Design and Development of multi-biometric systems

Galdi Chiara

▶ **To cite this version:**

Galdi Chiara. Design and Development of multi-biometric systems. Signal and Image Processing. Telecom ParisTech; Universita degli studi di Salerno, 2016. English. NNT : . tel-01483609

HAL Id: tel-01483609
https://theses.hal.science/tel-01483609

Submitted on 6 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0015

EDITE - ED 130

Università degli Studi di Salerno

Doctorat ParisTech

Cotutelle de thèse
entre
Télécom ParisTech *(France)*
et
L'Université de Salerno *(Italie)*

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Traitement du signal et des images »

présentée et soutenue publiquement par

Chiara GALDI

le 22 Fevrier 2016

Conception et développement de systémes biométriques multimodaux

Directeur de thèse (France) : **Jean-Luc DUGELAY**
Directeur de thèse (Italie) : **Genoveffa TORTORA**

Jury

M. Jean-Luc DUGELAY, Professeur, Eurécom	Directeur de thèse
Mme Genoveffa TORTORA, Professeur, Université de Salerno	Directeur de thèse
Mme Bernadette DORIZZI, Professeur, Télécom SudParis	Rapporteur
M. Massimo TISTARELLI, Professeur, Université de Sassari	Rapporteur
Mme Maria DE MARSICO, Professeur associé, La Sapienza - Université de Rome	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E

Università degli Studi di Salerno
Corso di Dottorato in Economia e Direzione delle Aziende Pubbliche - XIV ciclo

in cotutela con
Télécom ParisTech
Ecole doctorale EDITE - ED 130

Tesi di Dottorato

Design and Development of multi-biometric systems

Candidata

Chiara Galdi

Tutors
Prof.ssa Genoveffa Tortora
Università degli Studi di Salerno
Prof. Jean-Luc Dugelay
Télécom ParisTech

Coordinatore
Prof.ssa Paola Adinolfi

Anno Accademico 2014/2015

Contents

1 Introduction **11**

Biometric recognition on mobile devices 11

Thesis outline 13

1.1 Authentication overview 13

1.2 Biometrics overview 14

1.3 Iris recognition overview 16

1.4 Challenges 20

1.4.1 ICE 20

1.4.2 NICE 20

1.4.3 MICHE 26

MICHE I - Evaluation protocol 27

MICHE II - Challenge protocol 29

1.5 Commercial applications 31

2 State of the Art **33**

Behavioural biometric recognition on mobile devices . . . 33

Physical biometric recognition on mobile devices 34

Iris detection and recognition on mobile devices 35

3 MICHE dataset **37**
3.1 Acquisition Protocol . . . 38
3.2 Database composition . . . 39
3.2.1 MICHE Iris . . . 40
3.2.2 MICHE Iris - Fake . . . 41
3.2.3 MICHE Iris - Video . . . 41
3.2.4 MICHE Face . . . 42
3.2.5 MICHE Eyes . . . 43
3.3 Metadata . . . 43
3.4 Noise Factors . . . 44

4 Iris segmentation on mobile devices **47**
4.1 ISIS . . . 48
4.1.1 Preprocessing . . . 48
4.1.2 Pupil location . . . 48
4.1.3 Linearisation . . . 50
4.1.4 Limbus location . . . 50
4.2 BIRD . . . 51
4.2.1 Pre-processing . . . 52
4.2.2 Watershed transform and binarization . . . 54
Watershed transform, region merging and colour quantization . . . 55
Binarization of watershed transform . . . 56
4.2.3 Iris Detection . . . 57
Circle detection . . . 57
Limbus boundary refinement . . . 59
4.2.4 Experimental results . . . 61
Tan et al. technique . . . 62
4.3 Conclusive remarks . . . 67

5 Multi-biometric and multi-modal authentication on mobile devices **69**
5.1 Combining iris and periocular area . . . 70

5.1.1 Periocular region segmentation 70
5.1.2 Iris and periocular area feature extraction 71
CSUM algorithm 72
5.1.3 Experimental results 73
5.2 Combining Iris and Face 77
5.2.1 Face recognition 79
Acquisition and segmentation 79
Spoofing detection 81
Best template selection 82
Feature extraction and matching 84
5.2.2 Iris recognition 85
Acquisition and segmentation 85
5.2.3 Score level fusion of face and iris biometrics 86
Confidence values 87
Fusion schema 89
5.2.4 Experimental results 90
Data acquisition and preprocessing 90
Performance of single biometrics 90
Performance of FIRME with different confidence functions 92
5.3 Combining Biometry and Hardwaremetry 94
5.3.1 Hardwaremetry 95
5.3.2 Score Normalization 99
Max-Min normalization technique 99
Z-score normalization technique 100
Median/MAD normalization technique 100
TanH normalization technique 100
Sigmoidal normalization technique 101
5.3.3 Experimental results 101
Data acquisition and preprocessing 101
Sensor recognition 101
Different sensors of the same model 102
Reference Sensor Pattern Noise Extraction 103

"Thesis_GALDI_TelecomParisTech" — 2016/5/19 — 15:17 — page 8 — #8

Iris recognition . . . 104
Fusion at feature level . . . 106
Fusion at score level . . . 107
Noise response . . . 107
5.3.4 Conclusive Remarks . . . 108
Future Implications and Open Issues . . . 111

6 Gaze Analysis 113
6.1 Method . . . 114
6.1.1 Data acquisition . . . 114
6.1.2 Data normalization . . . 115
Face fiducial points detection . . . 116
6.1.3 Feature extraction . . . 117
Features graph . . . 119
6.1.4 Comparison . . . 123
6.2 Experimental results . . . 124
6.2.1 Experimental protocol . . . 124
6.2.2 Results . . . 125
6.2.3 Single feature experiments . . . 128
6.2.4 Combined features experiments . . . 128
6.2.5 Weighted combined features experiments . . . 130
6.2.6 Comparison with previous experiments . . . 130
6.3 Conclusive remarks . . . 131

7 Conclusions 135
7.1 Future perspectives . . . 136

A Résumé en français 139
A.1 Introduction . . . 139
A.1.1 Reconnaissance biométrique sur les appareils mobiles . . . 140
A.1.2 L'authentification . . . 141
A.1.3 La biométrie . . . 142
A.1.4 La reconnaissance de l'iris . . . 142

A.2 La base de données biométrique MICHE 146
A.2.1 Protocole d'acquisition 147
A.3 Segmentation de l'iris sur les appareils mobiles 149
A.3.1 ISIS 149
A.3.2 BIRD 150
A.4 Authentification multi-biométrique et multi-modale sur les appareils mobiles 152
A.4.1 Combinaison de l'iris avec la zone périoculaire 152
A.4.2 Combinaison de l'iris avec le Visage 153
A.4.3 Combinaison de la biométrie avec l'hardwaremétrie 153
A.5 Analyse du regard 155
A.6 Conclusions 156

Chapter 1

Introduction

Biometrics can provide a higher level of security compared to other authentication systems based on passwords or cards, however there are some issues related to the characteristics of biometrics themselves (some change substantially over time) or devices used to capture them (some can be fooled or they may have difficulties in acquiring the biometric trait) that deter their spread. Biometrics mostly used for the automatic recognition of people are fingerprints and face. The first is highly reliable but computationally expensive, while the second requires a well-controlled setting. We will see that the iris lends itself much better than other biometrics to reliable identification, but that applications on the market until today have been limited by the need to acquire the iris at a close distance and with cooperation from the user. For this reason, recent research investigates the use of iris recognition systems in the presence of noise in order to develop reliable systems that can acquire the iris at a distance and with little cooperation from the user. [27]

Biometric recognition on mobile devices

Biometric recognition for a long time has been used in confined spaces, usually indoor, where security-critical operations required high accuracy recognition

systems, e.g. in police stations, banks, companies, airports (usually for frequent flyers, so just for a limited number of voyagers). Field activities, on the contrary, required more portability and flexibility leading to the development of devices for less constrained biometric traits acquisition and consequently of robust algorithms for biometric recognition in less constrained conditions [10]. However, the application of "portable" biometric recognition, was still limited in specific fields e.g. for immigration control, and still required dedicated devices. A further step would be to spread the use of biometric recognition on personal devices, as personal computers, tablets and smartphones.

Some attempts in this direction were made embedding fingerprint scanners in laptops or smartphones[1]. However, so far biometric recognition on personal devices has been employed just for a limited set of tasks, as to unlock the screen using fingerprints instead of passwords, PINs, or patterns. One of the reasons is that systems presented so far can be easily spoofed, as the well-known hacking of the Touch ID on iPhone6.

In this thesis, the results of the study of new techniques for biometric recognition on mobile devices are presented. In particular, because of the background knowledge of the PhD candidate, the use of iris recognition has been investigated. Many aspects of iris recognition on mobile devices have been analysed, starting from the study of the issues related to the iris images acquisition process using mobile devices, that has led to the formulation of an acquisition protocol and the collection of an iris image database, namely the MICHE dataset, then analysing the challenges of iris segmentation on mobile devices, exploring the benefits of combining iris with other biometric traits or authentication items, and finally investigating its possible combination with a relative new biometric trait, i.e. the gaze.

It is worth noticing that each analysed aspect, has been then employed in the development of the others, e.g. the database has been used to test all the techniques developed, and the iris segmentation methods have been used as part of the recognition systems presented.

[1] iPhone 6 Touch ID: https://www.apple.com/iphone-6/touch-id/

Thesis outline

In chapter 2 "State of the art", an overview of the State of the Art in the field of the biometric recognition on mobile devices is presented; in chapter 3 "MICHE dataset", the MICHE dataset is presented, and its collection building process is described; in chapter 4 "Iris segmentation on mobile devices", two iris segmentation techniques, namely ISIS and BIRD, are presented; the works presented in chapter 5 "Multi-biometric and multi-modal authentication on mobile devices" illustrate two different approaches to combine iris recognition with another biometric trait or to combine iris recognition with sensor recognition; in chapter 6 "Gaze Analysis" the use of Gaze analysis for human recognition and for gender/age classification in order to verify its possible fusion with iris recognition has been investigated, since both biometric traits, iris pattern and gaze, come from the analysis of the eyes and thus can be captured at the same time.

The Thesis ends in chapter 7 with the conclusions and future perspectives.

1.1 Authentication overview

Authentication can be performed based on one or a combination of the following items:

- Something the user knows (e.g., password, personal identification number (PIN), secret answer, pattern);
- Something the user has (e.g., smart card, ID card, security token, software token);
- Something the user is or does (e.g. fingerprint, face, gait).

The last are known as biometrics and will be discussed in more detail later. For now we want to briefly analyse the security level associated with each type of authentication item or to combinations as well. As a premise, it is worth considering that passwords can be forgotten or snatched by malicious people,

physical objects such as badges and ID documents can be lost or stolen, while biometrics can hardly be stolen and also the process of falsification is much more complicated (e.g. plastic surgery). The most recent biometric recognition systems also embed mechanisms to recognize live biometrics (liveness detection) and fakes (spoofing detection). If we consider all possible combinations of the three factors of authentication, we obtain the following ranking, from lower to higher security:

1. Something the user knows;
2. Something the user has;
3. Something the user knows + something the user has (e.g. ATM card + PIN);
4. Something the user is or does;
5. Something the user has + something the user is or does (e.g. biometric passport);
6. Something the user knows + something the user is or does;
7. Something the user knows + Something the user has + Something the user is or does.

Figure 1.1 shows the relative degrees of security. Biometrics by itself ensures an adequate level of security which may be increased by combining it with the other factors.

1.2 Biometrics overview

Biometric authentication is the process of human identification by their physiological or behavioural characteristics. These characteristic have to be distinctive and measurable in order to perform recognition.

Recognition can be performed in verification mode (1:1 matching, when the subject claims an identity that must be verified), or in identification mode (1:N

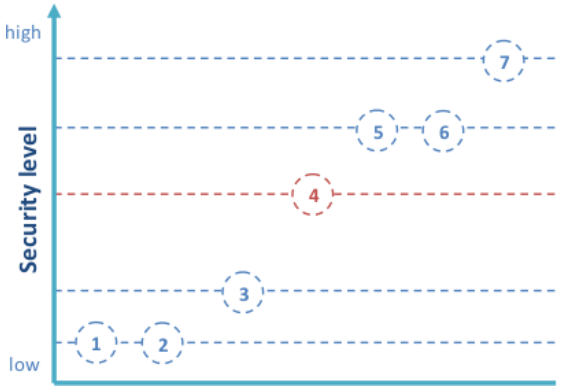

Figure 1.1: Security levels. (1) Something the user knows; (2) Something the user has; (3) Something the user knows + something the user has; (4) Something the user is or does; (5) Something the user has + something the user is or does; (6) Something the user knows + something the user is or does; (7) Something the user knows + something the user has + something the user is or does.

matching, or one against all, when there is no preliminary claim and the system must return the identity of a probe subject). Physiological biometrics include fingerprints, face, hand geometry, retina, DNA, veins pattern, ear and iris. Behavioural biometrics are related to the particular behaviour of a person and may be influenced by one's mood; they includes signature, speech, keystroke and gait. A good biometrics has to satisfy the following features: uniqueness, permanence, ease of use, good performance, accuracy, low cost, positive public perception. Iris optimally satisfies almost all of them.

Uniqueness: iris complex pattern can contain many distinctive features such as arching ligaments, furrows, ridges, crypts, rings, corona, freckles, and a zigzag collarette. It is proved statistically that iris is more accurate than even DNA matching as the probability of two irises being identical is 1 in 10 to the power of 78.

Permanence: the iris begins to form in the third month of gestation and the structures creating its pattern are largely complete by the eighth month, although pigment accretion can continue into the first postnatal years. Then it

remains almost unchanged lifelong. Its position behind the cornea protects it from the environment.

Ease of use: iris is externally visible and automatic eyes detection is a relatively simple operation.

Performance: template obtained from iris are small and feature extraction and matching on iris are very fast operations.

Accuracy: the iris has the great mathematical advantage that its pattern variability among different persons is enormous.

Low cost: iris recognition devices may have limited costs and new noisy iris recognition systems that use simple camera devices will have even lower prices.

Positive public perception: although iris image can be acquired without direct contact, iris based recognition systems are still perceived as intrusive [37].

1.3 Iris recognition overview

The identification process can be seen as a classification problem. A biometric trait can be reliably classified only if the variability among different instances of a given class is less than the variability between different classes. For example images of the same face have a high variability (intra-class variability) due to expressions, as well as being an active three-dimensional (3-D) object whose image varies with viewing angle, pose, illumination, accoutrements, and age. On the other hand face has a limited inter-class variability because different faces possess the same basic set of features, in the same canonical geometry. On the opposite iris inter-class variability is enormous and intra-class variability is low: as a planar object its image is relatively insensitive to angle of illumination, and changes in viewing angle cause only affine transformations. Even the non-affine pattern distortion caused by pupillary dilation is readily reversible [11].

The weakest element of iris recognition is the relatively low public acceptance. Although iris acquisition is performed in a contactless way, related applications are perceived as intrusive. Many system use NIR (Near Infra Red) illumination. Such kind of illumination is used because it is not visible and allows to light eyes up without annoying the users. However, even though studies

confirm that a few seconds exposure to NIR ray does not damage eyes in normal conditions, it is not clear what could happen to eyes or skin with pre-existing pathologies, or what if a subject is accidentally exposed to NIR ray for a long time.

Almost all commercial systems, visible or NIR light-based, require the users to stand at a distance of up to 1 m (usually much less), in order to capture an high quality iris image. The need of standard conditions and cooperation of users, still limits the application fields for iris-based recognition. Therefore, new techniques for noisy iris recognition have been proposed. "Noisy Iris" relates to the quality of iris images on which recognition is performed [43]. They could present the following problems:

- Occlusions: eyelids, eyelashes, glasses, hair, etc.;
- Reflections;
- Different size;
- Low resolution (due to device or distance);
- Different dominant colours in images of the same iris (due to different conditions during acquisition).

Such problems may especially arise if recognition is performed on subjects at a distance, on the move, unconscious of ongoing acquisition, as well as if there is not a standard illumination, or simply if a lower level of user's cooperation is desired in order to speed up the identification process. Noisy iris recognition phases are the same used in controlled conditions, and therefore in "traditional" systems, even if they require different approaches due to image characteristics mentioned before. Such phases are: acquisition; segmentation; normalization; coding; matching.

Acquisition: with respect to traditional systems, acquisition is not necessarily performed with dedicated devices or high quality video camera. Iris images may be obtained from simple cameras, or standard acquisition equipment built in computers or mobile devices. Acquisition conditions (illumina-

tion, distance, pose, etc.) are not strictly controlled, contrarily to traditional systems.

Segmentation: it is the process of identification of iris boundaries in order to extract only iris information from eye images. In traditional systems this is a relatively simple operation consisting in finding two circles matching with pupillary-iris and pupillary-sclera boundaries. With noisy iris, segmentation is much more complicated. It has to take into account the possible presence of occlusions or reflections, which must be discarded, in the sense that the corresponding area has not to be considered for coding and matching. The identification of the boundaries is further hindered by the low resolution or noise presence, which make boundaries less clear. For this reason noisy iris segmentation methods usually implement a preprocessing phase in which smoothing filter (to reduce noise) and/or enhance filter (to enhance feature such as iris boundaries) are applied [36][28].

Normalization: in traditional systems, due to the controlled acquisition condition, it is only needed to normalize the segmented iris form. Usual normalization implies transforming Cartesian coordinates into Polar ones. If colour information is taken into account, colour correction, histogram normalization or similar operations may be also useful.

Coding: this phase produces a template or feature vector, i.e., a compact representation of an iris image. Differences in the feature extraction algorithms when noisy irises are processed depend on the fact that in high quality images even tiny iris texture details are easily visible. On the contrary, noisy images may present altered or less characteristics to observe. The adopted methods for feature extraction in noisy iris images mainly analyse iris texture e.g. colour distribution, presence of lighter or darker region and can also combine a number of operator each computing a particular feature [29].

Matching: matching phase only depends on the kind of templates used.

Figure 1.2 is an illustration of the five phases described above.

In the following we will describe some research initiatives aimed at evaluating results of current research on iris recognition.

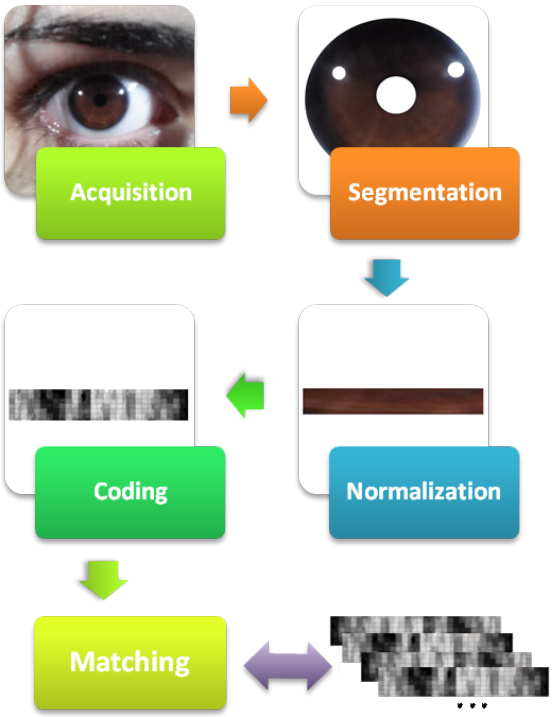

Figure 1.2: Iris recognition phases.

1.4 Challenges

1.4.1 ICE

The National Institute of Standards and Technology (NIST) conducted and managed the Iris Challenge Evaluation (ICE) projects. ICE 2005 was the first public iris recognition contest. The goals were to promote the development of iris recognition algorithms and to evaluate the submitted solutions with a standard protocol in order to obtain a meaningful comparison of their performance [39]. For ICE 2005 a standard set of items were provided:

- a publicly-available dataset for algorithm development;
- an experimental protocol for measuring performance;
- the irisBEE baseline algorithm.

The database provided by ICE, including 2953 iris images from 132 subjects, was one of the largest publicly-available iris image database at that time. However its images were captured with the aim of obtaining high quality samples and simulating the users' cooperation in the image capturing process. Therefore, the noise factors in the ICE database are almost exclusively occlusions and poor focused images [39]. For ICE 2006 the performance evaluation of iris recognition algorithms was performed on sequestered data (data not previously seen by the researchers or developers).

1.4.2 NICE

NICE (Noisy Iris Challenge Evaluation) was born to promote the development of noisy iris recognition solutions. This iris biometric evaluation initiative of the SOCIA Lab. (Soft Computing and Image Analysis Group) of the University of Beira Interior (Portugal), received worldwide participations [44]. The competition was performed in two phases:

- NICE.I (2007-2009): evaluated iris segmentation and noise detection techniques;

- NICE.II (2009-2011): evaluated encoding and matching strategies for biometric signatures.

The proposed methods were tested on a database provided by NICE itself: UBIRIS.v2 [42]. The UBIRIS database is one of the few iris image databases that contains realistic noise factors that make it suitable for the evaluation of robust iris recognition methods [39]. It was developed within the SOCIA Lab. and released in September, 2004. The main feature of UBIRIS.v2 database is that eye images contain an high level of noise in order to simulate less constrained capturing conditions, e.g., acquisition at-a-distance, on-the-move, with minor cooperation or within dynamic imaging environments. Another important aspect of this database is that iris images are taken on the visible wavelength in spite of the controlled databases in which acquisition is usually performed under controlled NIR illumination. NICE iris image database contains:

1. Out-of-focus iris images. Due to the limited depth-of-field of the camera.
2. Off-angle iris images. Obtained when the subject is not looking straight at the acquisition device. In this kind of images pupil and iris have an elliptical shape that needs to be taken into account during the detection of pupil and iris boundaries.
3. Rotated iris images. When the subject's body/head is not in the vertical (natural) position.
4. Motion blurred iris images. Due to on-the-move acquisition or eyelids movements.
5. Iris occlusion due to eyelashes.
6. Iris occlusion due to eyelids.
7. Iris occlusion due to glasses, in particular to eyeglass frames and/or to reflections on the lenses.
8. Iris occlusion due to contact lenses. Contact lenses with high optical power can cause non-linear deformations of the iris texture.

9. Iris with specular reflections. These reflections appear as small spots that obstruct the iris texture and are relatively easy to remove in the segmentation phase because they are usually much lighter than the iris.

10. Iris with diffuse reflections. These reflections are due to reflected information from the environment where the subject is located or is looking at. They can obstruct a large part of the iris.

11. Partial captured iris. The acquisition at-a-distance and on-the-move does not guarantee to capture the entire iris.

12. Out-of-iris images. In this case the system failed to capture the iris. However the recognition process has to be able also to understand that in the image taken there is not an iris and for example require to repeat the acquisition.

This database was downloaded by over 500 users (individuals and academic, researchers and commercial institutions) from over 70 different countries of the world [41]. The acquisition device used in capturing eye images of the UBIRIS database is a simple camera. Details of the imaging framework set-up are given in Table 1. It was installed in a lounge under both natural and artificial lighting sources. Volunteers were of different ethnicities:

- Latin Caucasian (approximately 90%);
- Black (8%);
- Asian (2%).

They were only required to walk at a speed slightly slower than normal, in an area comprised between three and ten meters away from the acquisition device, and to look at few marks, located laterally with respect to the field of view of the camera, to simulate the behaviour of a non-cooperative subject that does not look straight at the camera. The large distance between the subject and the acquisition device is one of the main differences between the UBIRIS.v2 database and most of the remaining ones.

Two distinct acquisition sessions were performed, each lasting two weeks and separated by an interval of one week. From the first to the second session, both the location and orientation of the acquisition device and artificial light sources were changed. Approximately 60% of the volunteers participated in both imaging sessions, whereas 40% participated exclusively in one or the other.

Image Acquisition Framework and Set-Up	
Camera = Canon EOS 5D	Colour Representation = sRGB
Shutter Speed = 1/197 sec.	Lens Aperture = F/6.4 - F/7
Focal Length = 400 mm	F-Number = F/6.3 - F/7.1
Exposure Time = 1/200 sec.	ISO Speed = ISO-1600
Metering Mode = Pattern	
Details of the Manually Cropped Resultant Images	
Width = 400 pixels	Height = 300 pixels
Format = tiff	Horizontal Resolution = 72 dpi
Vertical Resolution = 72 dpi	Bit Depth = 24 bit
Volunteers	
Totals = Subjects 261; Irises 522; Images 11 102	Gender = Male: 54.4%; Female: 45.6%
Age = [0,20]: 6.6% [21,25]: 32.9% [26,30]: 23.8% [31,35]: 21.0% [36,99]: 15.7%	Iris Pigmentation = Light: 18.3% Medium: 42.6% Heavy: 39.1%

Table 1.1: UBIRIS.v2 imaging framework

Both in NICE.I and NICE.II participants were required to submit an application executable written in any programming language and running in standalone mode. Evaluation for NICE.I (segmentation and noise detection) was performed using the following sets:

1. Alg denoted the submitted executable, which performs the segmentation of the noise free regions of the iris.
2. $I = \{I_1, ..., I_n\}$ was the data set containing the input close-up iris images.
3. $O = \{O_1, ..., O_n\}$ were the output images corresponding to the above described inputs, such that $Alg(I_i) = O_i$.

4. $C = \{C_1, ..., C_n\}$ were the manually classified binary iris images, given by the NICE.I Organizing Committee. It must be assumed that each C_i contains the perfect iris segmentation and noise detection result for the input image I_i.

All the images of I, O and C had the same dimensions: c columns and r rows. Two measures of evaluation were used:

- The classification error rate (E^1);
- The type-I and type-II error rate (E^2).

The classification error rate (E^1) of the Alg on the input image $I_i(E_i)$ is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator) over the whole image:

$$E_i = \frac{1}{(c \times r)} \sum_{c'} \sum_{r'} O(c', r') \otimes C(c', r')$$

where $O(c', r')$ and $C(c', r')$ are, respectively, pixels of the output and class images. The classification error rate (E^1) of the Alg is given by the average of the errors on the input images E_i:

$$E = \frac{1}{n} \sum_{i} E_i$$

The value of (E^1) belongs to the [0, 1] interval and was the measure of evaluation and classification of the NICE.I participants. In this context, "1" and "0" will be respectively the worst and optimal values.

The second error measure aims to compensate the disproportion between the a-priori probabilities of "iris" and "non-iris" pixels in the images. The type-I and type-II error rate (E^2) of the image E_i is given by the average between the false-positives (FPR) and false-negatives (FNR) rates:

$$E_i = 0.5 * FPR + 0.5FNR$$

Similarly to the E^1 error rate, the final E^2 error rate is given by the average of the errors (E_i) on the input images. The best 8 participants, that achieved the lowest test error rates, were invited to publish their approach in a Special Issue on the Segmentation of visible Wavelength Iris Images Captured At-a-distance and On-the-move, Elsevier Image and Vision Computing 28 (2010).

The evaluation procedure for NICE.II (encoding and matching strategy) was the following:

- Let P denote the submitted application, which gives the dissimilarity between segmented iris images.
- Let $I = \{I_1, ..., I_n\}$ be the data set containing the input iris images and let $M = \{M_1, ..., M_n\}$ be the corresponding binary maps that give the segmentation of the noise-free iris region.

    1. P receives two iris images (and the corresponding binary maps) and outputs the dissimilarity value between the corresponding irises: $P(I_i, M_i I_j, M_j) \rightarrow D.D$ should be a real positive value.
    2. Performing a "one-against-all" comparison scheme for each image of I gives a set of intra-class dissimilarity values $D^I = \{D_1^I, ..., D_k^I\}$ and a set of inter-class dissimilarity values $D^E = \{D_1^E, ..., D_m^E\}$, whether the captured images are from the same or from different irises.
    3. The decidability value $d'(D_1^I, ..., D_k^I, D_1^E, ..., D_m^E) \rightarrow [0, \infty[$ was used as evaluation measure:

    $$d' = |avg(D^I) - avg(D^E)|/sqrt(0.5 * (std(D^I)^2 + std(D^E)^2))$$

    where $avg(D^I)$ and $avg(D^E)$ denote the average values of the intra-class and inter-class comparisons and $std(D^I)$ and $std(D^E)$ the corresponding standard deviation values.

Participants of the NICE:II contest were ranked from the highest (best) to the lowest (worst) decidability values [45].

1.4.3 MICHE

A more challenging problem is faced by MICHE (Mobile Iris CHallenge Evaluation). As the name suggests, MICHE is an iris recognition technology evaluation that requires all the steps of the iris recognition algorithm to be performed on a mobile device (smartphones or tablets).

Performing biometric recognition on mobile devices is an important issue due to the need of a secure use of critical services (e.g. home banking) and to protect sensitive data that nowadays are mostly stored on our personal smartphones or tablets.

Iris is a natural candidate for mobile biometric recognition for two main reasons: iris acquisition is little intrusive, and iris codes are among the less expensive templates from the storage point of view.

MICHE contest, is the result of the collaboration of the BIPLab (Biometric and Image Processing Lab) from the University of Salerno (Italy) and the SOCIA Lab. (Soft Computing and Image Analysis Group) of the University of Beira Interior (Portugal). MICHE will include two phases:

- MICHE I (2014-2015): participants are required to provide both their executable programs and the dataset they used for their experiments, as well as the characteristics of the devices used for acquisition and testing. They can submit their results related to one or all of the steps of an iris recognition system performed on a mobile device (detection, segmentation, recognition) as well as applications of iris biometrics on mobile devices;

- MICHE II (2015-2016): the collected dataset is used to as test-bed for a challenge which is accessible for both original authors and new groups [46].

Performing iris recognition on mobile devices may introduce many noise factors while performing the acquisition due to the fact that:

- the user may need to get authenticated in any moment and in any place, with different illumination conditions, while walking, standing or sitting;

- the user holds the mobile device by his hand and may involuntarily move the device;
- the acquisition device characteristics may influence the acquisition: resolution of the sensor, presence of the frontal camera, possibility of using voice control to take the picture, etc.

In order to develop a robust solution for iris recognition on mobile devices, the database used for testing should simulate the uncontrolled acquisition conditions described above. An example of such images acquired by mobile devices is shown in Figure 1.3.

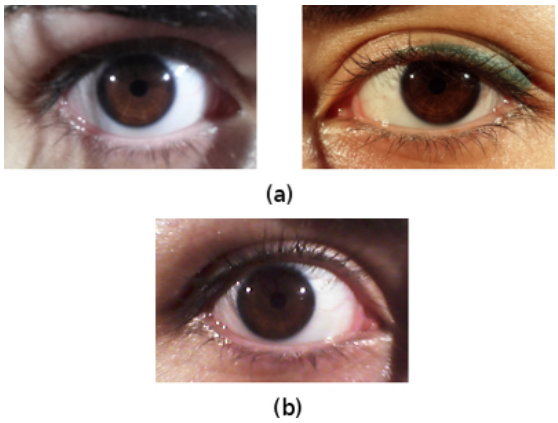

Figure 1.3: MICHE I iris images example. (a) Images acquired with Samsung Galaxy S4 rear camera (left) and front camera (right); (b) Image captured with Samsung Galaxy Tab 2 front camera.

MICHE I - Evaluation protocol

The algorithms presented to *MICHE I* deal with both iris segmentation and iris recognition. Two evaluations were carried out in order to compare the algorithms and to identify the best segmentation and the best recognition methods.

Four iris segmentation algorithms have been tested on a subset of MICHE database. Here are summarized the metrics used to evaluate the segmentation

quality of the methods submitted to MICHE I. Each method was evaluated on a subset of MICHE database for which ground-truth data have been made available:

- **PRATT**: This metric is a formulated function of the distance between correct and measured edge positions, but it is also indirectly related to the false positive and false negative edges.
- **F1 Score**: It is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score. The $F1score$ can be interpreted as a weighted average of the precision and recall, where an $F1$ score reaches its best value at 1 and worst at 0.
- **Rand Index**: RI counts the fraction of pairs of pixels whose labelling are consistent between the computed segmentation and the ground truth.
- **Global Consistency Error**: The Global Consistency Error (GCE) [Martin2001] measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this manner are considered to be consistent, since they could represent the same natural image segmented at different scales.
- **E1 Score**: The classification error rate ($E1$) of the algorithm on the input image is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator) over the whole image.
- **Pearson Correlation Coefficient**: It is a measure of the linear correlation between two variables X and Y, giving a value between $+1$ and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation.

The iris recognition methods submitted to MICHE I have been evaluated in terms of decidability, area under ROC curve and equal error rate. The algorithms have been tested on iris images segmented by the segmentation methods proposed to MICHE I and discussed above.

MICHE II - Challenge protocol

MICHE II [2] is a challenge for iris recognition on the MICHE database (see 3 for details). Each executable should therefore be able to receive from command line a pair of images from the dataset and a pair of corresponding segmentation masks and should produce a score in terms of dissimilarity between the two irises.

The order of inputs is strictly defined. Let:

`I1 = image1Filename.ext` be the first RGB image containing an iris;

`M1 = mask1Filename.ext` be the binary mask of `I1`;

`I2 = image2Filename.ext` be the second RGB image containing an iris;

`M2 = mask2Filename.ext` be the binary mask of `I2`;

`path` be the directory for matching results;

Let `APP` be the executable application, then by running:

```
APP I1 M1 I2 M2 path
```

a TXT file containing the `dissimilarity_score` is created. Such TXT file must have the following properties:

a. it is saved in `path` (preferably something like "./results");

b. its filename is `image1Filename_image2Filename.txt` (NOTE. filenames without file extensions);

c. its content is `image1Filename [whitespace] image2Filename [whitespace] dissimilarity_score` (NOTE. filenames without file extensions).

The dissimilarity score $d : \{d \in \Re : 0 \leq d \leq 1\}$ is meant as the probability that two irises are from two different subjects. The higher is the dissimilarity the higher is the probability that the two irises are not from the same person. Let I be set of images from MICHE database, the dissimilarity function D is defined as:

[2] http://biplab.unisa.it/MICHE/MICHE-II/Protocol.html

$$D : I_a \times I_b \rightarrow [0, 1] \subset \Re$$

where $I_a, I_b \in I$
and satisfies the following properties:

a. $D(I_a, I_a) = 0$;

b. $D(I_a, I_b) = 0 \rightarrow I_a = I_b$;

c. $D(I_a, I_b) = D(I_b, I_a)$.

The participants can use the whole MICHE database for developing and performing experimentations of their proposed algorithm. The participants should take into account that the dataset is going to be extended with new acquisitions by new mobiles and of new subjects according to the same acquisition protocol applied to the current version of the database. The challenge will be run on a subset of the new version of the MICHE database that will be revealed together with the final ranking.

Participants must consider that the best segmentation algorithm submitted to *MICHE I* (The "Unsupervised detection of non-iris occlusions", by Haindl *et al.*, is available for download on MICHE II website) will be used to generate the binary masks. Since it will be used also for the final ranking of submitted algorithms, participants are invited to use it for testing their proposal. Given an RGB image in input, the segmentation algorithm gives in output (a) the binary mask; (b) the normalised mask of the iris region; (c) the normalised RGB iris extracted from the image (see figure 1.4 below).

Each executable is supposed to be self-contained and it will not have access to the Internet. No any additional download has to be expected to run the application. The submitted proposal must therefore contain all supporting files (dlls, libraries and so on) useful to its proper running.

The executable can be written in any programming language and should run on one of the following operating systems: (1) Windows 7 64/32 bit, (2) Linux Ubuntu 14.04. Code written in Matlab is also acceptable at condition that

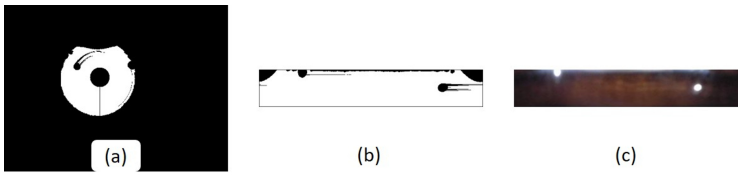

Figure 1.4: *MICHE II*: (a) the binary mask; (b) the normalised mask of the iris region; (c) the normalised RGB iris extracted from the image.

it runs on Matlab 2013. In case of any special setting needed for the proper running of the algorithm, a README file is expected.

Executables that do not match the requirements above could be discarded from the contest at the discretion of the Evaluating Committee.

1.5 Commercial applications

Current commercial applications mostly address access control to restricted areas. Iris recognition systems in less controlled situation were installed in some major international airports, especially in the UK and the Arab Emirates. For example both Gatwick airport and Dubai airport adopted an AOptix solution.

AOptix InSight®VM iris recognition system was integrated into 34 automated e-Gates at the Gatwick Airport South Terminal in order to speed up the process of passport control, formerly performed manually. The iris recognition system allows passengers to be acquired at a distance of two meters, and they are only required to look at a specific point indicated on the device. A monitor gives passengers a textual feedback indicating them where to look at the device and eventually to open their eyes if there is an occlusion problem. The recognition process lasts few seconds. The illumination employed is a NIR lamp. The system can perform recognition whether passenger is in a wheelchair or above 2.15 meter tall.

AOptix InSight®Duo maintains the same characteristics of AOptix InSight®VM but provides a combination of iris and face recognition. It was adopted in the

Dubai airport and probably will be soon integrated also at Gatwick [3].

It is interesting to know that before using AOptix products, Gatwick and many UK airports adopted an iris recognition system in controlled condition. However because of the high percentage of false rejections and the difficulties for the passengers in lining up eyes with the iris recognition equipment, the identification process took a lot longer than it was supposed to and the system was abandoned. This demonstrates the need to develop iris recognition systems requiring less and less users' collaboration and that are suitable to any type of environment in order to be a valid and faster but also more secure tool than existing authentication systems.

Iris can be used also in multi-biometric systems, as the already mentioned AOptix InSight®Duo which combines face and iris biometric traits, but noisy iris recognition may be also combined with other biometrics or soft-biometrics, for example periocular information, gaze analysis, etc. [47][6]. Finally, thanks to the ever increasing technological development in a near future iris may be captured at a considerable distance, we can then imagine to integrate the iris biometric in video surveillance systems for people recognition or re-identification [38].

[3] AOptix, Identity Solutions: http://www.aoptix.net/identity-solutions/overview

Chapter 2

State of the Art

Among the issues linked to biometric recognition on mobile devices, there are undoubtedly the technological limits of nowadays devices: the classic configuration for a biometric recognition system, for example for face recognition, is composed by one or more high-quality cameras placed in fixed positions, possibly in an environment with controlled lighting, and connected to a processor. The use of a mobile device, e.g. a smartphone, for its nature, implies very different usage scenarios, limitations on the quality of the hardware, nonconformity with the software requirements of a biometric recognition system.

However, the opportunity offered by these devices is to make biometric recognition portable, replacing or strengthening the existing authentication systems, a good reason to further investigate in this direction.

In the following a quick overview on the biometric traits employed for the recognition on mobile devices is given.

Behavioural biometric recognition on mobile devices

Among the behavioural biometrics being studied, there are the user's gait modelled using the smartphone embedded accelerometer, as described in [12], the

recognition of the typing style on virtual keyboards, presented in [13], and the recognition of the arm movement, recorded by the embedded accelerometer when answering a call, proposed in [14], where the idea is to minimize user involvement in the authentication process, making it transparent to the user by using the arm movement that instinctively the user does to answer his/her phone.

Physical biometric recognition on mobile devices

For what concerns recognition on mobile devices based on physical biometrics, there are many examples, exploiting in particular the mobile devices embedded cameras and microphone. In [15], another approach for transparent user authentication is presented, but in this case, the ear is used and captured when the user answers (or places) a call. A multi-modal framework employing ear and speaker recognition is proposed in [16], based on the use of a hybridization of DWT (Discrete Wavelet Transform, using haar wavelet) and GLCM (Gray-Level Co-Occurrence Matrix) to extract both shape and texture information from ear images, and MFCC (Mel Frequency Cepstral Coefficient) technique to extract features from speech signal.

In [17] the system presented is based on creating a 3d face model by reconstructing depth information from videos recorded by the phone camera. One of the most recent works on mobile biometrics presents the MOBIO system, combining face and voice biometrics [18]. It is preferable to combine face with iris to simplify capture and processing, while at the same time we avoid heavy normalization procedures, which are performed in [18].In general the quality of captured signals, both audio and video, depends on factors, which are internal (device dependent) and external (environment dependent). Even if the concept of quality is often quite vague, we can assume for sure that resolution is one of the internal factors, which concur to it. The image resolution depends on the size and technology of the sensor, while the resolution of captured audio signals depends on the precision of the sensor and on sampling frequency. It is well known that during interaction, low quality in images is better tolerated

than low quality in sound [19], because the human visual system is able to cope with more interferences than the auditory one, and this often dictates the compression parameters, which are used during signal processing.

Iris detection and recognition on mobile devices

For what concerns iris detection, in [20] and [21] methods for pupil and iris boundaries detection are presented; in these two works however, the databases employed were collected respectively with a Samsung SPH-S2300 and Samsung SPH-2300[1] (in [20] only 132 images were captured with the mobile phone and the others were from CASIA database[2]) which embed a 3.2 megapixel digital camera with a 3X optical zoom, which is a very specific imaging sensor that cannot commonly be found in the most popular smartphones.

One of the first works investigating the possibility of optimizing iris segmentation and recognition on mobile phones is [22], Jeong *et al.* propose a method for computing the iris code based on Adaptive Gabor Filter. In [23], Park *et al.* present a recognition method based on corneal specular reflections, while Kang in [24] presents a method to pre-process iris in order to remove the noise related to occlusions of eyelids and improve system performances. In [25][26] an iris recognition system based on Spatial Histograms is presented.

[1]Samsung. http://www.samsung.com/
[2]CASIA database. http://www.sinobiometrics.com.

Chapter 3

MICHE dataset

In order to reliably test an approach, it is necessary to have a dataset that replicates the real conditions in which the approach should be applied. The database described in this chapter simulates the acquisition of iris and face on mobile devices.

The MICHE dataset [30] was collected on the occasion of the Mobile Iris CHallenge Evaluation (see paragraph 1.4.3) at the University of Salerno, by Chiara Galdi (at that time Ph.D. candidate at University of Salerno, Italy) and Silvio Barra (at that time Ph.D. candidate at University of Cagliari, Italy), members of the BIPLab (Biometric and Image Processing Lab) that promoted the Challenge and the collection of the MICHE biometric database.

The images were acquired with three different mobile devices, representative of the current top market category:

- iPhone 5 (hereafter IP5)

  Operating System: Apple iOS;

  Front Camera: FaceTime HD Camera with 1.2 Megapixels;

  Rear Camera: iSight with 8 Megapixels.

- Samsung Galaxy S4 (hereafter GS4)

Operating System: Google Android;

Front Camera: CMOS with 2 Megapixel;

Rear Camera: CMOS with 13 Megapixel.

- Samsung Galaxy Tablet II (hereafter GT2)

  Operating System: Google Android;

  Front Camera: VGA for Video Call;

  Rear Camera: 3 Megapixel Camera.

3.1 Acquisition Protocol

Biometry is very suitable for human recognition on mobile devices in fact the users are used to employ the frontal camera of their personal mobile devices to capture pictures of themselves, the so called "selfie". The subjects involved in the database acquisition process are only asked to take self-pictures of their face, eyes, and single iris, sometimes with both frontal and rear camera and sometimes with the frontal camera only. Further details on the procedure to acquire each sub-database are given later in this chapter in the respective paragraphs.

Since the goal of the acquisition process is to achieve a realistic simulation of the data capture process, users are left free to hold the devices (smartphones or tablet) in their hands, to use the voice commands (when available) and, for the subjects wearing eyeglasses, to decide if wear them or not during the acquisition according to what they would have done during a real use of such an application.

The standard acquisition process was the following:

1. Smartphones (GS4 and IP5):

   (a) **indoor:**

   - 4 shots of the face taken with the rear camera;
   - 4 shots of the two eyes (landscape format) taken with the rear camera;
   - 4 shots of the iris (left or right, the chosen one will be used for the further acquisition too) taken with the rear camera;

- 4 shots of the iris taken with the frontal camera.

(b) **outdoor:**

- 4 shots of the iris taken with the rear camera;
- 4 shots of the iris taken with the frontal camera.

2. Tablet (GT2):

(a) **indoor:**

- 4 shots of the iris taken with the frontal camera.

(b) **outdoor:**

- 4 shots of the iris taken with the frontal camera.

For a total of at least 56 pictures per person (there were few exceptions). During the indoor acquisition mode various sources of artificial light, sometimes combined with natural light sources, are used, while during the outdoor acquisition mode data capture takes place using natural light only. The resulting captured images are affected by different noise factors that we will discuss later in paragraph 3.4.

3.2 Database composition

The database currently consists of 92 different subjects with age ranging between 20 and 60 years, among them 66 are males and 26 are females and all of them are of Caucasian ethnicity. MICHE is composed by different sections:

1. **MICHE Iris**;
2. MICHE Iris - Fake;
3. MICHE Iris - Video.
4. **MICHE Face**;
5. **MICHE Eyes**.

Part of the database, a subset of MICHE Iris containing the subjects from ID number 1 to 75, the MICHE Iris - Fake, and MICHE Iris - Video, are available on-line for research purposes[1].

3.2.1 MICHE Iris

MICHE Iris is the main database section, in its current version it contains more than 3500 pictures of irises of 92 different subjects, around 40 pictures per subject.

Only one iris per subject has been captured but in different modalities, i.e. indoor, outdoor, and with all the three different mobile devices mentioned before. Some examples of iris images contained in this database are showed in figure 3.1. Due to the fact that users were let free to take a self-picture and because of the use of different devices, the resulting images are very different from each other.

Figure 3.1: *MICHE Iris* images example. First row: pictures acquired by GS4; Second row: pictures acquired by IP5; Third row: pictures acquired by GT2

[1]MICHE I database: http://biplab.unisa.it/MICHE/database/

Figure 3.2: *MICHE Iris* acquisition modalities.

In figure 3.2 some pictures of the same eye captured in indoor and outdoor modalities are given.

3.2.2 MICHE Iris - Fake

This database is suitable for the testing of anti-spoofing techniques, i.e. techniques to identify if the iris in front of the camera is a real iris or an artefact reproducing the real iris (e.g. a photo, a video, etc.).

This small database contains 40 photos of grey-scale printed eye images (by a LaserJet Printer) of 10 different subjects. Photos were taken by the Samsung Galaxy S4 rear camera. For each of the 10 subjects, 4 indoor images were selected from the *MICHE Iris* database, 2 pictures acquired by the GS4 and 2 acquired by the IP5 smartphone. Figure 3.3 shows some samples from *MICHE Iris - Fake* database.

3.2.3 MICHE Iris - Video

MICHE Iris - Video is made up of 113 videos of about 15 seconds, recording the eye of 10 subjects both in indoor and outdoor mode. Also this database is

Figure 3.3: *MICHE Fake* images example.

suitable to test anti-spoofing techniques and in particular those aimed to detect iris liveness.

3.2.4 MICHE Face

Although the *MICHE Face*, as the name suggests, is a face database, it was acquired with the purpose of providing a database suitable for iris and face recognition at once. In fact, the photos were taken in the best illumination conditions and with the highest-resolution cameras (smartphones rear cameras) in order to provide an adequate iris resolution for analysing its features. Some images contained in *MICHE Face* database are shown in figure 3.4.

Figure 3.4: *MICHE Face* images example.

3.2.5 MICHE Eyes

The idea behind the collection of this database is to provide images that contain both irises captured in one shot. This allows the study of advantages/disadvantages behind the use of both irises in a human recognition system. This kind of photos were captured only in indoor modality, by only the two smartphones and with their rear camera. The reason behind this choice is that since the relative iris size in these pictures is small an adequate resolution has to be assured, that is why the irises were collected in the most controlled scenario (indoor) and by the cameras with best resolution. Examples of *MICHE Eyes* are given in figure 3.5.

Figure 3.5: *MICHE Eyes* images example.

3.3 Metadata

MICHE database, at the best of our knowledge, is the first database that along with the images provides metadata files with full information about the subjects acquired, the acquisition device characteristics, the acquisition conditions. An example of the metadata structure (xml file) is given in the following:

```
<img>
   <filename>img.jpg</filename>
   <img_type>[iris, face]</img_type>
   <iris>[left, right, both]</iris>
   <distance_from_the_device>[100cm, 10cm]
            </distance_from_the_device> (in centimetres)
   <session_number>[01, 02, 03, ...]</session_number>
```

```
   <image_number>[1, 2, 3, 4, ...]</image_number>
   <user id="[001, 002, ..., 022, ...]">
        <age>[20, 32, 55, ...]</age>
        <gender>[M, F]</gender>
        <ethnicity>[Afro American, Asian, Caucasian, Indians,
                    Latinos]</ethnicity>
    </user>
    <device>
        <type>[Smartphone, Tablet]</type>
       <name>[IPhone5, Galaxy S4, Galaxy Tab 2, ...]</name>
        <camera type="[front, rear]">
            <name>[VGA, CMOS, iSight, ...]</name>
            <resolution>[0.3MP, ...]</resolution>
            <dpi>[72, ...]</dpi>
        </camera>
    </device>
    <condition>
        <location>[indoor, outdoor]</location>
        <illumination>[artificial, natural, both]</illumination>
    </condition>
    <author>[BIPLab, University of ...]</author>
</img>
```

3.4 Noise Factors

The noise factors have been in part previously presented in section 1.4.2: out-of-focus iris images; off-angle iris images; rotated iris images; motion blurred iris images; iris occlusion due to eyelashes; iris occlusion due to eyelids; iris occlusion due to glasses; iris occlusion due to contact lenses; iris occlusion due to hairs; occlusion due to shadows; iris with specular reflections; iris with diffuse reflections; partial captured iris; out-of-iris images. Another factor that can affect the iris segmentation phase is the presence of strong make-up in the picture,

since dark areas can be confused with the pupillary and a wrong detection of the pupillary boundary can lead to a completely wrong iris segmentation. Some examples of iris images affected by the noise factors listed above, are given in figure 3.6.

Figure 3.6: MICHE I iris noise factors examples: (a) out-of-focus iris image; (b-g) examples of images affected by strong light/shadows; (h) eyelids and eyelashes occlusion; (i) hairs occlusion; (j) out-of-iris image; (k-m) off-angle iris image; (m-o) partial captured iris; (p-q) strong make-up; (r) specular reflections; (s-t) diffuse reflection on eyeglasses.

Chapter 4

Iris segmentation on mobile devices

Iris recognition on mobile devices is a challenging task, in fact, with respect to other dedicated iris acquisition devices, usually fixed on a desk or on a stand, the use of the smartphone embedded sensors introduce a number of noisy factors during the iris acquisition process [41] due to the fact that the device is hand-held by the users: out-of-focus, off-angle iris, rotated iris images, motion blurring, occlusions due to eyelashes, occlusions due to eyelids, occlusions due to eyeglasses, occlusions due to contact lenses, specular reflections, diffuse reflections, partially captured iris. In off-angle iris images, the iris aspect ratio is distorted. All these aspects have to be taken into account when designing an iris segmentation algorithm. In the following, two methods for iris segmentation are presented, specifically designed for iris segmentation in adverse conditions and on mobile devices. These two methods, namely ISIS and BIRD, are then employed in the systems presented in chapter 5 "Multi-biometric and multi-modal authentication on mobile devices".

4.1 ISIS

ISIS (Iris Segmentation for Identification Systems) is an iris segmentation algorithm proposed by the BIPLab - Biometric and Image Processing Lab [31], of the University of Salerno. It was especially devised and implemented to address under-controlled acquisition conditions, therefore it is well suited to be used on mobile devices. It is robust to the presence of reflections and requires a limited computational time. It has four main phases:

- Pre-processing;
- Pupil location;
- Linearisation;
- Limbus location.

ISIS algorithm uses operative (e.g. image window sizes) and decision parameters (e.g. thresholds) that have been experimentally tuned by using a training set of images.

4.1.1 Preprocessing

Eye image contains many disturbing details such as sclera vessels, skin pores, or eyelashes shape; these are complex patterns that can negatively interfere with edge detection. Moreover, the same internal characteristic patterns, which are fundamental for recognition, may hinder a correct segmentation. To avoid this problem, a posterization filter FE (Enhance) is applied: a square window W is moved over the whole image, pixel by pixel, a histogram is computed for the region contained in W, and the value with the maximum frequency is substituted for the central position.

4.1.2 Pupil location

In this phase, a first step implies to apply Canny filtering on the preprocessed image with ten different thresholds $th = 0.05, 0.10, 0.15, ..., 0.55$. We start from

the assumption that relevant circles can be detected at different thresholds (e.g. at 0.15, 0.20, and 0.25), while circles detected only for a single value of the threshold may be artefacts. Choosing a fixed step (0.05) for the threshold value allows to explore uniformly its overall admissible range, which was experimentally found. Processing the image (Canny filtering, circle fitting) with different thresholds has a little impact on the computational burden of the overall technique, but significantly increases the accuracy of the segmented result. On the other hand, an adaptive threshold technique may concentrate only on specific parts of this range; however, it is important to explore it completely, because we don't know in advance the dominant grey level of the pupil. The connected components are identified in each resulting image, and those containing a number of pixels greater than a threshold Th_C are all included in a unique list L of starting candidates. Since the pupil is not a perfect circle, many approaches search elliptical shapes possibly representing it. However the presence of noise (e.g., spurious branches by Canny filter) may cause erroneous results from ellipse fitting algorithms. Therefore, ISIS detects circular objects within the image by using a precise and fast circle detection procedure presented by Taubin in [72]. Taubin's algorithm is applied to each element in L to compute the approximating circle. All the components whose corresponding circles are not completely contained inside the image are removed from L. To extract the real pupil boundary from all candidates in the final list, each remaining circle undergoes a voting procedure, according to the sum of two ranking criteria:
Homogeneity: pupil is a darker, homogeneous area with respect to iris.
Separability: both limbus and pupil contour represent a boundary region with a pronounced step from a darker to a lighter zone.

To calculate the separability score, for each candidate circle a slightly smaller and a little larger circles are considered: let CL be a circle with radius r and let CL_1 and CL_2 be two additional circles having the same centre as CL and radii $0.9r$ and $1.1r$, respectively. For each point in CL having polar coordinates (ρ, θ), two pixels, p_1 and p_2, located at the same angle θ on CL_1 and CL_2 are considered. The score for CL is computed as the sum of the differences between each pair of pixels corresponding to p_1 and p_2. The circle with highest

homogeneity and separability scores is considered as the circular shape which better approximates the pupil.

4.1.3 Linearisation

In order to perform limbus location, the image is first transformed from Cartesian coordinates to polar coordinates. The pixel with the greatest distance ρmax from the centre of the identified pupil circle is selected as the starting point for the image transformation. See figure 4.1.

4.1.4 Limbus location

A median filter is applied to the polar image. For each column corresponding to a position θ_i on the horizontal axis, ranging over ρ_j on the vertical axis, the following weighted difference is calculated:

$$\Delta(\rho_j, \theta i) = \phi(\dot{I}, \rho_j, \theta_i) \cdot (\dot{I}(\rho_j + \delta, \theta_i) - \dot{I}(\rho_j - \delta, \theta_i))$$

where $\dot{I}$ is the image in polar coordinates, and

$$\phi(\dot{I}, \rho_j, \theta_i) = \begin{cases} 1 & \text{if } \dot{I}(\rho_j + \delta, \theta_i) - \dot{I}(\rho_j - \delta, \theta_i) > 0 \\ & \text{and } \min\left(\dot{I}(\rho_j - \delta, \theta_i), \dot{I}(\rho_j + \delta, \theta_i)\right) > \varepsilon_G \\ 0 & \text{otherwise} \end{cases}$$

This procedure allows to identify the points with the higher positive variation,which indicates the transition from a darker zone (the iris) to a lighter one (the sclera). The first inequality just selects points with a positive gradient, while the second one rules out those points between pupil and iris by prescribing that the darker pixel in the pair at hand must exceed a threshold $\varepsilon_G \in [0, 255]$ (here $\varepsilon_G = 50$). Points which maximize (2) for each column θ_i in $\dot{I}$ make up the limbus that we will call F. After this, it is possible to discard outliers by considering that, in the polar space, point of the limbus must lie approximately on a line, and therefore have an almost constant ρ component. To this aim,

points with a ρ_i producing a relative error above a threshold ε are cancelled (here $\varepsilon = 0.4$). The relative error err is calculated as follows:

$$err = \frac{|\rho_i - \rho_{med}|}{\max_i |\rho_i - \rho_{med}|}$$

where ρ_{med} is the median value over F.

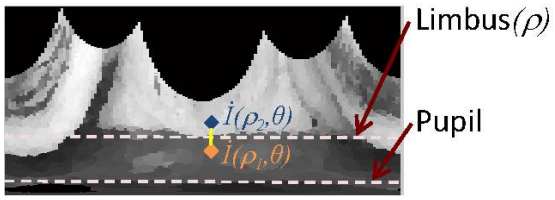

Figure 4.1: Illustration of ISIS algorithm.

At the end of acquisition and segmentation procedures, we obtain an image in polar coordinates where most useless information has been discarded.

4.2 BIRD

The iris segmentation method BIRD (watershed Based IRis Detection), has been presented in [2] along with a periocular area segmentation technique and a recognition approach based on the fusion of the iris and the periocular area. In this section only the iris segmentation method is presented, while the periocular segmentation and fusion approaches are presented in chapter 5 "Multi-biometric and multi-modal authentication on mobile devices" as one of the multi-biometric system proposed in this thesis.

BIRD[2] is a technique for smart mobile devices, which is the follow up of a technique presented in [1]. BIRD exploits the use of the watershed transform to identify more precisely the iris boundary and, hence, to obtain a more accurately computed code for iris recognition.

A positive feature of the watershed transform is that the contours delimit-

ing the regions into which an image is divided are mostly placed where human observers perceive them. In fact, the watershed transformation is a growing process performed generally on the gradient image, where the edges are enhanced. This feature should allow to correctly detect the limbus boundary. In turn, a negative feature is over-segmentation, i.e., the image may be partitioned into a number of parts that is remarkably larger than expected. Over-segmentation is particularly evident when all the regional minima in the gradient image are considered as seeds for the growing process. A common strategy to overcome this drawback is to adopt region merging and/or seed selection to reduce the number of watershed regions. However, in the case of eye images, processes for over-segmentation reduction cannot be stressed. Otherwise, some weak boundaries between sclera and limbus (light eye case) or between eyelashes and limbus (dark eye case) might be no longer present in the segmented image.

BIRD performs a binarization of the watershed transform to obtain an image where large portions of the limbus boundary are better enhanced. In this way BIRD is able to exploit the positive features of the watershed transform independently of over-segmentation problem. The boundaries of the foreground region are then given as input to a circle detection process, which aims at finding the circle that best approximates the limbus boundary (limbus circle).

4.2.1 Pre-processing

Uncontrolled iris acquisition may produce an image with local distortions due for example to shadows and different colour temperature. A colour/illumination correction is performed to reduce such local distortions, by processing separately the three RGB components of the eye image as grey level images. For each grey level image, a Gaussian filtered version is computed. A new image is built, where each pixel is set to the ratio between the value of the homologous pixels in the grey level image and in its filtered version. This ratio has the side effect to bring out the details in the image, so the kernel parameters of adopted Gaussian filter play a fundamental role. The parameters are the kernel size g_k, the average m_k and variance σ_k (they are mainly related to the resolution of the input image). In fact, a kernel too small excessively flattens the distribution of

the colours within the image, while one that is too large will not produce any substantial correction on lighting and colour distortions in it. In order to find a viable relationship between Gaussian kernel parameters to be adopted and the resolution of the image, was considered a set of pictures of irises at different resolutions $w_k \times h_k$ where $k = 1, 2, ..., n$, $w_k + 1 > w_k$ and $h_k + 1 > h_k$. The image resolution was represented by considering the value of the diagonal $d_k = \sqrt{w_k^2 + h_k^2}$. The optimal parameters for the Gaussian kernel were determined in terms of segmentation and recognition accuracy obtained on the set of images. It was observed that the relationship between g_k and d_k is quadratic, i.e. $g_k = \alpha_2 d_k^2 + \alpha_1 d_k + \alpha_0$, while $m_k = g_k$ and $\sigma_k = 0.1 \cdot g_k$. In this case, it was found that $\alpha_2 = -0.0001$, $\alpha_1 = 0.3064$ and $\alpha_0 = 11.1351$.

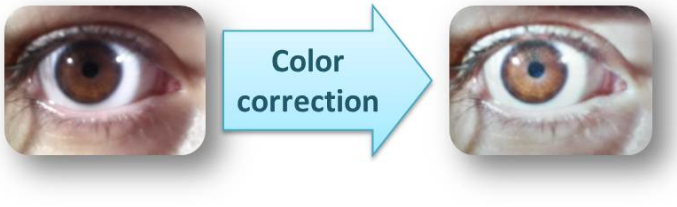

Figure 4.2: Colour correction.

A normalization process of pixel values is performed to map the values in the range $[0, 255]$. The combination of three obtained grey level images originates the colour/illumination corrected image. The left image in figure 4.2 shows the original image, while the right one corresponds to the colour/illumination corrected image. As BIRD is able to work even on low resolution images, it is possible to limit the computational cost of the method. The colour/illumination corrected image is resized by using a linear interpolation method without changing the aspect ratio, in order to get an image of the eye in the foreground with a horizontal resolution of 200 pixels (vertical resolution depends on aspect ratio). As previously, the process of correcting lighting/colour enhances the details in the image and these details are irrelevant for the segmentation. Thus a median filter is applied with a fixed-size window 7×7. The window size can be

regarded as fixed, because the image is first brought to a standard resolution by the resizing step. The resulting image I is shown in figure 4.3 (a).

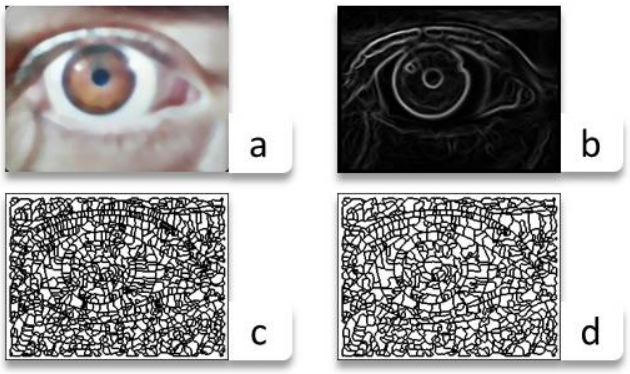

Figure 4.3: a) resized and smoothed image; b) gradient image; c) watershed transform; d) result of region merging.

4.2.2 Watershed transform and binarization

The effect of colour/illumination correction is to generate an almost uniform image, independently of the acquisition conditions that are highly uncontrolled. To extract the region of interest, i.e., the iris, the watershed transformation is used to partition the image into regions, based on gradient information. As already pointed out, over-segmentation is likely to affect the obtained partition. Thus, a successive process is necessary to merge adjacent regions, characterized by a certain homogeneity. Though the watershed transformation is computationally heavy, its use significantly reduces the processing time of the remaining steps, which will involve operations to be applied to a small number of regions rather than to their individual pixels. A colour quantization process is also performed to associate a unique colour to all pixels of the same region.

Watershed transform, region merging and colour quantization

The watershed transformation performs a partition of an image into a number of disjoint regions; each of them is characterized by a certain degree of homogeneity according to a specific uniformity criterion. The partitioning process relies on a region growing strategy, where starting seeds are chosen as the regional local minima of the gradient image.

BIRD computes the gradient image from the colour/illumination corrected image I. It first decomposes the colour image I in its three RGB components and then applies the 3×3 Sobel edge filter to each of them, separately. The final gradient image is obtained by averaging gradients computed on the three channels (figure 4.3 (b)). Then, the watershed transform W (figure 4.3 (c)) is obtained by adopting the topographical distance approach [49].

Due to the very large number of sub-regions in W produced by the watershed transformation, a merging process is mandatory to reduce significantly their number. To this aim, a representative colour is assigned to each watershed region and the merging criterion is based on the difference in representative colour of adjacent regions. Let R_i be a region of W and let $C_i(r_i, g_i, b_i)$ be the representative colour assigned to R_i, which is computed as the arithmetic mean of the colours of the pixels belonging to R_i. In the following, the representative colour of R_i is denoted as C_i. Two adjacent regions R_i and R_j are merged if it results:

$$d(C_i, C_j) < \delta$$

where $d(C_i, C_j)$ represents the Euclidean distance between C_i and C_j. Thanks to the process of correcting the distortions of colour and lighting applied during the preprocessing phase, the value for the threshold δ can be set permanently for all images and, in this application, it is experimentally set to 50. The result of merging is shown in figure 4.3 bottom right.

After merging, the representative colours of the so obtained watershed regions are updated originating a new version, Q, of the colour-quantized image (figure 4.4 (a)).

Binarization of watershed transform

It is worth to notice that, in general, the colour of pupil and iris are darker than that of sclera and eyelids in an eye image. More precisely, the former two show colours closer to black, while the latter a colour closer to white. BIRD strongly relies on this property to derive the binary image BW from the watershed transform. Indeed, all regions R_i whose representative colour C_i is closer to black could be tentatively ascribed to the foreground, while regions with a representative colour closer to white could be tentatively associated with the background. To this aim, information on the differences of the representative colours with respect to black and white should be properly taken into account to fix the threshold value able to cause the correct assignment of the regions to foreground or background.

In the RGB space, black is represented by (0, 0, 0) and white by (255, 255, 255), respectively. The Euclidean distances db_i and dw_i of all the representative colours C_i from black and white are then computed, as well as their arithmetic means, db and dw. Finally, the distance between black and white, dbw, is also computed.

In principle, if $db_i \leq d$b, R_i might be ascribed the foreground status. In turn, if $dw_i \leq dw$, R_i might be associated with the background. However, a decision on the status of R_i cannot be taken in the following two cases:

a) $db_i > db$ and $dw_i > dw$

b) $db_i \leq db$ and $dw_i \leq dw$

The former case occurs if it results $db + dw < dbw$, while the latter case when it results $db + dw \geq dbw$. Thus, there are cases in which R_i would remain unassigned (case a), or R_i might be ascribed both the foreground and background status (case b).

To overcome the above problems, a binarization threshold T is used whose value is set to $db \times dbw/(db + dw)$, where the ratio $dbw/(db + dw)$ is a multiplicative weight for the arithmetic mean db. Thus, any region R_i of W such that $db_i \leq T(dbi > T)$ is tentatively assigned to the foreground (background).

Let F and B be the sets of the regions tentatively ascribed the foreground status and the background status, respectively. Any region belonging to B is definitely considered as belonging to the background of BW, while regions belonging to F are further processed to take the final decision on their status. In particular, any region R_i belonging to F changes its status from foreground to background if it results:

$$d(C_i, c_B) \leq d(C_i, c_F)$$

where c_F and c_B, are the average foreground colour and the average background colour, respectively. The values c_F and c_B are computed as the arithmetic means of the colours associated the regions of F and B, respectively. The final binary image BW is shown in figure 4.4 (b).

4.2.3 Iris Detection

Circle fitting, see next paragraph for details, is accomplished only on smooth contour components of the foreground F of BW. The obtained circles undergo the voting procedure presented in section 4.1 ISIS, for the approximation of the limbus boundary, and then the procedure is applied again to locate the pupil boundary. Iris is then represented by the annular region enclosed by the limbus and the pupil circle.

Circle detection

The circle fitting procedure [44] searches for circumferences fitting the contour of the foreground region in BW. As it can be observed in figure 4.4 (c), the limbus represents only a small portion of the whole foreground. Thus, before performing the circle fitting procedure, a split operator is applied to the whole contour to partition it in smaller components. It is worth to notice that limbus boundary has a very smooth circular shape, whose end points are characterized by a very high curvature. It comes out that curvature analysis may represent a suitable tool to detect end points of the limbus boundary, so as to isolate it from the whole contour just by breaking it at those points. In other words, the whole

contour is divided into parts having a very smooth curvature (yellow parts in figure 4.4 (c)), which are delimited by parts characterized by strong curvature changes (blue parts in figure 4.4 (c)).

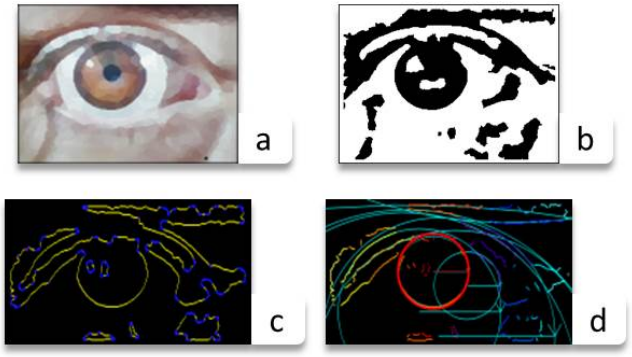

Figure 4.4: a) colour quantized image; b) binarized image; c) foreground contours; d) best fitting circle (red).

To estimate the curvature of a contour component CC, the sequence $S = p_1, p_2, ..., p_{|S|}$ of pixels of CC, detected by contour tracing, is considered. For each pixel p_i of S, a pixel p_{i+t}, with $t = 4 \times \lfloor log_2(|S|) \rfloor$ is selected. Let p_k be the contour point midway along the contour arc delimited by p_i and p_{i+t}, and let p_m be the mid-point of the straight line segment joining p_i and p_{i+t}. The curvature at p_i is estimated by dividing the distance $d(p_k, p_m)$ by the greatest of all distances computed for the pixels of CC. In this way the curvature values are normalized in the range $[0, 1]$, so that the points p_i, whose curvature value is greater than 0.5, are considered to be points of separation of two smooth contour components of CC.

For circle fitting, we follow the strategy described in [44] taking into account only contour components representing smooth curves and whose dimension is at least 5% of the whole contour. Many circles are generated, but only those included for at least 80% in the image are taken into account for the selection of the best fitting circle.

The best fitting circle is represented by the circle with the maximal separability score (the limbus approximation), shown in red in figure 4.4 (d).

Limbus boundary refinement

The best fitting circle might not coincide completely with the limbus, because it has been generated starting from only a part of the limbus boundary. Moreover, the limbus is not always characterized by circular shape. Thus, the selected circle might include parts external with respect to the limbus (sclera, eyelashes and eyelids).

To correctly identify the pixels actually belonging to iris and pupil, all the regions of W at least partially overlapping the circle (shown in purple in figure 4.5 (a)) are again taken into account to achieve a new segmentation BW' of W, where the foreground of BW' will include only iris and pupil. The remaining regions of W will be no more analysed and assume the background status in BW'. The regions of W overlapping the limbus circle are divided in two different subsets RT and RP. RT is the set of regions of W totally overlapping the circle (red and blue regions in figure 4.5 (b)), while RP is the set of regions of W only partially overlapping the circle (green regions in figure 4.5 (b)). Note that not all the regions of RT and RP belong to the foreground of BW; for example the blue regions and some green regions in figure 4.5 (b) belong to the background of BW.

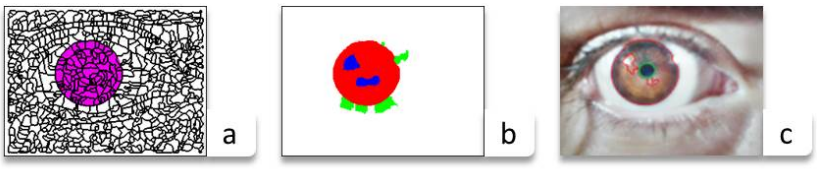

Figure 4.5: a) circle detected by circle fitting (purple) superimposed on the watershed transform; b) watershed regions totally overlapping the circle (red and blue) and partially overlapping the circle (green); c) red and green curves denote limbus and pupil boundaries.

To compute BW', the following features of the regions of RT and RP are taken into account: status in BW, representative colour, and (total or partial) degree of overlapping with the circle.

Let H be the set of pixels within the circle and belonging to foreground of BW. Let K be the set of pixels that are outside the circle and have been assigned to the background in BW. Let c_{limbus} and c_{back} be the arithmetic means of the colours that the pixels of H and of K assume in the quantized image Q, respectively. The values c_{limbus} and c_{back} are taken as representing the average colours of limbus (iris and pupil) and of the background, respectively.

The first step of segmentation implies the analysis of the regions belonging to RT. In particular, any region R_i belonging to RT is ascribed to the foreground of BW' if at least one of the following conditions holds:

1) $d(C_i, c_{back}) > d(C_i, c_{limbus})$;

2) R_i belongs to the foreground in BW, and at least one of its adjacent regions belongs to RT, while no adjacent region belongs to RP.

Any region R_i belonging to RT and that is not yet been ascribed a status in BW', is analysed again. In particular, R_i is ascribed to the foreground of BW' if the following condition is satisfied:

3) It exists at least one adjacent region R_j assigned to the foreground of BW' such that $d(C_i, c_{back}) \geq d(C_i, C_j)$;

Otherwise, R_i assumes the status of background in BW'.

The second step of segmentation involves a change of W in correspondence of the regions belonging to RP. Any region R_i belonging to RP is divided into two sub-regions, respectively including pixels within the circle, and pixels outside the circle. The former sub-region replaces R_i in RP and the latter sub-region is ascribed to the background of BW'. Since W has been modified, the quantized image Q and BW are updated, in order to update also the values c_{limbus} and c_{back}. Any region R_i belonging to RP is assigned to the foreground if at least one of its adjacent regions has already been assigned to the foreground

in BW', and the above condition 1) or condition 3) holds. Otherwise, the region R_i is assigned to the background in BW'. The boundary of the foreground in BW' is shown in figure 4.5 (c) as a red curve superimposed on I.

In figure 4.6, some examples of the performance of BIRD are shown, where red and green curves delimit the limbus boundary and the pupil boundary, respectively.

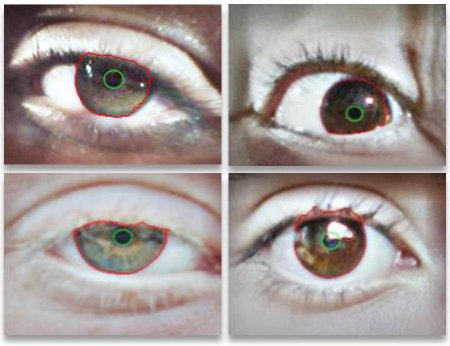

Figure 4.6: Irises segmented by BIRD.

4.2.4 Experimental results

The experiments have been conducted on the *MICHE Iris* dataset (see section 3.2.1). BIRD has been tested on a subset of *MICHE Iris* (we refer to as MICHEsub), which includes the first two images per subject in all 10 acquisition conditions, resulting in a total of 1500 iris snapshots. A manual segmentation of the whole MICHEsub dataset has been performed, in order to create a ground truth, which has been used for comparing **BIRD** performances with those of two state of the art methods, which are **ISIS** [31] (see section 4.1 for details) and our re-implementation of that proposed by Tan *et al.* in [52] that will be referred as **NICE-I** from now on. Experiments have been carried out to assess performances in terms of segmentation precision.

Tan et al. technique

In this section, the iris segmentation technique presented by Tan *et al.* in 2010, and winning algorithm in NICE I contest, is presented. Hereinafter this technique will be referred as NICE-I: "After reflection removal, a clustering based coarse iris localization scheme is first performed to extract a rough position of the iris, as well as to identify non-iris regions such as eyelashes and eyebrows. A novel integrodifferential constellation is then constructed for the localization of pupillary and limbic boundaries, which not only accelerates the traditional integrodifferential operator but also enhances its global convergence. After that, a curvature model and a prediction model are learned to deal with eyelids and eyelashes, respectively" [52].

In the tables presented below, the following notation has been adopted:

- **out** - outliers;
- **nout** - not outliers;
- **Iris CX** - iris centre X coordinate;
- **Iris CY** - iris centre Y coordinate;
- **Iris Rad.** - iris radius;
- **Pupil CX** - pupil centre X coordinate;
- **Pupil CY** - pupil centre Y coordinate;
- **Pupil Rad.** - pupil radius;

To evaluate the segmentation precision, the segmented iris masks provided by the automatic methods ISIS, NICE-I and BIRD are compared with those provided by the manual segmentation into the ground truth, in terms of percentage of errors measured in pixels. In more details, circle fitting is applied to the binary masks in order to approximate the centre and radius of both iris and pupil. The discrepancy in pixels is then computed between homologous

Device	Method		Iris CX	Iris CY	Iris Rad.	Pupil CX	Pupil CY	Pupil Rad.
IP5	ISIS	out	3.16	5.05	5.09	3.21	4.47	2.81
	out 33%	nout	0.62	1.77	0.79	0.70	1.24	0.67
	NICE-I	out	1.033	3.11	1.89	1.26	2.07	1.03
	out 10%	nout	0.32	2.77	0.98	0.55	1.46	0.57
	BIRD	out	3.30	4.82	2.76	3.54	4.66	1.14
	out 12%	nout	0.61	1.91	1.21	1.50	1.56	0.87
GS4	ISIS	out	4.51	5.33	6.26	4.61	5.49	3.57
	out 13%	nout	0.92	2.06	1.28	0.75	1.37	0.75
	NICE-I	out	1.27	3.18	2.03	1.51	2.26	1.13
	out 33%	nout	0.33	2.61	0.93	0.62	1.53	0.65
	BIRD	out	4.09	4.93	2.79	4.32	4.87	1.16
	out 36%	nout	2.52	2.85	2.14	1.48	2.62	0.94
GT2	ISIS	out	5.91	7.57	10.76	6.02	7.39	5.94
	out 54%	nout	3.28	5.69	5.50	3.46	5.20	2.47
	NICE-I	out	1.60	3.67	2.71	1.79	3.47	1.36
	out 21%	nout	0.36	2.57	1.12	0.61	2.00	0.66
	BIRD	out	3.54	4.16	2.55	4.04	4.36	1.24
	out 34%	nout	2.86	2.70	2.22	2.19	2.60	0.97

Figure 4.7: Segmentation accuracy measured in terms of percentage of error with respect to manual segmentation for the three tested approaches (ISIS, NICE-I, and BIRD) on probe/gallery images without colour/illumination correction.

Device	Method		Iris CX	Iris CY	Iris Rad.	Pupil CX	Pupil CY	Pupil Rad.
IP5	ISIS	out	2.85	3.98	4.31	3.03	3.79	2.04
	out 33%	nout	0.62	1.75	0.88	0.71	1.43	0.70
	NICE-I	out	4.84	6.73	10.93	4.84	6.99	5.49
	out 46%	nout	3.27	4.11	2.51	3.39	3.34	1.07
	BIRD	out	3.33	3.69	1.43	3.51	3.46	0.82
	out 25%	nout	0.43	1.58	0.54	0.58	1.15	0.59
GS4	ISIS	out	4.12	5.00	4.39	4.19	4.77	1.92
	out 32%	nout	0.62	1.87	0.83	0.77	1.35	0.63
	NICE-I	out	5.89	6.69	12.78	5.76	6.71	6.46
	out 50%	nout	3.23	5.22	2.74	3.15	4.66	1.27
	BIRD	out	3.58	4.07	1.68	3.70	4.01	0.81
	out 30%	nout	0.46	1.69	0.61	0.68	1.61	0.58
GT2	ISIS	out	6.03	7.89	9.46	6.13	7.92	5.23
	out 54%	nout	2.67	4.67	3.69	2.96	4.59	1.81
	NICE-I	out	7,45	6.90	7.86	7.50	7.62	3.57
	out 39%	nout	2.90	4.60	2.40	3.04	3.97	0.93
	BIRD	out	3.25	4.18	2.28	4.40	6.30	1.09
	out 32%	nout	1.05	2.06	0.92	1.30	2.17	0.83

Figure 4.8: Segmentation accuracy measured in terms of percentage of error with respect to manual segmentation for the three tested approaches (ISIS, NICE-I, and BIRD) on probe/gallery images with colour/illumination correction.

parameters (centres and radii) extracted from each ground truth mask and the corresponding ones obtained by the testing approaches. A global estimation is obtained by averaging the discrepancy values over all images, after they have normalized with respect to the image resolution and multiplied by 100.

There are cases in which the detection is completely wrong, so we can consider them as outliers when computing the global segmentation precision. Outliers can be detected as they generally provide a segmentation error higher than a threshold ϵ_{out}, which is computed as $\epsilon_{out} = 6 \cdot md$, where md is the median value computed over all the errors in the subset of tested images. Since the process of colour/lighting correction plays an important role for the subsequent stages of segmentation, precision was evaluated in two cases, in which it is included or not in the process of preprocessing. Table 4.7 and Table 4.8 show the results obtained in terms of accuracy of the three methods of segmentation (ISIS, NICE-I and BIRD) with three different devices (IP5, GS4 and GT2), excluding or including the colour/lighting correction process from the pipeline. In both Table 4.7 and Table 4.8, we provide results accounting for both cases, in which outliers have been considered or not in the evaluation. It comes out from these results that colour correction improves the segmentation accuracy of both ISIS and BIRD almost always. There are only few cases in which a very low decrease in precision is registered after colour correction of the input image. In particular this mostly happens for ISIS on images acquired by GT2, and can be ascribed to the lower resolution of the camera mounted on this device. Indeed, on low resolution images after colour correction ISIS undergoes more difficulties to correctly locate the pupil, so undermining all the subsequent steps. On the contrary, colour correction seems to negatively affect the performance of NICE-I. This can be explained by considering that the first step in the NICE-I pipeline consists in a coarse clustering of pixels with low brightness to be considered as iris candidates. The colour/illumination correction produces significant changes in the histogram of the iris image, so that clustering fails even jeopardizing subsequent steps in the remaining pipeline. Indeed, by inspecting images for which NICE-I fails, it comes out that the method do not output any circle neither for the iris or for the pupil. This is also the motivation for which the number of

outliers produced by NICE-I increases when colour/illumination correction is included.

It is worth to notice that the amount of outliers produced by a testing approach is in itself a measure of its effectiveness. From results in Table 4.8, we observe that BIRD always outperforms ISIS in terms of both segmentation precision and number of outliers. From both tables, it is also noted that in most cases, when outliers are kept out of the computation, the increment in segmentation precision observed for BIRD is higher than that obtained by ISIS. However, BIRD produces more outliers than NICE-I. This happens because some images are not correctly segmented by the watershed, so that the iris boundary presents some cuts, which mislead the curvature analysis process and result in a mislocated iris/pupil circle. Furthermore, BIRD provides a higher accuracy in locating the iris than the pupil; this is because pupil is smaller than iris and then it is more affected by noise factors like bad illumination and specular reflection. This can also explain why BIRD outperforms ISIS. Indeed, BIRD first locates iris and then searches for pupil by restricting the region of interest (for the pupil) to the area included into the iris circle. On the contrary, ISIS performs in the real opposite way by first searching for pupil and then extending its searching window to neighbouring regions. This jeopardizes the performances of ISIS, since for several images into the dataset the pupil is not completely visible, due to large occluded areas or severe out of focus conditions, which are caused by user holding the mobile device in his/her hands during the acquisition.

Results in Table 4.7 and Table 4.8 show that NICE-I provide better performance when colour/illumination correction is not included in the segmentation process. In this case, it also outperforms BIRD in terms of segmentation accuracy. Conversely, colour/illumination significantly improves the performance of BIRD, so it must be considered as a fundamental part of the segmentation process implemented by BIRD. Indeed, by comparing results obtained by BIRD in Table 4.8 (its best performance) with those of NICE-I in Table 4.7 (its best performance) it is worth to notice that in most cases BIRD outperforms NICE-I.

Another important aspect in the evaluation of the performance of BIRD is the execution time given by the sum of the time needed to carry out the various stages of processing. In order to provide an estimate of the impact that each one has on the total time, the time necessary to accomplish each stage is reported as a percentage of the weight that it has on the entire running time:

i) colour/illumination correction (20%);

ii) image resizing (4%);

iii) median filtering (1%);

iv) watershed (11%);

v) merging (6%);

vi) binarization (2%);

vii) iris detection (40%);

viii) pupil detection (16%).

The tests were carried out on a 64-bit system with Genuine Intel®processor U7300 1.3GHz, with 4GB of RAM and the total time was about 1.2s, which is comparable with that required by ISIS while it is of an order of magnitude smaller than that required by NICE-I.

4.3 Conclusive remarks

BIRD, that can be seen as an evolution of ISIS, is principally devised for mobile devices, as it is designed to be robust with respect to many kinds of image distortions that typically occur when the user is holding a smartphone/tablet in his/her hands to perform acquisition (out of focus, occlusions, motion blur, non-uniform illumination, off-axis angles). BIRD exploits the watershed transform at two different levels. First it applies the watershed transform to binarize the image and then it further analyse the sub-regions produced during the first

application of the watershed in order to refine their classification as either foreground or background.

With respect to ISIS algorithm, BIRD produces less outliers and detects more refined pupil and limbus boundaries, aspects that assume an important role in the following phases of an iris recognition system.

The results demonstrates that, in particular when the colour/illumination correction is performed, BIRD outperforms also its main competitor, NICE-I. In addition, its execution time is much lower than the one of NICE-I. This is another important aspect, since the limited computational power of the mobile devices, requires a good trade-off between accuracy of the segmentation and the execution time.

In the following chapter, it is shown how ISIS and BIRD are employed in the iris recognition systems presented.

Chapter 5

Multi-biometric and multi-modal authentication on mobile devices

To improve biometric recognition robustness against attacks it is worth combining it with another authentication item (see paragraph 1.1 "Authentication overview" for reference): something the user knows (e.g., password, personal identification number (PIN), secret answer, pattern); something the user has (e.g., smart card, ID card, security token, software token, phone, or cell phone); something the user is or does (biometric trait). In the following chapters two main strategies are presented:

- Multi-biometric recognition. Iris recognition has been combined with another biometric trait: the periocular area, and the face;

- Multi-modal recognition. Iris recognition has been combined with "something that the user has": the smartphone;

5.1 Combining iris and periocular area

In this section we present a biometric recognition method based on the fusion of iris and periocular area biometric traits. The results of this study are presented in the second part of the article "BIRD: Watershed Based IRis Detection for mobile devices"[2].

The information regarding position and size of the iris (its centre and radius) extracted by BIRD (see section 4.2 "BIRD" for reference) constitute the starting point for the delimitation of the periocular region. Recent studies [44] showed how the latter could be considered itself a biometry. Once the periocular region is selected a transformation from Cartesian to polar coordinates is applied on it. In this way, it is possible to apply to the periocular area a process of extracting and matching characteristics similar to that used for the iris. Iris and the periocular region are then fused at score level through a simple sum criterion in order to increase the accuracy of the recognition system.

5.1.1 Periocular region segmentation

It has been demonstrated that features detected in the periocular region can be used as a soft-biometric to authenticate people [48] and significantly improve the recognition accuracy of iris recognition systems when dealing with iris images acquired in visible light [50],[47].

Starting from the approximating circle detected during the iris segmentation process, the centre coordinates $C_{iris}(x_c, y_c)$ and radius r of the circle are used to construct two concentric ellipses E_1 and E_2 that enclose part of the area around the iris. Both E_1 and E_2 are centred in C_{iris}, but they are characterized by different parameters (major and minor axes, a and b), which are respectively $(a, b) = (2 \times r, r)$ for E_1 and $(a, b) = (3 \times r, 2.5 \times r)$ for E_2 (see figure 5.1).

The area enclosed by E_1 and E_2 is mapped to a rectangular region $I_{periocular}$ by adopting a procedure that is similar to the Daugman's rubber sheet model [11]. For each angle θ (ranging between 0 and 2π), two homologous points $P_1(\theta)$ and $P_2(\theta)$ are considered on E_1 and E_2 respectively, and pixels lying on the line ρ joining these points are mapped on the θ^{th} column of $I_{periocular}$. The

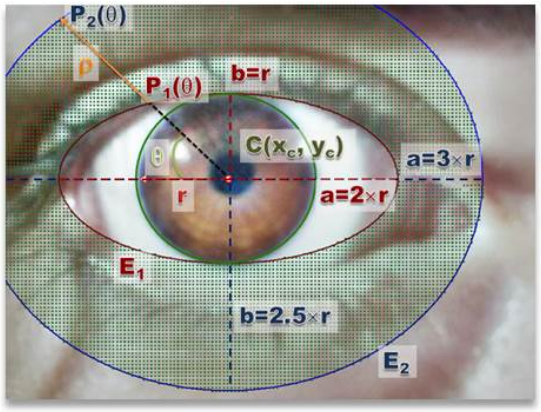

Figure 5.1: Periocular segmentation.

rectangular region resulting from this remapping procedure shows a resolution (256×32 pixel in this case), which depends on the granularity (discretization) chosen for θ and ρ (see figure 5.2).

Figure 5.2: Cartesian to polar coordinates mapping of the periocular area.

5.1.2 Iris and periocular area feature extraction

Two kinds of features from the reshaped regions of the iris and the periocular area are extracted, which characterize both the iris/periocular area colour and texture. First, it performs a colour conversion from the RGB colour space to HSV, with the aim of separating colour information from the one regarding luminance, which are then treated separately. Indeed, the cumulative sums

(CSUM) [51] technique (see the following section 5.1.2 "CSUM algorithm" for details) is applied to create an iris code for the structural pattern contained into the luminance channel, while a histogram based approach is used for colour channels. As there are two colour channels, they are combined through a pixel by pixel product, so to obtain a single channel. A 64-bits histogram is computed on this channel and it is used as a colour feature vector. The distance between two colour vectors is computed by using the cosine dissimilarity, while textural feature vectors produced by the CSUM approach are compared by means of the Hamming distance.

In order to assess the contribution of the periocular region to the recognition task, we considered two different scenarios. In the first one, only iris features are exploited to match two irises, while in the second one it fuses both iris and periocular biometrics at a score level, by implementing a simple sum approach.

CSUM algorithm

The feature extraction phase has been implemented by the CSUM (Cumulative SUMs) algorithm proposed in [51]. The method is based on the analysis of local variations of grey levels in the image. It is simple to implement, and does not require high computational costs. Last but absolutely not least, it is robust to under-controlled iris acquisition conditions. These features make it appropriate for mobile processing, therefore it was chosen for FIRME. The algorithm is applied to the polar image obtained in the previous phase, and entails the following steps:

- Step 1. Division of normalized iris image into basic cell regions for calculating cumulative sums. A cell region is $3(row) \times 10(col)$ pixels; the mean grey value X is used as a representative value of a basic cell region for the following calculation.

- Step 2. Horizontal and vertical grouping of basic cell regions (in our case the group size is of 4 cells); the mean value X over cell representative values is calculated for each group.

- Step 3. Computation of cumulative sums over each group according to the following :

$$S_0 = 0$$

$$S_i = S_{i-1} + (X_i - \bar{X}) \text{for i} = 1, 2, 3, 4$$

- Step 4. Generation of iris feature codes.

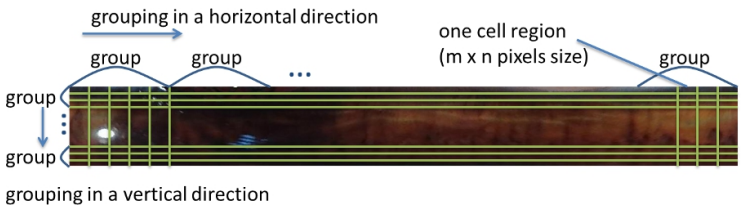

Figure 5.3: Cumulative Sums algorithm illustration.

The iris code is obtained by comparing, for each group, two consecutive cumulative sums. Values 1 or 2 are assigned to a cell if the value S_i contributes respectively to an upward slope or to a downward slope. Otherwise, value 0 is assigned to S_i. An illustration of an iris image subdivides in cells and groups is given in figure 5.3. Matching of iris codes is computed by Hamming distance.

5.1.3 Experimental results

Several experiments have been conducted to assess the recognition performances. We also tested the influence of the iris segmentation method by testing three different approaches: ISIS, NICE-I and BIRD. Hereinafter the recognition approach using BIRD as iris segmentation technique will be simply indicated as "BIRD". In each test, two images per subject have been considered to build a gallery set (used to enrol subjects) and a probe set (used to test the recognition system). System performance was measured in terms of Decidability, Equal Error Rate (EER) and Area Under Curve (AUC), the former being calculated

according to the definition given by Proenca and Alexandre in [44] (see section 1.4.2 "NICE" for more details about the Decidability value computation).

In the tables presented below, the following notation has been adopted:

- **I** - iris recognition;
- **P** - periocular recognition;
- **F** - fusion of iris and periocular recognition;
- **Dec.** - decidability;
- **EER** - Equal Error Rate;
- **AUC** - Area Under ROC Curve;

Both EER and AUC are single descriptions related to the Receiving Operating Characteristic Curve (ROC), which is a graphical plot that characterizes the performance of a biometric verification system as its discrimination threshold is varied. The EER represents the value where False Accept Rate and False Reject Rate equal and can be considered such as a steady state point for the system. While ROC is a two-dimensional representation of a model's performance, the AUC distils this information into a single scalar and represents the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one.

In the first experiment, both gallery and probe images have been acquired with the same mobile device using the rear camera. The results were produced for iris recognition, for periocular region recognition and for merging of both. In Table 5.4, we report the numeric values for the various performance measures.

From the obtained results, it comes out that the segmentation approach implemented by BIRD induces a higher performance than ISIS in all cases. The reason underlying the larger increase in recognition accuracy produced by BIRD is that the precision it offers in segmenting the iris boundary (limbus) is generally higher than that provided by ISIS. Indeed, ISIS only approximates both iris and pupil boundaries by circles, while circle fitting represents only a preliminary step

Method		Device								
		IP5			GS4			GT2		
		Dec.	EER	AUC	Dec.	EER	AUC	Dec.	EER	AUC
ISIS	I	0.562	0.393	0.651	0.938	0.335	0.742	0.692	0.388	0.664
	P	0.626	0.414	0.637	0.894	0.349	0.707	0.681	0.402	0.671
	F	0.634	0.446	0.632	1.039	0.336	0.729	0.765	0.378	0.673
NICE-I	I	0.787	0.386	0.6674	0.869	0.409	0.714	0.672	0.413	0.681
	P	0.366	0.470	0.581	1.087	0.334	0.753	0.829	0.334	0.715
	F	0.366	0.470	0.581	1.086	0.334	0.753	0.829	0.334	0.715
BIRD	I	1.151	0.266	0.784	1.083	0.314	0.778	0.880	0.326	0.707
	P	0.949	0.295	0.736	1.340	0.278	0.798	1.159	0.342	0.789
	F	1.156	0.260	0.764	1.290	0.322	0.775	1.103	0.277	0.765

Figure 5.4: Recognition accuracy measured in terms of decidability, EER and AUC for the three tested approaches (ISIS, NICE-I and BIRD) on probe/gallery images acquired by the same mobile device.

for the segmentation procedure implemented by BIRD. Moreover, BIRD further refines the representation of the limbus by considering sub-regions, which cross the iris circle and re-assigning them to the foreground/background depending on the classification of neighbouring regions. Since results provided by BIRD outperform those obtained by ISIS, only the former has been considered in the following experiments, whose main goal is to assess also the robustness of BIRD with respect to adverse acquisition conditions as well as its interoperability among different mobile devices. It can be observed that, in many cases, the results provided by iris and periocular area are comparable, with regard to the other measures of performance. Moreover, it can be observed that the fusion not always leads to an improvement. This result can be explained by considering that a simple sum as been adopted for the score fusion step. However, resorting to a more refined approach [53] or postponing the fusion process to the decision level [54] could lead to larger improvements.

In the second experiment, the probe and gallery sets come from different devices, which show significant variation in image resolution, as they are equipped with very different sensors. Results when only the iris region is considered are reported in Table 5.5, while those obtained including also periocular information are reported in Table 5.6.

Method		Gallery								
		IP5			GS4			GT2		
		Dec.	EER	AUC	Dec.	EER	AUC	Dec.	EER	AUC
Probe	IP5	1.151	0.266	0.784	0.834	0.333	0.745	0.016	0.485	0.525
	SG4	0.255	0.414	0.601	1.083	0.314	0.778	0.887	0.729	0.248
	SGT	0.213	0.534	0.451	0.317	0.430	0.593	0.880	0.326	0.707

Figure 5.5: Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach (without periocular information) on probe/gallery images acquired by different mobile devices.

Method		Gallery								
		IP5			GS4			GT2		
		Dec.	EER	AUC	Dec.	EER	AUC	Dec.	EER	AUC
Probe	IP5	1.156	0.260	0.764	0.858	0.309	0.738	0.374	0.397	0.614
	SG4	0.375	0.397	0.630	1.290	0.322	0.775	0.296	0.565	0.406
	SGT	0.124	0.584	0.466	0.502	0.367	0.642	1.103	0.277	0.765

Figure 5.6: Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach (with periocular information) on probe/gallery images acquired by different mobile devices.

Method		Device								
		IP5			GS4			GT2		
		Dec.	EER	AUC	Dec.	EER	AUC	Dec.	EER	AUC
BIRD	I	0.529	0.401	0.652	0.583	0.386	0.665	0.078	0.472	0.535
	P	0.246	0.470	0.581	0.633	0.373	0.672	0.596	0.391	0.658
	F	0.506	0.399	0.646	0.647	0.367	0.689	0.432	0.430	0.619

Figure 5.7: Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach on probe/gallery images captured in different acquisition settings (outdoor/indoor).

From results in Table 5.5 and Table 5.6, it emerges that testing on cross-datasets is particularly difficult when images captured with an IP5 mobile device are used as probe. This can be explained by visually inspecting images acquired by different devices. Indeed, before the lighting/colour correction, IP5 iris images show a higher sharpness and brightness than those acquired by both GS4 and GT2, independently of the used camera (either frontal or rear). Even so, BIRD achieves better accuracy than ISIS further confirming that a higher precision in segmenting iris images greatly affects the recognition process.

The last experiment is devoted to assess performances of both testing approaches with respect to the capturing settings. In particular, images acquired indoor constitute the gallery, while those captured outdoor (probe images) are used to test the methods. This scenario is more plausible than the opposite one, since one can assume that the enrolment of a person is rarely performed (may be just once and in very controlled conditions), while testing occurs more often either in controlled or adverse conditions. Numerical results in terms of decidability, EER and AUC are reported in Table 5.7. As expected, in this more challenging scenario, the performances drop, but we can notice that the trend is confirmed since GS4 performances outperform those of the other devices and in particular GS4 obtains better performances that IP5 when considering iris recognition only, with respect to the performances presented in table 5.4.

When only iris features are used to recognize people, there is no appreciable gap in performance of BIRD when using different device. However, when periocular features are also integrated, the accuracy obtained with IP5 remains almost the same, while that achieved with GS4 and GT2 slightly improves.

5.2 Combining Iris and Face

In this section iris recognition is combined with face recognition to provide a robust authentication system, namely Face and Iris Recognition for Mobile Engagement (FIRME)[10], on a mobile device. Face anti-spoofing, continuous re-identification and best template selection are some of the important features that FIRME provides. The architecture of FIRME is illustrated in figure 5.8.

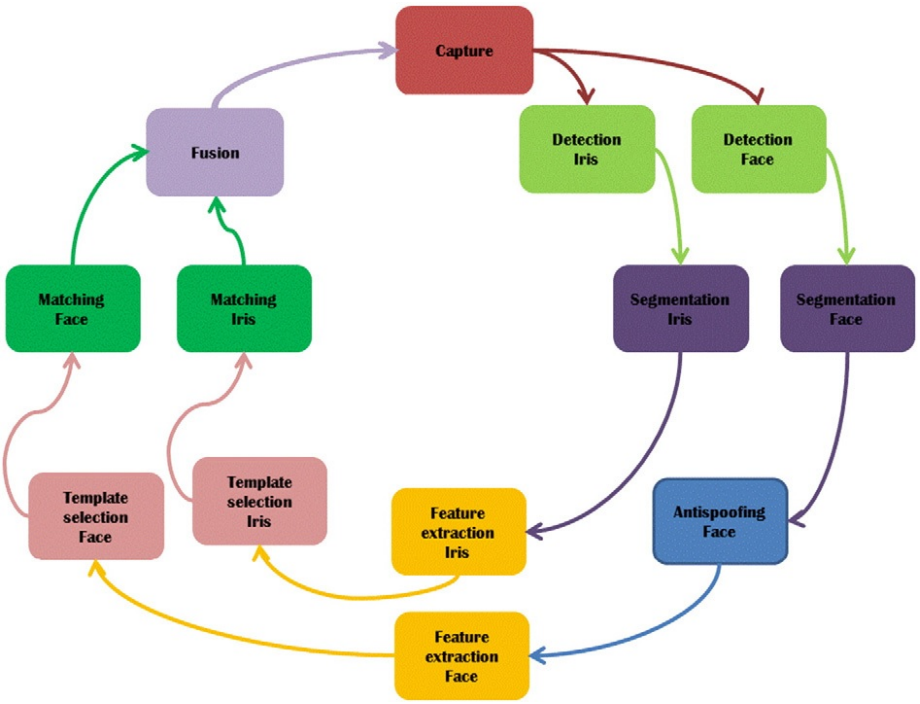

Figure 5.8: Architecture of FIRME [10].

5.2.1 Face recognition

Acquisition and segmentation

Nowadays, any mobile OS provides both routines and drivers to access an on-board device camera. Using them, it is quite easy to acquire and store face images, possibly coming from a continuous clip. On the other hand, specialized routines must be implemented to handle the following steps. In particular, a robust face segmentation and possible correction is critical both because it must provide the best starting point for feature extraction, and because it can provide crucial information for spoofing detection. The optimization of all procedures is of paramount importance for the efficiency and also for the usability of the application. It is therefore necessary to optimize phase wise implementation in terms of both CPU and memory.

Figure 5.9: Localization of reference points.

The face detection module presently included in FIRME also locates a number of reference points used to determine if it presents a pseudo-frontal pose (see figure 5.9). The module first implements the Viola-Jones algorithm [66] for the localization of the relevant subregion of the image containing the face. Since the user is usually looking at the device screen during such operation,we can reasonably assume a frontal pose.

Therefore, we can avoid computationally heavy pose normalization techniques. Those include the ones based on Active Shape Model (see for example [32]),which provides a good compromise between accuracy and computational weight, but is still too demanding for an average mobile device. However, a fast sample selection still seems appropriate.

To this aim, once the face region has been cropped, Viola-Jones algorithm is reused, with obviously different parameter configurations, to locate the eyes and mouth: their positions and distances allow to discard those samples which would require pose correction before proceeding with the recognition. An image correction routine is used instead of normalizing illumination. FIRME implements a Self Quotient Image (SQI) algorithm [74]. SQI relies on the Lambertian reflectance model of an ideal diffusely reflecting surface. It states that the image of a 3D object (and therefore its 2D projection) is subject to intrinsic factors, namely surface normal and albedo, and the extrinsic factor of illumination intensity. Differently from the quotient image [73] technique, which follows the same model though without addressing shadows, SQI uses two different versions of the same image (self), therefore no alignment process is needed. Moreover it is robust to shadows and to different kinds of light source, does not need training to estimate illumination direction, and does not assume that all objects (faces) have the same surface normal. This is a great advantage to the specific features of faces and of operating settings. Last but not least, SQI is also computationally much lighter than quotient image. The illumination-corrected image is obtained by applying a smoothing filter.

Then the ratio between the original image and its smoothed version is computed, providing a light-invariant representation of the face image. In practice, the value of each image pixel is divided by the mean of the values in its neighbourhood, represented by a square mask of size $k \times k$ (in our case $k = 8$). These selection/correction procedures also return quality indices, which allow to discard those biometric samples that are expected to poorly support if not hinder the next recognition steps.

Spoofing detection

Traditionally, applications requiring maximal security due to the critical nature of handled data, are also those which may attract most spoofing efforts. In the mobile scenario, this is for instance the case of mobile banking. In this context, accurate recognition may not be enough, and anti-spoofing countermeasures might be worth being adopted. To address these issues, Google has recently implemented a recognition system on Android which is extended with a liveliness detection routine, which exploits eye blink detection. It processes the entire part of the image containing the eyes, and the cost of this operation is affected by the image resolution and by the kind of analysis. To avoid a too demanding computation, FIRME exploits the technique of 3D geometric invariants, which is both simpler and more robust, to estimate the structure of the face [33]. The elective application of 3D geometric invariants is the classification of 3D objects from a single 2D view. Given a 2D image of a 3D object, the technique measures the distances between pre-defined landmarks (reference points) on the object. The distances between those points, of course, change with the capture viewpoint.

However, some relationships between them, under specific conditions, remain invariant to the viewpoint, and are therefore called geometric invariants. They are intrinsic to a certain object and can also be measured on a single image of it. The invariance conditions which may hold regarding, for instance, collinearity or co-planarity of the points under consideration. The anti-spoofing technique implemented in FIRME applies this principle in the reverse direction. It exploits a set of five points on the face, for which the co-planarity constraint is normally strongly violated, namely the outer corner of the right and left eyes, the extreme left and right of the face, and the nose tip. When the face is moved, the spoofing detection routine estimates the geometric invariant relative to the identified schema of points; if the invariant holds, the points comply with the constraint of co-planarity: since this should not be the case, this means that the captured face image must be a photo (spoofing); otherwise the points are not coplanar and the three-dimensionality of the face is guaranteed and thus the captured image corresponds to a real (live) user (figure 5.10). This technique calculates

a ratio of determinants of matrices (more details are available in [33]), therefore it is extremely straightforward and fast.

Figure 5.10: Detection of a real live user.

Best template selection

Among the factors that make up the effort required by the user of a biometric system, and the possibly deriving discomfort, a crucial role concerns the degree of control required during the acquisition phase; considering presently available systems, iris recognition is the one generally requiring maximum control, since the user must stay immobile looking at the camera during image capture; on the other hand, face recognition is often used also in systems, such as video surveillance ones, where the subject is most often unaware of image capturing. Choosing this pair of traits, we had to handle their different requirements. For this reason, we implemented a fast strategy for best template selection, transparent to the user and able to discard samples of low quality.

Especially for currently available mobile applications, this aspect is seldom considered or is only addressed indirectly. One approach to provide the user with a greater freedom would be to implement some robust method to correct image distortion caused by pose and/or movements (Google, and Facebook tagging system). In FIRME, the acquisition phase is implemented so that the camera

actually captures a high number of frames, aiming at maximizing the accuracy of the recognition. As a matter of fact, after a suitable choice procedure, only the best one will be used for the recognition process. FIRME implements a sample selection mechanism based on entropy. Of course, we avoid any further processing before such selection, to save computational resources. We also avoid any correction/normalization procedure for both pose which might be exploited in this step, as the one described in [32], to improve the quality of samples. In practice, this step occurs immediately after the operation of face localization (figure 5.11). The selection requires to calculate the correlation between all pairs of faces in the currently acquired sequence of frames. The obtained value of the correlation index (usually in the interval $[-1, 1]$) is normalized to the range $[0, 1]$. Given a pair of samples, this index can be interpreted as the probability that they conform to each other, and can be used in a way similar to [34]. The difference with the algorithm for best sample selection in [34], is the localized version of the correlation index that is used as similarity measure.

Figure 5.11: Best template selection.

This provides more accurate results, but is much more computationally expensive. FIRME uses this more complex version only for face matching, which will be discussed below. For this reason, though sharing a common basic technique, Best Sample Selection module and Face Recognition module are imple-

mented in two distinct procedures.

The innovation with respect to [34], is the implementation of different similarity measures according to the current goal. Returning to the presently described best sample selection, after a further appropriate normalization, the values sum up to 1. In this way, they can be treated as a probability distribution on the whole (small) set of samples under examination. This distribution can be used to calculate the entropy of that set. Afterwards, the selection module "discards" a sample at a time from the collection and recalculates the entropy of the remaining set; in addition, it calculates the difference between the entropy before and after a sample "elimination". This is done for all the samples in the sequence ("set") captured. That sample which, when extracted, produces the minimum difference, is permanently removed from the set, and the process iterates. The last sample left is selected as the best sample, and is submitted to the rest of the workflow towards the recognition phase.

Feature extraction and matching

The earliest techniques for face classification, such as Linear Discriminant Analysis [71], are too sensitive to image distortions to be reliably used in commercial applications, which imply partially or completely uncontrolled settings. Given two images I_1 and I_2 and their respective mean values $\bar{I_1}$ and $\bar{I_2}$ for pixel intensity, their spatial correlation is computed as:

$$s(I_1, I_2) = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (I_1(i,j) - \bar{I_1})(I_2(i,j) - \bar{I_2})}{\sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (I_1(i,j) - \bar{I_1})^2 \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (I_2(i,j) - \bar{I_2})^2}}$$

In FIRME recognition step, such correlation is adapted to work locally, on individual sub-regions r_1 and r_2 of the images I_1 and I_2 [32][34]. For each sub-region r_1 in I_1, the region r_2 that maximizes the correlation coefficient $s(r_1, r_2)$ is searched, extending this search in a narrow window around the corresponding position in I_2. The global correlation $S(I_1, I_2)$ between the two images I_1 and I_2 is the sum of these local maxima. As mentioned above, this approach

achieves better accuracy, but it is also more computationally expensive. This is the reason for using it in a very late phase. However, it is worth noticing that the pre-calculation of some quantities involved in the formulas, some code optimization, and reduced resolution allow to perform a considerable number of verification checks (in the order of tens) in less than a second even when using a mobile architecture. These pre-calculated factors extracted from the face are stored in what is called biokey. Differently from a verification operation, where the user claims an identity, an identification protocol requires that each image must be matched against all those stored in a certain gallery. For each registered identity g_k in the gallery $G(k = 1, ..., |G|)$, we assume to store one or more images I_j, $j > 0$. When a new query image q is submitted, FIRME compares it with all images in the system and computes the corresponding correlation indices. After that the final list of values is reorganized in decreasing order, FIRME returns the identity g_k with more images in the first n positions. If the gallery contains only one image per subject, FIRME returns the first retrieved identity. To be consistent with the following processing, the value of global spatial correlation, which is of course a similarity measure, is normalized in the range $[0, 1]$ and then transformed into a distance in the same range by subtracting it from 1.

5.2.2 Iris recognition

Acquisition and segmentation

As in face recognition, eye images acquisition and storage exploit routines as well as drivers provided by the mobile SO at hand. Eye capture procedure does not present special constraints, since the interaction pattern itself implicitly guides the user to maintain a frontal pose, so to avoid off-axis problems. The user has to get close to the camera till reaching a maximum distance of 10-12 cm. The used segmentation algorithm is ISIS (see section 4.1 for details). The feature extraction process has been implemented by the CSUM (Cumulative SUMs) algorithm, see section 5.1.2 "CSUM algorithm" for reference.

5.2.3 Score level fusion of face and iris biometrics

Mobile systems are a true challenge in terms of accuracy of the result from a matching module. This originates from two important factors which characterize mobile settings: a) the reduced resolution of images captured by on-board sensors equipping the mobile device; b) the high degree of freedom of the user in handling the device. These two factors interact so that a recognition system dealing with a single biometric trait, though robust, might not be sufficient to address the security requirements of a number of mobile applications. On the other hand, the fusion of multiple biometric traits is very often considered as one of the best, if not the only solution in those settings, where under-controlled acquisition conditions may negatively affect the performance of a unimodal biometric system at a significant extent [61]. FIRME adopts a multi-biometric fusion scheme, where two different biometric traits are integrated to improve the recognition performance of the system. When dealing with such kind of fusion, different aspects must be taken into account, among which fusion policy and fusion scheme play a crucial role. Fusion may happen in different steps of the recognition process. Some systems fuse the biometric features, so cutting the cost of the next steps and maximizing the amount of information available in the biometric signature. However, this is only possible when the features extracted from the different traits are compatible, i.e. can be expressed in a consistent form allowing to fuse them in a single vector. Alternatively, information from the different biometric systems may be fused immediately after matching, i.e., at score level. At present, this is considered as the best compromise between the complexity of the fusion process and the amount of information still available to guarantee a good result [56]. As a matter of fact, pushing fusion forward towards the decision step would make the fusion scheme even simpler (responses are of binary form and fusion rules of Boolean type), however at the expense of losing an excessive amount of information, which might be still effectively used for a more accurate result.

FIRME performs fusion of face and iris information at matching level. This allows to handle scalar values, the scores, which are easy to analyse but at the same time still retain rich information. A special attention has been devoted

to the fusion scheme, that in such an under-controlled system becomes a key factor to assure effectiveness and accuracy of the fusion itself.

In FIRME, each subsystem, together with an identity, returns a pair of values. The first one is the distance, i.e. a value normalized in the interval [0,1] which indicates how much two biometric templates (from face or iris) are far from each other (0 = identical templates, 1= completely different templates). The second value represents the confidence to be associated to the response from the single system. In other terms, a high distance with a low confidence, indicates that the templates are very different, but the system considers the probability of a wrong answer high as well. In such a situation, the response would probably become a false reject. On the contrary, a low distance, with a low confidence, may become a false accept. The fusion scheme implemented by FIRME exploits distance and confidence from both biometrics to produce a single value, which is compared with an acceptance threshold: if it is lower, the subject is accepted, otherwise is rejected. Distance measures used for face and iris have been introduced in detail in the respective sections, we focus here on the definition of the confidence function and on how the obtained result for the current probe is integrated into the fusion scheme.

Confidence values

There is a substantial difference between a quality measure of an input sample and a confidence value measured as a function of the response of the biometric system. If defined independently from the distance metric used for matching, the second one can be used for any recognition system and takes into account the context in which the metric is computed, namely the whole gallery of the recognition system. Within the proposed framework of fusion in a multi-biometric identification setting, it is appropriate to have a biometrics-independent confidence function. We considered two possible formulations of the confidence function ϕ defined in [35]. We used the relative distance and the density ratio, as well as a combination, to obtain the final function (for details see [35]). Given a probe p and an identification system A with a gallery G, and the set of gallery subjects returned by the system ordered by their distance from the probe, the

relative distance is defined as:

$$\phi_1(p) = \frac{d(p, g_{i_2}) - d(p, g_{i_1})}{d(p, g_{i_{|G|}})} \tag{5.1}$$

where d is a distance function with codomain [0, 1]; if the values were not included in that range, it is possible to use a normalization function like the Quasi Linear Sigmoidal (QLS), defined in [35] (used in this work). Relative distance is based on the principle that if a person is genuine, there is a great distance between the scores assigned to the first retrieved identity and the score assigned to the immediately closest one. The density ratio, is instead defined as:

$$\phi_2(p) = 1 - |N_b|/|G|,$$

where

$$N_b = \{g_{i_k} \in G | d(p, g_{i_k}) < 2 \cdot d(p, g_{i_1})\} \tag{5.2}$$

In practice, we consider the identities returned in a process of identification as a cloud; the more this cloud thickens near to the first returned identity, the more unreliable is the answer, as there are many individuals that might be potential candidates as well; on the contrary, the response is reliable, if the cloud is scattered. We also adopt a variation of the density ratio. As one can observe in the definition of N_b in equation 5.2.3, the factor 2 (the cloud radius) is constant. We define here a new version N_c of the term used to compute ϕ where this factor is proportional to the distance of the first returned identity, so that the cloud radius depends on the relative distance according to the following definition:

$$\phi_3(p) = 1 - |N_c|/|G|,$$

where

$$N_c = \left\{ g_{i_k} \in G | d(p, g_{i_k}) < \frac{(1 + d(p, g_{i_2}))(1 + d(p, g_{i_2}) - d(p, g_{i_1}))}{4} \right\} \quad (5.3)$$

Since the value of d ranges in [0,1], the maximum (absolute) value for the numerator in the second term in equation 5.2.3 is 4, and this explains the normalization factor.

Fusion schema

The fusion schema adopted by FIRME exploits the confidence values produced by face and iris systems to fuse their returned distances. The two values are $\phi_F(p)$ for the face and $\phi_I(p)$ for the iris, where ϕ_F and ϕ_I represent the same function applied to the respective galleries. They are used as weights to be assigned to the respective distances, in order to produce a weighted sum, which corresponds to the global distance produced by the system using equation 5.2.3. Of course, in order to make consistent such weighted sum, the sum of the weights must be 1, so that the single values ϕ_F and ϕ_I are normalized with respect to their sum $(\phi_F + \phi_I)$. The obtained global distance is then compared with an acceptance threshold th, so to divide recognized subjects (with a distance lower or equal to the threshold) from the unrecognised ones (distance higher than the threshold).

$$d(p, g_j) = \frac{\phi_F}{(\phi_F + \phi_I)} \cdot d_F(p, g_j) + \frac{\phi_I}{(\phi_F + \phi_I)} \cdot d_I(p, g_j) \quad (5.4)$$

where p represents the probe at hand and g_j is the $j - th$ element in the system gallery. After this, the subjects are ordered according to the increasing distance from the probe. The first one is returned is one sample per subject id present in the gallery, otherwise the one with more samples in the first n positions.

5.2.4 Experimental results

Data acquisition and preprocessing

The dataset employed is composed by 49 different subjects. Four pictures of the face and four pictures of the iris of each subject have been acquired by a Samsung Galaxy Tab 2. The acquisitions have been carried out both indoor and outdoor.

Iris pictures have been segmented by ISIS algorithm, ISIS outputs iris images in a rectangular shape, result of the mapping of the Cartesian coordinates in polar ones.

For what concerns face images preprocessing, they have been cropped to select a close-up of subjects' face.

For the experiments, the dataset has been divided in Probe and Gallery sets, in particular all the 49 subjects have been included in the Probe set, while only 21 subjects (but different images with respect to those included in the Probe set) have been put in the Gallery set. The fact that not all the subjects are included in the Gallery, implies a more challenging scenario since it simulates the case in which an impostor, not enrolled in the system, try to get authenticated.

Performance of single biometrics

The first experiment aims at evaluating the performance obtained by the single biometrics, both in terms of Cumulative Match Curve (CMC), and Receiving Operating Characteristic (ROC) curve. The curves in figure 5.12 represent the performance of the system when using samples acquired indoor as Probe.

We can observe from figure 5.12 that face offers much better results than iris. This seems to be in contrast with general reported trends, which indicates iris as a much more reliable biometrics than face. However, this result can be explained. First of all, regions containing the two biometric traits are captured at a resolution that can be sufficient for face recognition, but less appropriate for iris, whose patterns require a better resolution. We have to consider, that, in general, iris is captured using dedicated sensors, in infra-red spectrum and at a distance that assures an optimal resolution in the relevant portion of the image.

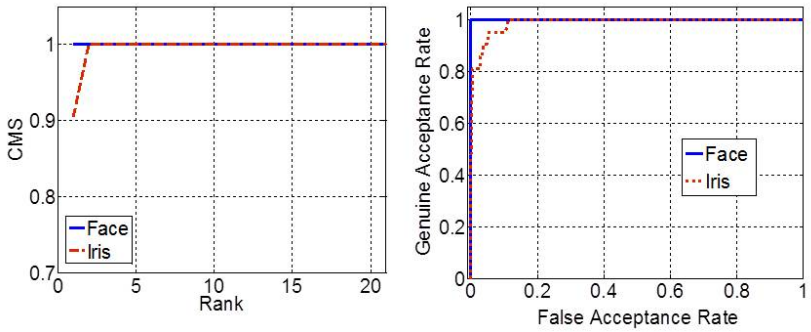

Figure 5.12: Performance of face and iris biometrics with indoor probes.

In this specific case, the on-board camera does not provide an acquisition with the resolution usually exploited by iris-based systems, so that the localization and normalization processes are especially affected, and as a consequence the following matching process. A second consideration is that, while face is usually processed in the visible spectrum, iris is most often processed in infra-red, so that processing in visible light and under-controlled conditions is a less usual setting for this biometrics. Last but not least, face recognition achieves good results even because the natural interaction pattern with the device takes the user to spontaneously adopt a frontal (and therefore less disturbed) pose, with normalization and best sample selection further intervening to address possible pose problems. figure 5.13 shows the same type of experiment, with samples acquired outdoor used as Probe.

Curves in figure 5.13 confirm the trend observed in the previous experiment, further highlighting that the available resolution better supports the face with respect to iris, which occupies a much smaller region in the image. A further factor to consider is the difficulty in focusing by the sensor, which hinders a correct acquisition of all features in the iris. However, as a final remark, we may consider that the increasing availability of high-performance sensors even on medium level mobile devices will soon turn this situation in favour of the most reliable iris biometrics.

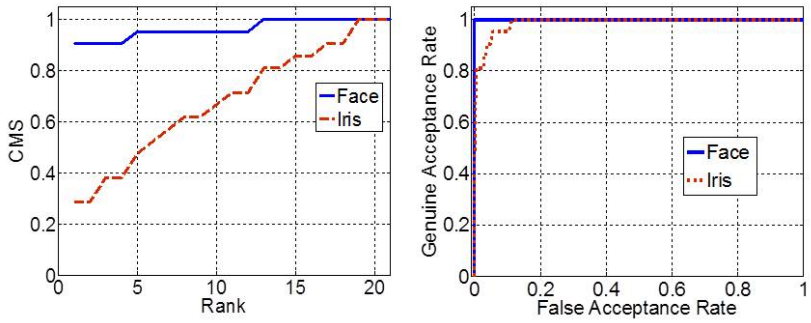

Figure 5.13: Performance of face and iris with outdoor probes.

Performance of FIRME with different confidence functions

In this experiment, modules for face and iris have been fused at score level, based on confidence values. The fusion policy is the one described in section 5.2.3. The same experiment has been performed both for indoor and outdoor Probe. Results with indoor Probe are reported in figure 5.14, while those with outdoor Probe are reported in figure 5.15.

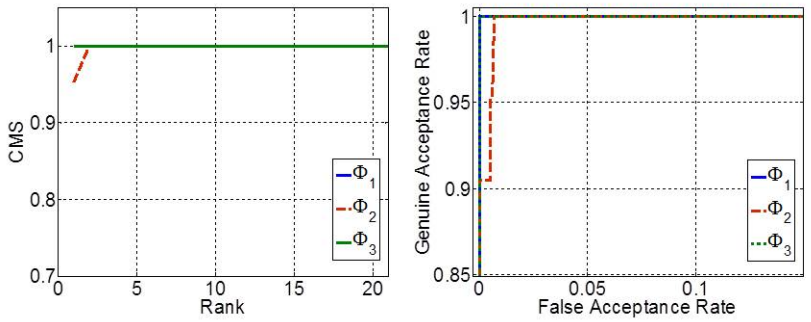

Figure 5.14: Performance of FIRME with indoor probes. ϕ_1, ϕ_2, and ϕ_3 indicate the confidence values *relative distance*, *density ratio*, and a combination of the two, respectively.

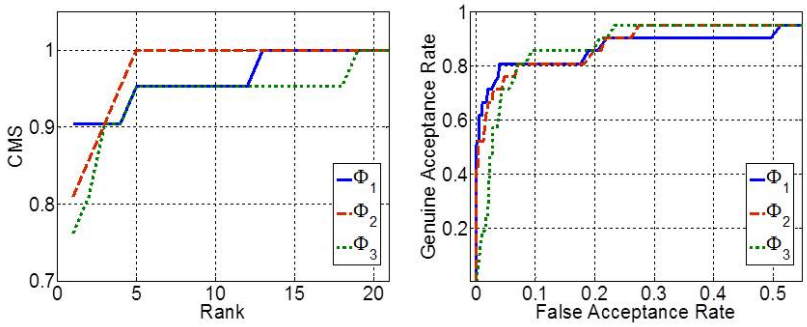

Figure 5.15: Performance of FIRME with outdoor probes. ϕ_1, ϕ_2, and ϕ_3 indicate the confidence values *relative distance*, *density ratio*, and a combination of the two, respectively.

Each graph in figure 5.14 and figure 5.15 reports three curves. In some cases, two of them completely overlap, so that one is not visible. The experiments with outdoor probes show how the three different confidence functions ϕ provide very different performance. The worst one is ϕ_3. Since it is based on a cloud of subjects around the first returned one whose radius depends (is proportional) on distance between the first and second responses from the same biometric system, we can deduce that this distance is not sufficiently significant to capture enough information about the behaviour of the rest of the gallery to return a useful reliability measure. The density ratio, i.e. function ϕ_2, considers a cloud of responses around the top ranked one provided by the system with a fixed radius, which seems to give a better account of how the template of the returned identity is related with the overall Gallery. As a matter of fact, if we observe the corresponding CMC and ROC curves, we can notice that FIRME performance improves significantly. Among the three, ϕ_1 is the one giving more rambling results. For the indoor tests, FIRME achieves optimal performance. On the contrary, in outdoor tests, though satisfying, performance with ϕ_2 are slightly worse than those with face alone. However, the element to consider is the robustness of the adopted fusion technique, besides its versatility. As a matter of fact, it is able to efficaciously contain the scarce performance provided in

this case by the iris. This demonstrates that the combination of more biometric techniques, when appropriately designed, is able to neutralize, on a given device, a negative behaviour that might be observed from time to time, possibly due to bad acquisition, sensor problems, incorrect device orientation, etc. In this way the overall response from the system, in this case FIRME, is still reliable. This also supports the belief that, with the predictable improvement of acquisition devices and better iris images, the recognition accuracy can significantly improve, and therefore it is worth following the multi-biometric approach.

5.3 Combining Biometry and Hardwaremetry

In this section, a multi-modal authentication system that combines a biometric trait, namely the iris, with a personal object owned by the user, namely the smartphone is presented[8][9]. This approach has several advantages:

- The genuine sample consists in the couple user-device, making more difficult the spoofing process;
- The two recognition processes are applied on a single photo of the eye captured by the user with his/her smartphone;
- Good trade-off between accuracy and ease of use;
- Performances of iris recognition and, in particular, of sensor recognition, are very high.

The system we propose is therefore a multi-modal recognition system based on the combination of sensor recognition (hardwaremetry) and iris recognition (biometry), i.e. something the user has + something the user is. If we analyse the authentication systems security levels shown in figure 5.16, we can see that the degree of security assured by the combination of biometry and a physical object is higher with respect of the use of biometry only[1]. The second aspect that we address in this work, is the sensor interoperability problem [55]. This

[1] http://www.ffiec.gov/pdf/authentication_guidance.pdf

problem rises up when the data to be compared (e.g. the pictures of the eye) are acquired with different sensors and thus contain differences depending on the sensor characteristics. As we will show later, this can affect the biometric algorithm performances since two pictures of the same eye can appear different because if they were captured by different devices. Our approach can be seen as a way to bypass the sensor interoperability problem, instead of focusing on the development of an algorithm able to operate regardless of the sensor employed, we leverage on the differences introduced by different sensors on photos in order to obtain a more robust recognition system.

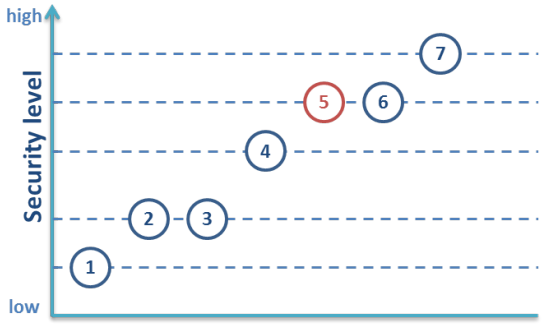

Figure 5.16: Authentication systems security levels: (1) Something the user knows; (2) Something the user has; (3) Something the user knows + something the user has; (4) Something the user is or does; (5) Something the user has + something the user is or does; (6) Something the user knows + something the user is or does; (7) Something the user knows + something the user has + something the user is or does.

5.3.1 Hardwaremetry

In order to recognize the sensor that captured a given photo, we implemented the Enhanced Sensor Pattern Noise (ESPN) based algorithm presented by Li in [57]. This method extracts from a picture the noise pattern of the sensor, it can also be used to distinguish cameras of the same model [58][59]. The approach

presented by Li, is based on a previous work by Lukás *et al.* [58] in which the authors present the algorithm for extracting the Sensor Pattern Noise (SPN).

The ESPN is extracted from the Sensor Pattern Noise (SPN) by applying a filter that removes the details of the image located in the highest frequencies. The SPN is obtained using the following formula:

$$n = DWT(I) - F(DWT(I))$$

where $DWT()$ is the discrete wavelet transform to be applied on image I and $F()$ is a denoising function applied in the DWT domain. For $F()$ we used the filter proposed in appendix A of [58]. In figure 5.17 the denoising process is illustrated: figure 5.17 (a) shows a sample of the MICHE database. We selected this image because it contains many "strong details", e.g. the frame of the glasses or the dark hair on a light background, that in the frequency domain are located in the high frequencies and can affect the Sensor Patter Noise extraction process. In figure 5.17 (c) and figure 5.17 (e), we can see how the denoising process has mitigated the presence of those details. SPN is then enhanced as suggested in [57] according to the following formula:

$$n_e(i,j) = \begin{cases} e^{\frac{-0.5n^2(i,j)}{a^2}}, & \text{if } 0 \leq n(i,j) \\ -e^{\frac{-0.5n^2(i,j)}{a^2}}, & \text{otherwise} \end{cases}$$

where n_e is the ESPN, n is the SPN, i and j are the indices of the components of n and n_e, and a is a parameter that we set to 7, as indicated in [57]. An example of the difference between the SPN and the ESPN is shown in figure 5.18, the original picture contains many details that influence the SPN but they can be mitigated by the enhancing step.

The process shown above, allows us to obtain the ESPN, i.e. the "fingerprint" of the sensor that captured the given photo. To associate then the extracted fingerprint to the correct sensor, we have to compare this fingerprint to the Reference Sensor Pattern Noise (RSPN) of the sensor. To extract the RSPN n_r of a sensor, we compute the average SPN over N photos acquired with

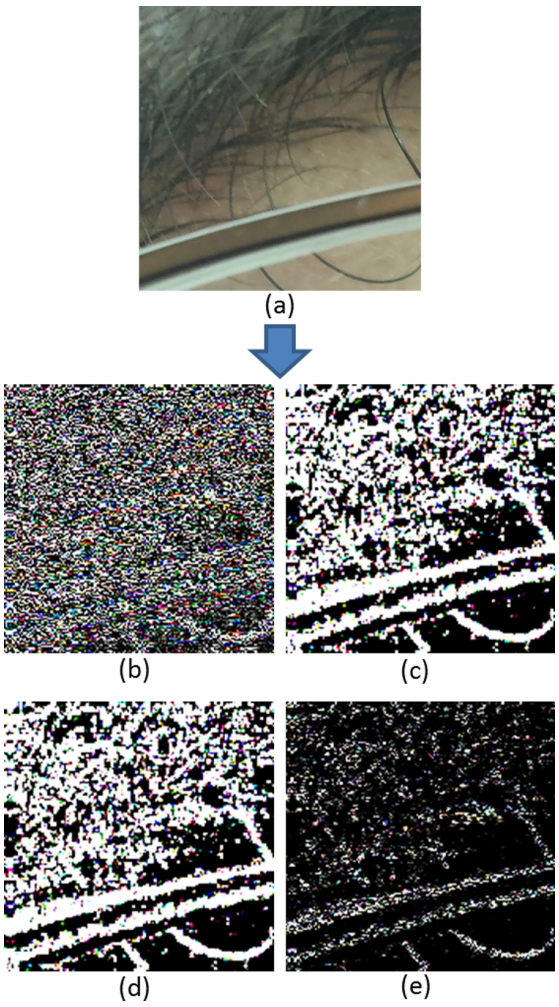

Figure 5.17: Denoising process: (a) original image; (b) original wavelet coefficients; (c) local variance; (d) selection of the minimum variance; (e) denoised wavelet coefficients.

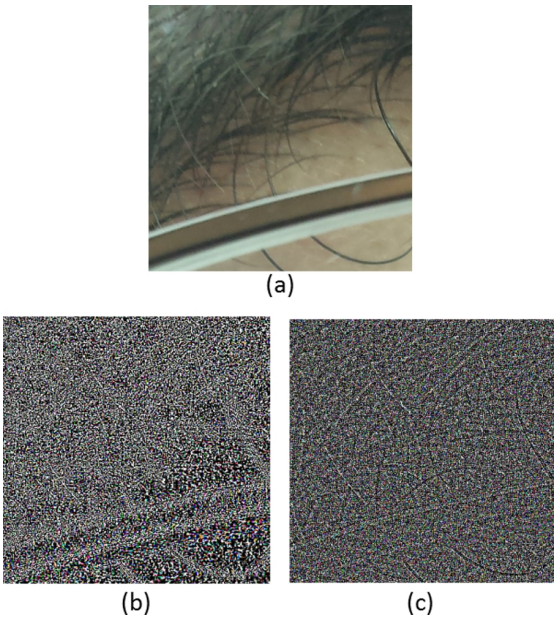

Figure 5.18: Sensor Pattern Noise enhancing: (a) original image; (b) the SPN extracted from the image contains image details (e.g. hairs, part of the eyeglasses frame); (c) ENSP, after the enhancing step the influence of image details is mitigated.

the given camera:

$$n_r = \frac{1}{N} \times \sum_{k=1}^{N} n_k$$

Where n_k is the SPN extracted from the k^{th} image. To compare the ESPN extracted from a photo with a RSPN of a sensor, we compute their correlation as follows:

$$corr(n_e, n_r) = \frac{(n_e - \overline{n_e}) \times (n_r - \overline{n_r})}{\|n_e - \overline{n_e}\| \|n_r - \overline{n_r}\|}$$

where the bar above a symbol denotes the mean value.

5.3.2 Score Normalization

Score normalization is a necessary step when combining different modules. The algorithms employed by each module can generate scores that are different in terms of distribution and numerical range. In the past several different methods of score normalization have been proposed [60], addressing different issues that can emerge during the fusion process. In our experiments, we tested five different normalization techniques, namely: Max-Min, Z-score, Median/MAD, TanH, and Sigmoidal. We will briefly explain these techniques in the following. Let's denote the set of K scores as: $S \rightarrow s_k, k = 1, 2, ..., K$, and the resulting set of normalized scores as: $S' \rightarrow S'_k, k = 1, 2, ..., K$.

Max-Min normalization technique

With the Max-Min technique, the scores are normalized based on the maximum and the minimum values in the scores set. The advantage of this simple technique is that the resulting scores set has a fixed numerical range: [0, 1]. In addition the shape of the original scores distribution is preserved. The Max-Min normalization technique can be implemented using the following formula:

$$s'_k = \frac{s_k - min_k s}{max_k s - min_k s}$$

Z-score normalization technique

The Z-score normalization technique is based on the calculation of the arithmetic mean and the standard deviation of the scores set. Thus, the resulting normalized scores set has a mean of zero and a standard deviation of one. However, this technique does not assures that the resulting scores set has a common numerical range, and it can also be sensitive to the presence of outliers. The Z-score normalization technique can be implemented using the following formula:

$$s'_k = \frac{s_k - mean(s)}{mean(s) - std(s)}$$

Median/MAD normalization technique

This technique is based on the median and median absolute deviation (MAD) that are insensitive to outliers. The Median/MAD normalization technique can be implemented using the following formula:

$$s'_k = \frac{s_k - median(s)}{MAD}$$

where $MAD = median(|S_k - median(s)|)$. This technique too has the disadvantage that does not preserve the input distribution and does not transform the scores into a common numerical range.

TanH normalization technique

The tanH technique was introduced by Hampel *et al.* [62]. It is robust and highly efficient and the normalization formula is:

$$s'_k = \frac{1}{2}\{tanh(0.01(\frac{s_k - \mu_{GH}}{\sigma_{GH}})) + 1\}$$

where μ_{GH} and σ_{GH} are the mean and standard deviation estimates, respectively, of the genuine score distribution as given by Hampel estimators [62]. An advantage of this method is that it is not sensitive to outliers.

Sigmoidal normalization technique

Cappelli *et al.* [63] adopted this technique in order to combine the scores of a multi-modal biometric system based on the combination of different fingerprint classifiers. The normalized score can be obtained by the following double sigmoid function:

$$s'_k = \begin{cases} \frac{1}{1+exp(-2(\frac{s_k-t}{r_1}))} & \text{if } s_k < t, \\ \frac{1}{1+exp(-2(\frac{s_k-t}{r_2}))} & \text{otherwise} \end{cases}$$

where t is the reference operating point and r_1 and r_2 denote the left and right edges of the region in which the function is linear, i.e., the double sigmoid function exhibits linear characteristics in the interval $(t - r_1, t - r_2)$. This technique guarantees that the set of normalized scores has a common numerical range [0, 1]. But, it requires careful tuning of the parameters t, r_1, r_2 to obtain good efficiency.

5.3.3 Experimental results

Data acquisition and preprocessing

For sensor recognition no preprocessing is needed, the picture as it is, is submitted to the ESPN extractor. For iris recognition some further steps are required, in fact we need to extract the iris from the whole picture that contains also other information, e.g. the periocular area and part of the face, that we do not need in the following steps. In order to focus on recognition performances and avoid the negative influence of a wrong segmentation, we manually segmented the images. After iris segmentation we performed a transformation from Cartesian to polar coordinates in order to obtain a rectangular shape of the iris on which is easier to apply the CSUM algorithm.

Sensor recognition

It is well known [58][59] that device recognition based on SPN extraction is a very robust technique. However we investigated its use on mobile devices, which

are limited in terms of memory and computational power. In this section we will present different experiments on sensor recognition in order to show the robustness of this technique even if applied on a small part of the image. In appendix A of [58], it is suggested to process large images by blocks of 512×512 pixel, but during our experiments we observed that using just one block, the same for all images, is sufficient to obtain a RR of 98%, for this reason, in our experiments we extracted from all the images a block of size 512×512 starting from the top-left corner of the photo. As stated before, with the SPN-based technique it is possible to distinguish which is the device that captured a given photo even among different devices that embed sensors of the same model. In order to test the performance of the sensor recognition algorithm, we extracted the RSPNs of four cameras employed to acquire MICHE images: Galaxy S4 front camera, Galaxy S4 rear camera, iPhone 5 front camera and iPhone 5 rear camera. While collecting the images for MICHE database, the iPhone 5 was changed with another device of the same model, this means that from the subject with ID = 49, photos were acquired with the new device, and since we extracted the RSPN from the new device, pictures relative to IDs less than 49, should be detected as impostors by the system. Moreover, the presence of unrolled IDs in the probe, i.e. pictures captured with a device of which we do not have the RSPN ("old" iPhone 5) in the Gallery, makes the system performance assessment more reliable. We used the RSPNs extracted from the four cameras as Gallery set and the ESPNs extracted from 579 photos selected from the MICHE database as Probe set. The system obtained a RR of 98% and a very low average FAR of about 5%, AUC is equal to 0.99. Results for sensor recognition are shown in Figure 5.19.

Different sensors of the same model

In the following experiment, we employed three smartphones of the same model, namely Samsung Galaxy S4. We compared a total of 6 different sensors, as we extracted the RSPNs from both the frontal and the rear cameras of the three devices. Gallery set is thus composed by the 6 RSPNs while the Probe consists in 1297 images from MICHE database, the ones captured with Galaxy

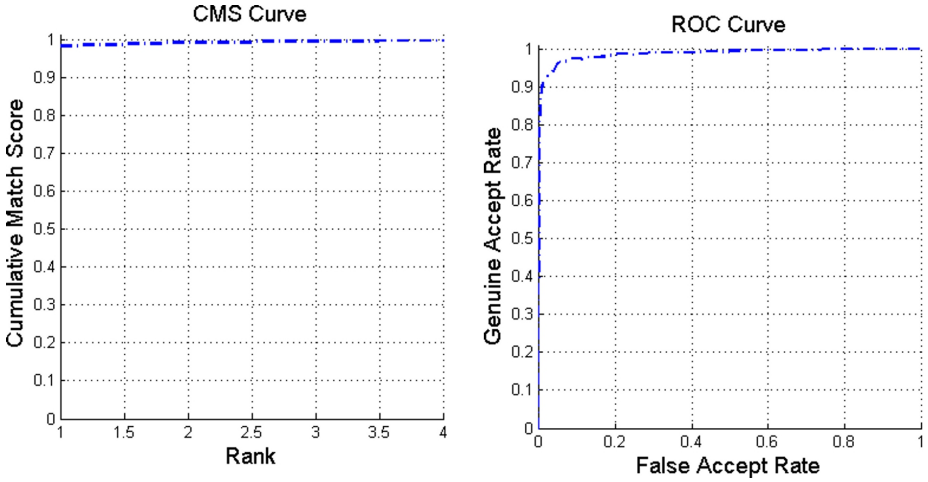

Figure 5.19: Sensor recognition performances.

S4 smartphone. As just one of the three smartphones was actually employed in MICHE acquisition, the system should correctly assign all the 1297 images to the correct device. Results, see figure 5.20, shows that the Recognition Rate (RR) is 97% and the Area under the ROC curve (AUC) is equal to 0.99.

Reference Sensor Pattern Noise Extraction

In order to extract the RSPN of a sensor it is worth employing an high number on images (recommended more than 50) of the blue sky because this kind of pictures do not contain details that, as the sensor noise, are located in the high frequencies of the images and can be confused with it [58]. However, we imagined that for a user, collecting images of the blue sky could be difficult, for this reason and as suggested in [64], we compared the performance of the sensor recognition system when using blue sky images or using any kind of pictures to extract the RSPN of the sensor, results are presented in figure 5.21. Values obtained for RR and AUC are very close, with RR = 98% for both the experiments and AUC = 0.92 and AUC = 0.93 for the case in which RSPN is extracted from

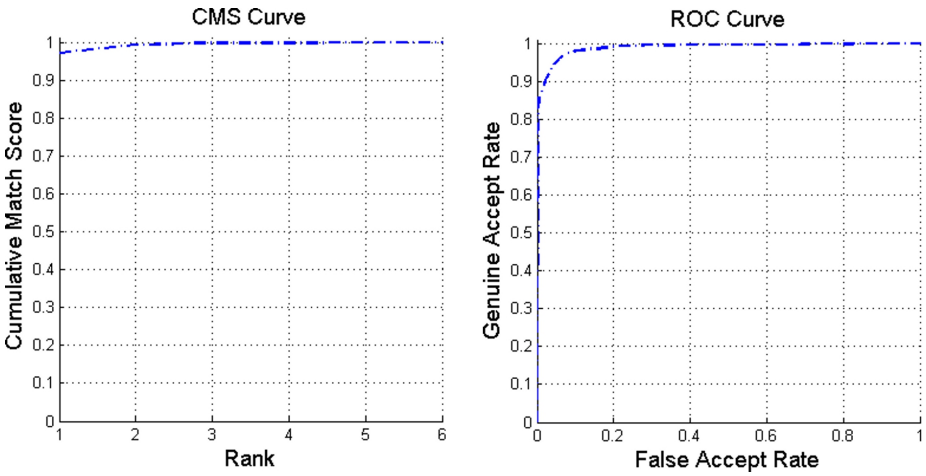

Figure 5.20: Sensor recognition performance: experiment on same camera model.

blue sky images and the case in which it is extracted from any kind on pictures, respectively. We performed these experiments on 1297 images captured by the Galaxy S4, and compared them with 3 RSPNs from three frontal cameras of three different Galaxy S4.

Iris recognition

In this section we present the performances of the iris recognition module. The CSUM algorithm has been employed to extract iris features (see section 5.1.2 for details). Images selected from MICHE dataset were split in Gallery and Probe sets. Probe set is composed by 298 iris images belonging to 75 subjects. The Gallery is composed by 150 iris images, we selected only a part (half) of the 75 individual subjects composing MICHE dataset, we did it in order to simulate the attempt of not enrolled subjects to access the system. Results are shown in Figure 5.22, with RR = 85% and AUC = 0.77. Performances are poor due in part to the noise introduced by the acquisition in uncontrolled settings (e.g. specular reflections, eyelids and eyelashes occlusions, etc.) and in part to the

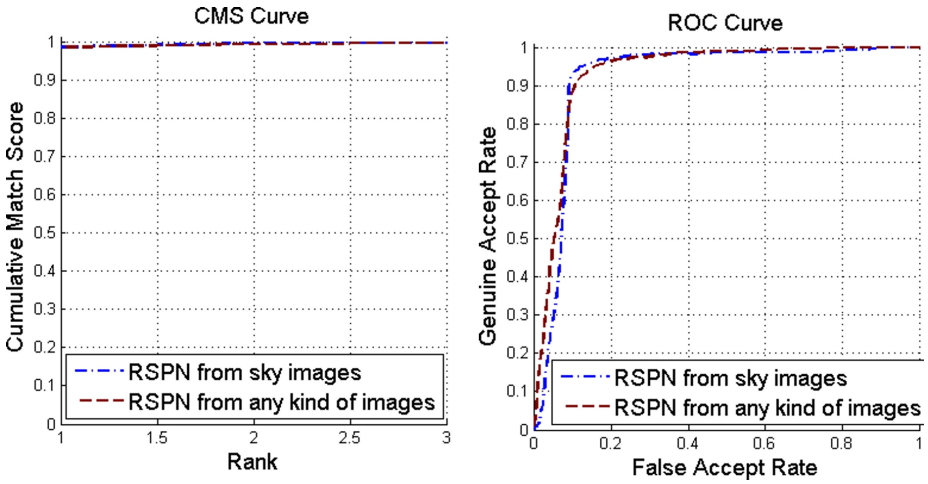

Figure 5.21: Sensor recognition performances: in red the curves relative to the case in which the RSPN is extracted from blue sky images and in blue the curves relative to the case in which the RSPN is extracted from any kind of images captured by the sensor.

sensor interoperability problem, in fact MICHE images of the same iris often appear very different because they were acquired by different sensors.

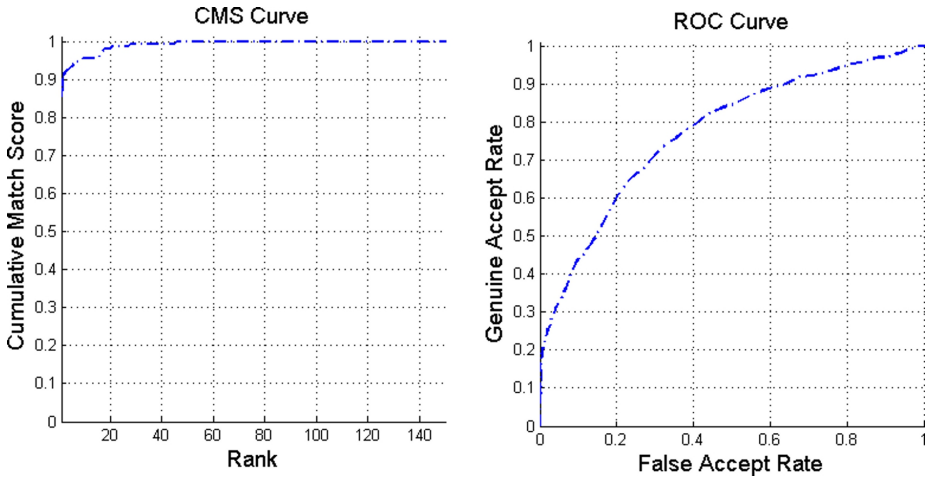

Figure 5.22: Iris recognition performances affected by sensor interoperability problem.

Fusion at feature level

In this paragraph we present the experiments relative to the combination of iris recognition and sensor recognition. We first tested the fusion at feature level, concatenating the feature vectors extracted from the two recognition modules. To compare the new feature vectors obtained, as the two algorithms employed for iris and sensor recognition use different matching techniques, namely hamming distance for iris recognition and correlation for sensor recognition, we tested both approaches. Results are presented in figure 5.23: performances are very close with AUC of about 0.93 for both distance metrics and RR of 23% obtained by Hamming distance and RR of 20% when using Correlation.

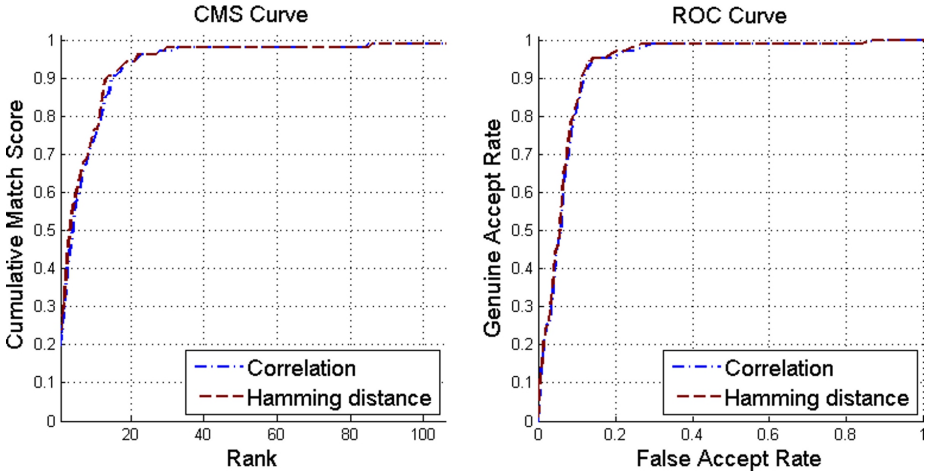

Figure 5.23: Fusion at feature level performances.

Fusion at score level

We also tested fusion at score level, we computed the distance matrices of the two recognition modules and then we combined the scores obtained averaging them. Before combining the score, a score normalization step is required. We tested different normalization techniques and we report the results obtained in Figure 5.24. The best performances are obtained via fusion at score with Max-Min score normalization: the AUC value is equal to 0.98 while the RR is 86%. The results obtained show that fusion at score level is more suitable than fusion at feature level for this kind of system. Since the system recognize a couple of entities very different in nature, we assigned the same weight both to iris module scores and sensor module ones, in order to avoid the system to be biased towards recognizing the iris or the device.

Noise response

In this section we want to highlight the advantage in using the sensor recognition module in combination with biometric recognition. It has really high and

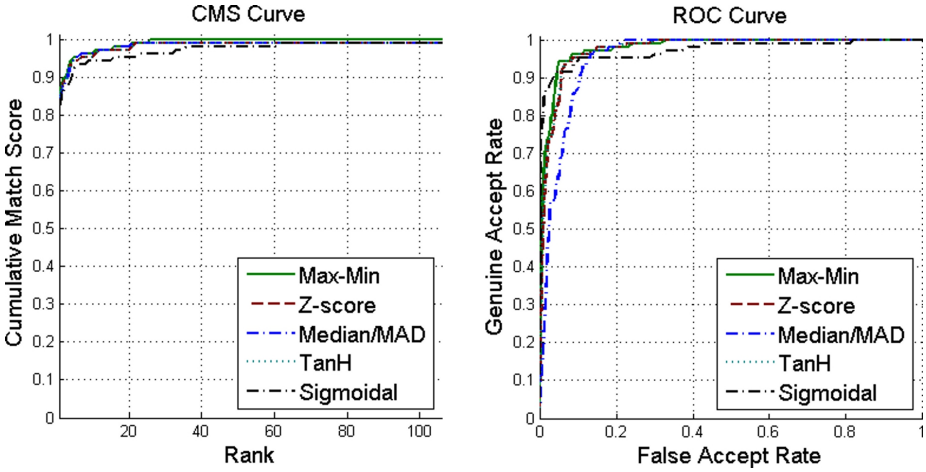

Figure 5.24: Fusion at score level performances.

robust performances as shown in section 5.3.3. Here we present an example that confirms what stated before: if we submit to the system more challenging pictures, e.g. eye pictures with strong noise due to large specular reflections, important occlusions etc., the iris recognition module performances drop while sensor recognition performances remain the same. We employed pictures from MICHE database acquired outdoor, in figure 5.25 we can see that, with respect to performances obtained on indoor photos, the RR drops from 85% to 21% and the AUC from 0.77 to 0.67. Figure 5.26 shows that even on outdoor pictures, sensor recognition performances remain high, with RR = 98% and AUC = 0.99.

5.3.4 Conclusive Remarks

An extensive series of experiments has been performed to show that the SPN-based technique can be reliably applied on smartphones. This technique is based on the discrete wavelet transform. Large images, as those captured by nowadays smartphones, should be processed by blocks of 512×512 pixel. E.g. for a picture of MICHE Iris database acquired by Galaxy S4, of size 2322×4128, the ESPN extraction process should be applied around 36 times. This would require a

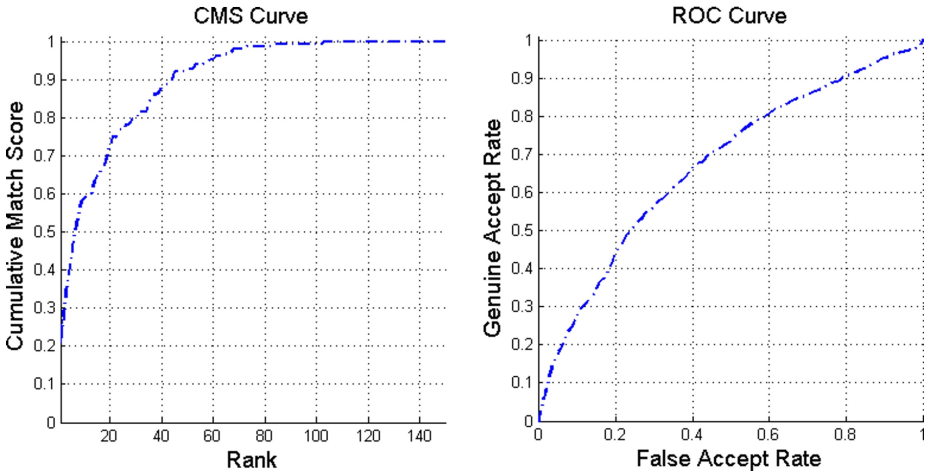

Figure 5.25: Iris recognition performances on outdoor images.

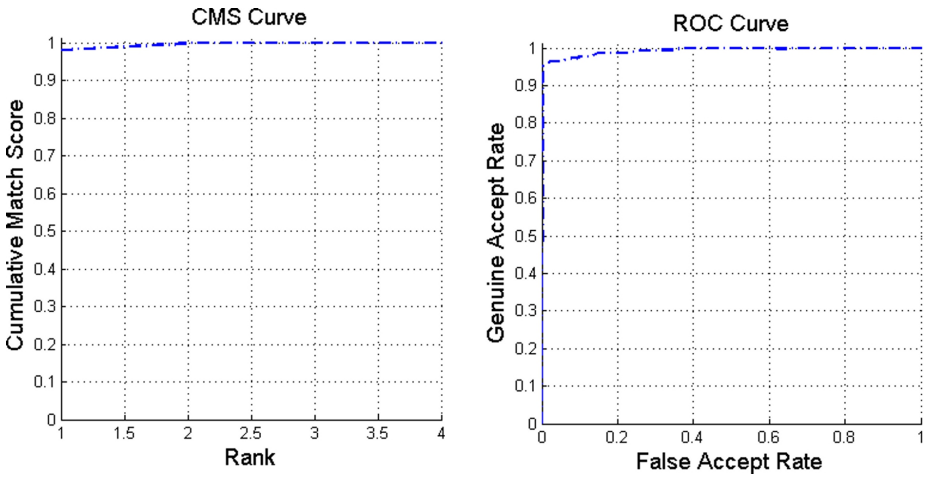

Figure 5.26: Sensor recognition performances on outdoor images.

high computational cost not suitable for the application of this technique on smartphones, that are still limited in terms of memory and computational power. To speed up the process and reduce the computational cost we used just a small block of the image and we obtained a Recognition Rate (RR) of 97% and an Area under the ROC curve (AUC) equal to 0.99.

	EER	**avg FAR**	**avg FRR**	**RR**	**AUC**
Iris	0.2951	0.2747	0.6044	0.8553	0.7723
Sensor	0.0447	0.0537	0.5592	0.5592	0.9883
Fusion	0.0569	0.2758	0.3590	0.8585	0.9797

Table 5.1: Experimental results summary.

For what concerns the performances obtained by the fusion of device and iris recognition modules, we want to clear up that the performances were not expected to outperform the single modules. The reason is that the two modules recognize two different entities and their fusion recognize a combination of entities. This is different from a multi-biometric system were two user's traits are combined to recognize his/her identity (same entity) and thus the performances should outperform the single modules. Indeed the performance were expected to be limited by the weakest module: the iris recognition module. However, we can see that, as reported in Table 5.1, fusion performances greatly outperform iris recognition ones. The reason is that the CSUM algorithm employed for iris recognition, suffers for the sensor interoperability problem introduced before. Thus, when it is required to recognize the iris despite the sensor that acquired it, its performances drop down. On the contrary, on the fusion scenario, it is required to distinguish between irises acquired with different sensors, getting around the sensor interoperability problem, and obtaining better performances from the iris recognition module.

Finally we demonstrated that the more the quality of acquired iris degrades the more the SPN is important in a verification process. In fact the experiments presented show that sensor recognition has very high and robust performances.

Future Implications and Open Issues

This novel system can provide a more secure authentication process without the disadvantage of requiring dedicated sensors. The authentication process is fast and easy, in one single shot the user can get authenticated via his/her iris and his/her smartphone. The smartphones are nowadays strictly related to the owners, and in many companies smartphones are provided to the employees and they are required to bring them during the working hours.

This is the perfect scenario in which passwords, tokens or badges can be replaced by the authentication system proposed here. And it is worth noticing that this kind of system is particularly suitable for this scenario because it can distinguish devices of the same model with high accuracy, as shown in paragraph 5.3.3, and it is very likely that the devices provided by a company are of the same model or belong to a restricted set of models.

The use of the recognition of the smartphone in addition to the iris, makes the spoofing attacks more difficult to be carried on. The opponent has to spoof both modules, that even if still possible, it is more complicated than spoofing a system based only on biometric recognition.

Chapter 6

Gaze Analysis

In this chapter novel Gaze ANalysis Technique, namely GANT[6][4] is presented, exploiting a graph-based representation of fixation points obtained by an eye tracker during human computer interaction. The main goal is to demonstrate the conjecture that the way an individual looks at an image might be a personal distinctive feature.

The experimental results obtained are very interesting, confirming that gaze can be used as behavioural biometric for human recognition, and further investigations have been carried out in order to test gaze analysis as a gender and age classifier[5][9].

Even if this topic differs from the others discussed in this thesis, it has been included since the gaze is extracted from the analysis of the eyes (their movements), and can be easily combined with iris recognition, and in particular they could be acquired by a single sensor at the same time. This is an important aspect when considering the development of biometric systems easy to use and requiring less cooperation from the user.

6.1 Method

Data were acquired through the Tobii 1750 remote eye tracker, which integrates all its components (camera, infra-red lighting, etc.) into a 17" LCD monitor (1280×1024 resolution). As experimental stimuli, images representing human faces were used.

6.1.1 Data acquisition

In the Tobii eye tracker, five NIR-LED (Near Infra-Red Light Emitting Diodes) light eyes up producing reflection patterns. An image sensor records pupil position and corneal-reflections to determine eyes position and the gaze point. For correct use of the system, at least one eye (better if both) must stay within the field of view of the infra-red camera, which can be represented as a box with size $20 \times 15 \times 20$ cm placed about 60 cm from the screen. The accuracy of the device is 0.51.

The Tobii ClearView gaze recording software was employed to define stimuli (still images, slideshows, videos, etc. to be presented to the subject), as well as to record and manipulate gaze data. The system acquires 50 raw gaze coordinates per second, from which fixations are then obtained (characterized by coordinates, duration and timestamps). For the purpose of our experiments, a fixation was considered as a sequence of successive samplings detected within a circle with a 30 pixel radius, for a minimum duration of 100 ms. The ClearView software also allows to obtain two useful graphical depictions, namely gazeplots and hotspots. While a gazeplot displays the sequence of fixations of a user, in the form of circles with areas proportional to fixation times, a hotspot uses colour codes to highlight those screen portions in which there are high concentrations of fixations and consequently have been watched most. Gazeplot circles are numbered, thus clearly indicating the fixation sequence. A further output of the eye tracker is the gaze replay, which dynamically shows the evolution of fixations and saccades over time.

6.1.2 Data normalization

To analyse how a particular observer looks at faces, all his or her observations i.e. recordings of fixation point sequences while looking at each one of the 16 faces, are merged together. To do this, a normalization step is required, because in different face images a particular area of interest such as right or left eye and mouth, may be located in slightly different positions in the image.

Even a deviation of only few pixels in the position of a face in an image with respect to another may imply that in the first image a fixation at coordinates (x, y) falls in the right eye area, while in the second the fixation falls in the right eyebrow area. Moreover, face images used in our experiments have little different sizes. In figure 6.1, for example, the point at coordinates $(x = 280, y = 420)$ corresponds to different face areas in different face images.

Figure 6.1: In the left image, the red point at coordinates (280, 420) is exactly in the centre of the right eye, while in the right image the point is on the right cheekbone.

The normalization of fixation point coordinates is performed with respect to scale and shift, by means of an affine transformation. Let be dx the distance between the eyes and dy the distance between the mouth and the eyes middle point (X_M, Y_M) (see figure 6.2), the scaling factors with respect to the horizontal

Figure 6.2: dx (left) and dy (right) distances. M indicates the eyes middle point.

and vertical directions are $s_x = 1/dx$ and $s_y = 1/dy$ respectively. The new normalized coordinates (x_0, y_0) are simply obtained as follows:

$$P(x', y') = \begin{bmatrix} s_x & 0 \\ 0 & -s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -X_M \\ Y_M \end{bmatrix} = \begin{bmatrix} s_x x - X_M \\ Y_M - s_y y \end{bmatrix}$$

After normalization, the coordinates (x_0, y_0) will correspond to the same face area for all 16 subjects.

Once all coordinates of the fixation points are normalized for each of the 16 subjects observed, all fixation points of a given observer, in a single test session, are merged in a single plot. As an example, the result of the normalization phase for observer #15 is shown in figure 6.3. Now it is possible to analyse the cloud of fixation points for each observer and try to extract some features useful to distinguish an observer from another.

Face fiducial points detection

To automatically obtain the distances dx and dy and the coordinates of the eyes middle point M, a face detector system presented in [65], namely, the extended

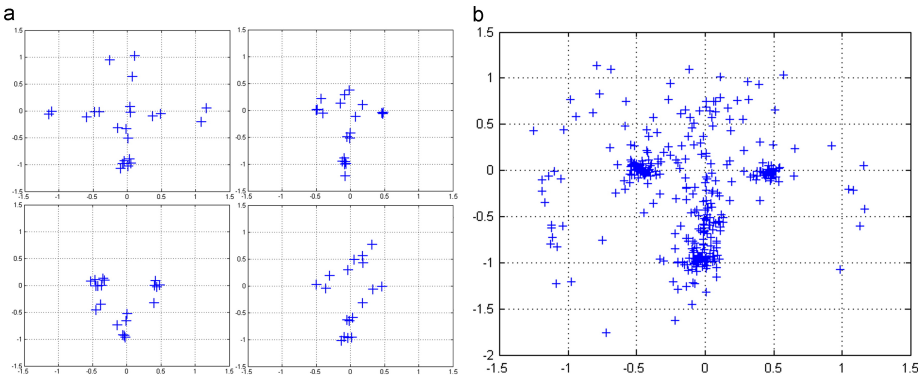

Figure 6.3: Cloud of normalized fixation points of observer #15 in session 1: (a) observation points on four different face images; (b) merging of observation points coming from all the 16 observed face images.

Active Shape Model (STASM) algorithm, is used. First, faces are detected by a global face detector (Viola-Jones [66] or Rowley [67]), which extracts all regions of interest (ROI) from an image that contains at least one face. The detected ROIs are then submitted to the STASM algorithm, which searches for facial landmarks by minimizing a global distance between candidate image points and their homologues using a general model (shape model), which is precomputed ("learned") over a wide set of training images. The algorithm locates 68 interest points. The precision of the location procedure depends on the amount of face distortion. For instance, the output of STASM on subject #3 is shown in 6.4. Among the 68 points extracted by STASM we only need those representing the pupils central points and the mouth central point in order to compute dx and dy distances. Since the eyes middle point M is not returned by STASM, its coordinates have to be computed starting from those of the pupils.

6.1.3 Feature extraction

For each observer in each session, we have now the set of all his or her fixation points. In our previous work "GAS - Gaze Analysis System", presented

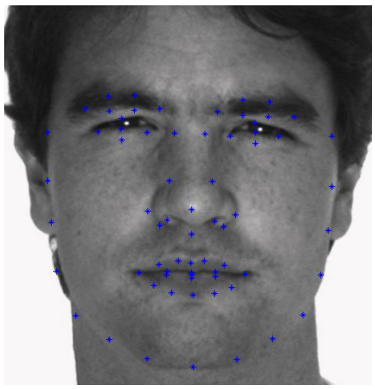

Figure 6.4: STASM output on subject #3.

in [6], we subdivided faces into 17 ROIs such as right eye, left eye, forehead, lips, and background, and then a feature vector was built with each element corresponding to a ROI.

The values in the feature vector corresponded to the sum of durations of fixations contained in the corresponding ROI. However, with the approach described in [6] we only analysed the time spent by the observer in each ROI. Moreover, as each face is different from another, a specific ROI mask was built for each face.

In this work, instead of using a different mask for each face, we decided to use a standard subdivision in ROI for all faces, that we will discuss later.

The objective of this work is to obtain a fully automatic gaze analysis system capable to analyse which areas of a face the observer is used to look at, for how much time and in which order. The idea is to build a graph from the set of fixation points that contains information in the form of weights associated with nodes and arcs, related to the density of fixations and to the time spent in a specific area of the face, and to the observation path.

Another graph based approach for eye movement based biometric recognition is presented in [68]. The authors present a method based on the construction of a joint minimal spanning tree graph structure between a reference and a test

sample of fixation points. The distance between the two samples is measured by a multivariate generalization of the Wald-Wolfowitz random runs test. However, with respect to our approach, only coordinates of fixation points are taken into account, while GANT, through the use of weights associated with nodes and arcs, analyses other important distinctive features such as duration, density and trajectory of the fixation points.

Features graph

Because of the high number of points in the fixation cloud, they are first aggregated in a smaller number of nodes. In order to do this, the cloud is subdivided using a grid. Since in our experiments we have observed that after normalization all fixation point x coordinates ranged in $[-1.5, 1.5]$ and y coordinates ranged in $[-2, 1.5]$, we have chosen a cell size of 0.5×0.5, obtaining a 7×6 grid (see figure 6.3). For each cell of the grid, we consider a node centred in that cell with a default weight equal to zero. Weights are then associated with each node based on:

- Density: the number of fixations in the corresponding cell;
- Duration: the sum of durations (in milliseconds) of each fixation in the corresponding cell.

Figures 6.5 and 6.6 show some examples of density-graphs and duration-graphs.

The method used to create and associate weights to the arcs of the graph is a little more complex due to the fact that we have merged together different observations (i.e. observations of the 16 different subjects). We cannot build a unique path over all fixation points because they belong to different face images, so it is not correct to build an arc from the last node observed on an image and the first node observed on another image. For each face image, we want to increment the weight of an arc linking two nodes A and B each time the observer gaze passes from node A to node B and vice versa.

The arcs of the graph are defined as follows:

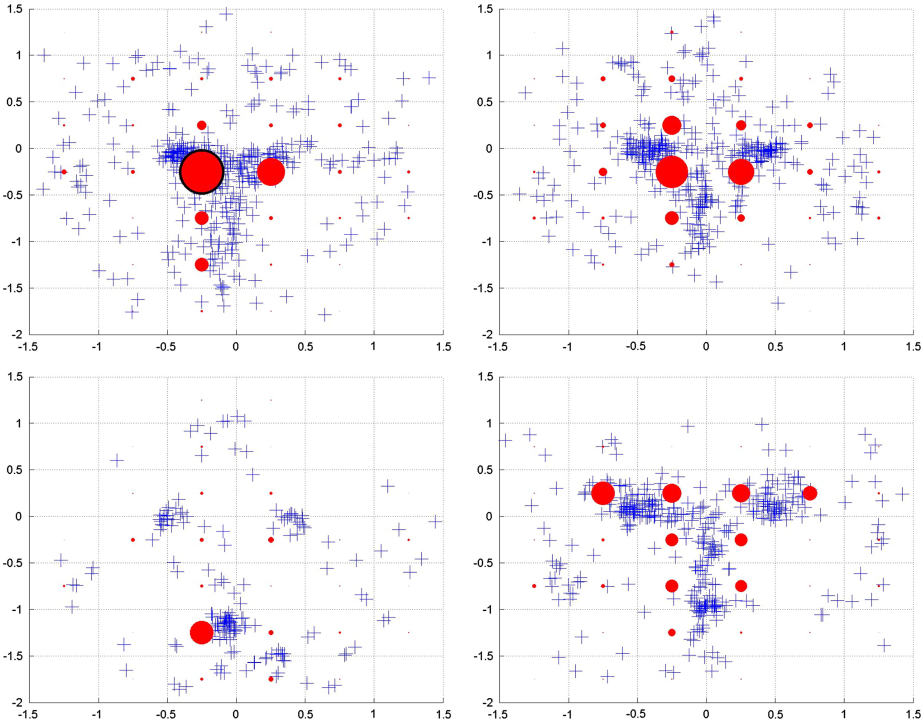

Figure 6.5: Density graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The size of red circles indicates the weight associated with the node. Black bordered circles indicate a weight > 95 (fixation points).

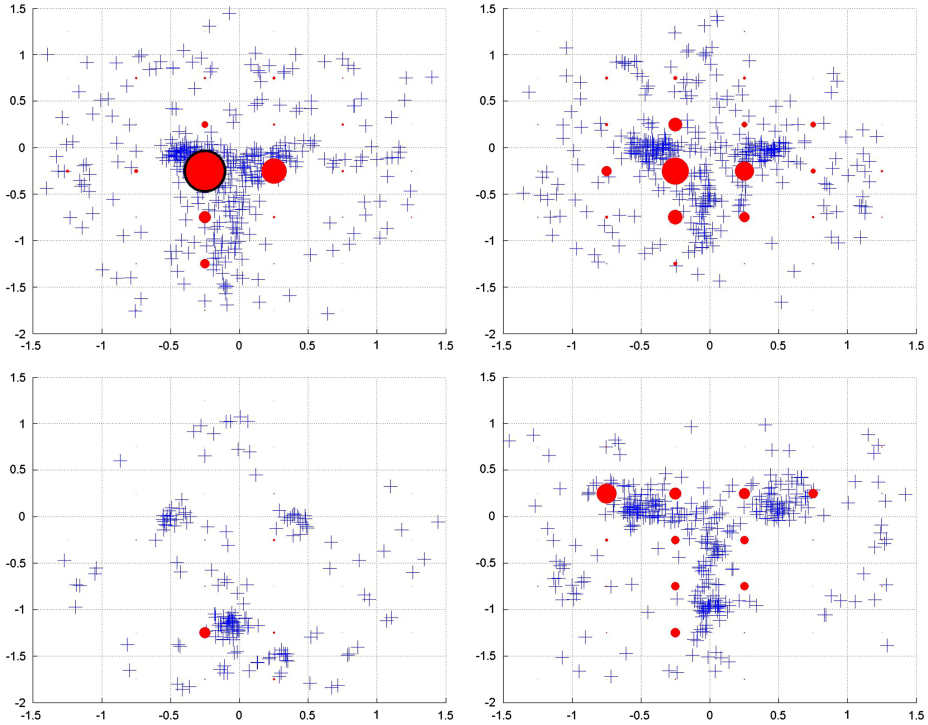

Figure 6.6: Duration graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The size of red circles indicates the weight associated with the node. Black bordered circles indicate a weight $>$ 45,000 (milliseconds).

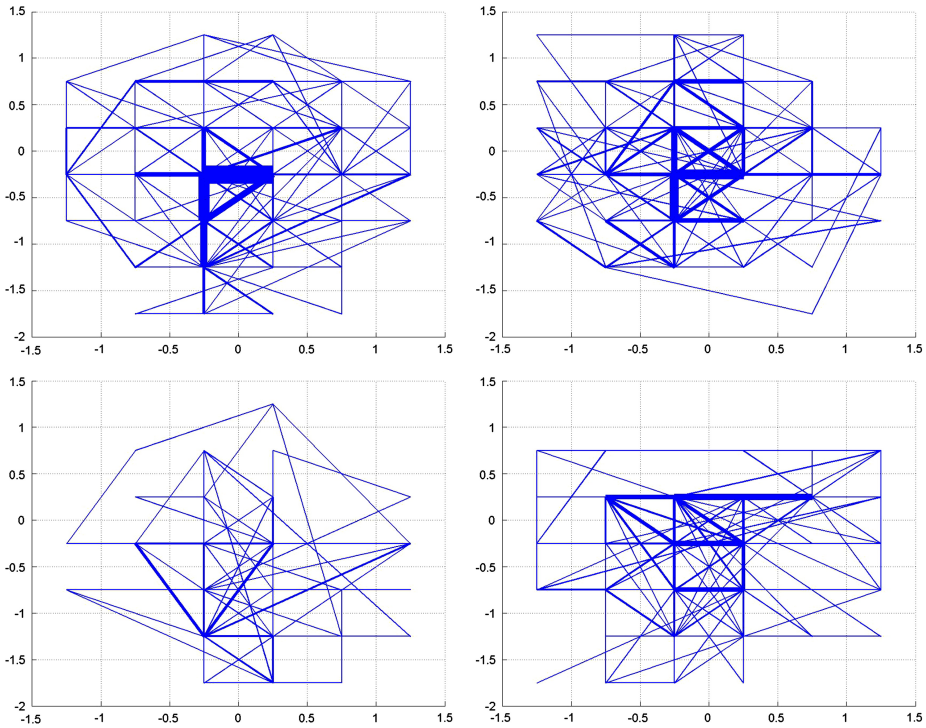

Figure 6.7: Weighted arcs graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The thickness of arcs specifies the associated weight.

Algorithm 1 Weight association

```
Let 0 be the initial weight for all arcs;
Let f_{i,k} be the K^{th} fixation (1 ≤ k ≤ n) of an observer on image i (1 ≤ i ≤ 16);
Let A, B be the first and second endpoints of an arc, A|f_{i,1} ∈ A;
for each f_{i,k} of the fixation sequence do
    B|f_{i,k+1} ∈ B
    if B <> A then
        increment the weight of arc AB by one;
        A = B;
        k = k + 1
        if k = n then
            break;
        end if
    end if
end for
```

Figure 6.7 shows some examples of weighted arcs. The thickness of arcs specifies the associated weight: a ticker arc indicates that the gaze of the observer shifted more frequently between the two nodes connected by that arc. With this approach we do not consider the order in which nodes are observed, but, rather, which nodes are more closely related each other. This representation allows us to easily combine arc graphs, obtained from an observer of different face images, by summing up the weights of corresponding arcs, and also to make the comparison process described in the following easier.

6.1.4 Comparison

The graph obtained for each observer in each session represents the observer fixation model. Weights associated with nodes are described by two matrices 7×6 with each element corresponding to a node, one for densities and one for durations. To represent weighted arcs, an adjacency matrix 42×42 is used. To compare different observers based on densities, durations or arcs, the distance between couples of matrices of the same feature is measured through the Frobenius norm of the matrices' difference. Frobenius norm is a matrix defined as the square root of the sum of the absolute squares of its elements [69].

6.2 Experimental results

6.2.1 Experimental protocol

A total of 112 volunteer observers (73 males and 39 females) took part in the trials, subdivided into the following age groups: 17-18 (11 persons), 21-30 (58), 31-40 (9), 41-50 (16), 51-60 (8), 61-70 (9) and 71-80 (1). All participants reported normal or corrected-to-normal vision.

Prior to the beginning of the experiments, carried out in a quiet environment, participants were informed about the fact that some images, without specifying their kind, would appear on the eye tracker's display in full screen mode (to prevent the user from getting distracted during the gaze recording procedure). Face images were interleaved with blank white screens with a small cross at their centre, to ensure a common starting location for stimulus exploration. The first blank screen was displayed for 5 s, while the others for 3 s. Each test was also preceded by a short and simple calibration procedure, lasting about 10 s and consisting in following a moving circle on the screen. Participants were then instructed to look at the cross when the blank screen was displayed, and to freely watch wherever they wanted when the images were presented. Each stimulus image was shown for 10 s.

Sixteen black-and-white pictures were employed in the experiments, which contained close-up faces of eight males and eight females. Half of the faces (four males and four females) were of famous persons (mostly actors and actresses), while the others were of people unknown to the observers. Images were mostly taken from the AR Face Database [70]. Examples of the stimulus image set are shown in figure 6.8.

The presentation order of the 16 images was random. Behind the eye tracker there was a wall painted in neutral grey and the illumination of the room was uniform and constant. Also, all images had similar grey-level distributions. On average, a single test session, including task explanation, device calibration, lasted a little more than 5 min.

A first set S_1 of tests was carried out with 88 participants. Of these, 36 were involved in a second session (with the same images), and 16 other participants

Figure 6.8: Examples of stimuli.

were also involved in a third test session. One hundred forty tests were therefore carried out in the three sessions. Time intervals between the first and the second session, and between the second and the third session, ranged from a minimum of 5 days to a maximum of 9 days.

A second set S_2 of tests, with 34 participants, was implemented after 1 year from S1. Ten observers in this group had been involved in S_1 as well. Also in this case, three sessions were organized: 17 observers out of 34 were involved in a second session (nine of whom had participated in S_1), and 13 took part in a third session (six of whom had participated in S_1). Sixty-four tests were therefore carried out in the three sessions. Time intervals between the first and the second session, and between the second and the third session, ranged from a minimum of 1 day to a maximum of 21 days.

6.2.2 Results

The accuracy of the GANT system has been assessed in terms of Area Under Curve (AUC - the area under the ROC curve), Equal Error Rate (EER) and Cumulative Match Curve (CMC). The ROC is a curve relating the Genuine Acceptance Rate and False Accepting Rate according to an acceptance threshold

δ varying in the range [0,1]. The Equal Error Rate represents a sort of steady state for the system, as it corresponds to the point where False Acceptance Rate equals False Recognition Rate. The Cumulative Match Score at a rank n of a biometric identification system represents the likelihood that the correct identity is returned by the system among its top n answers. Thus the CMC is a curve representing the CMS with the rank ranging from 1 to N, where N is the number of enrolled subjects into the system gallery. CMS(1) indicates the value at rank 1, namely the Recognition Rate. The fixation models have been subdivided into Gallery and Probe sets as follows:

- Gallery: the Gallery contains the fixation models obtained from data acquired in the first session (111 observers, where 87 observers comes from S_1 and 24 are the new observers coming from S_2. Observer #80 was discarded because the fixation points relative to one of the 16 images were missing);
- Probe I: the first probe set is composed of the fixation models obtained from data acquired in the second session (44 observers, where 36 comes from S_1 and eight are the new observers coming from S_2);
- Probe II: the second probe set is composed of the fixation models obtained from data acquired in the third session (23 observers, where 16 comes from S_1 and seven are the new observers coming from S_2).

We have carried out three kinds of experiments based on:

- Single features;
- Combined features;
- Weighted combined features.

For tests based on single features, we separately tested density, duration and arc features following the test scheme presented in Table 6.1.

For tests based on combined and weighted combined features, we tested all features combinations following the test scheme presented in Table 6.2. Before combining single features, scores have been normalized to the range [0, 1].

Feature	Experiment	Gallery	Probe
Density	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II

Table 6.1: Test scheme for single feature experiments.

Feature	Experiment	Gallery	Probe
Density-duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs-density	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs-duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Density-duration-arcs	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II

Table 6.2: Test scheme for combined feature experiments.

Experiment	EER	AUC	CMS(1)
Density I	0.3078	0.7893	0.2093
Density II	0.2785	0.7727	0.3043
Duration I	0.2783	0.8032	0.1860
Duration II	0.2628	0.7960	0.3043
Arcs I	0.3158	0.7330	0.1860
Arcs II	0.3484	0.7445	0.1739

Table 6.3: Single feature experiments results.

6.2.3 Single feature experiments

Results of the tests on single features show that the duration represents the best discriminating feature among the three, and in particular it has the best EER, AUC and CMS values. The results of the single feature experiments are summarized in Table 6.3.

The performance obtained when the duration is used as a discriminant feature is shown in figure 6.9.

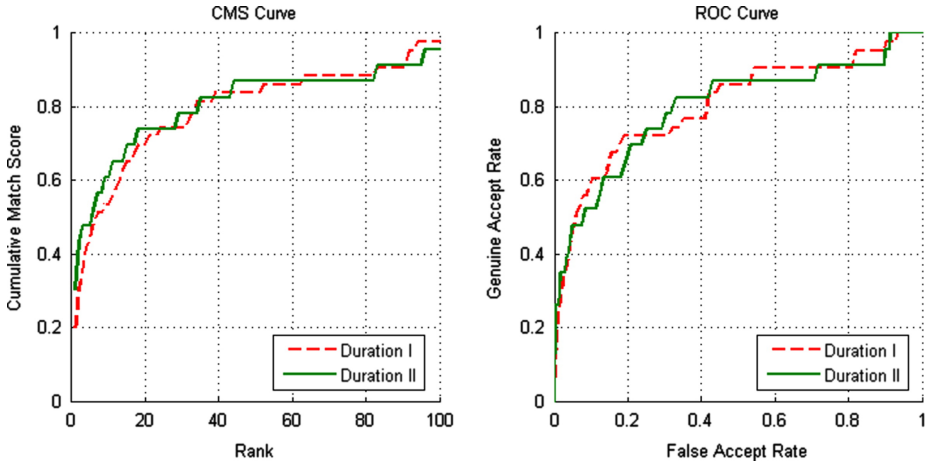

Figure 6.9: Performance graphs for the duration feature.

6.2.4 Combined features experiments

We tested all the combinations of the three features that are density, duration and observation path. The final score of each combination is computed by averaging scores obtained by GANT with single feature. The results of this experiment are reported in Table 6.4.

The graphs related to the combination of features which achieved the best performance are presented in figure 6.10.

Experiment	EER	AUC	CMS(1)
Den-Dur I	0.2878	0.7917	0.1395
Den-Dur II	0.2901	0.7922	0.3043
Arc-Den I	0.3093	0.7824	0.1628
Arc-Den II	0.2984	0.7879	0.2174
Arc-Dur I	0.2786	0.7892	0.1395
Arc-Dur II	0.2688	0.7928	0.3043
Arc-Den-Dur I	0.2859	0.7905	0.1395
Arc-Den-Dur II	0.2911	0.7935	0.3043

Table 6.4: Combined features experiments results.

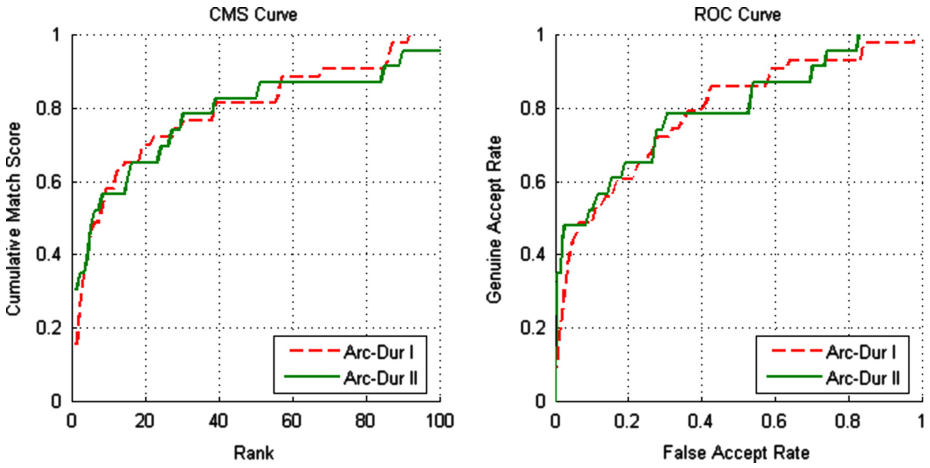

Figure 6.10: Performance graphs for the experiment on the combination of arcs and duration features.

Experiment	EER	AUC	CMS(1)
Den-Dur I	0.2819	0.7924	0.1395
Den-Dur II	0.2933	0.7922	0.3043
Arc-Den I	0.3047	0.7828	0.1628
Arc-Den II	0.3038	0.7867	0.1739
Arc-Dur I	0.2803	0.7906	0.1395
Arc-Dur II	0.2706	0.7917	0.3043
Arc-Den-Dur I	0.2719	0.7932	0.1395
Arc-Den-Dur II	0.2739	0.7930	0.3043

Table 6.5: Weighted combined features experiments results.

6.2.5 Weighted combined features experiments

We have finally tested all the weighted combinations of features in order to improve the performance of the system.

Weights have been chosen proportionally to the results of single features in order that their sum was equal to 1. Higher weights have been given to scores relative to duration features which performed better in our experiments, medium weights to scores relative to density features and lower weights to scores relative to arc features. The scores fusion has been performed through the weighted average. Results are reported in Table 6.5.

Figure 6.11 shows the graphs of the performance of the combination of features Arcs, Density and Duration which obtained the best results.

6.2.6 Comparison with previous experiments

As mentioned in Section 6.1.3, in [6] we presented "GAS - Gaze Analysis System", and a first series of experiments on a part of the database (88 individuals in total, the set S1 of tests described in Section 6.2.1), their performance graphs are shown in figure 6.12.

Table 6.6 shows the comparison, on the same dataset (88 observers), between the best results obtained in [6] (GAS) and the best results (duration feature) of the system presented in this work (GANT) in terms of EER, AUC and CMS.

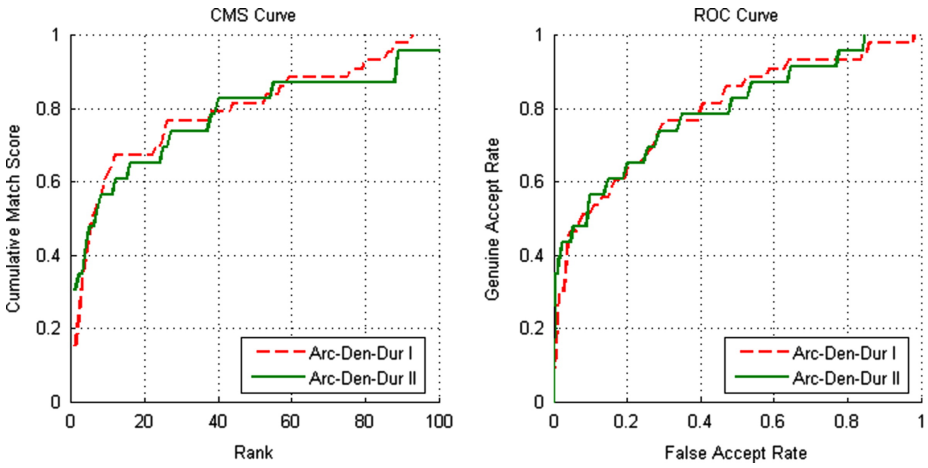

Figure 6.11: Performance graphs for the experiment on the weighted combination of arcs, density and duration features.

With GAS I and GANT I we indicate experiments in which the second session is used as Probe, while GAS II and GANT II indicate experiments in which the third session in used as Probe. With respect to the previous experiments, EER drops from 0.361 to 0.250 when using the third session as Probe set, while AUC and CMS(1) values have a significant improvement. The reason is that we provided a more robust representation of users' fixation model, increasing the number of AOIs and using the same standard subdivision in AOIs for all fixation sequences. Finally, in computing system performance we took into account users' observations over all the 16 faces, both for the Gallery and the Probe sets, while in the previous work testing was performed by considering, in turn, one of the 16 observations acquired during sessions 2 and 3 separately.

6.3 Conclusive remarks

The recent interest of the research community in soft biometric led us to design and test a novel technique for gaze analysis, namely GANT. The GANT

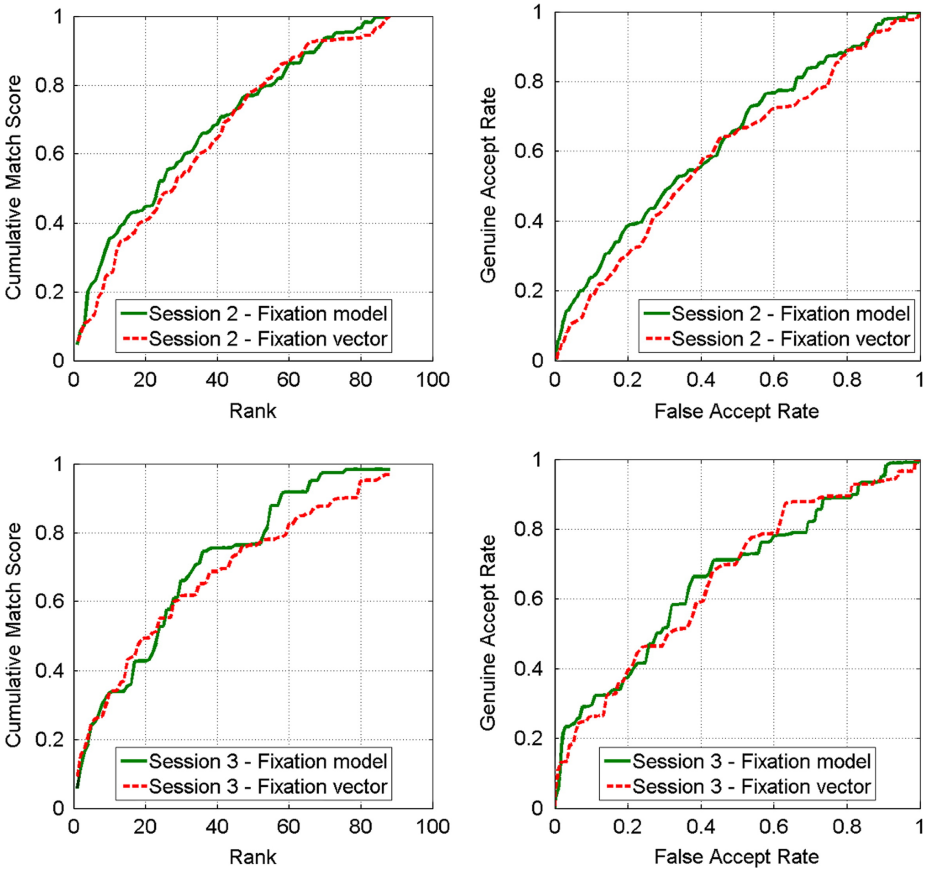

Figure 6.12: GAS performance.

Experiment	EER	AUC	CMS(1)
GAS I	0.424	0.6318	0.0476
GAS II	0.361	0.6617	0.0589
GANT I	0.224	0.8179	0.2286
GANT II	0.250	0.7901	0.3125

Table 6.6: Comparison between experiments presented in [6] (GAS) and best results of the system presented in this work (GANT).

approach, applied to a wide dataset composed of 112 volunteer observers acquired through the Tobii 1750 remote eye tracker, verified the conjecture that the way an individual looks at an image might be a personal distinctive feature. However, significant improvements are required, specifically to allow the application of gaze analysis in large-scale identification scenarios. We plan to include GANT in a multi-biometric framework including iris and face, in a first attempt by using a multi-view camera for acquisition and then we plan to study how to migrate this technology on mobile devices.

Chapter 7

Conclusions

In this Thesis, we presented our works mainly related to iris recognition on mobile devices. We first collected a database simulating real acquisition conditions via mobile devices and then we developed several techniques suitable for iris recognition on mobile devices, testing them on our database.

In Chapter 3, we presented the MICHE database and described in details the characteristics of the photos it contains. This database, and in particular its section *MICHE Iris*, allowed us to develop diverse techniques for iris segmentation and/or recognition on mobile devices, presented in this thesis. In addition, this database contains photos of the same irises acquired by different devices, allowing us to develop techniques for sensor recognition to be combined with iris recognition.

In Chapter 4, the *MICHE Iris* database is employed to test few techniques for iris segmentation suitable for images captured by mobile devices in unconstrained conditions. Iris segmentation is a very sensitive stage, an error in this phase can cause a chain reaction that will lead to a failure at the final stage (decision phase) of the system. In addition, this stage is made even more complex

because of the images employed that are affected by many and sometimes very strong noise factors. The results presented in section 4.2.4, show that BIRD in many case outperforms the other methods it is compared with, one of whom achieved the first place in NICE-I Challenge (see section 1.4.2).

In Chapter 5, three approaches for iris recognition are presented, one based on the combination of iris and the periocular area, one based on the combination of iris and face recognition, namely "multi-biometric" systems, and the third one based on the combination of iris and sensor recognition that we indicated as a "multi-modal" system. Thanks to the MICHE database it was possible, in particular, to replicate the real conditions of use of the latter technique, since to test it, a database of iris images acquired with different devices is needed. Both strategies, multi-biometric and multi-modal, have the aim of providing a more robust and secure system for human recognition on mobile devices.

Finally, in Chapter 6, we gave an overview on a new trending behavioural biometric trait: the gaze; and on its use for biometric recognition. Since the gaze is a biometric trait that can be acquired by recording eye movements, it is easily combinable with iris recognition. We also studied its applicability for gender and age classification in [5] and [9].

The two most challenging issues that came out from our studies are the iris segmentation and the development of a robust and secure mobile authentication system based on the iris biometry. We studied all the steps that make up an iris recognition system and we developed new algorithms and techniques specifically for iris recognition on mobile devices.

7.1 Future perspectives

Thanks to the knowledge gained through our studies, and the development of new solutions for iris recognition on mobile devices, we are now able to develop a complete system suitable for the daily use of biometric recognition on personal

smartphones.

The possible applications of such system include but are not limited to: home banking, e-commerce, home automation, border control, access control both in the sense of controlling access to restricted areas and also to access homes, cars or computers. Still to be investigated is which is the best item (e.g. biometric trait, physical object, password or code) to be combined with the iris in order to obtain a system with a good trade-off between ease-of-use and security.

Last but not the least, the study of demographic categorization, e.g. gender, age and ethnicity, has many possible applications, starting from commercial applications (customized advertising alerts), to more relevant application such as parental control.

Appendix A

Résumé en français

A.1 Introduction

La biométrie peut fournir un niveau de sécurité plus élevé par rapport à d'autres systèmes d'authentification basés sur des mots de passe ou des badges. Cependant, certaines caractéristiques de la biométrie peuvent changer sensiblement au fil du temps et les dispositifs utilisés pour les capturer peuvent se tromper ou peuvent avoir des difficultés d'acquisition de la biométrie, ce qui empêchent donc son utilisation.

Les biométries principalement utilisées pour la reconnaissance automatique des personnes sont les empreintes digitales et le visage. La première est très fiable, mais coûteux en calcul, tandis que la seconde nécessite un réglage bien contrôlé. Nous verrons que l'iris est plus fiable pour une identification par rapport à d'autres données biométriques, mais les applications sur le marché, jusqu'à aujourd'hui, sont limitées par la nécessité d'acquérir l'iris à une distance proche et d'obtenir la coopération de l'utilisateur. Pour cette raison, la recherche récente se focalise sur l'utilisation des systèmes de reconnaissance de l'iris en présence de bruit sur l'image obtenue, afin de développer des systèmes fiables qui peuvent acquérir l'iris à distance et avec peu de coopération de l'utilisateur.

A.1.1 Reconnaissance biométrique sur les appareils mobiles

La reconnaissance biométrique a été utilisée longtemps dans des espaces confinés, généralement à l'intérieur, où les opérations de sécurité exigeaient des systèmes de haute précision, par exemple dans les postes de police, les banques, les entreprises, les aéroports. Les activités de terrain, au contraire, exigent plus de flexibilité et de portabilité ce qui conduit au développement de dispositifs d'acquisition des traits biométriques et des algorithmes pour la reconnaissance biométrique dans des conditions moins contraintes. L'application de la reconnaissance biométrique "portable" est limitée dans des domaines spécifiques, par exemple pour le contrôle de l'immigration, et exige des dispositifs dédiés.

Pour étendre l'utilisation de la biométrie sur les appareils personnels, des scanners d'empreintes digitales ont été intégrés dans les ordinateurs portables ou les smartphones. Mais la reconnaissance biométrique sur les appareils personnels a été utilisée seulement pour un nombre limité de tâches, comme le déverrouillage d'écrans à l'aide des empreintes digitales au lieu de mots de passe.

Les activités décrites dans cette thèse se sont portées sur le développement de solutions pour la reconnaissance biométrique sur les appareils mobiles. En particulier, en raison des connaissances du doctorant, l'utilisation de la reconnaissance de l'iris a été étudiée. De nombreux aspects de la reconnaissance de l'iris sur les appareils mobiles ont été analysées à partir de l'étude des problèmes liés à l'acquisition des images de l'iris en utilisant des dispositifs mobiles. Les recherches ont conduit à la formulation d'un protocole d'acquisition et à la collection d'une base de données d'images d'iris, c'est-à-dire la base de données MICHE, puis à analyser les difficultés liées à la segmentation de l'iris sur les appareils mobiles, en explorant les avantages de combiner l'iris avec d'autres traits biométriques ou des éléments d'authentification, et enfin à étudier sur son éventuelle association avec un nouveau trait biométrique: le regard.

Il convient de noter que chaque travail réalisé, a été utilisé dans le développe-

ment d'autres travaux, par exemple, la base de données a été utilisée pour tester toutes les techniques mises au point, et les méthodes de segmentation développées ont été utilisées dans le cadre des systèmes de reconnaissance présentés.

A.1.2 L'authentification

L'authentification peut être effectuée en se basant sur un ou plusieurs des éléments suivants:

- Quelque chose que l'utilisateur connaît (par exemple, le mot de passe, numéro d'identification personnel (NIP), réponse secrète);
- Quelque chose que l'utilisateur dispose (par exemple, carte à puce, carte d'identité, jeton de sécurité, jeton logiciel);
- Ce que l'utilisateur est ou fait (par exemple les empreintes digitales, le visage, la démarche).

Le dernier est connu comme la biométrie et sera discuté en détail plus tard. Pour l'instant, nous voulons analyser brièvement le niveau de sécurité associé à chaque type d'authentification. En principe, il est intéressant de considérer que les mots de passe peuvent être oubliés ou obtenus par des personnes malintentionnées, des objets physiques tels que les badges et les documents d'identité peuvent être perdus ou volés, alors que données biométriques ne peuvent pas être volés et leur falsification est beaucoup plus compliquée (par exemple la chirurgie plastique). De plus, les systèmes les plus récents de reconnaissance biométrique intègrent des mécanismes pour reconnaître si la biométrie est fausse (détection de l'usurpation d'identité).

En considérant toutes les combinaisons possibles des trois facteurs d'authentification, nous obtenons le classement, du niveau de sécurité plus bas au niveau le plus élevé:

- Quelque chose que l'utilisateur connaît;
- Quelque chose que l'utilisateur dispose;

- Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose (par exemple la carte bancaire + NIP);
- Ce que l'utilisateur est ou fait;
- Quelque chose que l'utilisateur dispose + quelque chose que l'utilisateur est ou fait (par exemple passeport biométrique);
- Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur est ou fait;
- Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose + ce que l'utilisateur est ou fait.

A.1.3 La biométrie

L'authentification biométrique est le processus d'identification humaine par leurs caractéristiques physiologiques ou comportementales. Ces caractéristiques doivent avoir un caractère distinctif et mesurable afin d'effectuer une reconnaissance.

La reconnaissance peut être effectuée en mode de vérification (lorsque le sujet revendique une identité qui doit être vérifiée), ou en mode d'identification (un contre tous, quand il n'y a aucune réclamation préalable et le système doit retourner l'identité d'un sujet). Les données biométriques physiologiques comprennent les empreintes digitales, le visage, la géométrie de la main, la rétine, l'ADN, l'oreille et l'iris. Les comportementales sont liés au comportement particulier d'une personne et peuvent être influencées par l'humeur; ils comprennent la signature, la voix, et la démarche. Une bonne biométrie doit répondre aux caractéristiques suivantes: l'unicité, la permanence, la facilité d'utilisation, une bonne performance, la précision, un faible coût, une perception positive par les personnes. L'iris répond de manière optimale à la quasi-totalité d'entre eux [37].

A.1.4 La reconnaissance de l'iris

L'identification peut être considérée comme un problème de classification. Un trait biométrique peut être classifié de manière fiable seulement si la variabilité

entre les différentes instances d'une classe est inférieure à la variabilité entre les différentes classes. Par exemple, les images du même visage ont une forte variabilité (variabilité intra-classe) en raison d'expressions, ainsi que le visage est un objet en trois dimensions (3-D) dont l'image varie selon l'angle de vision, la pose, et l'illumination. D'autre part la variabilité inter-classe du visage est limitée parce que différents visages possèdent le même ensemble de caractéristiques de base, la même géométrie. Au contraire, la variabilité inter-classe de l'iris est énorme et la variabilité intra-classe est faible: l'iris peut être considéré comme un objet en deux dimensions et son image est relativement insensible à l'angle d'éclairage, et les changements dans l'angle de vision causent des transformations affines seulement. Même les distorsions non affines provoquées par la dilatation pupillaire sont facilement réversibles [11].

L'élément le plus faible de la reconnaissance de l'iris est que l'acquisition de l'iris, même si effectuée d'une manière sans contact, est perçue comme intrusive. Beaucoup de système utilise l'éclairage NIR (Near Infra Red). Ce type d'éclairage est utilisé parce qu'il n'est pas visible et permet d'éclairer les yeux sans ennuyer les utilisateurs. Cependant, même si les études confirment que l'exposition de quelques secondes aux rayons NIR n'endommage pas les yeux dans des conditions normales, on ne sait pas ce qui pourrait arriver pour les yeux ou la peau avec des pathologies préexistantes, ou si un sujet est accidentellement exposé aux rayons NIR pendant longtemps.

Presque tous les systèmes commerciaux, basés sur la lumière visible ou NIR, imposent aux utilisateurs de se tenir à une distance maximale de 1 m (généralement beaucoup moins), afin de capturer une image de l'iris de haute qualité. La nécessité de ces conditions standard et de la coopération des utilisateurs, limitent encore les domaines d'application pour la reconnaissance de l'iris. Par conséquent, de nouvelles techniques de reconnaissance de l'iris en présence de bruit ont été proposées. "Noisy Iris" se rapporte à la qualité des images de l'iris [43], ils peuvent présenter les problèmes suivants:

- Occlusions: paupières, cils, lunettes, cheveux, etc .;
- Réflexions;

- Taille différente;
- Basse résolution (à cause de l'appareil ou de la distance);
- Différentes couleurs dominantes dans les images de la même iris (dues à des conditions différentes lors de l'acquisition).

Ces problèmes peuvent survenir si la reconnaissance est effectuée sur des sujets à distance, en mouvement, inconscient de l'acquisition en cours, aussi bien s'il n'y a pas un éclairage standard, ou simplement si moins de coopération de l'utilisateur est souhaitée pour accélérer l'identification.

Les phases de la reconnaissance de l'iris en présence de bruit sont les mêmes utilisées dans les systèmes "traditionnels", même si elles nécessitent des approches différentes en raison des caractéristiques de l'image mentionnées précédemment. Ces phases sont: acquisition; segmentation; normalisation; codage; comparaison.

Acquisition: par rapport aux systèmes traditionnels, l'acquisition n'est pas nécessairement réalisée avec des appareils de haute qualité. Les images de l'iris peuvent être obtenues à partir de simples caméras, ou d'acquisition standard intégrée dans les ordinateurs ou les appareils mobiles. Les conditions d'acquisition (éclairage, distance, pose, etc.) ne sont pas strictement contrôlées, contrairement aux systèmes traditionnels.

Segmentation: c'est le processus d'identification des contours de l'iris. Dans les systèmes traditionnels ceci est une opération relativement simple, consistant à trouver deux cercles correspondant aux contours de la pupille et de la sclérotique. En présence de bruit, la segmentation de l'iris est beaucoup plus compliquée. Il doit tenir compte de la présence éventuelle d'occlusions ou de réflexions. De plus, la basse résolution ou la présence de bruit, rendent les frontières moins claires. Pour cette raison, les méthodes de segmentation de l'iris en présence de bruit mettent généralement en œuvre une phase de prétraitement dans laquelle un filtre de lissage (pour réduire le bruit) et/ou un filtre pour

améliorer les contours de l'iris, sont appliqués [36] [28].

Normalisation: dans les systèmes traditionnels, en raison de la condition d'acquisition contrôlée, il est seulement nécessaire de normaliser la forme de l'iris segmentée. La normalisation habituellement adoptée implique la transformation de coordonnées cartésiennes en coordonnées polaires. Si la couleur est prise en compte, la correction des couleurs, la normalisation de l'histogramme ou des opérations similaires, peuvent également être utiles.

Codage: cette phase produit un vecteur de caractéristiques, c'est-à-dire, une représentation compacte d'une image de l'iris. Dans les images de haute qualité même les minuscules détails de la texture de l'iris sont facilement visibles. Au contraire, les images bruitées peuvent présenter moins de caractéristiques à observer ou des altérations. Les méthodes adoptées pour l'extraction de caractéristiques dans les images d'iris bruitées analysent principalement la texture de l'iris, par exemple la distribution de la couleur, et peuvent également combiner un certain nombre d'opérateurs différentes[29].

Comparaison: ne dépend que de la nature des modèles utilisés.

A.2 La base de données biométrique MICHE

Afin de tester de manière fiable une approche, il est nécessaire d'avoir un ensemble de données qui reproduit des conditions réelles dans lesquelles l'approche doit être appliquée. La base de données décrite dans ce chapitre simule l'acquisition de l'iris et du visage sur les appareils mobiles. La base de donnée MICHE [30] a été recueillie à l'occasion de l'évaluation "Mobile Iris CHallenge Evaluation", à l'Université de Salerne, par Chiara Galdi (à cette époque doctorante à l'Université de Salerne, Italie) et Silvio Barra (à cette époque doctorant à l'Université de Cagliari, Italie), membres de Biplab (Biometric and image Processing Lab) qui a favorisé le Challenge et la collection de la base de données biométrique MICHE.

Les images ont été acquises avec trois appareils mobiles différents:

- iPhone 5 (ci-après IP5)

  Système d'exploitation: Apple iOS;

  Caméra avant: Caméra FaceTime HD avec 1,2 mégapixels;

  Caméra arrière: iSight avec 8 mégapixels.

- Samsung Galaxy S4 (ci-après GS4)

  Système d'exploitation: Google Android;

  Caméra avant: CMOS avec 2 mégapixels;

  Caméra arrière: CMOS avec 13 mégapixels.

- Samsung Galaxy Tablet II (ci-après de GT2)

  Système d'exploitation: Google Android;

  Caméra avant: VGA pour appel vidéo;

  Caméra arrière: 3 mégapixels.

La biométrie est très appropriée pour la reconnaissance des personnes sur les appareils mobiles, en fait, les utilisateurs sont habitués à utiliser la caméra frontale de leurs appareils mobiles personnels pour capturer des images d'eux-mêmes, qu'on appelle "selfie". Les sujets impliqués dans le processus d'acquisition

de la base de données sont invités à prendre des auto-photos de leur visage, leurs yeux et de leur iris, parfois avec les deux caméras et parfois avec la caméra frontale seulement. De plus amples détails sur la procédure d'acquisition sont donnés dans ce chapitre dans le paragraphe suivant.

A.2.1 Protocole d'acquisition

Étant donné que le but du processus d'acquisition est de réaliser une simulation réaliste du processus d'acquisition des données d'un système de reconnaissance biométrique, les utilisateurs sont libres de maintenir les appareils (smartphones ou tablettes) dans leurs mains, d'utiliser les commandes vocales (si disponible) et, pour les sujets qui portent des lunettes, il peuvent décider de les porter ou non lors de l'acquisition en fonction de ce qu'ils auraient fait au cours d'une utilisation réelle d'une telle application.

Le processus d'acquisition standard était ce qui suit:

1. Smartphones (GS4 and IP5):
   (a) **à l'intérieur:**
      - 4 photos du visage avec la caméra frontale;
      - 4 photos des avec la caméra arrière;
      - 4 photos de l'iris avec la caméra arrière;
      - 4 photos de l'iris avec la caméra frontale.
   (b) **en plein air:**
      - 4 photos de l'iris avec la caméra arrière;
      - 4 photos de l'iris avec la caméra frontale.
2. Tablet (GT2):
   (a) **à l'intérieur:**
      - 4 photos de l'iris avec la caméra frontale.
   (b) **en plein air:**

- 4 photos de l'iris avec la caméra frontale.

Il y a environ 56 images par personne (il y avait quelques exceptions). Pendant l'acquisition a l'intérieure, diverses sources de lumière artificielle, parfois combinés avec des sources de lumière naturelle, ont été utilisées. L'acquisition en plein air a lieu en utilisant la lumière naturelle seulement. Les images capturées sont affectées par différents facteurs de bruit.

A.3 Segmentation de l'iris sur les appareils mobiles

La reconnaissance de l'iris sur les appareils mobiles est une tâche difficile, en fait, par rapport à d'autres dispositifs d'acquisition de l'iris dédiés, généralement fixé sur un bureau ou sur un support, l'utilisation des dispositifs d'acquisition embarqués sur les smartphones introduit un certain nombre de facteurs de bruit pendant le processus d'acquisition de l'iris [41],en raison du fait que le dispositif est dans les mains de l'utilisateurs et qu'ils ont tendance à bouger.

Dans ce qui suit, deux méthodes de segmentation de l'iris sont présentées. Elles sont spécialement conçues pour la reconnaissance de l'iris dans des conditions difficiles et sur les appareils mobiles. Ces deux méthodes, ISIS et BIRD, sont ensuite utilisées dans les systèmes présentés dans le chapitre A.4 "Authentification multi-biométrique et multi-modale sur les appareils mobiles".

A.3.1 ISIS

ISIS (Iris Segmentation for Identification Systems) est un algorithme de segmentation de l'iris proposé par le Biplab - Biometric and Image Processing Lab [31], de l'Université de Salerne. Il a été spécialement conçu et mis en œuvre pour remédier aux problèmes liés à l'acquisition en conditions sous-contrôlé. Il est donc possible de l'utiliser sur les appareils mobiles. Il est robuste à la présence de reflets et a un temps de calcul limité. Il dispose de quatre phases principales:

- Pré-traitement;
- Localisation de la pupille;
- Linéarisation;
- Localisation du limbus.

A.3.2 BIRD

La méthode BIRD (watershed Based IRis Detection), a été présentée dans [2] avec une technique de segmentation de la zone périoculaire et une approche de reconnaissance basée sur la fusion de l'iris et de la zone périoculaire.

BIRD [2] BIRD est une technique pour les appareils mobiles, qui est une extension d'une technique présentée dans [1]. BIRD exploite l'utilisation de la transformation watershed pour identifier plus précisément le contour de l'iris et, par conséquent, pour obtenir un codage de l'iris plus précis pour la reconnaissance de l'iris.

Un aspect positif de la transformation watershed est que les contours délimitant les régions d'une image sont principalement placés là où des observateurs humains perçoivent. En fait, la transformation watershed est un processus de croissance généralement effectuée sur le gradient de l'image, dont les bords sont renforcés. Cette fonctionnalité devrait permettre de détecter correctement la limite du limbus (contour extérieur de l'iris). À son tour, une caractéristique négative est l'over-segmentation, c'est-à-dire, l'image peut être divisée en plusieurs parties qui sont remarquablement en nombre plus élevé que prévu. L'over-segmentation est particulièrement évident lorsque tous les minima régionaux dans le gradient de l'image sont considérés comme des semences pour le processus de croissance. Une stratégie commune pour remédier à cet inconvénient est d'adopter la fusion des régions et/ou la sélection des semences afin de réduire le nombre de régions. Cependant, dans le cas des images de l'œil, ces procédures de réduction des régions, ne peuvent pas être adoptées. Sinon, certains contours faibles entre la sclérotique et l'iris (yeux claires) ou entre les cils et l'iris (yeux sombres) pourraient ne plus être présents dans l'image segmentée.

BIRD effectue une binarisation de la transformation watershed pour obtenir une image où une grande partie du contour du limbus est mis en évidence. De cette façon BIRD est capable d'exploiter les aspects positifs de la transformation watershed indépendamment du problème de l'over-segmentation. Les contours de la région de premier plan sont alors donnés en entrée à un processus de

détection de cercle, qui vise à trouver le cercle qui se rapproche le plus au contour du limbus.

A.4 Authentification multi-biométrique et multi-modale sur les appareils mobiles

Pour améliorer la robustesse des systèmes de reconnaissance biométrique contre les attaques, il convient de combiner la biométrie avec un autre élément d'authentification (voir paragraphe A.1.2 "L'authentification"): quelque chose que l'utilisateur connaît (par exemple, le mot de passe, numéro d'identification personnel (NIP), réponse secrète); quelque chose que l'utilisateur dispose (par exemple, carte à puce, carte d'identité, jeton de sécurité, jeton logiciel, téléphone ou téléphone cellulaire); quelque chose que l'utilisateur est ou fait (biométrie). Dans les chapitres suivants deux stratégies principales sont présentées:

- Reconnaissance multi-biométrique. La reconnaissance de l'iris a été combinée avec un autre trait biométrique: la zone périoculaire, et le visage;
- Reconnaissance multi-modale. La reconnaissance de l'iris a été combinée avec «quelque chose que l'utilisateur dispose": le smartphone;

A.4.1 Combinaison de l'iris avec la zone périoculaire

Dans cette section, nous présentons une méthode de reconnaissance biométrique basée sur la fusion de l'iris et de la zone périoculaire. Les résultats de cette étude sont présentés dans la deuxième partie de l'article "BIRD: Watershed Based IRis Detection for mobile devices" [2].

Les informations concernant la position et la taille de l'iris (son centre et rayon) extraites par BIRD (voir la section A.3.2 "BIRD")constituent le point de départ pour la délimitation de la région périoculaire. Des études récentes [44] ont montré comment ce dernier pourrait être considéré comme une biométrie. Une fois que la région périoculaire est sélectionnée une transformation des coordonnées cartésiennes en coordonnées polaires est appliquée sur elle. De cette façon, il est possible d'appliquer à la zone péri-oculaire un processus d'extraction des caractéristiques similaire à celui utilisé pour l'iris. L'iris et la région périoculaire sont ensuite fusionnés au niveau du score grâce à un critère de somme simple afin d'augmenter la précision du système de reconnaissance.

A.4.2 Combinaison de l'iris avec le Visage

La reconnaissance de l'iris est combinée avec la reconnaissance du visage pour fournir un système robuste d'authentification. Les résultats de ces études ont été présentés dans l'article: Face and Iris Recognition for Mobile Engagement (FIRME)[10]. Ce système présente les importantes caractéristiques suivantes: anti-spoofing pour le visage; ré-identification continue; la sélection de meilleures modèles.

A.4.3 Combinaison de la biométrie avec l'hardwaremétrie

Dans cette section, un système d'authentification multi-modale qui combine une biométrie, en ce cas l'iris, avec un objet personnel appartenant à l'utilisateur, c'est-à-dire le smartphone, est présenté [8] [9]. Cette approche présente plusieurs avantages:

- Le couple utilisateur-appareil doit être authentifié, ce qui rend plus difficile le processus d'usurpation d'identité;
- Les deux processus de reconnaissance sont appliqués sur une seule photo de l'œil, capturée par l'utilisateur avec son téléphone;
- Bon compromis entre la précision et la facilité d'utilisation;
- Les performances de la reconnaissance de l'iris et, en particulier, de la reconnaissance du capteur, sont très élevées.

Le système que nous proposons est donc un système de reconnaissance multi-modale basé sur la combinaison de la reconnaissance du capteur (hardwaremétrie) et la reconnaissance de l'iris (biométrie), c'est-à-dire quelque chose que l'utilisateur dispose + ce que l'utilisateur est. Si nous analysons les niveaux de sécurité des systèmes d'authentification indiqués dans la figure A.1,nous pouvons voir que le degré de sécurité assuré par la combinaison de la biométrie avec un objet physique est plus élevé que lorsque la biométrie, seule, est utilise [1].

[1] http://www.ffiec.gov/pdf/authentication_guidance.pdf

Le deuxième aspect que nous abordons dans ce travail, est le problème de l'interopérabilité du capteur [55]. Ce problème se pose lorsque les données à comparer (par exemple, les images de l'œil) sont acquises par des capteurs différents et contiennent ainsi des différences en fonction des caractéristiques des capteurs. Comme nous le verrons plus tard, cela peut affecter les performances des algorithmes biométriques, car deux photos du même œil peuvent apparaître différentes parce que elles ont été capturées par des différents dispositifs. Notre approche peut être considérée comme un moyen de contourner le problème de l'interopérabilité du capteur, au lieu de se concentrer sur le développement d'un algorithme capable de fonctionner indépendamment du capteur utilisé, nous misons sur les différences introduites par différents capteurs sur les photos pour obtenir un système de reconnaissance plus robuste.

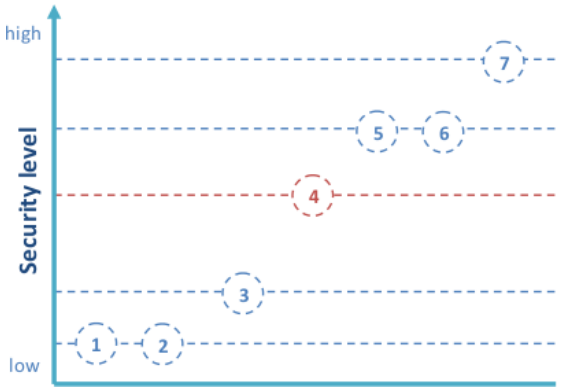

Figure A.1: Niveaux de sécurité. (1) Quelque chose que l'utilisateur connaît; (2) Quelque chose que l'utilisateur dispose; (3) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose; (4) Ce que l'utilisateur est ou fait; (5) Quelque chose que l'utilisateur dispose + quelque chose que l'utilisateur est ou fait; (6) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur est ou fait; (7) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose + ce que l'utilisateur est ou fait.

A.5 Analyse du regard

Dans ce chapitre, une nouvelle technique pour l'analyse du regard, à savoir GANT [6] [4] est présenté, en exploitant une représentation à base de graphes de points de fixation obtenus par un oculomètre lors de l'interaction homme-machine. L'objectif principal est de démontrer la conjecture que la façon dont une personne regarde une image pourrait être un trait distinctif personnel.

Les résultats expérimentaux obtenus sont très intéressants, confirmant que le regard peut être utilisé comme une biométrie comportementale pour la reconnaissance humaine, et d'autres études ont été effectuées afin de tester l'analyse du regard pour la classification du genre et de l'âge [5] [9] .

Même si ce sujet diffère des autres abordés dans cette thèse, il a été inclus car le regard est extrait de l'analyse des yeux (leurs mouvements), et peut être facilement combiné avec la reconnaissance de l'iris, et en particulier, le deux pourraient être acquis par un seul capteur à la fois. Ceci est un aspect important lorsque l'on considère le développement de systèmes biométriques faciles à utiliser et nécessitant moins de coopération de l'utilisateur.

A.6 Conclusions

Dans cette thèse, nous avons présenté nos études, principalement liées à reconnaissance de l'iris sur les appareils mobiles. Nous avons d'abord recueilli une base de données simulant les conditions réelles d'acquisition via les appareils mobiles, puis nous avons développé plusieurs techniques pour la reconnaissance de l'iris sur les appareils mobiles et nous les avons testées sur notre base de données.

List of Figures

1.1 Security levels. (1) Something the user knows; (2) Something the user has; (3) Something the user knows + something the user has; (4) Something the user is or does; (5) Something the user has + something the user is or does; (6) Something the user knows + something the user is or does; (7) Something the user knows + something the user has + something the user is or does. 15
1.2 Iris recognition phases. 19
1.3 MICHE I iris images example. (a) Images acquired with Samsung Galaxy S4 rear camera (left) and front camera (right); (b) Image captured with Samsung Galaxy Tab 2 front camera. 27
1.4 *MICHE II*: (a) the binary mask; (b) the normalised mask of the iris region; (c) the normalised RGB iris extracted from the image. 31

3.1 *MICHE Iris* images example. First row: pictures acquired by GS4; Second row: pictures acquired by IP5; Third row: pictures acquired by GT2 40
3.2 *MICHE Iris* acquisition modalities. 41
3.3 *MICHE Fake* images example. 42
3.4 *MICHE Face* images example. 42

3.5 *MICHE Eyes* images example. 43
3.6 MICHE I iris noise factors examples: (a) out-of-focus iris image; (b-g) examples of images affected by strong light/shadows; (h) eyelids and eyelashes occlusion; (i) hairs occlusion; (j) out-of-iris image; (k-m) off-angle iris image; (m-o) partial captured iris; (p-q) strong make-up; (r) specular reflections; (s-t) diffuse reflection on eyeglasses. 45

4.1 Illustration of ISIS algorithm. 51
4.2 Colour correction. 53
4.3 a) resized and smoothed image; b) gradient image; c) watershed transform; d) result of region merging. 54
4.4 a) colour quantized image; b) binarized image; c) foreground contours; d) best fitting circle (red). 58
4.5 a) circle detected by circle fitting (purple) superimposed on the watershed transform; b) watershed regions totally overlapping the circle (red and blue) and partially overlapping the circle (green); c) red and green curves denote limbus and pupil boundaries. 59
4.6 Irises segmented by BIRD. 61
4.7 Segmentation accuracy measured in terms of percentage of error with respect to manual segmentation for the three tested approaches (ISIS, NICE-I, and BIRD) on probe/gallery images without colour/illumination correction. 63
4.8 Segmentation accuracy measured in terms of percentage of error with respect to manual segmentation for the three tested approaches (ISIS, NICE-I, and BIRD) on probe/gallery images with colour/illumination correction. 64

5.1 Periocular segmentation. 71
5.2 Cartesian to polar coordinates mapping of the periocular area. . 71
5.3 Cumulative Sums algorithm illustration. 73

5.4 Recognition accuracy measured in terms of decidability, EER and AUC for the three tested approaches (ISIS, NICE-I and BIRD) on probe/gallery images acquired by the same mobile device. . . 75
5.5 Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach (without periocular information) on probe/gallery images acquired by different mobile devices. 76
5.6 Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach (with periocular information) on probe/gallery images acquired by different mobile devices. 76
5.7 Recognition accuracy measured in terms of decidability, EER and AUC for the BIRD approach on probe/gallery images captured in different acquisition settings (outdoor/indoor). 76
5.8 Architecture of FIRME [10]. 78
5.9 Localization of reference points. 79
5.10 Detection of a real live user. 82
5.11 Best template selection. 83
5.12 Performance of face and iris biometrics with indoor probes. . . . 91
5.13 Performance of face and iris with outdoor probes. 92
5.14 Performance of FIRME with indoor probes. ϕ_1, ϕ_2, and ϕ_3 indicate the confidence values *relative distance*, *density ratio*, and a combination of the two, respectively. 92
5.15 Performance of FIRME with outdoor probes. ϕ_1, ϕ_2, and ϕ_3 indicate the confidence values *relative distance*, *density ratio*, and a combination of the two, respectively. 93
5.16 Authentication systems security levels: (1) Something the user knows; (2) Something the user has; (3) Something the user knows + something the user has; (4) Something the user is or does; (5) Something the user has + something the user is or does; (6) Something the user knows + something the user is or does; (7) Something the user knows + something the user has + something the user is or does. 95

5.17 Denoising process: (a) original image; (b) original wavelet coefficients; (c) local variance; (d) selection of the minimum variance; (e) denoised wavelet coefficients. 97

5.18 Sensor Pattern Noise enhancing: (a) original image; (b) the SPN extracted from the image contains image details (e.g. hairs, part of the eyeglasses frame); (c) ENSP, after the enhancing step the influence of image details is mitigated. 98

5.19 Sensor recognition performances. 103

5.20 Sensor recognition performance: experiment on same camera model. 104

5.21 Sensor recognition performances: in red the curves relative to the case in which the RSPN is extracted from blue sky images and in blue the curves relative to the case in which the RSPN is extracted from any kind of images captured by the sensor. 105

5.22 Iris recognition performances affected by sensor interoperability problem. 106

5.23 Fusion at feature level performances. 107

5.24 Fusion at score level performances. 108

5.25 Iris recognition performances on outdoor images. 109

5.26 Sensor recognition performances on outdoor images. 109

6.1 In the left image, the red point at coordinates (280, 420) is exactly in the centre of the right eye, while in the right image the point is on the right cheekbone. 115

6.2 dx (left) and dy (right) distances. M indicates the eyes middle point. 116

6.3 Cloud of normalized fixation points of observer #15 in session 1: (a) observation points on four different face images; (b) merging of observation points coming from all the 16 observed face images. 117

6.4 STASM output on subject #3. 118

6.5 Density graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The size of red circles indicates the weight associated with the node. Black bordered circles indicate a weight > 95 (fixation points). 120
6.6 Duration graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The size of red circles indicates the weight associated with the node. Black bordered circles indicate a weight > 45,000 (milliseconds). 121
6.7 Weighted arcs graphs of four different observers: observer #10 (top-left); observer #41 (top-right); observer #56 (bottom-left); observer #86 (bottom-right). The thickness of arcs specifies the associated weight. 122
6.8 Examples of stimuli. 125
6.9 Performance graphs for the duration feature. 128
6.10 Performance graphs for the experiment on the combination of arcs and duration features. 129
6.11 Performance graphs for the experiment on the weighted combination of arcs, density and duration features. 131
6.12 GAS performance. 132

A.1 Niveaux de sécurité. (1) Quelque chose que l'utilisateur connaît; (2) Quelque chose que l'utilisateur dispose; (3) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose; (4) Ce que l'utilisateur est ou fait; (5) Quelque chose que l'utilisateur dispose + quelque chose que l'utilisateur est ou fait; (6) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur est ou fait; (7) Quelque chose que l'utilisateur connaît + quelque chose que l'utilisateur dispose + ce que l'utilisateur est ou fait. . 154

List of Tables

1.1 UBIRIS.v2 imaging framework 23

5.1 Experimental results summary. 110

6.1 Test scheme for single feature experiments. 127
6.2 Test scheme for combined feature experiments. 127
6.3 Single feature experiments results. 127
6.4 Combined features experiments results. 129
6.5 Weighted combined features experiments results. 130
6.6 Comparison between experiments presented in [6] (GAS) and best results of the system presented in this work (GANT). 132

Bibliography

[1] **M. Frucci, C. Galdi, M. Nappi, D. Riccio, G. Sanniti di Baja, *IDEM: Iris DEtection on Mobile devices*, 22nd International Conference on Pattern Recognition, IEEE(2014), pp. 1752-1757.**

[2] **A. F. Abate, M. Frucci, C. Galdi, D. Riccio, *BIRD: Watershed Based IRis Detection for mobile devices*, Pattern Recognition Letters, Volume 57, 1 May 2015, pp. 43-51, ISSN 0167-8655.**

[3] **S. Barra, M. De Marsico, C. Galdi, D. Riccio, H. Wechsler, *FAME: Face Authentication for Mobile Encounter*, Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2013 IEEE Workshop on, pp. 1-7.**

[4] **V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio, *GANT: Gaze analysis technique for human identification*, Pattern Recognition, Volume 48, Issue 4, April 2015, pp. 1027-1038.**

[5] **V. Cantoni, C. Galdi, M. Nappi, M. Porta, H. Wechsler, *Gender and age categorization using gaze analysis*, Proceedings of the 10th International Conference on Signal Image Technology & Internet Systems (SITIS 2014), Marrakech, Morocco (November 23-27, 2014), pp. 574-579.**

[6] **C. Galdi, M. Nappi, D. Riccio, V. Cantoni, M. Porta, *A New Gaze Analysis Based Soft-Biometric*, Mexican Conference on Pattern Recognition (MCPR) 2013: 136-144, 2013.**

[7] **C. Galdi, M. Nappi, J.-L. Dugelay, *Combining Hardwaremetry and Biometry for Human Authentication via Smartphones*, Image Analysis and Processing - ICIAP 2015, Volume 9280 of the series Lecture Notes in Computer Science, pp. 406-416.**

[8] **C. Galdi, M. Nappi, J.-L. Dugelay, *Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity*, Pattern Recognition Letters, Available online 3 October 2015, ISSN 0167-8655, doi:10.1016/j.patrec.2015.09.009.**

[9] **C. Galdi, H. Wechsler, V. Cantoni, M. Porta, M. Nappi, *Towards demographic categorization using gaze analysis*, Pattern Recognition Letters, Available online 16 September 2015, ISSN 0167-8655, doi:10.1016/j.patrec.2015.08.018.**

[10] **M. De Marsico, C. Galdi, M. Nappi, D. Riccio, "FIRME: face and iris recognition for mobile engagement", Image Vis. Comput. (2014) Volume 32, Issue 12, December 2014, pp. 1161-1172.**

[11] J. Daugman, *How Iris Recognition Works*, IEEE Transactions on Systems for Video Technology, vol. 14, no. 1, pp. 21-30, 2004.

[12] C. Nickel, T. Wirtl, C. Busch, *Authentication of smartphone users based on the way they walk using k-NN algorithm.*, Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on. IEEE, 2012. pp. 16-20.

[13] L. JAIN, J. V. Monaco, M. J. Coakley, C. C. Tappert, *Passcode keystroke biometric performance on smartphone touchscreens is superior to that on hardware keyboards.* International Journal of Research in Computer Applications & Information Technology, 2014, Vol2, Issue 4, pp. 29-33.

[14] M. Conti, I. Zachia-Zlatea, B. Crispo, *Mind how you answer me!: transparently authenticating the user of a smartphone when answering or placing a call.*, Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security. ACM, 2011. pp. 249-259.

[15] P. N. A. Fahmi, E. Kodirov, C. Deok-Jai, L. Guee-Sang, *Implicit authentication based on ear shape biometrics using smartphone camera during a call*, Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on. IEEE, 2012, pp. 2272-2276.

[16] T. K. Mohanta, S. Mohapatra, *Development of Multimodal Biometric Framework for Smartphone Authentication System*, International Journal of Computer Applications, 2014, pp. 102-7.

[17] R. Raghavendra, K. B. Raja, A. Pflug, Bian Yang, C. Busch, *3d face reconstruction and multimodal person identification from video captured using smartphone camera*, Technologies for Homeland Security (HST), 2013 IEEE International Conference on. IEEE, 2013. pp. 552-557.

[18] P. Tresadern, T.F. Cootes, N. Poh, P. Matejka, A. Hadid, C. Levy, C. McCool, S. Marcel, *Mobile biometrics: combined face and voice verification for a mobile platform*, IEEE Pervasive Comput., 12 (1) (2013), pp. 79-87.

[19] M.A. Sasse, U. Bilting, C.-D. Schulz, T. Turletti, *Remote seminars through multimedia conferencing: experiences from the Mice project*, B.R. Platter, J.P.A. Kiers (Eds.), Proc. of INET'94/JENC5 Conference, 1994, (251-1 - 251-8).

[20] D.H. Cho, K.R. Park, D.W. Rhee, Y.G. Kim, J.H. Yang, *Pupil and iris localization for iris recognition in mobile phones*, Proceedings of International Conference on Software Engineering, Networking and Parallel/Distributed Computing (SNPD) (2006), pp. 197-201.

[21] D.H. Cho, K.R. Park, D.W. Rhee, *Real-time iris localization for iris recognition in cellular phone*, Proceedings of International Conference on

Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (2005), pp. 254-259.

[22] D.S. Jeong, H.A. Park, K.R. Park, J. Kim, *Iris recognition in mobile phone based on adaptive Gabor filter*, Proceedings of International Conference on Advances on Biometrics (ICB'06), Volume 3832 of Lecture Notes in Computer Science (2006), pp. 457-463.

[23] K.R. Park, H.A. Park, B.J. Kang, E.C. Lee, D.S. Jeong, *A study on iris localization and recognition on mobile phones*, EURASIP J. Adv. Signal Process (2008), pp. 1-12.

[24] J.S. Kang, *Mobile iris recognition systems: an emerging biometric technology*, Proceedings of International Conference on Computational Science (ICCS 2010), Procedia Computer Science (May 2010), pp. 475-484.

[25] S. Barra, A. Casanova, F. Narducci, S. Ricciardi, *Ubiquitous iris recognition by means of mobile devices*, Pattern Recognit. Lett., 57 (2015), pp. 66-73.

[26] A. F. Abate, M. Nappi, F. Narducci, S. Ricciardi, *Fast iris recognition on smartphone by means of spatial histograms*, Biometric Authentication, Lecture Notes in Computer Science, Springer International Publishing (2014), pp. 66-74.

[27] M. De Marsico, C.Galdi, M. Nappi, D. Riccio, G. Mastronardi, *Riconoscimento dell'iride in condizioni critiche*, Mondo Digitale - rassegna critica del settore ICT, ANNO XIII N.53 - Settembre 2014.

[28] M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, *Iris segmentation using pupil location, linearization, and limbus boundary reconstruction in ambient intelligent environments*, Journal of Ambient Intelligence and Humanized Computing 2(2), pp. 153-162, 2011.

[29] M. De Marsico, M. Nappi, D. Riccio, *Noisy Iris Recognition Integrated Scheme*, Pattern Recognition Letters 33(8): pp. 1006-1011, 2012.

[30] M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, *Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols*, Pattern Recognition Letters, Vol. 57, pp. 17-23, 2015.

[31] M. De Marsico, M. Nappi, D. Riccio, *IS_IS: Iris Segmentation for Identification Systems*, Pattern Recognition, 20th International Conference on Pattern Recognition, 2010, pp. 2857-2860.

[32] M. De Marsico, M. Nappi, D. Riccio, *Face: face analysis for Commercial Entities*, 17th IEEE International Conference on Image Processing (ICIP '10), Honk Kong, China, 2010, pp. 1597-1600.

[33] M. De Marsico, M. Nappi, D. Riccio, J.-L. Dugelay, *Moving face spoofing detection via 3D projective invariants*, in 5th IAPR International Conference on Biometrics (ICB '12), New Delhi, India, 2012, vol., no., pp. 73-78.

[34] M. De Marsico, M. Nappi, D. Riccio, *ES-RU: an entropy based rule to select representative templates in face surveillance*, Multimedia Tools and Applications November 2014, Volume 73, Issue 1, pp. 109-128.

[35] M. De Marsico, M. Nappi, D. Riccio, G. Tortora, *NABS: Novel Approaches for Biometric Systems*, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.41, no.4, pp. 481-493, July 2011.

[36] M. Frucci, M. Nappi, D. Riccio, G. Sanniti di Baja, *Using the Watershed Transform for Iris Detection*, ICIAP (2) 2013, pp. 269-278, 2013.

[37] M. Nappi, D. Riccio, *Moderne Tecniche di Elaborazione di Immagini e Biometria*, C.U.A. "Cooperativa Universitaria Athena", 2008.

[38] M. Nappi, H. Wechsler, *Robust re-identification using randomness and statistical learning: Quo vadis*, Pattern Recognition Letters 33(14), pp. 1820-1827, 2012.

[39] R. Parashar, S. Joshi, *Comparative Study of Iris Databases and UBIRIS Database for Iris Recognition Methods for Non-Cooperative Environment*,

International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 5, 2012.

[40] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, W. Todd Scruggs, *The Iris Challenge Evaluation 2005*, IEEE Second International Conference on Biometrics: Theory, Applications and Systems, 2008.

[41] H. Proença, L. A. Alexandre, *The NICE.I: Noisy Iris Challenge Evaluation - Part I*, IEEE First International Conference on Biometrics: Theory, Applications and Systems, 2007.

[42] H. Proença, S. Filipe, R. Santos, J. Oliveira, L. A. Alexandre, *The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-the-Move and At-a-Distance*, IEEE Trans. Pattern Anal. Mach. Intell. 32(8), pp. 1529-1535, 2010.

[43] H. Proença, *Quality Assessment of Degraded Iris Images Acquired in the Visible Wavelength*, IEEE Transactions on Information Forensics and Security 6(1), pp. 82-95, 2011.

[44] H. Proença, L. A. Alexandre, *Toward Covert Iris Biometric Recognition: Experimental Results From the NICE Contests*, IEEE Transactions on Information Forensics and Security 7(2), pp. 798-808, 2012.

[45] H. Proença, L. A. Alexandre, *Introduction to the Special Issue on the Recognition of Visible Wavelength Iris Images Captured At-a-distance and On-the-move*, Pattern Recognition Letters 33(8), pp. 963-964, 2012.

[46] *Special Issues on Mobile Iris Challenge Evaluation (MICHE I and II)*, http://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/special-issues-on-mobile-iris-challenge-evaluation/

[47] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, A. Ross, *On the Fusion of Periocular and Iris Biometrics in Non-ideal Imagery*, Pattern Recognition (ICPR), 2010 20th International Conference on , vol., no., pp. 201-204, 23-26 Aug. 2010.

[48] D. L. Woodard, S. J. Pundlik, J. R. Lyle, P. E. Miller, *Periocular Region Appearance Cues for Biometric Identification*, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2010, pp.162-169.

[49] J.B.T.M. Roerdink, A. Meijster, *The watershed transform: definitions, algorithms and parallelization strategies*, Fundamenta Informaticae, vol.41, no.1-2, pp. 187-228, 2001.

[50] S. Bharadwaj, H. S. Bhatt, M. Vatsa, R. Singh, *Periocular biometrics: When iris recog-nition fails*, IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS) 2010, pp.1-6.

[51] E. Raj, M. Chirchi, R.D. Kharadkar, *Biometric Iris Recognition for Person Identification using Cumulative Sum Algorithm*, International Journal of Scientific & Engineering Research, vol. 3, no. 5, May 2012.

[52] Tan, T., He, Z., Sun, Z., *Efficient and Robust Segmentation of Noisy Iris Images for Non-Cooperative Iris Recognition*, Image and Vision Computing 28, pp. 223-230, 2010.

[53] N. Srinivas, K. Veeramachaneni, L. A. Osadciw, *Fusing Correlated Data from Multiple Classifiers for Improved Biometric Verification*, 12th International Conference on Information Fusion, pp. 1504-1511, 2009.

[54] K. Kryszczuk, J.Richiardi, P. Prodanov, and A. Drygajlo, *Reliability-Based Decision Fusion in Multimodal Biometric Verification Systems*, EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 86572, pp. 1-9, 2007.

[55] A. Ross, A. Jain, *Biometric Sensor Interoperability: A Case Study In Fingerprints*, International ECCV Workshop on Biometric Authentication (BioAW), (Prague, Czech Republic), Springer Publishers, May 2004, LNCS Vol. 3087, pp. 134-145.

[56] A. Ross and A. K. Jain, *Information Fusion in Biometrics*, in Pattern Recognition Letters, Vol. 24, No. 13, pp.2115-2125, September, 2003.

[57] C.-T. Li, *Source camera identification using enhanced sensor pattern noise*, IEEE Trans-actions on Information Forensics and Security 5(2): pp. 280-287, 2010.

[58] J. Lukás, J. Fridrich, and M. Goljan, *Digital camera identification from sensor pattern noise*, IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 205-214, Jun. 2006.

[59] M. Chen, J. Fridrich, M. Goljan, J. Lukás, *Determining image origin and integrity using sensor noise*, IEEE Trans. Inf. Forensics Security, vol. 3, no. 1, Mar. 2008, pp. 74-90.

[60] A. Jain, K. Nandakumar, A. Ross, *Score normalization in multimodal biometric systems*, Pattern Recognition, Volume 38, Issue 12, December 2005, pp. 2270-2285.

[61] Anil K. Jain, Arun Ross. *Multibiometric systems*, Commun. ACM, Vol. 47, n. 1, pp. 34-40, January 2004

[62] F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti, W.A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York (1986).

[63] R. Cappelli, D. Maio, D. Maltoni, *Combining fingerprint classifiers*, Proceedings of First International Workshop on Multiple Classifier Systems (2000), pp. 351-361.

[64] J. A. Redi, W. Taktak, and J.-L. Dugelay, *Digital image forensics: a booklet for beginners*, Multimedia Tools and Applications, vol. 51, no. 1, 2011, pp. 133-162.

[65] S. Milborrow and F. Nicolls, *Locating facial features with an extended active shape model*, Proceedings of the Tenth European Conference on Computer Vision (ECCV '08), 2008, pp. 504-513.

[66] P. Viola and M. Jones, *Rapid object detection using a boosted cascade of simple features*, Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001), 2001, pp. 511-518.

[67] H.A. Rowley, S. Baluja, T. Kanade, *Neural network-based face detection*, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998), pp. 23-38.

[68] I. Rigas, G. Economou, S. Fotopoulos, *Biometric identification based on the eye movements and graph matching techniques*, Pattern Recognition Letters, 33 (6) (2012), pp. 786-792.

[69] G.H. Golub, C.F. Van Loan, *Matrix Computations*, third ed., Johns Hopkins, Baltimore, MD, 1996.

[70] A.M. Martinez and R. Benavente. *The AR face database*. CVC Technical Report n.24, June 1998.

[71] K. Etemad, R. Chellappa, *Discriminant Analysis for Recognition of Human Face Images*, Journal of the Optical Society of America A, Vol. 14, No. 8, pp. 1724-1733, Aug. 1997.

[72] G. Taubin, *Estimation Of Planar Curves, Surfaces And Nonplanar Space Curves Defined By Implicit Equations, With Applications To Edge And Range Image Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 1115-1138, 1991.

[73] A. Shashua, T. Riklin-Raviv, *The quotient image: Class-based re-rendering and recognition with varying illumination*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23(2), pp. 129-139, 2001.

[74] H. Wang; S. Z. Li, Y. Wang, and J. Zhang, *Self quotient image for face recognition*, International Conference on Image Processing (ICIP '04), Singapore, 2004, pp. 1397-1400.