DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

XVIII CICLO

UNIVERSITY OF CAGLIARI

Design of a Multi-biometric Platform, based on physical traits and physiological measures: Face, Iris, Ear, ECG and EEG.

Author:
Silvio BARRA

Ph.D. Chairman:
Prof. Michele Pinna

Supervisor:
Prof. Andrea CASANOVA

Department of Mathematics and Computer Science

Declaration of Authorship

I, Silvio BARRA, declare that this thesis titled, "Design of a Multi-biometric Platform, based on physical traits and physiological measures: Face, Iris, Ear, ECG and EEG." and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

UNIVERSITY OF CAGLIARI

Abstract

Faculty of Science
Department of Mathematics and Computer Science

Doctor of Philosophy

Design of a Multi-biometric Platform, based on physical traits and physiological measures: Face, Iris, Ear, ECG and EEG.

by Silvio BARRA

Security and safety is one the main concerns both for governments and for private companies in the last years so raising growing interests and investments in the area of biometric recognition and video surveillance, especially after the sad happenings of September 2001. Outlays assessments of the U.S. government for the years 2001-2005 estimate that the homeland security spending climbed from $56.0 billions of dollars in 2001 to almost $100 billion of 2005. In this lapse of time, new pattern recognition techniques have been developed and, even more important, new biometric traits have been investigated and refined; besides the well-known physical and behavioral characteristics, also physiological measures have been studied, so providing more features to enhance discrimination capabilities of individuals. This dissertation proposes the design of a multi-modal biometric platform, FAIRY, based on the following biometric traits: ear, face, iris EEG and ECG signals. In the thesis the modular architecture of the platform has been presented, together with the results obtained for the solution to the recognition problems related to the different biometrics and their possible fusion. Finally, an analysis of the pattern recognition issues concerning the area of videosurveillance has been discussed.

Contents

List of Figures

List of Tables

List of Abbreviations

BICA	Block Independent Component Analysis
CMS	Cumulative Match Score
ECG	ElectroCcardioGram
EEG	ElectroEncephaloGram
EER	Equal Error Rate
EMD	Empirical Mode Decomposition
EBGM	Elastic Bunch Graph Matching
FAIRY	Face eAr IRis phYsiological measures
FAR	False Acceptance Rate
FRR	False Rejection Rate
FBG	Face Bunch Graph
GAR	Genuine Accept Rate
GMM	Gaussian Mixture Model
LDA	Linear Discriminant Analysis
LBP	Local Binary Pattern
LPP	Locality Preserving Projections
LR	Likelihood Ratio
MBP	Multi Biometric Platform
MIMA	Mutual Information Multi-Anchor Tracking
OLPP	OrthogonalLocality Preserving Projections
PCA	Principal Ccomponent Analysis
PIE	Pose Illumination Expression
PTZ	Pan Tilt Zoom
QLS	Quasi Linear Sigmoid
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RR	Recognition Rate
SRR	System Response Reliability
WCS	World Coordinates System

Dedico questo lavoro di tesi alla mia famiglia, a mia nonna Rosa e a tutte le persone che sono state presenti e che mi hanno permesso di superare con azioni, o solo parole, i momenti più complicati di questi ultimi tre anni. Erika, Paolo, Tozzi, Alessandro, Claudio, Gabbo, Fabrizia e tutti i VERI amici, vecchi e nuovi. I ringraziamenti vanno a loro e a tutti i professori che mi hanno seguito e con cui ho avuto il piacere e l'onore di lavorare. Innanzitutto, il professore Casanova, che mi ha aperto nuovi spunti di ricerca, indirizzandomi verso l'ambito medico che fino a poco tempo fa mi era totalmente ignoto; i Professori Nappi e Riccio che sono stati guide e confidenti nei momenti in cui ero lontano da casa e sono stati sempre pronti ad aiutarmi qualora ne avessi avuto il bisogno. Un ringraziamento speciale va al mio collega Fabio Narducci, con cui ho trascorso sei mesi di internship in Portogallo e ci sostenevamo l'un l'altro nei momenti più duri, vista anche la lontananza da casa. A tal proposito, ringrazio i colleghi del SOCIALab dell'Università di Beira Interior, Gil, Joao, Juan, Silvio e Emmanuel ed il professor Proenca per aver messo al mio servizio la sua conoscenza. Ringraziamenti vanno anche al professore Fenu e al professore Pinna dell'università di Cagliari per la loro gentilezza. L'ultimo ringraziamento, ma non per importanza, va alla musa che è sempre stata presente in ogni lavoro, paper e pensiero: la Musica.

Chapter 1

Introduction

The homeland security has always been one of the main concerns of most governments all over the world. The need of protecting a country and its boundaries constantly leads to increase the security level and to improve the tools to guarantee it. For this reason, increasing funds are allocated for the defense of several sensible areas: nowadays, customs, airports and borders can count on a really high level of protection. A similar situation happens in the private sector, where many investments are devoted to the safeness and security in all the fields (financial, customers' security, etc...) (Bernat, 2012). All the aforementioned issues have raised growing interests in the areas of biometric recognition and video analysis, especially after the sad happenings of September 2001. In (Hobijn and Sager, 2007) an outlays assessment of the U.S. government is done concerning the years 2001-2005; the estimation reports describe that the homeland security spending climbed from \$56.0 billions of dollars in 2001 to almost \$100 billion of 2005. In this lapse of time, new pattern recognition techniques have been developed and, even more important, new biometric traits have been investigated and refined; besides the well-known physical and behavioral characteristics, also physiological measures have been studied, so providing more features to enhance discrimination capabilities of individuals.

1.1   The Biometry

The biometry is the science that studies the physical, behavioral and physiological measures and provides a quantitative representation of them by means of mathematical and statistical models. The biometric traits can be grouped in three areas:

- *Physical traits*: fingerprint, palmprint, face, iris, retina, ear. This kind of characteristics is also known as *hard* biometrics;

- *Behavioral traits*: signature, gait, gaze (eye movement), voice, keystroke. Usually this kind of traits have less discriminative capabilities than the former ones, given their very emotional nature; they're also known as *soft* biometrics;

- *Physiological measures*: signals like Electroencephalogram (EEG) and Electrocardiogram (ECG) belong to this group. Their nature is physical (brain

for EEG and heart for ECG), but the signal is strongly affected by the emotions and the state of mind of the subject in enrollment. Many aspects related to this kind of characteristics are still being investigated.

In order to asses a biometric trait, seven characteristics are typically taken into account (AlMahafzah and AlRwashdeh, 2012):

- *universality*: every person should possess the biometric trait;

- *distinctiveness*: two persons should be distinguishable by their own biometric trait;

- *permanence*: the biometric trait should be invariant over time;

- *collectability*: the biometric trait should be measurable by means of some practical device;

- *performance*: the enrollment, the feature extraction and the matching processes should take a limited amount of time;

- *acceptability*: the enrollment procedure should be accepted by a wide portion of the users;

- *resistance to spoofing*: the biometric trait should be strong against tries of spoofing.

The Figure 1.1 shows a comparison among the characteristics of several biometric traits. The target of the biometry is the development of biometric systems

Biometric identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

FIGURE 1.1: Comparison among several biometrics in terms of their characteristics.

whose main objective is secure and as-fast-as-possible identification of a person. The typical phases of a biometric system are:

- *Enrollment*: a user biometric trait is acquired by the system by means of proper sensors (cameras, microphones, fingerprint sensors...);

- *Detection and segmentation*: if needed, the enrolled trait is segmented and the area of interest is isolated;

- *Feature Extraction*: the template of the trait is extracted and a vector representing the trait is created;

- *Matching*: two feature vectors are compared and a similarity (or a distance) score is given as output.

A biometric system can be either aimed at *recognition* or at *verification*. The recognition process is a 1:N comparison; it takes as input a feature vector and finds a matching subject in a dataset; the verification process is a 1:1 comparison; it takes as input both a feature vector and an identity and verifies if the submitted biometric trait belongs to the claimed identity or not. In 1.2, enrollment phase, verification and recognition processes are shown.
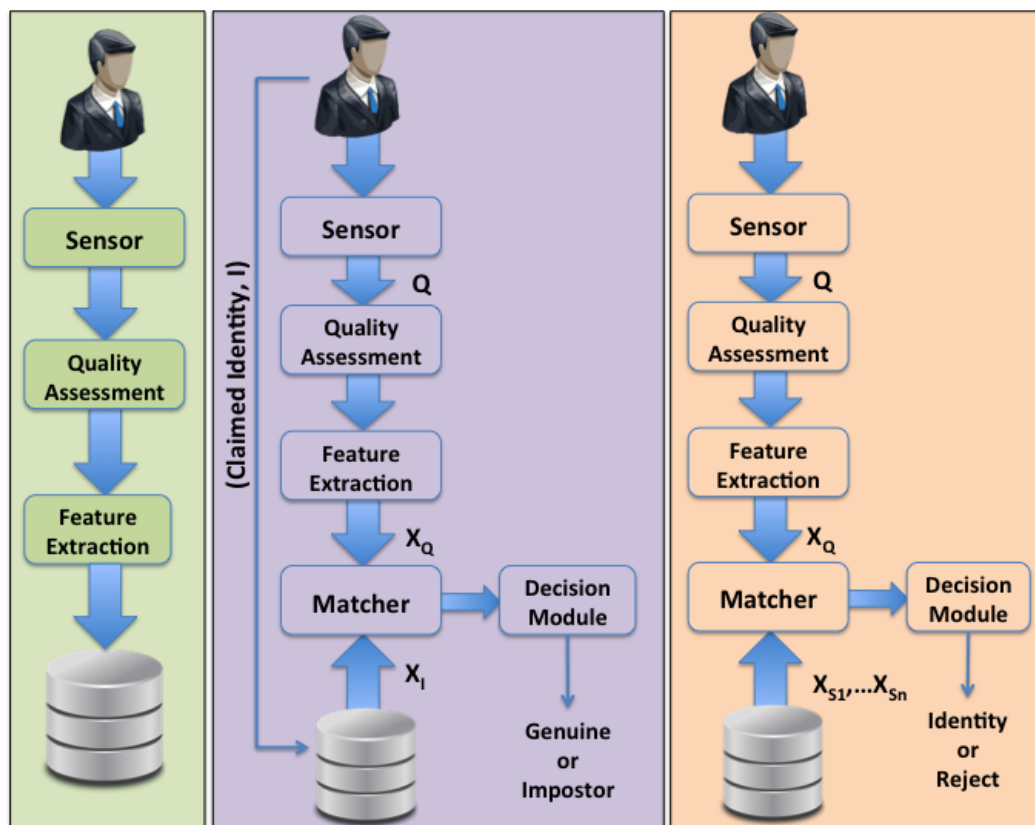
FIGURE 1.2: Enrollment, Verification and Recognition processes.

1.2   The Multi Biometric systems

The simultaneous use of more biometric traits provides further security to a recognition system. Such systems are named Multi Biometric Platforms (or

multi modal systems), and are gaining even more popularity since they contribute to overcome some limitations that affect a single biometric trait alone. Examples of these limitations are listed below:

- the iris is sensible to the distance from the camera in the acquisition process; moreover, eyelashes and eyelids usually partially cover the area of interest;

- the face can be spoofable without many efforts(masks or make up) and is not stable over time; in terms of occlusion, scarfs, hats and glasses reduce the area of interest;

- the DNA feature extraction process takes too much time for being complete;

- gait, keystroke and many other soft biometrics are affected by the emotions of the subject and, in these cases, a signature of the biometric trait is not always acquirable.

The fusion of biometric data tries to solve such drawbacks. It can be performed at several levels, sensor, feature, score and decision (see Figure 1.3), depending on the phase of the system where the fusion takes place.
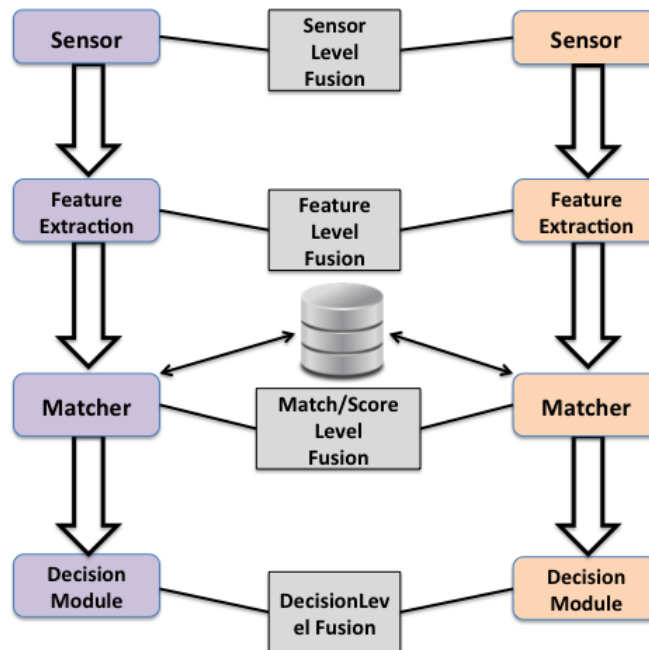
FIGURE 1.3: Levels of fusion.

Definitely, the score level fusion is the most used methodology; in fact, in the first two, problems above merging heterogeneous data may arise, whereas in the last one we have too much loss of information. In (Fathima et al., 2014) a two-level fusion approach has been proposed; a sensor level one, based on the combination of visible and thermal images using EMD (Empirical Mode

Decomposition) and a score level one performed fusing the output of three algorithms:

- Block-Independent Component Analysis (BICA)(Zhang, Gao, and Zhang, 2007);

- Discrete Cosine Transform with Fisher Linear Discriminant Classifier(Joo Er, Chen, and Wu, 2005)

- Kalman Filter(Eidenberger, 2006)

Further score level fusion approaches have been exploited in (Islam et al., 2013), where the Iterative Closest Point algorithm have been used in order to combine 3D information of face and ear, and in (Dehache and Souici-Meslati, 2012), where a SVM has been used over fingerprint and signature traits.

1.3   Thesis Contribution

This dissertation proposes the design of a multi biometric platform, based on the following biometric traits: ear, face, iris EEG and ECG signals. The major contributions of the thesis are:

- we propose a platform in which the customization of a multi biometric system is possible; the scalability of the whole system also allows the introduction of new biometric traits with minimum effort;

- the system offers two different fusion methodologies, depending on the biometric traits involved; the first one is based on the combination of the reliability measure and score. It is used for fusing physical traits (ear, face and iris). The second one exploits the weighted sum and is used for fusing EEG and ECG traits. Both approaches are performed at score level;

- the system allows the insertion of new classifiers and new segmentation/feature-extraction/matching methods. The classifier is stored in the warehouse of the platform. The combination of different functions into a new classifier is also possible;

- the system offers the tools for the configuration of the parameters of the classifiers;

- the platform provides the tools for populating the biometric datasets by adding a new database or by generating new chimera subjects.

- for each biometric trait, i.e. ear, face, iris, EEG and ECG, new algorithms have been proposed and compared with existing ones. In particular:

  - the ear recognition process uses a method that we experimented for the solution of the problem of newborn cradles'swap. The use of newborns' ears does not limit the generality of the method; on the

contrary, it represents a special case in which the potentialities of the method are toughly stressed. The method has really good performance since it has been developed to be used on mobile devices to allow the nurses to simply achieve the newborn recognition (Barra et al., 2014a);

– the face recognition process implemented in FAIRY uses a methodology that we proposed to make the process affordable on mobile devices, so paving the way to the porting of the FAIRY platform on such kind of devices; in fact the recognition process is optimized with respect to the required computational resources Barra et al., 2013.

– the iris recognition process exploits spatial histogram to extract features; similarly to the face recognition process, also this algorithm is optimized for running on mobiles Barra et al., 2015b.

– The ECG and EEG recognition modules exploit two methodologies for achieving the recognition of these physiological signals. The first is based on the detection of peaks in the ECG . The peaks are points of local maxima and local minima in the signal. The second method is based on the power spectral method and has been used for extracting the features from the EEG. The fusion of both signals is performed at score-level by means of the weighted sum, a simple as well as efficient approach (Barra et al., 2015a);

– in FAIRY, the fusion module uses a multibiometric fusion process based on different subsystems (ear recognition, face recognition and iris recognition modules) that produce their own recognition results. The method that we proposed evaluates the reliability and the decision score of each subsystem (module) and computes a fusion score (Barra et al., 2014c).

1.4   Outline of the thesis

The thesis presents the design of a multi biometric platform aimed at recognizing individuals, named FAIRY (Face, eAr, IRis, phYsiological measures). In the second chapter, the platform architecture has been presented, together with the modules and their functionalities. In chapters 3, 4 and 5 we have presented the study conducted for the solution of the recognition for ear, face and iris, respectively. Then in chapter 6, the studies related to the use of physiological measure for recognition purposes have been described, and the issue of fusing EEG and ECG signals has been faced. The fusion problem for the physical traits has been discussed in chapter 7. Chapter 8 completes the investigation on the pattern recognition issue involved in the videosurveillance problem. Conclusions and further works are drawn in chapter 9.

Chapter 2

FAIRY: a MultiBiometric Platform aimed at face, ear, iris and physiological measures recognition

2.1 Introduction

In this thesis the design of a Multi-Biometric platform, named FAIRY, has been proposed. FAIRY offers the possibility of using several biometric traits: face, ear, iris, physiological measures, independently or a fusion of some of them, as chosen by the user. FAIRY has a modular architecture, presented in Figure 2.1, so resulting in a fully updatable biometric system. Moreover, its modularity makes FAIRY also highly customizable. In fact:

- new modules can be added;

- out-of-date modules can be upgraded;

- new biometric classifiers can be created by composing different segmentation/feature extraction/matching techniques.

Besides offering several built-in algorithms for different biometrics, the FAIRY platform is provided with a Control Setting Module, by means of which it is possible to create, experiment and configure new recognition methods. From a functional point of view, FAIRY provides enrollment/recognition capabilities for each of the considered biometric traits, taken either singularly or in a fused fashion.

2.2 The Architecture of FAIRY

The architecture of FAIRY is composed of three subsystems (see Figure 2.1):

- *Configuration Subsystem*;

- *Biometry Subsystem*;

- *Enrollment Subsystem*.
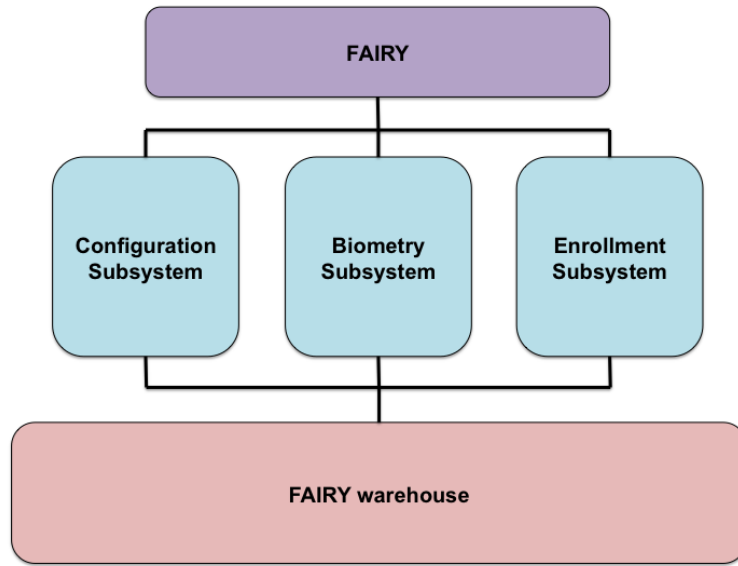
FIGURE 2.1: FAIRY modular architecture.
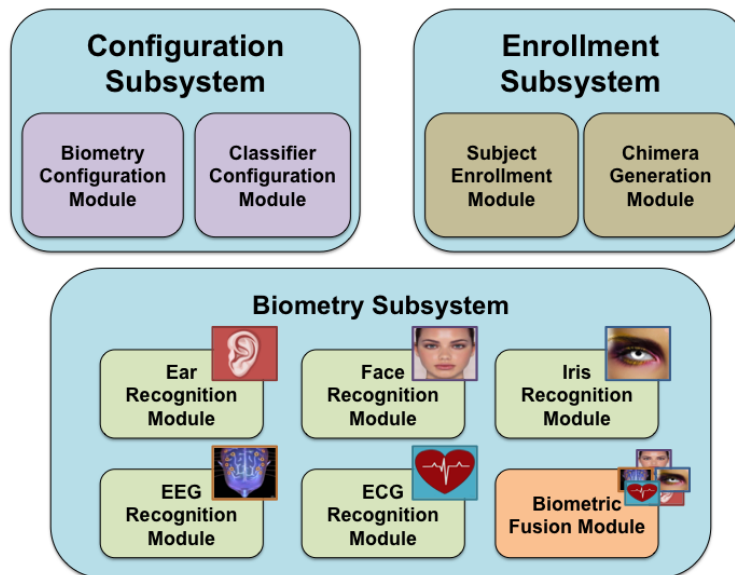
FIGURE 2.2: FAIRY's subsystems.

The Configuration Subsystem consists of two modules: the *Biometry Configuration Module* and the *Classifer Configuration Module*; the first module enables the choice of the biometric traits to be used whereas the second one deals with the creation of new classifiers and the addition of new segmentation/feature extraction/matching functions. The created classifiers are maintained in the Classifier Database. The Biometry Subsystem is the core of the FAIRY platform, and is composed of six modules:

- *Ear Recognition Module*;

- *Face Recognition Module*;

- *Iris Recognition Module;*

- *EEG Recognition Module;*

- *ECG Recognition Module;*

- *Biometric Fusion Module.*

A deep presentation of the problems involved and the solutions adopted for the previous modules have been extensively discussed in the next chapters. Finally, the Enrollment Subsystem consists of the *Subject Enrollment Module* and the *Chimera Generation Module*. The Subject Enrollment Module is in charge of the creation/management of all the biometric datasets, whereas the Chimera Generation Module supports the generation of new chimeric data for biometric fusion. The created datasets, together with the aforementioned Classifier Database, compose the FAIRY warehouse (see Figure 2.3), which is accessed by all the FAIRY subsystems. In this section they will be analyzed one-by-one.
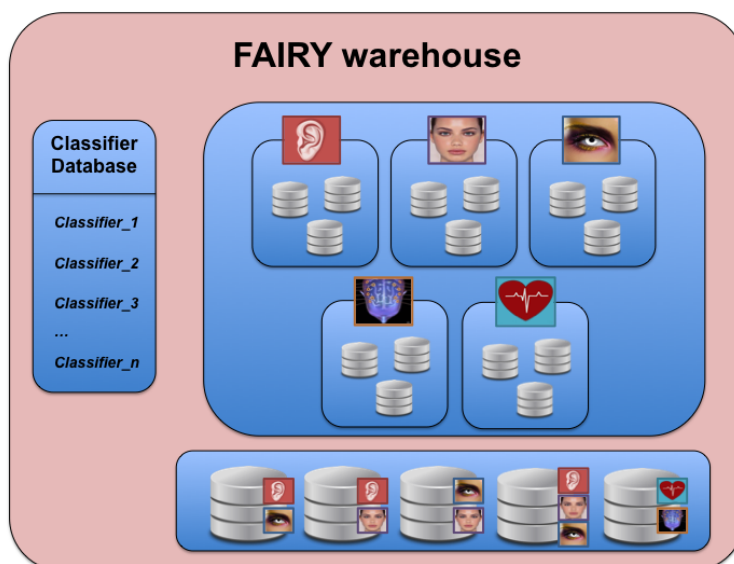
FIGURE 2.3: FAIRY warehouse.

2.3 The FAIRY Platform

In Figure 2.4, the FAIRY main screenshot is shown; from left to right, the three available options allow the access to the Biometric Recognition, the Subjects Enrollment and the General Settings functionalities (see Figure 2.4).

2.3.1 The Biometry Subsystem

The Biometry Subsystem functionalities are accessed by selecting the Biometric Recognition function; they offer the possibility of achieving the recognition of

FIGURE 2.4: FAIRY main screenshot.

FIGURE 2.5: Biometric Recognition functionalities.

the five biometrics considered or the fusion of a group of them. The Figure 2.5 shows the functions provided by the Biometry Subsystem. The ear recognition module is activated as the "ear" icon is single-clicked. The screenshot in Figure 2.6 is shown to the user. The research we conducted to investigate the ear recognition issue has been presented in (Barra et al., 2014b), where we experimented the ear recognition algorithms for identifying newborns in the Hospital of Cagliari so as to face the problem of cradles' swap. In the leftmost part of the view, the specific segmentation/feature-extraction/matching algorithms used are indicated. In Chapter 3 the results of this investigation have been reported. The ear recognition module exploits:

FIGURE 2.6: The user needs to select an image containing an ear.

- *Viola-Jones+STASM* as segmentation method;

- *Multiscale LBP* as feature extractor;

- *Euclidean Distance* as matching method.

The user is requested to select a file containing the image of an ear and then to select a database from those available in the FAIRY warehouse (see Figure 2.7). As the user presses the "GO" button (Figure 2.7, bottommost image) the
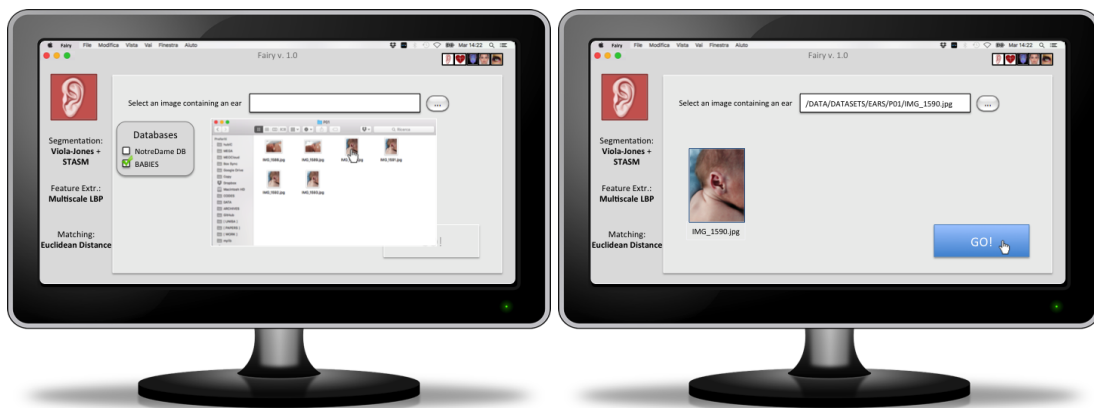
FIGURE 2.7: The user selects an image containing an ear (leftmost image) and the image is loaded (rightmost image).

recognition process starts. If the ear is correctly detected and segmented, it will be shown to the user (see Figure 2.8, leftmost image) and a "waiting" icon will inform the user that the recognition process is in progress. Once the process is terminated, on the left part of the screen a response will appear (see Figure 2.8, rightmost image).

FIGURE 2.8: The ear is correctly detected and segmented (leftmost image) and the recognition is successfully completed (rightmost image).

The iris recognition module works in a similar way (see Figure 2.9). The research we conducted has been presented in (Barra et al., 2015b), where the whole chain of an iris based biometric system has been implemented for mobile devices. In the leftmost part of the view, the specific algorithms used are indicated. In Chapter 5 the results of the study has been reported. The iris recognition module exploits:

- *IS_IS* as segmentation method;

- *Spatial Histogram* as feature extractor;

- *Weighted Sum* as matching method.

As for the face recognition module, FAIRY allows the user to select the built-in live camera of the computer as source of the image to be processed. In such cases the "Spoofing Detection" option is available (see Figure 2.10). As explained further in chapter 4, this method needs a live video stream in order to evaluate potential deceptions aimed at cheating the system. The research conducted in this face recognition field have led to the implementation of a mobile-oriented biometric system. The algorithms exploited in this recognition module are:

- *Viola Jones + Self Quotient Image* as segmentation method;

- *Spatial Correlation* as feature extractor;

- *Spatial Correlation* as matching method.

Two further modules allow the recognition of physiological measures, i.e. the ECG (electrocardiogram) and the EEG (electroencephalogram). As regard the ECG, several studies in this field led to the conclusion that this measure provides many discriminative features that may help to recognize individuals. As further explained in chapter 6, the conducted research exploited a very simple method for achieving ECG recognition: first a detrending operation normalizes
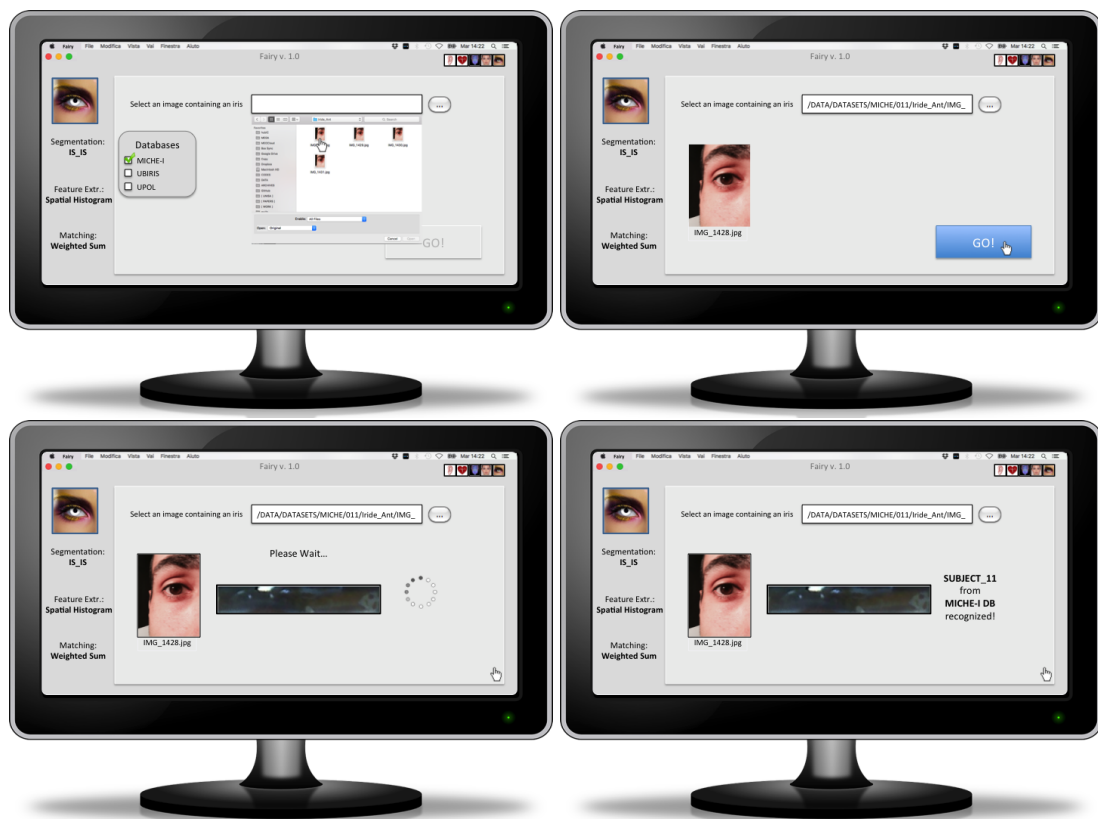
FIGURE 2.9: The iris recognition module

FIGURE 2.10: The "Spoofing Detection" option becomes available
when the user selects a live source.

the baseline of the signal; then, a peak detection method extracts the local minima and maxima from the ECG. The matching phase is done by means of the norm-1 method (Barra et al., 2015a). The user needs to select and EDF or a DAT file from the file system, since in this module the file to be processed is a signal and not an image (see Figure 2.12). The EEG recognition module allows the recognition, by analyzing the brain activity of an individual. The features of the EEG signal are still being investigated, especially from the point of view of the biometric identification. Nonetheless, recent researches have confirmed some discriminative capabilities of the signal. In chapter 6, a method for extracting features by means of the *power spectral density* (PSD) method is described. The case of study has been proposed in (Barra et al., 2015a). As well as for the ECG recognition method a detrending operation normalized the signal before the extraction of the features (see Figure 2.13).

The biometric fusion module allows the merge of different biometrics, in order to improve the recognition process and to make the final response safer. In literature, four fusion approaches have been investigated, depending on the

FIGURE 2.11: The user needs to select an image containing a face
(uppermost image) or to activate the built-in camera for a live shot
(bottommost images).

FIGURE 2.12: The selection of the ECG signal and the recognition
process.

FIGURE 2.13: The selection of the EEG signal and the recognition process.

level they operate (Ross and Govindarajan, 2005):

- *sensor* level fusion: the raw data acquired from multiple sensors can be processed and integrated to generate new data from which features can be extracted;

- *feature* level fusion: the feature sets extracted from multiple data sources can be fused to create a new feature set to represent the individual;

- *score* level fusion: multiple classifiers output a set of match scores which are fused to generate a single scalar score;

- *decision* level fusion: the responses of multiple biometric systems are combined in order to obtain a decision for the multibiometric system.

Due to difficulties implied in the former two, and in the loss of supplementary information in the last one, most schemes in literature opt for score level fusion. In FAIRY, two score level fusion methodologies have been inserted: the first combines the SRR (*system response reliability*) of a biometric system with a recognition score and computes a complex value that represents the reliability of the resulting response (Barra et al., 2014c); the second one simply fuses two scores by computing a weighted sum (Barra et al., 2015a). The reason behind the use of two different fusion algorithms depends on the type of biometrics involved. The first technique is used for fusing physical traits (ear, face, iris) whereas the second one is used for the physiological signals (EEG and ECG). Obviously, according to the technique and to the kind of feature vectors extracted from the single biometrics, both fusion techniques may be used for the fusion of any biometric presented in FAIRY; notwithstanding, since this option has not be tested, it has not been considered in the platform. However, in 9 it is considered as a possible upgrade of FAIRY. In Figure 2.14, the views of the fusion module are shown: in the uppermost image, the user have chosen to execute the fusion of ear and face; according to the involved technique, the reliability scores are indicated. In the bottommost image, the user executes the fusion of the physiological signals. In both cases, the user is asked to select a *.mat* file containing the biometric traits

FIGURE 2.14: The available options for the fusion module: the fusion of physical (uppermost image) and physiological traits (bottommost).

2.4 The Enrollment Subsystem

The Enrollment Subsystem functionalities are accessed by clicking on the Subjects Enrollment button in the main view; they provide the tools for populating the FAIRY warehouse. Two options are available:

- the addition of a new dataset to the FAIRY warehouse ;

- the creation of a chimera subject.

As for the first option, the user selects which kind of trait he is going to enroll, and checks the corresponding checkbox; then, he navigates the file system until the root of the dataset. As regard the second option, the user selects more

checkboxes depending on the traits he means to fuse; then, by means of two or more drop-down menus (depending on the number of traits involved) he can choose the samples to be fused. Figure 2.15 shows this scenario.

FIGURE 2.15: In the uppermost image the user adds the Notre Dame dataset to the FAIRY warehouse; in the bottommost, instead, a new chimera is generated.

2.5 The Configuration Subsystem

The Configuration Subsystem consists of two modules that provide functionalities to configure the whole platform:

- the Biometry Configuration module allows the choice of the biometric traits to be used in FAIRY;

- the Classifier Configuration module allows the insertion of new classifiers in the platform and the configuration of the segmentation/feature-extraction/matching methods parameters.

The customization of the platform is achieved by selecting the biometric traits; the selection is made by checking the corresponding boxes. The Figure 2.16 shows a scenario of this functionality. In order to add/compose a new classifier,
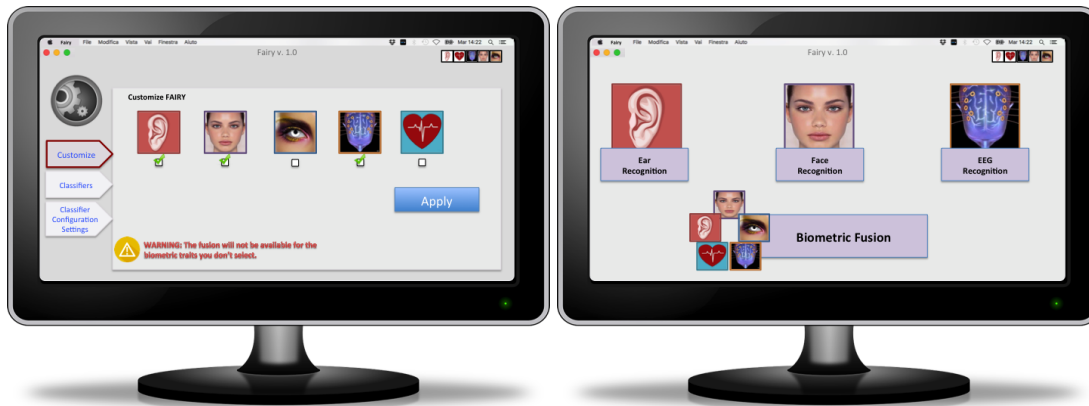
FIGURE 2.16: Customization of the platform: the user selects ear, face and EEG for his customized platform.

the user needs to access the "classifier" (see figure 2.17) section, by means of which he has the possibility to modify an existing classifier for a given biometric trait, or to upload a new method to be used in the recognition process. By checking the desired radio buttons, he can compose a new classifier. Otherwise, new methods can be uploaded: in such cases, the signatures (input and output parameter types) of the method are set. The type constraints about the functions are listed below:

- the segmentation function (see Figure 2.18, leftmost image) takes an image (or a signal) in input and returns the segmented image (or a portion of the signal);

- the feature extraction method (see Figure 2.18, middle image) takes a segmented image as input (or the portion of the signal) and returns a feature vector;

- the matching method (see Figure 2.18, rightmost image) takes two feature vectors in input and returns a distance score.

Given the signature of the methods that can be loaded in FAIRY, each of them can have a configuration file attached. The configuration file (in *.xml* format) provides further parameters needed from the method to work properly. The third section in the General Settings Subsystem provides a form that loads the configuration file and allows the user to modify the parameters.

FIGURE 2.17: The user can add a new classifier to the platform or can compose a new classifier by cascading a segmentation method, a feature extraction method and a matching method.

FIGURE 2.18: The segmentation, feature extraction and matching templates.

FIGURE 2.19: The form by means of which is possible to modify the configuration file.

Chapter 3

Ear Biometrics

3.1 Introduction

Not many people are aware of the uniqueness of the individual ear features, but actually, in the far 1890, Bertillon stated: *"We will end the exam of the profile with the study of the ear which, thanks to multiple small valleys and hills which furrow across it, is the most significant factor from the point of view of identification. Immutable in its form since birth, resistant to the influences of environment and education, this organ remains, during the entire life, like the intangible legacy of heredity and of the intra-uterine life."* This discriminative capability is mostly due to the external ear "flap", technically defined as *pinna*, with its morphological components. The resulting structure, though relatively simple, varies significantly across different individuals, as shown by examples in Figure 3.1. Ear biomet-
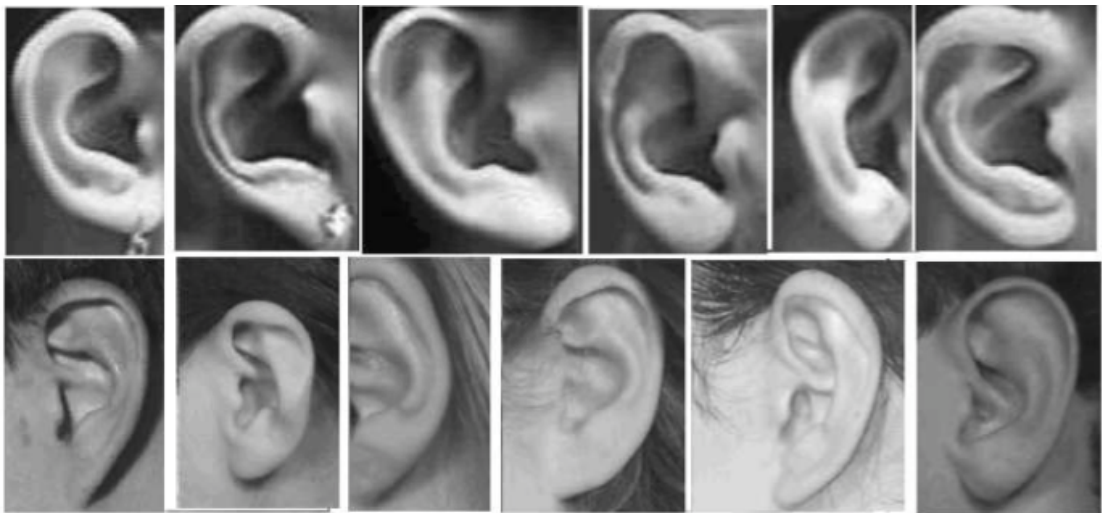
FIGURE 3.1: Different shapes of adult ears. Differences in the morphology can be visually appreciated.

rics, compared with other physical traits, presents both advantages and limits. First of all, the small surface and the quite simple structure play a controversial role. On the positive side, they allow faster processing than, say, face recognition, as well as less complex recognition strategies than, say, fingerprints. On the negative side, the small ear area itself makes recognition systems especially sensitive to occlusions. Moreover, the prominent 3D structure of distinctive elements like the pinna and the lobe makes the same systems sensible to changes in

illumination and viewpoint. Overall, the best accuracy results are still achieved in conditions that are significantly more favorable than those found in typical (really) uncontrolled settings. This makes the use of this biometrics still difficult to propose in real world applications, since a commercial use requires a much higher robustness. Notwithstanding the mentioned limits, ear is still an attractive topic for biometrics research, due to other positive aspects. In particular, it is quite easy to acquire ear images remotely, and these anatomic features are also relatively stable in size and structure along time. Of course, as any other biometric trait, they also call for some template updating. This is mainly due to age, but not in the commonly assumed way. The apparent bigger size of elders' ears with respect to those of younger subjects, is due to the fact that aging causes a relaxation of the skin and of some muscle-fibrous structures that hold the pinna, that is the most evident anatomical element of the ear. In FAIRY, the ear recognition process exploits the method proposed by (Barra et al., 2014a) for the solution of the problem of newborn cradles'swap. The use of newborns' ears does not limit the generality of the method; on the contrary, it represents a special case in which the potentialities of the method are toughly stressed. The method has really good performance since it has been developed to be used on mobile devices to allow the nurses to simply achieve the newborn recognition. This chapter is organized as follows: the section 3.2 shows the related works; the section 3.3 describes the methodology used in the ear recognition module of FAIRY. Finally, in section 3.4, experimental results are presented.

3.2   Related Works

A first rough classification of human ear identification techniques using outer anatomical elements distinguishes two approaches: the first one exploits detection and analysis of either feature points or relevant morphological elements, and derives geometric measures from them; the second one relies on global processing (Moreno, Sanchez, and Velez, 1999). Taking into further consideration the dimensionality of captured data and the kind of feature extraction, we can also identify the following categories:

- Recognition from 2D images

    - Geometrical approaches (interest points)
    - Global approaches
    - Multiscale/multiview approaches
    - Thermograms

- Recognition from 3D models

We will briefly sketch some examples here and return on more details in the following paragraphs.

3.2.1 Recognition from 2D images

In geometrical approaches, a set of measures are computed over a set of interest points and/or interest contours, identified on a 2D normalized photo of the ear. The pioneering work related to ear recognition, namely "Iannarelli system" of ear identification, is a noticeable example along this line. In Iannarelli's system, the ear image is normalized with respect to dimensions. The point named Crux of Helix is identified (see Fig. 3.2) and becomes the center of a relative space. All measures are relative to it, so that a wrong identification of such point compromises the whole measurement and matching processes. Burge
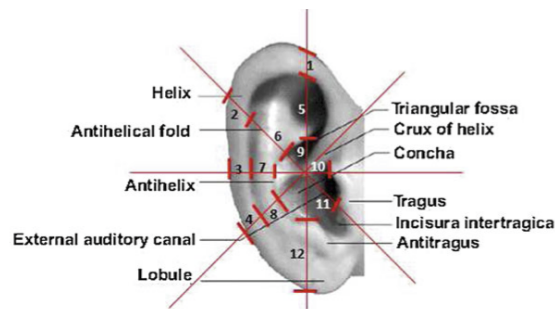
FIGURE 3.2: Anatomy of the pinna of external ear and Iannarelli's measures.

and Burger (Burge and Burger, 1997; Burge and Burger, 1996) are among the first to try more advanced techniques. They use Voronoi diagrams to describe the curve segments that surround the ear image, and represent the ear by an adjacency graph whose distances are used to measure the corresponding features (Fig. 3.3). The method has not been extensively tested, however we can assume that Voronoi representation suffers from the extreme sensitiveness of ear segmentation to pose and illumination variations. Global approaches con-
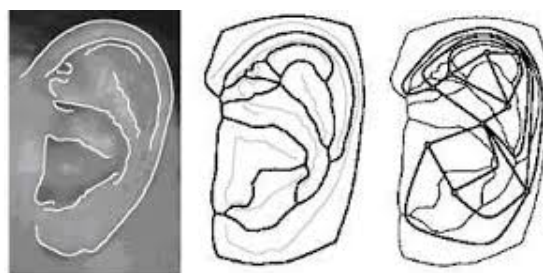
FIGURE 3.3: Relevant edges on ear external surface (left) and derived Voronoi diagram (center) and adjacency graph (right).

sider the whole ear, instead of possible relevant points. Hurley et al. (Hurley, Nixon, and Carter, 2000; Hurley, Nixon, and Carter, 2002) address the problem of ear recognition through the simulation of natural electromagnetic force fields. Each pixel in the image is treated as a Gaussian attractor, and is the source of a spherically symmetric force field, acting upon all the other pixels in a way that is directly proportional to pixel intensity and inversely proportional

to the square of distance. The ear image is so transformed into a force field. This is equivalent to submit it to a low pass filter transforming it into a smooth surface, where all information is still maintained. The directional properties of the force field can support the location of local energy peaks and ridges to be used to compose the final features. The technique does not require any explicit preliminary description of ear topology, and the ear template is created just following the force field lines. It is invariant to the translation of the initializing position and to scaling, as well as to some noise. However, the assessment of the recognition accuracy has been attempted only with images presenting variations on the vertical plane and without hair occlusion (Hurley, Nixon, and Carter, 2005). Since face and ear present some common features (e.g., sensitiveness to pose and illumination, sensitiveness to occlusions) it is natural to inherit face related techniques. In particular, Principal Component Analysis (PCA) has been widely used for face template coding and recognition. In (Chang et al., 2003), PCA is applied to both biometrics for comparison purposes, demonstrating similar performance as well as limitations. Despite the apparently easy, for ear recognition images must be accurately registered, and extraneous information must be discarded by a close cropping. Last but not least, this method suffers from very poor invariance to those factors which also affect face recognition, in particular pose and illumination. Further global approaches, still borrowing from face recognition, are presented in (Moreno, Sanchez, and Velez, 1999; Sanaa, Guptaa, and Purkaitb, 1999). The former applies neural network strategy to ear recognition, testing Compression Networks, Borda Combination, Bayesian and Weighted Bayesian Combinations techniques. The latter exploits Haar wavelets transformation. In multiscale/multiview approaches, the set of features used for recognition is enriched by considering more scales for the same image, or more acquisitions, possibly from slightly different points of view, for the same trait of the same subject. While multiview approach is often used to obtain a 3D model of the anatomical element, it can also be used in 2D techniques. Though acquired by different equipment and containing information different from pixel intensities, also thermograms are 2D images. Through a thermographic camera, a thermogram image captures the surface heat (i.e., infrared light) emitted by the subject. These images are not sufficiently detailed to allow recognition, but can rather be used for ear detection and segmentation, especially in those cases where the ear is partially covered and passive identification is involved (i.e. when the user does not cooperate or might be unaware of the acquisition). In such cases, texture and color segmentation should allow to discard hair region. As an alternative, Burge and Burger (Burge and Burger, 1996) propose to use thermogram images. The pinna usually presents an higher temperature than hair, so that the latter can be segmented out. Moreover, if the ear is visible, the Meatus (i.e., the passage leading into the inner ear) is the hottest part of the image, which is clearly visible and allows to easily detect and localize the rest of the ear region. Disadvantages include sensitiveness to movement, low resolution and high costs.

3.2.2   Recognition from 3D images

Like other techniques, 3D processing for ear has followed the success of three-dimensional techniques applied to face. As a matter of fact, they solve similar problems of sensitiveness of 2D intensity images to pose and illumination variations, including shadows which may sometime play a role similar to occlusions. On the other hand, the outer part of the ear presents even richer and deeper 3D structure than face, with a very similar discriminating power, which can be profitably modeled and used for recognition purposes. Among the first and most significant works along this direction, we mention Chen and Bhanu (Chen and Bhanu, 2007; Bhanu and Chen, 2003), and Yan and Bowyer (Yan, Bowyer, and Chang, 2005; Yan and Bowyer, 2007b). Actually, both research lines have an articulated development in time, and we will only mention the main achievements. Both use acquisitions by a range scanner. In the first approach, ear detection exploits template matching of edge clusters against an ear model; the model is based on the helix and antihelix, which are quite extended and well identifiable anatomical elements; a number of feature points are extracted based on local surface shape, and a signature called Local Surface Patch (LSP) is computed for each of them. This signature is based on local curvature, and is used together with helix/antihelix to compute the initial translation/rotation between the probe and the gallery model. Refined transformation and recognition exploit Iterated Closest Point (ICP). ICP is quite simple and accurate, therefore it is widely used for 3D shape matching. The reverse of the medal is its high computational cost. The authors also test on 3D models of the ear a more general approach presented in (Chen and Bhanu, 2009), which integrates rank learning by SVM for efficient recognition of highly similar 3D objects. In the second mentioned approach, an efficient ICP registration method exploits enrollment data, assuming that biometric applications include a registration phase of subjects before they can be recognized. Moreover, ear extraction exploits both 2D appearance and 3D depth data. A detailed comparison of the two methods can be found in (Chen and Bhanu, 2007). It is worth mentioning that a number of attempts are made to reduce the computational cost of ICP. As an example, in (Yan and Bowyer, 2007a) Yan and Bowyer use a k-d tree structure for points in 3D space, decompose the ear model into voxels, and extract surface features from each of these voxels. In order to speed up the alignment process, each voxel is assigned an appropriate index so that ICP only needs to align voxel pairs with the same index.

3.3   The Ear Recognition Module approach

The process pipeline of the ear recognition module can be divided into the usual three general steps:

- Localization of the zone containing the ear;

- Segmentation and extraction of the features;

- Matching.

The overall processing is depicted in Figure 3.4.
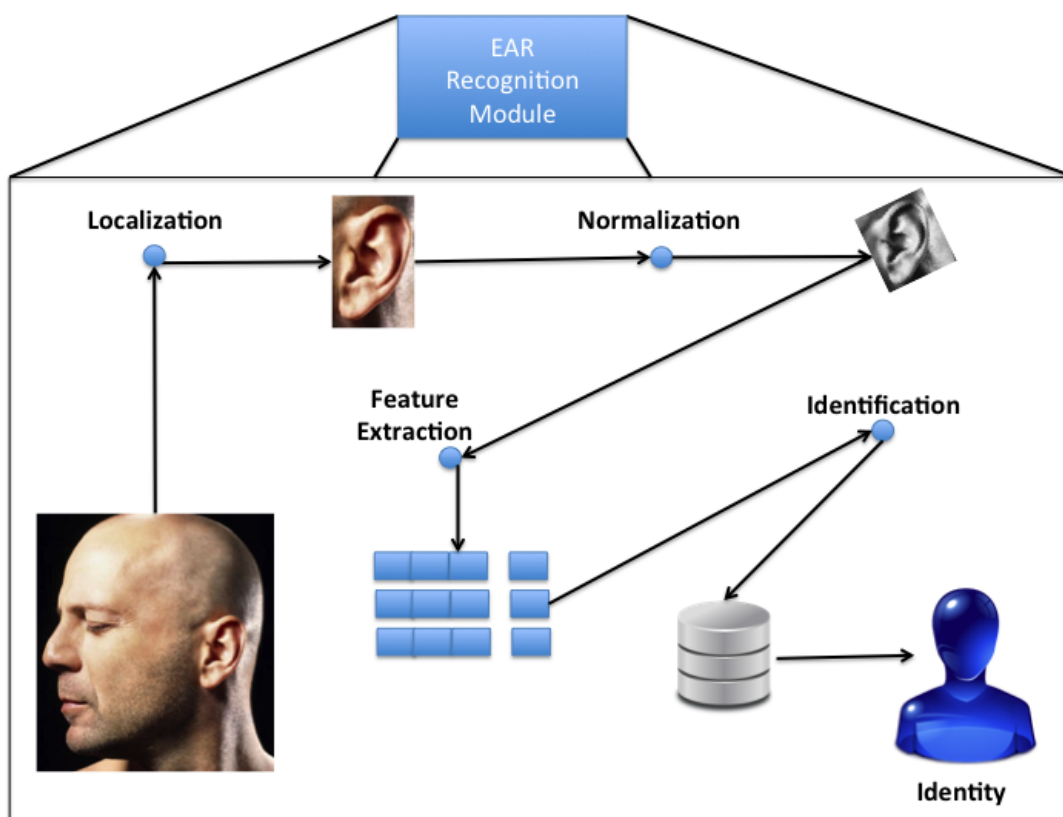
FIGURE 3.4: The overall processing for the module.

3.3.1   Detection

The image is first resized so that the height and the width are not greater than 800 pixels. After this, the Viola-Jones algorithm (Viola and Jones, 2001) is exploited to identify one or more candidate regions of interest (ROIs), possibly containing the ear. Starting from the original position and rotating the image of 10 and 20 degrees both clockwise and counterclockwise, we look for a positive response from the detector. For each image rotation, a growing threshold is then iteratively applied to finally possibly select a single region of sufficient size for each image. The initial and last iterations are shown in Figure 3.5 for a sample image. In case of successful detection, the result is a ROI similar to that shown in Figure 3.6.

3.3.2   Segmentation

The next step is to identify the shape of the ear as precisely as possible. We use an approach based on the Active Shape Model (Cootes et al., 1995), namely STASM (Milborrow and Nicolls, 2008). Before doing segmentation, a training
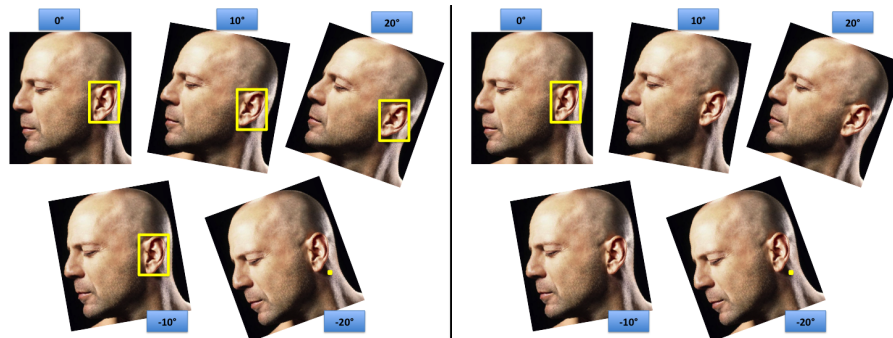
FIGURE 3.5: The first iteration of ROI detection (five leftmost images) and the last (five rightmost images).

FIGURE 3.6: The result of a correct detection.

phase is needed which takes a set of images where the points that eventually we would like to identify are manually indicated. In this way, we obtain a classifier that searches for the desired shape in the image. Landmark points are always labeled such that landmark 0 is the base of the ellipse, landmark 1 is the base of the lobe and so on. It is worth noticing that ASM-based techniques, like STASM, may fail in case of variations in the position/orientation of the object. In this case, we can rely on an acquisition in a quite canonical position. On the other hand, the model used for training can significantly affect the obtained results. We used in turn two different model, which are shown in Figure 3.7(a) and Figure 3.7(b). A first approach was to train the classifier with 44 landmark points distributed over the ear that covered in good part the internal forms and also touched the outer contour (Figure 3.7(a)). While the points on the outer perimeter of the ear were found with a fair success, the internal points did not match with the result expected always. This is because the internal structure of the ears has a more variable conformation than the outer one, so that it is very difficult to approximate it in a uniform manner. For this reason, we changed the model using a more dense distribution of points on the outer contour. The new landmarks are 35 (Figure 3.7(b)), 33 of which are on the outer contour of the ear. The last 2 are respectively at the vertex of the triangular fossa and on the tip of the tragus. Once obtained the landmark points defining the ear shape, we transform the image of the ear to a standard format to facilitate the recognition process. First of all, images are rotated so that they are all angled the same way. In order to do this, we approximate the first 33 points of the

(a)   (b)
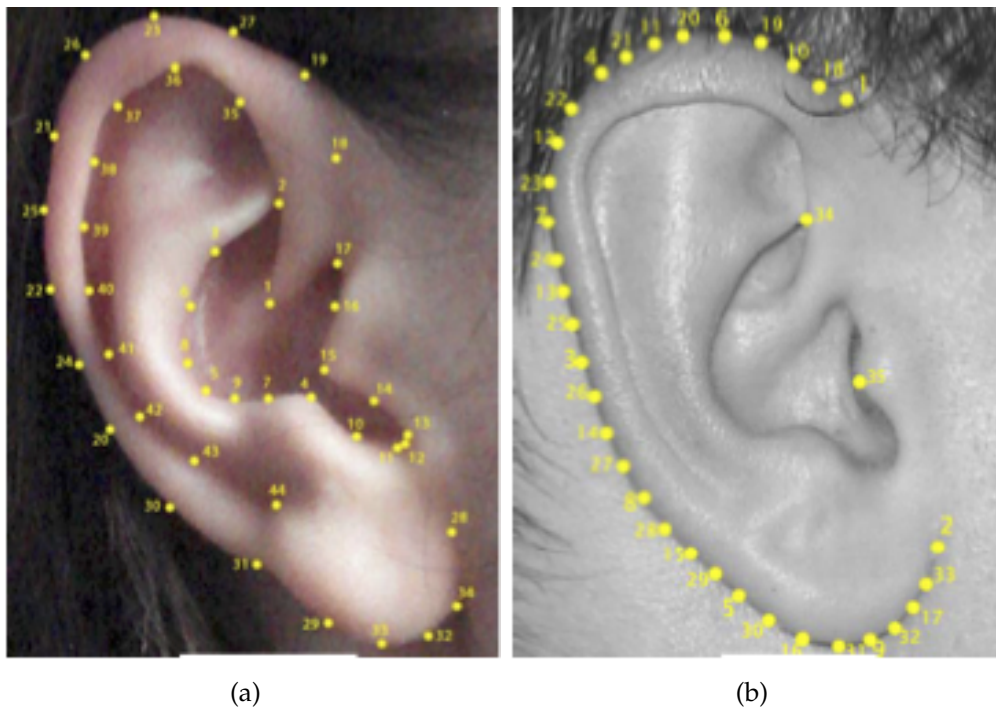
FIGURE 3.7: The models used for STASM training: the leftmost
image shows the model with 44 outer and inner points; the right-
most shows the one with 33 outer and 2 inner points.

shape with an ellipse. Let us remind that the last 2 points belong to the inner
region. They were excluded since they usually represent a deviation from the
quasi-elliptical shape of the ear and therefore would disturb the approximation
process. The center of the found ellipse is taken as the pivot for the rotation of
the image, which is performed in such a way that the major axis of the ellipse
is parallel to the y axis (Figure 3.8). Further normalization steps regard size
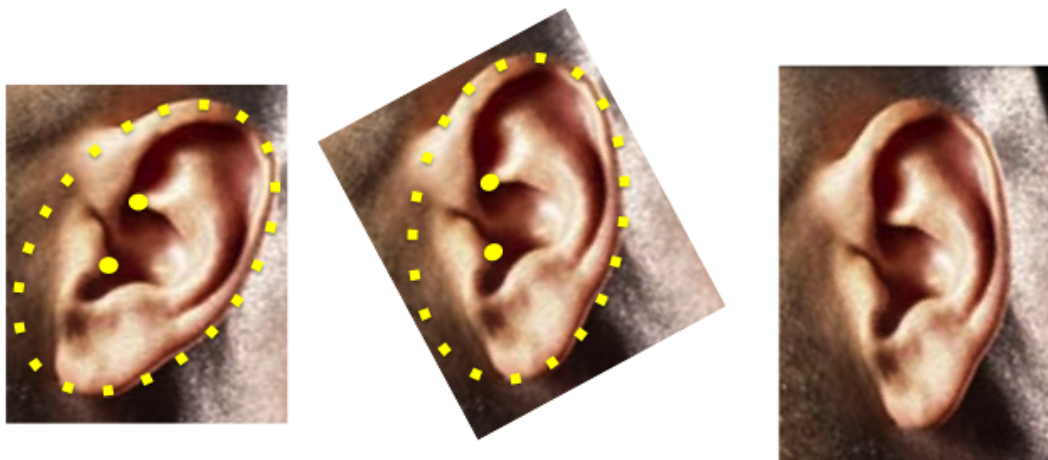
FIGURE 3.8: Normalization of ear Rotation.

and color. It was decided to use an image size of 90x144, with a height/width

ratio of 1.6, which is a good approximation of the average size for the ears. As for color, we apply grayscale conversion, histogram equalization to improve contrast, and median filter for noise reduction. Color normalization is worth since, in general, ear structure is sufficient in itself for recognition, while skin color, though being a useful soft biometrics, can be influenced by environment and capture conditions. Some examples are shown in Figure 3.9.
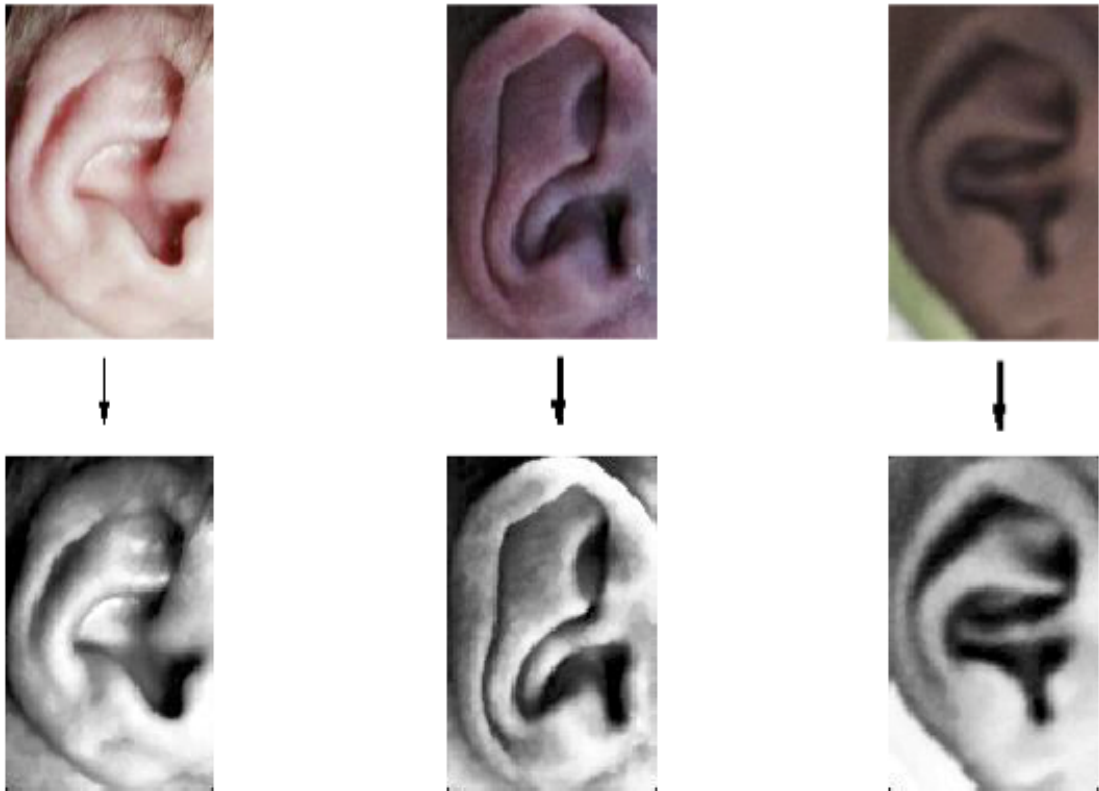
FIGURE 3.9: Normalization of ears presenting different skin color. The images on the top are the original ones, whereas those on the bottom are the normalized ones.

3.3.3 Feature Extraction

The last step of processing is the computation of a feature vector. We exploit multiscale LBP (Mäenpää and Pietikäinen, 2003) after dividing the image in 32x32 square regions. Then, in order to reduce the dimension of the feature vector, we applied four different techniques and compared the recognition performance obtained by Euclidean distance between probe and gallery vectors: Principal Component Analysis (PCA) (Daultrey, 1976; Turk and Pentland, 1991), Linear Discriminant Analysis (LDA) (Lachenbruch and Goldstein, 1979; Etemad and Chellappa, 1997), Neighborhood preserving embedding (NPE) (He et al., 2005b), and Orthogonal Locality Preserve Projections (OLPP) (Cai et al., 2006).

3.4   Experimental Results

We tested recognition using the four mentioned reduction techniques, and with a varying number of images in the gallery of each subject. This should positively influence recognition, since having more images to compare can help catching some variations and improve both verification and identification. In practice, when a probe is submitted, it is compared with all the images in the gallery (pertaining to the supposed subject if in verification mode). Let us remind that Euclidean distance among templates is used for matching. We measured performance by setting FAR at 10^{-3}, which is an acceptable value in this context. With respect to this value we computed Genuine Acceptance Rate (GAR) and False Rejection Rate (FRR). We further computed the equal Error Rate (ERR). Table 3.1 reports the results obtained at 10^{-3} in the four cases in which the gallery for each subject includes 1,2,3 and 4 images, and for the four algorithms for dimensionality reduction applied to LBP feature vectors. Table 3.2 reports the values for EER for the same cases. Figure 3.10 shows the varia-

TABLE 3.1: GAR and FRR at FAR=10^{-3} obtained by the four reduction methods applied on LBP feature vectors, with a varying number of images per subject in the gallery.

number of images	NPE		OLPP		PCA		LDA	
	GAR	FRR	GAR	FRR	GAR	FRR	GAR	FRR
1	0.80	0.20	0.83	0.17	0.83	0.17	0.63	0.37
2	1.00	0.00	0.92	0.08	0.92	0.08	1.00	0.00
3	0.93	0.07	0.87	0.13	0.87	0.13	0.93	0.07
4	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

TABLE 3.2: EER obtained by the four reduction methods applied on LBP feature vectors, with a varying number of images per subject in the gallery.

number of images	NPE	OLPP	PCA	LDA
1	0.08	0.09	0.09	0.11
2	0.07	0.08	0.08	0.11
3	0.07	0.08	0.08	0.10
4	0.06	0.08	0.07	0.07

tion of Recognition Rate for the four reduction techniques when the number of images per subject in the gallery increases. As expected, increasing the number of images in the gallery from 1 to 4 for each subject positively affects performance of all four methods, in both verification and identification operations. The differences among dimensionality reduction techniques are often negligible, and this can be assumed to depend on the characteristics of multiscale LBP when applied to the images on a per-region basis.
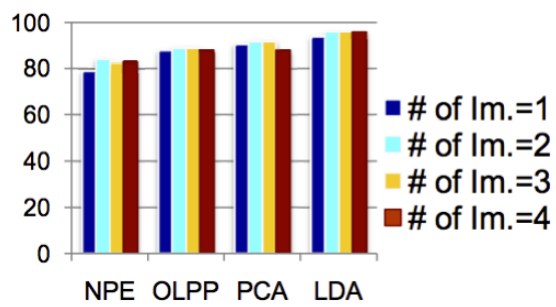
FIGURE 3.10: Performance variation in terms of Recognition Rate as the number of images per subject in the gallery increases.

Chapter 4

Face Biometrics

4.1 Introduction

The face is definitely one of the most used biometrics in recognition processes, as well as one of the most investigated. Without any doubt, it's the most immediate physical characteristic that people usually face with, and therefore the first that they use in order to recognize other individuals. For this reason it has a really high degree of acceptability among people, due to the fact that people show their face without any problem and, generally, they don't complain if photographed. Over the years lots of aspects of the face have been studied to make the recognition of this trait even more precise and context independent. This is the case of the *In-The-Wild* face recognition, where the images to be analyzed are often affected by several disturbing factors in order to simulate real scenarios of applications. These factors are also known as *PIE (pose, illumination, expression) issues*. In order to toughly stress the detection and recognition methods aimed at recognizing people in real scenarios, several datasets have been collected, like FERET (Phillips et al., 1998) and CMU-PIE database (Sim, Baker, and Bsat, 2002).

4.2 Related Works

Face recognition techniques may undergo different classifications. In (Zhao et al., 2003) psychophysicists and neuroscientists theories are taken into account to identify three classes: *holistic*, which uses the whole face region as input to the recognition system; *feature-based methods*, which first extract local features such as the eyes, nose, and mouth and then use their locations and local statistics for classification; *hybrid*, which work just as the human perception system is deemed to do, using both local features and the whole face region for recognition. A different classification was used in (Abate et al., 2007), which mostly takes into account the main techniques underlying the literature approaches: linear/nonlinear projection methods (for dimensionality reduction), neural networks, Gabor filters and wavelets, fractals and Iterated Function Systems (IFSs), use of thermal and hyperspectral images. 3D face recognition is also often discussed, but it requires special equipment and/or computationally expensive processing. A detailed discussion of the methods in literature is presented in (Nappi and Riccio, 2008) and (Jafri and Arabnia, 2009). Notwithstanding the

adopted classification, it is widely accepted that a mixture of local and global methods often seems the most appropriate and PIE-robust approach to address the problem of face recognition. However, it is to consider that even the most traditional methods underwent research efforts to improve their robustness to uncontrolled settings. A first example is related to PCA (Principal Component Analysis), an application of the Karhunen-Loeve procedure for space dimensionality reduction. (Kirby and Sirovich, 1990) were among the first to adopt the PCA in characterizing human faces. The addition to the database of images pertaining to a new face appears to be very fast, because PCA-related procedure is only performed for the training phase of the system. However, when a large number of new subjects is added to the system, it is necessary to train the system again (retraining). Moreover, the method suffers from a high sensitivity to PIE variations and occlusions. A possible solution to address such limitations was presented in (Koh, Ranganath, and Venkatesh, 2002), where PCA is used in a modified version intended to address the complexity of recognizing a person in a video frame, with multiple views of the same face. Proposed as a better alternative to PCA, LDA (Linear Discriminant Analysis) (Martínez and Kak, 2001; Lu, Plataniotis, and Venetsanopoulos, 2003) expressly provides a better discrimination between the classes. However, LDA too achieves poor robustness to PIE variations to be used in uncontrolled settings. Despite this, in (Lee and Kim, 2004) the authors claim to use an enhanced combination of PCA/ LDA for the face recognition process in a video summarization and retrieval system. Still in the context of proposals for MPEG-7 face recognition descriptors, a further proposal involving LDA was presented by (Kim et al., 2005). They devise a component-based LDA representation which allows to enhance the generalization capability of the LDA representation with respect to changes underrepresented in the training data. In practice, the face is divided into a number of facial regions and a separate LDA is learnt for each region. As a matter of fact, this is an example of how local processing can even improve the performance of more classical approaches. Neural Networks are a nonlinear solution to the pattern recognition problem. The basic idea of face recognition based on neural networks is to associate each neuron of the net with each pixel in the image, but because of the high number of neurons needed, a dimensionality reduction technique is first performed and neural networks are trained on the reduced image space. In practice, the vector representing the input image is often first submitted to a first neural network that reduces it to a (much) shorter vector h. Afterwards, the vector h is submitted to the second net (classification) for recognition. Actually, the neural network approach is not suited for massive face recognition tasks in uncontrolled conditions, but related methods can give good results in preliminary face region detection instead. For instance, in (Garcia and Delakis, 2002) a convolutional neural network architecture was presented to recognize highly variable face patterns with no preprocessing. Detection and precise localization of semi-frontal human faces is performed making no assumption on colors, on the illumination conditions, or on the size of the face. The approach of using Laplacianfaces for recognition was first introduced by (He et al., 2005a). Laplacianface use Locality Preserving Projection

(LPP) to learn a locality preserving subspace. Its aim is to capture the intrinsic geometry of the data and its local structure. Each face image is mapped onto the computed subspace that is characterized by a set of feature images, called Laplacianfaces. The basis functions obtained with the original Laplacianface approach are not orthogonal, making it difficult to reconstruct the data. For this reason the authors in (Cai et al., 2006) developed the OLPP method that produces orthogonal basis functions that preserve the metric structure of the face space and achieves more locality preserving power than LPP. Among traditional feature based methods, Elastic Bunch Graph Matching (EBGM), first introduced by (Wiskott et al., 1997) is very popular. It is an extension of Elastic Graph Matching for object classes with a common structure, for example faces in the same pose. It uses an approach based on Gabor Wavelets. A two-dimensional face image is represented by a full connected graph in which fiducial points on the face (pupils, mouth, etc.) are the nodes. Each node is then described by linear combinations of a set of wavelets named jets, and each arc is labeled with the distance between the two nodes it connects. A jet describes a small set of gray values (the local texture) around a certain pixel and then is stored in a feature vector that is further used for recognition. From these individual graphs a bunch graph (Face Bunch Graph - FBG) with the same structure is created, with the nodes representing local textures of any face in the class, e.g. all the detected variants of a left eye, and the edges represent the mean distances between the node locations, e.g. the mean distance between the two eyes. A bunch graph represents classes rather than individual objects. Comparison is performed in two steps: a rigid alignment of the grid is performed to handle global transformations, such as translations and scale, then the local misplacement of nodes is evaluated by a Graph Similarity Function. EBGM-based recognition requires that the objects share a common structure. As for faces, this means to match faces in frontal pose, sharing a common set of landmarks. Therefore EBGM can only handle small pose changes. Larger pose differences excessively modify the local features, and, some of them may also be occluded. To address this limitation, it is possible to create bunch graphs for different poses and match all of them to the probe image. This method is robust with respect to variations in pose and illumination, and does not require retraining (unlike Eigenfaces) when new graphs are added. On the other hand, the training as well as the recognition phases are quite slow. (Westphal and Würtz, 2009) have later proposed a graph dynamics that allow an object representation emerge from a collection of arbitrary objects. The underlying idea is to extract typical local texture arrangements from the objects and provide the rules to compose them as needed to represent new objects.

4.3 The Face Recognition Module approach

The method used in the Face Recognition Module of FAIRY has been presented in Barra et al., 2013, where we have proposed a whole mobile oriented biometric system, from the acquisition to the recognition phase. The face recognition module architecture is depicted in the Figure 4.1. It is composed of five steps:
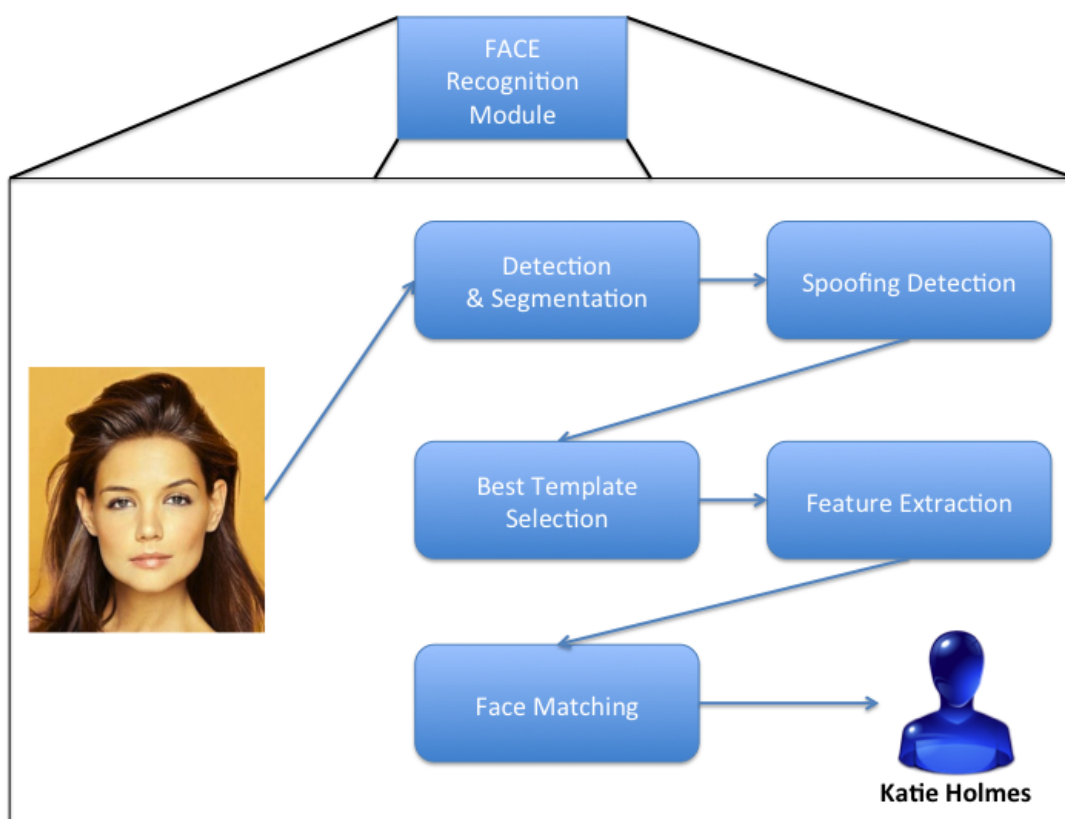
FIGURE 4.1: The architecture of the face recognition module.

a) detection and segmentation, b) spoofing detection (applicable in case of live acquisition of the image), c) best template selection, d) feature extraction and e) matching.

4.3.1 Detection and Segmentation

Once an image is given as input to the module, the detection and segmentation phases begin. This procedure is of paramount importance, since it also provides essential information for the spoofing detection step. Face detection is not limited solely to the localization of the face, but using cascade refining localizes additional reference points on the face. In particular, the module implements the Viola-Jones algorithm (De Marsico, Nappi, and Riccio, 2010a) for the localization of the entire face, so that the subsequent operations are based solely on that sub-region of the image. Once the face region has been cropped, the same Viola-Jones algorithm is reused, obviously with different parameter configurations, to locate the eyes and mouth. A further image correction process is used to overcome varying illumination issues; the Self Quotient Image (SQI) algorithm (De Marsico, Nappi, and Riccio, 2014) is exploited. The basic idea for SQI (see Figure 4.2) is that an image, under suitable conditions, can be decomposed into the product between the illumination component and the shape of the illuminated object. Actually, in the case of the face, the conditions for such an assumption do not appear to be fully met; nevertheless, one can
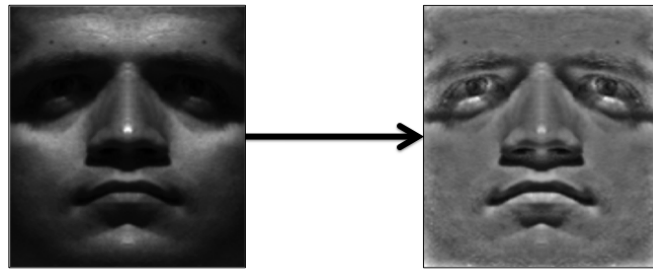
FIGURE 4.2: The SQI approach for the illumination correction.

still obtain a good approximation, which does not compromise the final result. The illumination-corrected image is therefore obtained by applying a smoothing filter to the acquired image and dividing the input image pixel-wise (at one location at a time) to filter the image. A natural benefit of these approaches relies on the integration of quality indices, which exploit additional information produced by the face correction procedures. This allows to discard those biometric samples whose contribution to the subsequent recognition stages might be severely limited if not nil.

4.3.2 Spoofing Detection

The use of liveness detection aims to detect any potential spoofing attempt. Recently this phase has became of basic importance in the last biometric systems, due to the concrete possibility of an individual to attack a biometric system in order to obtain some privileges on sensitive data. In the face recognition module a spoofing detection module is considered. It exploits an invariant 3D technique (De Marsico et al., 2012) which is both simpler and more robust at the same time with the objective to estimate the 3D structure of the face. While on surface it seems to be a much more complicated approach, in fact it is much more efficient. 3D geometric invariants have been traditionally employed for the classification of 3D objects from a single 2D view. In other words, given a 2D image of an object in 3D, it is possible to measure the distances between landmark / reference points lying on the object. When changing the viewpoint from which the 2D image is captured, the distance between those points, of course, would change. For instance, the length of an edge of a cube, in a 2D view, varies depending on the point of view. However, it is possible to identify some relationships between the distances, which remain invariant despite the different point of view from which the object is captured, namely geometric invariants. These relationships are intrinsic to the object and can also be measured on a single image of the object itself. The invariance of these relationships is, however, linked to very specific constraints, which may hold, e.g., collinearity or coplanarity of the points under consideration. The anti-spoofing technique implemented in the face recognition module of the platform, is based precisely on this principle. It identifies a set of five points on the face (outer corner of the right and left eyes, extreme left and right of the face, nose tip), for which the coplanarity constraint is strongly violated. When the user moves her face, the

geometric invariant relative to the identified schema of points is estimated; if the invariant holds, the points comply with the constraint of coplanarity, therefore the captured face image must be a photo (spoofing), otherwise the points are not coplanar and the 3D of the face is guaranteed and thus corresponds to a real (live) user. This technique is extremely straightforward and fast, being implemented by calculating a ratio of determinants of matrices. More details can be found in (De Marsico et al., 2012)).

4.3.3   Best Template Selection

The ease of use of a system is linked to the effort required by the user to employ it. In general, in a biometric system the factor that most affects this aspect is the degree of control required from the user during the acquisition phase; maximum control is generally required by a system for iris recognition, in which the user must stay immobile looking at the camera, while the minimum is characteristic of a face recognition system in a video surveillance application, where the subject is most often unaware of her biometrics being captured. In current solutions this aspect is not considered or is addressed only indirectly. Most of the times some form of robust methods to correct image distortion is pursued, in order to give the user a greater degree of freedom for both pose and movements (Google). Tagging systems, such as those available from Facebook, employ the same type of solution. In the platform, during the acquisition phase, the camera actually acquires a high number of frames, although only one is then used for the recognition process. Through a suitable module, a best sample selection mechanism, based on entropy, is implemented, in order to maximize the accuracy of the recognition operation. This selection mechanism is applied following the operation of face localization, but before all the other processing stages, in order to keep down the computational load of the system. After acquiring the sequence of frames to be examined, the set of faces is passed to the module responsible to select the best sample. Notice that a correction/normalization procedure for both pose and illumination might be exploited in this step, as the one described in (De Marsico, Nappi, and Riccio, 2010a), but we avoided it in order to maintain a lower computational demand. The selection module calculates the correlation between all pairs of faces in the recently acquired sequence of frames / faces and normalizes the value of the correlation index (usually in the interval [-1, 1]) to the range [0,1]. Similarity for the selection of the best template is computed by the (global) correlation between the face images. This technique avoids extracting the biometric keys for samples that will be discarded later on, and allows to evaluate sample similarity by an index, which can be computed faster than the one used for matching, which is discussed later on. Given a pair of samples, the correlation index can be interpreted as the probability that the first sample conforms to the second one (De Marsico, Nappi, and Riccio, 2014). After an appropriate normalization, we obtain a probability distribution on the whole restricted set of samples under examination, which is used for the calculation of the entropy of that set.

The module "discards" a sample from the collection and recalculates the entropy of the remaining set; in addition, it calculates the difference between the entropy before and after that sample "elimination." This is done for all the samples in the sequence ("set") captured. One removes permanently from the set that sample which, when extracted, produces the minimum difference, and the process iterates. The last sample left is selected as the best sample for the recognition phase. It is worth noticing that the algorithm in (De Marsico, Nappi, and Riccio, 2014) for best sample selection rather exploits a parts-based localized version of the correlation index as similarity measure, which provides more accurate results but is much more computationally expensive. The face recognition module uses this more complex version only for face matching as shown below.

4.3.4   Feature Extraction and Matching

Many of the most popular techniques for face classification, such as Linear Discriminant Analysis (Etemad and Chellappa, 1997), are still too sensitive to image distortions to be profitably used in commercial applications in partial or complete uncontrolled contexts. The face recognition module implements a matching technique based on a localized version of the spatial correlation (De Marsico, Nappi, and Riccio, 2010a; De Marsico, Nappi, and Riccio, 2014). In its global form, given two images A and B and their respective mean values $\overline{A}$ and $\overline{B}$ for pixel intensity, correlation is computed as:

$$s(A,B) = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}(A(i,j)-\overline{A})(B(i,j)-\overline{B})}{\sqrt{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}(A(i,j)-\overline{A})^2\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}(B(i,j)-\overline{B})^2}}$$

In the recognition phase implemented in the face recognition module, correlation is adapted to work locally, on individual subregions r_A and r_B of the images A and B. In particular, for each sub-region r_A in A, the region r_B that maximizes the correlation coefficient $s(r_A, r_B)$ is searched in a narrow window around the corresponding position in B. The global correlation $S(A, B)$ between the two images A and B is then obtained as the sum of the local maxima found. This approach, although more precise, is more expensive from a computational point of view. However, the pre-calculation of some quantities involved in the comparison formulas, code optimization, and reduced resolution allow to perform a considerable number of verification checks (in the order of tens) in less than a second using a mobile architecture. The features extracted from the face and stored in what is called *biokey*, are represented by the pre-calculated factors, which are necessary to calculate the correlation sought. When we are not dealing with a verification operation, where the user claims an identity, but rather with an identification protocol, each image must be matched against those stored in a certain gallery. For each identity g_k that is registered in the gallery $G(k = 1, ..., |G|)$, one or more images I_j, $j > 0$ are stored. When

a new query image q is submitted to the face recognition module, it compares it with all images in the system and computes the corresponding correlation indices. The obtained list of values is reorganized in decreasing order, and the identity g_k with more images in the first n positions, is returned as the retrieved identity.

4.4   Experimental Results

As mentioned in Section 4.1, the presented face recognition approach has been proposed for being used on mobile devices. Therefore, the tests have been run on a Samsung Galaxy Tab 2.0 with 10.1-inch screen and 1280x800 pixel resolution. The tests were carried out with twenty-five subjects. Each experiment was divided into a registering enrollment operation and four access attempts. In order to have an overall assessment of the accuracy of the system, the mean values are reported, namely FAR = 0.055 and FRR = 0.177; these results have been obtained for all authentication attempts made during the tests. Analyzing the tests performed by users, there is a higher presence of false negatives; however, while these might be disturbing for the user, they do not undermine the security of the system. Information regarding the execution time are also provided. Such information is based on an estimate of the time required by the application to complete the operations of registration (4.82 sec.) and authentication (5.04 sec).

Chapter 5

Iris Biometrics

5.1 Introduction

Among the physical biometrics, the iris is without any doubt the most discriminative trait, due to the irregular and random pattern composed by muscle fibers, needed for regulating the dilation of the pupil, and blood vessels. Three factors make the iris perfectly suitable for biometric systems:

- the high interclass variability;

- the low intraclass variability;

- the permanence over time.

The recognition approach described in this chapter has been proposed in (Barra et al., 2015b). The authors implemented a biometric system aimed at iris recognition on mobile devices. The feature vector of the iris is based on the spatial histogram (*spatiogram*) and the method has been tested over three datasets: MICHE-I, UBIRISv1 and UPOL. The first is the most challenging since it has been acquired in both controlled and uncontrolled conditions, by means of built-in cameras of three diffused mobiles. In the 5.2 the related works are shown; in the section 5.3 the algorithm is detailed, and the experimental results will be reported together with the related curves.

5.2 Related Works

Iris has been explored as a biometric firstly in the seminal work by J. D. Daugman (Daugman, 1993) who assessed the statistical independence of two coded patterns originated from different eyes and described a method for iris localization based on integro-differentials operators, the use of 2D Gabor filters to extract iris texture features and a statistical-based approach to iris codes matching. This work has been further expanded in (Daugman, 2003; Daugman, 2004). In 1996, Wildes patented a non-invasive system for iris recognition and compared that to the one from Daugman (Wildes, 1997). These pioneering works were mostly focused on achieving maximum accuracy in iris recognition under controlled conditions including specific enrollment protocols for the user to undergo. The following studies on this biometric have progressively targeted

from the one side iris segmentation and from the other side its recognition under less predictable acquisition conditions, involving uncontrolled lighting and environmental factors which can result in noisy iris images (for instance due to strong reflections over the cornea surface, blur, low contrast, etc.). To this regard (Boles and Boashash, 1998) proposed an approach to represent iris features by means of wavelet transform zero crossing. This descriptor is translation, rotation, and scale invariant and very slightly affected by variations in illumination and noise levels. In (Lim et al., 2001) the Haar wavelet transform is used to optimize the dimension of feature vectors to 87 bits, to the aim of reducing processing time without affecting accuracy of recognition. The authors combined this descriptor with a method of initializing weight vectors and a method of determining winners for recognition in a competitive learning neural network, resulting in a method accurate even for "real world" applications. In 2007, (Proenca and Alexandre, 2007) presented the results of the NICE-I contest for the performance evaluation of recognition algorithms on noisy iris images, that resulted in a measurable robustness improvement of the state of the art, particularly for iris segmentation. More recently, Bowyer resumed the results of the NICE.II Iris Biometric Competition (Bowyer, 2012), focused on performance in feature extraction and matching, arguing that "since the top-ranked algorithms seem to have relatively distinct technical approaches, it is likely that a fusion of the top algorithms would result in further performance improvement". In this line of research, (Jeong et al., 2010) presented a new iris segmentation method that can be used to accurately extract iris regions from non-ideal quality iris images while (Shin et al., 2012) proposed an integrated iris recognition method that discriminates the left or right eye on the basis of the eyelash distribution and specular reflection and exploits iris region color and texture information to achieve a reliable classification.

5.3   The iris recognition module approach

The iris recognition module is mainly composed by two phases: iris detection/segmentation and feature extraction/matching. The first phase exploits IS_{IS}, the iris segmentation system introduced in (De Marsico, Nappi, and Riccio, 2010b). This method is composed by four steps:

- Iris image pre-processing

- Pupil localization

- Image linearization

- Limbus localization

The second phase is based on features extracted from the spatial histogram of the iris. An overview of the iris recognition module is depicted in Figure 5.1
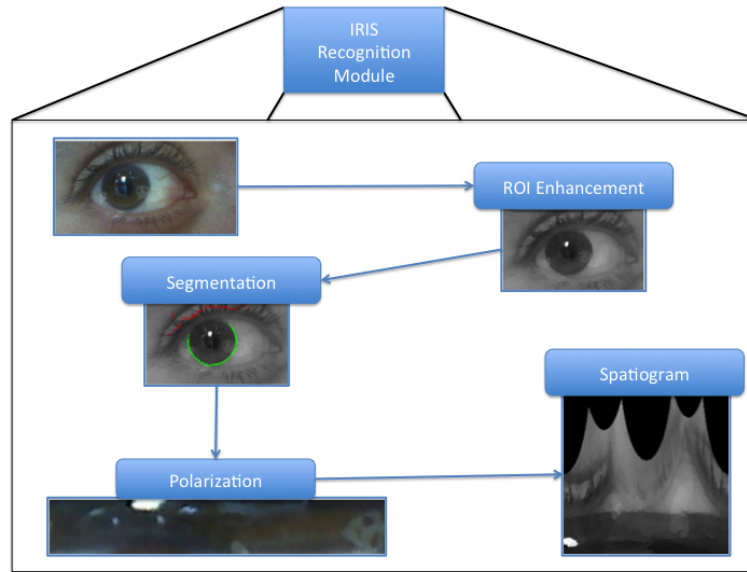
FIGURE 5.1: The feature extraction process in the Iris Recognition Module.

5.3.1 The iris detection and segmentation phase

Homogeneity and separability are important concepts used for the selection of the best circle, which is a candidate to represent the pupil. The pupil can be considered as a circular region with a homogeneous distribution of pixels. Some approaches are based on the assumption that the darker region within the image is the pupil. Nevertheless, this is not always true because the pupil may change its appearance in relation to the lighting. The proposed scoring function is based on the gray-scale histogram of the pupil. Each circular region receives a score according to the homogeneity degree of the pixels it contains. Let H be the region histogram, s_H will be the maximum number of occurrences of the same value in that region:

$$s_H = max_i[H(i)]/_{i=1}^{255}H(i)) \tag{5.1}$$

As for the limbo, also the outline of the pupil has a zone in which it passes from a dark color to a lighter one; of course if you analyze an iris dark in color, this transition is more subtle. Therefore we define an index of separability. Given a candidate circle C with center $c = (c_x, c_y)$ and radius ρ in the image I, the Cartesian coordinates are given by:

- $x_C(\rho, \Theta) = c_x + \rho * cos\Theta$

- $y_C(\rho, \Theta) = c_y + \rho * sin\Theta$ where $\Theta \in [0, 2\pi]$

Considering the circle C_{in}, internal to C, with radius $\rho_1 = 0.9\rho$ and the circle C_{out}, external to C, with radius $\rho_2 = 1.1\rho$, measuring the difference of gray tones, on the edge of the circle for each angle Θ_i, using an operator similar to

the Daugman's integro-differential operator, structured as follows:

$$D(i) = I(x_C(\rho_2, \Theta_i), y_C(\rho_2, \Theta_i)) - I(x_C(\rho_1, \Theta_i), y_C(\rho_1, \Theta_i)) \qquad (5.2)$$

where $i = 1, ..., 360$ represents the discrete value of the angle and then the index within the gradient vector h; while $\Theta_i = i\pi/180$ is the same angle, in radians. At the pupil, we expect a high and constant value for D, i.e. an average high and low variance $\sigma(D)$. Based on these observations, the index of separability can be defined as:

$$s_D = \frac{\overline{D}}{\sigma(D) + 1}. \qquad (5.3)$$

By analyzing the polarized image of eye in the vertical direction, (see Figure 5.2), it is possible to accurately define the region of separation between the iris and the sclera, which corresponds to limbo. However, the captured iris image
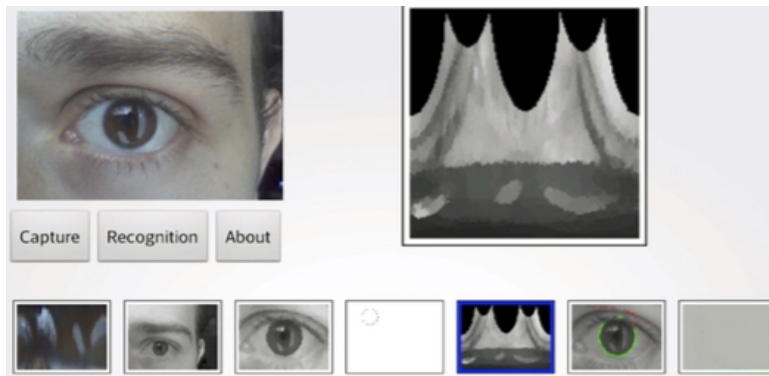
FIGURE 5.2: The polarization of the eye with the limits that separate the sclera and iris pupil. From the left, subimages show the iris extracted, the grayscale converted frame, the enhanced ROI, the edges of the iris, the polarized iris and, a graph showing the distribution of pixels in the grabbed frame

contains some information that is not useful for the localization of the pupil. This is the case for features such as the pores of the skin, eyelashes and eyelids, which negatively interfere with edge detection. The first stage of IS_{IS} eliminates interference through an enhancement filter. A square window W of size kxk, is passed over the entire image pixel by pixel; for the window W, a histogram H_W is calculated and the value with the highest occurrence is replaced in a central position. A "canny" filter is applied to the image resulting from the previous stage to locate the pupil. The canny filter is applied with ten different thresholds $th = 0.05, 0.010, 0.015, ..., 0.055$. For each of the eleven images the connected components are identified. The components containing a number of pixels above a threshold th_C are included in a list L. Then, the algorithm of Taubin (Taubin, 1991) is applied to each element of the list to calculate the corresponding circle. The circles that are not completely within the image are eliminated from the list L, in order to obtain a final list L_C. To identify the pupil, the circles of the list L_C undergo a procedure based on the homogeneity and separability criteria. For each circle the value: $s = s_H + s_D$ is calculated.

Finally, the circle C_{max} with the highest s_{max} value is considered as the circular shape that best approximates the pupil. At this stage, the pixels with the ρ_{max} distance from the center of the localized pupil are searched. Then, the image is transformed from Cartesian coordinates to polar ones, so obtaining a new image $\overline{I}$ (Figure 5.2). This transformation is made to locate the boundary between the sclera and the iris. The image $\overline{I}$ is further filtered with a median filter. If R is a row of the image, for each pixel P contained in R we consider the neighborhood including $2q+1$ pixels (i.e. the q previous pixels, the pixel p itself and the q following pixel). Then the neighborhood pixels are sorted and the pixel P takes the median value. For each column, which is located beyond ρ_J and the corresponding position on the horizontal axis of i and θ_i, the following weighted difference is calculated pixelwise:

$$\Delta(\rho_j, \Theta_i) = \varphi(I, \rho_j, \Theta_j) \cdot (I(\rho_j + \delta, \Theta_i) - I(\rho_j - \delta, \Theta_i)) \tag{5.4}$$

where:

$$\varphi(\overline{I}, \rho_j, \Theta_i) = \begin{cases} 1 & \text{if } \overline{I}(\rho_j + \delta, \Theta_i) - \overline{I}(\rho_j - \delta, \Theta_i) > 0 \\ & \text{and } min(\overline{I}(\rho_j + \delta, \Theta_i), \overline{I}(\rho_j - \delta, \Theta_i)) > \varepsilon_G \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

It is worth noting that the pupil occupies the lower part ρ_J of the polarized image $\overline{I}$, followed by the iris and the sclera. The sign of the difference is relevant as it is expected that the sclera is brighter than the iris. This indicates that one should look for changes with a positive sign, which represent the transition region between iris and sclera. In formula 5.5: the first inequality imposes a positive gradient; the second inequality excludes the pixels of the border between the pupil and the iris, as it requires the darkest pixel in the pair to have a gray level greater than a threshold $\varepsilon_G \in [0, 255]$. The area on the limbo is composed of points that maximize the weighted difference 5.4 for each column Θ_i in I.

5.3.2 Feature extraction and matching

For a given discrete function $f : x \to v$, where $x \in X$ and $v \in V$, a histogram of f counts the number of occurrences for each element in the range of f. In particular, the histogram is $h_f : v \to Z*$, where $v \in V$ and $Z*$ is the set of positive integers, and $h_f(v)$ is the number of elements $x \in X$ such that $f(x) = v$. The histogram h_f can be seen as a binary function $g_f(x, v)$, where

$$g_f(x, v) = \begin{cases} 1 & \text{if } f(x) = v \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

The moment of zero order of g on the dimension v is:

$$h_f(v) = \sum_{x \in X} g_f(x, v) \tag{5.7}$$

Histograms are relevant to iris segmentation because they discard all the information about the domain, so as to create the respective invariant for the one by one transformations of domain of the original function. A limited amount of information regarding the domain, can be extrapolated by means of higher order moments to the binary function g, where the $i - th$ order moment is given by:

$$h_f^{(i)}(v) = \sum_{x \in X} x^i g_f(x, v) \tag{5.8}$$

To define this concept, we use the term *spatial histogram* or simply *spatiogram*; this because it captures not only the occurrences of information relating to the range of the function, such as the histogram, but also the information relating to the spatial domain. We define the $k - th$ order spatiogram as a tuple containing all the moments up to k:

$$\langle h_f^{(0)}(v) \dots \dots h_f^{(k)}(v) \rangle \tag{5.9}$$

A histogram is thus only the zero-order moment of a spatiogram. The spatiogram can also be seen as a geometric model that allows arbitrary transformations and more specific models such as: translation, similarity, etc. As histograms, the spatiograms are efficient to calculate the differences between the correspondences of the images, without precisely calculate the geometric transformation between them. However, even spatiograms as more specific models, retain information about the geometry of the region of the image. The spatiograms, differently from the comparison of co-occurrences between arrays, capture the global position of the pixels instead of the relation between their pairs. An image is a two-dimensional map $I : x \to v$ of pixel $x = [x, y]^T$ with v values. The pixel value may represent an arbitrary value: gray tones, colors, or the result of a preprocessing (quantization, the color transformation of the space, etc.). The second order spatiogram of the image can be represented as:

$$h_I^{(2)}(b) = \langle n_b, \mu_b, \Sigma_b \rangle, \text{ with } b = 1, \dots, B \tag{5.10}$$

where n_b is the number of pixels whose values are represented by the $b - th$ bin, μ_b is the mean vector and $\sum b$ are the covariance matrices. $B = |V|$ is the number of bins in the spatiogram. Given h and h' as the enrolled spatiogram and the currently acquired one respectively, the similarity between them can be calculated as the following weighted sum:

$$\rho(h, h') = \sum_{b=1}^{B} \Psi_n \rho_n (n_b n'_b) \tag{5.11}$$

For a zero-order spatiogram $\psi_b = 1$. For a second order spatiogram, ψ_b can be seen as the probability that x_b is calculated by a Gaussian distribution described

by multiplying the probability in the reverse direction:

$$\Psi_b = \eta \exp\{-\frac{1}{2}(\mu_b - \mu'_b)^T \hat{\sum_b}^{-1}(\mu_b - \mu'_b)\} \tag{5.12}$$

where η is the Gaussian normalization constant, while $\hat{\sum_b}^{-1} = (\sum_b^{-1} + (\sum'_b)^{-1})$ is a covariance matrix. It should be noted that the values of the summation are the average of the two Mahalanobis distances, the first one calculated between x and x' and the other one calculated between x' and x.

5.4   Experimental Results

The approach has been tested over three iris datasets: MICHE-I, UBIRIS and UPOL. MICHE-I (De Marsico et al., 2015) is totally composed by images acquired mobile devices; this aspects provides randomness to the images (due to the fact that the shot is done by the subject itself) and set the testing phase in a more realistic environment. Three different mobile devices have been used, in order to achieve a cross sensor comparison and provide further indicative values concerning the kind of cameras:

- Apple iPhone5

    - Operating system: Apple iOs;
    - Rear camera: iSight, 8 Megapixels;
    - Front camera: Facetime HD camera, 1.2 Megapixels;

- Galaxy Samsung IV

    - Operating system: Google Android;
    - Rear camera: CMOS, 13 Megapixels;
    - Front camera: CMOS, 2 Megapixels;

- Galaxy Tablet II

    - Operating system: Google Android;
    - Rear camera: N/A;
    - Front camera: 0.3 Megapixels;

An example of the acquired images is shown in Figure 5.3. The whole acquisition protocol is detailed in (De Marsico et al., 2015). UBIRISv1 (Proença and Alexandre, 2005) is composed by 1877 images in two sessions. While in the first one the noise factors are minimized, in the second one the images are affected by different kind of disturbs (reflections, blurring, different light conditions, ...). This provides reliability and robustness to the recognition method. Examples of irises from UBIRISv1 dataset are shown in Figure 5.4. UPOL (Dobeš et al., 2004; Dobeš et al., 2006; Dobeš, 2008) images have been captured through an
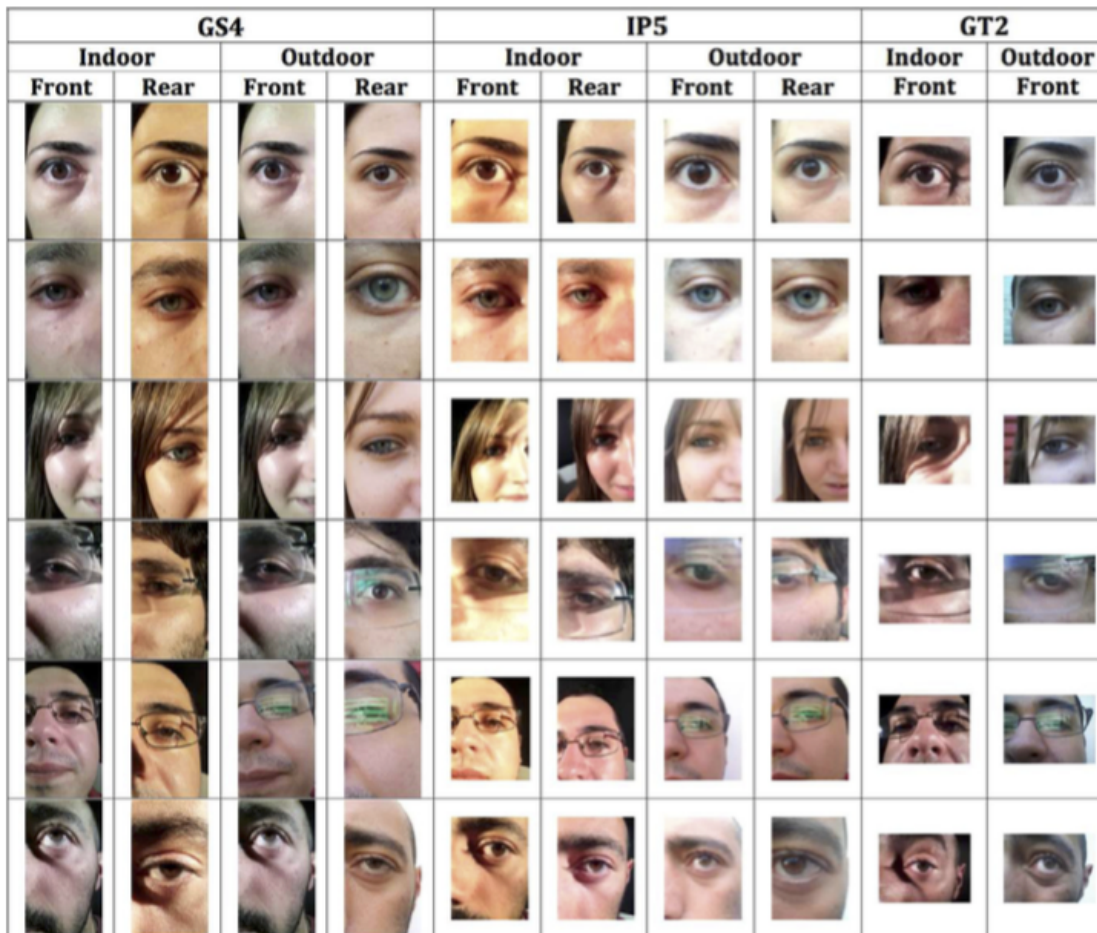
FIGURE 5.3: Samples from MICHE-I dataset.
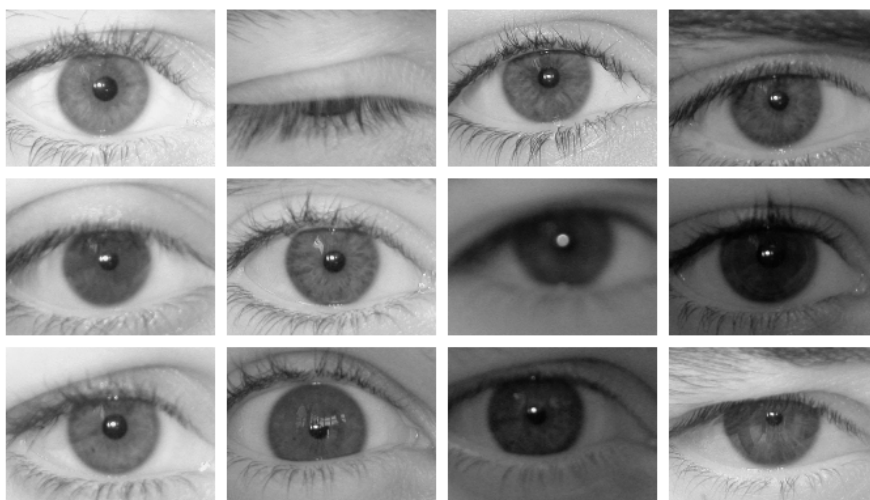
FIGURE 5.4: Samples from UBIRISv1 dataset.

optometric framework (TOPCON TRC50IA) and due to this, are of extremely high quality. Therefore they are suitable to be used for the evaluation of iris

recognition methods in completely noise-free environments. The dataset contains 384 images acquired from 64 subjects. An example of the irises from the UPOL database is shown in 5.5. We first wanted to assess the performance of
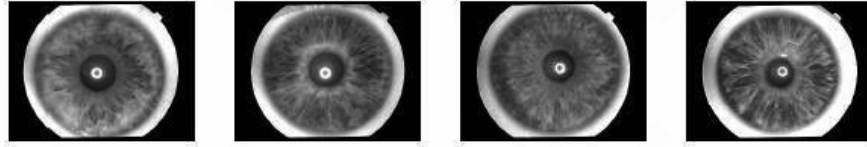
FIGURE 5.5: Samples from UPOL dataset.

the proposed iris recognition method described in section 5.3 on UBIRIS and UPOL. Figure 5.6 shows respectively the CMS and the ROC curves resulting for both the aforementioned datasets as a combined view (actually a subset of UPOL composed by 173 elements as probe and 173 as gallery and a subset of UBIRIS including 114 elements as probe and 228 elements as gallery). In these first experiments the method performed well; in particular, with regard to the ROC curve, spatiograms performed better on UPOL than on UBIRIS until FAR 0.6 is reached, while after that point the two curves are very similar. For what concerns the CMS, the situation is almost reversed, as in this case the performance on UBIRIS is clearly better with a CMS value near 0.7 for rank 1 compared to rank 9 on UPOL. This first testing is needed in order to have a baseline to compare the further experiments with. Three groups of experiments
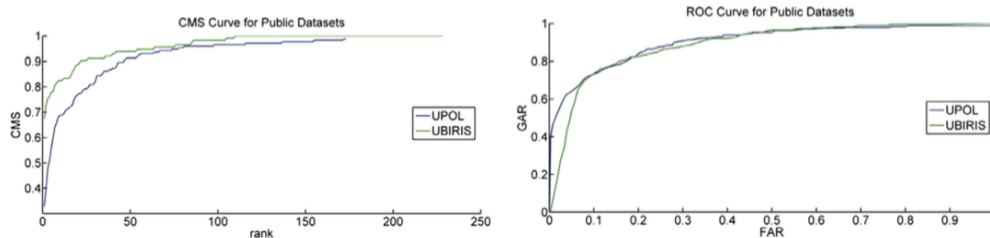
FIGURE 5.6: Comparison of CMS and ROC curves for UPOL and UBIRIS datasets.

have been considered to fully assess the performance of the method. The first group of experiments is aimed to measure performance in case the device that the probe has been acquired from and the device of the gallery images are the same; therefore, IP5, GS4 and GT2 are compared in indoor environment with regard to ROC and CMS curves, by using for both probe and gallery images captured by the same camera, i.e. front camera for both probe and gallery (see Figure 5.7) and rear camera for both probe and gallery(see Figure 5.8). The second group of experiments is aimed to measure performance in case the device the probe has been acquired from and the device of the gallery images are different. IP5, GS4 and GT2 are compared in indoor environment with regard to ROC and CMS curves, by using for both probe and gallery images captured by the same camera, i.e. front camera for both probe and gallery and rear camera for probe and gallery (see Figure 5.9). The third group of experiments is
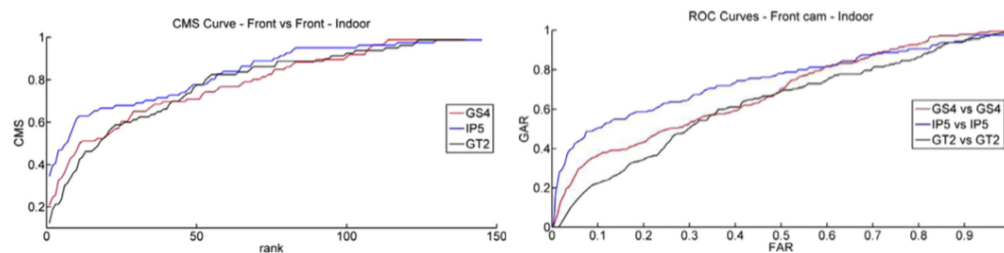
FIGURE 5.7: Experiment 1, comparison of CMS and ROC curves. Probe and gallery acquired indoor, both through rear camera of Galaxy S4 and iPhone 5.
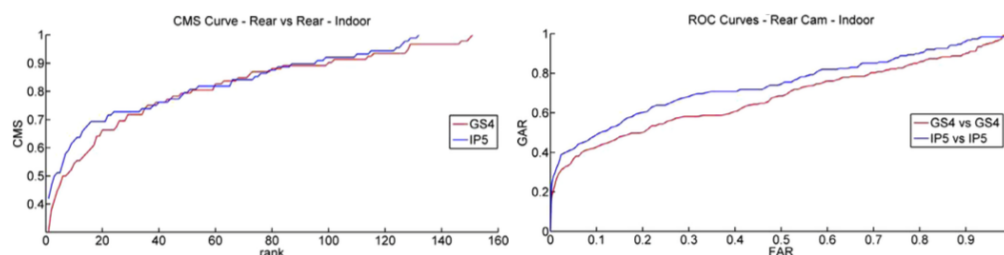
FIGURE 5.8: Experiment 1, comparison of CMS and ROC curves. Probe and gallery acquired indoor, both through front camera of Galaxy S4, iPhone 5 and Galaxy Tab 2.
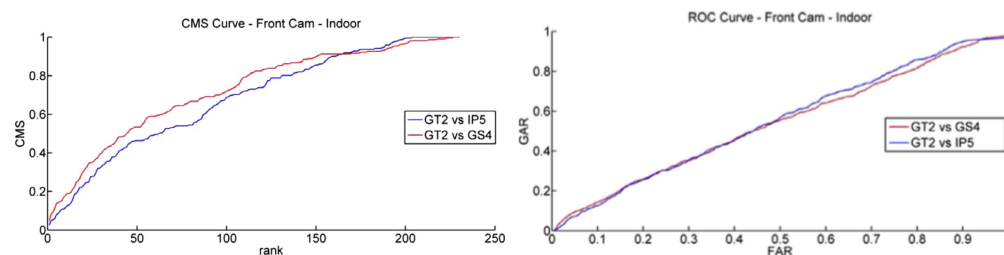
FIGURE 5.9: Experiment 2, comparison of CMS and ROC curves. Probe and gallery acquired indoor, respectively through front camera of Galaxy Tab 2 and iPhone5 or through front camera of Galaxy Tab 2 and Samsung S4.

aimed to measure indoor-vs-outdoor performance in case the devices the the probe and the gallery have been acquired from are the same IP5, GS4 and GT2 are compared by using for probe and gallery images captured by the same camera but in different environments, i.e. probe captured outdoor, gallery captured indoor (see Figure 5.10) and vice-versa (see Figures 5.11). Overall, the results of experiments 1, 2 and 3 are substantially worse than those achieved on UPOL and UBIRIS datasets. This appears (at different extents) in both ROC and CMS performance, whose curves are characterized by a clear and constant drop (compared to their counterparts for the reference datasets) throughout the various experiments. By carefully analyzing the intermediate results of the processing pipelines (i.e. the results of iris segmentation and the features extracted) and also considering these results with regard to each type of device and its
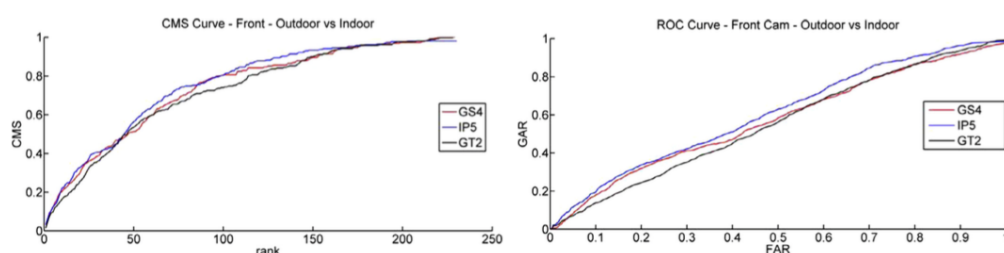
FIGURE 5.10: Experiment 3, comparison of CMS and ROC curves. Probe acquired outdoor and gallery acquired indoor, both through front camera of Galaxy S4, iPhone 5 and Galaxy Tab 2.
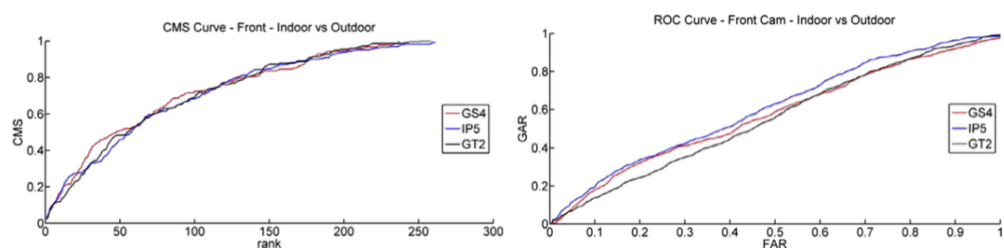
FIGURE 5.11: Experiment 3, comparison of CMS and ROC curves. Probe acquired indoor and gallery acquired outdoor, both through front camera of Galaxy S4, iPhone 5 and Galaxy Tab 2.

technical specifications, it seems reasonable to explain them in the light of the challenging characteristics often exhibited by the samples collected in MICHE-I. This is partly due to the subject's self-operated capture of the eye-region and partly related to the environmental conditions, sometimes featuring strong ceiling lighting or a mix of sunlight and artificial lighting. For what concerns the performance of each device involved in the experimentation, it is interesting to note that the top performer resulted to be Apple's iPhone 5, despite the lower resolution of its front and rear imaging sensors compared to Samsung's S4 cameras, while, not so surprisingly, the Samsung's Galaxy Tab 2 is constantly the worst performer. These findings could be related to the IP5's built-in image processing that could have affected more favorably the iris localizations, segmentation and feature extraction than the GS4 did. Moreover, a typical digital photography rule, implies that for similarly sized image sensors, the highest the resolution the higher the noise resulting in the image. But the bad GT2 performance also proves that, noise level apart, a good level of detail in the image captured is almost mandatory for accurate features extraction in iris images.

Chapter 6

EEG and ECG Biometrics

6.1 Introduction

The introduction of physiological measures into a multi biometric platform is one of the novelties of this dissertation. In FAIRY, two modules concerning EEG and ECG recognition have been inserted. These signals have been widely studied from a diagnostic and medical point of view. In the last years, they have also been investigated as possible biometric markers. The ECG and EEG recognition modules exploit two methodologies for achieving the recognition of these physiological signals. The first is based on the detection of peaks in the ECG. The peaks are points of local maxima and local minima in the signal. The second method is based on the power spectral method and has been used for extracting the features from the EEG. The fusion of both signals is performed at score-level by means of the weighted sum, a simple as well as efficient approach. Tests of the ECG and EEG recognition methods have been executed over the PTB and the EEGMMI databases respectively, both available on the Physionet bank. For testing the fusion, a chimera dataset has been built. The related works, the method and the experimental results will be further detailed in the following.(Barra et al., 2015a)

6.2 Related Works

Taken separately, EEG and ECG signals have been widely investigated in the last years. For both signals, most of the researches lies in the field of diagnostic. Applications for heart attack prevention (Prittopaul, Sathya, and Jayasree, 2015), (Ravish et al., 2014), as well as for arrhythmia detection (Noh, Hwang, and Jeong, 2011) have been developed. On the EEG side, the signal has been mostly studied for prevention and detection of epileptic events and seizure (Shantha Selva Kumari and Prabin Jose, 2011), (Pan et al., 2008). For a comprehensive review of the relevance of EEG in neuroscience, see (Lopes Da Silva, 2013). On the contrary, the joined analysis of both signals has been scarcely investigated. Most applications that exploit the information brought by ECG and EEG combined together (Soria-Frisch, Riera, and Dunne, 2010), (Riera et al., 2008), (Riera et al., 2009), mainly lie in the HCI (Human Computer Interface) field, like in (Shahid, Prasad, and Sinha, 2011) and (Sakai and Wei, 2008). In the

following subsections, the state of art in the biometric field is presented, w.r.t. the use of the two signals separately and w.r.t. their combined use.

6.2.1 ECG in Recognition

The interclass variability of the ECG has always been widely investigated since the very first researches in this field (Draper et al., 1964), (Kozmann, Lux, and Green, 1989), (Hoekema, Uijen, and Van Oosterom, 2001). Basing on these preconditions, the interest moved to the biometric aspects of the signal and to its discriminative capability. Therefore, in the last ten years, several recognition methods have been developed. The main difference among these techniques regards the type of characteristics used for the construction of the feature set. In the literature there's an equitable distribution among methods that use fiducial points and fiducial features in the signal (Biel et al., 2001), (Kyoso and Uchiyama, 2001), rather than basing on non-fiducial points (Agrafioti and Hatzinakos, 2008), (Plataniotis, Hatzinakos, and Lee, 2006). Some hybrid approaches have also been presented (Shen, Tompkins, and Hu, 2002), (Matos, Lourenço, and Nascimento, 2014), but although they obtained good results, this is not a very explored branch. Another categorization regards the number of leads used for the feature vector construction. Two very detailed and comparative analysis about the ECG in recognition processes are presented in (Odinaka et al., 2012) and (Fratini et al., 2015).

6.2.2 EEG in Recognition

As cited above, the brain waves have always been extensively investigated in medical field, especially as regards the diagnosis of cerebral diseases and the prevention of epileptic events. Only recently, the biometric point of view of the EEG's signals has been studied. It seems that the problem of subject identification by EEG signals is harder than expected and the fusion of information from different biometrics may represent an important solution. Up to now, it has been shown that power spectrum features extracted from EEG contain individual-specific information that allow to reach high classification performance. For a detailed overview of the related techniques see (Pozo-Banos et al., 2014) and (Campisi and La Rocca, 2014). More recently methods based on functional connectivity (La Rocca et al., 2014) and network organization (Fraschini et al., 2015) have been proposed.

6.2.3 Physiological Signals Fusion in Literature

Physiological signals fusion is mainly achieved for medical purposes. In the neurological field, for example, fusion of EEG and ECG signals has been exploited in order to detect epileptic events (Bermudez, Lowe, and Arlaud-Lamborelle, 2009). In (Rosli et al., 2014), the authors used a Bayesian network to combine

ECG and EMR(Electromyographic) signal for the development af a rehab application. Further interesting applications have been developed in the psychological field: in (Koné et al., 2015) and (Verma and Tiwary, 2014) biomedical signals have been combined for applications of emotion classification and recognition. As regards the biometric field, (Soria-Frisch, Riera, and Dunne, 2010) shows the result of the fusion of EEG and ECG using 5 different operators over 29 subjects, whereas in (Riera et al., 2008) the authors proposed a wearable wireless biometric system based on EEG and ECG.

6.3 The ECG Recognition Module approach

In order to obtain the feature vector representing the ECG's signature, the signal of the lead i is taken into account. Five random non-overlapping epochs (twelve seconds each) have been randomly selected. Each epoch contains 12.000 values, representing the ECG wave, given the sampling rate of 1000Hz. For each signal, one epoch is inserted in the probe set, and the remaining four in the gallery. Based on the sampling rate, an ad-hoc detrend operation is executed to normalize the baseline of the signal, by subtracting the best-fit straight line. Such an operation is done in order to focus only on the fluctuation of the signal and not on the inclination of the curve. In order to better explain the effect of this operation, the signals, before and after the detrend, are shown in Figure 6.1 Then, a cutoff of the frequencies less than 3Hz and greater than 15Hz, has been

a) b)
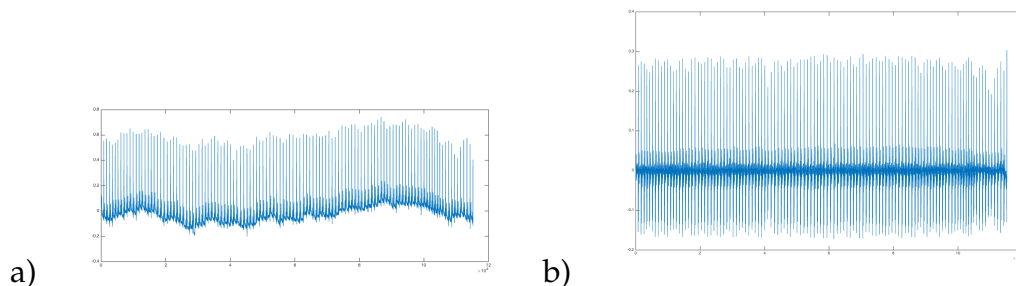
FIGURE 6.1: An ECG signal before (a) and after (b) the detrending operation.

done. The feature extraction process has been based on a simple peak detection method, that selects the local minima and local maxima from a given signal. The use of this technique has the advantage that is both very precise and fast in the execution. Based on some preliminary experiments reported in (Barra et al., 2015a), the vector is sorted in descending order. Finally a selection of the 48 most prominent peaks has been done. In figure 6.2 the extraction process of the features of the ECG is shown. The distance between probe and gallery samples is calculated by means of the *norm1* method.
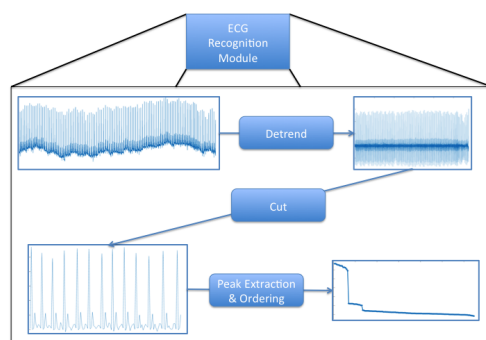
FIGURE 6.2: The feature extraction process for the ECG.

6.4 The EEG Recognition Module approach

As for the EEG signal, each recording was off-line band-pass filtered into the following frequency bands:

- delta (0.5 - 3.5 Hz);

- theta (4 - 8 Hz);

- alpha (8 - 13 Hz);

- beta low (13- 20 Hz);

- beta high (20 - 30 Hz);

- gamma (30 - 45 Hz)

For each band, five 12-seconds epochs have been randomly selected. Each epoch is composed of 1.920 values, given the low sampling rate of the EEG signal, i.e. 160 Hz. A detrend operation normalizes the baseline of the signal. As for the EEG feature extraction procedure, a technique based on spectral features computed with power spectral density (PSD) method has been implemented. In figure 6.3 the extraction process of the features of the EEG is shown. The relevance of using a method based on Fourier transform is due to the inherently rhythmic trend of EEG signal. The PSD of each EEG epoch was extracted by computing Welch's averaged modified periodogram method, using a Hamming window with 50 percent overlap. The procedure was accomplished for each subject, epoch and frequency band. Also for the EEG, the matching method used for comparing probes and gallery samples has been the *norm1* method.

6.5 EEG and ECG fusion

In order to perform the fusion of the vectors, three operators have been exploited: sum, product and weighted sum. We preferred to perform a score level fusion, for the following reasons:
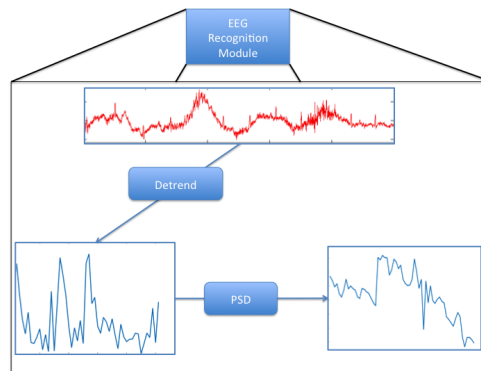
FIGURE 6.3: The feature extraction process for the EEG.

- a sensor level fusion is not possible, since we operate on well-known datasets (see section 6.6) and we may not access any sensor;

- performing signal level or feature level fusion means to combine heterogeneous data, since EEG and ECG have different ranges of values;

- the decision level fusion implicates too loss of information; moreover is not fair as regards the contribution of both the biometric systems.

6.6 The Datasets

The ECG and EEG recognition approaches have been tested over two datasets published in the Physionet bank(Goldberger et al., 2000): the *PTB Diagnostic ECG Database* and the *EEG Motor Movement/Imagery Dataset* respectively. Then, in order to test the fusion approach, an extended test set has been constructed by combining the cited datasets. In the following paragraphs, the datasets are presented together with the *Chimera* Subjects, used for testing the fusion approach.

6.6.1 The PTB Diagnostic ECG Database

The PTB Diagnostic ECG Database (*PTB*) (Bousseljot, Kreiseler, and Schnabel, 1995) contains 549 ECG recordings from 290 subjects of different ages, each represented by one to five registrations. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) and the 3 Frank lead ECG's (vx, vy and vz). Each signal is digitized at 1000 samples per second. A detailed clinic summary is available for most of the ECG records within a header file (.hea) attached to the data file (.dat). Besides the generic information about the subject (age and gender), the clinic summary particularizes the diagnosis, and, where applicable, medications, interventions and other data on the medical history. The data files have been analyzed with the WFDB (WaveForm Database) Library, under the MATLAB environment.

6.6.2   EEG Motor Movement/Imagery Dataset

EEG Motor Movement/Imagery Dataset (*EEGMI*) (Schalk et al., 2004) consists of over 1500 one and two minutes EEG 64-channel recordings, acquired from 109 volunteers and digitized at 160 samples per second. For each subject, 14 recordings are provided: two baselines recordings, six imagery actions recordings and six motor actions recordings. The two baseline recordings correspond to acquisitions while the subjects are in resting state conditions with opened and closed eyes respectively. The six imagery actions recordings are acquired while the subject was imagining to perform an action, whereas the six motor action recordings are acquired while the subject was really performing the same actions. The data are provided in EDF+ extension and therefore analyzed with PhD toolbox running under the MATLAB environment.

6.6.3   The *Chimera* Subjects

Starting from the two datasets described in their respective paragraphs, in order to properly test the proposed biometric trait, 52 *chimeras* have been created. Given an EEG signal eeg_i belonging to the EEGMMI dataset and an ECG signal ecg_j, belonging to the PTB dataset, a *chimera* subject c_{ij} is defined as follows:

$$c_{ij} = (eeg_i, ecg_j); \tag{6.1}$$

Since it is meant to test ECG recordings free of any artifact, arrhythmia or heart disease that may affect the ECG wave, only the healthy subjects in the PTB database have been taken into account. For this reason, among the 249 subjects of the PTB database, only 52 have been selected. As a consequence, 52 subjects out of the 109 of the EEGMMI database have been randomly chosen. Finally, 52 *chimeras* are created according to 6.1, by randomly associating 52 different EEG's to the 52 ECG's.

6.7   Results Discussion

In Table 6.1, the EER values for ECG and for each of the bands of the EEG are reported. The results claim that the ECG has the strongest discriminative capability, in compliance with the related ROC curves in Figure 6.4. In the Table 6.1, the EER values together with the AUC values are presented.   The fu-

TABLE 6.1: EER values and AUC for both ECG and EEG signals.

	ECG	Alpha	LBeta	HBeta	Gamma	Delta	Theta
EER	**1.33%**	29.36%	26.58%	22.97%	23.0%	23.08%	29.11%
AUC	**96.15**	71.13	77.34	76.40	74.95	80.01	70.84

sion process has been achieved exploiting the basic fusion operators: the sum, the product and the weighted sum. According to these operators, the distance

a

b
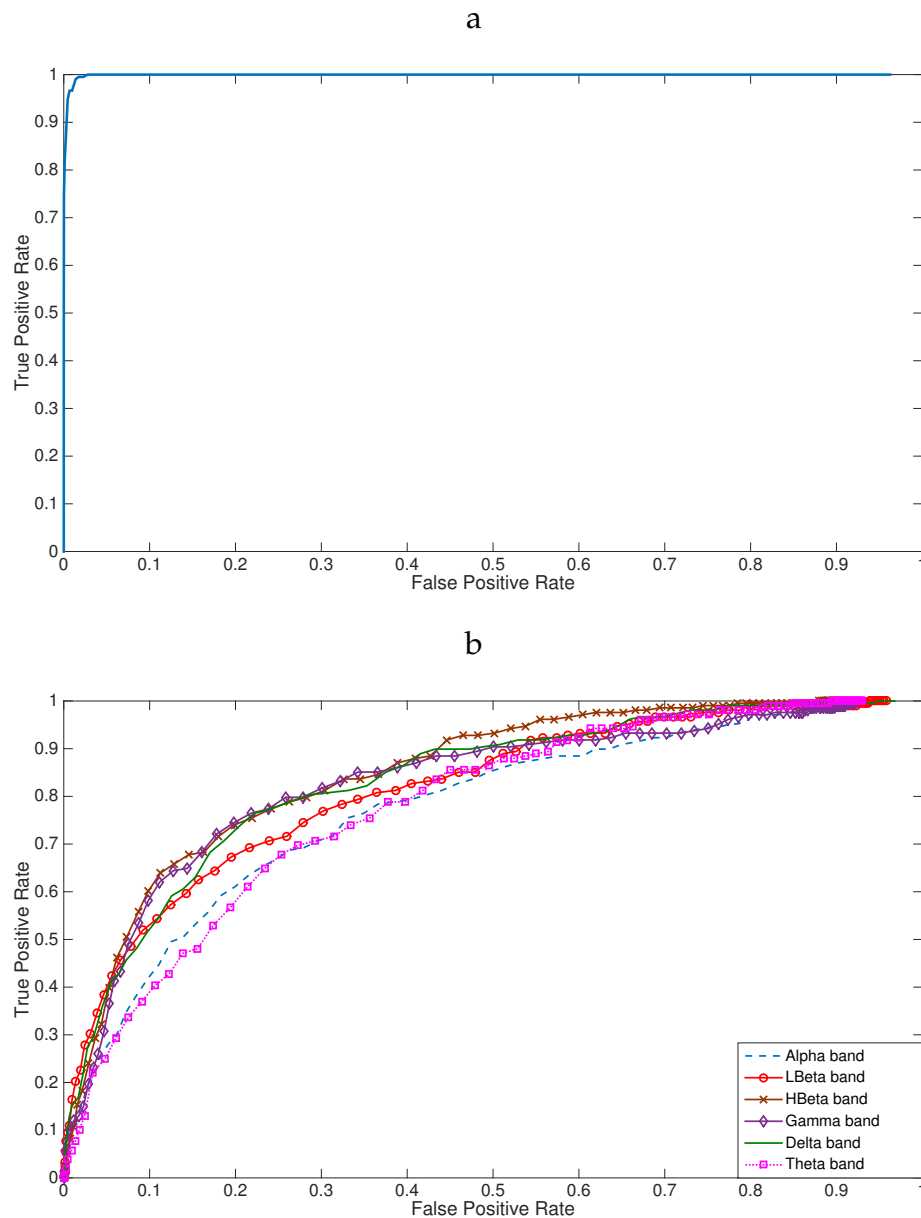
FIGURE 6.4: ROC curve related to the ECG channel(a) and to EEG channels(b).

matrix related to the ECG has been added (and respectively, multiplied and weighted) to the matrices related to each of the EEG channels. The comparison between our approach and the one proposed in the preliminary results presented in (Barra et al., 2015a) are depicted in Table 6.2. Improvements for all the bands are visible. The weighted sum is the most effective method, due to the fact that it takes into consideration the strength of the ECG with respect to the EEG. The AUC values of the fusion methods applied to the vectors are shown in 6.3. The ROC curves for each fusion achieved are shown in Figure 6.5, 6.6, 6.7, 6.8, 6.9 and 6.10.
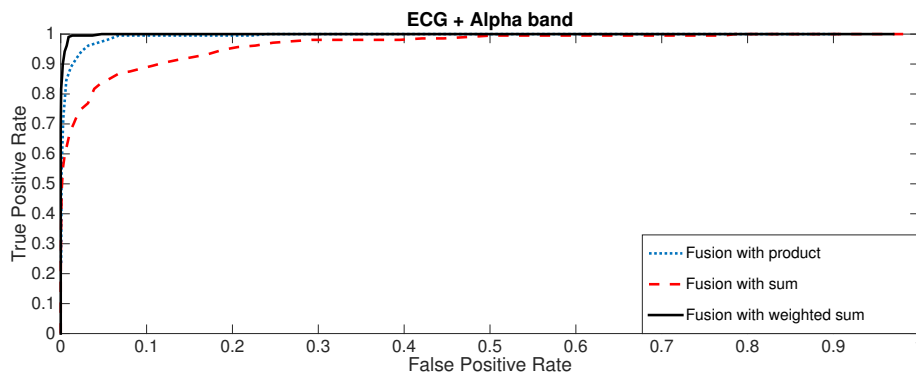
FIGURE 6.5: ROC curves for the fusion between ECG lead i and the alpha band of the EEG.
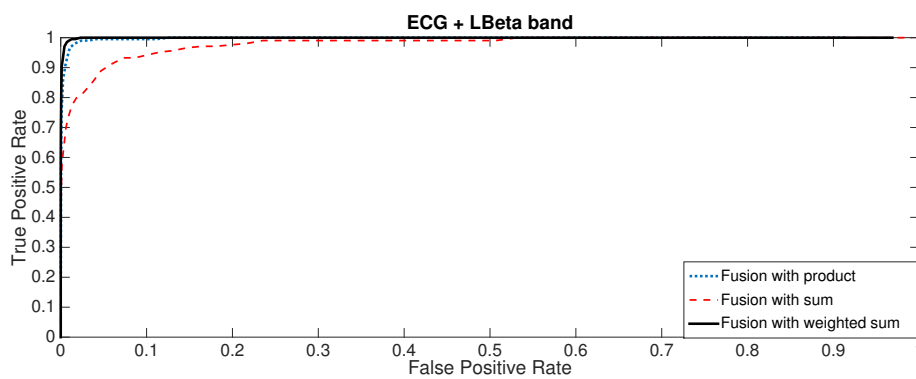
FIGURE 6.6: ROC curves for the fusion between ECG lead i and the low beta band of the EEG.
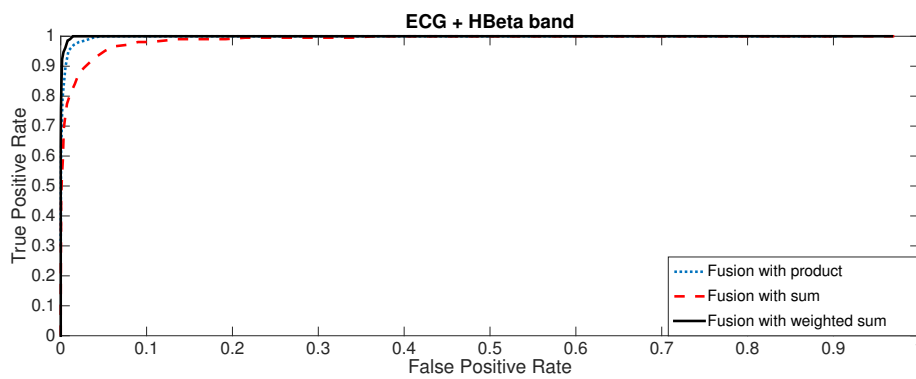
FIGURE 6.7: ROC curves for the fusion between ECG lead i and the high beta band of the EEG.
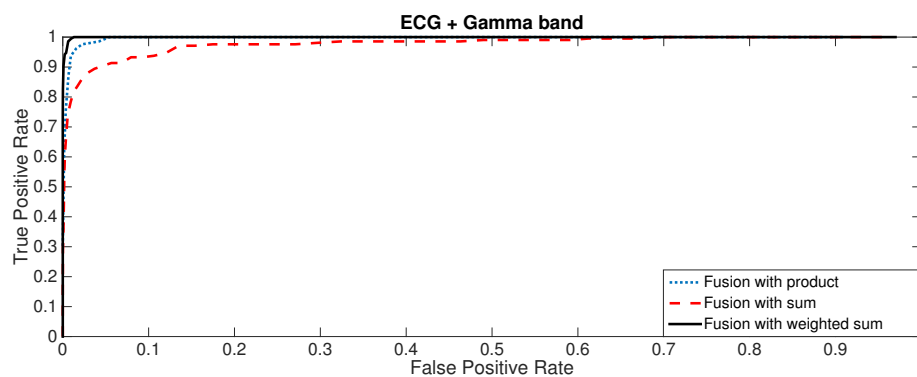
FIGURE 6.8: ROC curves for the fusion between ECG lead i and the gamma band of the EEG.
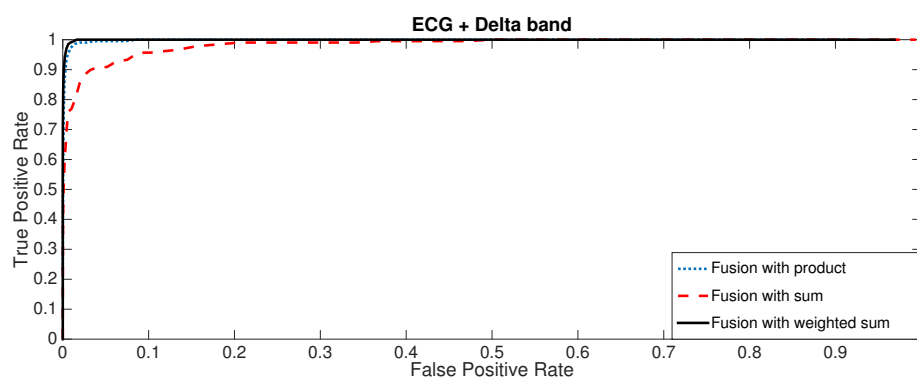
FIGURE 6.9: ROC curves for the fusion between ECG lead i and the delta band of the EEG.
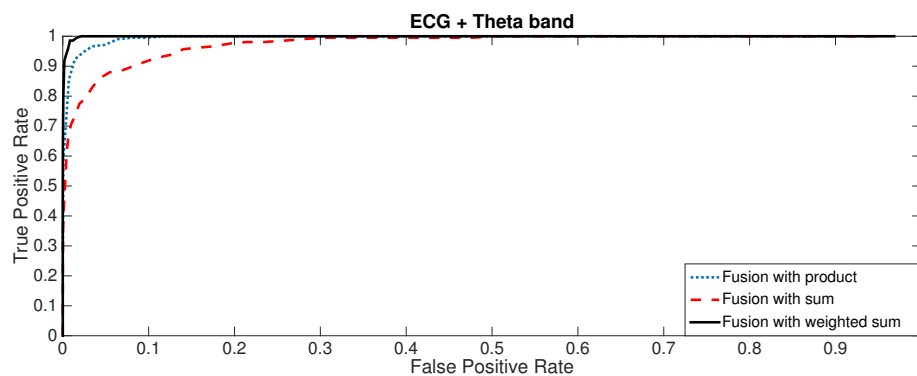
FIGURE 6.10: ROC curves for the fusion between ECG lead i and the theta band of the EEG.

TABLE 6.2: EER values related to the recognition process between the fused vectors.

| | Proposed Approach | | | (Barra et al., 2015a) |
	sum	weighted sum	product	K-means Clustering
ECG+Alpha	10.42%	**0.93%**	3.51%	2.94%
ECG+LBeta	7.04%	**0.97%**	1.77%	2.94%
ECG+HBeta	5.02%	**1.03%**	2.00%	2.94%
ECG+Gamma	7.34%	**0.84%**	2.19%	2.94%
ECG+Delta	6.83%	**0.92%**	1.44%	2.94%
ECG+Theta	8.89%	**1.23%**	3.43%	2.94%

TABLE 6.3: AUC values related to the recognition process between the fused vectors.

	sum	weighted sum	product
ECG+Alpha	95.23	**96.87**	90.03
ECG+LBeta	**97.07**	96.80	91.26
ECG+HBeta	**96.89**	96.74	90.18
ECG+Gamma	94.27	**96.95**	90.71
ECG+Delta	**97.74**	96.82	91.71
ECG+Theta	93.45	**96.75**	89.66

Chapter 7

Biometric Fusion Module

7.1   Introduction

Multi-biometric systems (Ross, Nandakumar, and Jain, 2006) are considered as one of the best viable solutions to overcome limitations of systems dealing with single biometrics. In a well-designed multi-expert architecture, a winning strategy is to balance weaknesses of a sub-system with strengths of a companion one, offering complementary features. The implementation of an effective fusion strategy of the results from different sources is among the most relevant issues raised by the combined approach. Let us recall that information fusion in a multi-biometric system can be performed at different times (Jain, Nandakumar, and Ross, 2005):

- *sensor* level;

- *feature* level;

- *score* level

- *decision* level

Due to difficulties implied in the former two, and in the loss of supplementary information in the last one, most schemes in literature opt for score level fusion (Ross and Jain, 2003). Score normalization is one of the important aspects to consider during this kind of fusion. A comparison of different algorithms is reported in (Jain, Nandakumar, and Ross, 2005), while a technique overcoming most limitations of popular ones, namely the Quasi Linear Sigmoid (QLS) function is presented in (De Marsico et al., 2011). As a possible alternative strategy, fusion schemes may also combine scores into a unified feature vector. This requires to design and train a further classifier, or to transform them in *a posteriori* probabilities (Kittler et al., 1998). An important point to consider is that flaws of a single sub-system may not affect all its responses, but be rather caused by specific temporary conditions (e.g. a change in illumination). A well-devised fusion strategy must take this into consideration. Some systems are designed to return additional complementary information to qualify their recognition scores. This can be represented by an assessment of input quality (Fierrez-Aguilar et al., 2005; Nandakumar et al., 2006) or by an evaluation of the confidence margin corresponding to the response (Kittler et al.,

1998). The former (e.g. sharpness, lighting) can be used either before classification, to possibly discard too problematic samples, or after classification, as a measure to further appraise the final obtained score. The latter are used after classification to possibly discard responses that do not assure a sufficient accuracy. For simplicity, and since they can be treated in the same way, from now on we will refer to all types of additional response information as "reliability", unless differently required by the ongoing discussion. The introduction of an additional reliability measure in the system response arises a new problem to address during fusion, regarding the strategy to effectively exploit it. Two trends are currently developing, based respectively on response weighting or selection. In the first approach, all subsystems participate in all fusion operations, and single responses are weighted by their respective reliability. In the second one, reliability values are exploited to select, from time to time, only a subset of sub-systems to be enabled to take part in the fusion. In both cases, the reliability measure is mostly handled as a separate value. FAIRY's fusion module focuses on the first approach, by combining the recognition score and the corresponding reliability value into a single complex number. The main usefulness of a compact representation of the pairs score/reliability is in the possibility to speed-up and simplify the fusion of results from different systems. Furthermore, it can allow to compare in one shot the responses of different systems. Having two pairs (score, reliability) from two different systems, it is quite difficult to obtain an obvious ordering. On the other hand, the different nature of the considered measures calls for a careful choice of the method to combine them. The aforementioned factors fully fit the requirements of FAIRY, where the multibiometric fusion process is based on different subsystems (ear recognition, face recognition and iris recognition modules) producing their own recognition results. The method evaluates the reliability and the decision score of each subsystem (module) and computes a fusion score.

7.2   Related Works

Input quality is used by (Kryszczuk et al., 2007). The authors design appropriate Bayesian networks to evaluate the probability of a correct verification decision by a classifier. The propagated available evidence comes from several sources: a vector of signal-domain quality measures gathers significant features of each biometrics and of each classification method. A Bayesian approach to quality-based fusion is also adopted in (Poh and Kittler, 2012). While the framework aims at modeling the expert outputs and quality measures for a single one modality, the combination of the information from multiple modalities can be obtained applying the Naive Bayes principle. In both of these approaches, the setup of the Bayesian network (variables, probabilities and propagation) depends on the application context and may require a complex investigation of interrelated factors. Confidence margins as well as reliability measures can be used after classification to decide if the system response is sufficiently trustworthy. Each margin stems from the observation over time of the scores produced

by a single subsystem, and measures the amount of risk associated to its response. The authors in (Poh and Bengio, 2005) introduce a confidence margin that is computed from False Acceptance Rate (FAR) and False Rejection Rate (FRR) of a biometric system, while (De Marsico et al., 2011) introduce a reliability measure based on the composition of the system gallery, and returned on a single- response basis. A number of authors claim that simple sum offers the best compromise between simplicity and performance among the many simplest score fusion rules, e.g. sum, weighted sum, product, min, or max (Kittler et al., 1998). These methods are attractive, since they entail a low computational demand. On the other hand, significantly better results can be provided by more complex techniques (Abate et al., 2007). Likelihood Ratio (LR) is one of the most investigated and appraised candidates. As an example, according to the experiments presented in (Ulery et al., 2006), in the fusion of results of a multibiometric system, the product of LR provides the highest Genuine Accept Rate (GAR) for a fixed FAR, given that a number of conditions are met, among which the Gaussian distribution of scores. This rule assumes an accurate preliminary estimate of the joint distribution (across all the subsystems) of the conditional probabilities of the scores achieved by genuine and impostor users. Though providing optimal results, this strategy presents a number of drawbacks too. First, it requires a complex modeling phase. As an example, in (Nandakumar et al., 2008) genuine and impostor score densities are modeled by a finite Gaussian Mixture Model (GMM). Second, it requires a significant number of training samples, to sufficiently represent possible combinations of single results. Last but not least, performance of systems whose operational parameters are tuned up with a preliminary estimation of score distributions, may degrade if these significantly change along time. Nevertheless, given the optimality of LR in stable conditions, we assume it as an asymptotic limit for which to strive when devising a new fusion rule, and at the same time we aim at avoiding its limitations. The FAIRY's BFM (Biometric Fusion Module) approach is different from what is presently found in literature. As an example, the technique used in (Yang et al., 2003) works at feature level, and fills the real and imaginary part of a complex vector by mapping onto them the (real) feature vectors from two different biometric systems. The BFM rather deals with score level fusion, by fusing the results returned by different subsystems voting for the same identity. To this aim, the score and the corresponding reliability are considered, together with the identification result produced by the single subsystems. These values are used to derive the module and the anomaly in the exponential representation of a complex number. Fusion relies on complex product.
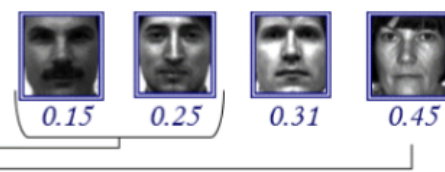
7.3 The Biometric Fusion Module approach

The reliability measure exploited to assess our combination strategy is SRR (System Response Reliability) introduced in (De Marsico et al., 2011). SRR

is independent from both the exploited biometric trait(s), and from the estimated distributions of system scores, but rather takes into account the composition of the gallery underlying the recognition system. It does not need any knowledge about acquisition quality, extracted features, and classification methods. For this reason, it can be used with all (off-the-shelf) identification modules, which return an ordered list of distances (similarities). Given a gallery $G = g_1, g_2, \cdots, g_{|G|}$ of size $|G|$, the system returns the whole ordered list of scores $L = d(p, g_{i_1}), d(p, g_{i_2}), \cdots, d(p, g_{i_{|G|}})$ obtained by matching each gallery template with the probe, where the subscript indicates the permutation of gallery indexes induced by the similarity order. For simplicity, a gallery with a single template per subject is considered; however, the approach is adaptable to galleries with more templates per subject. SRR can be computed according to three different versions of function ϕ defined in (De Marsico et al., 2011). All three functions provide a measure of the amount of "confusion" among possible candidates, in turn affected by different factors. Given a probe p and a system A with gallery G, the first function ϕ_1 is:

$$\phi_1 p = \frac{d(p, g_{i_2}) - d(p, g_{i_1})}{d(p, g_{i_{|G|}})}, \tag{7.1}$$

where d is a distance function with co-domain $[0, 1]$. In galleries with more templates per subject, g_{i_2} is the closest template belonging to a subject different from the first retrieved one. We can also exploit distance metrics returning values falling in a different co-domain, since these can be suitably normalized. Here we use the *QLS* introduced in (De Marsico et al., 2011). It better preserves the original distribution of data, and is robust to a missing reliable estimate for the maximum value. Relative distance measures the "confusion" generated by the presence of a subject close to the first returned one, where closeness is estimated relatively to the maximum measured distance. With relative distance, the higher the relative difference between the first two distances from the probe (the distance of the first retrieved subject and the distance of the second retrieved one), the more reliable is the response. In other words, it is inversely proportional to the "confusion" deriving from the presence of two very similar subjects in the gallery. Nevertheless, reliability is not a synonym for correctness. Correct responses may have low reliability just because the first two retrieved subjects might be confused. An example of computation of ϕ_1 is in Figure 7.1 The second function ϕ_2 is based on *density ratio* and is defined as:

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

Relative Distance = (0.25 − 0.15)/0.45 = 0.22

FIGURE 7.1: An example of computation of relative distance ϕ_1.

$$\phi_2(p) = 1 - \frac{|N_b|}{|G|}, \tag{7.2}$$

where

$$N_b = \{gi_k \in G | d(p, g_{i_k}) < 2d(p, g_{i_1})\}. \tag{7.3}$$

The formula for ϕ_2 considers the identities returned during identification as a cloud around p; the higher is the density of subjects in this cloud, the more unreliable is the answer. In other words, there are many individuals as potential candidates. Even in this case, the definition is slightly changed for galleries with more templates per subject, since in N_b we do not count templates belonging to the same subject of the first retrieved one. Notice that this is a different situation than the one considered for ϕ_1. In this case, the possible problem is not represented by a single individual very close to the first, but by a number of individuals concentrated around p, even if not that close. Figure 7.2 shows an example of computation of ϕ_2. We also adopt a variation of the density ratio.
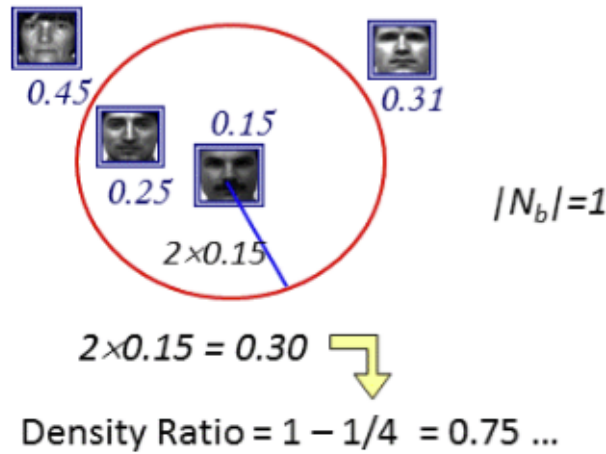
FIGURE 7.2: An example of computation of density ratio ϕ_2.

As one can observe in the definition of N_b in 7.3, the radius of the considered cloud depends on the distance between the probe and the first returned identity and from a constant. This function is less sensible to outliers than ϕ_1. However, it considers narrower clouds when the first retrieved identity is closer to the probe, while a distance which is large in itself takes to an even wider cloud. The latter can be expected to be more crowded anyway. To avoid this anomaly we define ϕ_3, and include the term N_c such that the cloud radius depends on the difference between the first two distances:

$$\phi_3(p) = 1 - \frac{|N_c|}{|G|}, \tag{7.4}$$

where

$$N_c = \left\{ g_{i_k} \in G | d(p, g_{i_k}) < \frac{(1 + d(p, g_{i_2}))(1 + d(p, g_{i_2}) - d(p, g_{i_1}))}{4} \right\}. \tag{7.5}$$

The new radius increases with the second distance, and with the difference between the first and the second one. In practice, the farthest the second returned subject from the probe, also with respect to the first one, the wider the cloud we inspect. However, being all distances in $[0,1]$, we add 1 to both terms to maintain direct proportionality. We also add the appropriate normalization factor: since the value of d is in $[0,1]$, and the maximum value for the numerator in 7.5 is 4, this is the value that we use for the denominator. In Fig. 3, we show an example for ϕ_3. Once chosen the function ϕ to use, some more steps are required
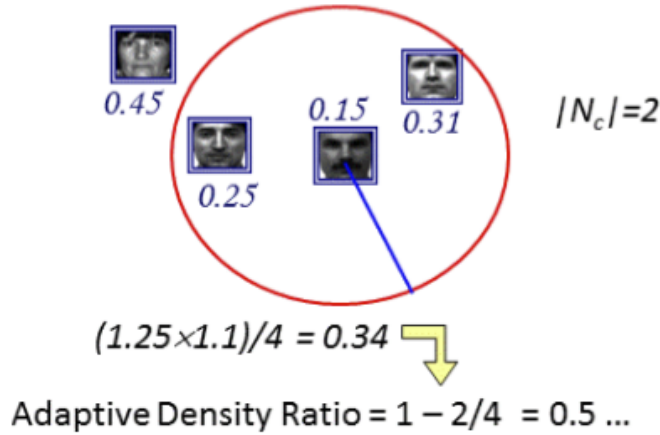
FIGURE 7.3: An example of computation of density ratio ϕ_3.

to pass from ϕ_p to the value of SRR for the probe at hand, namely $srr(p)$. First, for each ϕ, we identify a value $\bar{\phi}$ fostering a correct separation between correct and incorrect responses. We also define $S(\phi(p), \phi)$ as the width of the subinterval from $\bar{\phi}$ to the proper extreme of the overall $[0,1)$ interval of possible values, which contains the current $\phi(p)$:

$$S(\phi(p), \bar{\phi}) = \begin{cases} 1 - \bar{\phi} & \text{if } \phi(p) > \bar{\phi} \\ \bar{\phi} & \text{otherwise.} \end{cases} \tag{7.6}$$

This allows to compare reliability of different responses and of responses from different systems. As a matter of fact, the value of $\phi(p) - \bar{\phi}$ might not be sufficient to fully qualify the reliability of a single response with respect to other ones, which is especially the case with multi-biometric systems. We clarify the point by an example. Let us assume $\bar{\phi} = 0.1$; a reject with $\phi(p) = 0.02$ (absolute distance from $\bar{\phi}$ is 0.08 and $srr(p) = 0.8$) is to be considered more reliable than a reject with $\phi(p) = 0.08$ (absolute distance from $\bar{\phi}$ is 0.02 and $srr(p) = 0.2$), and this emerges even from the absolute distances alone. However, the former response must also be considered more reliable than an accept with $\phi(p) = 0.19$ (absolute distance from $\bar{\phi}$ is 0.09 and $srr(p) = 0.1$), which in fact lies at a proportionally lower distance from the critical point toward the other end of the overall interval. Figure 7.4 depicts an example of distribution of srr values from a number of probes, and of the meaning of $\bar{\phi}$ and $S(\phi(p), \bar{\phi})$ SRR index can
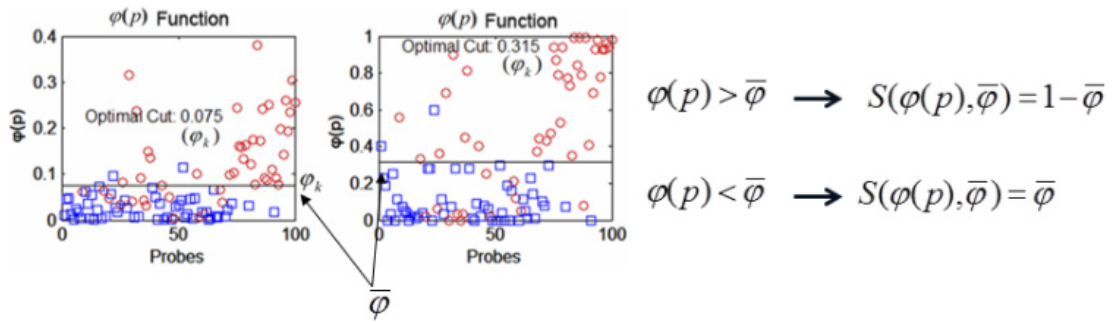
FIGURE 7.4: An example of distribution of srr values for a number of probes, of $\bar{\phi}$ and $S(\phi(p), \bar{\phi})$.

finally be defined as:

$$srr(p) = \frac{\phi(p) - \phi}{S(\phi(p), \phi)}. \tag{7.7}$$

Notice that even a "not recognized" answer, which can be returned in open set settings (not all probes surely belong to the gallery), can be marked either as reliable or unreliable. In detail, we measure the distance between $\phi(p)$ and the "critical" point $\bar{\phi}$, which gets higher values for $\phi(p)$ much higher than $\bar{\phi}$ (unambiguously recognized), or for $\phi(p)$ much lower than $\bar{\phi}$ (unambiguously not recognized). However, it is also important to take into account how much such distance is significant with respect to the subinterval over which it is measured. SRR gets values in $[-1, 1]$. More details on computation and its motivations can be found in (De Marsico et al., 2011). Also extensive experimental results about the increase of recognition accuracy due to the use of the different alternative version of $\bar{\phi}$, and of derived SRR, can be found in the referenced paper; since these measures are not the focus of the present work, but are rather only used as examples, the presentation of such kind of results would not add any value to the present proposal. On the other hand, it is interesting to show how their combination with similarity scores can simplify the handling of results. In a multimodal architecture, each subsystem returns both a similarity score s and a quality/confidence/reliability value, in our case srr, which pertain to the retrieved identity. We refer to such value in general as *reliability* and indicate it as w. Eventually groups of subsystems return the same identity (see Figure 7.5), therefore in order to identify the winning one, we have to devise a strategy to consistently fuse the pairs (score, reliability) pertaining to the same identity. This is the issue addressed in the next subsection.

7.3.1 The Combination of Score and Reliability in a Single Complex Value

As underlined above, the use-case of a compact representation of the pairs score/ reliability is the comparison or the fusion of results from different systems. The nature of the combined measures lends itself to the proposed kind
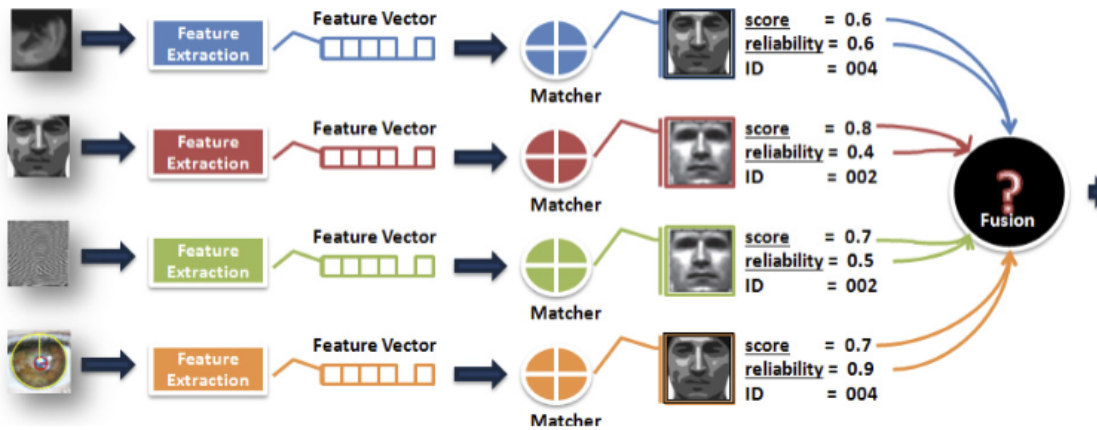
FIGURE 7.5: An example of responses from different biometric systems.

of approach. In fact, having two pairs (score, reliability) from two different systems, it is quite difficult to obtain an obvious ordering. As a matter of fact, the two measures express a completely different kind of information: the score indicates how much the probe template and the returned gallery template are similar, while either quality (indirectly) or reliability (directly) indicate how much we can trust the response. As already noticed, a high similarity can also be combined with a low reliability, and vice versa. This is exactly what happens with complex numbers too. Numbers in the complex field can be represented as $a + ib$, with $a, b \in R$, where a is the real part, ib is the imaginary one, and i is the imaginary unit ($i^2 = -1$). While this is the most used form to represent a complex number, in some cases it can be advantageous to use an alternative representation. In the exponential representation $z = \rho \cdot e^{i\Theta}$, ρ is the modulus, i.e. the absolute value of the complex number $\rho = \sqrt{a^2 + b^2}$ and Θ is the anomaly given by $\Theta = tan^{-1}\frac{b}{a}$. In the proposed fusion protocol, the score s and the reliability measure w from a single subsystem are first combined, by using them to derive the real and imaginary part of a complex number. In this way, a unified representation is obtained. The use of the most intuitive representation $a + ib$ (score as a and reliability for b) has been given up due to the fact that the product of two complex values in this form, namely $(a + ib)(c + id) = (ac - bd) + i(bc + ad)$, would combine scores with reliability values, therefore causing a misleading cross-influence between heterogeneous parts which represent information of very different nature. The exponential representation allows overcoming the limitation discussed above, though after using a simple yet necessary trick. Score s and reliability w are both modified to derive the ρ and the Θ of the exponential representation, respectively. However, in most cases, scores produced by a recognition system vary in the range $[0, 1]$, while reliability ranges in $[-1, 1]$. We want to avoid a product among scores lower than 1, since the product would be lower than the starting factors. Therefore we set $\rho = (1 + s)$, so that $\rho \in [1, 2]$. Moreover, in order to handle always positive reliability values, we set $\Theta = (w + 1)/2$, so that we also have $\Theta \in [0, 1]$. Given this, the combined response from classifier j is $z_j = p_j \cdot e^{i\Theta_j}$. The value

resulting from the fusion of k responses z_j from classifier C_j voting for a same identity is defined as:

$$z = \prod_{j=1}^{k}(z_j)^{\frac{1}{k}} = \sqrt[k]{\prod_{j=1}^{k}(\rho_j)} \cdot e^{i\frac{1}{k}\sum_{j=1}^{k}\Theta_j}. \tag{7.8}$$

As it is possible to observe from 7.8, in the final value resulting from fusion, the (modified) score part is given by the geometric mean of single starting (modified) scores, while the (modified) reliability part is given by the arithmetic mean of the single starting (modified) reliability values. The expression in 7.8 can be easily exploited to fuse the responses of k classifiers voting for the same identity (see Figure 7.6). Of course, responses of classifiers voting for an identity with no further supporters are converted in the exponential complex representation yet without being fused. The next problem regards the possibility to order the
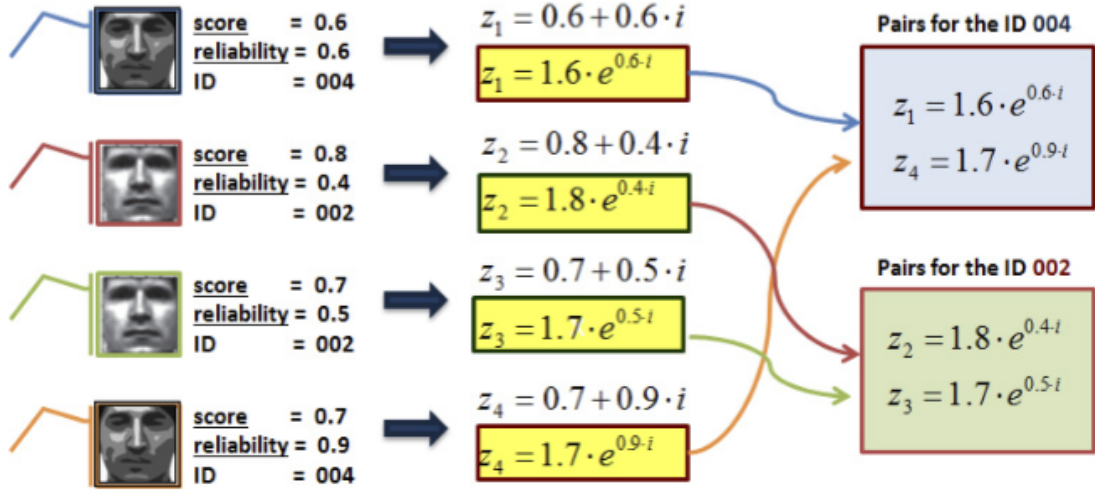
FIGURE 7.6: An example of responses to fuse for different identities.

(reduced) list of responses after fusion (see Figure 7.7), since in the field of complex numbers it is not possible to devise an ordering consistent with the ordering of real numbers, which are a subset of complex ones. In order to overcome this further problem too, the logarithm of the values obtained from fusion is considered:

$$\ln(z) = \ln\left(\prod_{j=1}^{k}(z_j)^{\frac{1}{k}}\right) = \frac{1}{k}\left[\sum_{j=1}^{k}\ln\left(\rho_j e^{i\Theta_j}\right)\right] = \frac{1}{k}\sum_{j=1}^{k}\ln\left(\rho_j\right) + i\frac{1}{k}\sum_{j=1}^{k}(\Theta_j) \tag{7.9}$$

The real and imaginary parts in 7.9 can be linearly combined according to the formula:

$$\lambda\left(\frac{1}{k}\sum_{j=1}^{k}\ln(\rho_j)\right) + (1-\lambda)\left(\frac{1}{k}\sum_{j=1}^{k}(\Theta_j)\right), \tag{7.10}$$
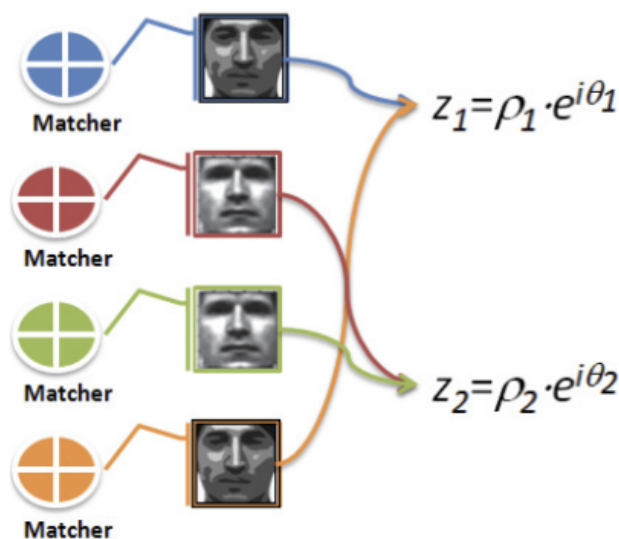
FIGURE 7.7: An example of fused responses for different identities.

where the value $\lambda \leq 1$ can be appropriately chosen to adjust the relative importance of score versus reliability. The obtained values can finally be ordered to determine a winning identity.

7.4 Experimental Results

The proposed protocol for combining score and reliability for a single response, and for fusing different responses, has been tested in a multi-biometric setting including face, ear and iris traits. The results have been compared on one side with very simple yet effective techniques, namely Simple Sum and Simple Product, which, tough being suboptimal, have the advantage of being very fast and immediate to implement. On the opposite side, they have been compared with LR, according to the optimality considerations discussed in the introduction. As a matter of fact, given also the achieved theoretical results, the authors of (Ulery et al., 2006) demonstrate that, in the given conditions, it is not possible to obtain better performance, therefore it is an optimal methodology. We adopted the formulation discussed in (Nandakumar et al., 2008), using the same implementation for the estimation of the GMM model. It is worth underlining that, though representing an upper bound to the obtainable performance, LR has the limit to require an accurate estimation of score distributions, as it happens for the measures based on FAR and FRR, while the approach that we use allows to evaluate the reliability of each single response without any preliminary and specific training or learning phase. The multi-biometric database was built creating Chimeric users whose biometric traits were taken from three different datasets. We underline that it is presently accepted that results obtained in this way are worthy of full reliability (Garcia-Salicetti et al., 2005). Of course each image in each dataset was exclusively assigned to a given

subject. Therefore, the number of subjects in the database is constrained by the size of the smallest one, namely 100 subjects in the Notre Dame Ear Database presented in (Notre Dame, 2005; Yan and Bowyer, 2005). We will refer to the latter simply as Ear. We consider an open set identification setting, i.e. a situation where not all users are enrolled, and as a consequence also impostors can occur. For this reason, the gallery consists of 75 enrolled subjects, with a single image each, while the probe includes 100 subjects, each with a single image. The faces are from a subset (50 males and 50 females) of AR-Faces database (Martínez, 2002), consisting of four different datasets: gallery (normal), Face-2 (smile), Face-5 (left-light) and Face-11 (scarf). The irises belong to the first 100 subjects in UBIRISv1s1 database introduced in (Proença and Alexandre, 2005). We will refer to the latter simply as Iris. Performance were measured in terms of RR and EER (Bolle et al., 2005). A first aim of our experiments was to understand the relation between the behavior of the presented protocol and the classifier used. With respect to the work in (De Marsico, Nappi, and Riccio, 2013) we substituted the module implementing Principal Component Analysis (PCA), with one implementing Local Binary Pattern (LBP). We continued using Linear Discriminant Analysis (LDA) and the classifier based on local correlation which is part of FACE system (De Marsico, Nappi, and Riccio, 2010a), indicated from now on as FACE for short. It is worth reminding that in these new experiments we use a different operation with complex numbers to perform fusion of different responses. Table 7.1 shows the performance on each dataset. As expected, results are quite heterogeneous, but this will help later to better underline how the fusion technique works. The most interesting case is not when all classifiers provide optimal results, but when one or more of them fail. Table 7.1 shows that the different approaches have very heterogeneous behavior, and therefore make up a significant basis to test our fusion protocol. When considered singularly, we can observe that LDA achieves good performance on datasets particularly free from distortions (difficulties), such as Iris and Face-2, while it achieves lower performance with Face-5 and Face-11. In particular, performance is very bad with Face-11 set, where the lower part of the face is completely occluded by a scarf. In combination with other biometrics, this condition may be particularly stressing for the fusion process. On the contrary, LBP shows a more balanced behavior. It never reaches very high performance in terms of RR and ERR; however, it is able to address even more difficult datasets like Face-11. FACE is the best among the three, since it achieves very good performances on almost all datasets. This mixture is very interesting, since it allows to understand, in a fusion setting exploiting our protocol, how much this is influenced by the presence of very poor classifiers. As a matter of fact, in very simple fusion schemes, it often happens that the worst classifiers significantly decrease the performance of the overall system. The following experiments will use Iris, Ear and Face-2 datasets, therefore when we refer to Face we mean this latter one. The second experiment enters the core of the assessment, and aims at evaluating the performance of different fusion schemes, when different classifiers process the same biometric trait. This kind of experiment was not reported in (De Marsico, Nappi, and Riccio, 2013). For

TABLE 7.1: Performances of the single classifiers on single datasets, in terms of RR and EER.

	LDA		LBP		FACE	
Dataset	RR	EER	RR	EER	RR	EER
Iris	0.97	0.020	0.93	0.071	1.00	0.072
Ear	0.91	0.065	0.93	0.063	0.94	0.058
Face-2	1.00	0.003	0.94	0.058	1.00	0.016
Face-5	0.87	0.075	0.94	0.054	1.00	0.003
Face-11	0.07	0.345	0.75	0.135	1.00	0.019

each single biometrics (iris, ear and face), the three classifiers return a fused response. Fusion was performed with different techniques: Simple Sum, Simple Product, the proposed Complex Fusion using either ϕ_1, or ϕ_2, or ϕ_3 for reliability, and with logarithmic (log) to support ordering, and finally LR. The different techniques are compared in terms of RR and ERR achieved. Since in Complex Fusion with logarithmic combination it is necessary to specify the value for λ, Table 7.3 reports value of λ providing optimal performance. Some interesting considerations can be derived from Tables 7.2 and 7.3. First of all, if we exclude the fusion scheme based on LR, which demonstrated to be an optimal strategy (we can consider it as an asymptotic performance to strive), Complex Fusion (CF) with ϕ_1 achieves the best performance of all and in all cases. The better behavior of CF with ϕ_1 with respect to CF with ϕ_2 can be ascribed to a better performance of ϕ_1. As a matter of fact, by observing values in Table 7.3, we can see that CF with ϕ_1 with log requires lower values for λ with respect to CF with ϕ_2 with log, giving a higher weight to the reliability of single responses. We can further observe that ϕ_2 and ϕ_3 present a similar behavior. Tables 7.4 and 7.5

TABLE 7.2: Performance in terms of RR and EER of the different fusion schemes in multi-classifier setting (all classifiers process the same trait).

	Iris		Ear		Face	
Method	RR	EER	RR	EER	RR	EER
Simp.Sum	0.79	0.188	0.77	0.158	0.81	0.171
Simp.Prod	0.56	0.376	0.65	0.371	0.68	0.285
CFΦ_1+Log	**0.97**	**0.108**	**0.81**	**0.136**	**1.00**	**0.045**
CFΦ_2+Log	0.93	0.228	0.80	0.200	1.00	0.131
CFΦ_3+Log	0.93	0.228	0.80	0.200	1.00	0.131
Likelihood ratio	1.00	0.028	1.00	0.017	1.00	0.011

confirm the same trend observed in multi-classifier context. Again CF with ϕ_1 achieves better performance than the others. Moreover, the weight assigned to reliability through the λ parameter is actually very low when using ϕ_2 and ϕ_3.

TABLE 7.3: Values for λ adopted for Complex Fusion with logarithmic combination in multi-classifier setting.

Method	LBP	LDA	FACE
CFΦ_1+Log	0.55	0.80	0.25
CFΦ_2+Log	0.90	0.90	0.95
CFΦ_3+Log	0.90	0.90	0.95

Though by different experiments, the results achieved in (De Marsico, Nappi, and Riccio, 2013) is confirmed. Using a robust classifier aligned with the state-of-the-art, e.g. FACE, the proposed fusion technique can provide better results than simple sum, and only slightly lower than the optimum LR. However, it is to consider that the latter requires an accurate estimate of (joint) score distributions, so that is both computationally more expensive and less stable in time. The proposed Complex Fusion is simple like the sum, since it does not require such preliminary estimation, but is able to achieve very satisfying results, comparable to LR.

TABLE 7.4: Performance in terms of RR and EER of the different fusion schemes in multi-biometrics setting (each classifier process all traits).

Method	LBP		LDA		FACE	
	RR	EER	RR	EER	RR	EER
Simp.Sum	0.86	0.153	0.98	0.088	0.97	0.045
Simp.Prod	0.62	0.165	0.77	0.080	0.97	0.045
CFΦ_1+Log	**0.92**	**0.120**	**0.99**	**0.080**	**1.00**	**0.028**
CFΦ_2+Log	0.90	0.188	0.97	0.108	1.00	0.080
CFΦ_3+Log	0.90	0.188	0.98	0.108	1.00	0.080
Likelihood ratio	1.00	0.028	1.00	0.017	1.00	0.011

TABLE 7.5: Values for λ adopted for CF with logarithmic combination in multi-biometrics setting.

Method	LBP	LDA	FACE
CFΦ_1+Log	0.65	0.55	0.80
CFΦ_2+Log	1.00	0.95	0.90
CFΦ_3+Log	1.00	1.00	0.85

Chapter 8

Video Surveillance

8.1   Introduction

A natural completion of investigation of pattern recognition issues related to security and privacy is the one regarding the wider problem of videosurveillance. In this chapter, three issues have been faced. The first regards the development of a tracking method based on the calculation of the mutual information (Barra et al., 2014d). This method introduces the concept of *anchor*, defined as the point or the region of interest we want to track, to the aim of achieving behaviour and action recognition. An anchor is not forced to be a person, but, may be also a part of a person(head, hands, ...). Two more works in the same field were born during my one-year internship in Portugal, as a visiting researcher in the SOCIALab of the UBI (Universitade da Beira Interior, Covilhà) directed by professor Hugo Proença. The first work regards a master-slave camera calibration algorithm aimed at the face detection in surveillance systems for biometric recognition purposes (Neves et al., 2015a). The latter, instead, regards a fully automated surveillance system with *In The Wild* biometric recognition (Neves et al., 2015b). In the following section the aforementioned works will be further detailed.

8.2   Using Mutual Information for Multi Anchor Tracking

The tracking of human beings has always represented a challenging problem that is increasingly attracting the researchers attention, as it is essential for many applications in video analytics. Significant progresses have been achieved in many recent works in literature along this research line (Bregler and Malik, 1998; Hogg, 1983; Sidenbladh, Black, and Fleet, 2000). Most approaches from the present state of the art focus their attention on the estimation of the human body configuration, since they are aimed at recognizing human actions and activities. The first step of the tracking process consists in segmenting the human shape from the background. It represents a very complicated problem, as real-world applications work with complex and possibly moving backgrounds, large changes in illumination conditions and self-occlusions. In order to address these issues, several approaches have been proposed, which are based on

background subtraction (Sato and Aggarwal, 2001), optical flow (Okada, Shirai, and Miura, 2000) or statistical modeling of the human appearance (Wren et al., 1997; Haritaoglu, Harwood, and David, 2000). All those methods try to segment video frames to extract edges, silhouettes or blobs. Blobs are often preferred with respect to other features, as they are more suited to the task of detecting human body parts. As a matter of fact, in connection with a suitable model of body structure, tracking single body parts allows to compensate for partial occlusions due to mutual people overlapping during interaction, or caused by objects in the scene that may partially hide a person for a few frames. Blobs are also more suitable than other descriptors, like 2D contour, for modeling the articulated motion of human body parts. Models for approximating the articulated motion generally rely on a stick figure, which was first defined by Johansson (Johansson, 1975) as the union of segments linked by joints. Thus, detection methods can either locate blobs composing the human shape or simply the joints and end points of a stick figure (elbows, knees, head, hands, feet). Yilmaz et al. categorized all detection methods in four main classes: i) background subtraction, ii) segmentation, iii) supervised learning, iv) point detector. Approaches based on background subtraction suffer from illumination changes and partial occlusions, since precisely separating the human silhouette from a cluttered background with several moving objects may turn in a very difficult task. Similarly, segmentation methods also encounter problems with complex scenes, even if they further rely on additional features like color or gradient direction. Supervised learning overcomes all these limitations, but requires a training phase, which binds the tracker to a specific application context. A point based method detects interest points that are considered representative in terms of a specific feature, like contour, intensity or color. Interest points have been largely used in computer vision to solve a wide range of problems like image registration, image retrieval or tracking systems. Due to the higher robustness of this kind of approach, we adopt it here. We track interest points defined as anchors, without requiring a prior separation of foreground from background. As a matter of fact, this latter task is often very complex and its precision strongly influences tracking results. In our case, anchor tracking is performed by locally processing information right in the color frame, without any kind of foreground detection. To this aim, we exploit concepts borrowed from Information Theory. The core idea of the algorithm, named MIMA (Mutual Information Multi- Anchor), is to use mutual information (Sato and Aggarwal, 2001) for multi-anchor tracking of human figures. Mutual information finds several uses in the context of computer vision, such as the detection of cut/fade in video sequences (Bregler and Malik, 1998). Dame and Marchand used it with good results, for the tracking of feature points in the context of augmented reality (Dame and Marchand, 2010). Probabilistic measures are used in (Loutas, Pitas, and Nikou, 2004) to track multiple faces in scenes with a simplified setting with respect to the one addressed here, and tracking cues are provided by mutual information. The above works suggest that the mutual information can be a valuable tool, even for multi-anchor tracking of people. As a matter of fact, MIMA uses the mutual information to track relevant interest

points tied to the human body (head, hands , ...). Compared to classical methods , it has the additional advantage of working on color images. This is crucial to assure greater precision also on regions, such as the hands, which in a video with a limited resolution have a structure not easily distinguishable and, therefore, difficult to track in grayscale images. Furthermore, the lack of a geometric reference model, difficult to formalize as in the case of the hands, makes it difficult to apply techniques such as the one described in (Dame and Marchand, 2010), which is strongly model-based.

8.2.1 Mutual Information

Shannon theory is a mathematical abstract one that has very important applications in many fields, such as physics (thermodynamics), economics (Stock Market), computer science (data compression and transmission). Entropy is a key measure of information, which quantifies the uncertainty involved in predicting the value of a random variable. Starting from it, the joint entropy, the conditional entropy and the mutual information can be derived. Given a random variable X over a set of values χ with distribution function $p_X(x)$, entropy is defined as:

$$H(X) = -\sum_{x \in \chi} p_X(x) \cdot \log p_X(x) \tag{8.1}$$

Given two random variables X and Y defined over two alphabets χ and ν with respective marginal distribution function $p_X(x)$ and $p_Y(y)$ and with joint distribution function $p_{XY}(x,y)$, mutual information is defined as:

$$MI(X;Y) = \sum_{x \in \chi} \sum_{y \in \nu} p_{XY}(x,y) \cdot \frac{p_{XY}(x,y)}{p_X(x) \cdot p_Y(y)} \tag{8.2}$$

The above formulation for mutual information (MI) is the generic one applied to the communication theory. In the following we will rely on a more specific definition better bound to the problem at hand. In the context of image processing, the random variables X and Y refer to the intensity values of the pixels in two images, denoted as I and I^*. these intensity values are respectively denoted as i and j. For instance, if I and I^* are two gray-level images , i and j can take values in the interval $\Omega_I = \Omega_{I^*} = [0, 255] \subset \mathbb{N}$. The probability $p_I(i)$ is related to the frequency with which the intensity value i appears in image I. In general, such probability is estimated through histograms:

$$p_I = \frac{1}{W \cdot H} \sum_{x,y} \delta(i - I(x,y)) \tag{8.3}$$

where (x,y) is the position within the image, W and H represents the image width and height, respectively, $W \cdot H$ is the total number of pixels in the image

and $\delta(k)$ is the Kronecker function defined as:

$$\delta(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & otherwise \end{cases} \tag{8.4}$$

In histogram computation, every time $I(x) = i$ the value of the $i - th$ bin of the histogram is incremented by 1. Likewise, the joint probability $p_{II*}(i, j)$ of the pair (i, j) indicates the frequency of joint occurrence of (i, j) in the pair of images (I, I^*), in the sense that when i appear in I, j appears in I^* in the same position. Its value is obtained by computing the joint histogram of the two images normalized to the same number $W \cdot H$ of pixels:

$$p_{II*} = \frac{1}{W \cdot H} \sum_{x,y} \delta(i - I(x, y)) \cdot \delta(j - I^*(x, y)) \tag{8.5}$$

As a consequence, the MI for a pair of images can be expressed as:

$$MI(I, I^*) = \sum_{i \in \Omega_I} \sum_{j \in \Omega_{I*}} p_{II*} \cdot (i, j) \cdot \log \left(\frac{\log p_{II*} \cdot (i, j)}{p_I(i) \cdot p_{I*}(j)} \right) \tag{8.6}$$

8.2.2 MIMA System

MIMA is a multi-anchor tracking system which exploits the mutual information to follow the movement of the anchors in a video sequence. The system pre-processes the single frames so that the individual channels of the color image undergo a process of quantization, and the resulting bits are interleaved. In this way, though converting an image, originally represented in an RGB color space, into one with 8 bits depth, part of the information concerning the color is still preserved. The mutual information is the base of the tracking process, which has also been made more robust through further expedients such as the use of a weight matrix to discard outliers from the selection of the current position for an anchor, and the integration of a skin detection algorithm to increase the accuracy of hands tracking.

8.2.2.1 Image Processing

The first step in the MIMA operations pipeline is the transformation of a 24-bit RGB image in a new 8-bit representation. The conversion is meant to preserve part of the information given by the color, which would be irretrievably lost with a trivial conversion of the image in grayscale. MIMA divides the image into the three fundamental channels R, G and B and for each of them considers only the most significant bits: namely three for red, three for green and two for blue. The bits selected for the individual channels are interleaved in order to form a string of 8 bits of the form <R_1 G_1 B_1 R_2 G_2 B_2 R_3 G_3>. The resulting image has the same aspect of a grayscale image, but carries information related to both luminance and chrominance. An example is provided in Figure 8.1.

FIGURE 8.1: An example of a frame before and after reduction to 8 bits.

8.2.2.2 Anchor Selection and Tracking

A correct initial selection of each anchor is an important condition for the general performance of tracking. Actually, this is a key problem in any tracking algorithm, because a wrong selection leads to a faster loss of the anchor. In a real application, the anchor would be selected automatically and in the shortest possible time, so as not to delay the start of the tracking process. Since the focus of our present work is on tracking rather than detection, at the current state MIMA provides manual selection for the initialization of the anchors, but it is possible to adopt any algorithms in the literature for automatic detection of head and hands, and use its results for automatic initialization of the anchors. MIMA represents each anchor A_k with a data structure, which contains the first and last frame in which the anchor has been detected, and the list of the coordinates of the upper left corner of its bounding box B_k, in the consecutive frames in this interval. MIMA works on pairs of consecutive frames (F_{i-1}, F_i). Given the position of each anchor A_k in F_{i-1}, it looks for its position in the following frame F_i. The search process assumes that, although an anchor changes its position in two consecutive frames, the movement is limited within a neighborhood of the original position. Therefore, MIMA calculates the MI between the region of F_{i-1} contained within the bounding box B_k, starting from the upper left point (x, y), and with width w and height h, and all its possible homologous ones contained in the rectangular region of F_i delimited by the vertices $(x - w, y - h)$ and $(x + 2w, y + 2h)$, as shown in Figure 8.2. The values for MI
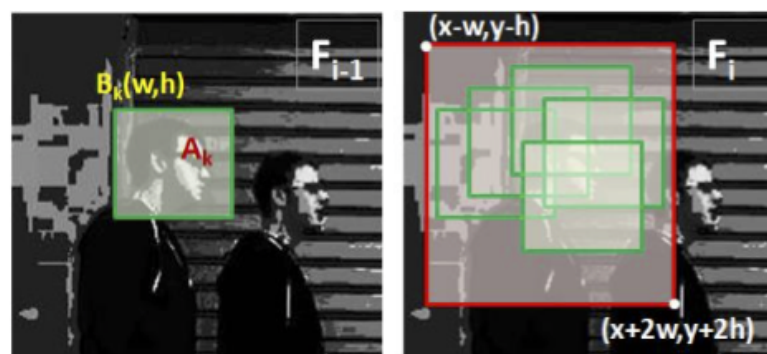
FIGURE 8.2: Searching the position of an anchor in a frame, given its position in the preceding one.

computed for homologous bounding boxes are included in a matrix M, with dimension $2w * 2h$. In particular, position $M(l, m)$ contains the value of MI computed B_k in $Fi - 1$ in position (x, y) and the homologous in F_i in position $(x-w+l, y-h+m)$. The more similar the two homologous bounding boxes, the higher the value of the MI. As a consequence, to determine the position of the anchor in the frame F_i it is sufficient to find the cell (a, b) in M corresponding to the maximum value of MI.

8.2.2.3   Outlier discarding and error correction

In the calculation of the matrix M, we can observe the possible presence of more maxima, some of which are relatively far away from the actual position of the anchor. The selection of one of these maxima, in place of the correct one, causes the bounding box move to an incorrect position. Since the latter is considered afterwards for the search of a further new position for the anchor, the error tends to propagate frame by frame leading to a completely wrong position of the anchor (drift problem, see Figure 8.3). We can notice that, while

FIGURE 8.3: Propagation of anchor location error.

the correct maximum of matrix M is usually in the centre of a cloud of relatively high values, maxima different (far) from the correct one are isolated and can be therefore considered as outliers. MIMA weights the values in matrix M with respect to the distance of the new candidate point in frame Fi from the anchor position in frame F_{i-1}. Each weight is computed as:

$$\gamma(a, b) = 1 - \frac{\sqrt{(x-a)^2 + (y-b)^2}}{\sqrt{w^2 + h^2}} \tag{8.7}$$

Therefore, the new weighted matrix M_γ is in the form:

$$M_\gamma(a, b) = \gamma(a, b) \cdot M(a, b), \tag{8.8}$$

where $0 \leq a \leq 2w$ and $0 \leq b \leq 2h$. A further increase in the precision of the tracker can be obtained by considering the specific context in which the MIMA system operates, namely, the tracking of human body parts. In fact, both the hands and the face (frontal and side pose) are usually characterized by a high content of skin. MIMA integrates a skin detector to increase the tracking accuracy of anchors attached to hands and face. The method adopted by MIMA for skin detection is the Explicitly Defined Skin Region, i.e. it defines the threshold

values for the region of skin in the $YCbCr$ colorspace. During the search for the new position of the bounding box B_k associated to the anchor A_k, MIMA does not select the absolute maximum in the weighted matrix M_γ, but considers the m highest values (in the present implementation, $m = 5$). In this phase, the original color frames are considered. For each of the corresponding positions, MIMA measures the amount of skin present in the bounding box and selects the one with more skin. In practice, each out of the m candidate bounding boxes is transformed from the RGB color space to the $YCbCr$ color space and a skin map is computed for it, i.e. a binary image where white pixels (value 1) represent skin and black pixels (value 0) represent a no-skin regions (Figure 8.4), according to the following equation:

$$skinmap(l,m) = \begin{cases} 1 & \text{if } Cb(l,m) \in [77,127] \\ & \text{and } Cr(l,m) \in [137,155] \\ 0 & \text{otherwise} \end{cases} \qquad (8.9)$$

The amount of skin for the bounding box B_k is computed by adding the number of pixels set to one inside it.
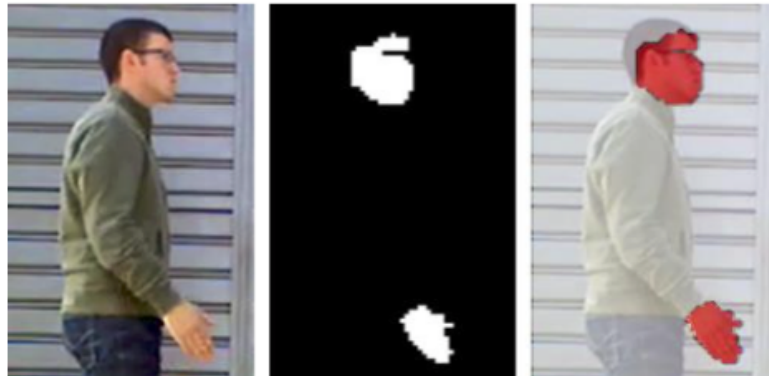
FIGURE 8.4: Skin map for the anchors of hand and face.

8.2.2.4 Anchors Overlapping and Algorithm for Conflict Solutions

Occlusion and self-occlusion represent an important critical element for all tracking systems, since the overlap of two anchors may cause that when they separate the system continues to track only one of the two. MIMA solves this problem by using an algorithm to resolve conflicts between anchors. Even for this specific procedure, color frames are considered. The process of conflict resolution is limited to the analysis of only the portion of the image occupied by the bounding boxes involved in the collision (Figure 8.5) and is divided into two steps: i) recognition of the frame in which two disjoint bounding box merge/collide, ii) reallocation of the right anchors to the two bounding boxes detected. To this aim, when MIMA detects the collision between two anchors, it starts storing the centroid c_k of each bounding box involved in the collision. In addition, MIMA calculates and stores the corresponding color histogram H_k

for the same bounding boxes. Specifically, the histogram H_k on the three channels R, G, and B is obtained by concatenating the three histograms calculated on individual channels. MIMA needs this information to be able to reassign correctly the anchors when they separate again. In the following we provide a more detailed description of the two steps of the algorithm for conflict resolution.

FIGURE 8.5: Collision between two hands shaking.

Detection of the separation between two overlapping bounding boxes   To better understand conflict resolution, assume two bounding boxes B_1 and B_2, associated respectively to anchors A_1 and A_2, and with centroids c_1 and c_2. The algorithm which determines the moment when the two anchors separate again works as follows:

1. determine the segment joining the two centroids and computes its midpoint $c_m(x_m, y_m)$;

2. using the method described in Section 8.2.2.3, compute the skin map for the portion of image delimited by the vertices $(x_m - s, y_m - t)$ e $(x_m + s, y_m + t)$, where s and t are the dimensions of the smaller bounding box between the two conflicting ones; then apply to the skin map the cascade of morphological operators opening, closing and hole filling;

3. identify the connected components C in the skin map obtained (regions with only 1s); the result of this step is the staring point to identify a possible anchor separation (i.e., the separation of a formerly merged region);

4. if at the previous point two connected components at least, say C_A and C_B, have been identified with a number of pixels higher than a threshold δ (in present implementation, δ=100 pixel and depends on image resolution), then execute the procedure to reassign anchors; in case of more candidates, the biggest ones are chosen.

Reassignment of anchors to the bounding boxes   The procedure to reassign anchors to the corresponding bounding boxes computes the Pearson correlation coefficient. The procedure uses the two histograms H_1 and H_2, that MIMA stored for the two bounding boxes B_1 and B_2, when it detected the collision, and the two histograms H_A and H_B computed for the two connected components identified when the bounding boxes separated. MIMA computes the four

correlation coefficients, one for each possible coupling $H_{1,2}/H_{A,B}$ and selects the greatest one to determine the first final coupling; the second one is a mere consequence.

8.2.3 Experimental Results

The test set (BIPLab, 2013) includes 24 videos (720x480) with different length and presenting different challenges. It was necessary to create a new dataset, since publicly available ones are not suited to carry out tests on the specific problem of body-part tracking, and in particular on the detection and recognition of interactive actions. As an example, the dataset described and made available at `http://www-prima.inrialpes.fr/FGnet/data/03-Pointing/index.html#Scene%20` setup is limited to video sequences of hand gestures, while we address a more complex setting were gestures are immersed in a real scenario involving full-body images of more subjects. Human Activity Video Datasets (`https://www.cs.utexas.edu/~chaoyeh/web_action_data/dataset_list.html`) includes either datasets with higher "resolution" actions, like running, walking, etc., or with finer action classes that are out of the scope of our study. The whole dataset is partitioned in two groups, characterized by different kinds of problems. The first group includes videos where anchors do not undergo occlusions; the second group includes videos with partial or total occlusions of the anchors to track. Examples of these groups are shown in Figure 8.6. Videos were manually annotated to build a ground truth to com-

FIGURE 8.6: Example frame with no anchor collision (left) and with anchor collision (right).

pare the anchor positions determined by MIMA. For each frame F_i, the coordinates of each anchor A_k in the ground truth are denoted with $GTi, k(x_{i,k}^{GT}, y_{i,k}^{GT})$, while those determined by MIMA are denoted as $B_{i,k}(x_{i,k}^B, y_{i,k}^B)$. Similarly, rectangular regions corresponding to the anchors from the ground truth and MIMA are denoted by $R_{i,k}^{GT}$ and $R_{i,k}^B$, respectively. Performance were measured according to the *Pascal index*, which offers an assessment of the validity of the determined anchor position. It is defined as:

$$P_{i,k} = \frac{Area(R_{i,k}^{GT} \cap R_{i,k}^B)}{Area(R_{i,k}^{GT} \cup R_{i,k}^B)} \tag{8.10}$$

An anchor tracking is considered incorrect when the corresponding Pascal index falls below a threshold of 1/3. A frame where all anchors are tracked correctly is considered valid, while if for one anchor at least the index falls below the threshold, the frame is considered invalid. Though useful to determine the correctness of an estimated anchor position, Pascal index provides no information about the amount of error. For this reason we use a further indicator which evaluates the error between the position of the bounding boxes estimated by the algorithm with respect to the ground truth of each anchor. The position error is given by the Euclidean distance computed between the coordinates B_{ik} estimated by MIMA for the anchor and the ground truth GT_{ik}:

$$E_{i,k} = \| B_{ik} - GT_{ik} \|^2 . \tag{8.11}$$

In Table 8.1 we report the results for videos in the first group, where anchors are never occluded: From Table 8.1, Table 8.2 and Table 8.3 we can observe

TABLE 8.1: Results for videos without anchor occlusion (Group I).

Video	Frames	Anchors	Valid Frames	Position Error
Video_01	65	4	87.7%	0.56%
Video_02	45	4	100%	0.34%
Video_03	180	4	98.3%	0.46%
Video_04	100	4	95%	0.46%
Video_05	100	4	100%	0.23%
Video_06	125	4	100%	0.34%
Video_07	120	4	100%	0.46%
Video_08	75	4	89.3%	0.46%
Video_09	70	4	98.6%	0.34%
Video_10	85	4	100%	0.34%
Video_11	130	4	100%	0.34%
Video_12	40	4	100%	0.34%
Video_13	55	4	100%	0.46%

that the percentage of valid frames is firmly around 98% for videos without occlusion, and 92% for videos with partial or total anchor occlusion. Therefore, MIMA is able to achieve good results even with problematic anchor tracking. A slightly lower percentage of valid frames is obtained for videos in the second group, because the positioning of bounding boxes on the respective anchors during overlap is not optimal; this is due to the fact that the bounding boxes are located on the area pertaining to the anchor appearing in the foreground in the video. However, for such videos the position error is lower on the average, because after overlap the bounding boxes are positioned precisely from the conflict resolution procedure.

TABLE 8.2: Results for videos with anchor occlusion (Group II).

Video	Frames	Anchors	Valid Frames	Position Error
Video_15	82	4	98.8%	0.23%
Video_16	50	2	88%	0.46%
Video_17	95	2	87.4%	0.56%
Video_18	180	4	98.3%	0.23%
Video_19	130	4	94.4%	0.56%
Video_20	95	4	92.6%	0.46%
Video_21	110	4	92.7%	0.46%
Video_22	100	4	83%	0.56%
Video_23	90	4	86.9%	0.46%
Video_24	82	4	98.8%	0.46%

TABLE 8.3: Summary of results.

Video	Frames	Valid Frames	Position Error
Group I	1280	98%	0.46%
Group II	1024	92.3%	0.34%

8.3 Acquiring High-resolution Face images in Outdoor Environments: A Master-Slave Calibration Algorithm

Facial recognition at-a-distance in surveillance scenarios remains an open problem, particularly due to the small number of pixels representing the facial region. To address this issue, several authors have defended the use of PTZ cameras (Jain et al., 2006; Wheeler, Weiss, and Tu, 2010; Choi, Park, and Jain, 2010; Park et al., 2013), which are capable of acquiring high resolution imagery on arbitrary scene locations. In PTZ-based systems, a master-slave configuration is usually adopted, i.e., a static camera is responsible both for detecting and tracking subjects in the scene so that it can instruct the PTZ camera to point to subject faces. While several advantages can be outlined, inter-camera calibration is the major bottleneck of this configuration, since determining the mapping function from static image coordinates to pan-tilt parameters requires depth information. To address this problem, most master-slave systems use 2D-based approximations, but, in turn, they are compelled to rely on different assumptions (e.g., similar points-of- view (Wheeler, Weiss, and Tu, 2010), intermediate zoom states (Del Bimbo et al., 2010; Marchesotti et al., 2005)) to alleviate pan-tilt inaccuracies. The use of multiple optical devices has been pointed as a solution to infer depth information through triangulation. Choi et al. (Choi, Park, and Jain, 2010) and Park et al. (Park et al., 2013) were the first to exploit this alternative without using stereographic reconstruction, which is not feasible in real-time applications. Instead, they disposed the cameras in a coaxial configuration to

ease triangulation. In addition, the authors ascertained the feasibility of facial recognition at-a-distance using the proposed calibration method. However, the highly stringent disposal of the cameras restrains its use in outdoor environments as well as its operational range (up to 15m). The work that is about to be presented aims at improving the existing master-slave systems, in particular those described in (Choi, Park, and Jain, 2010) and (Park et al., 2013), by extending PTZ-assisted facial recognition to surveillance scenarios. The calibration algorithm presented is capable of accurately estimating pan-tilt parameters, without resorting to intermediate zoom states, multiple optical devices or highly stringent configurations. The approach exploits geometric cues, i.e., the vanishing points available in the scene, to automatically estimate subjects height and thus determine their 3D position. Furthermore, we have built on the work of Lv et al. (Lv, Zhao, and Nevatia, 2002) to ensure robustness against human shape variability during walking. Considering that the proposed calibration algorithm is intended to be integrated in an automated surveillance system, we have also assessed the performance of the proposed algorithm using two challenging scenarios: 1) automatic estimation of head and feet locations using a tracking algorithm; and 2) incorrect vanishing point estimation. A comparative analysis between the most relevant master-slave systems is presented in Table 8.4 .

TABLE 8.4: Comparative analysis between the existing master-slave systems and the proposed method.

Master-slave Systems	Pan-Tilt Estim.	Camera Disposal	Interm. Zoom State	Multiple Devices	Calib. Marks
Zhou et al. 2003	Approx.	Arbitrary	No	No	Yes
Liao 2008	Approx.	Arbitrary	No	No	Yes
Marchesotti et al. 2005	Approx.	Arbitrary	Yes	No	Yes
Robert Bodor 2004	Approx.	Specific	No	No	Yes
Liu et al. 2014	Approx.	Specific	No	No	No
Wheeler, Weiss, and Tu 2010	Approx.	Specific	No	No	Yes
Del Bimbo et al. 2010	Approx.	Arbitrary	Yes	No	No
Hampapur et al. 2003	Exact.	Arbitrary	No	Yes	Yes
Choi, Park, and Jain 2010	Exact.	Specific	No	Yes	No
Park et al. 2013	Exact.	Specific	No	Yes	No
Senior, Hampapur, and Lu 2005	Exact.	Arbitrary	No	No	Yes
Fiore et al. 2008	Exact.	Arbitrary	No	No	Yes
Our Approach	Exact.	Arbitrary	No	No	No

8.3.1 Proposed Method

Let us introduce some notations that we will use in the presentation of the proposed method:

- (X, Y, Z): the 3D world coordinates;

- (X_S, Y_S, Z_S): the 3D coordinates in the static camera world referential;

- (X_p, Y_p, Z_p): the 3D coordinates in the PTZ camera world referential;

- (x_S, y_S): the 2D coordinates in the static camera image referential;

- (x_t, y_t): the 2D coordinates of a held in the static camera image referential;

- x_p, y_p: the 2D coordinates in the PTZ camera image referential;

- $(\Theta_p, \Theta_t, \Theta_z)$: the pan, tilt and zoom parameters of the PTZ camera.

In the pin-hole camera model, the projective transformation of 3D scene points onto the 2D image plane is governed by:

$$\lambda \begin{pmatrix} x_t \\ y_t \\ 1 \end{pmatrix} = \underbrace{\mathbf{K}[\mathbf{R}|\mathbf{T}])}_{:=\mathbf{P}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{8.12}$$

where λ is a scalar factor, $\mathbf{K}$ and $[\mathbf{R}|\mathbf{T}]$ represent the intrinsic and extrinsic camera matrices, which define the projection matrix $\mathbf{P}$. Let $\mathbf{p}_t = (x_t, y_t)$. Solving the equation 8.12 for (X, Y, Z) yields an under-determined system, i.e., infinite possible 3D locations for the face. As such, we propose to solve equation 1 by determining one of the 3D components. By assuming a world coordinate system (WCS) where the (XY) plane corresponds to the reference ground plane of the scene, the Z component of a subject's head corresponds to its height h. The use of height information reduces the equation 8.12 to:

$$\lambda \begin{pmatrix} \mathbf{p}_t \\ 1 \end{pmatrix} = [\mathbf{p}_1 \ \mathbf{p}_2 \ h\mathbf{p}_3 + \mathbf{p}_4] \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{8.13}$$

where $\mathbf{p}_i$ is the set of column vectors of the projection matrix $\mathbf{P}$. As such, our algorithm works on the static camera to extract x_t, y_t and infers the subjects position in the WCS using its height.

8.3.2 Height Estimation

To perform height estimation, we rely on the insight that surveillance scenarios are typically urban environments with useful geometric information that can be exploited, such as vanishing points and vanishing lines. As in (Criminisi, Reid, and Zisserman, 2000), three vanishing points $(\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z)$ are used for the X, Y and Z axis, in order to infer the height of a subject, which is vertical to a planar surface. $\mathbf{v}_x$ and $\mathbf{v}_y$ are determined from parallel lines contained in the reference plane, so that the line $\mathbf{l}$ defined by these points represents the plane vanishing line. The point $\mathbf{v}_z$ corresponds to the intersection of two lines perpendicular to the reference plane. Given $\mathbf{l}, \mathbf{v}_z$, the head ($\mathbf{p}_t$) and feet ($\mathbf{p}_b$) points in an image,

the height of a person can be obtained by:

$$h = -\frac{\|\mathbf{p}_b \times \mathbf{p}_t\|}{\alpha (\mathbf{l} \cdot \mathbf{p}_b) \|\mathbf{v}_z \times \mathbf{p}_t\|},$$ (8.14)

where $\alpha = -\frac{\|\mathbf{p}_{rb} \times \mathbf{p}_{rt}\|}{h_r (\mathbf{l} \cdot \mathbf{p}_{rb}) \|\mathbf{v}_z \times \mathbf{p}_{rt}\|}$, whereas $\mathbf{p}_{rt}$ and $\mathbf{p}_{rb}$ are top and base of a reference object in the image with height equal to h_r.

8.3.3   Pan-Tilt Angle Estimation

Considering the referential depicted in Figure 8.7, the center of rotation of the PTZ camera is given by $C = (0, \rho \sin \Theta_t, \rho \cos \Theta_t)$, being ρ the displacement between the mechanical rotation axis and the image plane (which can be approximated by the camera focal distance f). Given the 3D coordinates (X, Y, Z) of an interest point in the WCS, the location of that point with respect to the PTZ referential is obtained by:

$$\begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} = [\mathbf{R}|\mathbf{T}] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$ (8.15)

and the corrected coordinates are given by:

$$\begin{pmatrix} X_p^t \\ Y_p^t \\ Z_p^t \end{pmatrix} = \begin{pmatrix} X_p \\ Y_p - \rho \sin \Theta_t \\ Z_p + \rho \cos \Theta_t \end{pmatrix}.$$ (8.16)

The corresponding pan and tilt angles are given by:

$$\Theta_p = \arctan\left(\frac{X_p^y}{Z_p^t}\right),$$ (8.17)

and

$$\Theta_t = \arcsin \frac{Y_p'}{\sqrt{(X_p')^2 + (Y_p')^2 + (Z_p')^2}}.$$ (8.18)

8.4   Quis-Campi: Extending in the Wild Biometric Recognition to Surveillance Environments

Quis-Campi is a fully automated surveillance system for human recognition purposes, attained by combining human detection and tracking, further enhanced by a PTZ camera that delivers data with enough quality to perform biometric recognition. The system is devised over three main layers (Figure 8.8):
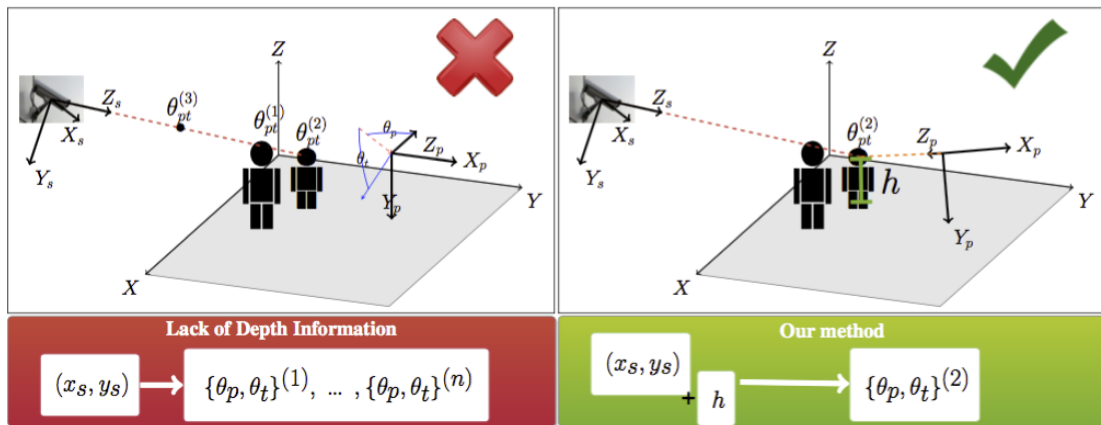
FIGURE 8.7: Illustration of the principal bottleneck of master-slave systems and the proposed strategy to address this problem. The same image pixel (x_s, y_s) corresponds to different 3D positions and consequently to different pan-tilt Θ_p, Θ_t values. The work is based on the premise that human height can be exploited to infer depth information and avoid that ambiguity.

scene understanding, camera control/synchronization, and recognition modules. Scene understanding refers to the detection and tracking of human beings. This phase should be supported by the wide-view camera so that it provides head location of persons in the scene, allowing the PTZ camera to zoom-in on those regions. Following the PTZ image acquisition, the recognition modules are responsible to infer the identify of the subject.
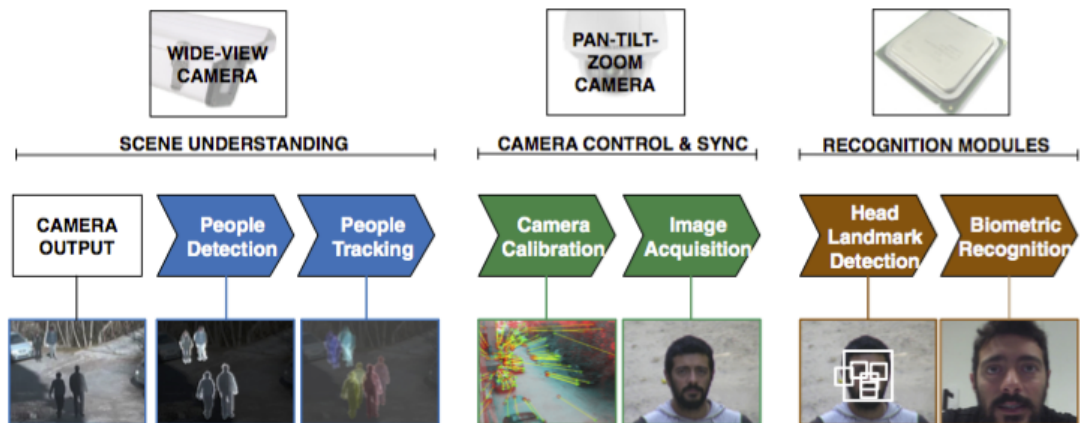
FIGURE 8.8: Working diagram of the proposed system, and the three-layer architecture: scene understanding, camera control/synchronization and recognition modules.

8.4.1 Scene Understanding

The scene understanding layer has two main modules: people detection and tracking. The first module locates persons as they enter the scene and tracks

them until they are no longer visible, taking as input the video feed from the wide-view camera, and has three main steps: background subtraction, upper-body detection and tracking (Figure 8.9).
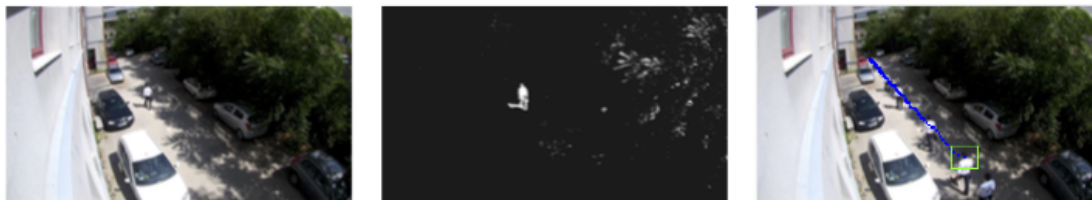
FIGURE 8.9: Illustration of the preliminary results obtained by the people detection and tracking module: sample image acquired with the wide-view camera (leftmost image); foreground regions attained by background subtraction (middle image); people tracking module results (rightmost image).

8.4.2 Camera Control and Synchronization

Considering that the wide-view and the PTZ can be disposed arbitrarily in the scene, a calibration algorithm is required to relate the image coordinates of both devices. However, due to the lack of depth information, this problem is ill-defined and thus several approximations have been proposed to alleviate the inaccuracies of 2D-based methods. With a view to determine a precise mapping the devices, different solutions have been proposed to infer 3D information from the scene. The system relies on (Neves et al., 2015a) where the subjects height is inferred and used as an ancillary measure to define a precise mapping between the cameras. Additionally, it is necessary to plan, in real-time, the sequence of PTZ observations when multiple subjects are in the scene. Despite a random walk could be adopted, this strategy would lead to failures in the observation of some targets as the number of subjects increases. For this purpose, we rely on (Neves and Proenca, 2015) where an algorithm for maximizing the observed number of targets has been devised.

8.4.3 Recognition Modules

After a successful acquisition of a PTZ shot, the recognition module should be supported by a head landmark detection phase. This strategy improves recognition performance since it determines which facial landmarks are visible, and thus decides the weight of each recognition module. Being able to describe which facial traits are visible and where, is far more important than actually getting a close estimation of the head's pose, as we can tell to which extent the trait is reliable or not. For recognition purposes, the proposed system relies on a multi- modal biometric approach that combines face, iris, periocular, ear shape and gait information.

Chapter 9

Conclusion and Future Works

In the thesis a multi biometric platform, FAIRY, has been presented. The interest in concentrating several biometric traits and related recognition techniques is twofold: from the researcher point of view the possibility of stressing different biometric traits, variously fused, can try to approximate the human behavior in recognizing individuals and help the researcher in finding better solutions. On the other hand such a platform can be exploited by users to understand which biometry or a fusion of them can best solve his own problem, before deciding to adopt a given biometric trait. Further researches will deal with the investigation of new fusion techniques. In FAIRY two different fusion approaches have been adopted. The first deals with the combination between the SRR value and the score outputted by the matching between two samples. The SRR can be calculated according to three different functions, ϕ_1, ϕ_2 and ϕ_3, defined in 7. FAIRY exploits the first function since better results have been obtained. The second approach performs fusion of physiological measures (EEG and ECG) by means of the weighted sum. The reason behind the use of two different fusion algorithms depends on the type of biometrics involved. The first methodology is used for fusing physical traits (ear, face, iris) whereas the second one is used for the physiological signals (EEG and ECG). Obviously, according to the techniques and to the kind of feature vectors extracted from the single biometrics, both techniques may be used for the fusion of any biometric presented in FAIRY. Further experimentation have been planned, so as to integrate in the FAIRY platform the possibility of choosing the fusion methodology to be adopted in each specific scenario.

Bibliography

Abate, Andrea F et al. (2007). "2D and 3D face recognition: A survey". In: *Pattern Recognition Letters* 28.14, pp. 1885–1906.

Agrafioti, F. and D. Hatzinakos (2008). "Fusion of ECG sources for human identification". In: *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, pp. 1542–1547. DOI: `10.1109/ISCCSP.2008.4537472`.

AlMahafzah, Harbi and Maen Zaid AlRwashdeh (2012). "A Survey of Multi-biometric Systems". In: *arXiv preprint arXiv:1210.0829*.

Barra, S. et al. (2014a). "Babies: Biometric authentication of newborn identities by means of ear signatures". In: *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on*, pp. 1–7. DOI: `10.1109/BIOMS.2014.6951528`.

Barra, Silvio et al. (2013). "FAME: Face Authentication for Mobile Encounter". In: *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2013 IEEE Workshop on*. IEEE, pp. 1–7.

Barra, Silvio et al. (2014b). "Babies: Biometric authentication of newborn identities by means of ear signatures". In: *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on*. IEEE, pp. 1–7.

Barra, Silvio et al. (2014c). "COMPLEX NUMBERS AS A COMPACT WAY TO REPRESENT SCORES AND THEIR RELIABILITY IN RECOGNITION BY MULTI-BIOMETRIC FUSION". In: *International Journal of Pattern Recognition and Artificial Intelligence* 28.07, p. 1460003.

Barra, Silvio et al. (2014d). "Using Mutual Information for Multi-Anchor Tracking of Human Beings". In: *Biometric Authentication*. Springer International Publishing, pp. 28–39.

Barra, Silvio et al. (2015a). "EEG/ECG Signal Fusion Aimed at Biometric Recognition". In: *New Trends in Image Analysis and Processing–ICIAP 2015 Workshops*. Springer International Publishing, pp. 35–42.

Barra, Silvio et al. (2015b). "Ubiquitous iris recognition by means of mobile devices". In: *Pattern Recognition Letters* 57, pp. 66–73.

Bermudez, T., D. Lowe, and A.-M. Arlaud-Lamborelle (2009). "EEG/ECG information fusion for epileptic event detection". In: *Digital Signal Processing, 2009 16th International Conference on*, pp. 1–8. DOI: `10.1109/ICDSP.2009.5201231`.

Bernat, Scott M. (2012). *Biometrics: Enhancing security in the public and private sectors*. Available on line. URL: `http://www.asiapacificsecuritymagazine.com/biometrics-enhancing-security-in-the-public-and-private-sectors/`.

Bhanu, Bir and Hui Chen (2003). "Human Ear Recognition in 3D". In: *Multimodal User Authentication Workshop (MMUA)*, pp. 91–98.

Biel, L. et al. (2001). "ECG analysis: a new approach in human identification". In: *Instrumentation and Measurement, IEEE Transactions on* 50.3, pp. 808–812. ISSN: 0018-9456. DOI: 10.1109/19.930458.

BIPLab (2013). *MIMA_Annotations*. URL: http://biplab.unisa.it/InterActions.zip.

Boles, Wageeh W and Boualem Boashash (1998). "A human identification technique using images of the iris and wavelet transform". In: *IEEE transactions on signal processing* 46.4, pp. 1185–1188.

Bolle, R.M. et al. (2005). "The relation between the ROC curve and the CMC". In: *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, pp. 15–20. DOI: 10.1109/AUTOID.2005.48.

Bousseljot, R, D Kreiseler, and A Schnabel (1995). "Nutzung der EKG Signaldatenbank CARDIODAT der PTB über das Internet". In: *Biomedizinische Technik/Biomedical Engineering* 40.s1, pp. 317–318.

Bowyer, Kevin W. (2012). "The Results of the NICE.II Iris Biometrics Competition". In: *Pattern Recogn. Lett.* 33.8, pp. 965–969. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2011.11.024. URL: http://dx.doi.org/10.1016/j.patrec.2011.11.024.

Bregler, C. and J. Malik (1998). "Tracking people with twists and exponential maps". In: *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pp. 8–15. DOI: 10.1109/CVPR.1998.698581.

Burge, Mark and Wilhelm Burger (1996). "Ear biometrics". In: *Biometrics*. Springer, pp. 273–285.

— (1997). "Ear biometrics for machine vision". In: *21st Workshop of the Austrian Association for Pattern Recognition*, pp. 275–282.

Cai, Deng et al. (2006). "Orthogonal Laplacianfaces for Face Recognition". In: *Image Processing, IEEE Transactions on* 15.11, pp. 3608–3614. ISSN: 1057-7149. DOI: 10.1109/TIP.2006.881945.

Campisi, P. and D. La Rocca (2014). "Brain waves for automatic biometric-based user recognition". In: *Information Forensics and Security, IEEE Transactions on* 9.5, pp. 782–800. ISSN: 1556-6013.

Chang, Kyong et al. (2003). "Comparison and combination of ear and face images in appearance-based biometrics". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.9, pp. 1160–1165. ISSN: 0162-8828.

Chen, Hui and Bir Bhanu (2007). "Human Ear Recognition in 3D". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 29.4, pp. 718–737. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1005. URL: http://dx.doi.org/10.1109/TPAMI.2007.1005.

— (2009). "Efficient recognition of highly similar 3D objects in range images". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.1, pp. 172–179.

Choi, Hyun-Cheol, Unsang Park, and A.K. Jain (2010). "PTZ camera assisted face acquisition, tracking amp; recognition". In: *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pp. 1–6. DOI: 10.1109/BTAS.2010.5634539.

Cootes, Timothy F et al. (1995). "Active shape models-their training and application". In: *Computer vision and image understanding* 61.1, pp. 38–59.

Criminisi, Antonio, Ian Reid, and Andrew Zisserman (2000). "Single view metrology". In: *International Journal of Computer Vision* 40.2, pp. 123–148.

Dame, Amaury and Eric Marchand (2010). "Accurate real-time tracking using mutual information". In: *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, pp. 47–56.

Daugman, John (2003). "The importance of being random: statistical principles of iris recognition". In: *Pattern recognition* 36.2, pp. 279–291.

— (2004). "How iris recognition works". In: *Circuits and Systems for Video Technology, IEEE Transactions on* 14.1, pp. 21–30.

Daugman, John G (1993). "High confidence visual recognition of persons by a test of statistical independence". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15.11, pp. 1148–1161.

Daultrey, S. (1976). *Principal Component Analysis*.

De Marsico, M., M. Nappi, and D. Riccio (2010a). "Face: face analysis for Commercial Entities". In: *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 1597–1600. DOI: 10.1109/ICIP.2010.5650758.

De Marsico, Maria, Michele Nappi, and Daniel Riccio (2010b). "IS_IS: Iris Segmentation for Identification Systems". In: *2010 International Conference on Pattern Recognition*. IEEE, pp. 2857–2860.

— (2013). "Fusion of Multi-biometric Recognition Results by Representing Score and Reliability as a Complex Number". English. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Ed. by José Ruiz-Shulcloper and Gabriella Sanniti di Baja. Vol. 8259. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 302–309. ISBN: 978-3-642-41826-6. DOI: 10.1007/978-3-642-41827-3_38. URL: http://dx.doi.org/10.1007/978-3-642-41827-3_38.

— (2014). "ES-RU: an entropy based rule to select representative templates in face surveillance". English. In: *Multimedia Tools and Applications* 73.1, pp. 109–128. ISSN: 1380-7501. DOI: 10.1007/s11042-012-1279-6. URL: http://dx.doi.org/10.1007/s11042-012-1279-6.

De Marsico, Maria et al. (2011). "NABS: Novel approaches for biometric systems". In: *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 41.4, pp. 481–493.

De Marsico, Maria et al. (2012). "Moving face spoofing detection via 3D projective invariants". In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, pp. 73–78.

De Marsico, Maria et al. (2015). "Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols". In: *Pattern Recognition Letters* 57, pp. 17–23.

Dehache, Ismahene and Labiba Souici-Meslati (2012). "A multibiometric system for identity verification based on fingerprints and signatures". In: *Complex Systems (ICCS), 2012 International Conference on*. IEEE, pp. 1–5.

Del Bimbo, A. et al. (2010). "Exploiting Distinctive Visual Landmark Maps in Pan-tilt-zoom Camera Networks". In: *Comput. Vis. Image Underst.* 114.6, pp. 611–623. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2010.01.007. URL: http://dx.doi.org/10.1016/j.cviu.2010.01.007.

Dobeš, M et al. (2004). "Human eye iris recognition using the mutual information". In: *Optik-International Journal for Light and Electron Optics* 115.9, pp. 399–404.

Dobeš, M et al. (2006). "Human eye localization using the modified Hough transform". In: *Optik-International Journal for Light and Electron Optics* 117.10, pp. 468–473.

Dobeš M., Machala L. (2008). *UPOL Iris Image Database*. [Available Online] at http://www.inf.upol.cz/iris/.

Draper, Harold W et al. (1964). "The corrected orthogonal electrocardiogram and vectorcardiogram in 510 normal men (Frank lead system)". In: *Circulation* 30.6, pp. 853–864.

Eidenberger, Horst (2006). "Kalman filtering for pose-invariant face recognition". In: *Image Processing, 2006 IEEE International Conference on*. IEEE, pp. 2037–2040.

Etemad, Kamran and Rama Chellappa (1997). "Discriminant analysis for recognition of human face images". In: *JOSA A* 14.8, pp. 1724–1733.

Fathima, A Annis et al. (2014). "Fusion framework for multimodal biometric person authentication system". In: *IA ENG Int. J. Comput. Sci* 41.1.

Fierrez-Aguilar, Julian et al. (2005). "Discriminative multimodal biometric authentication based on quality measures". In: *Pattern Recognition* 38.5, pp. 777–779.

Fiore, Loren et al. (2008). "Multi-camera human activity monitoring". In: *Journal of Intelligent and Robotic Systems* 52.1, pp. 5–43.

Fraschini, M. et al. (2015). "An EEG-Based Biometric System Using Eigenvector Centrality in Resting State Brain Networks". In: *Signal Processing Letters, IEEE* 22.6, pp. 666–670. ISSN: 1070-9908. DOI: 10.1109/LSP.2014.2367091.

Fratini, A. et al. (2015). "Individual identification via electrocardiogram analysis". In: *BioMedical Engineering Online* 14.1. cited By 0. DOI: 10.1186/s12938-015-0072-y.

Garcia, Christophe and Manolis Delakis (2002). "A neural architecture for fast and robust face detection". In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Vol. 2. IEEE, pp. 44–47.

Garcia-Salicetti, Sonia et al. (2005). "A generic protocol for multibiometric systems evaluation on virtual and real subjects". In: *Audio-and Video-Based Biometric Person Authentication*. Springer, pp. 494–502.

Goldberger, A. L. et al. (2000). "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals". In: *Circulation* 101.23, e215–e220.

Hampapur, A. et al. (2003). "Face cataloger: multi-scale imaging for relating identity to location". In: *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, pp. 13–20. DOI: `10.1109/AVSS.2003.1217896`.

Haritaoglu, Ismail, Davis Harwood, and Larry S. David (2000). "W4: Real-Time Surveillance of People and Their Activities". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.8, pp. 809–830. ISSN: 0162-8828. DOI: `10.1109/34.868683`. URL: `http://dx.doi.org/10.1109/34.868683`.

He, Xiaofei et al. (2005a). "Face recognition using Laplacianfaces". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.3, pp. 328–340.

He, Xiaofei et al. (2005b). "Neighborhood preserving embedding". In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. Vol. 2. IEEE, pp. 1208–1213.

Hobijn, Bart and Erick Sager (2007). "What has homeland security cost? an assessment: 2001-2005". In: *Current Issues in Economics and Finance* 13.2.

Hoekema, Rudi, Gérard JH Uijen, and Adriaan Van Oosterom (2001). "Geometrical aspects of the interindividual variability of multilead ECG recordings". In: *Biomedical Engineering, IEEE Transactions on* 48.5, pp. 551–559.

Hogg, David (1983). "Model-based vision: a program to see a walking person". In: *Image and Vision Computing* 1.1, pp. 5 –20. ISSN: 0262-8856.

Hurley, David J, Mark S Nixon, and John N Carter (2000). "Automatic ear recognition by force field transformations". In: *Visual Biometrics (Ref. No. 2000/018), IEE Colloquium on*. IET, pp. 7–1.

— (2002). "Force field energy functionals for image feature extraction". In: *Image and Vision computing* 20.5, pp. 311–317.

— (2005). "Force field feature extraction for ear biometrics". In: *Computer Vision and Image Understanding* 98.3, pp. 491–512.

Islam, Syed MS et al. (2013). "Multibiometric human recognition using 3D ear and face features". In: *Pattern Recognition* 46.3, pp. 613–627.

Jafri, Rabia and Hamid R Arabnia (2009). "A survey of face recognition techniques". In: *journal of information processing systems* 5.2, pp. 41–68.

Jain, Anil, Karthik Nandakumar, and Arun Ross (2005). "Score normalization in multimodal biometric systems". In: *Pattern recognition* 38.12, pp. 2270–2285.

Jain, Ankur et al. (2006). "Using stationary-dynamic camera assemblies for wide-area video surveillance and selective attention". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Vol. 1. IEEE, pp. 537–544.

Jeong, Dae Sik et al. (2010). "A new iris segmentation method for non-ideal iris images". In: *Image and vision computing* 28.2, pp. 254–260.

Johansson, Gunnar (1975). "Visual motion perception." In: *Scientific American*.

Joo Er, Meng, Weilong Chen, and Shiqian Wu (2005). "High-speed face recognition based on discrete cosine transform and RBF neural networks". In: *Neural Networks, IEEE Transactions on* 16.3, pp. 679–691.

Kim, Tae-Kyun et al. (2005). "Component-based LDA face description for image retrieval and MPEG-7 standardisation". In: *Image and Vision Computing* 23.7, pp. 631–642.

Kirby, Michael and Lawrence Sirovich (1990). "Application of the Karhunen-Loeve procedure for the characterization of human faces". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12.1, pp. 103–108.

Kittler, Josef et al. (1998). "On combining classifiers". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20.3, pp. 226–239.

Koh, Lian Hock, Surendra Ranganath, and YV Venkatesh (2002). "An integrated automatic face detection and recognition system". In: *Pattern Recognition* 35.6, pp. 1259–1273.

Koné, Chaka et al. (2015). "Multimodal Recognition of Emotions Using Physiological Signals with the Method of Decision-Level Fusion for Healthcare Applications". In: *Inclusive Smart Cities and e-Health.* Springer, pp. 301–306.

Kozmann, Gyorgy, Robert L Lux, and Larry S Green (1989). "Sources of variability in normal body surface potential maps." In: *Circulation* 79.5, pp. 1077–1083.

Kryszczuk, Krzysztof et al. (2007). "Reliability-based decision fusion in multimodal biometric verification systems". In: *EURASIP Journal on Applied Signal Processing* 2007.1, pp. 74–74.

Kyoso, M. and A. Uchiyama (2001). "Development of an ECG identification system". In: *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE.* Vol. 4, 3721–3723 vol.4. DOI: 10.1109/IEMBS.2001.1019645.

La Rocca, D. et al. (2014). "Human Brain Distinctiveness Based on EEG Spectral Coherence Connectivity". In: *Biomedical Engineering, IEEE Transactions on* 61.9, pp. 2406–2412. ISSN: 0018-9294. DOI: 10.1109/TBME.2014.2317881.

Lachenbruch, Peter A and M Goldstein (1979). "Discriminant analysis". In: *Biometrics*, pp. 69–85.

Lee, Jae-Ho and Whoi-Yul Kim (2004). "Video Summarization and Retrieval System Using Face Recognition and MPEG-7 Descriptors". English. In: *Image and Video Retrieval.* Ed. by Peter Enser et al. Vol. 3115. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 170–178. ISBN: 978-3-540-22539-3. DOI: 10.1007/978-3-540-27814-6_23. URL: http://dx.doi.org/10.1007/978-3-540-27814-6_23.

Liao H.C., Cho Y.C. (2008). "A new calibration method and its application for the cooperation of wide-angle and pan-tilt-zoom cameras". In: *Information Technology Journal* 7.8, pp. 1096–1105.

Lim, Shinyoung et al. (2001). "Efficient iris recognition through improvement of feature vector and classifier". In: *ETRI journal* 23.2, pp. 61–70.

Liu, Yu et al. (2014). "A Master-Slave Surveillance System to Acquire Panoramic and Multiscale Videos". In: *The Scientific World Journal* 2014.

Lopes Da Silva, F. (2013). "{EEG} and MEG: Relevance to Neuroscience". In: *Neuron* 80.5, pp. 1112 –1128. ISSN: 0896-6273. DOI: http://dx.doi.org/10.1016/j.neuron.2013.10.017. URL: http://www.sciencedirect.com/science/article/pii/S0896627313009203.

Loutas, Evangelos, Ioannis Pitas, and Christophoros Nikou (2004). "Probabilistic multiple face detection and tracking using entropy measures". In: *Circuits and Systems for Video Technology, IEEE Transactions on* 14.1, pp. 128–135.

Lu, Juwei, Kostantinos N Plataniotis, and Anastasios N Venetsanopoulos (2003). "Face recognition using LDA-based algorithms". In: *Neural Networks, IEEE Transactions on* 14.1, pp. 195–200.

Lv, Fengjun, Tao Zhao, and R. Nevatia (2002). "Self-calibration of a camera from video of a walking human". In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Vol. 1, 562–567 vol.1. DOI: 10.1109/ICPR.2002.1044793.

Mäenpää, Topi and Matti Pietikäinen (2003). "Multi-scale binary patterns for texture analysis". In: *Image Analysis*. Springer, pp. 885–892.

Marchesotti, Luca et al. (2005). "Cooperative multisensor system for real-time face detection and tracking in uncontrolled conditions". In: *Electronic Imaging 2005*. International Society for Optics and Photonics, pp. 100–114.

Martínez, Aleix M (2002). "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.6, pp. 748–763.

Martínez, Aleix M and Avinash C Kak (2001). "Pca versus lda". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.2, pp. 228–233.

Matos, André Cigarro, André Lourenço, and José Nascimento (2014). "Embedded System for Individual Recognition Based on {ECG} Biometrics". In: *Procedia Technology* 17. Conference on Electronics, Telecommunications and Computers – {CETC} 2013., pp. 265 –272. ISSN: 2212-0173.

Milborrow, Stephen and Fred Nicolls (2008). "Locating facial features with an extended active shape model". In: *Computer Vision–ECCV 2008*. Springer, pp. 504–513.

Moreno, B., A. Sanchez, and J.F. Velez (1999). "On the use of outer ear images for personal identification in security applications". In: *Security Technology, 1999. Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on*, pp. 469–476.

Nandakumar, Karthik et al. (2006). "Quality-based score level fusion in multibiometric systems". In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 4. IEEE, pp. 473–476.

Nandakumar, Karthik et al. (2008). "Likelihood ratio-based biometric score fusion". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30.2, pp. 342–347.

Nappi, M. and D. Riccio (2008). *Moderne Tecniche di Elaborazione di Immagini e Biometria*.

Neves, J.C. and H. Proenca (2015). "Dynamic camera scheduling for visual surveillance in crowded scenes using Markov random fields". In: *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*, pp. 1–6. DOI: 10.1109/AVSS.2015.7301790.

Neves, JC et al. (2015a). "A calibration algorithm for multi-camera visual surveillance systems based on single-view metrology". In: *Pattern Recognition and Image Analysis*. Springer International Publishing, pp. 552–559.

Neves, João C et al. (2015b). "Quis-Campi: Extending in the Wild Biometric Recognition to Surveillance Environments". In: *New Trends in Image Analysis and Processing–ICIAP 2015 Workshops*. Springer International Publishing, pp. 59–68.

Noh, Yun-Hong, Gi-Hyun Hwang, and Do-Un Jeong (2011). "Implementation of real-time abnormal ECG detection algorithm for wearable healthcare". In: *Computer Sciences and Convergence Information Technology (ICCIT), 2011 6th International Conference on*, pp. 111–114.

Notre Dame, UND University of (2005). *Notre Dame Ear Database*. URL: http://www.nd.edu/cvrl/UNDBiometricsDatabase.html..

Odinaka, I. et al. (2012). "ECG Biometric Recognition: A Comparative Analysis". In: *Information Forensics and Security, IEEE Transactions on* 7.6, pp. 1812–1824. ISSN: 1556-6013. DOI: 10.1109/TIFS.2012.2215324.

Okada, R., Y. Shirai, and J. Miura (2000). "Tracking a person with 3-D motion by integrating optical flow and depth". In: *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp. 336–341. DOI: 10.1109/AFGR.2000.840656.

Pan, Yaozhang et al. (2008). "Detection of seizures in EEG signal using weighted locally linear embedding and SVM classifier". In: *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pp. 358–363. DOI: 10.1109/ICCIS.2008.4670889.

Park, Unsang et al. (2013). "Face Tracking and Recognition at a Distance: A Coaxial and Concentric PTZ Camera System". In: *Trans. Info. For. Sec.* 8.10, pp. 1665–1677. ISSN: 1556-6013. DOI: 10.1109/TIFS.2013.2261061. URL: http://dx.doi.org/10.1109/TIFS.2013.2261061.

Phillips, P Jonathon et al. (1998). "The FERET database and evaluation procedure for face-recognition algorithms". In: *Image and vision computing* 16.5, pp. 295–306.

Plataniotis, K.N., D. Hatzinakos, and J.K.M. Lee (2006). "ECG Biometric Recognition Without Fiducial Detection". In: *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, pp. 1–6. DOI: 10.1109/BCC.2006.4341628.

Poh, Norman and Samy Bengio (2005). "Improving fusion with margin-derived confidence in biometric authentication tasks". In: *Audio-and Video-Based Biometric Person Authentication*. Springer, pp. 474–483.

Poh, Norman and Josef Kittler (2012). "A unified framework for biometric expert fusion incorporating quality measures". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34.1, pp. 3–18.

Pozo-Banos, Marcos Del et al. (2014). "Electroencephalogram subject identification: A review". In: *Expert Systems with Applications* 41.15, pp. 6537 –6554. ISSN: 0957-4174. DOI: http://dx.doi.org/10.1016/j.eswa.2014.05.013. URL: http://www.sciencedirect.com/science/article/pii/S0957417414002930.

Prittopaul, P., S. Sathya, and K. Jayasree (2015). "Cyber Physical System approach for heart attack detection and control using wireless monitoring and

actuation system". In: *Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on*, pp. 1–6. DOI: 10.1109/ISCO.2015.7282352.

Proenca, H. and L.A. Alexandre (2007). "The NICE.I: Noisy Iris Challenge Evaluation - Part I". In: *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pp. 1–4. DOI: 10.1109/BTAS.2007.4401910.

Proença, Hugo and LuísA. Alexandre (2005). "UBIRIS: A Noisy Iris Image Database". English. In: *Image Analysis and Processing – ICIAP 2005*. Ed. by Fabio Roli and Sergio Vitulano. Vol. 3617. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 970–977. ISBN: 978-3-540-28869-5. DOI: 10.1007/11553595_119. URL: http://dx.doi.org/10.1007/11553595_119.

Ravish, D.K. et al. (2014). "Heart function monitoring, prediction and prevention of Heart Attacks: Using Artificial Neural Networks". In: *Contemporary Computing and Informatics (IC3I), 2014 International Conference on*, pp. 1–6. DOI: 10.1109/IC3I.2014.7019580.

Riera, Alejandro et al. (2008). "Starfast: a wire-less wearable EEG/ECG biometric system based on the enobio sensor". In: *Proceedings of the International Workshop on Wearable Micro and Nanosystems for Personalised Health*.

Riera, Alessandro et al. (2009). "1 Multimodal Physiological Biometrics Authentication". In: *Biometrics: Theory, Methods, and Applications*, pp. 461–482.

Robert Bodor Ryan Morlok, Nikolaos Papanikolopoulos (2004). "Dual-camera system for multi-level activity recognition". In: *in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 643–648.

Rosli, N.A.I.M. et al. (2014). "Electrocardiographic (ECG) and Electromyographic (EMG) signals fusion for physiological device in rehab application". In: *Research and Development (SCOReD), 2014 IEEE Student Conference on*, pp. 1–5. DOI: 10.1109/SCORED.2014.7072965.

Ross, Arun and Anil Jain (2003). "Information fusion in biometrics". In: *Pattern recognition letters* 24.13, pp. 2115–2125.

Ross, Arun A and Rohin Govindarajan (2005). "Feature level fusion of hand and face biometrics". In: *Defense and Security*. International Society for Optics and Photonics, pp. 196–204.

Ross, Arun A, Karthik Nandakumar, and Anil K Jain (2006). *Handbook of multibiometrics*. Vol. 6. Springer Science & Business Media.

Sakai, M. and Daming Wei (2008). "Wavelet shrinkage applications of EEG-ECG-based human-computer interface". In: *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on*, pp. 538–543. DOI: 10.1109/CIT.2008.4594732.

Sanaa, Anupam, Phalguni Guptaa, and Ruma Purkaitb. "Ear Biometrics: A New Approach". In: *Biometrics* 1, 1ntroduction.

Sato, Koichi and JK Aggarwal (2001). "Tracking and recognizing two-person interactions in outdoor image sequences". In: *Multi-Object Tracking, 2001. Proceedings. 2001 IEEE Workshop on*. IEEE, pp. 87–94.

Schalk, Gerwin et al. (2004). "BCI2000: A general-purpose brain-computer interface (BCI) system". In: *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING* 51.6, p. 2004.

Senior, A.W., A. Hampapur, and M. Lu (2005). "Acquiring Multi-Scale Images by Pan-Tilt-Zoom Control and Automatic Multi-Camera Calibration". In: *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*. Vol. 1, pp. 433–438. DOI: 10.1109/ACVMOT.2005.16.

Shahid, S., G. Prasad, and R.K. Sinha (2011). "On fusion of heart and brain signals for hybrid BCI". In: *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*, pp. 48–52. DOI: 10.1109/NER.2011.5910486.

Shantha Selva Kumari, R. and J. Prabin Jose (2011). "Seizure detection in EEG using time frequency analysis and SVM". In: *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*, pp. 626–630. DOI: 10.1109/ICETECT.2011.5760193.

Shen, T.W., W.J. Tompkins, and Y.H. Hu (2002). "One-lead ECG for identity verification". In: *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*. Vol. 1, 62–63 vol.1. DOI: 10.1109/IEMBS.2002.1134388.

Shin, Kwang Yong et al. (2012). "New iris recognition method for noisy iris images". In: *Pattern Recognition Letters* 33.8, pp. 991–999.

Sidenbladh, Hedvig, Michael J Black, and David J Fleet (2000). "Stochastic tracking of 3D human figures using 2D image motion". In: *Computer Vision—ECCV 2000*. Springer, pp. 702–718.

Sim, Terence, Simon Baker, and Maan Bsat (2002). "The CMU pose, illumination, and expression (PIE) database". In: *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, pp. 46–51.

Soria-Frisch, A., A. Riera, and S. Dunne (2010). "Fusion operators for multimodal biometric authentication based on physiological signals". In: *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pp. 1–7. DOI: 10.1109/FUZZY.2010.5584121.

Taubin, Gabriel (1991). "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 13.11, pp. 1115–1138. ISSN: 0162-8828. DOI: 10.1109/34.103273.

Turk, Matthew and Alex Pentland (1991). "Eigenfaces for recognition". In: *Journal of cognitive neuroscience* 3.1, pp. 71–86.

Ulery, Brad et al. (2006). *Studies of biometric fusion*. US Department of Commerce, National Institute of Standards and Technology.

Verma, GK and US Tiwary (2014). "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals." In: *NeuroImage* 102, p. 162.

Viola, Paul and Michael Jones (2001). "Rapid object detection using a boosted cascade of simple features". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, pp. I–511.

Westphal, Günter and Rolf P Würtz (2009). "Combining feature-and correspondence-based methods for visual object recognition". In: *Neural Computation* 21.7, pp. 1952–1989.

Wheeler, Frederick W, Richard L Weiss, and Peter H Tu (2010). "Face recognition at a distance system for surveillance applications". In: *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, pp. 1–8.

Wildes, Richard P (1997). "Iris recognition: an emerging biometric technology". In: *Proceedings of the IEEE* 85.9, pp. 1348–1363.

Wiskott, Laurenz et al. (1997). "Face recognition by elastic bunch graph matching". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19.7, pp. 775–779.

Wren, C.R. et al. (1997). "Pfinder: real-time tracking of the human body". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19.7, pp. 780–785. ISSN: 0162-8828. DOI: 10.1109/34.598236.

Yan, Ping and Kevin W Bowyer (2007a). "A fast algorithm for ICP-based 3D shape biometrics". In: *Computer Vision and Image Understanding* 107.3, pp. 195–202.

— (2007b). "Biometric recognition using 3D ear shape". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.8, pp. 1297–1308.

Yan, Ping, Kevin W Bowyer, and Kyong J Chang (2005). "ICP-based approaches for 3D ear recognition". In: *Defense and Security*. International Society for Optics and Photonics, pp. 282–291.

Yan, Ping and KevinW Bowyer (2005). "Empirical evaluation of advanced ear biometrics". In: *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, pp. 41–41.

Yang, Jian et al. (2003). "Feature fusion: parallel strategy vs. serial strategy". In: *Pattern Recognition* 36.6, pp. 1369–1381.

Zhang, Lei, Quanxue Gao, and David Zhang (2007). "Block independent component analysis for face recognition". In: *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. IEEE, pp. 217–222.

Zhao, Wenyi et al. (2003). "Face recognition: A literature survey". In: *ACM computing surveys (CSUR)* 35.4, pp. 399–458.

Zhou, Xuhui et al. (2003). "A Master-slave System to Acquire Biometric Imagery of Humans at Distance". In: *First ACM SIGMM International Workshop on Video Surveillance*. IWVS '03. Berkeley, California: ACM, pp. 113–120. ISBN: 1-58113-780-X. DOI: 10.1145/982452.982467. URL: http://doi.acm.org/10.1145/982452.982467.