, Gil Santos; Fusing Color and Shape Descriptors in the Recognition of Degraded Iris Images Acquired at Visible Wavelength, Elsevier Computer Vision and Image Understanding, volume 116, pag. 167-178, ISSN 1077-3142, Digital Object Identifier 10.1016/j.cviu.2011.10.008</p>
---	--

We proposed a recognition strategy that uses techniques that are substantially different from those traditionally used in iris recognition, making use of *color* and *shape*. Color is a major visual feature in image and video analysis because color features are considered robust to viewing angle, translation, rotation and scale. Furthermore, for many situations, the shape of image objects often provides important clues for recognition, although shape is sensitive to geometric distortions. The proposed method begins by partitioning the iris into coherent regions in terms of space and color, using data self-organization techniques that tend to compensate for global changes in data. Our method then makes use of a set of well-known color and shape MPEG.7 descriptors to extract both global and local information from the iris data. According to the experiments performed, two types of conclusions are substantiated: 1) the proposed approach achieves performance close to the state-of-the-art methods, and 2) because the data encoding and matching techniques are radically different from the state-of-the-art approaches, the proposed method exhibits low levels of linear correlation with the outputs, which allows it to obtain significant improvements in performance when performing evidence fusion.

A cohesive perspective of the proposed recognition strategy is given in figure 4.26. A color constancy technique is used for regularization purposes, and data are normalized into a Polar coordinate system of constant dimensions, from which global MPEG.7 color descriptors are extracted. Next, a self-organizing data technique divides the noise-free iris data into spatially and color coherent regions that feed the local color and shape MPEG.7 descriptors. Finally, fusion is performed by means of a weighted sum rule.

4.7.1 Retinex

The original Retinex model was proposed by Land [92]. Its key insight is that color is determined by three independent retinal-cortical systems that use intensity information from different spectral regions of the input data. Each system determines a lightness quantity that is superimposed, yielding the output

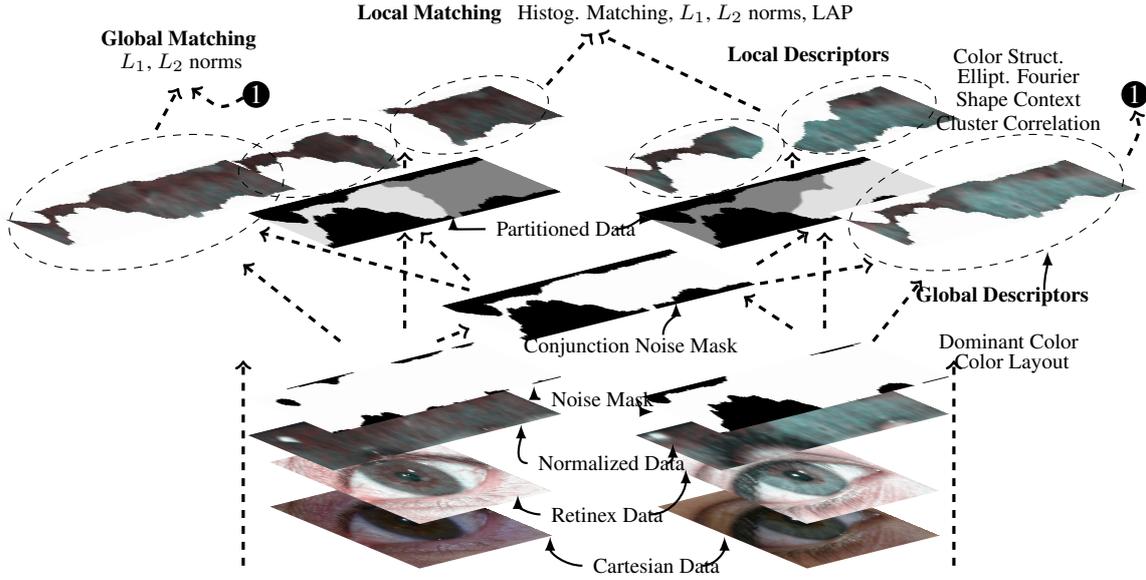


Figure 4.26: Cohesive perspective of the proposed method to recognise visible-wavelength iris images.

color for each point. As detailed by Provenzi *et al.* [133], given an image I , $|\gamma_{ki}|$ ordered chains of pixels can be obtained, starting at k and ending at i , where $|\cdot|$ denotes cardinality. Let x_t and x_{t+1} be subsequent pixels of a chain. Let R_t be the ratio between the intensity of consecutive pixels in the chain, $R_{t_k} = \frac{I(x_t)}{I(x_{t+1})}$, with respect to each image channel. Lightness at position i is given by

$$L(i) = \frac{1}{N} \sum_{k=1}^N \prod_{t=1}^{|\gamma_{ki}|-1} \delta_k(R_{t_k}), \quad (4.55)$$

being δ_k given by:

$$\delta_k(R_{t_k}) = \begin{cases} R_{t_k} & , \text{if } 0 < R_{t_k} \leq 1 - \epsilon \\ 1 & , \text{if } 1 - \epsilon < R_{t_k} < 1 + \epsilon \\ R_{t_k} & , \text{if } 1 + \epsilon \leq R_{t_k} \leq \frac{1+\epsilon}{\alpha} \\ \frac{1}{\alpha} & , \text{if } R_{t_k} > \frac{1+\epsilon}{\alpha}, \end{cases}$$

where $\epsilon > 0$ and $\alpha = \frac{1+\epsilon}{\prod_{m_k=0}^{t_k-1} \delta_k(R_{m_k})}$.

4.7.2 Noise-Free Iris Segmentation

The segmentation of the noise-free iris data acquired in uncontrolled setups has motivated significant research efforts. He *et al.* [151] used a clustering-based scheme to roughly perform iris localization followed by an integro-differential constellation method for fine detection of each boundary, which not only accelerates the traditional integro-differential operator but also enhances its global convergence. Finally, parametric models were trained to deal with eyelids and eyelashes. Du *et al.* [45] used a high-pass filter to detect specular reflections inside the pupil and performed a coarse-to-fine segmentation scheme using a least-squares ellipse fitting strategy. A gradient-based technique detected noisy regions that corresponded to diffuse reflections inside the iris. Li *et al.* [95] used the Viola and Jones method to roughly detect eyes and normalized their region of interest by a K-means-based technique. These

data fed the subsequent processing combining traditional iris segmentation methods with RANSAC-like techniques. Concerned about the computational requirements of previously published iris segmentation methods, Proença [129] considered the sclera the most easily distinguishable part of the eye in degraded VW images and fed a neural network with a feature set based in the local proportion of sclera in different directions, resulting in a process that runs in deterministically linear time with respect to the size of the image. Regarding all of the experiments carried out in the scope of this work, it was observed that, although with noticeably higher computational requirements, the segmentation method of He *et al.* [65] outperforms the other strategies. Because we aim to obtain performance indicators that are as unbiased as possible, we chose to use this method as the basis for our recognition experiments. Fig. 4.27 gives examples of eye images and the corresponding noise-free iris segmentation masks, obtained by He *et al.*'s [65] method.



Figure 4.27: Examples of degraded VW iris images and the corresponding noise-free segmentation masks obtained according to the method of He *et al.* [65]. The binary masks discriminate between the non-occluded pixels of the iris (white regions) and all of the remaining types of data (black regions).

4.7.3 Parameterization of Iris Boundaries

Subsequent to segmentation, efficient parameterization of the iris boundaries that are *behind* occlusions was a key issue, especially regarding the normalization of the iris data into a pseudo-polar coordinate system of constant dimensions. As detailed in [130], this phase was divided into two steps: (1) discriminating between the boundary segments that correspond to biological iris borders and the boundary segments that delimit noisy regions and (2) reconstructing the full biological iris boundaries according to the former segments. The key insight in this step is that biological boundaries can be faithfully described by periodic signals, which justifies the use of Fourier series for such purposes. The cumulative angular function was used as a shape descriptor, defined as the amount of angular change from an arbitrary starting point:

$$\gamma(t) = \int_0^{\frac{Lt}{2\pi}} k(r) dr - k(0) + t, \quad (4.56)$$

where $t \in [0, 2\pi]$ and $k(r)$ describe changes in direction at point t with respect to changes in arc length L . As illustrated in figures 4.28a and b, biological boundaries have smoother angular descriptor values with lower energy, which leads to the following objective function:

$$O(t) = \beta_0 \gamma(t) + \sum_{i=1}^2 \beta_i \frac{\partial^i \gamma(t)}{\partial t^i}, \quad (4.57)$$

where β_i were empirically obtained regularization constants. Arguments of the first quartile of $O(t^*)$ — t^* regularly spaced in $[0, 2\pi]$ — were deemed to belong to the biological border and their coordinates (column and row), illustrated by the dot and cross data points of figure 4.28b. Finally, the reconstruction of the biological border used the selected coordinates and was regarded as a nonlinear regression of a Fourier series of order r , with a fundamental frequency constrained to $\omega = 1$, which assures closure and completeness of the contour:

$$c(x) = \frac{a_0}{2} + \sum_{k=1}^r (a_k \cos(x\omega k) + a_{r+k} \sin(x\omega k)). \quad (4.58)$$

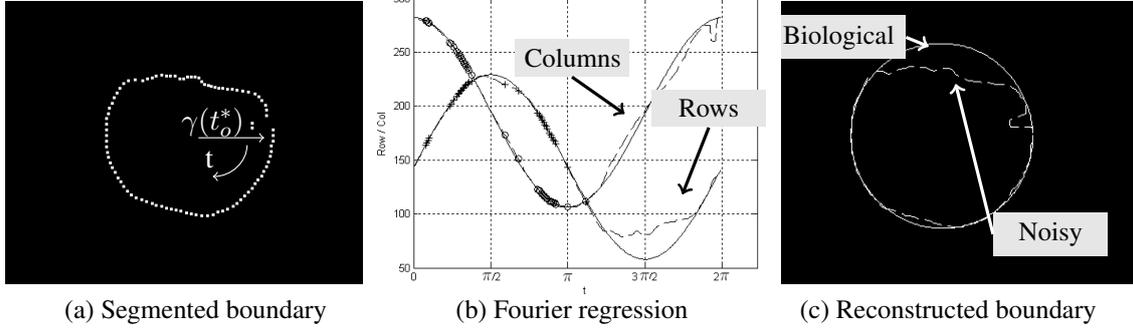


Figure 4.28: Parameterization of the biological iris boundaries. According to the values of (4.57), smoother regions with low energy of the cumulative angular descriptor (Fig. a) were deemed to belong to the biological boundaries (cross and circular data points of Fig. b) and used to reconstruct the deemed biological border through a regression of a Fourier series (Fig. c).

Using the deemed biological iris boundaries, in the next step we convert data into a pseudo-polar coordinate system of fixed dimensions, using the well known *Daugman rubber sheet* model [38].

4.7.4 Partitioning the Iris into Regions

Partitioning the iris into regions is one of the roots of the proposed recognition method and aims to divide the noise-free pixels of the iris into k disjoint sets $C = \{C_1, \dots, C_k\}$ such that elements within each C_i are as *homogenous* as possible in terms of both their position and their color. Considering the demands of a perceptually uniform color space and aiming to preserve the connectivity between pixels of each cluster, each element was represented by the feature set $f = \{r, |W/2 - c|, L^*, a^*b^*\}$, with r and c the row and column coordinates with respect to the normalized iris image. W is the width of the normalized image, and L^*, a^*b^* are the color coordinates in the CIELAB color space, using a reference white provided by illuminant $D65$, 2^{nd} observer, as described in ¹. In this 5D space, *distance* corresponds to the metric:

$$d(f_1, f_2) = \sqrt{\sum_i \omega_i (f_1^i - f_2^i)^2}, \quad (4.59)$$

where f^i denotes the i^{th} feature of f . Using a partition-based clustering scheme (fuzzy c-means [15]), partitions were found by maximizing an objective function that considered both the within and between cluster variation:

$$J(C) = \sum_{i=1}^k \left(\sum_{j=1}^k d(C_i, C_j) - \sum_{j \in C_i} d(C_i, f_j) \right), \quad (4.60)$$

being $d(C_i, C_j)$ the sum of L_2 distances (4.59) between every combination of elements of C_i and C_j and $d(C_i, f)$ the sum of L_2 distances (4.59) between every element of C and the feature point f .

¹<http://www.csse.uwa.edu.au/du/Software/graphics/xyz2lab.m>

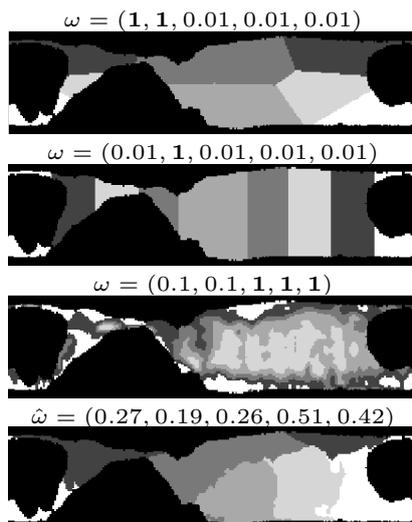


Figure 4.29: Comparison between the regions resulting from different clustering processes with respect to weights given to each feature. The image at the top resulted from large weights (denoted by bold font) for spatial features, whereas in the case of the second image at the top, a large weight was given exclusively for one of those spatial features (column). The second image at the bottom resulted from low weight values for spatial features, and clusters were formed, accounting for the color values. The weights used in the case of the image at the bottom were obtained by (4.61); this type of cluster is used in all subsequent processing phases.

Fig. 4.29 illustrates how the typical appearance of the generated clusters would vary with respect to different ω_i values. Black pixels denote regions that were classified as *noisy* by the segmentation method and, as such, were not considered in the clustering process. The remaining intensities represent the clusters assigned to each pixel of the normalized iris data when privileging the (ω_1, ω_2) weights (associated with spatial features) and the $(\omega_3, \omega_4, \omega_5)$ (associated with color features) (bottom left image). The image at the bottom illustrates clusters generated for the *optimal* weight values $\hat{\omega}_i$, which constitutes a trade-off between space and color, as follows:

$$\hat{\omega}_i = \arg \min_{\omega_i} \sum_k \sum_j \alpha \phi_k(\omega_i, \{C_{k,j}\}) + (L_{k,j}^2 - 4\pi A_{k,j}), \quad (4.61)$$

where $L_{k,j}$ and $A_{k,j}$ represent the perimeter and the area of the region delimited by the j^{th} cluster of the k^{th} image ($C_{k,j}$), ϕ_k corresponds to the total of connected components in that cluster [39] and α is a regularization term that was empirically found according to the training data set of k images ($\alpha = 10^{1.5}$, $k = 100$ in our experiments). To account for the dynamic conditions that propitiate occlusions in different regions of the irises, the clustering process was performed using as a noise-mask the conjunction of the noise-masks of images to be matched.

4.7.5 Color Descriptors

Most of the MPEG.7 descriptors have compression/reconstruction purposes and — consequently — tend to focus in the lowest frequency components of signals. When compared to other biometric traits, one of the most interesting features of the iris is that most of its discriminating information lies in the lowest and middle-low frequency components. For such, these descriptors would intuitively be useful for iris recognition purposes, which constituted the main key insight for their utilization in this work.

Dominant Color Descriptor

The dominant color descriptor summarizes the image content by extracting the most *important* colors in an image or region, naturally perceived as the most frequent. Let $I = \{\vec{x}\}$, $\vec{x} = (x_1^j, x_2^j, x_3^j)$ be a $r \times c$ image represented in the CIELAB color space, known to more closely fit the Euclidean difference between colors and the visual perception of color difference. Let k be the number of colors to extract from the image. Aiming to obtain deterministic results, the k geometric centroids (s_i) were used as initial values of the centers of clusters:

$$s_i = \frac{i \max\{\vec{x}\} + (k + 1 - i) \min\{\vec{x}\}}{k + 1}, \quad i = \{1, \dots, k\}. \quad (4.62)$$

The coordinates of these centroids were updated according to the generalized Lloyd algorithm, minimizing the objective function:

$$\arg_s \min \sum_{s=1}^k \sum_{\vec{x} \in s_i} \|\vec{x} - s_i\|^2. \quad (4.63)$$

The iterative procedure continues until the values of s_i at successive steps do not differ more than a positive value that acts as stopping criterium ($0 < \epsilon \ll 1$), i.e., $\|s_i^t - s_i^{t+1}\| < \epsilon$.

Color Layout Descriptor

This descriptor extracts the spatial distribution of the most representative colors of visual signals according to their position on a grid superimposed on data. Let I be an image of size $r \times c$ expressed in the YCbCr color space [27]. First, the most representative colors in each $n \times n$ region are obtained, as described in section 4.7.5, which yields a $r/n \times c/n$ array that was transformed using the 1D Discrete Cosine Transform type-II, performed first along the rows and then along the image columns, yielding a set of 2D components:

$$X_{i,j} = \sum_{n=0}^{c-1} \sum_{m=0}^{r-1} I(n, m) \cos\left[\frac{\pi}{r}(m + 0.5) j\right] \cos\left[\frac{\pi}{c}(n + 0.5) i\right].$$

Finally, the $X_{\{i,j\}}$ coefficients were zigzag ordered [118], resulting in a vector \vec{v} of $(r \times c)/n^2$ real components. Matching between vectors \vec{v}_1 and \vec{v}_2 was performed according to the L_2 norm of the vector $\vec{v}_1 - \vec{v}_2 = \sqrt{\sum_{i=1}^{(r \times c)/n^2} (v_1(i) - v_2(i))^2}$.

Color Structure Descriptor

The color structure descriptor [27] generalizes a simple color histogram and uses a structuring element that moves across data, defining a neighborhood where the dominant color values are analyzed and counted for each bin. In our experiments, we used a rectangular structuring element with $\frac{1}{12}$ of the image width and height. Because this descriptor is very similar to an image histogram, the same L_1 based matching functions were used in matching.

Scalable Color Descriptor

Scalable color descriptors [27] are global descriptors mainly used for image-to-image matching. The process starts by extracting a color histogram of k bins in the HSV color space, where the hue component is usually quantized to a larger number of bins compared to saturation and value layers. Such a histogram feeds a dyadic decomposition process based in a pair of Haar wavelet transforms. Let I be an image

represented in the HSV color space and let h be the corresponding normalized histogram with k bins. The convolution between h and the low-pass kernel of the Haar transform is equivalent to summing pairs of adjacent bins, whereas the high frequency components are obtained by the difference in adjacent bins. Such decomposition is repeated n times, using at each iteration the lower frequency components previously obtained. The default matching function is based on the L_1 metric (i.e., the sum of the absolute differences between corresponding elements): $S = \sum_{i=1}^k |h_A[i] - h_B[i]|$, where k denotes the number of extracted coefficients.

4.7.6 Order Statistics of Dominant Colors

Let \vec{c} be the dominant color of the noise-free iris portion of I . Let \vec{c}_i be the dominant color of each cluster C_i and $d : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$ the L_2 norm of the vector $\vec{c}_i - \vec{c}$. For any pair (\vec{c}_i, \vec{c}_j) , we define a pseudometric $d^*(\vec{c}_i, \vec{c}_j)$ given by

$$d^*(\vec{c}_i, \vec{c}_j) = |d(\vec{c}_i, \vec{c}) - d(\vec{c}_j, \vec{c})|.$$

Similarly, a binary relation \succ on $\mathbb{R}^3 \times \mathbb{R}^3$ is defined by

$$\vec{c}_i \succ \vec{c}_j \Leftrightarrow d(\vec{c}_i, \vec{c}) > d(\vec{c}_j, \vec{c}). \quad (4.64)$$

The rationale behind \succ is to consider as *greater* colors those that are more distant from the dominant color of the iris. From this definition, it is straightforward to infer that \succ is irreflexive, asymmetric and transitive, which is particularly useful for our purposes. Let $X = \{\vec{x}_1, \dots, \vec{x}_n\}$ be a random variable that represents the distance between the dominant colors inside each cluster and the whole iris. According to the elementary theory of rank tests and using (4.64), the k^{th} order statistic $x_{(k)}^{\vec{}}$ of a statistical sample $\{\vec{x}_1, \dots, \vec{x}_n\}$ is equal to its k^{th} smallest value. Let $x_{(1)}^{\vec{}}, \dots, x_{(n)}^{\vec{}}$ be the order statistics of a set of independent observations, that is, $x_{(1)}^{\vec{}} < x_{(2)}^{\vec{}} < \dots < x_{(n)}^{\vec{}}$. Assuming that \vec{x}_i is mutually independent, the distribution function of $x_{(k)}^{\vec{}}$ is equal to [59]:

$$\begin{aligned} F_{(k)}(y) &= P(x_{(k)}^{\vec{}} \leq y) \\ &= \sum_{i=k}^n \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i}, \end{aligned} \quad (4.65)$$

being $F(y)$ the cumulative distribution function of X . For a pair of images, having two vectors with the k^{th} , ($k = \{1, \dots, n\}$) order statistics of $x_{(k)}^{\vec{}}$, matching was performed according to the L_1 metric, i.e., the sum of the absolute differences between corresponding elements. As described in the experiments section, the k^{th} order statistics have evident discriminating information between individuals and was often one of the features automatically selected for the classification stage.

4.7.7 Linear Assignment Problem

The normalization of the iris data into a polar coordinate system propitiates invariance to translation and scale of the original data, but not to rotation, which appears as differences in translation of the normalized data and of the resultant clusters. For this, we used an automated method that seeks the maximal similarity between clusters, independent of their position in the normalized data, which was handled by a linear assignment strategy. Let $G = (U, V; E)$ be a bipartite graph with a separable set of vertices U and V ($|U| = |V| = n$) and a set of edges $E = \{e_{ij}\}$, such that e_{ij} denotes an edge from the i^{th} vertex of U to the j^{th} vertex of V . Let $c(e_{ij})$ denote the cost of the edge $c(e_{ij})$, such that $c(e_{ij}) \geq 0, \forall i, j \in \{1, \dots, n\}$. The linear assignment problem aims to find E^* , a subset of E that satisfies the following properties: 1) the accumulated cost of its edges is minimal, and 2) each vertex of

U and V appear exactly once in E^* . Let $\phi(i, j)_{\{\cdot\}} : \mathbb{N} \times \mathbb{N} \rightarrow \{0, 1\}$ be an indicator function, such that $\phi(i, j) = \mathbb{I}_{\{e_{ij} \in E^*\}}$. The *optimal* correspondence between elements of U and V is given by

$$\begin{aligned} \min \sum_{i=1}^n \sum_{j=1}^n \phi(i, j) c(e_{ij}) \\ \text{s.t. } \sum_{i=1}^n \phi(i, j) = 1, \forall j \in \{1, \dots, n\} \\ \sum_{j=1}^n \phi(i, j) = 1, \forall i \in \{1, \dots, n\}. \end{aligned} \quad (4.66)$$

Due to computational concerns, the problem was regarded as a shortest augmenting path algorithm with an implementation of the Dijkstra's shortest path method, which is known to run in time $O(n^3)$. Details can be found in the work of Jonker and Volgenant [79]. In practical terms, when matching two clustered iris images, the relative position of each cluster center is regarded as a vertex and included respectively in U and V . The cost (c) of edges E corresponds to the Euclidean distance between elements of U and V , which complies the above formalization.

4.7.8 Histogram Matching

In every phase of our method where the distance between histograms had to be obtained, several possibilities were tested, and the results were evaluated in a training data set. The best results were obtained with the cross-bin Quadratic-Chi distance histogram proposed by Pele and Werman [116]: let h_1 and h_2 be two non-negative bounded histograms, and let $A = [a_{ij}]$ be a non-negative and symmetric bib-similarity matrix, such that $a_{ii} \geq a_{ij}, \forall j \neq i$. The Quadratic-Chi histogram distance is given by (4.67), where $h_{\{\cdot\}}^i$ denotes the histogram value at position i , and m is a regularization factor (the best results were obtained in our experiments with 0.9).

$$QC(h_1, h_2) = \sqrt{\sum_{ij} \left(\frac{h_1^i - h_2^i}{(\sum_c (h_1^c + h_2^c) A_{ci})^m} \right) \left(\frac{h_1^j - h_2^j}{(\sum_c (h_1^c + h_2^c) A_{cj})^m} \right) A_{ij}}. \quad (4.67)$$

4.7.9 Shape Context Descriptor

Proposed by Belongie *et al.* [12], this descriptor provides an efficient way to measure the similarity between shapes, represented by a set of contour points $\{p_i\}$. For each p_i , we extract a histogram h_i of the relative coordinates of the remaining p_j points ($i \neq j$) with respect to p_i and represented in a log-polar coordinate system. Each h_i histogram is defined as the *shape context* of p_i and is used in all subsequent processing. Let p_1 and p_2 be boundary points of two shapes that are to be matched. The cost of matching p_1 with p_2 uses the χ^2 statistic:

$$C_{12} = \frac{1}{2} \sum_{k=1}^K \frac{(h_1(k) - h_2(k))^2}{h_1(k) + h_2(k)}, \quad (4.68)$$

where $h_1(k)$ and $h_2(k)$ denote the k^{th} bin of the histograms of p_1 and p_2 . The set of all costs C_{ij} between all pairs of points of two shapes is regarded as the cost matrix of a bipartite graph-matching problem and was solved as described in section 4.7.7. As illustrated in figure 4.30, this descriptor is an efficient way to extract discriminating information about the shape of the regions resulting from the data partitioning phase and is used as a soft biometric measure in the recognition process.

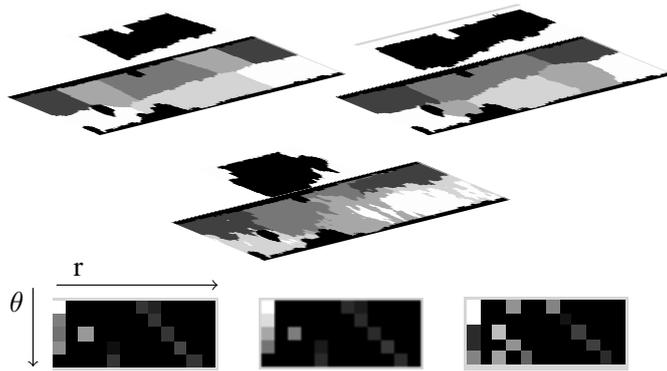


Figure 4.30: Illustration of the shape descriptor used to characterize each iris region. The upper row shows two similar shapes, from which shape context descriptors were extracted. The image at the center has a significantly different shape. Images in the bottom row illustrate the corresponding shape descriptors at point $L/4$, with L being the length of the contour and starting in the upper left pixel. Note the similarity between the far left and the centered descriptor and their dissimilarity to the far right image.

4.7.10 Robustness to Data Variation Factors

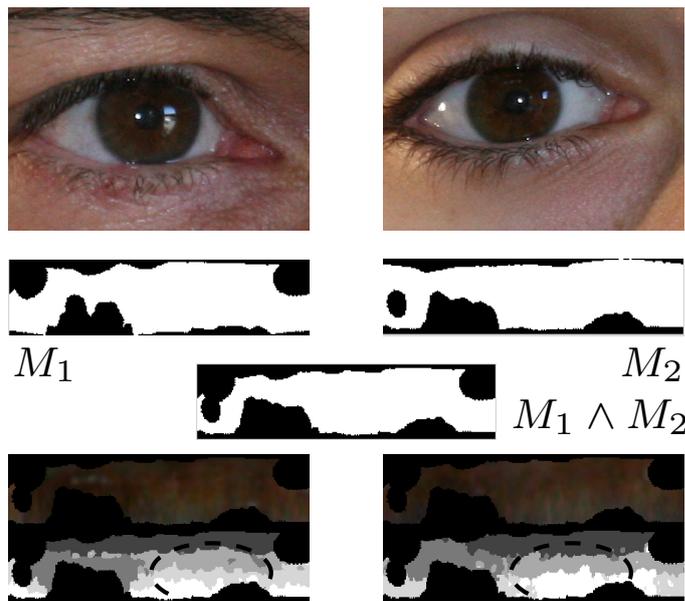


Figure 4.31: Clusters generated for two different heavy pigmented irises, where local contrast inside the iris ring is hardly perceived by a human observer. Even so, the appearance of the resulting group is evidently different. Images are "C_1_S1_I12.tiff" and "C_101_S1_I10.tiff" of the UBIRIS.v2 data set.

The basic premise of the proposed method is that the uniqueness of each iris texture determines that pixels are grouped in a specific way for each iris and compose clusters that are specific in terms of their positions and shapes, although these clusters cannot be expected to provide enough information for strong biometric recognition. Fig. 4.31 illustrates such discriminating ability, showing the clusters that

result from two different heavily pigmented irises. Here, the existence of four predominantly *horizontal* clusters in the left image is in opposition to the right image and is particularly evident in the regions delimited by the dashed ellipses.

Due to the dynamics of the acquisition setup, it is expected that the unoccluded regions of the iris will vary, which will affect the clustering results. This was overcome by obtaining the *conjunction noise-mask* of the pair of images to be matched (illustrated in Fig. 4.32), yielding two properties: 1) multiple biometric signatures are possible to extract from each image, depending on the other image that it will be matched against and 2) privacy concerns about the recognition process because it is required that the raw iris data and the corresponding noise-mask be stored in the database instead of the biometric signature.

Furthermore, it is important that the positions, sizes and shapes of regions are not subject to sudden or extreme changes as a result of the dynamics of the acquisition setup. Fig. 4.32 illustrates two images from the same eye acquired from different distances (nine and four meters). It can be seen that clusters remained relatively stable, essentially due to translation into the polar coordinate system and to the known property of invariance to color perception, as a result of moderate changes in scale.

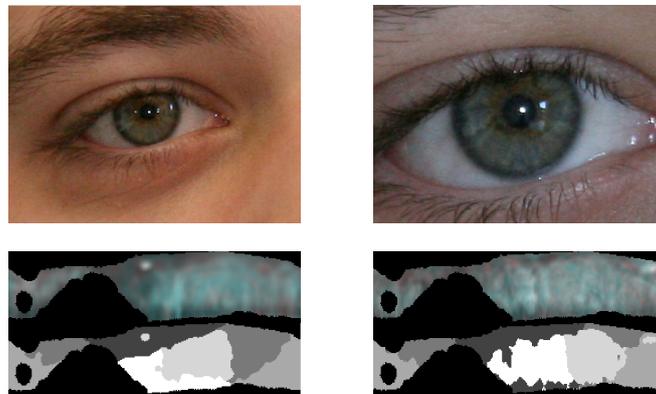


Figure 4.32: Robustness to changes in scale. Images are "C_111_S1_I4.tiff" and "C_111_S1_I13.tiff" of the UBIRIS.v2 data set.

The acquisition of a small moving target as the iris at relatively large and varying distances propitiates very different levels of image focus. Fig. 4.33 illustrates such variations. Although the similarity between the clustered images is evident, we observe that the shape of the clusters often becomes smoother in defocused data. If the defocus is exaggerated, the clustering process tends to augment the relevance given to spatial features, resulting in clusters with more regular shape.

Rotations in the original Cartesian space directly correspond to translations in the Polar coordinate system. However, significant changes in rotation are not expected due to the natural and biologically determined position of the head with respect to the neck and shoulders of stand-up subjects. Fig. 4.34 illustrates the behavior of the clustering process for a pair of images of the same eye where one of them was artificially rotated by $\frac{\pi}{6}$ (a value that is beyond the expected rotations). The relative position of clusters was shifted approximately $\frac{1}{12}$ of the width of the polar image. In this case, shapes remain roughly constant and the position of *corresponding* clusters varies significantly, which was handled by the Linear Assignment process described in section 4.7.7, which finds the optimal correspondence between clusters according to their shape.

Off-angle images are of special interest because gaze is known to be a primary source of error in traditional recognition strategies, particularly when circular iris parameterization techniques introduce differences in the phase of the normalized data and the bias phase-based in encoding/matching meth-

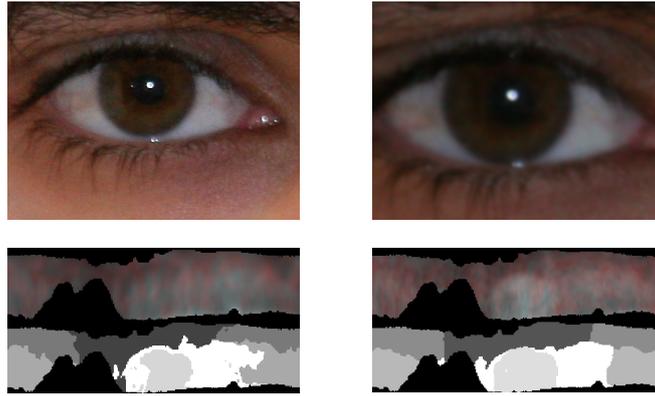


Figure 4.33: Robustness to defocused data. Images are "C_183_S2_I10.tiff" and "C_183_S2_I13.tiff" (defocused by a Gaussian kernel of $\sigma = 1.4$) of the UBIRIS.v2 data set.

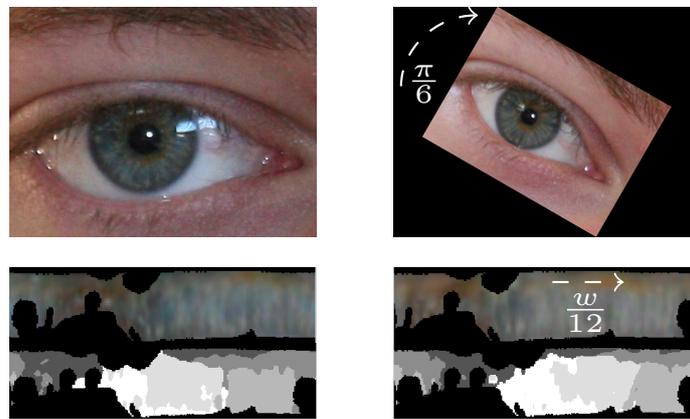


Figure 4.34: Robustness to changes in rotation. Images are "C_171_S1_I10.tiff" and "C_171_S2_I10.tiff" (rotated by $\frac{\pi}{6}$) of the UBIRIS.v2 data set.

ods. The translation into the polar coordinate system implies that the data are sampled at different rates with respect to the length of the iris ring at each angle ($a \gg b$ in the right image of figure 4.35, but $a \simeq b$ in the left image), which does not significantly affect the color perception of the resultant data. This relationship was observed even in cases where exaggerated deviations occlude portions of the iris. Furthermore, this figure gives a typical failure situation motivated by iris segmentation inaccuracies: the region delimited by the dashed ellipse in the right figure should have been classified as *noise* (corresponds to the upper part of the iris, partially occluded by eyelashes) but was erroneously considered for the clustering process and induced substantial differences in the resultant clusters near that region.

Lighting variations are due to the type of illuminants or to the amount of light in the environment and constitute a problematic factor, especially for local variations. The upper and middle row images of figure 4.36 were acquired from the same eye under substantially different lighting conditions but were mostly compensated by the *Retinex* process described in section 4.7.1 (compensated images are shown in the central column). Even so, higher variability in the shapes of the resulting clusters was observed, as highlighted by the regions delimited by the dashed horizontal ellipses. Finally, local lighting variations were observed to be the most problematic factor and to significantly bias the clustering process.

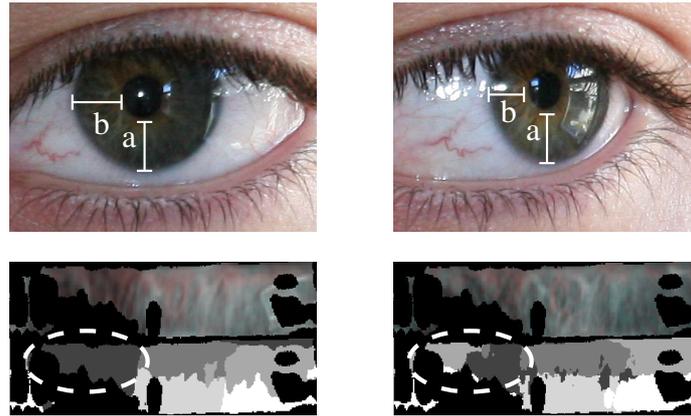


Figure 4.35: Robustness to off-angle image acquisition. Images are "C_24_S1_I13.tiff" and "C_24_S1_I15.tiff" of the UBIRIS.v2 data set.

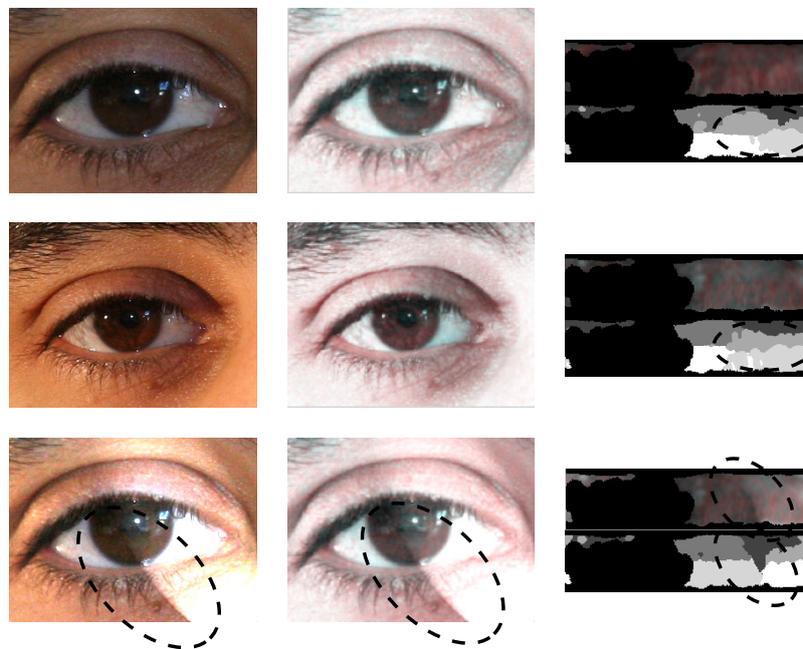


Figure 4.36: Robustness to global and nonuniform lighting changes. Images are "C_137_S1_I10.tiff" (top) and "C_137_S1_I7.tiff" (middle) and "C_137_S1_I10.tiff" (bottom, with a directional artificial light effect) of the UBIRIS.v2 data set.

Images at the bottom row illustrate such types of variations and, as highlighted by the diagonal dashed ellipses, the Retinex algorithm was not able to handle such variations, and the resulting clusters varied significantly.

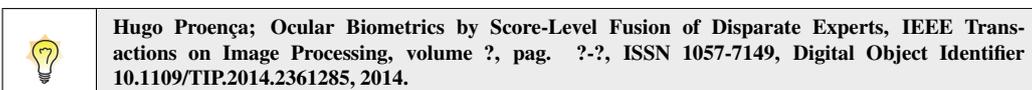
	<p>Iris Recognition Main Challenges:</p> <ul style="list-style-type: none"> Non-linear deformations in the iris texture Low entropy of heavily pigmented irises acquired in visible wavelengths Automatic selection of joint most discriminant bits.
---	--

4.8 Periocular Recognition

4.8.1 State-of-the-Art

The first work in this field was published in 2009, due to Park *et al.* [114]. They characterised the periocular region by local binary patterns (LBP), histograms of oriented gradients (HOG) and scale-invariant feature transforms (SIFT), fused at the score level. Subsequently, the same authors [115] described additional factors that affect performance, including segmentation inaccuracies, partial occlusions and pose. Woodard *et al.* [165] observed that fusing the responses from periocular and iris recognition modules improves performance with respect to each system considered individually. Bharadwaj *et al.* [16] fused a global descriptor based on five perceptual dimensions (image naturalness, openness, roughness, expansion and ruggedness) to circular LBPs. The Chi-square distances from both types of features were finally fused at the score level. Ross *et al.* [122] handled challenging deformed samples, using probabilistic deformation models and maximum-a-posteriori estimation filters. Also concerned about robustness, Woodard *et al.* [166] represented the skin texture and color using separate features, that were fused in the final stage of the processing chain. Tan *et al.* [153] proposed a method that got the best performance in the *NICE: Noisy Iris Challenge Evaluation*² contest. This method is actually a periocular recognition algorithm: texton histograms and semantic rules encode information from the surroundings of the eye, while ordinal measures and color histograms encode the iris data. Oh *et al.* [112] combined sclera and periocular features: directional periocular features were extracted by structured random projections, complemented by a binary representation of the sclera. Tan and Kumar [152] fused iris information (encoded by Log-Gabor filters) to an over-complete representation of the periocular region (LBP, GIST, HOG and Leung-Malik Filters). Both representations were matched independently and fused at the score level.

4.8.2 Fusion of Iris and Periocular Recognizers



As an attempt to increase the robustness of iris recognition in visible-light data, the concept of *periocular* biometrics has emerged, which compensates for the degradation in iris data by considering the discriminating information in the surroundings of the eye (eyelids, eyelashes, eyebrows and skin texture). Currently, the most relevant algorithms work in a holistic way: they define a region-of-interest (ROI) around the eye and use a feature encoding / matching algorithm regardless of the biological component in each point of the ROI. However, this augments the probability of sensitivity to some data covariate and the correlation between the scores extracted from the different points in the ROI.

We proposed a non-holistic approach to periocular recognition. Under an atomistic criterium, we devised two experts that use disjoint data, radically different recognition strategies and attain very different effectiveness. Here, we employ the term *weak* to refer to a recognition system that yields a poor separable decision environment, i.e., where the distributions of the genuine / impostor pairwise scores largely overlap. In opposition, the term *strong* refers to a system where the distributions of genuine and impostor scores almost don't overlap, resulting in a clearly separable decision environment and low error rates.

In our ensemble, the strong expert analyses the multi-spectral information in the iris texture, according to an automatically optimised set of multi-lobe differential filters (MLDF). Complementary, the weak expert parameterises the boundary of the visible cornea and defines a dimensionless ROI that comprises the eyelids, eyelashes and the surrounding skin. This expert helps to discriminate between individuals and has three interesting properties: 1) it analyses data that has an appearance independent

²<http://nice2.di.ubi.pt/>

of the iris texture; 2) it shows reduced sensitivity to the most problematic iris image covariates; and 3) it exclusively analyses traits that cannot be easily forged by anyone not willing to be recognised, which is in opposition to the traits classically used in periocular recognition (e.g., the shape of eyebrows). We encode the shape of eyelids, the distribution and shape of the eyelashes and the morphology of the skin wrinkles / furrows in the eyelids, which are determined by the movements of the *orbicularis oculi* muscles family. Fig. 4.37 overviews the proposed recognition ensemble and highlights some of its disruptive features with respect to the existing works.

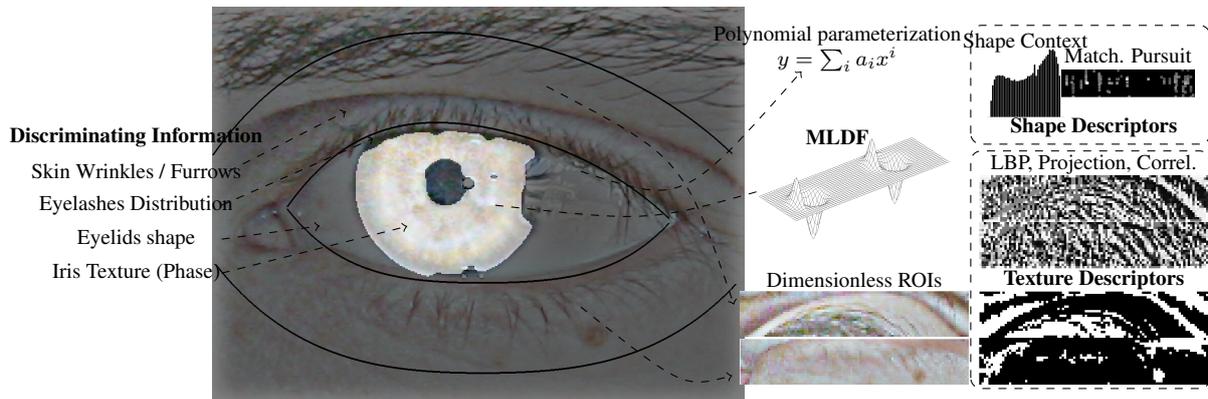


Figure 4.37: Cohesive overview of the ensemble recognition method proposed: a *strong* biometric expert encodes the information inside the iris by multi-lobe differential filters. The *weak* expert is based in the polynomial parameterisation of the shape of the visible cornea, from where two dimensionless regions-of-interest are defined. Shape and texture descriptors encode the discriminating information.

It is evident that using multiple sources for biometric recognition is not a new idea, and some controversy remains: is it actually an effective way to improve performance? It is argued that when a stronger and a weaker expert are combined, the resulting decision environment is averaged and the performance will be somewhere between that of the two experts considered individually [35]. Due to the way our ensemble was designed, our experiments support a radically different conclusion: even when the fused responses come from experts with very distant performance, the ensemble attains much better performance than the stronger expert (iris). This is due to the fact that both experts produce quasi-independent responses and are not particularly sensitive to the same image covariate, augmenting the robustness against degraded data.

4.8.3 Strong Biometric Trait: Iris Texture

Motivated by the Daugman’s pioneering method [34], there is a tradition of using phase-based techniques to encode iris data: phase is particularly discriminating between irises, if the alignment between gallery and probe samples is guaranteed. We also analyse the iris texture from the phase perspective, with three singularities: 1) to take advantage from the available multi-spectral data, the normalised iris is represented simultaneously in multiple perceptual color spaces; 2) inspired by the concept of ordinal filters [151], we extract not only the sign of coefficients but also consider their magnitude, using a sigmoid transfer function that eliminates the discontinuity of the sign function; and 3) we use a (filter) feature selection algorithm to find the optimal feature set, coming out with a compact yet effective representation of the iris.

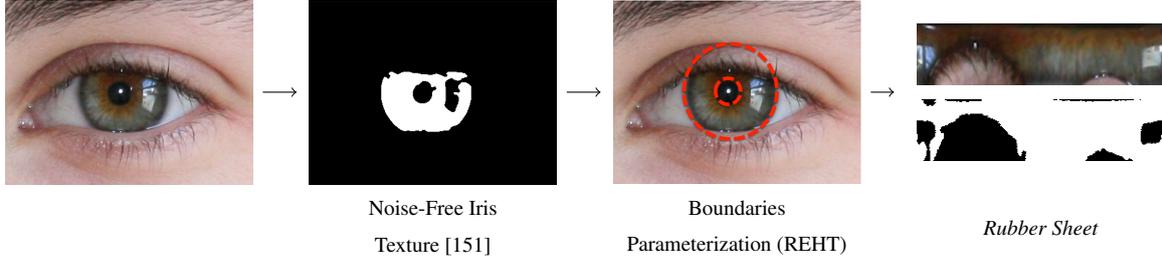


Figure 4.38: Processing chain for detecting the noise-free iris regions, parameterizing the iris boundaries and converting data into the polar domain.

4.8.4 Iris Segmentation and Parameterisation

There is a classical pattern recognition rule stating that “*weak data should be modelled with strong constraints*” [37]. Accordingly, an extremely robust algorithm for detecting the noise-free iris texture [151] was firstly used (producing the segmentation masks illustrated in the upper-right corner in Fig. 4.38) and non-concentric ellipses were considered to model the iris boundaries, according to the Random Elliptic Hough Transform (REHT). However, as the segmentation masks have shapes that are usually very far from elliptic, an objective function was designed to post-process the output of the REHT algorithm, and select the pair of ellipses that most likely corresponds to the biological (pupillary and scleric) iris boundaries. This function privileges the centre agreement between both ellipses and near-circular shapes. Let the i^{th} ellipse (out of t_e) be denoted by $\gamma_i = [x_i, y_i, a_i, b_i, p_i]$, being (x_i, y_i) the ellipse centre, a_i, b_i the major / minor axes and p_i the REHT score (the proportion of edge pixels in the segmentation mask that overlap γ_i). For $\binom{t_e}{2}$ pairs, the following objective function gives their goodness:

$$J(\gamma_i, \gamma_j) = \left[\frac{\|(x_i, y_i) - (x_j, y_j)\|_2}{\max(a_i, a_j)}, \frac{a_i - b_i}{a_i} + \frac{a_j - b_j}{a_j}, p_i + p_j \right] \cdot [\alpha_1, \alpha_2, \alpha_3]^T, \quad (4.69)$$

being α_i regularisation terms ($\alpha_1, \alpha_2 < 0, \alpha_3 > 0$). The deemed iris boundaries correspond to:

$$(\gamma_i^*, \gamma_j^*) = \arg \max_{i,j} J(\gamma_i, \gamma_j). \quad (4.70)$$

Converting the segmented data into a dimensionless pseudo-polar coordinated system (Daugman’s *rubber sheet* model) yields a set of normalised iris images and of segmentation masks that discriminate between the occluded and noise-free iris pixels (bottom-left images in Fig. 4.38). This enabled to obtain segmented images that were considered plausible (under visual inspection) in 92.60% for the UBIRIS.v2 and 95.40% for the FRGC data sets. The noise-free iris texture detection algorithm due to Tan *et al.* [153] proved to be remarkably effective against the typical data covariates (pose, gaze, iris occlusions and dynamic lighting conditions), which turned easier the tasks of the REHT and Rubber Sheet phases. Also, the most problematic case occurred in images severely occluded by eyelids where the pupillary and scleric boundaries are connected, which biased the results of REHT and, necessarily, of the Rubber Sheet phase.

However, in order to avoid that errors in segmentation bias the subsequent processing phases, and to perceive the strengths / weaknesses of the proposed ensemble, we filtered the wrongly / inaccurately segmented images and guaranteed that all images used were segmented in a plausible way.

Table 4.1: Summary of the perceptual color representations used to encode visible-light iris data.

Col. Sp.	Conversion (from RGB)	Description
RGB	—	No invariance properties
HSV	Algorithm described in [3]	Cylindrical-coordinate representations of points in RGB.
XYZ	Algorithm described in [157]	Using 2 nd observer, D65 illuminant.
LAB	Algorithm described in [66]	Using 2 nd observer, D65 illuminant. Close to perceptual uniformity.
RG	$\begin{pmatrix} R \\ G \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \end{pmatrix}$	Uses the normalised RGB color model (R+G+B=1). R and G are invariant to light intensity.
T-RGB	$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix}$	Invariant to scale and shift with respect to light intensity. Normalised against changes in light color and arbitrary offsets.
O-RGB	$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+\sqrt{2}G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}$	Based in the opponent color space. Intensity is represented in O ₃ channel and color in O ₁ and O ₂ channels that are shift-invariant with respect to light intensity.

4.8.5 Preliminary Selection of Color Spaces

To exploit the multi-spectral information available in visible-light images, they were represented in various perceptual color spaces, summarised in Table 4.1 (details about the conversions can be found in [141]). However, all these representations combined with the filter parameterisations lead to an intractable feature set. Hence, the data variability per color channel was assessed, and only a subset of the channels was considered for further processing, according to the concept of *eigeniris*. Let t be the number of images \mathbf{I} in a learning set, each one represented as a n -length column vector. Let $\bar{\mathbf{I}} = \frac{1}{t} \sum \mathbf{I}_i$ be the mean image, $\mathbf{D}_i = \mathbf{I}_i - \bar{\mathbf{I}}$ the residuals and $\mathbf{A} = [\mathbf{D}_1, \dots, \mathbf{D}_t]$ their concatenation ($n \times t$ matrix). Obtaining the eigenvectors of \mathbf{A} requires to work with $\mathbf{A}\mathbf{A}^T$, which has an intractable dimension. Let $\mathbf{A}^T\mathbf{A}$ be a $t \times t$ matrix and \mathbf{v}_i ($i \in \{1, \dots, t\}$) its eigenvectors. Then,

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i. \quad (4.71)$$

Multiplying both sides by \mathbf{A} , we have:

$$\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{v}_i) = \lambda_i(\mathbf{A}\mathbf{v}_i), \quad (4.72)$$

concluding that $\mathbf{A}\mathbf{v}_i$ are eigenvectors of $\mathbf{A}\mathbf{A}^T$. The *eigenirises* \mathbf{u}_i are given by $\mathbf{u}_i = \mathbf{v}_i\mathbf{A}$ and their magnitude gives the data variability in a color channel. Fig. 4.39 illustrates this concept by displaying (at the bottom histogram) the accumulated magnitude of the principal *eigeniris* \mathbf{u}_1 of each color channel (note that intensities are stretched for visualisation purposes). The bar plot at the top accumulates the magnitudes of the top-10 *eigenirises*, i.e., those associated with the largest eigenvalues. This plot highlights the variations among color channels: the RGB, Opposite-RGB, and intensity channel of HSV carry the predominating orthogonal variability, followed by the XYZ and L*AB. In opposition, the RG

and Transformed-RGB spaces had such small accumulated magnitudes that were disregarded from the subsequent phases of this work.

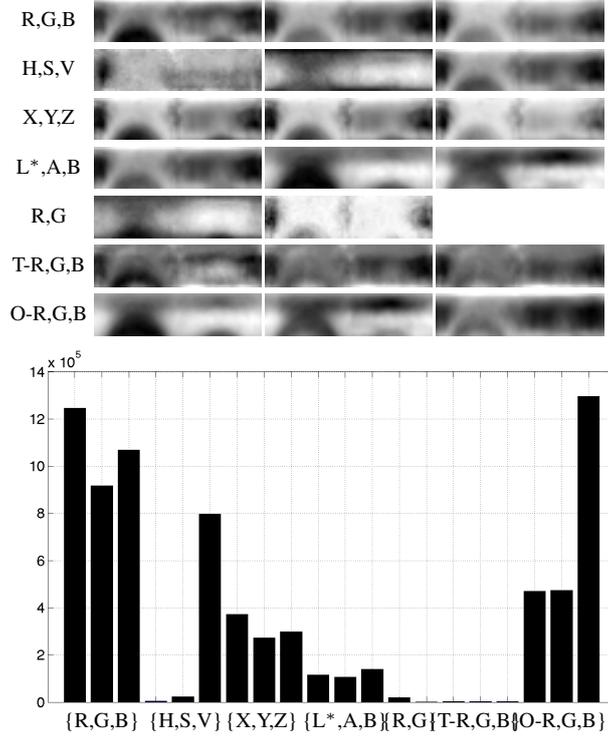


Figure 4.39: Top plot: Magnitude of the first eigenvector of each color channel (note that intensities are stretched for visualisation purposes). The bottom plot is the accumulated magnitude of the top-10 eigenvectors per color channel.

4.8.6 Feature Encoding

The iris codes were extracted by convolving the normalised data with a bank of Multi-Lobe Differential Filters (MLDF), recently reported as a relevant advance to the iris recognition field [150]. They are expressed in terms of the number of lobes, location, scale, orientation and inter-lobe distance. To keep the number of possibilities tractable, only filters with Gaussian kernels and equal number and scale of positive / negative lobes (1/1, 2/2, ...) were considered:

$$\mathbf{m}[\mathbf{x}_j, \mu_j, \sigma_j] = \sum_{j=1}^l (-1)^{j+1} \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left[\frac{-(\mathbf{x}_j - \mu_j)^2}{2\sigma_j} \right], \quad (4.73)$$

being $\mathbf{x}_j = (x_j, y_j)$ the center of each lobe and l the number of lobes. Fig. 4.40 illustrates examples of MLDFs, with varying number of lobes, scales and inter-lobes distances d .

To attenuate the reduced data resolution and amount of information available, not only the sign of the $I * m$ coefficients was considered but also their weighted magnitude. A transfer function with sigmoid shape was designed, mapping large magnitude values to 0/1, but also values near the vertical axis to the unit interval. This way, even considering values near the vertical axis as less reliable, it was accounted that they should contain *some* discriminating information and were still considered in the

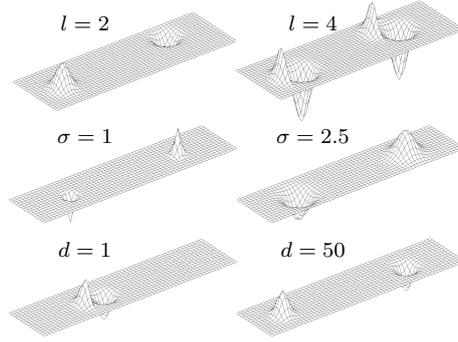


Figure 4.40: Examples of the filters used by the iris biometric expert, displaying varying number of lobes (top row), sigmas (middle row) and inter-lobes distance (bottom row).

matching process, with a smaller weight than for large magnitude values. According to this idea, the ℓ_2 norm was used as matching function between two iris codes. Fig. 4.41 compares the traditionally used sign-based strategy for codes quantisation (continuous line) and the proposed variant (dashed line). The horizontal axis corresponds to the values of $I * m$ and the vertical axis gives the corresponding weight in the matching process. It can be seen that, in both strategies, values with magnitude above ν are mapped equally to the $\{0,1\}$ values.

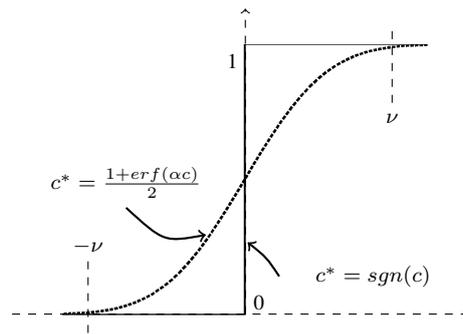


Figure 4.41: Comparison between the traditionally used sign-based function for codes quantisation (continuous line) and the sigmoid function proposed in this work (dashed line).

4.8.7 Learning Phase

The Sequential Floating Feature Selection (SFFS) algorithm [120] was used to select the best combination of features in a learning set. Other alternatives, such as the Fisher-score [47] and the Minimum Redundancy - Maximum Relevance (mRMR) algorithm [117] were tested, having obtained the following AUC values and corresponding 95% confidence intervals: 0.781 ± 0.020 (Fisher), 0.713 ± 0.018 (mRMR) and 0.950 ± 0.017 (SFFS) for the UBIRIS.v2 dataset and 0.815 ± 0.016 (Fisher), 0.860 ± 0.016 (mRMR) and 0.951 ± 0.018 (SFFS) for the FRGC. The values were obtained by selecting iteratively random samples of 90% of the available learning data and using the remaining 10% pairwise comparisons for performance evaluation.

According to these results, the SFFS algorithm was considered the most appropriate, with the following objective function:

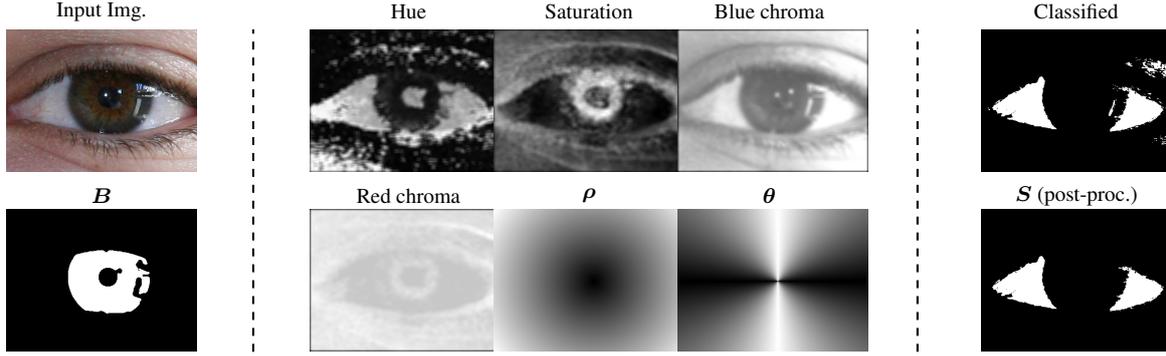


Figure 4.42: Phases of the sclera detection method. Using a feature set composed by image hue, saturation, blue and red chroma, the distance ρ and angle θ of each pixel to the iris centre of mass and major chord, a non-linear classifier detects the sclera pixels. Morphologic operators and analysis of connected components yields the result.

$$J(\{f, \mathcal{S}\}) = \frac{\sqrt{2}|\mu_{H_0}^{\{f, \mathcal{S}\}} - \mu_{H_a}^{\{f, \mathcal{S}\}}|}{\sqrt{(\sigma_{H_0}^{\{f, \mathcal{S}\}})^2 + (\sigma_{H_a}^{\{f, \mathcal{S}\}})^2}}, \quad (4.74)$$

being f a candidate feature, \mathcal{S} the set of selected features, μ and σ the mean and standard deviations of the ℓ_2 norm between two feature vectors. The subscript denotes the class (H_0 represents genuine comparisons) and $\{., .\}$ is the concatenation operator. Starting with the empty set $\mathcal{S} = \emptyset$, at each iteration the best feature was taken: $f^* = \arg_i \max J(\{f_i, \mathcal{S}\})$ and added to the selected set $\mathcal{S} \doteq \{\mathcal{S}, f^*\}$. After each insertion, the exclusion of features previously selected was considered: $f^* = \arg_i \max J(\{\mathcal{S} \setminus f_i\})$, where “ \setminus ” denotes set complement. If $J(\{\mathcal{S} \setminus f^*\}) > J(\mathcal{S})$, f^* was excluded from \mathcal{S} .

4.8.8 Weak Biometric Trait: Eyelids, Eyelashes and Skin

This section describes a biometric expert that analyses the surroundings of the human eye. The process is based in the segmentation of the iris and of the eyelids boundaries, defining a dimensionless ROI from where shape and texture descriptors are extracted. This ends up with an expert that is considered *weak*, in the sense that it cannot be used alone to reliably identify a subject, but is particularly useful to complement an iris biometric expert, due to its low correlation and reduced sensitivity to the most problematic iris data covariates.

4.8.9 Sclera Detection

The sclera can be detected at the pixel level [129], using as discriminating features the hue h and saturation s channels of the HSV color space, and the red cr and blue chroma cb values of the yCbCr space (Fig. 4.42). Geometrical information (angle and distance) of each position in the image with respect to the iris center of mass and major chord may also be helpful (the sclera is adjacent to the iris and spreads in opposite directions with respect to it), yielding a feature vector $[h_i, s_i, cr_i, cb_i, \rho_i, \theta_i]$, $\rho_i = \|(x_i, y_i) - (c_x, c_y)\|_2$, (c_x, c_y) is the iris centre of mass and $\theta_i = \arctan(|y_i - c_y, x_i - c_x|)$.

A binary non-linear classification model (feed-forward neural network) was learned, being its output illustrated in the upper-right corner of Fig 4.42. Next, morphologic operators smoothed the output of the classifier and only the two largest connected components were kept.

4.8.10 Eyelids Parameterisation

Let $M = B \oplus S$ be a mask combining the segmented iris B and sclera S (\oplus is the bitwise-or operator). Let $E = \{(x_i, y_i)\}$ be the set of t edge pixels in M (ordered clockwise) and $l^* = \arg \min_i x_i$, $r^* = \arg \max_i x_i$ the deemed positions of the eye corners. E can be divided into two subsets with indexes $e^{(1)} : \{l^*, \dots, r^*\}$ and $e^{(2)} : \{r^*, \dots, t, 1, l^*\}$, each one representing one of the eyelids. The following system of linear equations finds the coefficients of an interpolating polynomial of degree n :

$$\begin{bmatrix} x_{e_0}^n & x_{e_0}^{n-1} & \dots & x_{e_0}^0 \\ x_{e_1}^n & x_{e_1}^{n-1} & \dots & x_{e_1}^0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{e_n}^n & x_{e_n}^{n-1} & \dots & x_{e_n}^0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad (4.75)$$

being e_i an element of $e^{(\cdot)}$. The system above finds a solution in the least-squares sense, which most times is not acceptable due to the degradation of the masks M from where $e^{(\cdot)}$ are found. Instead, according to a RANSAC-like strategy, random samples of $n + 1$ distinct points in e were drew and a polynomial fitted to each sample (4.75). Every point was tested against that polynomial and in case it fits *relatively well* the model (ℓ_2 distance less than a threshold), its score was incremented (4.76) by a unit value. At the end, points with scores ϕ near the maximum value were considered inliers and the final polynomial found by least squares minimisation of inliers.

$$\phi^{(t+1)}(i) = \begin{cases} \phi^{(t)}(i) + 1, & \text{if } \|y_i - \sum_{j=0}^n a_j x_i^{n-j}\|_2 < \delta \\ \phi^{(t)}(i), & \text{otherwise} \end{cases}, \quad (4.76)$$

being $\delta \approx 1$ used in our experiments.

The next phase comprises the definition of the dimensionless ROI around the visible cornea. Let $d_c = \|r^* - l^*\|_2$ be the distance between the eye-corners, $\delta_x = \frac{d_c}{w_e}$, $\delta_y = \delta_x \frac{h_e}{w_e}$, and (h_e, w_e) the dimensions of the normalised ROI. Data were sampled from the Cartesian space according to the coordinates:

$$\begin{cases} x_{ik} = x_i + k\delta_x v_i^{(x)}, \\ y_{ik} = \sum_{j=0}^n a_j x_i^{n-j} + k\delta_y v_i^{(y)} \end{cases}, \quad (4.77)$$

for $k \in \{1, \dots, h_e\}$, $i \in \{1, \dots, w_e\}$, $\vec{v}_i = (v_i^{(x)}, v_i^{(y)})$ is the unit vector normal to the polynomial at $(x_i, \sum_{j=0}^n a_j x_i^{n-j})$. Fig. 4.43 illustrates this procedure: the leftmost column gives the initial image and, from left to right, its mask M and the polynomial parameterisations (red curves) are shown. Also, the ROIs around the eyelids are plotted in the Cartesian and dimensionless normalised spaces.

4.8.11 Feature Encoding and Matching

Two families of feature descriptors were considered: 1) Shape; to characterise the polynomial of each eyelid; and 2) Texture; to encode information in the ROIs.

4.8.12 Shape Descriptors

The shape descriptors used were based in the local accumulated curvature at the i^{th} point (out of t) in the eyelids boundary, given by $\sum_{j=1}^i \frac{\partial^2 y_j}{\partial x^2} / \sum_{j=1}^t \frac{\partial^2 y_j}{\partial x^2}$. Also, the shape context proposed by Belongie *et al.* [12] efficiently measures the similarity between shapes. For each (x_i, y_i) , a histogram h_i of $(x_i - x_j, y_i - y_j)$, $\forall j \neq i$ was represented in log-polar coordinates. The cost of matching h_i with h_j uses the χ^2 statistic:

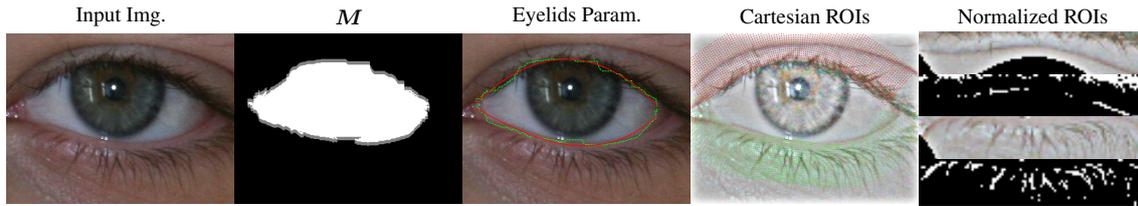


Figure 4.43: Example of a eyelids parameterisation (red curve in 'Eyelids Param. '), according to the boundary of the binary M masks. The right images give the regions-of-interest in the Cartesian and dimensionless normalised spaces.

$$R_{ij} = \frac{1}{2} \sum_{k=1}^K \frac{(h_i(k) - h_j(k))^2}{h_i(k) + h_j(k)}, \quad (4.78)$$

where $h_i(k)$ denotes the k^{th} bin of the histogram. The set of all costs R_{ij} between pairs of points was regarded as the cost matrix of a bipartite graph-matching problem and solved by a linear assignment algorithm.

Based in the work of Mallat and Zhang [104], $\mathbf{y} = \{y_1, \dots, y_n\}$ was considered a time-dependent 1D signal and decomposed into a linear expansion of signals taken from an over-complete dictionary:

$$\mathbf{y} = \sum_i a_i g_\gamma, \quad (4.79)$$

being g_γ the dictionary atoms and a_i the weighting factors. At each iteration, the atom g_{γ^*} that maximally correlates \mathbf{y} was subtracted from a residual, i.e., $r^{(t+1)} = r^{(t)} - a_n g_{\gamma^*}$. The process iterates until the ℓ_∞ norm of the residual is smaller than a threshold. The resulting \mathbf{a} values were matched by the χ^2 statistic (4.78).

Finally, Elliptical Fourier Descriptors [56] parameterise the \mathbf{y} coordinates by:

$$y_t = \sum_{i=0}^{\infty} [a_i, b_i] \left[\cos\left(\frac{2i\pi t}{T}\right), \sin\left(\frac{2i\pi t}{T}\right) \right]^T, \quad (4.80)$$

being $a_0 = 0$, $b_0 = \frac{1}{T} \int_0^T y(t) dt$, $a_i = \frac{2}{T} \int_0^T y(t) \cos\left(\frac{2i\pi t}{T}\right) dt$ and $b_i = \frac{2}{T} \int_0^T y(t) \sin\left(\frac{2i\pi t}{T}\right) dt$.

4.8.13 Texture Descriptors

Proposed by Ojala *et al.* [111], Local Binary Patterns (LBP) are among the most popular texture descriptors in the literature. The LBP value of an image pixel is actually a binary representation of the position of its neighbours with higher intensity, i.e., $f_{lbp}(x, y) = \sum_{p=0}^{t_p} s(I(x, y) - I(x', y')) 2^p$, being $s(\cdot)$ the Heaviside step function and (x', y') the coordinate of the t_p neighbours in a circular path. Histograms of the f_{lbp} values in image patches were concatenated and matched by the χ^2 statistic (4.78).

4.9 Analysis of *IrisCodes*: Bit Discriminability

	Hugo Proença; Iris recognition: What's Beyond Bit Fragility?, IEEE Transactions on Information Forensics and Security, volume 10, issue 2, pag. 321-332, ISSN 1057-7149, Digital Object Identifier 10.1109/TIFS.2014.2371691, 2015.
---	---

Most of the previous works that studied the effectiveness of the iris as a biometric trait concentrated in the levels of false rejections and in the concept of bit *fragility*, observing uneven levels of within-class variation among bits, i.e., the probabilities that bits "end up a 0 for some images of the iris and a 1 for other images of the same iris"[68] are uneven, as firstly formalised by Bolle *et al.* [17]. This work explores beyond the concept of fragility, by jointly considering the within-class and between-classes variabilities. The insight is that not only the probabilities of bits flipping among samples of one iris are uneven, but a similar phenomenon occurs for samples of different irises. i.e., some filters configurations used in particular regions of the irises augment the probability that bits predominantly take a particular value. We propose the concept of bit *discriminability*, which shares the roots of the Fisher discriminant and has an intuitive (visual) representation. A discriminant bit should: 1) keep a constant value among samples of one iris; and 2) have a value that is independent of the values in codes from different irises. According to this formulation, the *discriminability* can be regarded as an extension to *fragility*: a bit is fragile due to the high probability of flipping its value in genuine comparisons. To classify a bit as discriminant, we regard not only its fragility but also the probability of obtaining agreeing values in impostor comparisons.

Based on this concept, we infer the suitability of each region of the iris for biometric recognition. Three additional novelties are given: 1) results are shown not only for the classical Gabor-based texture description, but also for Multi-Lobe Differential Filters (MLDF) [150], which were reported as a relevant advance to the field; 2) we consider different levels of image quality, corresponding to a broad range of data acquisition protocols; and 3) we compare results for multi-spectral data (near-infrared (NIR) and visible wavelength (VW)), enabling to perceive the potential of each spectrum for biometric recognition.

All the conclusions inferred were based in four well known data sets: 1) the University of Bath, representing good quality NIR data; 2) the CASIA-Iris-Distance, representing NIR data of moderate quality; the 3) UBIRIS.v2 and 4) FRGC datasets, both representing VW data acquired in uncontrolled setups.

4.9.1 Study of Iris Codes

Several works studied the nature of the iris texture and the properties of the resulting codes. As relevant examples, Kong *et al.* [89] provided a deep understanding of the geometric structures of the codes, regarded as a clustering algorithm. These authors showed the relation between the Hamming distance used in matching and the bitwise phase distance, arguing that Gabor kernels are actually phase-steerable filters. Subsequently [90] [91], Kong focused on the geometrical relationships of bits in iris codes, regarded as convex polyhedral cones. The relationships detected imply that a property (central ray) is enough to reveal patterns among codes, which might be used to break systems without a liveness and quality checker.

The recognition performance respect to covariates were also previously studied: Bowyer *et al.* [20] tested three of these factors: 1) the effect of pupillary dilation in performance; 2) the iris stability over lifetime; and 3) the effect of contact lenses. They concluded that these factors bias the genuine distribution toward the impostors', but also confirmed that the probability for false acceptances is practically invariant to these factors. More recently, Mehrotra *et al.* [108] claimed that the movement of the genuine match scores toward the impostors distribution was due to other covariates (such as blur, occlusions and pupillary dilation), perhaps even at a higher degree than the ageing effect.

Concerns about the fragility of some bits in the iris codes date back to the earliest implementations of the acknowledged Daugman's recognition algorithm, by disregarding the bits with responses near the axes. Then, Bolle *et al.* [17] introduced the term of *fragile* bit and observed that, due to imaging noise, not all bits have equal possibilities to flip among samples of one iris. Hollingsworth *et al.* [67] analyzed the fragility in iris codes and a similar idea had been reported in [127]. Subsequently, Hollingsworth *et al.* [68] found that the middle bands of the iris are better than the inner parts and that large filters provide more consistent bits than small filters, due to the attenuated effect of acquisition artefacts. Finally,

the same authors used the notion of bit *fragility* to propose [69] a new matching distance based in the linear combination between the proportion of disagreeing bits and the *fragile bit distance* (FBD), that expresses the fraction of unoccluded bits masked for fragility in the comparison. They observed that the FBD carries complementary information to the traditional distance and that results obtained by fusing both measures are better than when using any of these alone. The discriminability of the bit coefficients due to the coarse quantization of the phase response was also studied [68], being suggested to ignore bits with amplitude in the lower quartile.

In terms of the selection of the most reliable bits of iris codes, Dozier *et al.* [43] based their work in the concept of bit fragility. When compared to the classical code of 2,048 elements, they were able to reduce the number of bits by 30%, without significantly increasing the error rates. They even reduced the number of bits in 90%, but in this case observed a significant increase in the error rates. Rathgeb *et al.* [134] obtained the bit-error occurrences and a corresponding global-rank of bit positions. Based on this information, the less reliable bits were discarded, which improved performance and simultaneously reduced the size of codes.

4.9.2 Bit Discriminability

For comprehensibility, we adopt a notation similar to the used by Bolle *et al.* [17]. Let $\mathcal{I}^{(p)}$ and $\mathcal{I}^{(q)}$ be two real world irises, from where the binary iris codes $\mathbf{C}^{(p)} = F(\mathcal{I}^{(p)})$ and $\mathbf{C}^{(q)} = F(\mathcal{I}^{(q)})$ are extracted (F is a composition of an imager and a feature encoding system). There are two hypotheses:

H_0 : $\mathbf{C}^{(p)}$ and $\mathbf{C}^{(q)}$ are from the same iris ($p = q$);

H_a : $\mathbf{C}^{(p)}$ and $\mathbf{C}^{(q)}$ are from different irises ($p \neq q$).

Let $C_k^{(p)}$ denote the k^{th} bit (out of t) of an iris code, i.e., $\mathbf{C}^{(p)} = \{C_1^{(p)}, \dots, C_t^{(p)}\}$. We are interested in defining a metric for each bit discriminability, in terms of its (a priori) effectiveness for biometric recognition.

Definition 4.9.1. Let \oplus denote the exclusive-or logical operation. The k^{th} bit is considered *discriminant* for biometric recognition if two conditions are met: 1) $1 - P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0) < \delta$; and 2) $|\frac{1}{2} - P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)| < \delta$, for a small $\delta \in \mathbb{R}^+$.

The k^{th} bit *contributes* for a *Type I* classification error (false match) with probability $P(H_a) P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$. Similarly, it contributes for a *Type II* classification error (false non-match) with probability $P(H_0) P(C_k^{(p)} \oplus C_k^{(q)} = 1 | H_0)$, i.e., $P(H_0) (1 - P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0))$. Hence, the probability that the bit contributes for a classification error ϵ is given by:

$$\begin{aligned} \epsilon(k) = & P(H_0)(1 - P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0)) \\ & + P(H_a)P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a), \end{aligned} \quad (4.81)$$

where $P(H_0)$ and $P(H_a)$ are the prior probabilities for genuine and impostors comparisons.

Definition 4.9.2. Let $x_k = P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0)$ and $y_k = P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$. We define (x_k, y_k) as the *visual representation* of the bit discriminability.

For the discussion below, two (readily satisfied) assumptions are made:

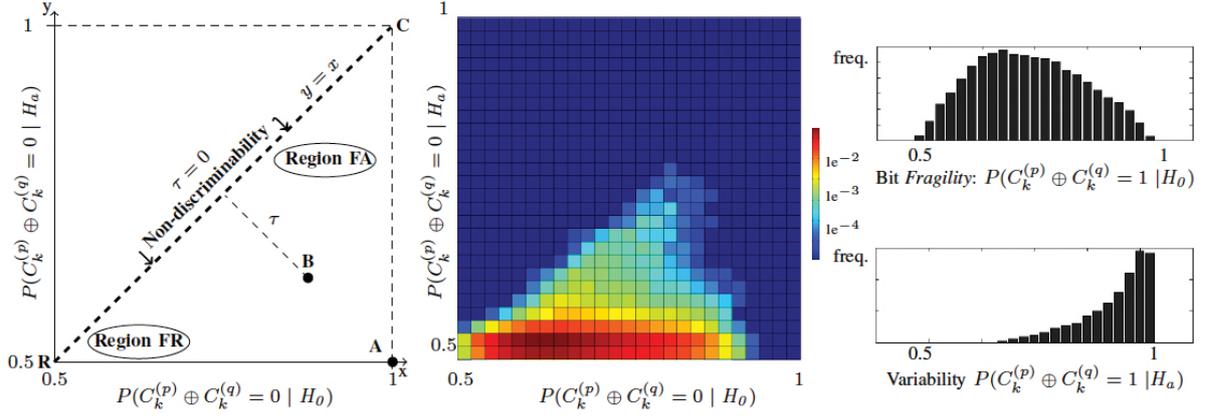


Figure 4.44: The left figure is a schematic representation of bit *discriminability*. The central plot is a 2D histogram of the discriminability of bits extracted from the University of Bath data set. The histograms at the right side are the vertical and horizontal projections of the center plot: the upper histogram evidence different levels of bits fragility, whereas the bottom histogram points for additional non-random variation.

- we assume that $P(H_0) = P(H_a) = 0.5$, i.e., for the purpose of our analysis we make no assumptions about the prior probabilities of genuine and impostor comparisons. However, it should be noted that in most practical scenarios $P(H_a) \gg P(H_0)$, i.e., for databases with a single template per eye, an identification process will require much more impostor than genuine comparisons (at most one).
- we assume that $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0) \geq P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$. In any practical biometric system, there is no reason for observing agreeing bits more likely in impostors than in genuine comparisons.

Definition 4.9.3. A quantitative *measure* of the bit discriminability is given by the Euclidean distance between (x_k, y_k) and the straight line $y = x$, $\tau(k) = |y_k - x_k|$.

Theorem 4.9.1. Let (x_i, y_i) and (x_j, y_j) be the visual representations of the discriminability of the i^{th} and j^{th} bits of iris codes, such that $\tau(i) > \tau(j)$, and $x_i \geq y_i \geq 0.5$. Then, $\epsilon(i) < \epsilon(j)$ and the i^{th} bit is less likely to contribute for a classification error than the j^{th} bit.

Proof. By hypothesis, $\tau(i) > \tau(j)$, i.e., $|y_i - x_i| > |y_j - x_j|$. Then, $y_i - x_i > y_j - x_j \vee x_i - y_i < x_j - y_j$. As $x_i \geq y_i$, it follows that $x_i - y_i \geq 0$. Then, $x_i - y_i > x_j - y_j$. Multiplying both sides by -1, we know that $y_i - x_i \leq y_j - x_j$. Adding a constant in both sides and dividing everything by another constant, we have $\frac{(1-x_i)+y_i}{2} \leq \frac{(1-x_j)+y_j}{2}$, i.e., $\epsilon(i) < \epsilon(j)$. \square

Fig. 4.44 illustrates the concept of bit discriminability: the central plot is the 2D histogram of the (x, y) visual representations for bits extracted from the University of Bath dataset. The left plot schematizes this histogram, and marks the non-discriminability line ($y = x$). The "A" symbol denotes an *optimal* feature and "C" denotes features that keep the same value both for genuine and impostor comparisons. At the other extreme, the "R" region corresponds to features that behave randomly (like dynamic noise). Features close to "R" predominantly contribute for Type-II classification errors (Region FR), and features in the upper-right corner contribute more for Type-I classification errors (Region FA).

Table 4.2: Results of the Kolmogorov-Smirnov [64] normality test, using the *null* hypothesis that $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$ follow a Binomial distribution $\mathcal{B}(n, 0.5)$, at the 5% confidence level. Results are given for all the data sets considered.

Dataset	n	<i>null</i> hypothesis	P-value
BATH	200,000	χ	$6.80e^{-4}$
CASIA	200,000	χ	$1.93e^{-6}$
UBIRIS.v2	200,000	χ	$6.71e^{-11}$
FRGC	200,000	χ	$5.27e^{-10}$

The fact that most bits fall in the region FR is the root for the extraordinary small probability of false acceptances in current recognition systems, but also justifies their relatively high false rejection rates.

The bar plots at the right side of Fig. 4.44 are particularly important for the context of this work: the upper plot (vertical projection) gives evidence of the levels of bit *fragility*. Complementary, the bottom plot (horizontal projection) shows the varying levels for $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$, and supports the concept of bit *discriminability*.

However, the concept of discriminability depends of whether the relative frequency of $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$ is not simply function of a random effect. In that case, values should follow a Binomial distribution $\mathcal{B}(n, p)$, where n is the number of pairwise bit comparisons, each of which yields success with probability p and probability mass function given by:

$$P\left((C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a) = \frac{s}{n}\right) = \binom{n}{s} p^s (1-p)^{n-s}, \quad (4.82)$$

where s is the number of successes (bits agreement). Using a large set of n pairwise bit comparisons and assuming that $p = 0.5$, the Normal distribution $\mathcal{N}(np, \sqrt{np(1-p)})$ may be used to closely approximate results from the afore binomial distribution. According to the Kolmogorov-Smirnov [64] normality test, the *null* hypothesis stating that values follow $\mathcal{B}(n, p)$ was rejected with asymptotic P-values lower than $1e^{-7}$ for all the datasets used in the experiments:

According to the results given in Table 4.2, it can be concluded that $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$ varies consistently with respect to some other factor apart randomness. Also, this phenomenon is more notorious for VW than for NIR data (substantially lower P-values for VW than for NIR data), which we believe to have roots in the corneal reflections determined by the ambient VW wavelengths that are not blocked in the camera. In this case, the images of the different subjects tend to display brighter intensities in similar positions of the iris, which has some influence in the $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$ value.

4.9.3 Datasets and Preprocessing

Four freely available data sets were used in the experiments, each one representing a data acquisition scenario. Fig. 4.45 illustrates some of the images considered: the upper row regards the BATH data set and the subsequent rows represent the CASIA-Iris-Distance, UBIRIS.v2 and FRGC data sets.

- The University of Bath data set³ contains 32,000 NIR images from 800 subjects. From these, 6,000 images from 1,000 different classes (eyes) with very good quality were considered, to represent the optimal conditions where a recognition system work. All irises are sharp, without relevant occlusions and in frontal view.

³<http://www.smartsensors.co.uk/products/iris-database/32-000-full-set/>

- The CASIA-Iris-Distance set⁴ was collected by the CASIA long-range device in a relatively unconstrained setup. Images feature blink, motion blur, off-axis gaze and other small anomalies, representing NIR data of moderate quality. A set of 9,521 images (127 subjects, 814 classes) was used, for which segmentation and noise detection was confirmed by visual inspection.
- The UBIRIS.v2 [128] dataset has 11,102 images from 261 subjects, acquired at visible wavelengths between three and eight meters away, under dynamic lighting conditions and unconstrained setups. Images are high heterogenous in terms of quality, with glossy reflections across the iris, significant occlusions due to eyelids and eyelashes, off-angle and blurred data. 5,340 images from 518 classes) were selected from this dataset, all of them accurately segmented. All these images were converted to grayscale.
- The FRGC [119] data set served initially for face recognition experiments and is a specially hard set for iris recognition, due to its limited resolution. The still images subset from both the controlled / uncontrolled setups was used. Images are typically frontal, with varying amounts of light, shadows and glossy reflections that occlude portions of the irises. 4,360 from 868 classes images were selected from this data set. All these images were reasonably segmented, according to visual inspection, and were converted to grayscale.
- The UBI_SPECTRAL is a set of iris data acquired in a synchronous way in the NIR and VW wavelengths, with a multispectral JAI AD080-GE camera, in a laboratorial controlled acquisition protocol. It contains data from 34 subjects, with 80 images per subject (20 NIR and 20 visible images, divided into 2 acquisition sessions). All images are frontal, sharp and practically noise-free. All the images were accurately segmented and VW images was converted to grayscale.

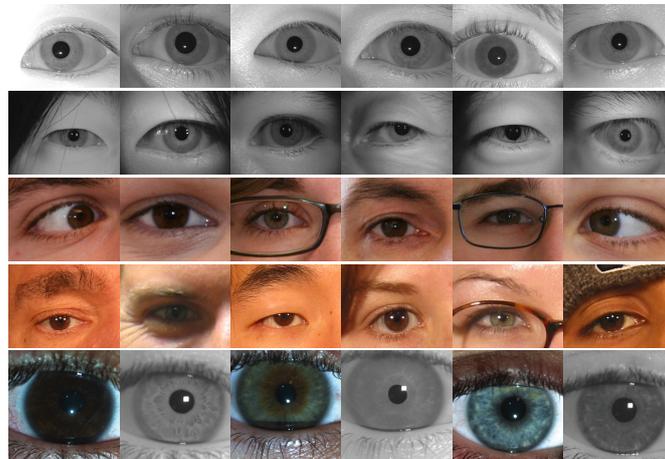


Figure 4.45: Examples of the data sets used in the experimental evaluation. From top to bottom rows: BATH, CASIA-Iris-Distance, UBIRIS.v2, FRGC and UBI_SPECTRAL datasets.

For the BATH, CASIA, UBIRIS and FRGC data sets, random samples composed by half of the within-class comparisons available and the same number of between-classes comparisons were created. Next, in an iterative way, fully disjoint sets of the learning data were used to evaluate the recognition performance and estimate $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_0)$ and $P(C_k^{(p)} \oplus C_k^{(q)} = 0 | H_a)$. Starting with $t_w = 1,000$ within-class and $t_b = 5,000$ between-classes comparisons, the recognition performance was

⁴<http://biometrics.idealtest.org/>

obtained. At each iteration t , the number of comparisons was increased by a constant factor ($t_w^{(t+1)} = 1.1 t_w^{(t)}$, $t_b^{(t+1)} = 1.1 t_b^{(t)}$) until the performance values converged (after $t = 49$ iterations).

The procedure described above can also be regarded as a way to mitigate the different number of degrees-of-freedom (DOF) in each sample and the way this factor might increase the correlation between bits and bias the subsequent results. Note that the data samples had varying number of classes, corresponding to different DOFs in the sets of pairwise comparisons. Even though, as the learning process was only stopped when performance was observed to converge, this implies that for large amounts of data, these changes in the number of DOF do not lead to substantial changes in the recognition performance. Finally, it should be stressed that none of these comparisons was used in feature selection, i.e., the learning and test sets were mutually exclusive.

As depicted in Fig. 4.46, for all the data sets considered in this work, the unoccluded regions of the irises were obtained according to the algorithm of Tan *et al.* [151] (leftmost image). Next, an elliptical parameterization was chosen for both iris boundaries, using the random elliptic Hough transform. Based on the parameterization of the pupillary and scleric iris boundaries, the translation into the dimensionless pseudo-polar coordinate system was carried out according to the Daugman's *rubber sheet* model (rightmost image).

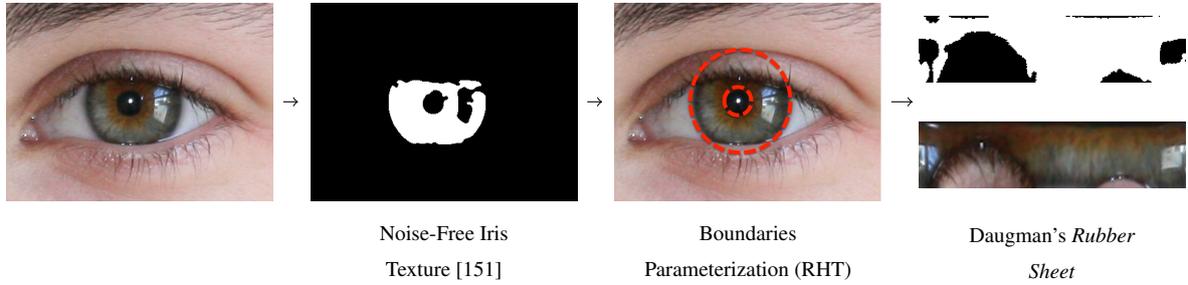


Figure 4.46: Processing chain for segmenting the irises, detecting the noise-free iris regions, parameterizing the boundaries and converting them into the polar domain.

4.9.4 Amount of Information in Iris Patches

The amount of information available in small iris patches was measured by the Shannon entropy criterion, quantifying (in terms of bits) the expected value for the amount of information in square regions $p \times p$ of the normalized image I :

$$h(\mathbf{I}_{p \times p}) = - \sum_i P(\mathbf{I}_{p \times p} = i) \log_2 (P(\mathbf{I}_{p \times p} = i)), \quad (4.83)$$

where $P(\mathbf{I}_{p \times p} = i)$ is the probability for the i^{th} intensity in the patch.

In order to fairly compare the local entropy between NIR and VW data, without concerns about the lighting conditions, the levels of iris pigmentation of the intrinsic features of the subjects in each dataset, the UBI_SPECTRAL data set was used. In this case, as all images were acquired in highly controlled lighting conditions and in synchronous way for the NIR and VW wavelengths, the effect of the above factors should be minimised. Also, we used the average intensity inside the iris of the grayscale version of the VW images as an estimator of the levels of iris pigmentation. This way, the highest values correspond to light pigmented irises (light blue), whereas the lowest intensities are from the heavily pigmented irises (dark brown / black). Fig. 4.47 compares the box plots of the local entropy values obtained for the NIR (left plot) and VW (right plot) data, with respect to the levels of iris pigmentation

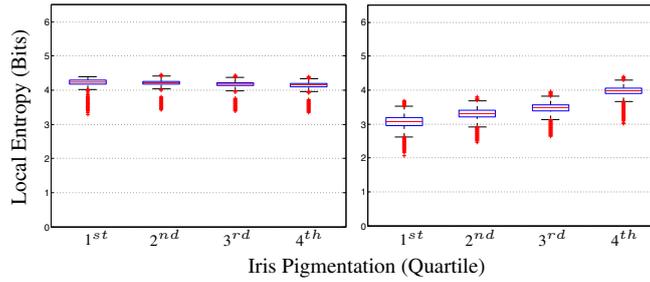


Figure 4.47: Comparison between the average entropy values (4.83) observed in iris patches of NIR (left plot) and VW data (right plot). Results regard the UBI_SPECTRAL set.

(horizontal axes). Four groups of pigmentation were considered, corresponding to dark brown / black (1^{th} quartile of the average intensities), light brown (2^{nd} quartile), green / dark blue (3^{rd} quartile) and light blue (4^{th} quartile) irises. We confirmed that values vary much more in VW than in NIR data, and, for the former wavelength, is notoriously higher for light pigmented than for dark irises. For NIR data, the heavy pigmented irises (3^{rd} and 4^{th} quartiles) have slightly higher local entropy than the remaining classes, which is in exact opposition of the VW case. Also, the entropy in patches across the iris is more heterogeneous for VW images than for NIR, which is particularly evident for light pigmented irises. With regard to local variations, heavy pigmented irises acquired in VW have not only a relatively low amount of local information, but also display low variability between patches. i.e., in practice provide much flatter distributions for VW than for NIR data.

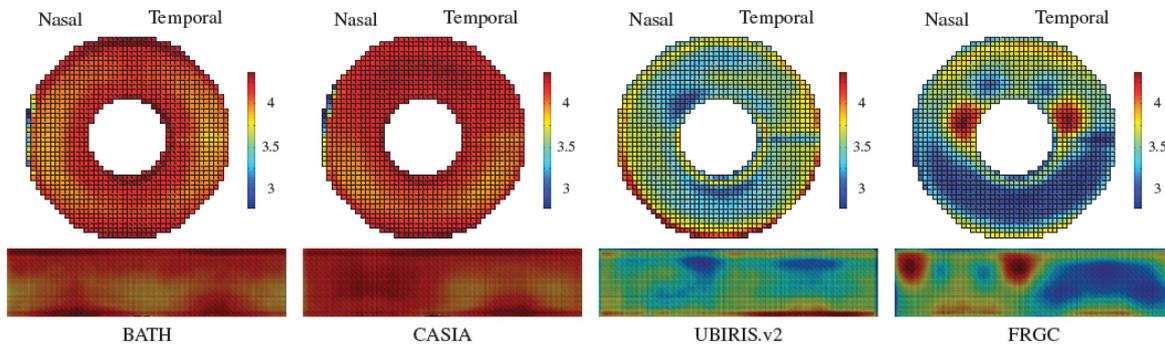


Figure 4.48: Average amount of information (Shannon entropy in 9×9 patches of the normalised images) across the different regions of the irises in the BATH, CASIA-Iris-Distance, UBIRIS.v2 and FRGC datasets. Values are expressed in bits, and enable to perceive the gap of information between NIR (BATH and CASIA) and VW (UBIRIS.v2 and FRGC) iris data.

For the remaining datasets, Fig. 4.48 quantifies the amount of information in $p = 9$ patches. Even noting that the comparison between data sets might be unfair (the original images have different resolution), the immediate conclusion is the higher homogeneity of values observed in NIR data than in the VW case. Note that the average values were also much higher in NIR than in VW data, which actually implies that the NIR images provide more heterogeneity in terms of intensities in iris patches than VW data.

Also, we observed that the pupillary regions are the most valuable in NIR images, which is not evident in VW. Regarding the FRGC dataset, there are two regions near the pupillary boundary with

Table 4.3: Types and range of the filters parameters varied in our experiments.

Gabor Filters $g[.,.]$	
Wavelength (px.)	$\omega : \{1 : 1 : 14\}$
Orientation	$\varphi : \{0, \pi/4, \pi/2, 3\pi/4\}$
Gaussian Sigma	$\sigma : 0.65\omega$
MLDF Filters $m[.,.]$	
Num. Lobes	$t_l : \{1/1, 2/2, 3/3, 4/4\}$
Gaussian Sigma	$\sigma : \{1, 2, 3, 4, 5, 6\}$

values notoriously higher than the remaining regions. We confirmed that they were due to frequent reflections not detected by the noise-free segmentation phase. Also, noted that in the FRGC set the bottom parts of the irises have evidently smaller amounts of information than the upper parts, probably due to the lighting sources from above that propitiate shadows in these regions.

4.9.5 Filters Parameterizations

The discriminating power provided by each region of the iris was assessed with respect to two families of filters: 1) Gabor kernels, which faithfully model simple cells in the visual cortex of mammalian brains [33] and are used in the most acknowledged iris recognition algorithm; and 2) Multi-lobe differential filters (MLDF), which were recently reported as a relevant advance in the iris recognition field [150].

The impulse response of a Gabor kernel is defined by the multiplication of a harmonic and a Gaussian function:

$$\mathbf{G}[x, y, \omega, \varphi, \sigma] = \exp\left[\frac{-x^2 - y^2}{\sigma^2}\right] \exp[2\pi\omega i\Phi], \quad (4.84)$$

where $\Phi = x \cos(\varphi) + y \sin(\varphi)$, ω is the spatial frequency, φ is the orientation and σ the standard deviation of a Gaussian kernel (isotropic in our experiments, $\sigma = 0.65\omega$). A more general form of Gabor filters can be found in the literature (e.g., [37]), allowing for different scales along the axes (σ_x and σ_y). To keep moderate the dimension of the parameterisation space, we decided to use exclusively filters with the same scale along the axes.

Regarding the MLDF filters, they can be parameterised in terms of the number of positive/negative lobes, location, scale, orientation and inter-lobe distance. To keep the number of possibilities moderately low, only Gaussian kernels with balanced number of positive / negative lobes (1/1, 2/2, ...) and equal scale for both types of lobes were considered. Hence, the MLDF filters are expressed by:

$$\mathbf{m}[\mathbf{x}_j, \mu_j, \sigma_j] = \sum_{j=1}^{t_l} (-1)^{j+1} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[\frac{-(\mathbf{x}_j - \mu_j)^2}{2\sigma_j}\right], \quad (4.85)$$

where $\mathbf{x}_j = (x_j, y_j)$ is the center of each of the t_l lobes. Next, $\mathbf{k} = \{\mathbf{m}, \mathbf{g}\}$ filters were convolved with each normalized iris image \mathbf{I} , providing a set of coefficients. The sign of the coefficients was obtained, i.e., \mathbf{C} is the vector representation of $\text{sgn}(\mathbf{I} * \mathbf{k})$. Fig. 4.49 illustrates the filters used and Table 4.3 summarizes the range of parameters considered ($\{a : b : c\}$ denotes values in the $[a, c]$ interval, with steps of size b).

Fig. 4.50 expresses the variations in discriminability with respect to each parameter of the filters. The continuous lines represent the BATH dataset, the dashed lines with the diamond marks regard the CASIA-Iris-Distance. The UBIRIS.v2 is given by the dotted lines with triangular marks and the FRGC dataset by the dashed lines with circular marks. Above each plot we illustrate a normalized iris image and

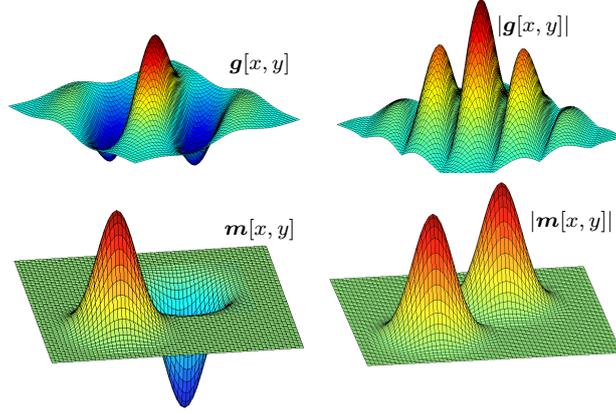


Figure 4.49: Illustration of the filters used in our experiments ($\{g, m\}$) and of the filters that give the contribution of each position in the iris to the coefficient in the iris code $\{|g|, |m|\}$.

represent the filters that correspond to the nearby positions in the plot. Generally, the discriminability was substantially higher for MLDF than for Gabor filters. In case of the latter filters, larger wavelengths consistently increased the discriminability, essentially because they have a reduced sensitivity to outlier values due to acquisition artefacts. Orientation is another relevant parameter for Gabor kernels, where filters that analyze features that spread radially in the normalized data provided much better results. Regarding MLDF filters, filters with more lobes got worse results, which might be due to the *cross-elimination* effect of differences between lobes. Surprisingly, the variation in results with respect to the sigma of the Gaussian kernel were not so evident as in the case of Gabor kernels.

4.9.6 Bit Discriminability

The discriminability τ of each bit extracted was obtained. Note that the iris patches used in the convolution for each bit contribute to the result in different degree, according to the magnitude of the kernel at each point, i.e., if a kernel has very small value at a specific position, the corresponding intensity on the patch almost does not affect the result. This way, the contribution of each location $[x, y]$ in the iris to the bit value is given by:

$$\Psi[x, y] = \frac{\sum_i \left(|\mathbf{k}_i[x - r_i, y - c_i]| \tau(i) \right)}{\sum_i |\mathbf{k}_i[x - r_i, y - c_i]|}, \quad (4.86)$$

where $[r_i, c_i]$ is the central position of the i^{th} filter \mathbf{k}_i and $\tau(i)$ is the discriminability of the i^{th} bit.

Fig. 4.51 gives the discriminability provided by each region of the iris in the Cartesian and polar coordinate systems: the maximal values were obtained for the NIR data sets, both for Gabor and MLDF filters. Interestingly, in all cases the lower parts of the iris were better than the upper parts, which are more frequently occluded by eyelids. Globally, MLDF filters provided more homogeneous values than Gabor filters. For VW data, regions nearby the pupillary boundary are worse than the middle and outer bands, probably due to the difficulty in obtaining reliable estimates of the pupillary boundary in VW images.

Regarding the radial bands in the iris, even though the maximal discriminability was observed for the middle bands, this might not be due to biological properties of the iris texture. Instead, the middle bands are the regions where the largest filters can be applied without surpassing the iris boundaries. As

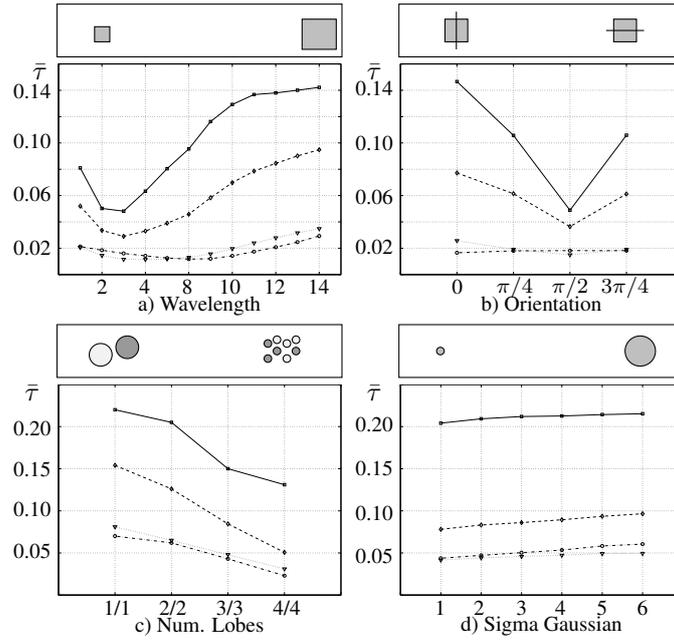


Figure 4.50: Average discriminability $\bar{\tau}$ of the bits in iris codes, regarding filters parameterization. The upper row regards the Gabor kernels (wavelength and orientation parameters) and the bottom row corresponds to the MLDF filters (number of lobes and sigma of the Gaussian kernel).

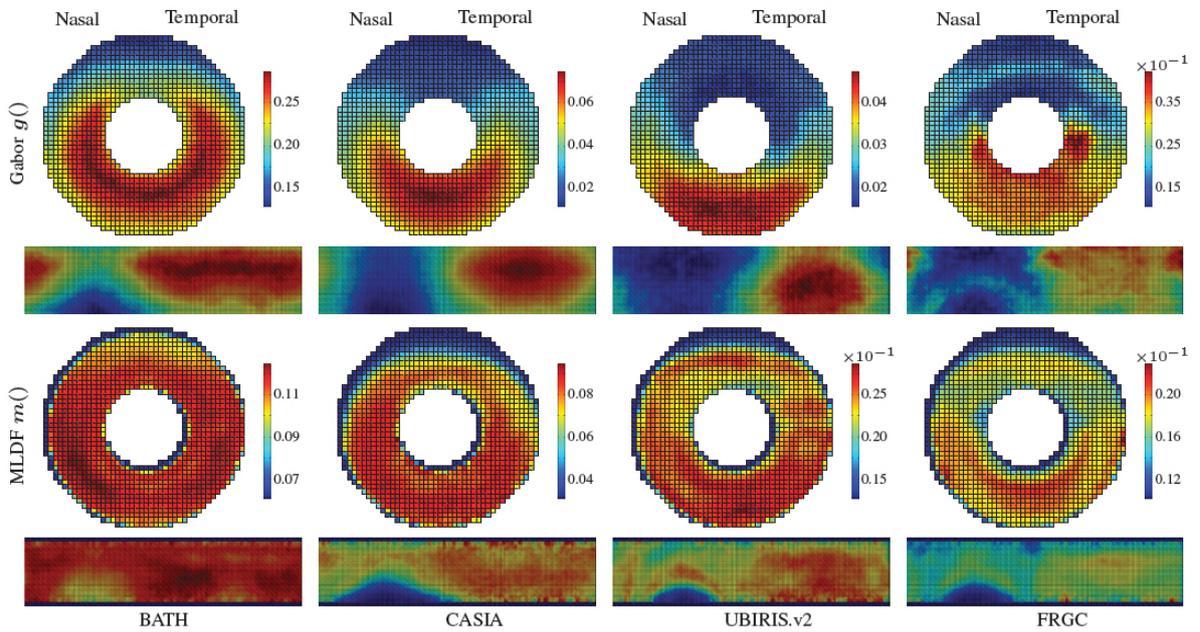


Figure 4.51: Average bit discriminability $\Psi[x, y]$ across the iris. Values are given for the Cartesian and polar coordinate systems, for the BATH, CASIA, UBIRIS and FRGC data sets.

illustrated in Fig. 4.50, large filters tend to produce more discriminant bits, which accords the results given in [68].

It is interesting to note the reduced correlation between the amounts of information in iris patches and the discriminability of each patch. For the BATH data set, the levels of linear correlation between variables $h[x, y]$ and $\Psi[x, y]$ were -0.12/-0.38 (Gabor/MLDF filters), and -0.40/-0.22 for the CASIA-Iris-Distance set. Regarding the VW data, values were 0.16/-0.02 for the UBIRIS.v2 and -0.34/-0.41 for the FRGC datasets. These low correlation values in terms of magnitude and sign (negative in 7/8 of the cases) give space for additional research about iris feature extraction / matching strategies that profit in a better way from the amount of information that is locally available.

4.9.7 Discriminability vs. Fragility

This section illustrates the differences between the previously reported concept of *fragility* and the concept of *discriminability* discussed in this work. In Fig. 4.52 we highlight the bits from one of the datasets used (BATH) where the largest differences in the fragility and discriminability z-scores $z(\cdot)$ were observed (using Gabor filters), i.e., $z(\Psi[x, y]) - z(1\text{-fragility})$. Here, red / orange regions are particularly discriminative but - even though- their bits have a relatively large fragility, whereas blue colours represent bits that are not fragile, but still have not particularly high discriminability. The less interesting cases (bits consistent and non-fragile or bits non-consistent and fragile) have values near to 0. This clearly distinguishes both concepts, i.e., apart fragility, there is a new family of bits (discriminability), which can be used to develop better iris recognition algorithms in the future.

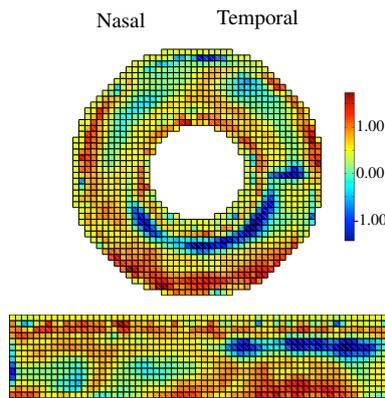


Figure 4.52: Illustration of the differences between the *fragility* and *discriminability* concepts. Regions in red / orange illustrate discriminant bits that - even though - are relatively fragile, whereas blue regions denote the opposite case, i.e., bits that are not particularly fragile but still have not the highest discriminability (results obtained using Gabor $g(\cdot)$ kernels, in the BATH dataset).

4.9.8 Codes Quantization: How Much Discriminating Information Is Lost?

In the most acknowledged iris recognition algorithm, only phase information is used in recognition. Amplitude information is not considered reliable, as it depends of imaging contrast, illumination and camera gain. Accordingly, Hollingsworth *et al.* [68] observed that most inconsistencies in iris codes are due to the coarse quantization of the phase response, and disregarded bits from filter responses near the axes.

Even considering the afore arguments reasonable, we assessed the amounts of discriminating information contained in the filter responses near the axes. With respect to the traditional strategy of keeping

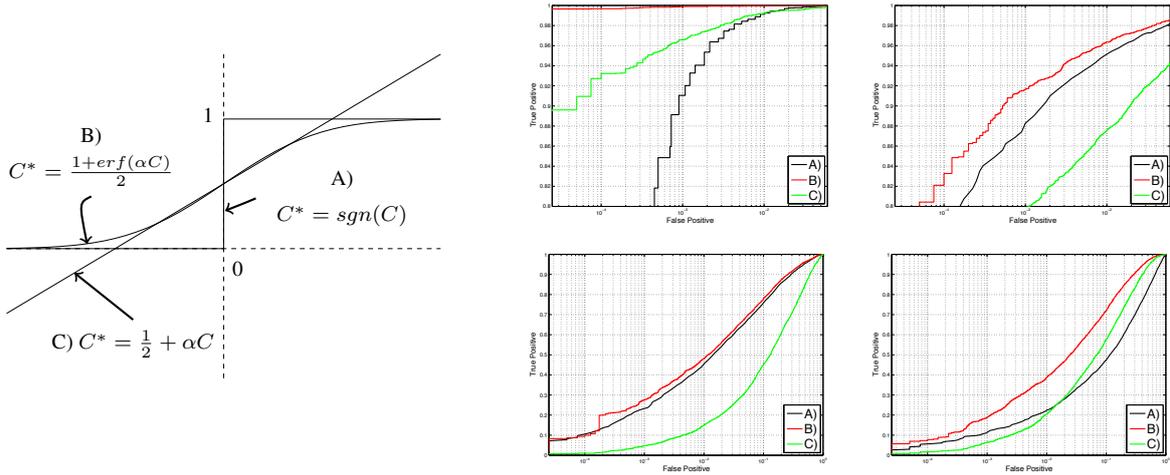


Figure 4.53: At left: Three different strategies for code quantization: A) binary; B) sigmoid function; and C) linear mapping. At right: recognition performance with respect to A), B) and C) code quantization strategies for BATH (upper-left plot), CASIA-Iris-Distance (upper-right), UBIRIS.v2 (bottom-left) and FRGC data sets (bottom-right).

only the sign of coefficients (function A) in Fig. 4.53), two other strategies were considered: a linear mapping of the magnitude of the responses, yielding real-valued coefficients matched by the ℓ_2 norm (function C) in Fig. 4.53); and a trade-off of both strategies, according to a sigmoid-based transform that maps large magnitude values to the 0/1 values, but weights values near the axes to real values in the $[0,1]$ interval. In this case, the ℓ_2 norm was also used as matching function.

The ROC curves given at the right side of Fig. 4.53 compare the recognition performance with respect to each quantization strategy and Table 4.4 summarizes the results, giving the Area Under Curve (AUC) and the decidability index d' that, as suggested by Daugman [35], measures how well separated the genuine / impostor distributions are:

$$d' = \frac{|\mu_G - \mu_I|}{\sqrt{\frac{1}{2}(\sigma_I^2 + \sigma_G^2)}}, \quad (4.87)$$

where $\mu_I = \frac{1}{k} \sum_i d_i^I$ and $\mu_G = \frac{1}{m} \sum_i d_i^G$ are the means of the genuine (G) and impostor (I) scores and $\sigma_I = \sqrt{\frac{1}{k-1} \sum_i (d_i^I - \mu_I)^2}$ and $\sigma_G = \sqrt{\frac{1}{m-1} \sum_i (d_i^G - \mu_G)^2}$ their standard deviations.

Two opposite conclusions were drawn: for Gabor filters, the best results were observed when using the traditional $\text{sign}()$ quantization function. In this case, using scalars instead of sign bits even decreased the recognition performance. Oppositely, for MLDF filters, the best results were observed when using the proposed sigmoid function, i.e., when the coefficients of small magnitude were also considered for the matching process. This points for the conclusion that there is actually reliable discriminating information in the coefficients near the origin. However, these coefficients are less reliable than those with large magnitude, as in no case the linear mapping strategy got results close to any of the remaining strategies.

Note that the above conclusions were drawn based on the reported AUC and d' values, which in the large majority of the cases were observed to be in agreement. The exceptions occurred mostly in cases where the shape of the genuine / impostor distributions were the farthest from Gaussian distributions. For these particular cases, we relied mostly in the AUC value, as it does not require a specific data distribution to report meaningful results.

Table 4.4: Variations in recognition performance with respect to different strategies for code quantization.

Dataset	Features	A) sign()		B) sigmoid()		C) linear (no quantization)	
		d'	AUC	d'	AUC	d'	AUC
BATH	Gabor	8.79 ± 0.01	0.994 ± 0.001	7.08 ± 0.01	0.992 ± 0.001	6.52 ± 0.01	0.990 ± 0.001
BATH	MLDF	9.15 ± 0.01	0.994 ± 0.001	8.82 ± 0.01	0.993 ± 0.001	5.89 ± 0.01	0.988 ± 0.001
CASIA-Iris-Distance	Gabor	3.20 ± 0.01	0.982 ± 0.001	3.16 ± 0.01	0.982 ± 0.001	3.05 ± 0.02	0.971 ± 0.001
CASIA-Iris-Distance	MLDF	3.89 ± 0.01	0.990 ± 0.001	4.12 ± 0.01	0.984 ± 0.001	3.13 ± 0.01	0.982 ± 0.001
UBIRIS.v2	Gabor	1.23 ± 0.01	0.813 ± 0.006	1.16 ± 0.02	0.793 ± 0.007	0.82 ± 0.02	0.720 ± 0.006
UBIRIS.v2	MLDF	1.88 ± 0.01	0.904 ± 0.003	1.96 ± 0.01	0.917 ± 0.003	1.02 ± 0.01	0.766 ± 0.009
FRGC	Gabor	1.12 ± 0.02	0.792 ± 0.006	1.01 ± 0.02	0.770 ± 0.008	0.83 ± 0.01	0.731 ± 0.007
FRGC	MLDF	1.74 ± 0.01	0.892 ± 0.006	1.88 ± 0.02	0.908 ± 0.002	1.47 ± 0.02	0.849 ± 0.007

	<p>Main Challenges in <i>IrisCodes</i> study:</p> <p>Perceive the biological properties in the human eye that lead to non-linear deformations in the iris texture</p> <p>Perceive the biological features in the human iris that cause fragile and non-discriminant bits</p>
---	---

Chapter 5

Ethics / Privacy in Non-Cooperative Biometric Recognition

The origin of the *ethics* term dates back from the Greek civilisation [49], as part of a more general domain: the *philosophy*. The word itself comes from the ancient Greek work *eché*, which refers to a person's character. Hence, the main goal in ethics is to discriminate between what is *right* and *wrong* in human conduct, governed by a set of rules called *codes of ethics*, which depending on the specific domain, can take one or more of the following norms [87]:

Principles. These are the moral bases, and act as guidelines or references for classifying a human action.

Public policies. Define the acceptable behaviour and practices for the interaction between humans in a society.

Codes of conduct. Are based in the principles and public policies to formalize a desirable / non-desirable conduct.

Legal instruments. Enforce good conduct through courts and penalties for any violation.

Undoubtedly, the type of recognition system focused in this seminar raises evident ethical concerns in terms of the citizens privacy and the morality behind recognising someone without asking his permission, and mostly sure, without even his awareness. As illustrated in Fig. 5.1, an appropriate balance between the right to privacy acknowledged by all modern societies and the procedures that such same societies must carry out to assure that citizens are safe and protected from others with harmful intents.

As the term itself suggests, *Biometric* recognition is usually said to analyse our *bios* for recognition purposes. However, maybe the most correct term would be *zoemetrics*. *Zoe* is the Greek term to refer human life according to the major properties that humans share with all living beings, among which are the family of patterns analysed by biometric recognition algorithms. In opposition, *Bios* refers to the human beings from a cultural perspective, related to the properties of the soul, that turn him a unique and non-repeatable identity.

However, in a practical scenario, the *zoe* and *bios* components are interconnected, in the sense that by knowing the identity of a human (even if analysing only its *zoe*), we find an injective correspondence to its *bios*, in the sense that no soul is able to go to a supermarket without the corresponding bios part (body). Hence, from our perspective, the fact of analysing exclusively our *zoe* does not reduce the privacy concerns about biometrics and the key point remains the same: how to establish a threshold that faithfully delimitates what is *reasonable* for a particular *case*.



Figure 5.1: *QUIS-CAMPI* systems should find a trade-off between the citizens' right to privacy and the procedures required to assure security in modern societies.

For instance, several international airports worldwide are starting to install full-body scan devices (Fig. 5.2, image taken from¹), attempting to minimize the probability that someone enters a plane while carrying a weapon or explosive device. Most probably, the common sense **judges this procedure as acceptable** if (even in a single case) the system is able to catch a terrorist, that otherwise will cause the death of hundreds of human beings. Note that there is no doubt that this kind of scan is **evidently intrusive** and demands the complete exposure of a person, just to access a resource that a priori should not imply such exposition.



Figure 5.2: Full-body scan devices are running in several international airports worldwide.

A similar discussion applies to the case of non-cooperative recognition systems. Suppose that a covert surveillance system was able to detect that the 2013 Boston Marathon terrorists left bags of considerable dimension in a public and crowded place, which should obviously had raised a security incident probably handled by legal authorities. Even assuming that such type of system was invading citizens' privacy by confirming their presence in that event, isn't it obvious that such system will be *acceptable*?

There are different types of national and international regulations that attempt to regulate the functioning of biometric recognition systems, and in particular of non-cooperative recognition systems. The *Universal Declaration of Human Rights* was adopted by the United Nations General Assembly on 10 December 1948, after the second world war, and has an article (the 12th), stating that: "*No one shall be*

¹http://core.physicsinfo.co.uk/filestore/2034/Body_scan.jpg

subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks". However, another article (the 3rd) of the same declaration states that "Everyone has the right to life, liberty and security of person". Both articles assure the right to privacy and security, but what happens in cases where both rights are impossible to assure jointly? Also, isn't the right to privacy often confused with the right to anonymity?

The European Parliament issued a directive (95/46/EC) for the protection of persons and data, where the *notice* and *consent* properties appear to bias the overall policy to forbidding non-cooperative systems, while the U.S. government appears not to be so conservative, and issued a regulation for the use of biometric data (8 CFR 103.16), both for American and foreign citizens. In Portugal, the *Comissão Nacional de Proteção de dados* is the legal entity that regulates the setup of biometric recognition systems. Presently, anyone that wants to deploy a biometric system has to fill an electronic form, describing the conditions under which the system will run:

1. Who will process the information?
2. What's the main purpose of the system?
3. What's the biometric trait used?
4. Will the system store the biometric data or the signature?
5. Will it be used a centralized or a local database?
6. Are biometric data supposed to be interchanged with third-party entities?
7. What are the physical and logical security plan, regarding the access to the database?

From our viewpoint, all the existing regulations unavoidably contain subjective terms that make hard to objectively perceive the cases where non-cooperative biometric recognition is acceptable and where such type of systems should be forbidden. The fear of libertarian-biased movements is that the increasingly high concern about safety will push the balance toward the side of "security", instead of "privacy", citing Benjamin Franklin: "They that can give up essential liberty to obtain a little temporary safety deserve neither liberty nor safety". On the other side, one should account that the context in B. Franklin's life was completely different, as (for instance) stated by Neil Young (musician): "Benjamin Franklin said that anyone who gives up essential liberties to preserve freedom is a fool, but maybe he didn't conceive of nuclear war and dirty bombs" [19].

As concluding remarks, we argue that it's to the official entities in each country to consider the levels of threats in each place, and to establish precise limits for the use of non-cooperative (covert) recognition systems. It appears that the common sense privileges in some cases clearly the *security* perspective, and the counterpart *privacy* in many others, while leaving the uncertain cases for the individual (subjective) sense of each person. For instance, is it *reasonable* to automatically recognise human beings nearby a public building of sensitive security? And in a crowded public sports event? What about in a supermarket? And at a gas station?

	<p>Ethics / Privacy Main Challenges:</p> <p>Regulate the particular cases where / when covert biometric recognition is <i>reasonable</i></p> <p>Demythify concerns about what can be exactly inferred from biometric information (e.g., iridology)</p>
---	---

Chapter 6

Conclusions and Further Work

6.1 Concluding Remarks

This document focused the development of non-constrained biometrics systems, toward the development of fully covert recognition systems, able to work effectively in conditions that are presently exclusively associated to visual surveillance systems.

We started by stating objectively the research problem and the idea behind the concept of *QUIS-CAMPI*, illustrating the main phases that compose such type of computer vision / pattern recognition system. Next, apart from enumerating the main achievements of our group in the scope of this research work, a more detailed description of each phase of the recognition chain was given, focusing not only on publications authored by us, but also in the most relevant methods published for each phase.

Note that the extremely ambitious goal behind *QUIS-CAMPI* still remains to be fully achieved, keeping this subject as the main research topic of our group. For the most relevant phases of the recognition chain, we included a set of tables that summarise the corresponding open problems, and also serve as motivation for further research.

6.2 Further Work

As concluding remark, we feel appropriate to illustrate the state-of-the-art results in terms of non-cooperative iris recognition. As it is known that performance suffer significant variations according to the database size, we estimated the probability that a query on a sample S_i returns a cumulative rank k value, as a function of the database size. Then, we plotted the values observed for the state-of-the-art classification ensembles [131] against this theoretical performance line, in order to infer the performance for large scale identification scenarios. We assume that X and Y are independent random variables that represent the dissimilarity scores generated by a biometric classifier for *match* and *non-match* comparisons. Even though iris match distributions are usually asymmetric and that non-match distributions are often skewed with longer left tails, for evaluation purposes an oversimplification was made and both distributions assumed as normal, which is more plausible in less favourable environment conditions. Let $X \sim \mathcal{N}(\mu_x, \sigma_x)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y)$ denote such *match* and *non-match* distributions. According to the elementary theory of rank tests [59], the k^{th} order statistic $Y_{(k)}$ of a statistical sample Y_1, \dots, Y_n is equal to its k^{th} smallest value. Let $Y_{(1)}, \dots, Y_{(n)}$ be the order statistics of a set of independent observations Y_1, \dots, Y_n , that is, $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$. The distribution function of $Y_{(k)}$ is equal to

$$\begin{aligned}
F_{(k)}(y) &= P(Y_{(k)} \leq y) \\
&= \sum_{i=k}^n \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i}.
\end{aligned} \tag{6.1}$$

In practical terms, (6.1) can be very hard to calculate for very large n values and $i \approx \frac{n}{2}$. Thus, using the concepts of the complement of an event and of the probability of non-occurrence, (6.1) is equivalent to

$$F_{(k)}(y) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i}. \tag{6.2}$$

Using (6.2) and (6.1) it is possible to obtain the density of the k^{th} smallest value of the *non-match* comparisons, for a given population of size n . Considering that $f(y)$ is the density of $F(y)$, the density of $Y_{(k)}$ is given by

$$\begin{aligned}
f_{(k)}(y) &= n \binom{n-1}{k-1} [F(y)]^{k-1} [1 - F(y)]^{n-k} f(y) \\
&= \frac{n!}{(k-1)!(n-k)!} [F(y)]^{k-1} [1 - F(y)]^{n-k} f(y).
\end{aligned} \tag{6.3}$$

The probability that a *match* distance observation X returns a cumulative rank k is given by

$$\begin{aligned}
P(\text{rank}(X) \leq k) &= P(X \leq Y_k) \\
&= \int_{-\infty}^{\infty} f_{(k)}(y) \int_{-\infty}^y f(x) dx dy.
\end{aligned} \tag{6.4}$$

Accordingly, the probability that the *match* observation has exact rank k is given by

$$P(\text{rank}(X)=k) = P(X \leq Y_{(k)}) - P(X \leq Y_{(k-1)}). \tag{6.5}$$

The continuous line of Fig. 6.1 estimates the probability of a cumulative rank 10 ($P(\text{rank}(X) \leq 10)$) as a function of the gallery size n , and the dashed line is analogous for a cumulative rank 30. We used $X \sim \mathcal{N}(0.32, 0.1)$ and $Y \sim \mathcal{N}(0.5, 0.06)$, values obtained by Gaussian curve fitting to the best classification ensemble composed by the fusion (*) of the four best algorithms. The circular data points are the cumulative rank 10 values observed empirically, according to the different databases sizes we were able to test. For contextualization purposes, the four vertical solid lines illustrate identification scenarios at different scales: global (World population), continental (Europe), country (Portugal) and small city (Covilh,, a Portuguese city with approximately 50 000 inhabitants). From this analysis, it appears that the observed values adequately fit the probability line, which is a good indicator of the potential performance of this type of recognition system on larger scales. It should be concluded that this type of recognition technology is not yet sufficiently mature to be deployed in large scale identification scenarios, and further advances in the technology are needed to meet the full range of operational requirements at those operating scales. Currently, approximately 2% of the queries would return the correct identity in top-10, if the entire world population is enrolled in the system. This value rises to approximately 4% and 12% if the universe is reduced to the Continental or National population.

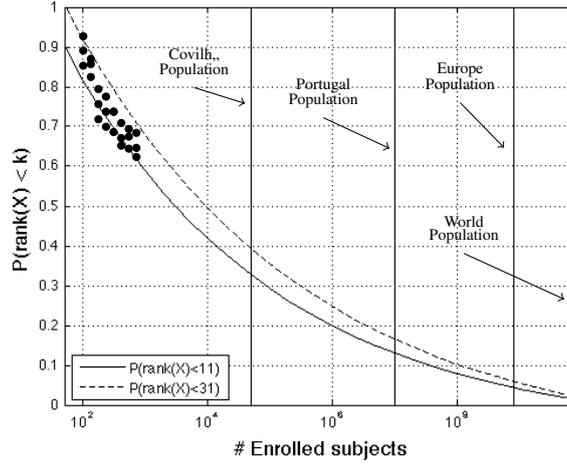


Figure 6.1: Estimated probability that an identification query returns the true identity out of the first 10 (continuous line) and 30 positions (dashed line), as a function of the number of enrolled identities, i.e., $P(\text{rank}(X) \leq 10)$ and $P(\text{rank}(X) \leq 30)$. The circular data points are the cumulative rank 10 values observed empirically, according to the different databases sizes.

However, as the type of systems discussed in this document work in a covert way, absolutely no human effort is demanded of the subjects during the recognition processes. This raises the possibility of using multiple recognition systems spaced across, for instance, in a city street. Simplifying the problem, we regard all subjects of a population $\mathbb{P} = \{s_1, \dots, s_n\}$ as *sheep*, i.e., assume that they follow the system averages: match relatively well against themselves and poorly against others. Let us consider k iris recognition systems with roughly similar performance, with a sensitivity of α at a false match rate of β . Here we introduce the concept of *exogenous independence*, hypothesizing that purposely changing the lighting conditions in the environment (by using different levels of light or types of illuminants) and the acquisition protocols (poses, distances) should potentiate the independence between the system outputs. Assuming that the independence of each system provides an upper bound on the performance that would be attainable by the fusion of multiple systems, the binomial distribution can be used to obtain the probability that a subject s_i is screened by k recognition systems and correctly recognized by k' of these, $1 \leq k' \leq k$:

$$P(R_{k'}) = \frac{k!}{k'! (k - k')!} \alpha^{k'} (1 - \alpha)^{k - k'}. \quad (6.6)$$

For different values of k' , the probability that a reported match is false is given by $\beta^{k'}$, assuming that false matches in each of the k recognition systems are independent events. Accordingly, a *match* will be reported *iff* a minimum of k' recognition systems output a match:

$$\begin{aligned} P(R_{\geq k'}) &= \sum_{j=k'}^k P(R_j) \\ &= \sum_{j=k'}^k \frac{k!}{j! (k - j)!} \alpha^j (1 - \alpha)^{k - j}, \end{aligned} \quad (6.7)$$

provided that all events are mutually exclusive. Considering the average state-of-the-art sensitivity value, Fig. 6.2 relates the expected sensitivity of such a *multipoint* biometric system to the number of

recognition systems used, considering different false match rates. From its analysis, one can conclude that approximately five independent recognition systems would be enough to attain almost full sensitivity at a false acceptance rate β of 0.01. This value substantially increases when a lower number of false alarms is convenient (large scale applications), requiring between thirteen and twenty three independent recognition systems to operate, respectively, at FAR $1e^{-4}$ and $1e^{-6}$.

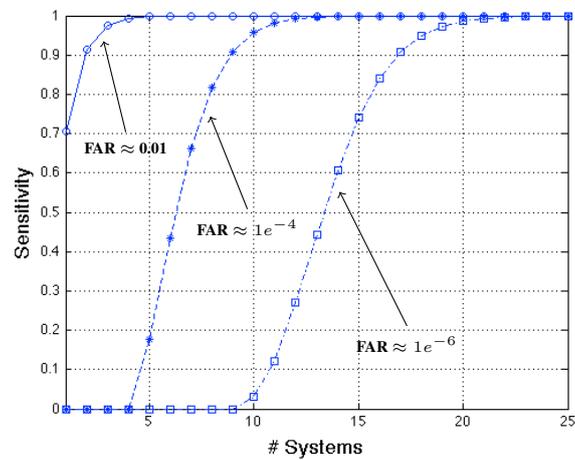


Figure 6.2: Expected sensitivity of a set of recognition systems placed consecutively and operating covertly under intentionally varying lighting conditions, with different required values for the false acceptance rates.

Bibliography

- [1] A. Agarwal and B. Triggs. 3D human pose from silhouettes by relevance vector regression. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 2, pag. 882–888, 2004.
- [2] A. Agarwal and B. Triggs. Recovering 3D human pose from monocular image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pag. 44–58, 2006.
- [3] M. Agoston. *Computer Graphics and Geometric Modeling: Implementation and Algorithms*. Springer, pag. 300-306, ISBN: 1-85233-818-0, 2005.
- [4] A. Albiol, L. Torres and E. Delp. Optimum color spaces for skin detection. In *Proceedings of the International Conference on Image Processing*, vol. 1, pag. 122–124, 2001.
- [5] A. Andriyenko and K. Schindler. Globally Optimal Multi-target Tracking on a Hexagonal Lattice. In *Proceedings of the 11th European Conference on Computer Vision (ECCV'10)*, vol. 1, pag. 466–479, 2010.
- [6] A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, pag. 1265–1272, 2011.
- [7] E. Arvacheh and H. Tizhoosh. *A study on Segmentation and Normalization for Iris Recognition*. Msc. dissertation, University of Waterloo, 2006.
- [8] B. Babenko, M-H. Yang and S. Belongie. Robust Object Tracking with Online Multiple Instance Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pag. 1619–1632, 2011.
- [9] David Barret (The Telegraph). One surveillance camera for every 11 people in Britain, says CCTV survey. <http://www.telegraph.co.uk/technology/10172298/>, July 10, 2006 (retrieved June 17, 2014).
- [10] C. Basca, M. Talos and R. Brad. Randomized Hough Transform for Ellipse Detection with Result Clustering. in *Proceedings of the The International Conference on Computer as a Tool (EUROCON'05)*, vol. 2, pag. 1397–1400, 2005.
- [11] A. Basit and M. Y. Javed. Iris localization via intensity gradient and recognition through bit planes. In *Proceedings of the International Conference on Machine Vision (ICMV'07)*, pag. 23–28, December 2007.
- [12] S. Belongie, J. Malik and J. Puzicha. Shape matching and Object Recognition Using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 24, pag. 509–522, 2002.

- [13] A. Bertillon. Identification Anthropométrique; instructions signalétiques. Leun Impr. Administrative, 1893.
- [14] A. Besbes, N. Komodakis, G. Lings and N. Paragios. Shape Priors and Discrete MRFs for Knowledge-based Segmentation. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, pag. 1295–1302, 2009.
- [15] J. C. Bezdek, J. M. Keller, R. Krishnapuram and N. R. Pal Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. *IEEE Transactions on Information Forensics and Security*, vol. 6, issue 1, pag. 82–95, 2011.
- [16] S. Bharadwaj, H. Bhatt, M. Vatsa and R. Singh. Periocular biometrics: When iris recognition fails. In Proceedings of the *4th IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS'10)*, pag. 1–6, 2010.
- [17] R.M. Bolle, S. Pankanti, J.H. Connell and N. Ratha. Iris Individuality: A Partial Iris Model. In Proceedings of the *17th International Conference on Pattern Recognition (ICPR'04)*, vol. 2, pag. 927–930, 2004.
- [18] N. Boulgouris, D. Hatzinakos and K. Plataniotis. Gait Recognition: A challenging signal processing technology for biometric identification. *IEEE Signal processing Magazine*, vol. 22, issue 6, pag. 78–90, 2005.
- [19] K.W. Bowyer. Face Recognition Technology and the Security Versus Privacy Tradeoff. *IEEE technology and Society*, pag. 9–20, 2004.
- [20] K.W. Bowyer, S.E. Baker, A. Hentz, K.P. Hollingsworth, T. Peters and P.J. Flynn. Factors that degrade the match distribution in iris biometrics. *Identity in the Information Society*, vol. 2, Issue 3, pag. 327–343, 2009.
- [21] M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. Van Gool. Online Multiperson Tracking-by-Detection from a Single, Uncalibrated Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pag. 1820–1833, 2011.
- [22] D. Butler, V. Bove and S. Sridharan. Real-Time Adaptive Foreground/Background Segmentation. *EURASIP Journal on Advances in Signal Processing*, vol. 14, 841926, doi: 10.1155/ASP.2005.2292, 2005.
- [23] Q. Cai and J. K. Aggarwal. Tracking human motion in structured environments using a distributed-camera system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pag. 1241–1247, 1999.
- [24] D. Cao, C. Chen, D. Adjeroh and A. Ross. Predicting gender and weight from human metrology using a copula model. In Proceedings of the *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS'12)*, pag. 162–169, August, 2012.
- [25] B. Caputo, S. Bouattour and H. Niemann. Robust appearance-based object recognition using a fully connected Markov random field. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'02)*, vol., 3, pag. 565–568, 2002.
- [26] C. Chen and J. Odobez. We are not contortionists: Coupled adaptive learning for head and body orientation estimation in surveillance video. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)*, pag. 1544–1551, 2012.

- [27] L. Cieplinski. MPEG-7 Color Descriptors and Their Applications. In *Lecture Notes in Computer Science*, vol. 2124, pag. 11–20, 2001.
- [28] R.T. Collins, R. Gross and J. Shi, Silhouette-based human identification from body shape and gait. In Proceedings of the *IEEE Conf. Automatic Face and Gesture Recognition*, pag. 351–356, 2002.
- [29] D. Comaniciu, V. Ramesh and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pag. 564–577, 2003.
- [30] S. Crihalmeanu and A. Ross. Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters*, vol. 33, no. 14, pag. 1860–1869, 2012.
- [31] DARPA. Human identification at-a-distance (Human-ID). <http://www.cc.gatech.edu/cpl/projects/hid/>, (retrieved October 5, 2014).
- [32] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pag. 886–893, 2005.
- [33] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America, Part A*, vol. 2, no. 7, pag. 1160–1169, 1985.
- [34] 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, pag. 1148–1161, 1993.
- [35] J. Daugman. Biometric decision landscapes. *University of Cambridge Technical Report*, UCAM-CL-TR-482, ISSN: 1476-2986, 2000.
- [36] J. G. Daugman. How Iris Recognition Works. *IEEE Transactions on Circuits, Systems and Video Technology*, vol. 14, issue 1, pag. 21–30, 2004.
- [37] J. Daugman. Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, vol. 94, no. 11, pag. 1927–1935, 2006.
- [38] J. G. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 37, no. 5, pag. 1167–1175, 2007.
- [39] M. Dillencourt, H. Samet and M. Tamminen. A general approach to connected-component labelling for arbitrary image representations. Springer London Heidelberg New York Dordrecht, 1999.
- [40] M. Dobes, J. Martineka, D. S. Z. Dobes, and J. Pospisil. Human eye localization using the modified hough transform. *Optik*, vol. 117, pag. 468–473, 2006.
- [41] D. Donoho and I. Johnstone. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, vol. 81, pag. 425–455, 1994.
- [42] V. Dorairaj, N. Schmid and G. Fahmy. Performance evaluation of nonideal iris based recognition system implementing global ICA encoding. In Proceedings of the *International Conference on Image Processing*, vol. 3, pag. 285–288, 2005.

- [43] G. Dozier, K. Frederiksen, R. Meeks, M. Savvides, K. Bryant, D. Hopes and T. Munemoto. Minimizing the Number of Bits Needed for Iris Recognition via Bit Inconsistency and GRIT. *Proceedings of the IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications (CIB'09)*, pag. 30-37, 2009.
- [44] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pag. 2270–2283, 2013.
- [45] Y. Du, E. Arslanturk, Z. Zhou, and C. Belcher. Video-based Non-cooperative Iris Image Segmentation. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 41, no. 1, pag. 64–74, 2011.
- [46] B. Duca, E. Bigiin, J. Bigiin, G. Maitre and G. Fischer. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, vol. 18, no. 9, pag. 835–843, 1997.
- [47] R. Duda, P. Hart and D. Stork. *Pattern Classification*. Wiley-Interscience Publication, ISBN: 978-0-471-05669-0, 2001.
- [48] A. Elgammal, D. Harwood and L. Davis. Non-parametric Model for Background Subtraction. In *Proceedings of the European Conference on Computer Vision (ECCV'00)*, LNCS, vol. 1843, pag. 751–767, 2000.
- [49] A. Fagothey. *Right and reason*. Tan Books and Publishers (2nd edition), Rockford, 1959.
- [50] P. Felzenszwalb and D. Huttenlocher. Efficient Belief Propagation for Early Vision. *International Journal of Computer Vision*, vol. 70, no. 1, pag. 41–54, 2006.
- [51] M.A. Fischler and R.A. Elschlager. The Representation and Matching of Pictorial Structures. *IEEE Transactions on Computers*, C-22, pag. 67–92, 1973.
- [52] B. Gloyer, H.K. Aghajan, K-Y. Siu and T. Kailath. Video-based freeway-monitoring system using recursive vehicle tracking. *Proceedings of the SPIE*, vol. 2421, pag. 173–180, 1995.
- [53] B. Glocker, D. Zikic, N. Komodakis, N. Paragios and N. Navab. Linear Image Registration Through MRF Optimization. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pag. 422–425, 2009.
- [54] M. Goffredo, I. Bouchrika, J. Carter and M. Nixon. Performance analysis for automated gait extraction and recognition in multi-camera surveillance. *Multimedia Tools and Applications*, vol. 50, no. 1, pag. 75–94, 2010.
- [55] H. Grabner, M. Grabner and H. Bischof. Real-Time Tracking via On-line Boosting. In *Proceedings of the British Machine Vision Conference (BMV'06)*, vol. 6, pag. 1–10, 2006.
- [56] G. Granlund. Fourier Preprocessing for Hand Print Character Recognition. *IEEE Transactions on Computers*, vol. 21, pag. 195–201, 1972.
- [57] K. Grauman, G. Shakhnarovich and T. Darrell. Inferring 3D structure with a statistical image-based shape model. In *Proceedings of the Ninth International Conference on Computer Vision (ICCV'03)*, vol. 21, pag. 641–647, 2003.
- [58] J. Gu, X. Ding, S. Wang and Y. Wu. Action and Gait Recognition From Recovered 3-D Human Joints. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 4, pag. 1021–1033, 2010.

- [59] J. Hajek and Z. Sidak. Theory of rank tests. Academic Press; 1st edition, 2000.
- [60] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pag. 316–322, 2006.
- [61] F. Hao, J. Daugman and P. Zielinski. A Fast Search Algorithm for a Large Fuzzy Database. *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pag. 203–211, 2008.
- [62] I. Haritaoglu, D. Harwood and L.S. Davis. W⁴: real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pag. 809–830, 2000.
- [63] K. Haskell and R. Hanson. An algorithm for linear least squares problems with equality and non-negativity constraints. *Math Program*, vol. 21, pag. 98–118, 1981.
- [64] M. Hazewinkel (ed.). Kolmogorov-Smirnov test. *Encyclopaedia of Mathematics*, Springer, 2001.
- [65] Z. He, T. Tan, Z. Sun and X. Qiu. Towards Accurate and Fast Iris Segmentation for Iris Biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pag. 1670–1684, 2009.
- [66] G. Hoffmann. CIELab Color Space. <http://docs-hoffmann.de/cielab03022003.pdf>, assessed on June 2013.
- [67] K.P. Hollingsworth, K.W. Bowyer and P.J. Flynn. All Iris Code Bits are Not Created Equal. in Proceedings of the First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS'07), pag. 1–6, 2007.
- [68] K.P. Hollingsworth, K.W. Bowyer and P.J. Flynn. The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pag. 964–973, 2009.
- [69] K.P. Hollingsworth, K.W. Bowyer and P.J. Flynn. Improved Iris Recognition Through Fusion of Hamming Distance and Fragile Bit Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pag. 2465–2476, 2011.
- [70] A. Hossain, Y. Makihara, J. Wang and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, vol. 43, no. 6, pag. 2281–2291, 2010.
- [71] N. Howe. Silhouette lookup for monocular 3D pose tracking. *Image and Vision Computing*, vol. 25, no. 3, pag. 331–341, 2007.
- [72] D. Huang, M. Storer, F. De la Torre and H. Bischof. Supervised local subspace learning for continuous head pose estimation. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (CVPR'11), pag. 2921–2928, 2011.
- [73] M. Husain, E. Saber, V. Mistic and S.P. Joralemon. Dynamic Object Tracking by Partial Shape Matching for Video Surveillance Applications. In Proceedings of the *International Conference on Image Processing*, (ICIP'06), pag. 2405–2408, 2006.
- [74] H. Iwama, D. Muramatsu, Y. Makihara and Y. Yagi. Gait-based person-verification system for forensics. In Proceedings of the *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems* (BTAS'12), pag. 113–120, 2012.

- [75] R. Jain and H-H. Nagel. On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pag. 206–214, 1979.
- [76] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong and A. Ross. Biometrics: A grand challenge In Proceedings of the 17th *International Conference on Pattern Recognition (ICPR'04)*, vol. 2, pag. 935–942, August, 2004.
- [77] O. Javed, Z. Rasheed, K. Shafique and M. Shah. Tracking across multiple cameras with disjoint views. in *Proceedings of the Proceedings of the IEEE International Conference on Computer Vision (ICCV'03)*, vol. 2, pag. 952–957, October, 2003.
- [78] Hao Jiang and D.R. Martin. Global pose estimation using non-tree models. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR'08)*, pag. 1–8, 2008.
- [79] R. Jonker and A. Volgenant. A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems *Computing*, vol. 38, pag. 325–340, 1987.
- [80] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-Learning-Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pag. 1409–1422, 2012.
- [81] N. Kalka, J. Zuo, N. Schmid and B. Cukic. Estimating and Fusing Quality Factors for Iris Biometric. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 40, no. 3, pag. 509–524, 2010.
- [82] Y. Kameda and M. Minoh. A human motion estimation method using 3-successive video frames. In Proceedings of the *International Conference on Virtual Systems*, 1996.
- [83] B. Kamgar-Parsi, W. Lawson and B. Kamgar-Parsi. Toward development of a face recognition system for watchlist surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pag. 1925–1937, 2011.
- [84] B. J. Kang and K. R. Park. A study on iris image restoration. In Proceedings of the *International Conference on Audio- and Video-Based Biometric Person Authentication*, pag. 31–40, 2005.
- [85] Z. Kato and T.C. Pong. A Markov random field image segmentation model for textured images. *Image and Vision Computing*, vol. 24, pag. 1103–1114, 2006.
- [86] K. Kim, T. Chalidabhongse, D. Harwood and L. Davis. Real-time foreground background segmentation using codebook model. *Real-Time Imaging*, vol. 11, no. 3, pag. 172–185, 2005.
- [87] J. Kizza. *Ethical and Social Issues in the Information Age*. Text in Computer Science, Springer London Heidelberg New York Dordrecht, 2013.
- [88] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pag. 1568–1583, 2006.
- [89] A. Kong, D. Zhang and M. Kamel. An Analysis of IrisCode . *IEEE Transactions on Image Processing*, vol. 19, no. 2, pag. 522–532, 2010.
- [90] A. Kong. IrisCode Decompression Based on the Dependence between Its Bit Pairs. *IEEE Transactions on Image Processing*, vol. 34, no. 3, pag. 506–520, 2012.
- [91] A. Kong. Modeling IrisCode and Its Variants as Convex Polyhedral Cones and Its Security Implications. *IEEE Transactions on Image Processing*, vol. 22, no. 3, pag. 1148–1160, 2013.

- [92] E.H. Land. The retinex. *American Scientist*, no. 52, pag. 247–264, 1964.
- [93] L. Lee, W. Grimson. Gait analysis for recognition and classification. In Proceedings of the *Fifth IEEE International Conference on Face and Gesture Recognition*, pag. 148–155, 2002.
- [94] L. Lee, R. Romano and G. Stein. Monitoring activities from multiple video streams: establishing a common coordinate frame. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pag. 758–767, 2000.
- [95] P. Li, X. Liu, L. Xiao and Q. Song. Robust and accurate iris segmentation in very noisy iris images. *Image and Vision Computing*, vol. 28, issue 2, pag. 246–253, 2010.
- [96] S-F. Lin, J-Y. Chen and H-X. Chao. Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 31, no. 6, pag. 645–654, 2001.
- [97] Z. Liu, H. Shen, G. Feng and D. Hu. Tracking objects using shape context matching. *INeurocomputing*, vol. 83, pag. 47–55, 2012.
- [98] X. Liu, K. W. Bowyer, and P. J. Flynn. Experiments with an improved iris segmentation algorithm. In Proceedings of the *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pag. 118–123, October 2005.
- [99] C. V. Loan. On the method of weighting for equally constrained least squares problems. *SIAM Journal of Numerical Analysis*, vol. 22, no. 5, pag. 851–864, 1985.
- [100] R. Luque, E. Domnguez, E. Palomo and J. Muoz. A Neural Network Approach for Video Object Segmentation in Traffic Surveillance. In Proceedings of the *International Conference on Image Analysis and recognition (ICIAR'08)*, LNCS, vol. 5112, pag. 151–158, 2008.
- [101] F. Lv and R. Nevatia. Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2007)*, pag. 1–8, 2007.
- [102] L. Maddalena and A. Petrosino. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE Transactions on Image Processing*, vol. 17, no. 7, pag. 1168–1177, 2008.
- [103] S. Mallat. A Wavelet Tour of Signal Processing. *Academic Press*, ISBN: 0-12-466606-X, 1999.
- [104] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal processing*, vol. 41, no. 12, pag. 3397–3415, 1993.
- [105] J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D.J. Lofacono, S. Mangru, M. Tinker, T.M. Zappia, and W.-Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. in *Proceedings of the IEEE*, vol. 94, no. 11, pag. 1936–1947, 2006.
- [106] N. McFarlane and C. Schofield. Segmentation and tracking of piglets in images *Machine Vision and Applications*, vol. 8, no. 3, pag. 187–193, 1995.
- [107] D. Moctezuma, C. Conde, IM. de Diego and E. Cabello. Person detection in surveillance environment with HoGG: Gabor filters and Histogram of Oriented Gradients. In Proceedings of the *IEEE International Conference on Computer Vision Workshops (CVPRW'11)*, no. 11, pag. 1793–1800, 2011.

- [108] H. Mehrotra, M. Vatsa, R. Singh and B. Majhi. Does Iris Change Over Time? *PLoS ONE*, vol. 8, no. 11, e78333, 2013.
- [109] K. Mikolajczyk, C. Schmid and A. Zisserman. Human Detection Based on a Probabilistic Assembly of Robust Part Detectors. In Proceedings of the *European Conference on Computer Vision (ECCV 2004)*, LNCS, vol. 3021, pag. 69–82, 2004.
- [110] K. Niinuma, U. Park and A. K. Jain. Soft Biometric Traits For Continuous User Authentication. *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pag. 771–780, 2010.
- [111] T. Ojala, M. Pietikainen and D. Harwood. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition*, vol. 29, pag. 51–59, 1996.
- [112] K. Oh, B-S. Oh, K-A. Toh, W-Y. Yau and H-L. Eng. Combining sclera and periocular features for multi-modal identity verification. *Neurocomputing*, vol. 128, pag. 185–198, 2014.
- [113] T. Ojala, M. Pietikainen and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, vol. 29, no. 1, pag. 51–59, 1996.
- [114] U. Park, A. Ross and A.K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In Proceedings of the *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS'09)*, pag. 1–6, 2009.
- [115] U. Park, R. Jillela, A. Ross and A.K. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pag. 96–106, 2011.
- [116] O. Pele and M. Werman. The Quadratic-Chi Histogram Distance Family. In Proceedings of the *European Conference on Computer Vision*, vol. 2, pag. 749–762, 2010.
- [117] H. Peng, F. Long and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pag. 1226–1238, 2005.
- [118] W. Pennebaker and J. Mitchell. JPEG Still Image Data Compression Standard. Van Nostrand Reinhold, New York, USA, 1993.
- [119] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pag. 947–954, 2005.
- [120] P. Pudil, J. Ferri, J. Novovičová. Floating search methods for feature selection with nonmonotonic criterion function. In Proceedings of the *12th International Conference on Pattern Recognition (ICPR'94)*, vol. 2, pag. 279–283, 1994.
- [121] R. Poppe. Evaluating Example-based Pose Estimation: Experiments on the HumanEva Sets. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, pag. 1–8, 2007.
- [122] U. Park, A. Ross and A. Jain. Matching highly non-ideal ocular images: an information fusion approach. In Proceedings of the *IEEE 5th International Conference on Biometrics (ICB'12)*, pag. 446–453, 2012.
- [123] J.K. Pillai, V.M. Patel, R. Chellappa and N.K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pag. 1877–1893, 2011.

- [124] N. Poh and S. Bengio. How Do Correlation and Variance of Base-Experts Affect Fusion in Biometric Authentication Tasks. *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pag. 4384–4396, 2005.
- [125] Priyadarshi, A. Nandy, P. Chakraborty and G. Nandi. Speed Invariant, Human Gait Based Recognition System for Video Surveillance Security. *Communications in Computer and Information Science*, vol. 276, pag. 325–335, 2013.
- [126] H. Proença and L.A. Alexandre. UBIRIS: a noisy iris image database. In Proceedings of the *International Conference on Image Analysis and Processing (ICIAP'05)*, vol. 1, pag. 970–977, 2005.
- [127] H. Proença and L.A. Alexandre. Iris Recognition: Measuring Feature's Quality for the Feature Selection in Unconstrained Image Capture Environments. Proceedings of the *2006 International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS'06)*, pag. 35-40, 2006.
- [128] H. Proença, S. Filipe, R. Santos, J. Oliveira and L.A. Alexandre. The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pag. 1529–1535, 2010.
- [129] H. Proença. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pag. 1502–1516, 2004.
- [130] H. Proença. Quality Assessment of Degraded Iris Images Acquired in the Visible Wavelength. *IEEE Transactions on Information Forensics and Security*, vol. 6, issue 1, pag. 82–95, 2011.
- [131] H. Proença and Luís A. Alexandre. Toward Covert Iris Biometric Recognition: Experimental Results From the NICE Contests. *IEEE Transactions on Information Forensics and Security*, vol. 7, issue 2, pag. 798–808, 2011.
- [132] H. Proença. Iris Biometrics: Indexing and Retrieving Heavily Degraded Data. *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pag. 1975–1985, 2013.
- [133] E. Provenzi, L. Carli, and A. Rizzi. Mathematical definition and analysis of the Retinex algorithm. *Journal of the Optical Society of America, Part A*, vol. 22, no. 12, pag. 2613–2621, 2005.
- [134] C. Rathgeb, A. Uhl and P. Wild. On Combining Selective Best Bits of Iris-Codes. In Proceedings of the *Biometrics and ID Management Workshop (BioID'11)*, LNCS 6583, pag. 227–237, 2011.
- [135] D. Reid, S. Samangoei, C. Chen, M. Nixon and A. Ross. Soft Biometrics for Surveillance: An Overview. *Handbook of Statistics*, vol. 31, Elsevier Publishing, 2013.
- [136] D. Reid, M. Nixon and V. Stevenage. Soft Biometrics; Human Identification Using Comparative Descriptors *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pag. 1216–1228, 2014.
- [137] R. Rosales, M. Siddiqui, Jonathan Alon, and S. Sclaroff. Estimating 3D body pose using uncalibrated cameras. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR'01)*, vol. 1, pag. 821–827, 2001.
- [138] A. Ross and A. Jain. Multimodal biometrics: An overview. In Proceedings of the *12th European Conference on Signal Processing*, pag. 1221–1224, 2004.

- [139] A. Ross and S. Shah. Segmenting non-ideal irises using geodesic active contours. In Proceedings of the *IEEE 2006 Biometric Symposium*, pag. 1–6, U.S.A, 2006.
- [140] E. Saber, Y. Xu and A. Tekalp. Partial shape recognition by sub-matrix matching for partial matching guided image labelling. *Pattern Recognition*, vol. 6312, no. 10, pag. 1560–1573, 2005.
- [141] K. vd Sande, T. Gevers and C. Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pag. 1582–1596, 2010.
- [142] B. Sapp, A. Toshev and B. Taskar. Cascaded Models for Articulated Pose Estimation. In Proceedings of the *European Conference on Computer Vision (ECCV'10)*, LNCS, vol. 6312, pag. 406–420, 2010.
- [143] W.R. Schwartz, A. Kembhavi, D. Harwood and L.S. Davis. Human detection using partial least squares analysis. in Proceedings of the *12th IEEE International Conference on Computer Vision (ICCV 2009)*, pag. 24–31, 2009.
- [144] M. Shah, O. Javed and K. Shafique. Automated visual surveillance in realistic scenarios. *IEEE Multimedia*, vol. 14, no. 1, pag. 30–39, August, 2007.
- [145] J. Shi and C. Tomasi. Good features to track. in Proceedings of the *12th IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pag. 593–600, 1994.
- [146] W. Shuigen, C. Zhen and D. Hua. Motion Detection Based on Temporal Difference Method and Optical Flow field. In Proceedings of the *Second International Symposium on Electronic Commerce and Security (ISECS'09)*, vol. 2, pag. 8–88, 2009.
- [147] L. Sloman, M. Berridge, S. Homatidis, D. Hunter and T. Duck. Gait patterns of depressed patients and normal subjects. *American Journal of Psychiatry*, vol. 139, no. 1, pag. 94–97, 1982.
- [148] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, vol. 2, pag. 246–252, 1999.
- [149] B. Subburaman, A. Descamps and C. Carincotte. Counting People in the Crowd Using a Generic Head Detector. in Proceedings of the *IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS'12)*, pag. 470–475, 2012.
- [150] Z. Sun and T. Tan. Ordinal Measures for Iris Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pag. 2211–2226, 2009.
- [151] T. Tan, Z. He and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, vol. 28, no. 2, pag. 223–230, 2010.
- [152] C-W. Tan and A. Kumar. Towards Online Iris and Periocular Recognition Under Relaxed Imaging Constraints. *IEEE Transactions on Image Processing*, vol. 22, no. 10, pag. 3751–3765, 2013.
- [153] T. Tan, X. Zhang, Z. Sun and H. Zhang. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters*, vol. 33, no. 8, pag. 970–977, 2012.
- [154] M. Tico and P. Kuosmanen. Weighted least squares method for the approximation of directional derivatives. In Proceedings of the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, vol. 3, pag. 1681–1684, U.S.A, 2001.

- [155] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. Nixon. Soft Biometrics and Their Application in Person Recognition at a Distance. *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pag. 464–475, 2014.
- [156] G. Toussaint. Solving geometric problems with the rotating calipers . In *proceedings of the IEEE MELECON'83*, Athens, Greece, May 1983.
- [157] P. Trezona. Derivation of the 1964 CIE 10 XYZ Colour-Matching Functions and Their Applicability in Photometry. *Color Research and Application*, vol. 26, no. 1, pag. 67–75, 2001.
- [158] M. Vatsa, R. Singh and A. Noore. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Transactions on Systems, Mans and Cybernetics - B*, vol. 38, no. 4, pag. 1021–1035, 2008.
- [159] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, vol. 57, no.2, pages 137–154, 2002.
- [160] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. in *Proceedings of the ComputerVision and Pattern Recognition Conference (CVPR'01)*, vol. 1, pag. I-511–I-518, 2001.
- [161] P. Viola, M. Jones and D. Snow. Detecting pedestrians using patterns of motion and appearance. in *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, vol. 2, pag. 734–741, 2003.
- [162] P. Viola, J.C. Platt and C. Zhang. Multiple Instance Boosting for Object Detection. in *Advances in Neural Information Processing Systems*, vol. 18, pag. 1417–1426, 2005.
- [163] C. Wang, M. Gorce and N. Paragios. Segmentation, Ordering and Multi-Object Tracking using Graphical Models. In *Proceedings of the in 12th International Conference on Computer Vision*, pag. 747–754, 2009.
- [164] X. Wang, T.X. Han and S.Yan. An HOG-LBP human detector with partial occlusion handling. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'09)*, pag. 32–39, 2009.
- [165] D. Woodard, S. Pundlik, P. Miller, R. Jillela and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *Proceedings of the in 20th International Conference on Pattern Recognition (ICPR'10)*, pag. 201–204, 2010.
- [166] D.L. Woodard, S.J. Pundlik, P.E. Miller and J. R. Lyle. Appearance-based periocular features in the context of face and non-ideal iris recognition. *Signal, Image and Video Processing*, vol. 5, no. 4, pag. 443–455, 2011.
- [167] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pag. 210–227, August, 2009.
- [168] C.R. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pag. 780–785, 1997.
- [169] B. Wu and R.Nevatia. Detection and Segmentation of Multiple, Partially Occluded Objects by Grouping, Merging, Assigning Part Detection Responses. *International Journal of Computer Vision*, vol. 82, no. 2, pag. 185–204, 2009.

- [170] J. Wu, J. Xia, J. Chen and Z. Cui. Adaptive Detection of Moving Vehicle Based on On-line Clustering. *Journal of Computers*, vol. 6, no. 10, pag. 2045–2052, 2011.
- [171] J. Yao and J-M. Odobez. Fast human detection from joint appearance and foreground feature subset covariances. *Computer Vision and Image Understanding*, vol. 115, no. 10, pag. 1414–1426, 2011.
- [172] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In Proceedings of the 17th *International Conference on Pattern Recognition (ICPR'04)*, vol. 2, pag. 28–31, 2004.
- [173] C.T. Zahn and R.Z. Roskies. Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers*, vol. 21, no. 3, pag. 269–281, 1972.
- [174] L. Zhang, S.Z. Li, X. Yuan and S. Xiang. Real-time Object Classification in Video Surveillance Based on Appearance Learning. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pag. 1–8, 2007.
- [175] J. Zhang, J. Pu, C. Chen and R. Fleischer. Low-resolution gait recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 4, pag. 986–996, 2010.
- [176] X. Zhang, W. Hu, H. Bao and S. Maybank. Robust Head Tracking Based on Multiple Cues Fusion in the Kernel-Bayesian Framework. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 7, pag. 1197–1208, 2013.
- [177] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pag. 1208–1221, 2004.
- [178] W. Zhong, H. Lu and M-H. Yang. Robust object tracking via sparsity-based collaborative model. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)*, pag. 1838–1845, 2012.
- [179] Q. Zhou and J.K. Aggarwal. Probabilistic recognition of human faces from video. *Image and Vision Computing*, vol. 24, no. 11, pag. 1244–1255, 2006.
- [180] S. Zhou, V. Krueger and R. Chellappa. Object tracking in an outdoor environment using fusion of features and cameras. *Computer Vision and Image Understanding*, vol. 91, no. 12, pag. 214–245, 2003.
- [181] J. Zuo, N. Kalka and N. Schmid. A robust iris segmentation procedure for unconstrained subject presentation. In Proceedings of the *Biometric Consortium Conference*, pag. 1–6, 2006.