UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Ocular Recognition in Uncontrolled Environments: Proof-of-Concept

Emanuel da Silva Grancho

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática
(2º ciclo de estudos)

Orientador: Prof. Doutor Hugo Proença

Covilhã, Outubro de 2014

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Hugo Proença, whose expertise, understanding and patience led me to successfully reach another stage of my academic career

I am also thankful to my colleagues at SOCIA Lab, who had the patience to transmit their knowledge. His good spirit and friendship were a fundamental pillar for the realization of this work.

To my parents Antonio and Conceição and sister Alexandra I would like to express my gratitude for encouragement and support.To my girlfriend Claudia, I am thankful for all the strength and support, and for always believing in me.

Finally, a special thanks to all the volunteers who in one way or another demonstrated curiosity to participate and learn more about this project.

Resumo

A biometria é uma área em grande expansão e é considerada uma possível solução para casos onde são exigidos parâmetros altos de autenticação. Embora esta área esteja bastante desenvolvia em termos teóricos, a passagem da mesma para a prática ainda apresenta alguns problemas. Os sistemas existentes no mercado ainda se encontram dependentes de um alto nível de cooperação, para que se possa alcançar níveis de performance aceitáveis. É com este objectivo que o seguinte trabalho é desenvolvido. Através do estudo do estado da arte, é proposto e provado um um novo sistema biométrico menos cooperativo que se enquadre nos parâmetros normalmente exigidos.

Resumo alargado

A constante necessidade de parâmetros mais elevados de segurança, nomeadamente ao nível de autenticação, leva ao estudo biometria como possível solução. Actualmente os mecanismos existentes nesta área tem por base o conhecimento de algo que se sabe ”password” ou algo que se possui ”codigo Pin”. Contudo este tipo de informação é facilmente corrompida ou contornada. Desta forma a biometria é vista como uma solução mais robusta, pois garante que a autenticação seja feita com base em medidas físicas ou compartimentais que definem algo que a pessoa é ou faz (”who you are” ou ”what you do”).

Sendo a biometria uma solução bastante promissora na autenticação de indivíduos, é cada vez mais comum o aparecimento de novos sistemas biométricos. Estes sistemas recorrem a medidas físicas ou comportamentais, de forma a possibilitar uma autenticação (reconhecimento) com um grau de certeza bastante considerável. O reconhecimento com base no movimento do corpo humano (gait), feições da face ou padrões estruturais da íris, são alguns exemplos de fontes de informação em que os sistemas actuais se podem basear. Contudo, e apesar de provarem um bom desempenho no papel de agentes de reconhecimento autónomo, ainda estão muito dependentes a nível de cooperação exigida. Tendo isto em conta, e tudo o que já existe no ramo do reconhecimento biometrico, esta área está a dar passos no sentido de tornar os seus métodos o menos cooperativos poss□veis. Possibilitando deste modo alargar os seus objectivos para além da mera autenticação em ambientes controlados, para casos de vigilância e controlo em ambientes não cooperativos (e.g. motins, assaltos, aeroportos).

É nesta perspectiva que o seguinte projecto surge. Através do estudo do estado da arte, pretende provar que é possível criar um sistema capaz de agir perante ambientes menos cooperativos, sendo capaz de detectar e reconhecer uma pessoa que se apresente ao seu alcance.O sistema proposto PAIRS (Periocular and Iris Recognition Systema) tal como nome indica, efectua o reconhecimento através de informação extraída da íris e da região periocular (região circundante aos olhos). O sistema é construído com base em quatro etapas: captura de dados, pré-processamento, extração de características e reconhecimento. Na etapa de captura de dados, foi montado um dispositivo de aquisição de imagens com alta resolução com a capacidade de capturar no espectro NIR (Near-Infra-Red). A captura de imagens neste espectro tem como principal linha de conta, o favorecimento do reconhecimento através da íris, visto que a captura de imagens sobre o espectro visível seria mais sensível a variações da luz ambiente. Posteriormente a etapa de pré-processamento implementada, incorpora todos os módulos do sistema responsáveis pela detecção do utilizador, avaliação de qualidade de imagem e segmentação da íris. O modulo de detecção é responsável pelo desencadear de todo o processo, uma vez que esta é responsável pela verificação da exist□ncia de um pessoa em cena. Verificada a sua exist□ncia, são localizadas as regiões de interesse correspondentes □ íris e ao periocular, sendo também verificada a qualidade com que estas foram adquiridas. Concluídas estas etapas, a íris do olho esquerdo é segmentada e normalizada. Posteriormente e com base em vários descritores, é extraída a informação biométrica das regiões de interesse encontradas, e é criado um vector de características biométricas. Por fim, é efectuada a comparação dos dados biometricos recolhidos, com os já armazenados na base de dados, possibilitando a criação de uma lista com os níveis de semelhança em termos biometricos, obtendo assim um resposta final do sistema. Concluída a implementação do sistema, foi adquirido um conjunto de imagens

capturadas através do sistema implementado, com a participação de um grupo de voluntários. Este conjunto de imagens permitiu efectuar alguns testes de desempenho, verificar e afinar alguns parâmetros, e proceder a optimização das componentes de extração de características e reconhecimento do sistema. Analisados os resultados foi possível provar que o sistema proposto tem a capacidade de exercer as suas funções perante condições menos cooperativas.

Abstract

Biometry is an area in great expansion and is considered as possible solution to cases where high authentication parameters are required. Although this area is quite advanced in theoretical terms, using it in practical terms still carries some problems. The systems available still depend on a high cooperation level to achieve acceptable performance levels, which was the backdrop to the development of the following project. By studying the state of the art, we propose the creation of a new and less cooperative biometric system that reaches acceptable performance levels.

Keywords

Biometrics, Iris, Periocular, Non-cooperative system.

Contents

List of Figures

List of Tables

Acronyms

UBI	Universidade da Beira Interior
DNA	Deoxyribonucleic acid
FAR	False acceptance rate
FRR	False Rejection Rate
FER	Failure to enroll rate
FTA	Failure to acquire rate
NSRI	Number of successful recognitions by impostors
NARI	Number of attempts at recognitions by impostors
NFARAU	Number of failed attempts at recognition by authorizes use
NRAU	Number of attempts at recognition by authorized users
CER	Cross Error Rate
CMC	Cumulative Match Characteristic
GST	Generalized Structure Tensor Algorithm
HD	Hamming distance
LOG	Laplacian of Gaussian Filters
HOG	Histogram of oriented gradients
LBP	Local Binary pattern
SIFT	Scale-invariant feature transform
SURF	Speeded-up robust features
ULBP	Uniform Local Binary Patterns
DCT	Discrete cosine transform
DWT	Discrete Wavelet Transform
SSGA	Steady-state Genetic Algorithm
NIR	Near-infrared
IR	Infrared
ROI	Region of intress

Chapter 1

Intoduction

Man has been able to distinguish and recognize himself through the human body's physical and/or behavioural properties since the dawn of civilization. Biometry has been used in several ways throughout history, such as in rock painting, the signing of important documents or in crime investigation based on fingerprint. Recognizing another person can also be shown less directly, for example, when recognizing someone through speech or gait.

As the population grows, the need to create more capable and autonomous biometric systems also arises. Automated biometric systems have only appeared in the past decades due to the considerable advances in the computational area. However, the basic idea has been developed before. The first systems began being used in order to control access, punctuality in the office or to solve criminal cases.

Daily life is currently associated with technologies that enable access to personal data or even to certain physical spaces, deposit boxes or server rooms in data centres, are some examples when it is important to safeguard access only to whoever possesses valid credentials. Although there are relevant forms of authentication, there is still the need to search for new solutions in the field in order to achieve a higher and less cooperative security level. Nowadays, not only is there a need to create authentication systems through identity verification systems, but also to create systems more directed at identification. Factors like the exponential growth of the world's population or the concentration of great masses in populated areas leads to the search for less cooperative systems that enable for the identification of a great number of people in less daily life scenarios.

Biometrics is a steadily increasing research area due to its promising security and authentication properties in vast areas of civil and military matters. However, it is still in a transition phase between investigation and large-scale production of systems that can be adapted to daily life. Aware of these factors, this project aims to implement a biometric system capable of working in a less cooperative way for the user.

1.1 Motivation and Objectives

Considering biometrics as a significantly promising field in the nearby future, the study of its aspects is quite alluring. Although this field is deeply mature in terms of research, there are still some obstacles regarding its practical aspect, namely in the creation of more dynamic systems capable of working in the most adverse conditions.

Given how important it is to create systems, which are more and more directed at less cooperative environments. This project proposes the study of the state of the art associated with this type of technology, aiming to project, implement and prove a new biometric system. More-

over, we propose a new system that belongs to the domain of identification systems capable of working in real time and based on biometric features of the area surrounding the eyes (periocular region) and the iris. By creating this kind of multimodal system (with several traits), the intention is to achieve an acceptable performance level, in which the user's data is acquired at a distance between 1.5 and 2.10 m.

1.2 Disertation Outline

The document then presents the following structure:

Chapter 2 This chapter aims to introduce the components of the state of the art that underlie this project. It begins with a historical introduction of this field of work and then presents some concepts associated with it. We then present what can be done best in terms of recognition through feature extraction of the iris, the periocular region and both. Finally, some concepts related to the implementation of less cooperative systems are approached.

Chapter 3 The presentation of the theoretical concepts and the methods is thus concluded. Chapter 3 presents and explains the system proposed. First, the structure of the implemented system is introduced from a higher-level point of view in order to understand which modules are implemented and how they interact with one another. Each module is elaborately explained throughout this chapter.

Chapter 4 This chapter basically presents and discusses the results obtained, giving equal importance to the work developed in terms of data acquisition and optimization that allow the development and perfecting of the system proposed.

Chapter 5 The project is summarized and conclusions are drawn as to the work developed. Some aspects that we consider worth investigating and implementing in the future are also presented in this chapter.

Chapter 2

State of The Art

This chapter addresses the biometric state of art, particularly those approaches that take into consideration the iris and periocular regions as sources of information. It begins by presenting a definition of biometrics and some historical concepts. It is essential, at this point, to introduce the basic concepts associated with biometric systems in general. Subsequently the chapter is directed to the biometric recognition through the iris, periocular and both. Each of these is presented according to authors who have already investigated this area. Finally, the concept of the non-cooperative biometric systems and the implications that it has is approached.

2.1 Biometrics

In order to obtain an optimal definition that would be useful to this dissertation's context, this section provides a set of possible definitions for biometry from different sources.Complementary to this, a list of the major milestones throughout the history of biometrics is presented to assist in the understanding of the evolution of this concept side by side with civilization.

The word *"biometric"* comes from the ancient Greek *bios (life)* and *metron (measure)*. By deconstructing the origin of the word it is possible to understand that biometry is a science that uses metrics to apply to living beings, and in this context Biometrics refers to a set of techniques or metrics for uniquely recognizing a person based on one or more behavioural or physical traits.

> *"Biometric identification, or isometrics, refers to identifying an individual based on his or her distinguishing characteristics. More precisely, biometric is the science of identifying, or verifying the identity of, a person based on physiological or behavioural characteristics" [1]*

> *"Biometrics is the science of establishing the identity of an individual based on the physical, chemical or behavioural attributes of the person." [2]*

> *"Biometrics is the automated recognition of individuals based on their behavioural and biological characteristics." [3]*

From a historical point of view, as will be seen below, biometric techniques are based on ideas that were originally conceived at the beginning of civilization. However, the emergence of automated systems dates back only a few decades due to the significant advances in the field

of computer processing.

- At the beginning of mankind over 30,000 years ago, prehistoric man painted the walls of his cave and signed his painting with hand prints [4].

- In Babylonian in 500 BC, fingerprints were used as a form of signature ”Babylonian business transactions are recorded in clay tablets that include fingerprints.”

- In the 14th century, a Spanish explorer and writer, João de Barros, reported that Chinese merchants utilized fingerprints for commercial transactions and Chinese parents differentiate their children by using fingerprints and footprints.

- In Egypt, commercial transactions were made taking into account the physical characteristics of the trader. Allowing them to recognize new elements in the profession or recognize merchants with a certain reputation.

- More recently, until the late 1800s, the photographic memory was a technique used in the Western world for the automatic recognition of individuals.

- In 1882, the French anthropologist Alphonse Bertillion created what many recognise as the first true biometric system, turning biometry into a distinguished field of study. His invention, named Bertillonage, was based on a set of records of body measurements, physical descriptions and photographs. Although this system was difficult to use the Bertillonage was an important advance in criminal and people identification. The increased use of the system allowed us to observe that many people shared the same anthropological measurements, leading to entitling this system as inaccurate.

- In 1880 an article written by Henry Faulds and William James describing the uniqueness and permanence of fingerprints was published in the British scientific journal ”Nature”.

- Sir Francis Galton in 1888, motivated by the work of Henry Faulds and William James, began to observe fingerprints as a means of identification.

- Juan Vucetich, an Argentinian Police Official, began the first fingerprint files based on the Galton pattern types. At first, Vucetich included the Bertillon System with the files.

- In 1900, The United Kingdom Home Secretary Office conducted an inquiry into ”Identification of Criminals by Measurement and Fingerprints”, where Mr. Edward Richard Henry appeared before a committee to explains the fingerprint system he developed and later published in ”The Classification and Use of Fingerprints”. The committee recommended the adoption of fingerprinting as a replacement for the relatively inaccurate Bertillon system of anthropometric measurement, which only partially relied on fingerprints for identification.

- Between 1905 and 1908, all branches of the North American forces adopted the use of fingerprinting systems.

- The nascent field experienced an explosion of activity in the 1990s and began to surface

in everyday applications in the early 2000s

Since then, companies and governments have invested in the evolution of biometrics to solve many issues related to social, military and public safety. Unlike standard systems of the past based on only one characteristic (e.g. fingerprint), systems now a days have become more complete and varied. A more current system can be composed of one or more traits depending on the purpose for which it is intended. A key to the development aspects of these systems is the constant evolution of computational power, allowing the expansion of this science in almost every type of technological equipment. Adding an important factor for portability and ease of acquisition for anyone interested in this area.

2.1.1 Standard Procedure

Biometric systems in their most basic form can be structured in four sequential phases as illustrated in fig. 2.1.

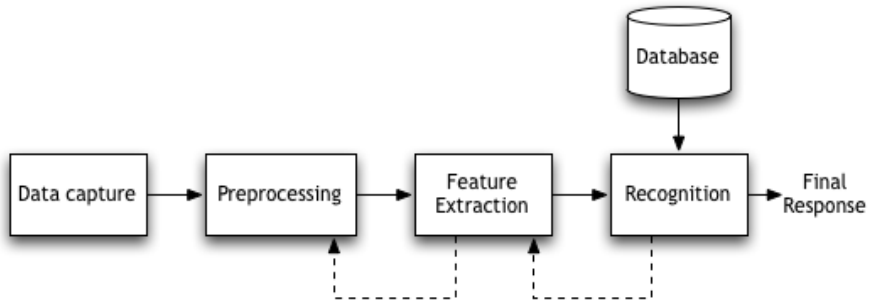


Figure 2.1: Biometrics standard procedure.

The first one is Data Capture, which is the interface between the real world and the system itself. The information captured can have several digital representations depending on the desired purpose. It can be in 1-D (e.g. sound from microphone), 2-D (e.g. image from camera), 3-D (e.g. point cloud from Kinect) depending on the sensors used on acquisition.

Pre-processing encompasses all kinds of methods that allow the processing of information that has been collected, which for reasons of the system design or external factors itself, needs to be treated. Normally this stage is the heaviest in the all biometric systems, including: enhancement for improving the quality of data, detection in order to detect Regions of interest, noise reduction to isolate only the useful information or normalization which makes the data invariant and more propitious for further processing.

After treating and selecting the information, the feature extraction stage comes next. This stage is responsible for the characterization of the information previously collected and treated. At this stage, it is assumed that information noise is as low as possible and therefore it is possible to make a characterization using a set of measures (features representation). In this stage, particular care is taken regarding invariance of the measures used for possible transformations (translation, scale, rotation changes, projective distortions, deformations in portions of the data), which adds a greater dynamism and strength to the system performance.

Lastly, the recognition stage which allows a final response to be given. This response consists of finding a degree of similarity between the features derived from the previous stage and bio-

metric templates stored into the database. Depending on the problem in question, several ways might be used to achieve this type of comparison, for example, using a simple distance between features (e.g. Hamming distance) or other more complex method that covers supervised learning (e.g. Neural Network). Sometimes the combination of several methods is also possible, taking maximum advantage of each.

2.1.2 Operating modes

Although biometric systems share the same basic procedures, these have different functionalities depending on the intended purpose. The systems can be characterized in three distinct modes: enrolment, verification and identification [5].

At the enrolment mode (fig. 2.2) the system can establish a digital representation between a person's identity and the corresponding biometric sample. This relation (template) is stored apart from the system (e.g. Database) and used whenever a verification or identification is requested.

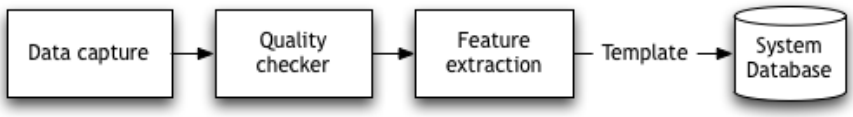


Figure 2.2: Enrollment mode.

In verification mode (fig. 2.3) the person's identity is validated based on the comparison made between newly acquired biometric data and a previously collected biometric template. In this way of operating it is assumed that the subject has in his possession some kind of secret information (e.g. Pin, password).The utilization of biometric data is used to verify that the secret really belongs to the person who submitted it. This approach aims to prevent that secrets can be stolen and presented by another person than the true owner. This type of approach gives the name of positive recognition.

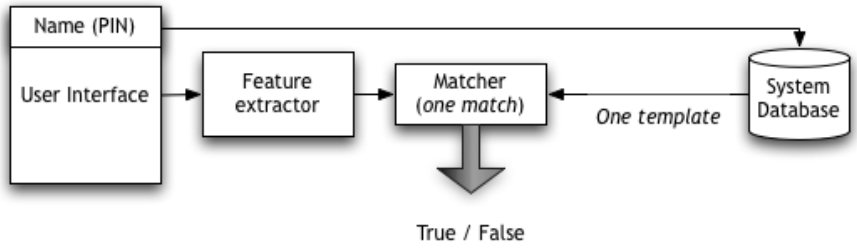


Figure 2.3: Verification mode.

On the other hand, a system that builds on a model of identification (fig. 2.4) intends to conclude what the identity of the person is without intentionally presenting itself to the system. The identification is usually done by negative recognition, although there are systems that work in positive recognition. The main difference between these two is the search of the person in question, taking into account a greater similarity grade in the positive recognition and a lower one in the negative recognition. The identification mode presupposes the comparison of collected biometric data in comparison with all the templates previously stored, in way so as to find a positive or negative relation of the person in the data base. This model is the foundation

to the creation of less cooperative biometric systems.

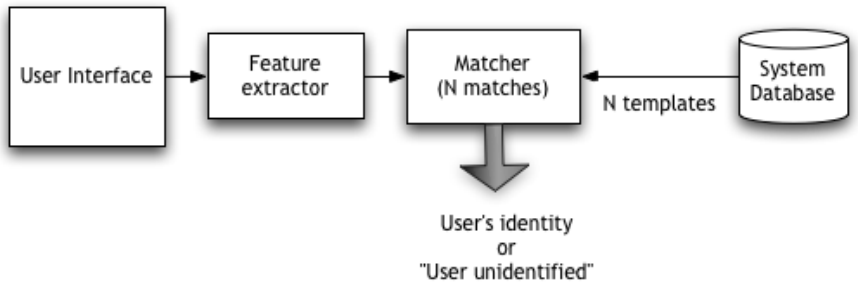


Figure 2.4: Identification mode.

2.1.3 Single Biometrics

A biometric system can be based on a wide range of physiological and behavioural traits. Each one has its advantages and disadvantages, and it can not be affirmed that a feature is better than any of the others. Although some of the traits are more valued by the scientific community, the choice depends largely on the purpose and conditions imposed on the system to be created.

2.1.3.1 Main traits

There is a fairly wide range of traits that can be applied to biometrics. Within this list of possibilities there are some that stand out more because of their use in large-scale systems or because they have major focuses at the scientific research level. The group of possibilities can be subdivided into two main groups; physical traits (e.g. DNA, Ear, Face, Facial Themogram, Finger Geometry, Fingerprint, Hand geometry, hand vein, Iris, Periocular, Palmprint, Retina) and behavioural traits (e.g. Gait, Keystroke, Signature, Voice).

- DNA - Deoxyribonucleic acid (DNA) is a molecule that contains information required for the development and functioning of all living beings. The pairs found in these chains are translated into a one-dimensional code that can be used for purposes of verification or identification. This characteristic is widely used in cases of mass accidents and solving criminal cases since it only requires a small proportion to obtain all the information needed. Although it has a high performance and easy acquisition, it is very propitious to contamination, very expensive and does not allow real-time processing. Its performance may decrease in situations of close Relatives (e.g. Identical twins).

- Ear - Recognition through the structure of the ear is still a topic that leads to a lot of research, taking into account that a conclusion has not yet been reached about its uniqueness. However it remains promising, and is currently seen as a complementary feature to other systems with more developed features. For example, by using information from the ears into facial recognition systems, it takes advantage of the same device to capture information from both traits.

- Face - The same way that humans use the face to subconsciously do recognition, the face is also widely used in biometric systems. Despite the accuracy being low in comparison to

other characteristics (e.g. iris, fingerprint) the acquisition of this source of information does not need advanced equipment (e.g. recognition through webcam or video vigilance camera) and allows for recognition in real time. However with the appearance of less controlled environments, this gives rise to an increase in computational weight and a decrease in performance.

- Facial Thermogram - The facial thermogram is very similar to facial recognition with the exception of using thermal cameras to capture the faces patterns. The heat patterns are due to the branching of existing blood vessels on the face, which are highly distinctive even between very similar people (e.g. twins). This trait is especially advantageous because it works in total darkness and with the natural light. Although it requires a lower degree of cooperation and operates in real time, the system itself is very expensive and the efficacy is not guaranteed near heat surfaces.

- Finger Geometry - Finger Geometry includes a set of physical metrics as a whole form of biometric trait. Based on the length, height, thickness and curvature of the fingers, it is possible to get some performance in recognition processes. It is an easy technique and not invasive, however it requires some level of cooperation as it requires the placement of the fingers at a default position. There are some high costs due to the equipment required for the acquisition.

- Fingerprint - The fingerprint can be considered the most common way to identification taking into account its modern history in biometric systems. Its uniqueness lies in the arrangement of ridges and valleys on the surface of the finger. Although it is a fairly mature trait, easy to use and with little cost, it is highly dependent on the cooperation of the subject and it is very sensitive. It does not work with small lesions or dirt in the finger and there are also certain people who do not have the needed characteristics to use the system (e.g. surgeon as they often wash their hands with strong detergents, builders, people with special skin conditions).

- Gait - The way we displace in space is increasingly taken into account. From a sequential set of images and analysis of time-series features, stride length, cadence and speed or silhouette it is possible to gather unique information in order to distinguish one subject from another. The low level of cooperation required makes this trait quite popular, although it is still sensitive to variations in respect of the surface where the person moves and the size of the person, which limits the variations of motion.

- Hand geometry - Consists of the use of metrics related to height and width of the back of the hand. The curves, thickness, length, weight of the fingers as well as the distance between the articulations and the bone structure can contain useful information. The fact of having lower values of discriminability makes this kind of system ideal for using on a restricted group of people and in cases where the need for precision is not one of the main criteria to the system selection (e.g. access to determinate areas in a factory by workers). This trait is highly dependent on cooperation, taking into account that the placement of the hand on the scanner surface is one of the sensitive points of this system.

- Hand Vein - This trait is based on the use of infrared light to identify the pattern formed

by the veins and blood vessels of the hand. Traditionally, the de-oxygenated haemoglobin appears as the black patterns in the captured image, while the hand or fingers have lighter patterns. It is believed that this pattern is unique among twins and even between the right and left hands of the same person. However this needs to be corroborated with the proper acquisition of information.

- Iris - The patterns of the iris develop during the fetal period and extend to approximately the age of two. The information from the iris is unmatchable between real twins and invariant over time. Furthermore, it is believed that it is very difficult to be surgically modified or cloned without being identified as false. Despite the major point in its favour, it still requires some level of cooperation, although it has worked to counteract this tendency.

- Periocular - The use of the periocular region is a technique emerging in biometrics and it is considered a middle term between facial recognition and the iris independently. The gathering of information of the regions that surround the eyes (eyelids, eyelashes, nearby skin area and eyebrows) is an alternative when it comes to systems where a demand of cooperation is at a lower level with an acceptable level of precision. In relation to the face, the information that comes from this region is not affected by ageing but when compared to the iris, the periocular has a lower level of precision.

- Keystroke - The way people write on keyboards may contain crucial information for the purpose of continuous authentication. The objective of this kind of system is to capture the existing time between pressing multiple keys in a way to obtain a unique behaviour pattern. Although this method allows for continuous authentication, it raises serious problems of privacy.

- Palmprint - In similar ways to the fingerprint, the palm print contains ridges, principal lines and wrinkles that can serve as a person's characteristic. Taking into account that the palm of the hand is a much bigger region than the fingerprint, it is expected that there is a large information volume that can be used in relation to the fingerprint. On the other hand, bigger scanners are required which raises the costs associated with this system.

- Retina - The pattern of veins beneath the back of the eyeball is called Retina. It is unique to each person and believed to be one of the safest biometric traits. On the other hand, the fact that it is a small internal region requires the use of very expensive equipment, and a high level of cooperation.

- Signature - Perhaps one of the most used features over time, the signature comprises a series of behavioural characteristics during the writing process and makes it unique. Normally your assessment can be done in a static way, where only geometric features are used or dynamically where the speed, acceleration and trajectory of the signature are also analysed. Despite having an acceptable level of performance in authentication processes, the signature is susceptible to physical and emotional variants.

- Voice - The voice is a mixture of physical and behavioural characteristics. For standard voice recognition two categories are created, text-dependent and text-independent.

Text-dependent recognition is based on reading from a pre-defined phrase. In text-independent category, there are no limitations as to what must be said, which makes the system more complex but provides more reliability. This type of system is not suitable on a large scale, since it is very sensitive to variations (e.g. room acoustics, misspoken Individuals emotional states or phrases), which considerably reduces its performance.

2.1.3.2 Traits performance

In summary, the existing traits and the corresponding systems can be classified according to some meaningful criteria. Jain et al. [6] propose to evaluate each trait having regard to its Universality, Distinctiveness, Permanence, Collectability, Performance, Acceptability, Circumvention.

Universality - Measures the amount of people where the trait is manifested.

Distinctiveness - Seeks to quantify how different a trait is between two people.

Permanence - Evaluates whether a certain feature remains unchanged over time.

Collectability - Establishes how easy it is to acquire relevant information for a particular trait.

Performance - Quantifies the precision and speed that a system can achieve.

Acceptability Establishes the level of social impact in the use of a system.

Circumvention - Measures the degree of reliability of a system against possible attacks.

2.1.4 Multimodal Biometric Systems

Given that each trait has strengths and weaknesses, the choice of the multimodal biometric system attempts to combine the properties of two or more traits in an attempt to attain even better performance. This approach can bring benefits to the level of security, information quality, universality, and better performance in terms of matching.

The level of security and the complexity of the system increases as the number of traits is used. If a feature is difficult to counterfeit, with two or more traits, the difficulty is increased.

Increased quality of information encompasses the scenario in which information from a source is compromised or for any reason is not enough. When using more than one source of information, this scenario can be compensated using the information from another source that was captured correctly. For example when a voice recognition system cannot process this information due to excessive noise, a fingerprint sample can be used. This compensation will reduce the chances of system malfunction.

Universality comprises if a person is not in a physical or emotional condition so that the system can capture a particular trait.

In multimodal systems, merging the characteristics of different origins can be done in two main forms, before and after the matching module of the system. When fusion takes place before the matching it can be at the sensor or feature levels. After this it can be at the score, rank or decision levels [7].

Pre-Matching Fusion

Sensor level It is applicable when there is a combination of information from various sensors.

Feature level fusion at this level can occur in two ways, when the merger is made with features of the same modality (e.g. instantaneous and transitional information for speaker recognition) and when it is performed based on features of different modalities. Bearing in mind that at this level the collected information is still raw, it is common to assume that this approach would be more efficient in comparison with other types of post-matching fusion. This, despite the fact that the literature is based on the post-matching methods. Fusion at this level is a difficult task, since some types of features may be incompatible and the fusion of several features results in the appearance of multi-dimensional vectors that involve more processing, more storage capacity and greater complexity in the matching process. To solve this problem, methods of transformation and selection for features can be applied.

Post-Matching Fusion

Score level By using different matchers that quantity the similarity or difference between the data acquired and stored in the database, it is possible to combine the different outputs. After performing the normalization, it is possible to apply some mathematical operations: Sum, Product, Min, Max and Median rules .

Rank level Is most widely used in identification systems and is based on the output matching characteristic of this type of system (rank list). By combining the different outputs from each list, is possible to reorder a final list.[8].

Decision level Given that in the verification systems each classifier gives a positive or negative answer. Or at the case of identification systems, where the output is an ordered list in which the person who most resembles is at the top. A final answer based on the response that has the most votes among all the classifiers can be obtained. This approach has the advantage of not requiring normalization as required by the score level fusion [8].

2.1.5 Performance Measures

The evaluation of a biometric system is very important to define the strengths and weaknesses of the system. By quantifying the performance against certain scenarios, it is possible to compare systems that are based on the same characteristics but which have been designed differently. The comparison is possible only if the same biometric data and the same metrics of performance

is used.

The evaluation of the system performance does not translate in the calculation of one metric in particular, but of a set of values that define the performance of parts of the system. The most common evaluation metrics for each phase are; data capture (Failure to Acquire Rate (FAR)), enrollment (Failure to Enroll Rate (FER)) and matching (False acceptance rate (FAR) and False Rejection Rate (FRR)). Only with a general analysis of all metrics, is it possible to obtain a coherent evaluation of a biometric system. Be noted that in biometrics; *Genuine* match scores denote the match scores computed between samples from the same identity and the *Impostor* match scores denote the match scores computed between two different identities.

The False Acceptance Rate (FAR) is the probability that the system incorrectly performs a correspondence between an input pattern and a non-corresponding database template (e.g. confuse two identities). This metric is calculated based on the total number of times that an impostor has been recognized (NSRI),and the total number of attempts, by the impostor, to be recognized (NARI).

$$FAR = \frac{NSRI}{NARI} \quad (2.1)$$

On the other hand, the False Rejection Rate (FRR) is the probability that the system does not recognise the identity of a user who owns a template in the database. This metric is calculated based on the number of failed attempts at recognition by a user known to the system (NFARAU) and the total number of attempts to recognize a known user (NRAU).

$$FRR = \frac{NFARAU}{NRAU} \quad (2.2)$$

The Crossover Error Rate (CER) is the rate where both accept and reject error rates are equal. Based on the values of FAR and FRR, it is possible to establish a threshold that allows to determinate the proximity that should exist between a template of the database and an input of the system. We call to this threshold the sensitivity (fig. 2.5). With the reduction of this value the system will be subjected to the occurrence of more false accept errors (higher FAR) and less false reject errors (lower FRR), consequently to increase the limit that will lead to lower FAR and higher FRR.

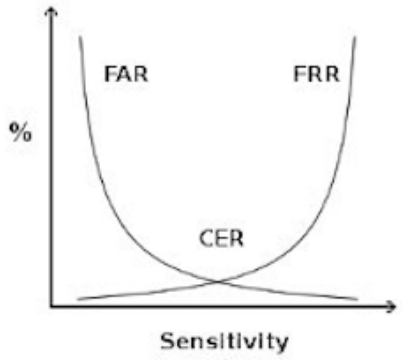


Figure 2.5: Crossover Error Rate.

Once more and, based on the values of FAR and FRR it is possible to characterize the variation of the relationship that exists between the two metrics. The area under curve from this type of representation illustrated at fig. 2.6 can be seen as a way of summarizing the system performance.

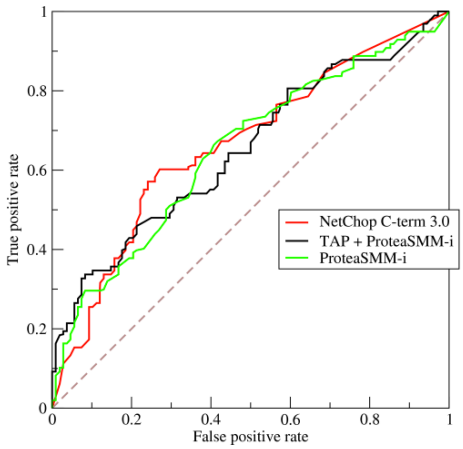


Figure 2.6: Receiver operating characteristic (Roc) curve.

The Failure to acquire (FTA) represents the probability of a system to fail in the detection and acquisition of a biometric input without the manifestation of external factors that can prevent the acquisition of the information in the right way. This metric is obtained by the number of attempts of acquiring data (NAC) and in the number of failed attempts in the data acquisition (NFCA).

$$FTA = \frac{NFCA}{NAC} \quad (2.3)$$

The Failure to enrol rate (FER) is the percentage of cases where the system failed the storage of biometric data. This metric is obtained from the total number of attempts at keeping the contact information of a user (NEA) and the number of failed attempts to carry out enrolment (NFEA).

$$FER = \frac{NFEA}{NEA} \quad (2.4)$$

In order to assess the degree of efficiency that a system has in terms of classification of inter and intra-class comparisons,we can calculate the degree of dissimilarity eq. (2.5). Between the scores obtained by the matching of features from the same person with different people, it is possible to obtain two distributions that respectively characterize the intra-class and inter-class comparisons. The distance calculation between the two distributions obtained, decidability, can serve as a performance measure.

$$d' = \frac{|\mu_{inter} - \mu_{intra}|}{\sqrt{\frac{\sigma^2_{inter} + \sigma^2_{intra}}{2}}}, \quad (2.5)$$

where μ_{inter} and μ_{intra} is the average of values of the scores obtained between inter-and intra-class comparisons and σ_{inter} e σ_{intra} are the respective standard deviations.

More directed to the evaluation of identification systems, where the output of the system is represented by a rank ordered comparisons, it is possible to analyse the performance through Cumulative Match Characteristic (CMC) curves representing the probability of being the correct matching between the observed *Top-K* first elements of the result list.

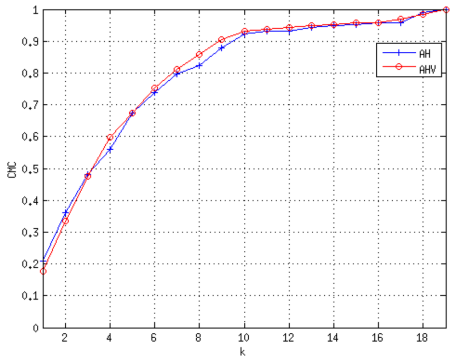


Figure 2.7: Cumulative Match Characteristic (CMC) curve.

2.2 Iris Recognition

In this section some of the most relevant methods from the literature with regard to biometric recognition through the iris is presented. A brief explanation of each step that make up this process will be done as well as the explanation of the methods that stand out in each one.

2.2.1 Typical Iris Recognition Stages

As a general rule and as in any other biometric trait used, iris recognition is also based on four steps previously explained in subsection 2.1.1 (Data capture, Pre-processing, Feature extraction and Recognition). But each step of the process comprises sub-methods focused only on iris

recognition in particular. Most of the published iris recognition methods can be illustrated according to fig. 2.8.

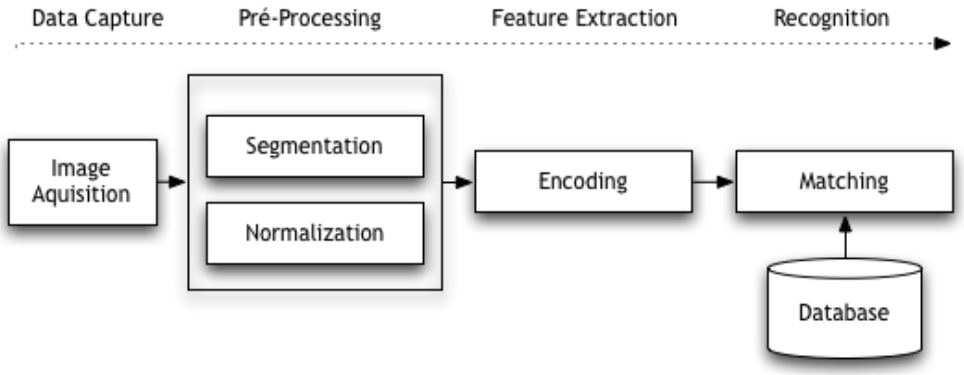


Figure 2.8: Standard Iris Recognition stages.

Data Capture

Image Acquisition Due to the characteristics of the iris, the acquisition is a task very sensitive to external factors (e.g. light variations, cooperation). Thus, the selection of sensors for proper acquisition is still a topic widely studied. Beside the difficulties intrinsic to the iris, biometrics has evolved to become less cooperative, witch makes even more complex to define an optimal solution for this first stage of the system.

Pre-Processing

Segmentation This step aims to estimate the internal (pupillary) and outer (scleric) boundaries of the iris essential to generate a biometric iris sample containing only solid information, that is subsequently used in the matching phase. In biometric systems this step is usually associated with a large computational load, which leads to a high demand for new ways to counteract this trend.

Normalization The normalization of the segmented region ensures a level of invariance against certain factors (e.g. iris size, position and rotation). The security of this type of invariance is essential for encoding and matching steps that follow.

Feature Extraction

Encoding The encoding step includes all kinds of methods to extract the most relevant features of the iris, creating a new representation of the same. These representations are used in the matching phase.

Recogntion

Matching The last step consists of comparing two iris signatures in order to conclude whether they are form the same person. The way the comparison is done may vary from a simple calculation of the distance between signatures or in more complex cases, the use of machine learning algorithms. In any case, the output is a measure of a similarity representing the degree of likeness between two biometric individuals.

2.2.2 Iris Image acquisition

The way information is acquired from the iris in uncooperative environments is still one of the most controversial issues in the iris recognition domain. The state of the art is divided between the use of sensors with the capability of acquiring iris images in the visible spectrum or in the near-infra-red (NIR) spectrum. This also includes cases where the use of both may be used, in particular for carrying out some specific steps (e.g. detection on visible spectrum and iris acquisition on NIR).

Matey and Kennell [9] studied the problems usually encountered when images are acquired from over one meter. Difficulties related to wavelength of light used, the type of light source, the amount of light reflected by the iris back to the sensor, the required characteristics of the lens, signal-to-noise ratio, eye safety, and image quality, capture volume, residence time, and sensitivity to subject motion were discussed. These authors applied the study of these points into marketable iris recognition systems dating between 1955 and 2008.

Boddeti and Kumar [10] studied the influence of wavefront coding in different stages of iris recognition (e.g. segmentation, feature extraction). Although this aspect had already been previously analysed in the literature, the author used a large database in their tests.

Wheeler et al. [11], Dong et al. [12] He et al. [13] propose to use iris images captured with NIR light. Wheeler et al. presents a system composed of two wide-field-of-view cameras to detect the face of the subject. A third camera and NIR illuminator is used to acquire iris images taking into account the location of the subject previously established. Dong et al. proposes a system for acquiring images of the iris in NIR able to operate at a distance of 3 m. He et al. in order to reduce the cost associated with the collection of high quality material for the iris images , designed a camera that consists of a CCD sensor, a custom glass with a fixed focus, one-pass filters NIR and an illuminator comprised of several NIR LEDS.

Although most systems presently on the market resort to NIR for acquiring images of the iris. Proença [14], argued that the use of sensors operating in the visible spectrum are more appropriate against factors such as distance and movement.

2.2.3 Segmentation

For segmenting the iris, two kinds of approaches can be considered. The first set is the more classical seen in restricted acquisition environment, while the second set is directed more to less cooperative environments.

Within the set of more classical methods, we have Daugman[15] which proposes the use of a integrodifrencial operator and Wildes [16], Mat et al [17], Tisse et all [18], Monro et al [19] who use the Hough transform. In the second family, we have the use of fuzzy algorithm (Proença and Alexander [20]), elastic model based on the Hooke's law (Zhaofeng He et al. [21]), concepts

of Grow-cut algorithm (Tan and Kumar [22]), Generalized Structure Tensor algorithm (Alonso-Fernandez and Bigun [23]) and spatial arrangement of boundary points (Xinyu et al. [24]).

2.2.3.1 Daugman's Integro-differential

In 1993, John Daugman [25] presented one of the most important methods in regards to iris recognition and what would be the first method implemented in functional biometric recognition systems. The author assumes that both the iris and the pupil have circular shapes. In this way the author uses an integrodifferential operator to search within the image domain, the boundaries of iris.

$$\max_{r,x_0,y_0} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right|, \quad (2.6)$$

where $I(x,y)$ is the input image, $G_\sigma(r)$ is a smoothing function such as a Gaussian of scale σ and $*$ denotes convolution. The variables x_0, y_0 and r, represent the center coordinates and the radius.

This method iteratively searches a space N^3 by circles with radius (r) and centre (x_0, y_0), where the response from the operator in eq. (2.6) is larger than a predetermined threshold value. This way, the author manages to approach the iris segmentation to find two circles representing the iris and pupil boundaries. A similar approach is used to find the eyelid boundaries with a small change, which consists of changing the contour integration form circular to arcuate. Later on, the author proposes some improvements to the method allowing for flexible shapes [26].

2.2.3.2 Hough's Transform

The Hough transform is an algorithm what is used to determine parameters of simple geometric objects, such as lines or circles in images. The circular Hough transform is often used to deduce the radius and the coordinates of the boundaries of the iris and pupil, assuming that the boundaries of iris resemble circles. Wildes [16], Ma et al. [17], Tisse et al. [18] and Monro et al. [19] apply the circular Hough transform.

Wildes [16], proposes the creation of a binary map from the detection of the boundary points in the captured image. This first step is achieved after establishing a threshold value in the magnitude of the gradient intensity calculated from the eq. (2.7) where $\bigtriangledown \equiv (\partial/\partial_x, \partial/\partial_y)$ and $G(x,y)$ is the 2D Gaussian filter centred in (x,y), defined by eq. (2.8). Once completed this step the Hough transform is applied.

$$|\bigtriangledown G(x,y) * I(x,y)| \quad (2.7)$$

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \quad (2.8)$$

Ma et al. [17] proposes an approach comprised of three phases. In a first phase an approximation of the pupil center is done based on minimum value found in two dimensional projections of the image (horizontal and vertical). Then it provides a region of 120x120 centred on (x_p, y_p), what is then passed to a binary representation based on a threshold value found by the histogram of the gray scale region. Finally, using the Canny edge detector and the Hough transform, the location of the circular border is found, together with the information from the pupil centre.

Monro et al. [19] suggests the use of a heuristic method to effect a reduction of the region of interest. Only then do the authors propose to use the Hough transform after using a Canny edge detector to reduce noise levels in the image.

2.2.3.3 Proença and Alexandre Method

Proença and Alexandre [20] surveyed the literature and found that the existing iris segmentation methods were based on essentially two ways of approaching the issue. The first using edge detection methods for the construction of maps of areas and the second by the analysis of the intensities varying radius of consecutive circles. Assuring that both kind of approaches are very susceptible to variations in the properties of the image acquired (e.g. brightness and contrast) and the existence of noise factors (e.g. reflections, eyelids or eyelashes and iris occlusions), these authors propose a new method illustrated in fig. 2.9, while ensuring the operability in less cooperative environments.

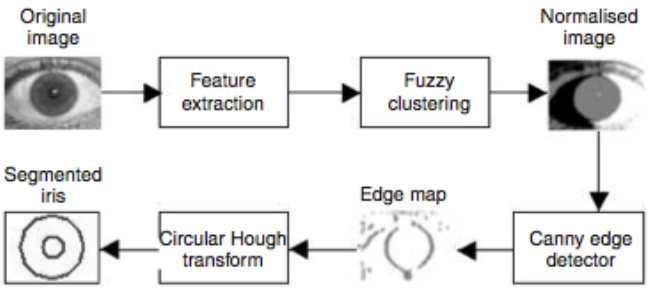


Figure 2.9: Main steps of the method proposed by Proença and Alexandre.

From a feature extraction phase in the region of interest and the application of a fuzzy clustering algorithm [27], a normalized image is obtained. Based on this new representation of the original image, the authors apply one edge detector and the Hough transform in order to estimate the targeted region of the iris.

2.2.3.4 Zhaofeng He et al. Method

Zhaofeng He et al. [21] found that the segmentation methods of the iris usually involve an exhaustive search in a large space of parameters. The authors present a new method with special attention to its computational speed, without bleaching the robustness required at this stage of iris recognition. The method can be presented with a set of sequential steps, as illustrated in fig. 2.10 and explained below.

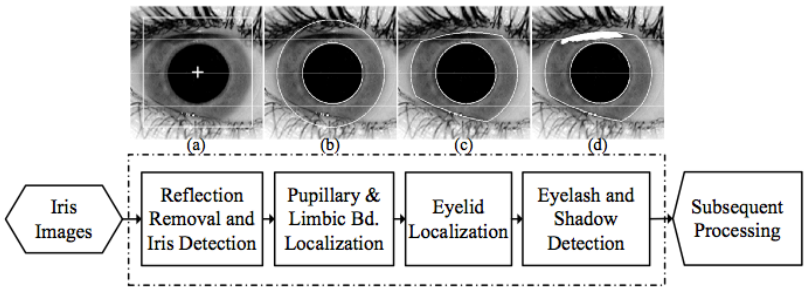


Figure 2.10: Illustrative figure of the various steps of Zhaofeng He et al. method.

1. This method starts by using the Haar-cascade detector proposed by Viola and Jones [28] for the detection of the iris. This method is widely used due to its speed of execution and the fact that the irises of different people share the same structure. However, the direct use of the images on the detector (learning and detection) is subject to failures resulting from specular reflection in the iris, witch is a problem often encountered when working with NIR illuminators. Thus, the authors propose to use a Specular Reflection Removal method prior to detection of the iris. From this step, it is possible to establish an approximate position of the pupil centre.

2. Detecting the boundaries of the iris using a new approach called pulling and pushing. Based on Hooke's law, the center and the radius of the circle of the iris are adjusted iteratively until reaching a point of equilibrium, preferably up to fit the true boundary of the iris. Being aware of the existence of a non-circular iris, the authors use a smoothing spline on the boundary points previously found. In this way, a real segmentation of iris is obtained and not just a circle that approximates the boundaries of iris.

3. After the inner and outer boundaries of the iris are set, the location of the eyelids is done. This step begins by eliminating the noise associated with the eyelashes through a rank filter. Later, the eyelid edge points are found through an edge detector. Based on the edge points found, a histogram is constructed in relation to its location in the image. Finally, an approximation of the true position of the eyelids is done based on the intersection between the edge points with three eyelid models established manually.

4. The last step consists in the detection of eyelashes and shadows via a learned prediction model. This model consists of an adaptive threshold by analysing the distribution of intensities in different regions of the iris.

2.2.3.5 Tan and Kumar (2012)

Tan and Kumar [22] propose a segmentation method of the iris with special attention to images acquired from large distances. Their method is based on the concept of Grow-cut, where the aim is to distinguish between the foreground (iris) and background (non- iris) in the input image. The results obtained in this first phase are later refined through set operations that allow to estimate the iris centre, inner and outer boundary, eyelashes, shadows and eyelids.

2.2.3.6 Alonso-Fernandez and Bigun (2012)

Alonso-Fernandez and Bigun [23] present a new algorithm to segment the iris based on a Generalized Structure Tensor algorithm (GST). In mathematics, a tensor structure (second- moment matrix) is a matrix that summarizes the predominant directions of gradient in the neighbourhood of a given point. After making the convolution of the region of interest with circular filters, the authors resort to the Generalized Structure Tensor, managing to obtain information about the magnitude and direction of each pixel edge found. This approach gives a level of additional information that significantly improves the detection of boundary points.

2.2.3.7 Xinyu et al. (2012)

Xinyu et al. [24] work with data acquired under less cooperative conditions, and propose the use of an algorithm able to work with images of different resolutions. The author first suggests finding the boundary points in the image by using the Canny edge detector. The non-connected components resulting from this previous step are considered nodes of a graph. Finally, based on the layout of the nodes in the graph, a structure that resembles the iris border more accurately is searched.

2.2.4 Normalization

Given the need of making the iris information invariant to scale and pupillary dilatation, the normalization of the segmented region is done.

Regarding the methods of literature addressing this question, the Daugman's rubber-sheet model [25] is the one with greater emphasis. The rubber sheet model assigns to each pixel of the image a couple of real coordinates in a polar domain (r, θ) in which r represents the radius and can have values between $[0, 1]$ and θ is the angle between $[0, 2\pi]$. The transition of the iris representation in Cartesian coordinates to polar coordinates translates into represented in eq. (2.9).

$$I(x(r,\theta), y(r,\theta)) \rightarrow I(r,\theta), \quad \textbf{(2.9)}$$

where $x(r, \theta)$ and $y(r, \theta)$ are defined as linear combinations, $(x_i(\theta), y_i(\theta))$ the set of pupillary boundary points and $(x_j(\theta), y_j(\theta))$ the set of limbus boundary points along the outer perimeter

of the iris.

$$x(r,\theta) = (1-r)x_i(\theta) + rx_j(\theta) \quad (2.10)$$
$$y(r,\theta) = (1-r)y_i(\theta) + ry_j(\theta), \quad (2.11)$$

where each point of the new representation is obtained using the following set of equations:

$$x_i(\theta) = x_{i0}(\theta) + r_i cos(\theta) \quad (2.12)$$
$$y_i(\theta) = y_{i0}(\theta) + r_i sin(\theta) \quad (2.13)$$
$$x_j(\theta) = x_{j0}(\theta) + r_j cos(\theta) \quad (2.14)$$
$$y_j(\theta) = y_{j0}(\theta) + r_j cos(\theta). \quad (2.15)$$

2.2.5 Encoding and Matching

As mentioned above (section 2.2.1), the encoding step is essential because it allows the coding of the most important features (pattern information) of the iris, so that these can be efficiently stored and used as templates for comparison in biometric systems. Regarding this step, the most classic literature addresses the question through an analysis of the iris patterns themselves. They are statistical approaches and are divided into three categories: phase-based methods (eg, Daugman [15]), zero-crossing methods (eg, Boles and Boashash [29] and Roche et al. [30]) and texture analysis based methods (eg, Wildes [16] Kim et al. [31] and Ma et al. [32]). Another approach (syntatic) focuses the existing structural information on iris pattern (Proença [33]).

2.2.5.1 Daugman's Iris-code

Daugman [15] proposes the coding standards of the iris through the convulsion of the segmented and normalized region with a 2D Gabor filter, defined by eq. (2.16).

$$h\{Re,Im\} = sgn_{\{Re,Im\}} \int_{\rho}\int_{\phi} I(\rho,\phi) e^{-i\omega(\theta_0-\phi)} . e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_{0-\phi})^2/\beta^2} \rho d\rho d\phi, \quad (2.16)$$

where $h\{Re,Im\}$ is a complex value in which the real and imaginary parts have value of 0 and 1 depending on the sign of the integral; $I(\rho,\phi)$ is the input image, normalized to the polar space as referred in section 2.2.4; α and β are the multi-scale 2D wavelet size parameters; ω is wavelet frequency; and (r_0,θ_0) represent the polar coordinates of each region of the iris for which the phasor coordinates $h\{Re,Im\}$ are computed .

The author proposes a new representation of the iris based on the quantification of the respective information from the response phase convulsion. The information from the amplitude is

discarded given that it is sensitive in terms of contrast, lighting and camera gain. In this way, each pixel of the normalized region takes on a binary representation.

In terms of matching, the author proposes the use of the Hamming distance (HD) eq. (2.17) to quantify the degree of similarity between two signatures.

$$HD = \frac{\|(codeA \otimes codeB) \bigcap maskA \bigcap maskB\|}{\|maskA \bigcap maskB\|}, \quad (2.17)$$

where the two binary iris signatures are represented by $codeA, codeB$ and $maskA, maskB$ are their masks. The application of the exclusive OR operator (XOR) between both signatures reflects the total of discordant bits, while the AND operator ensures that the information is not affected by noise derived from eyelashes, eyelids, specular reflections, or other noise.

2.2.5.2 Wilde's Pyramid

Whildes [16] suggests that after the segmentation of the iris corresponding region, the step of matching based on four steps: alignment, representation, comparison and decision is performed immediately.

1. Alignment. As a first step, the author proposes a new technique for registration, in order to compensate for possible variations associated with the rotation and scaling of images. The new images are mapped and subject to alignment with images already stored, to minimize their differences.

2. Representation. It is suggested a decomposition of the image through the application of Laplacian of Gaussian filters (LoG). The filtered image shall be represented by different scales on the Laplacian pyramid shape. This new representation allows the compression of information from the iris highlighting only the most significant characteristics.

3. Comparison. This step is based on a normalized correlation, where the aim is to quantify the degree of similarity between the different components of two irises. In this sense, the author applies this correlation model in the regions of 8x8 for each of the representations in the Laplacian pyramid. The output of this stage is based on a single value obtained through a statistical median.

4. Decision. The last step of the proposed method, intends to give a final response (accept / reject) through the Fisher linear discriminant.

2.2.5.3 Ma et al. Texture Analysis

Ma et al.[32] state that the essential information for the recognition is in the very transient patterns of the iris. Thus, authors propose the construction of a vector consisting of a set of

1-D intensity signals. Each signal is a combination of N successive horizontal lines obtained by scanning the image in order to represent local variations occurring in a horizontal direction. Then, a search is performed in the space of frequencies at different scales in order to find the local minimum and maximum. The authors hypothesized that the extremes that have been found are useful information, since they correspond to variations of the iris features in the original image. Finally, characteristics corresponding to different scales are concatenated into a feature vector. Regarding matching, the authors propose a two-step approach. Firstly, they encode the feature vector into binary where each point lies the value of 1 or -1, depending on the type of extreme point. Secondly, they calculate the degree of similarity between two binary sequences by exclusive OR operator.

2.2.5.4 Proença's Structural Pattern Analysis

Proenca [33] introduces a new method illustrated in fig. 2.11 that is based on the structural information of the iris. The author assumes that similar to what is visible on the iris pattern, the arrangement of the different regions of the iris itself is a useful property for this task. From the normalized iris image, the author proposes finding the most homogeneous regions. This step is done by searching for the minimum variation of intensities between each image pixel with its neighbour. After defining these regions, their centres are estimated to which the author gives the name of primitives. When all the points in the image are found, the author establishes a relationship between them based on the discrepancy of the value of each. This connection is represented by edges between each pair of points resulting in a representative graph of the data structure illustrated at **??**.

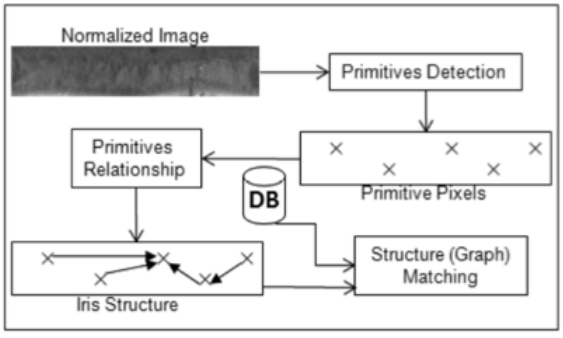


Figure 2.11: Steps that comprise the method proposed by Proença.

Considering a graph as a structural representation of the iris, the matching between two signatures is done through the calculation of the distance between graphs, resulting in a metric of similarity between them.

2.3 Periocular Recognition

This section aims to explain the methods of literature that stand out in relation to biometric recognition using the periocular region. A brief overview of the steps that constitute the process is given, even though this a quite new kind of recognition technology (initiated by Park et al. at 2009 [34]), and therefore still quite focused on the same kind of approach.

2.3.1 Standart Periocular Biometrics Stages

As well as in the case of the iris, the periocular recognition follows the standard steps characteristic of a common biometric system (Data Capture, Pre-Processing, Feature Extraction and Recognition). However, some differences lie in how some steps are addressed, especially the pre-processing and feature extraction phases (fig. 2.12). Regarding the feature extraction. it can be approached in three ways; extract information from the entire region (global descriptor), extracting information from the most relevant parts of the region (local descriptor) or using both types of extraction together.

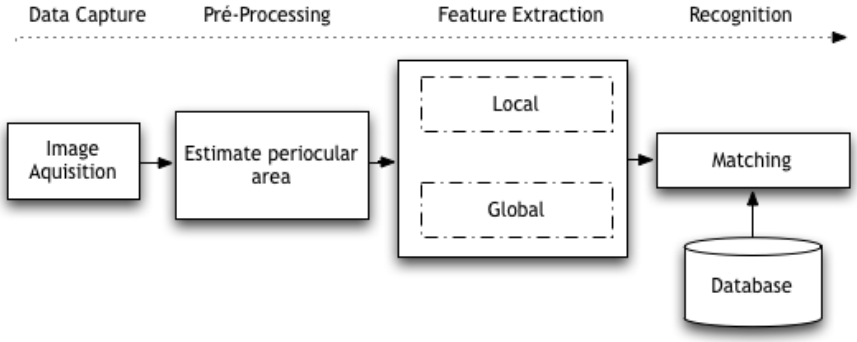


Figure 2.12: Steps on periocular recognition.

Data Capture

Image Acquisition - As well as in other biometric traits, is important to define the setup that best suits for the collection of biometric information, including whether to operate in the visible or NIR (woodard et al. [35]). However, the use of the periocular area can be considered a balance between using the face or the iris. Which means that in certain cases the capture is made in the NIR spectrum, in order to benefit iris acquisition.

Pre-Processing

Estimate the periocular area - After capturing directly the region surrounding the eyes, it is necessary to establish the exact area that will be used for feature extraction. The definition of a boundary that demarcates the periocular region is still a topic of some interest, as well the elements that should be included in the ROI (e.g. the iris, sclera and eyebrows).

Feature Extraction

Local and Global descriptors - Having established the periocular region and the elements that must be taken into account, the conditions for the extraction of biometric features which can be used for the next phase are gathered. The use of global descriptors (e.g. LBP - Local binary pattern [36] [37] and HOG - Histogram of oriented gradients [38]), local (SIFT - Scale-invariant feature transform [39]) or both, may vary depend-

ing on the type of region estimated in the previous step (e.g. Park et al.[34] by including eyebrows in this region denoted a marked improvement in the use of SIFT).

Recogntion

Matching - The final step is to find one or more metrics that indicate the degree of similarity between the collected biometric sample and the existing templates in the database, whether through the calculation of distances or for more complex algorithms.

2.3.2 Park et al.

Park et al. [34] are considered the pioneers in the use of the periocular region as a biometric trait. The proposed method uses two approaches to extract features: Local and Global, depending on whether the information is extracted from small regions or the complete region. The authors propose to estimate a region of interest (ROI) centred on the iris and divided by squares (5 by 7) in which each side is equal to the radius of the iris. For each square of the grid, LBP and HOG descriptors are computed, ensuring the extraction of information about the spatial arrangement and texture. In terms of the local descriptor, the authors propose the use of SIFT, allowing the detection of the relevant points and the description of its surroundings.

In terms of matching, global descriptors are easily comparable by calculating the Euclidean distance. While for the local descriptor, the authors propose the use of a matching distance ratio based scheme [39].

2.3.3 Miller et al.

Miller et al. [40] extracted the discriminating information from the periocular skin texture using a single descriptor: Uniform Local Binary Patterns (ULBP). The ULBP is a LBP- based method (global descriptor) with the particularity of ensuring greater invariance, especially in relation to rotation. The periocular region is estimated proportionally based on the distance between the eyes, and is set to a size of 100x 160 pixels. The estimated ROI is then subdivided, forming a grid of 7 by 4 blocks, centred on the eyes and its iris and sclera information are removed by overlaying a neutral elliptical mask. Each grid block is normalized and subsequently the ULBP is calculated. For each block a 59-bin histogram is created, representing information about the texture of that sub-region. Finally, all histograms are concatenated to form a single-dimensional array used as a biometric signature of the periocular region. For comparison of signatures (matching) the authors used the Manhattan distance.

2.3.4 Other importante works

Juefei-Xu et al. [41], Padole and Proença [42] and Jillela and Ross [43] were inspired by the work of Park et al [34], and try to find factors that influence the performance of the periocular region as a biometric trait. Hollingsworth, Bowyer and Flynn [44] investigated the elements of the periocular region to that the humans give more importance and Adams et al. [45] extended Miller et al. [40] work in terms of optimization.

Juefei Xu et al. (2010), were also inspired by the work of Park et al. [33], and analysed the performance of different feature schemes involving images acquired in the most difficult

environments. Beyond the LBP and SIFT, other descriptors were tested: Walsh masks [46], Law's masks [47], DCT [48], DWT [49], Force Fields [50], SURF [51], Gabor Filters [52] and Laplacian of Gaussian (LoG)[53].

Padole and Proen_cca [42] analysed to what extent pose variation, distance of the subject, pigmentation and occlusion can influence the performance of the recognition by periocular. The feature extraction scheme used was the same proposed by Park et al. [34]. With a single difference in how the periocular region was centred. The authors propose using eye-corners instead of iris centre considering that in less cooperative environments, this approach is more robust.

Jillela and Ross [43] , used the SIFT and LBP in the periocular region for extracting additional information.

Hollingsworth, Bowyer and Flynn [44] gathered a set of images of the periocular region acquired from 120 subjects. They deliberately concealed the iris region in all images and some specific elements of the periocular region (e.g. eyebrows) in a distributed manner. Finally, 80 pairs of images were presented against a set of 25 human observers. Where each observer answered two questions; if the pair was the same person or different persons; and what elements of the periocular led them to make the decision. The result of this research indicated that eyelashes are a feature which further helps in the recognition followed closely by the corners of the eyes and the shape of the eye. The skin was the one with less utility.

Adams et al. [45] proposes the use of a genetic algorithm (Steady-State Genetic Algorithm (SSGA)) in order to optimize the proposed method by Miller et al. [40].

2.4 Fusion of the Periocular and the Iris

The biometric recognition using the iris is severely affected in less cooperative environments or due to the existence of occlusions, motion and optical blur, poor contrast and illumination artefacts. This section discusses the idea of being able to use the periocular region as supplementary information.

2.4.1 Woodard et al.

Woodard et al. [35] studied the effect of fusion techniques on periocular and iris in non-ideal scenarios. The basis of this study involves using the method illustrated in fig. 2.13 in which images are captured in NIR. Note that the images used had difficult conditions for recognition by the iris.

Authors chose the approaches proposed by Daugman [25] to perform the steps of normalization (section 2.2.4) and encoding (section 2.2.5.1), while the segmentation was done manually. The periocular area was estimated based on manual annotation of the centre of the eyes. For the feature extraction they used only a global descriptor (LBP). After extracting features of both traits, the score-level fusion (section 2.1.4) was made. Through this study, the authors conclude that this type of fusion can be good, observing an increase in performance of rank-1 of 80% to 95.5%.

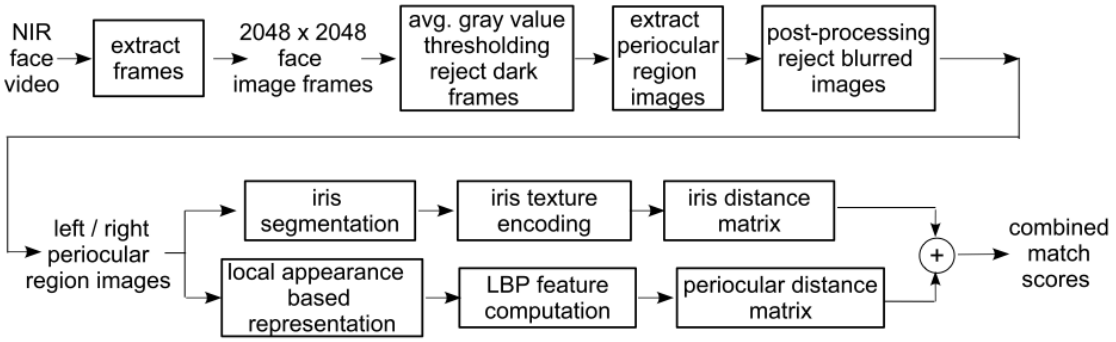


Figure 2.13: Woodard et al. steps to study a possible fusion between the iris and periocular.

2.5 Non-cooperative recognition

The level of cooperation required is an important factor in the creation of new biometric systems. In this perspective, the scientific community investigates new solutions in order to make them more robust. With respect to the iris and periocular biometric recognition fields, the effect of factors such as the distance at which the user is, and the quality of the information (e.g. spatial and motion blur) gathered from the surroundings to the system, can be minimized, particularly through the use of methods that allow the detection of objects in the image [28], analysis on the frequency domain [54] and the use of NIR illuminators to reduce the factors of the environment (e.g. variation in the natural lighting). In order to make systems work in less cooperative environments, it is expected that some of these methods are implemented in a sub-module in the stage of pre-processing, to ensure a lower level of cooperation with the same quality of performance.

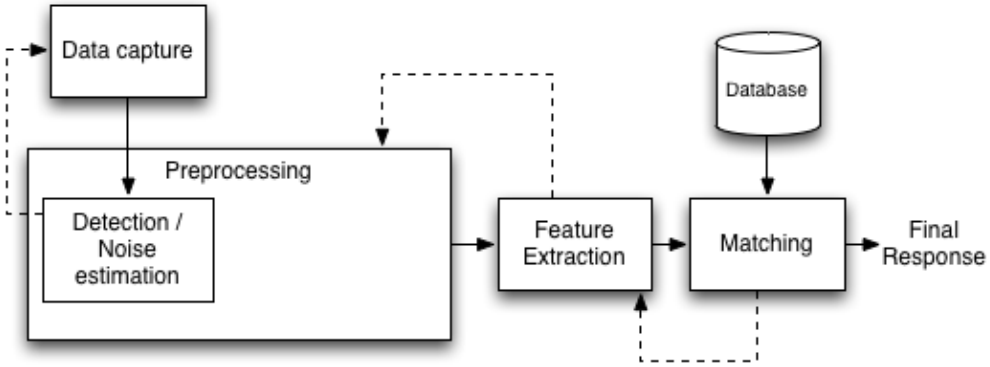


Figure 2.14: Non-cooperative system model.

Figure 2.14 represents the steps of a non-cooperative biometric system. There is a new sub-module which comprises detection and noise estimation that makes possible to ensures system robustness.

2.5.1 Region detection (Viola and Jones)

Paul Viola and Michael Jones [28] in 2001 presented one of the most used methods to detect regions in images due to its robustness and ability to operate in real time. Initially this approach emerged as a response in less cooperative environments. This method stands out mainly due to

three characteristics:

1. Uses a new type of image representation, which was given the name integral imaging.

2. Uses a variant of the AdaBoost algorithm to build the classifier.

3. Constructs a sort of weak classifiers with strong base cascade.

Based on these three points, the authors propose a method widely used in uncooperative systems.

2.5.1.1 Integral image

The integral image is an intermediate representation of the original image, which allows a greater speed in processing characteristics. In this type of image, each point (x, y) is the sum of pixels immediately above and to the left. For each pixel of the image is obtained eq. (2.18) , where $ii(x, y)$ is the integral image and $i(x', y')$ is the original image.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y}^{i} (x', y') \quad (2.18)$$

As illustrated in fig. 2.15, the sum of the intensity in rectangle D can be found based on four points. Since the value of point 1 in the integral image is the sum of the pixels inside the rectangle A, the value of point 2 is the sum of the pixels inside the rectangle A and B and the value of point 3 is the sum of the pixels in the rectangle A and C .

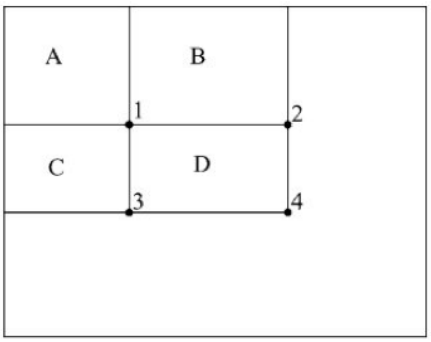


Figure 2.15: Illustration of the intensity calculation of a pixel in a integral image.

2.5.1.2 Features

The object detector classifies the images based on simple features, calculated over successive scans with windows of different sizes. They suggest three distinct types of features (fig. 2.16); two-rectangle features that represent the difference between the sum of the pixels within

two rectangular regions; three-rectangle features that compute the sum within two outside rectangles subtracted from the sum in the rectangle; and four-rectangle features compute the difference between diagonal centre pairs of rectangles. Since the image is already an integral representation and the calculation of the features translates into basic operations of addition and subtraction, extraction of features becomes much faster.

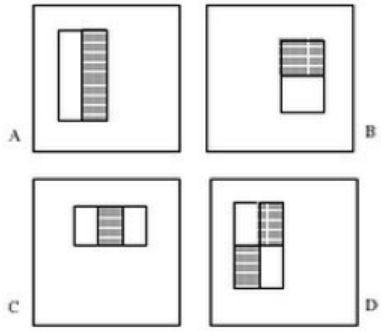


Figure 2.16: Feature types used in the object detector.

2.5.1.3 Adaboost

This method uses a variant of the machine learning algorithm called AdaBoost, which allows the construction of a classifier based on a set of weak features. The training algorithm is fed by a set of images represented by negative images (not including the object to find) and positive images (containing only the object to be detected). Then the algorithm will adapt to the shape and texture of the object that should be found and ensures certain robustness with respect to the noise that may be found under the same conditions of acquisition.

2.5.1.4 Cascade

The basic principle of the detection algorithm is to scan the same image several times, each time with a different size window. Even if the image contains more than once, the object, most of the sub-window don't contain the desired object. Authors chose a classifier cascade to discard the negative sub-windows, thus improving the speed of detection. The classifier cascade consists of several steps, each of which consists of a weak classifier that aims to determine whether a sub window is the object wanted or not. If the sub window is rejected by one of the classifiers, the process stop. If the sub-window is classified as a probable object, then it moves to the next stage of the cascade. The sub-window contains the object sought if it passes through all classifiers of the cascade without any failure.

2.5.2 Focus evaluation

The acquisition of focused images is a topic that arouses some research in less cooperative systems. Given that acquisition of unfocused images has impact on system performance and the cost associated with sensors that can automatically compensate this factor, is quite high. Daugman [15] and Byung Jun Kang and Park Ryoung [54] propose two methods to offset the impact of this factor in the system.

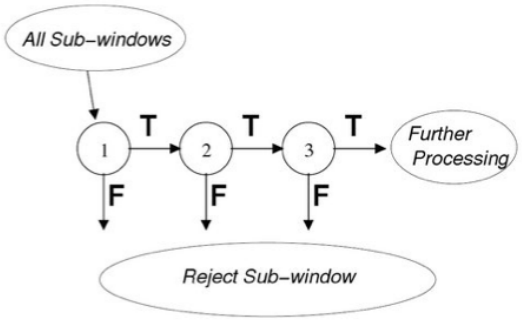


Figure 2.17: Cascade representation.

2.5.2.1 Daugman analysis

For the detection of out of focus data, Daugman [15] suggests that an analysis of high frequencies in the image domain, using $2D$ Fourier spectrum. To this end, the author proposes the convolution of each frame in real time with an 8x8 convolution kernel illustrated in fig. 2.18.

-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	+3	+3	+3	+3	-1	-1
-1	-1	+3	+3	+3	+3	-1	-1
-1	-1	+3	+3	+3	+3	-1	-1
-1	-1	+3	+3	+3	+3	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1

Figure 2.18: Daugman convolution kernel (8x8).

This convolution kernel results from the overlap of two box functions. The first has a size 8x8 pixels and an amplitude of -1, while the second has a size 4x4 pixels and amplitude of 4. By overlapping the two box functions, the central region of the filter acquires an amplitude of $+3$. The resulting filter can be represented by eq. (2.19).

$$K(u,v) = \frac{sin(u)sin(v)}{\pi^2 uv} - \frac{sin(2u)sin(2v)}{4\pi^2 uv}. \quad (2.19)$$

Having the convolution of the image with the filter, the magnitude is normalized by eq. (2.20) in order to obtain a representative score of the degree of focus between 0 and 100.

$$f(x) = \frac{100x^2}{x^2 + c^2}, \quad (2.20)$$

where x represents the value of the magnitude of the resulting convolution of the kernel (8x8) with an image and where c is a constant. The author argues that the process can be carried out in real time, thereby allowing a mechanism to be used to warn the user that should be placed in a more appropriate position.

2.5.2.2 Jun Kang and Kang Ryoung Park analysis

Jun Kang and Kang Ryoung Park [54] evaluated the approach proposed by Daugman, and concluded that the used filter would not be the best one to use in real time due, primarily to its size. They also found that the filter did not have the properties necessary to detect the finer texture of the iris. Considering that these factors are very important for the proper functioning of the least cooperative systems, the authors propose the construction of a new, smaller and refined filter. The new convolution kernel is shown in fig. 2.19 and has a size of 5x5, resulting from the superposition of a three box function. The first with a size 5x5 and amplitude of -1, other one with a size of 3x3 and amplitude of $+5$, and one of size 1x1 and amplitude of -5.

-1	-1	-1	-1	-1
-1	-1	+4	-1	-1
-1	+4	+4	+4	-1
-1	-1	+4	-1	-1
-1	-1	-1	-1	-1

Figure 2.19: Jun Kang and Kang Ryoung Park convolution kernel (5x5).

The representation of the filter can be expressed by eq. (2.21).

$$k^{'}(u,v) = \frac{sin(\frac{2}{3}u)sin(\frac{2}{3}v)}{\frac{9}{4}\pi^2 uv} - \frac{sin(\frac{5}{2}u)sin(\frac{5}{2}v)}{\frac{25}{4}\pi^2 uv} - \frac{sin(\frac{1}{2}u)sin(\frac{1}{2}v)}{\frac{1}{4}\pi^2 uv} \quad (2.21)$$

Similarly to the approach proposed by Daugman [15], the sum of the magnitudes of the resulting convolution is normalized. Again, using the eq. (2.20), the score is in a range between 0 and 100.

The authors state that compensation methods must be applied on the image if it is found that the resulting score is below a predetermined threshold. Otherwise, they admit that the image meets the conditions necessary to proceed with further processes, according to the desired end purpose.

Chapter 3

PAIRS - Periocular and Iris Recognition System

The following chapter aims to display the several modules used to build the biometric system proposed (PAIRS). This chapter begins with the global analysis of the system in order to establish a high level notion of its functioning. Each step of this procedure is fully explained throughout this chapter and the explanation is structured in order to follow the standard steps usually associated with the standard biometric systems (Image Acquisition, Pre-processing, Feature extraction, Recognition).

Figure 3.1: Periocular and Iris Recognition System.

3.1 Proposed Method Schema

In a high level perspective, the method proposed (the core of the system) follows a building line illustrated in fig. 3.2. It was conceived to take advantage of the best available methods in literature concerning the periocular region and the iris. That way, it will be possible to build a multimodal system (which uses information from two main traits: the periocular region and the iris) and a holistic system (which repeatedly extracts information from the same region).

The main purpose is to achieve an acceptable performance level in a less cooperative scenario. Therefore, during its construction, the required computational complexity and the ability to process in real time were taken into consideration.

The system presented here includes two types of fusion. The first type, which deals with the feature extraction, includes the concatenation of several descriptors in one single feature vector. The second type deals with the score, and is performed through the calculation of a final value, based on the several scores resulting from the matching phase.

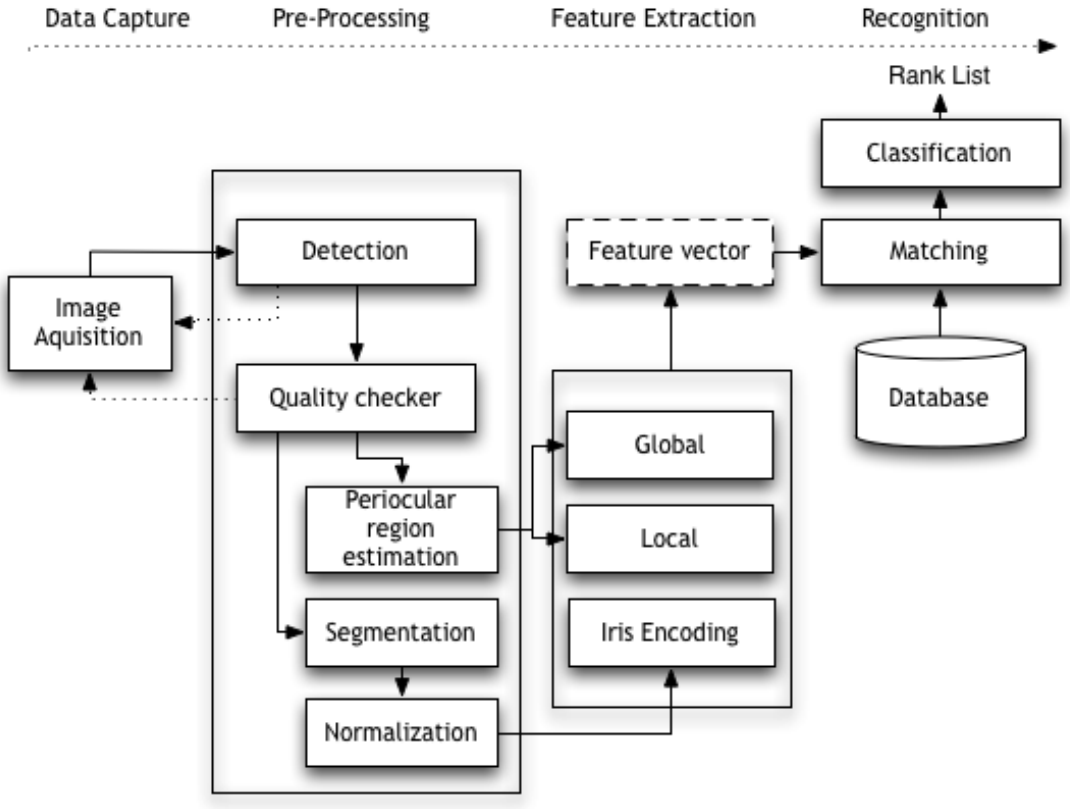


Figure 3.2: Proposed method schema.

Data Capture

Image Acquisition - Image acquisition is performed using an NIR illuminator and a high-resolution camera adapted with an infra-red band pass filter.

Pre-Processing

Detection - In all the data gathered (frames), it's used a detector to confirm the existence of a user. This sequence is essentially composed by a face and eyes detections. This is very important because it concerns the development of the whole process. The fact that the detection chain begins with face detection ensures a low level of cooperation. The user does not need to be in a single position to begin the process, since the system itself automatically confirms the existence of a possible user at its

reach.

Quality checker - Determines the quality of the images acquired, in order to ensure a good system performance in the most adverse conditions.

Periocular - region estimation The periocular region being used is estimated after the respective interest regions are established and verified. This estimate is performed based on the position of both eyes and the face.

Segmentation - This module includes the usual steps of the segmentation algorithms in the state of the art. The input parameter is an image of the interest region (left eye) and the output parameter is the localization concerning the iris.

Normalization - This step is directly associated with the segmentation of the iris. Its main purpose is the normalization of the region found for the iris [25].

Feature Extraction

Global features - In order to extract the global features from the periocular region, LBP and HOG descriptors suggested by Park et al.[34] were used. For this type of extraction, the region is subdivided in blocks, forming a grid. Each descriptor is then applied in each block to strengthen the information extracted.

Local features - The SIFT was used to extract global features from the periocular region [34]. Unlike the previously mentioned descriptors, this establishes the most important points of the region (keypoints), extracting only the information that comes from the area around these points.

Iris encoding - After segmenting and normalizing the region corresponding to the iris, the convolution of the resulting data is performed with a Gabor filter optimized for the system [25].From each convolution, a binary representation of the iris is created.

Feature vector - After using the several descriptors, fusion (feature level) is performed, allowing for biometric information to be represented using a single vector.

Recogntion

Matching - This step includes the obtainace of the similarity score between feature vectors. Therefore, the respective scores are established for each descriptor for all comparisons performed between the current user's data with all the templates stored in the database.

Classification - Using a neuronal network, it is performed a fusion at the score level, allowing the classification of the comparisons made during the matching between the interclass or intraclass. Finally, a rank-list is created based on the rearrangement of the values obtained.

3.2 Data capture

This section aims to present how the system was built, as well as the material used to capture data.

3.2.1 Image acquisition

With regarding to the image acquisition, three key elements are implied: a camera, an IR illuminator and an IR filter. By gathering these three components, it is possible to create an image acquisition model that works in real time (30 frames per second) and captures images in NIR, benefiting the recognition using the iris [11], [12], [13]. More specifically, the items used have the following descriptions:

Camera - Modelo BM-141GE Câmara GigE. High sensitivity 2/3" monochrome progressive scan CCD with NIR capabilities, 1392 (height) x 1040 (width) active pixels. Cell size 6.45 x 6.45 µm and 30 fps continuous operation with full resolution

Illuminatior - Raymax 25, 30 degree, 850nm, IR LED lamp

Camera filter - Bp850-46 infrared bandpass filter

Figure 3.3: (A) Camera, (B) Illuminatior, (C) Camera filter.

3.2.2 System Setup

As illustrated in fig. 3.4, the structure was designed to acquire images that allow recognition in less cooperative conditions. The system components are built in order for there to be an acquisition zone (AZ) between point $D1$ (1.5 m) and point $D2$ (2.10 m), having the camera lens as a reference source. Regarding the lens opening, it can vary between $W1$ (0.47m) and $W2$ (0.60m), considering its proximity to the camera. Finally, the focus of the lens was adjusted with a value between $D1$ and $D2$.

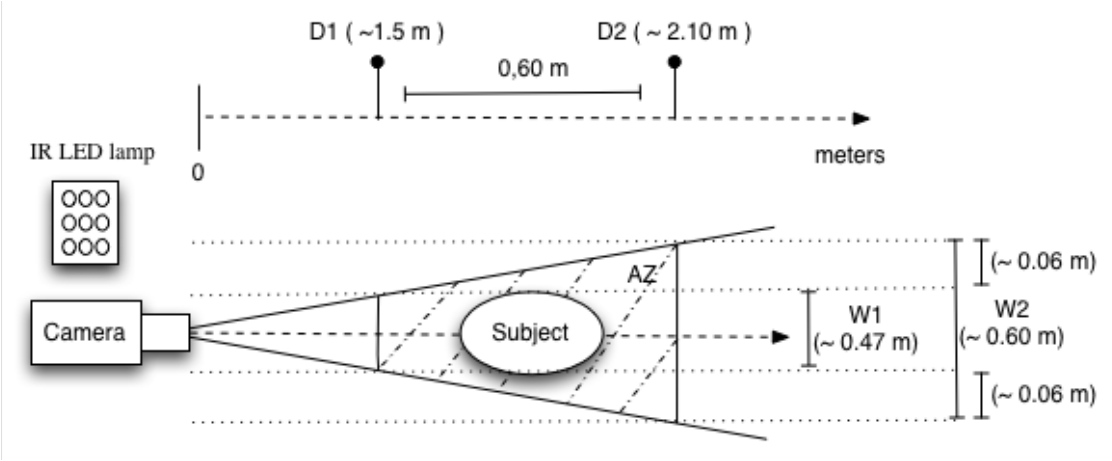


Figure 3.4: System layout and its capture range (seen from top to bottom)

It should be noted that everything that stands outside the acquisition zone but inside the camera's reach might also be detected. As we will demonstrate further along, these cases (frames), in which these conditions are confirmed, are discarded.

Figure 3.5: Data capture setup.

3.3 Pre-Processing

The pre-processing step includes a set of methods that makes easier the extraction of the biometric information needed to perform recognition. Considering this is a less cooperative system, this step also includes a small component to evaluate the quality of the acquired data and the detection of people. This same component is responsible for ensuring the presence of key properties for a good recognition performance. This phase is, without a doubt, the most demanding phase of the whole system, influencing the use of the strongest and with less computational cost methods in the detection [28], image quality measurement [54], segmentation [21] and normalization [25].

3.3.1 Detection of multiple regions of interest

The detection of multiple regions of interest in the image (face, left and right eye) has two main purposes: to confirm the existence of an individual, and to locate the necessary regions to extract the iris features and estimate the periocular region. This module begins with an automatic trigger of the camera that sends images with a pre-established frequency of fifteen frames per second to the computer. Each image received in the buffer is subjected to the procedure exemplified in fig. 3.6.

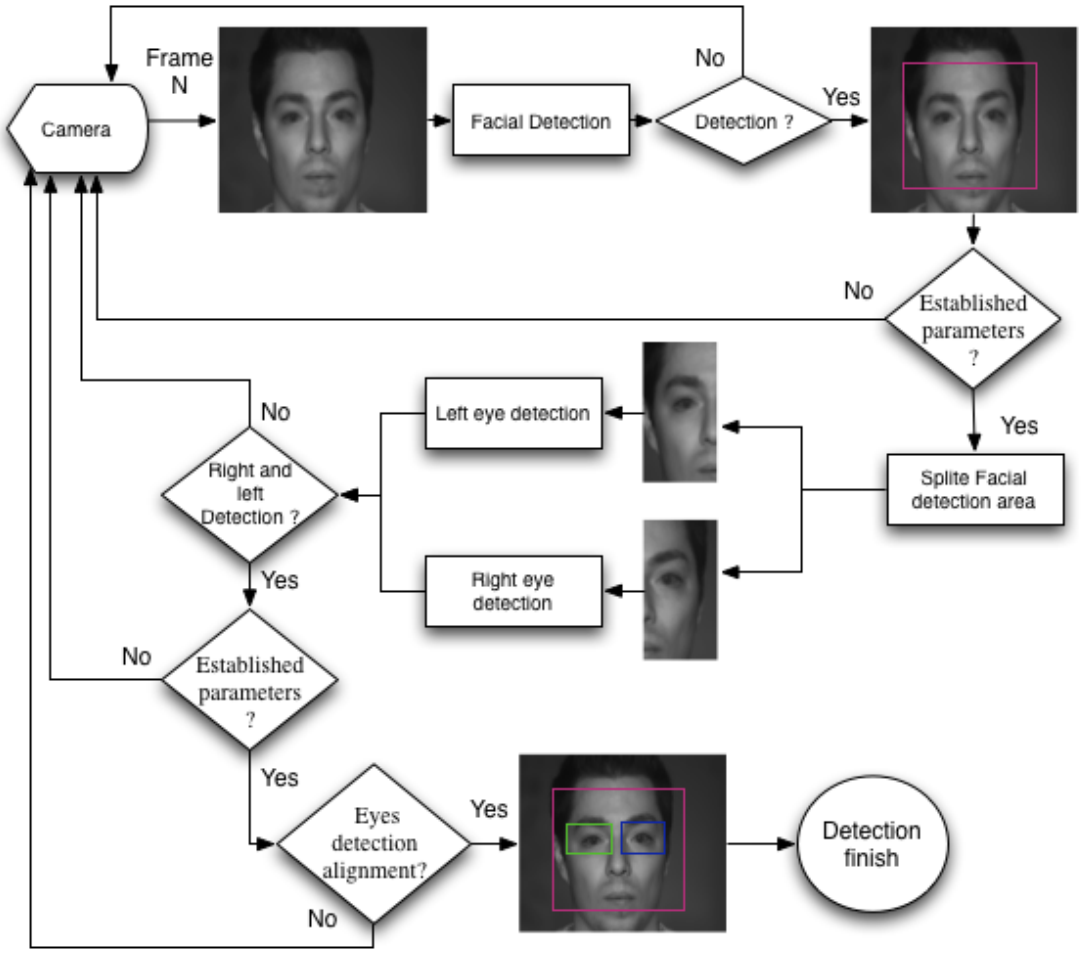


Figure 3.6: Detection module flowchart.

By using a face detector as a starting point, it is possible to establish the beginning of a possible

recognition. If the face detector establishes a region of the human face in the image and this region has an area larger than 40% of the input image size, the system confirms the existence of an individual for recognition. The need for this confirmation comes with the possible detection of false positives from the scenario's background and out of the acquisition zone. In order to apply both detectors at the same time, the region found is divided horizontally: the left eye detector remains on the left area and the right eye detector remains on the right area of the face. If the eyes are detected, both areas found are subject to property verification in order to avoid false detections (e.g. if the left eye and right eye areas have different proprieties, and vice-versa). Finally, it is also necessary to verify the horizontal alignment of the eye areas. This verification guaranties the discard of cases in which the user places their head in a less favorable position. Otherwise, those cases could influence the estimate of the periocular region in a negative way. It should be noted that, if any of the verifications or detections fails, the frame is automatically discarded from the buffer. The camera automatically resends a new frame to repeat the whole process until the camera is ordered to finish the process.

3.3.2 Quality checker

The scheme in fig. 3.7 corresponds to the development of the image quality evaluation module. This step can only be carried out if all of the verifications of the detection module (fig. 3.6) are positive. This module is based on the convolution of the detected left eye region with the kernel, in order to obtain a representative score of its quality [54].

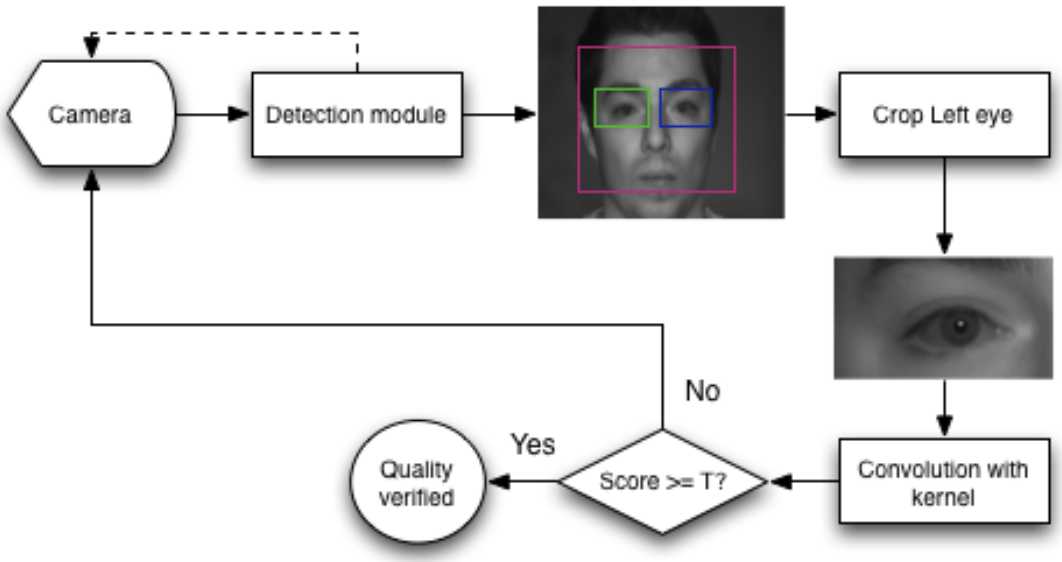


Figure 3.7: Quality checker module flowchart, were T is the minimum value established.

Several tests were performed in order to establish the quality limit value (T). During the first test the user was placed inside the acquisition zone to maximize the quality of the acquired data, and the resulting values of the method's application were registered. By doing so, the maximum average value in terms of quality towards the layout of the system itself and with the surrounding conditions were studied. The same analysis was done afterwards, in a second test, taking a less cooperative behavior from the user into consideration. This test allowed the analysis of the image's quality variations during its normal operation and the establishment of a quality interval. This interval is comprehended between the value obtained during the data

acquisition in great conditions (maximum quality) and the average quality during the system's normal operation, thus ensuring that all images within the established interval possess a constant and reasonable quality level. Otherwise, they are discarded.

Figure 3.8 illustrates the results obtained through the quality analysis during the system's normal operation. The set of points represents the relation between the total number of frames acquired and the corresponding quality scores. The values were then normalized, based on the maximum value found in the first test. Based on the points found, a behavioral approximation of the same was made in order to establish a quality threshold that could be used as a minimum value during the quality evaluation of future images. The $P0$, $P1$, $P2$ and $P3$ points seen in fig. 3.8 are important to visually understand the quality degree variation.

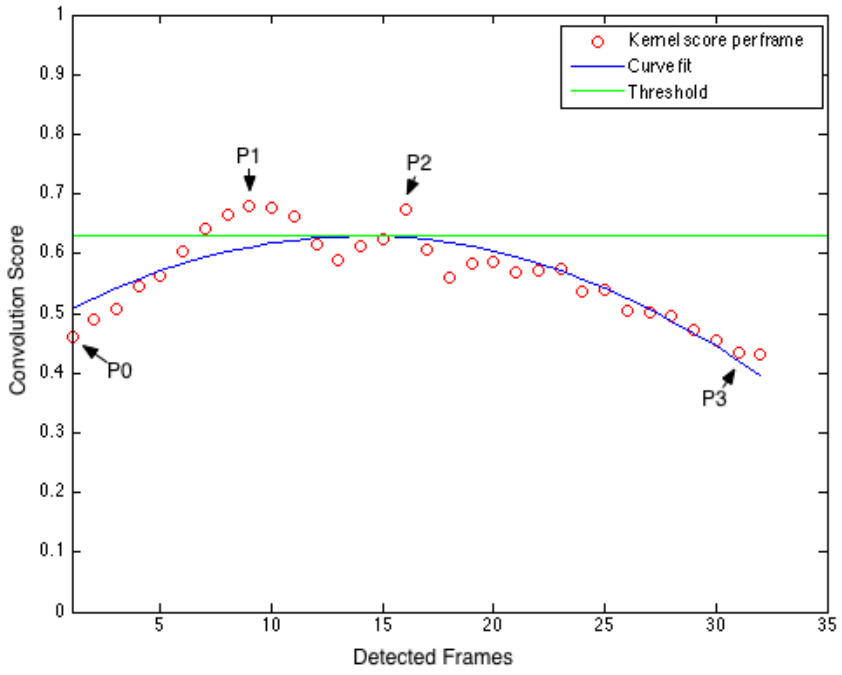


Figure 3.8: Quality analysis during the system's normal operation.

Using fig. 3.9 and fig. 3.8, it is possible to understand the behaviour of the quality degree associated with the distance covered by the user from the moment they enter the acquisition zone ($D2$ at fig. 3.4) to the nearest system acquisition point ($D1$ at fig. 3.4). As we can see in the image, the quality image reaches its highest values somewhere along the way. This happens mainly due to the camera lens being adjusted to an approximate distance in the middle of the acquisition zone.

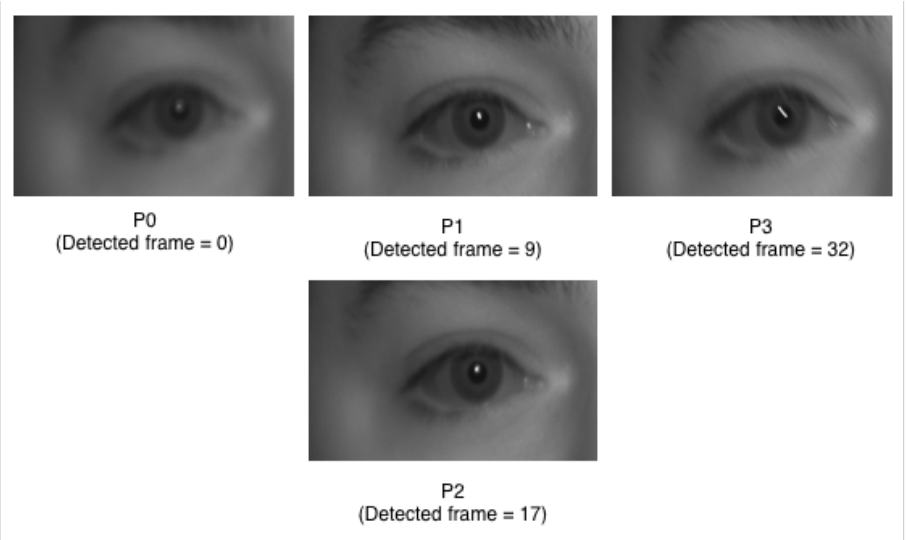


Figure 3.9: Respective frames for P0, P1, P2 and P3 points.

3.3.3 Periocular region estimation

This step aims to delimit the periocular region, based on the group of regions founded in the detection module. At this point in the system, the existence of a subject that has to be recognized and the quality parameters of the acquired data has already been assured, as we can see in fig. 3.10.

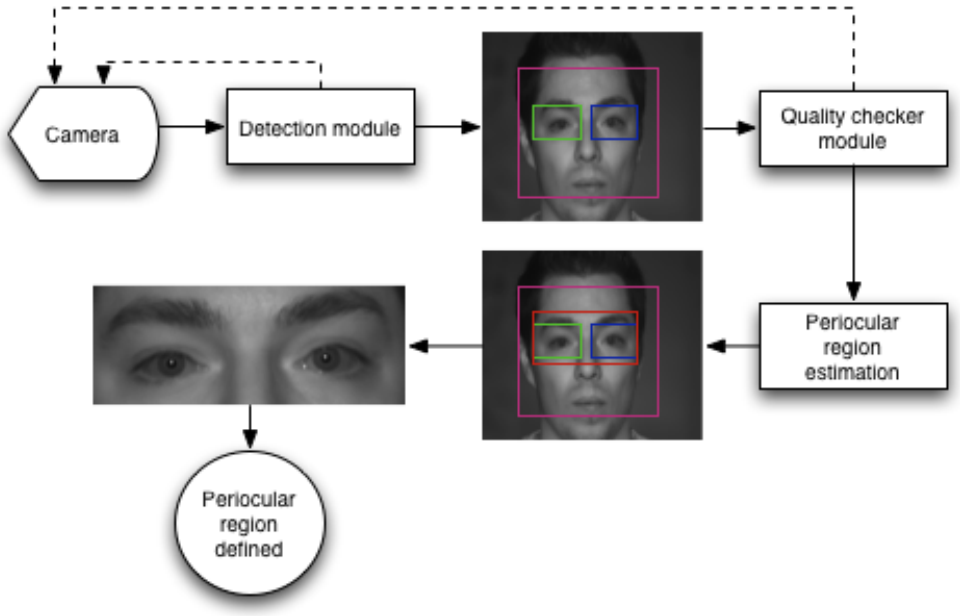


Figure 3.10: Periocular region estimation module steps.

In order to estimate the periocular region, several frames from multiple individuals and in multiple distances inside the acquisition zone were written down manually. From these notations of the periocular region, along with the detection properties (e.g. area and center of the regions), of the face and eyes, we were able to establish a percentage above and below the eye line. Allowing the definition of the periocular region for all future images presented to the system.

The region is thus delimited horizontally by the regions found through the detection of the eyes and must be delimited $D1$ and $D2$ percent, above and below the eye line.

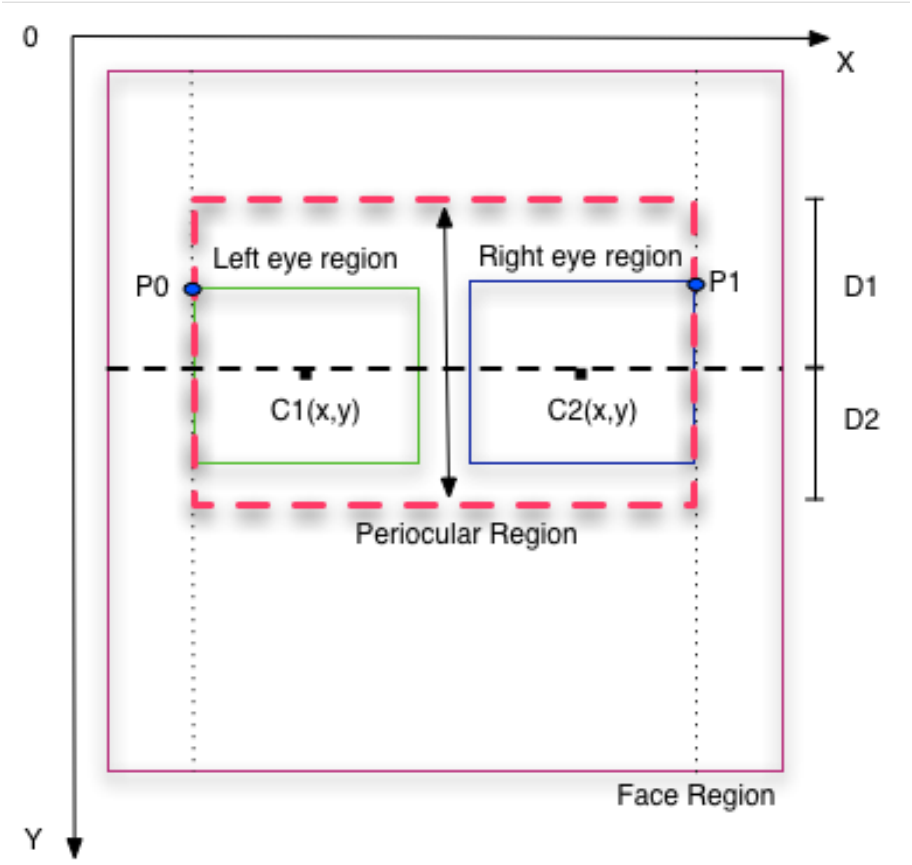


Figure 3.11: Periocular estimation based on the location of the face and eyes.

As we can see in fig. 3.11, it is possible to estimate the periocular region through the regions corresponding to the face and eyes. The $D1$ and $D2$ percentages above and below the eye line are established, based on the length x and height y of the face and the center of the eyes, $C1$ and $C2$. The region is then delimited vertically through $D1$ and $D2$ and horizontally by the corners points ($P0$ and $P1$) of the eye regions detected . The distances used are calculated through:

$$D1 = h.0.25 \quad (3.1)$$

$$D2 = h.0.20, \quad (3.2)$$

where h is a constant that defines the size of the region corresponding to the face in the vertical length.

3.3.4 Iris Segmentation

The segmentation of the iris is one of the key steps for recognition through the same, and is also one of the most complex and demanding steps, computational-wise. The fig. 3.12 shows how the building of this system module was approached.

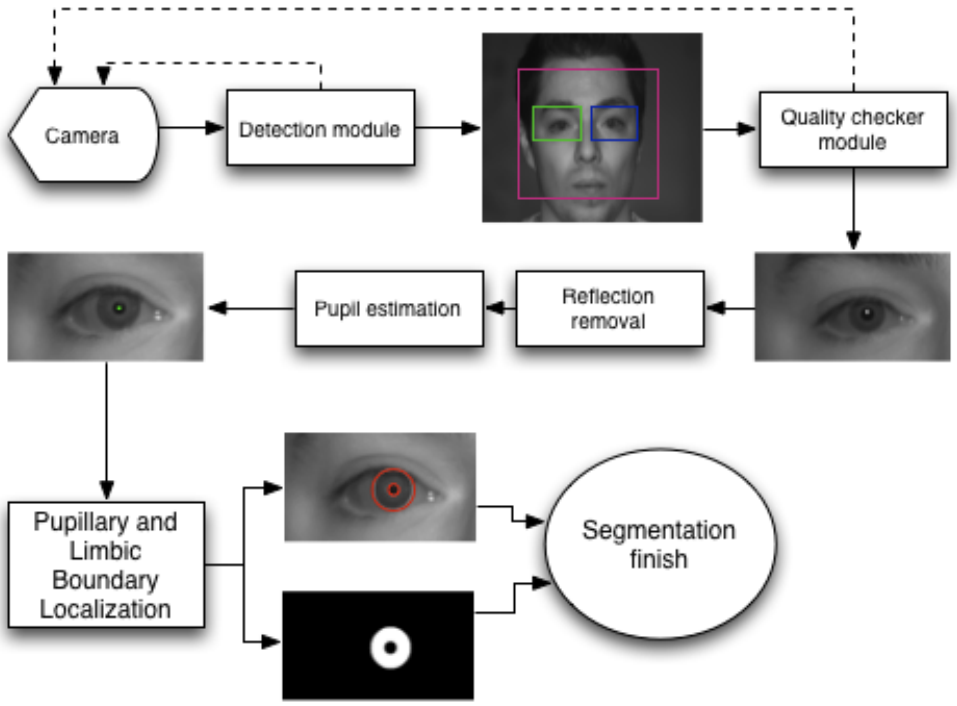


Figure 3.12: Iris segmentation steps.

We used the method proposed by Zhaofeng He et al. [21] as a starting point to address this question. In a similar way, this approach also uses a reflection removal and a pupillary and limbic boundary localization based on Hook's law. However, we propose to estimate the pupil center by analysing the image's darkest points

3.3.4.1 Reflection Removel

We use an infrared camera to have proper illumination for iris acquisition, although this approach introduces a specular reflection in the iris images in certain situations. Specular reflection appear in images due to several factors, as the intensity of the infrared illuminator or the environmental conditions of the acquisition. However, the main cause for this type of reflection is the use of glasses by the user. In order to avoid these cases, we propose the use of a specular reflection removal method [21]. The reflection does not disappear completely, but it is minimize, profiting a good performance of the rest of the iris segmentation.

This method proposes an adaptive threshold, T_{ref}, to locate the brightest point in the image, $I_{(x,y)}$, and to create a reflection map, $R_{(x,y)}$, of the original image. The areas corresponding to the reflections in $I_{(x,y)}$ will be felled with a bilinear interpolation method. For each $P^0_{(x_0,y_0)}$ point in the reflection area $R_{(x,y)}$, four points are calculated and defined as follows:

$$\left\{P^{left}_{(x_l,y_0)}, P^{rigth}_{(x_r,y_0)}, P^{top}_{(x_0,y_t)}, P^{down}_{(x_0,y_d)}\right\} \quad (3.3)$$

Each of these points represents the value of one point at a certain distance from one of four directions $(left, rigth, top, down)$ with an L distance from the last reflection point in that direction. When all four points are reached, the reflection point is filled by:

$$I(P^0) = \frac{I(P^l)(x_r - x_0) + I(P^r)(x_0 - x_l))}{2(x_r - x_l)} + \frac{I(P^t)(x_d - x_0) + I(P^d)(x_0 - x_t))}{2(y_d - x_t)} \quad (3.4)$$

3.3.4.2 Pupil estimation

In order to locate the boundaries of the iris, an approximation of the center of the pupil needs to be estimated. By doing so, the detected region of the left eye was minimized in order to reduce the effect of the eyebrow (10% of the top of the image is cropped). Afterwards, the darker points of the image are located through an established threshold. The obtained result is subject to an erosion morphological operator in order to minimize dark points on the eyelashes. Finally, all the points found horizontally and vertically are counted. When the peaks are found, they are intercepted to estimate the center of the pupil as illustrated in fig. 3.13.

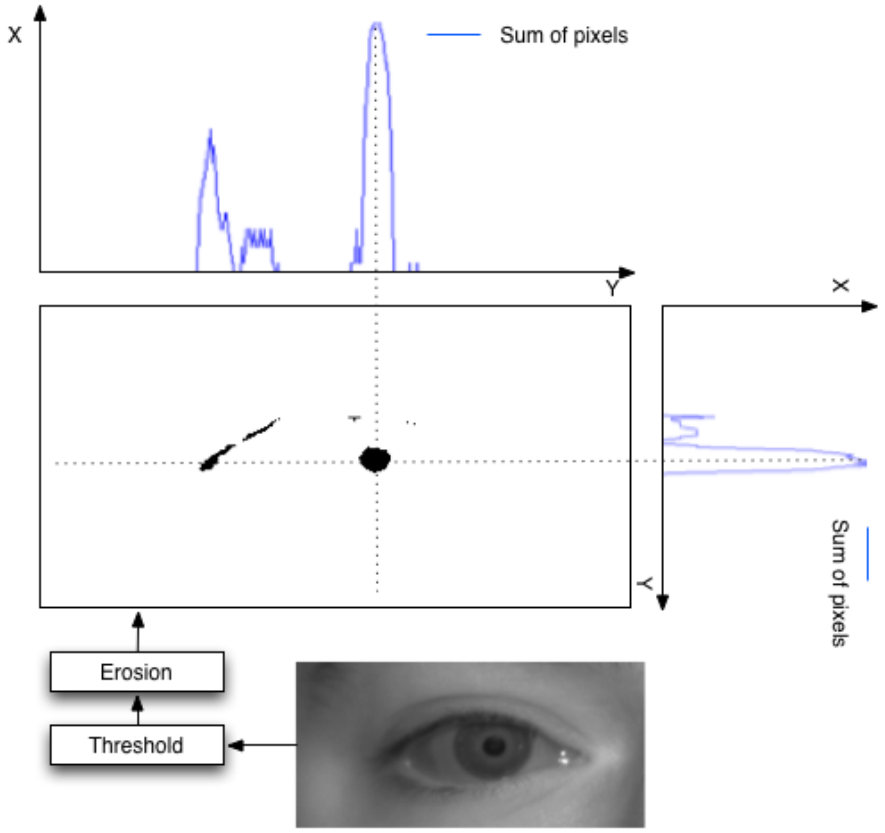


Figure 3.13: Pupil estimation steps. Localization of the darkest points in the input image and horizontal and vertical analysis.

3.3.4.3 Pupillary and Limbic Boundary Localization

The push and pull method [21] was used for the location of the limit between the iris and the pupil (pulpillary boundary) and between iris and sclera (scleric boundary). This method has its foundation in Hooke's law: the restoring force of a string is directly proportional to the deformation that is exerted on it,

$$\vec{f_i} = -k\,(R - r_i)\,\vec{e_i}, i = 0, 1, ..., N-1. \quad (3.5)$$

As illustrated at fig. 3.14, the restoring force $\vec{f_i}$ with the current length r_i is exert at each string S_i with the equilibrium length R with a constant k and direction $\vec{e_i}$ radiating the central point O'.Each point i of the total points N is associated one force $\vec{f_i}$.

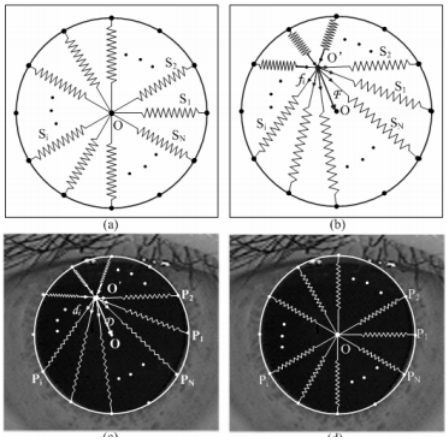


Figure 3.14: Basic idea of the Pulling-and-Pushing method, representing strings and force direction.

In this method, the point initially found in section 3.3.4.2 is taken as the central point O'. A set of points are then established $(P_i, i = 1, 2, ..., N)$, resulting from the application of an edge detector in the representation image in polar coordinates (fig. 3.15). The connection between each point and O' will represent a string $(S_i i = 1, 2, 3...N)$ that obeys Hooke's law eq. (3.6). Each string will create a radial force in O'. Then the sum of forces by eq. (3.7) is applied to O' and makes it move in gradually to its true center until it restores the balance between the strings.

$$\vec{d_i} = \vec{f_i} = -k\left(\bar{R} - r_i\right)\vec{e_i}, i = 0, 1, ..., N-1, \quad (3.6)$$

$$\vec{D_i} = \sum_{i=1}^{N} \vec{d_i} \quad (3.7)$$

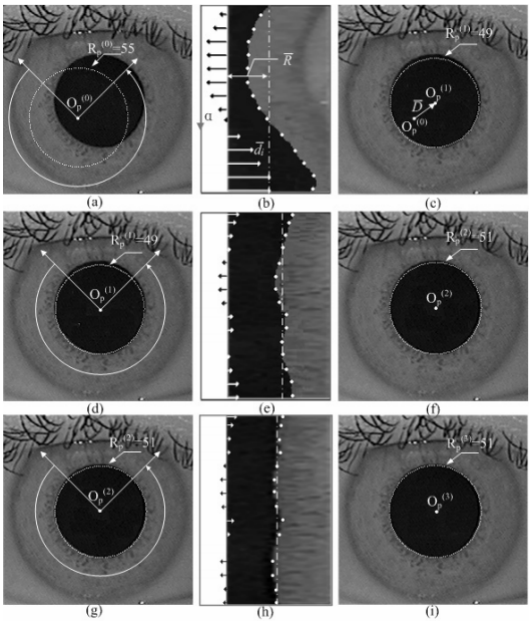


Figure 3.15: Illustration of the Pulling-and-Pushing procedure, where each row represents a iteration. Images in the left column illustrate the estimate of the pupil center and radius. The central column gives the detected edge points in polar coordinates and the right column gives the center displacement by the sum of forces.

Considering the representation of the image in polar coordinates in fig. 3.15, it is possible to obtain some representative points of the pulpillary boundary essential to the representation of $\{S\}_{i=0}^{N-1}$ imaginary springs. Each point is described radially $P_i\left(\alpha_i, r_i\right), i = 1, 2, ..., N$.The angles are in $\alpha\left[3\pi/4, 9\pi/4\right]$ to avoid possible occlusions generated by the eyelids. The distance from point O' is represented by r_i. Thus, only one P_i point can exist at a certain angle, thereby avoiding the influence of noise in the image. From the boundary points $P_i\left(\alpha_i, r_i\right), i = 1, 2, ..., N$, and O', it's possible to calculate the average value of the imaginary strings in the balance.

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} r_i \quad (3.8)$$

This process is done several times, interactively. To keep it from entering an infinite loop, it is established a threshold to the number of iterations and the value resulting from the convergence function (eq. (3.9)).

$$C(i) = \left|O_p^i - O_p^{(i-1)}\right| + \left|\bar{R}_p^i - \bar{R}_p^{(i-1)}\right| \quad (3.9)$$

The case $C(i)$ is inferior to the threshold set. The method may stop because it has converged.

In order to find the scleric boundaries, the same method is used, with the single difference that the initial r as a higher value and the information above the pulpillary boundary in a polar

representation is discarded.

3.3.4.4 Iris mask

Initially, the method represented here receives a left eye image as an input parameter, and the output is returned as a mask representing the fine location of the iris, fig. 3.16. This location allows the characterization of the information concerning the iris. The creation of a mask is important to the matching process, taking into account that only the concordant regions on both masks are compared.

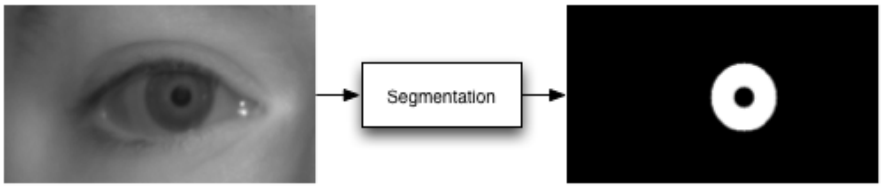


Figure 3.16: Resulting mask after completed the segmentation process.

3.3.5 Iris normalization

The normalization of the iris is done from where its border is located. This normalization was proposed by Daugman [15] and includes the transformation of a region's Cartesian coordinates into polar coordinates. The information in the iris is then represented by a determined radius (r) and angle (θ), as we can see in fig. 3.17.

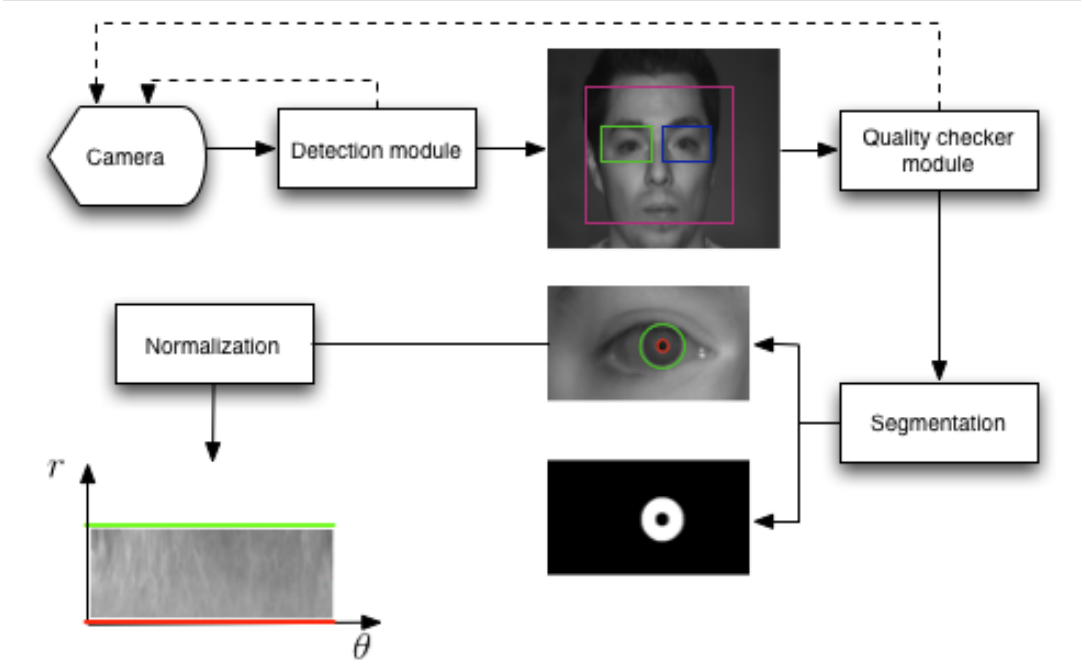


Figure 3.17: Iris normalization after the segmentation process.

The new iris representation ensures a level of invariance against certain factors (e.g. iris size and position). This type of invariance is essential for a good feature extraction and matching.

3.4 Feature extraction

When the pre-processing phase is concluded, all of the conditions required for extracting features from the periocular region and the iris are met. Global (LBP, HOG) and local (SIFT) features are extracted from the periocular region [34]. However, a grid was applied to the entire region during the global approach in order to strengthen the information's spatial connotation, in addition to the already acquired texture information. Regarding the iris, an optimized Gabor filter is used to perform a coding through the convolution of the its normalized region [15]. Considering the fact that the biometric information is sometimes extracted from the same region, this system can be called holistic. Finally, the feature level fusion is done and a vector is created with all of the concatenated biometric features, as shown in fig. 3.18.

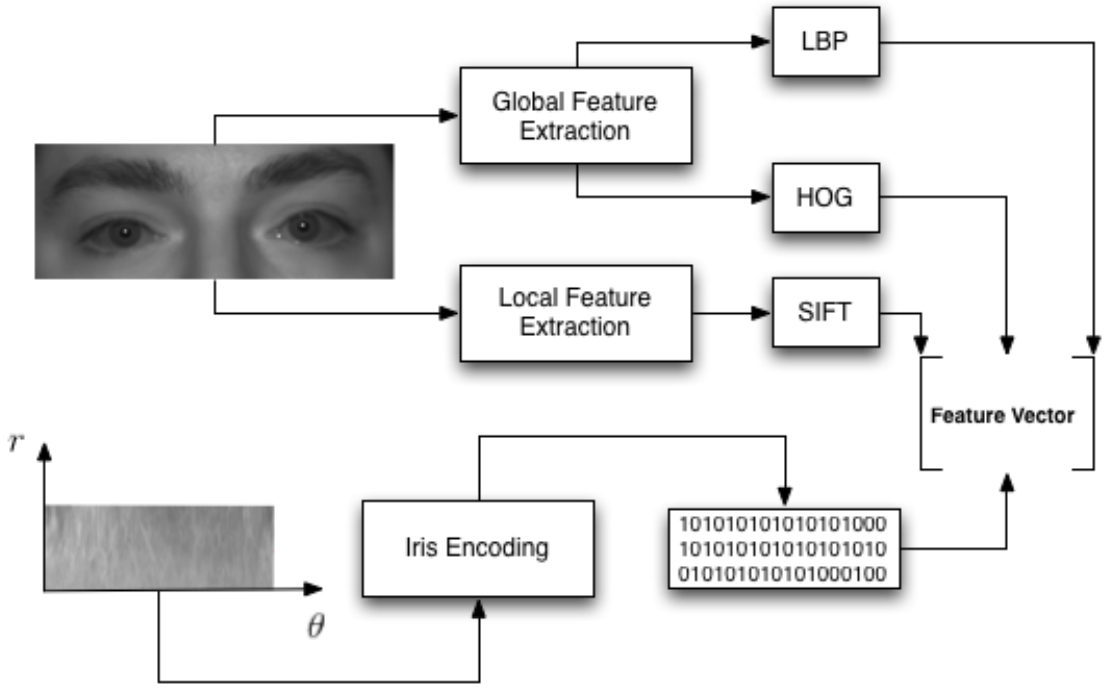


Figure 3.18: Feature extraction module with the respective descriptors.

3.4.1 Global feature extraction

As global descriptors, LBP and HOG describe the whole image texture, this can sometimes result in a low performance level. That way, the periocular region is subdivided by blocks, forming an auxiliary grid [34], in which both descriptors are applied for each block, instead of being applied directly to the whole image. The information extracted is thus reinforced, adding a spatial notion that did not exist before. Finally, the results obtained from the LBP and HOG for each block are concatenated in order to create two unique representations of the periocular region.

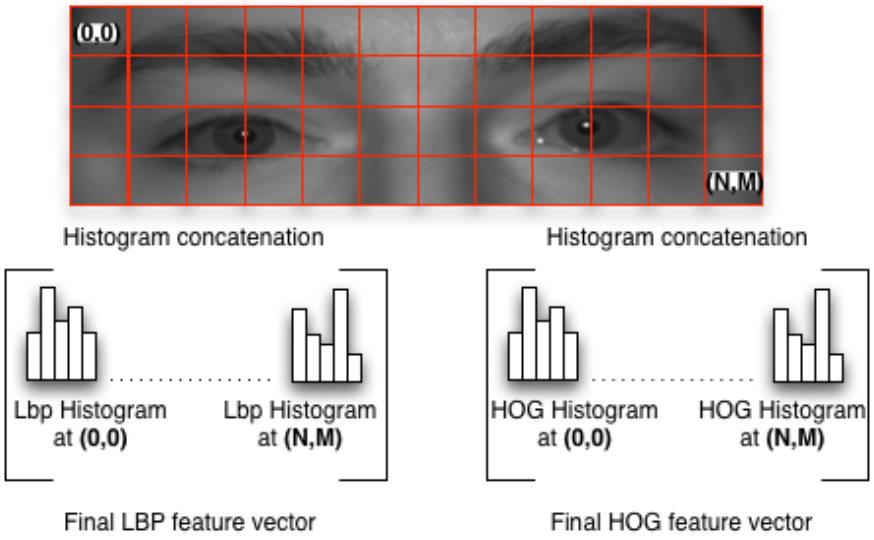


Figure 3.19: Feature vectors representation of HOG and LBP obtained through the regions imposed on periocualr region.

3.4.1.1 LBP - Local Binary Pattern

This method is widely used due to its simplicity and discriminative power. Because a mask of weights is used for each pixel, an LBP code is calculated by adding the highest values contained in the mask related to the pixel. In a simple analysis of all of the textures, generated codes would be represented by a histogram.

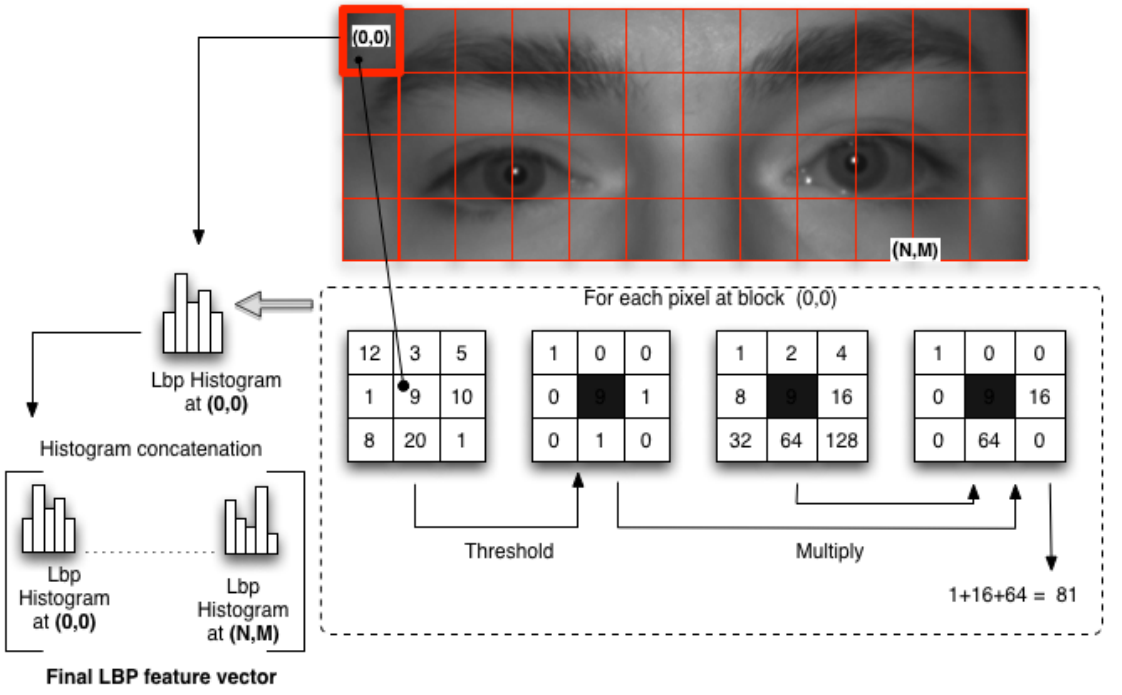


Figure 3.20: LBP code computation steps for each pixel inside a block.

3.4.1.2 HOG - Histogram of Oriented Gradients

The foundation of this method lies in the idea that the image can be described based on the gradient orientations. The method starts with a normalization of gamma and color before computing the gradients. Each block is subdivided in cells, where each pixel casts a weighted vote for a gradient orientation. The resulting histograms are then concatenated with the remaining histograms of other cells, forming a unique feature vector to each periocular block. The descriptor is formed by all the combined histograms in all regions.

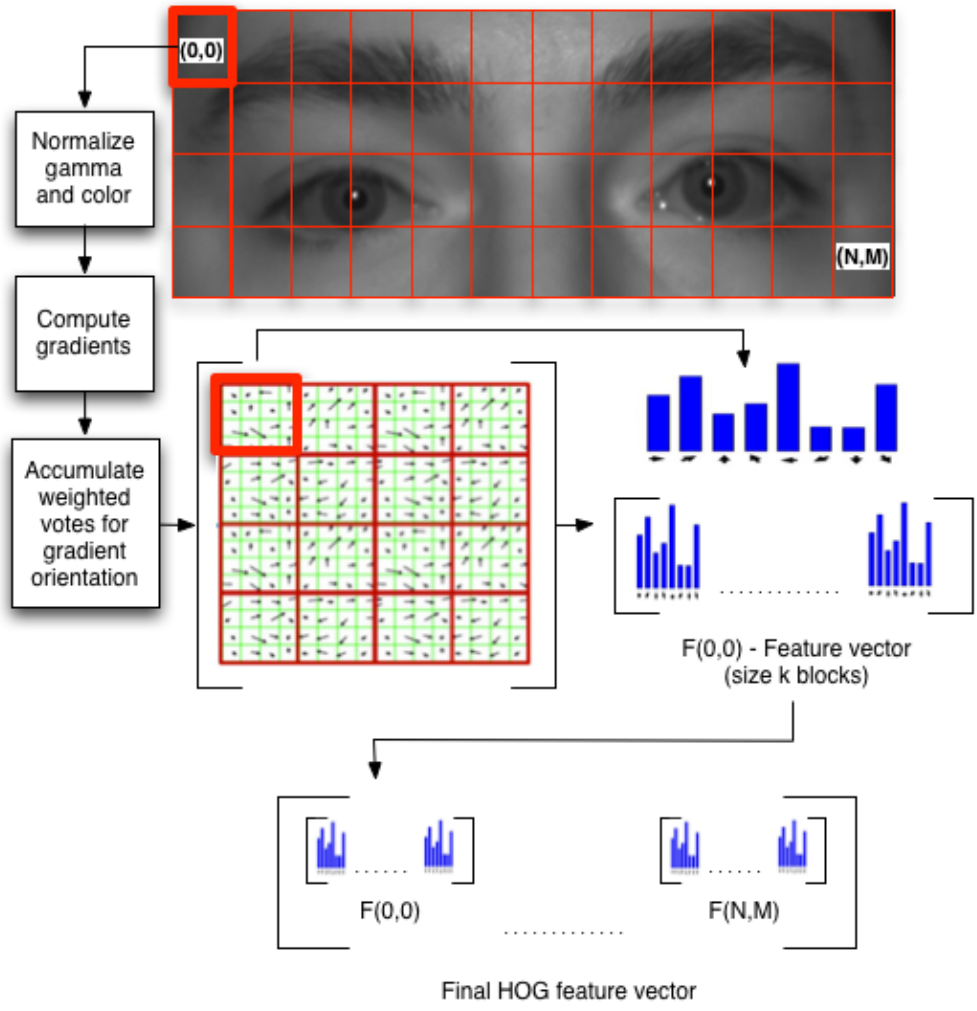


Figure 3.21: Gradient orientation analysis steps.

3.4.2 Local feature extraction

Unlike the extraction of global features from the periocular, the following kind of extraction is not subjected to any kind of condition (e.g. creation of a spatial representation). That way, the SITF is used, allowing the extraction of local features from all the periocular region.

3.4.2.1 SIFT - Scale-invariant feature transform

The SIFT is considered a local descriptor, since only small and relevant areas are used to describe the image. As an output, we obtained a feature vector invariant to geometric changes as translation, scaling and rotation. Similarly, this method shows robustness against affine distortion, change in 3D view-point (image from several perspectives), noise and lighting variations. This descriptor has the main advantage of being computationally fast, capable of extracting a large amount of features from small regions and insensitive to the occurrence of occlusion.

This method can be divided into two main parts: the detection of keypoints, which are scale invariant, and the building of a local descriptor. These are the following steps of the first phase:

Peak detection: Initially the convolution between the original image and a Gaussian filter is done within different scales. Then the Difference of Gaussian (DoG) are computed, resulting in a difference between filtered images at adjacent scales. These peaks are the possible candidates of key points.

Key points localization: Unwanted and outlier key points can be regarded as key point candidates. Similarly, the edges in the image tend to be detected as potential key points, which is not desirable for this method. For the outliers in the set of candidates, a Taylor expansion is used and applied to the DoG image, in order to eliminate these undesirable points along with a threshold. Using a Hessian matrix that represents a second derivative, it is possible to measurement the magnitude in DoG images, which in turn enables candidate points to be discarded.

Assignment of orientations: For the key points to become invariant to rotation, their information regarding the orientation is taken into account. Through the analysis of the gradient magnitude in the neighbourhood of the key point, it is possible to establish a histogram of gradient orientations. Then the peak of the histogram becomes the orientation associated to the key point. Certain key points may also have multiple peaks, thereby associated with multiple directions.

After the key points and their orientation are found in the image, and as a second phase of this method, the surrounding area of each key point is described. This description is done through the computation of the magnitudes of the gradients directions.

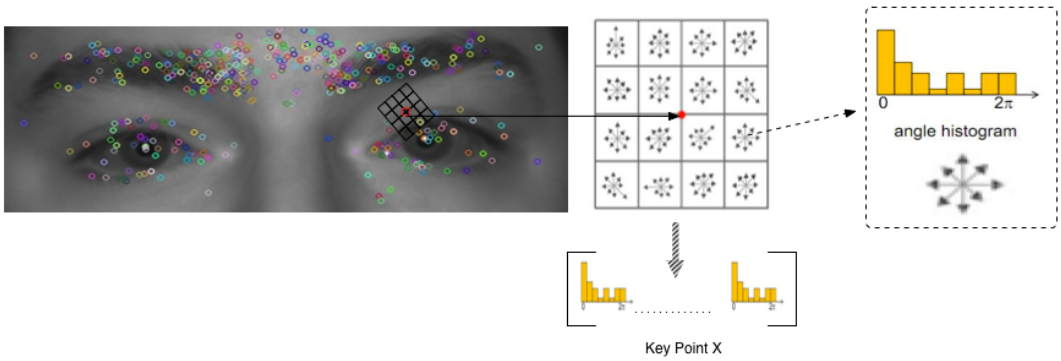


Figure 3.22: Key points detected and local descriptor representation.

Unlike the previously used descriptors for the periocular, the SIFT develops an array of features, which includes m lines, according to the number of key regions found, and n columns, according to the total number of directions times the number of cells of each grid associated with each key point.

$$A_{m,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \quad (3.10)$$

3.4.3 Iris Encoding

The method proposed by Daugman [25] was used to extract features from the iris. This method includes the convolution of the Gabor filter with the normalized image of the iris. Only the resulting information about the phase is taken into consideration, which is then converted into a binary representation and called the iris code.

3.5 Identification

This section aims to explain the last module of this system: recognition, which module can be subdivided into two main phases. The first phase aims to establish a similarity degree between the feature vector we want to recognize and all the biometric templates in the database. The second phase aims to classify each comparison as part of the interclass (same person) or intraclass (different people) set through the fusion of scores obtained.

3.5.1 Matching

When the process associated with the feature extraction is concluded, is necessary to establish a similarity degree between the subject's biometric features presented before the system and all the biometric templates in the database. This process is called matching in the biometric literature (fig. 3.23). Four scores are obtained from the comparison between feature vectors, and each one corresponds to similarity between LBPs, HOGs, SIFTs and iris codes. These scores are calculated through the hamming distance for the LBPs, HOGs, and iris codes. For SIFTs, a matching distance based scheme is used to obtain the score. Is important to say that on iris code matching, the hamming distance was used o obtain the proportion of disagreeing bits between two signatures extracted from the iris ring. We obtain an array after performing the matching with all of the database biometric templates. This array has as many lines as the number of comparisons made, and as many columns as number of descriptors used. Associated to each line of the array is associated the identification code (ID) of the biometric template owner.

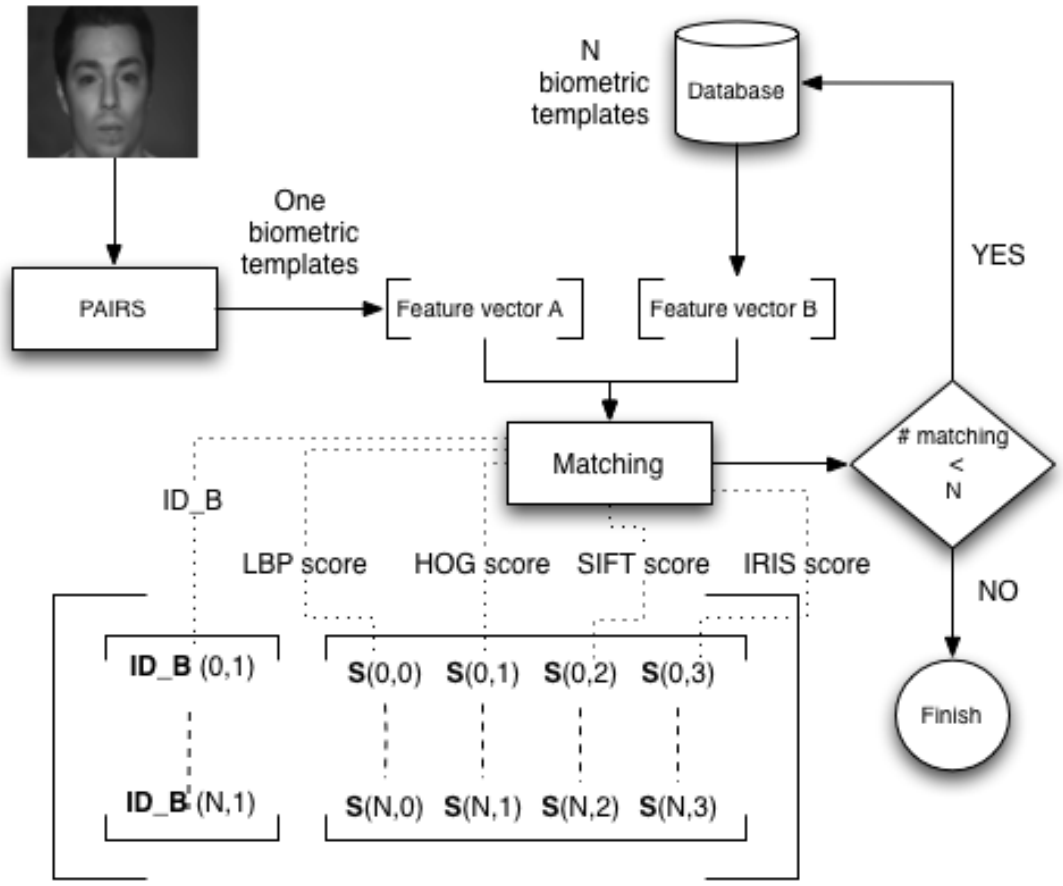


Figure 3.23: Features matching and scores array.

3.5.2 Classification

Finally the person identification is performed based on a method often used in the computational area of machine learning: the neural networks. It is a supervised learning method that, after being trained to a certain pattern, allows the comparison classification previously done regarding its kind. In our case we regard a binary classification problem, where 0 is a comparison with an different person, and 1 is a comparison between samples of the same person.

In order to create the sets needed for the training of the learning method, we asked several volunteers to take part in the biometric registration with a one-week interval. The data obtained was then crossed in order to create a considerable set of normalized comparisons between interclass and intraclass, allowing the system to be able to classify the biometric data comparisons that come from the same person or from different people as belonging to class 0 or 1, in which the closest values to zero correspond to impostors and the closest values to one correspond to genuine comparisons. Finally, the comparison list is sorted in descending order and presented by the system (rank list). If there are not comparisons above an established limit value, the system presents a warning as someone who has not made a registration and, therefore, is not in the database.

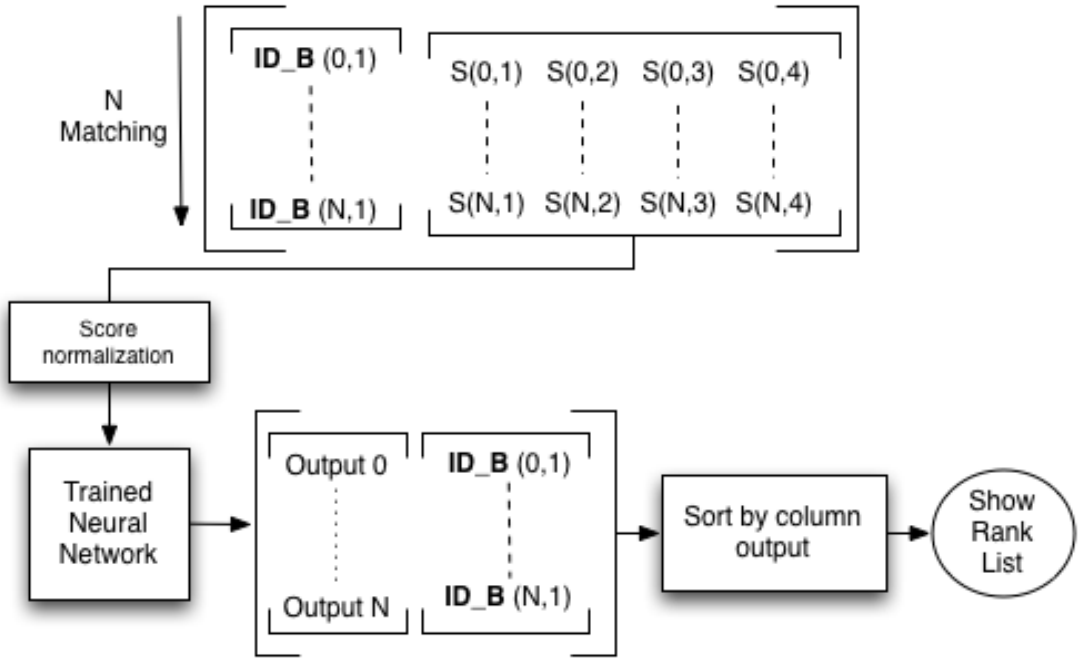


Figure 3.24: Score level fusion by classifying each matching score vector, where as final result a rank list is returned.

3.5.3 Enrollement

For the recognition to be done, the biometric data must be previously registered in a database. Therefore, the PAIRS also includes the registration feature, which implements, like the identification feature, all steps from the data capture to the feature extraction. With the user's biometric feature vector created, the system is built to have the participant's personal data inserted (e.g. name, age and sex). That way, along with the biometric data it is also stored the correspondent personal information into the database, and an identity card sized photo.

Chapter 4

Results and Discussion

The following chapter begins with the presentation and discussion of how important it is to acquire images to build a dataset. This set of images is indispensable for building and improving the system. Some work is then presented in terms of the descriptor optimization used to maximize the performance and then in terms of fusion. Finally, the final performance levels achieved by the proposed system (PAIRS) are presented and discussed.

4.1 Dataset

In order to enable the system's development, namely the classifier's training and testing (Neural Network) implemented in the recognition module, a set of images allusive to the use of the proposed system had to be created. The images were captured in conditions smiliar to the ones found during the system's normal operation, the same methods regarding the detection module and quality verification were used. All images belonging to this set have a 1392 x 1040 resolution and were captured in NIR at a distance that varies between 1.5 and 2.10 meters.

Ninety-three volunteers participated in this task, and each one of those participants appeared in two data capture sessions. Fifty images of each participant were captured in each session, which adds up to a total of 100 images per person, and to a total of 9300 images. There are 21 women and 73 men in each set of participants, with ages ranging from age 18 to 30. There was a one-week interval between the acquisition sessions in order to capture possible biometric variations in the participants and in the surrounding environment (e.g. lighting) over time.

Figure 4.1: Examples of captured images to the dataset.

Besides the data used in the identification module training, this dataset also aims to serve as a starting point to establish and verify parameters in detection modules, image quality evaluation and segmentation. It allowed us to establish several verification steps regarding the detected regions, to create an image quality score interval, to verify the performance of the segmentation method and its stopping parameters, and to create an optimized Gabor filter from the participant's segmented iris.

4.2 Iris segmentation and encoding

The use of a Gabor filter in the iris' encoding module is a key step to extracting information in the best possible way. Although it is possible to use the filter based on usually present parameters, suggested by some authors in the state of the art [15]. Understanding which are the best wavelength, orientation, phase and ratio values, is important in order to perform well in the imposed conditions for the proposed system. In order to analyse the filter's performance, the left-eye iris of each participant in the dataset were segmented. This allowed us to equally evaluate the segmentation module's performance in the conditions found throughout its normal operation.

4.2.1 Iris segmentation

Once again, the acquired dataset was essential to perform this task, considering that we were able to automatically detect the left eye and then segment the iris from the same. From a visual analysis of the obtained segmentations, we were able to verify the obtained performance by this method and to filter the cases in which that did not happen. This guaranties that the Gabor filter optimization subsequently only made use of good quality information and is not influenced by information that comes from poorly segmented regions.

The method was submitted to a set of 9300 left-eye images, from which 9076 images were considered well segmented. It was then concluded that the segmentation method used in the conditions imposed to the system had a failure rate of approximately 2%, and that nearly 98% of the images used showed a good segmentation of the iris.

Most of the images with negative results showed that this result was due to a significantly high level of eyelid occlusion. However, it must not be forgetten that there must have been other factors, such as a bad performance of the method responsible for the removal of reflexes.

4.2.2 Gabor optimization

After selecting the images that show a good segmentation of the iris, all the conditions are met to test the best parameters that should be acquired by the Gabor filter in order for the encoded information to reflect good quality in terms of biometric feature.

A set of 500 possible combinations of wavelenght, orientation, phase and ratio parameters was then created. From this set, we were able to compare each filter's behavior in terms of decidability (distance between the means of the intraclass and interclass distributions). Each filter was then used to extract information from all of the irises shown in the selected images (in which we can see a good segmentation). Afterwards, all signatures were compared, obtaining a score distribution for the interclass and intraclass comparisons. Therefore, each filter is associated with a decidability value (fig. 4.2), which can be seen in the filter's capacity to obtain good results when it comes to comparing the iris between the same people and different people.

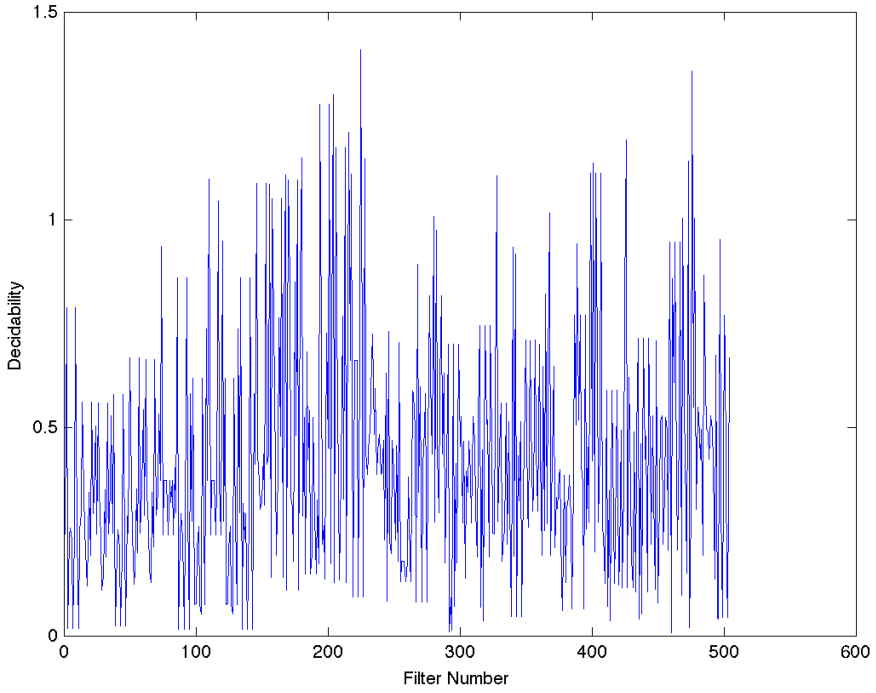


Figure 4.2: Gabor filters decidability.

Once this test was completed , it was possible to select the filter that performed best, taking into account maximum values found in terms of decidability. Thus, the selected filter is composed of the following properties:

Optimal gabor filter parameters					
Filter Number	Wavelength	Orientation	Phase	Ratio	Decidability
225	2	0	1.5708	1	1.409

Table 4.1: Gabor parameter combinations.

4.3 Performance evaluation

The individual evaluation of descriptors or through their fusion is always based on the results obtained in the classification of a testing set. This set includes images of the participants obtained in conditions similar to the ones used for the training of the implemented classifier in the system proposed. The use of the same testing set for all the tests made was taken into consideration, assuring that all tests have the same acquiring conditions.

4.4 Descriptores optimization

The extraction of biometric features from the periocular region includes a set of highly parameterized descriptors (e.g. number of histogram bins, thresholds, etc.), which makes us question which values should be assigned to the parameters and which are the best possible combinations to do so, taking into consideration the conditions imposed to the system. Several studies were done in order to evaluate the existing relationship between the parameters used and the error rate obtained in the recognition module. The descriptors are optimized individually, so the op-

timization reflects a system improvement through the fusions of the resulting scores. In order to make this kind of optimization, groups of possible combinations of values for the parameters of each descriptor were created.

Considering the fact that there are more parametrized descriptors than others, the size of the groups created may vary. Afterwards, and based on the images captured for the dataset, biometric features were extracted from all the participants, based on previously created variations. Thus ensuring that each descriptor is optimized to the same type of conditions imposed on the images.

4.4.0.1 LBP

Regarding the LBP, only two parameters were approached: the number of bins used in the histograms created and the number blocks associated with the grid estimated for the periocular. Five possible values were attributed to each parameter, obtaing a total of 25 combinations. The participants' biometric features were extracted from these combinations and used for the recognition process, obtaining an error value associated with each combination (fig. 4.3)).

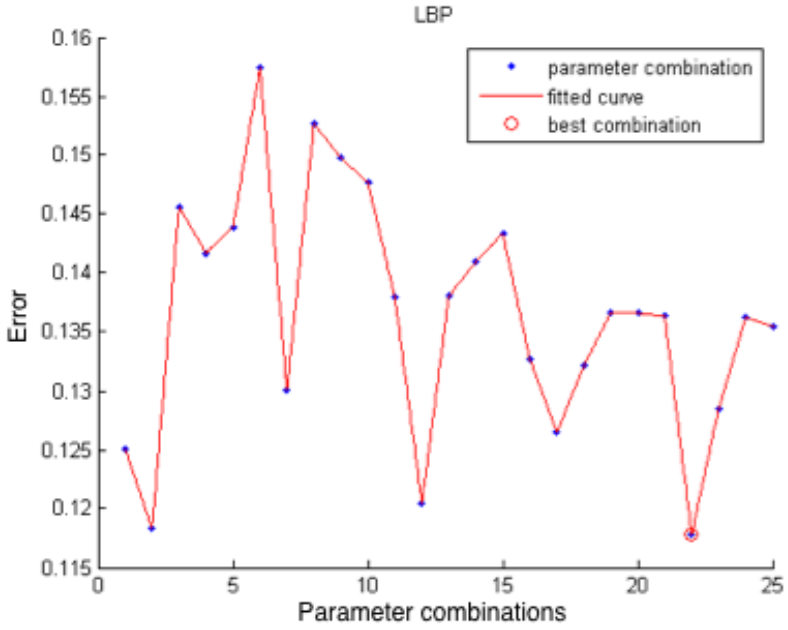


Figure 4.3: Local binary pattern parameter variation and respective error.

Being the horizontal axis the number of combinations created and the vertical axis the associated error, it is possible to see that combination # 22 is the combination with the lowest error in the entire set created.

Table 4.2 shows the initially combination and its error, in comparison to the combination found and its error. As can be seen, the new combination has a decrease of approximately 2% in the error associated with this descriptor.

Type	Combination number	Number of bins	Number of blocks	Error
Standard	22	256	32	0.141
Optimized	25	256	4	0.121

Table 4.2: Local an binary pattern, standard and optimal parameter combinations.

4.4.0.2 HOG

Unlike the LBP, the HOG descriptor is included in a set of parameters directed at the subdivision of the region from which one wants to extract features. That way, for this descriptor, five different parameters were approached: the number of bins for each histogram, the block size of each block in the periocular grid, the block stride for the overlaying value of the blocks and the cell size for the subdivision of each block, obtaing a total of 40 possible combinations. Once again, and from the created combinations, we obtained the error value associated with each combination (fig. 4.4).

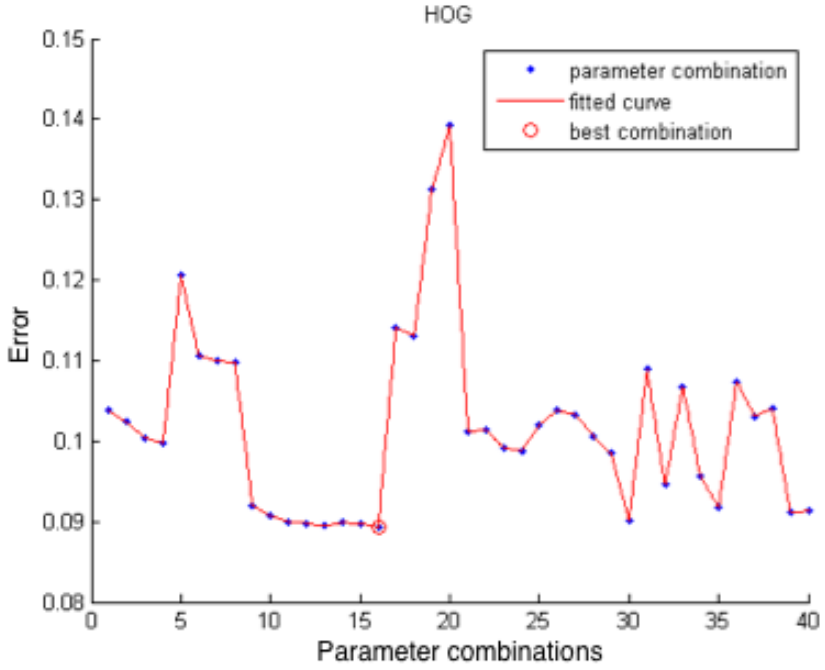


Figure 4.4: Histogram of Oriented Gradients parameter variation and respective error.

Being the horizontal axis the number of combinations created and the vertical axis the associated error, combination # 13 shows a minimum error in the entire set created.

Type	Combination number	Number of bins	Block size	Block stride	Cell size	Error
Standard	7	9	64	32	8	0.1144
Optimized	13	7	64	32	16	0.094

Table 4.3: Histogram of Oriented Gradients, standard and optimal parameter combinations.

Regarding the set of default parameters used, and in comparison with the parameters found (table 4.3), we were able to minimize the error by 2%.

4.4.0.3 SIFT

The parameters were varied for the SIFT descriptor: nOctaveLayers, contrastThreshold e edgeThreshold. A total of 80 possible combinations was obtained, and each one of these combinations is associated with its error (fig. 4.5).

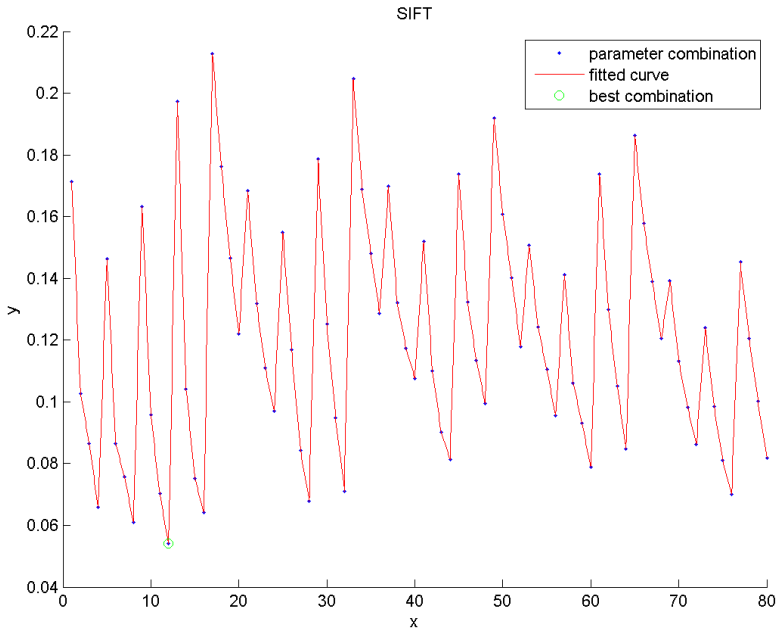


Figure 4.5: Scale invariant feature transform parameter variation and respective error.

Being the horizontal axis the number of combinations created and the vertical axis the associated error, combination # 12 shows the smallest error in the entire set created.

Type	Combination number	Octave Layers	Contrast Threshold	Edge Threshold	Error
Standard	36	3	0.01	10	0.1287
Optimized	12	1	0.03	10	0.0541

Table 4.4: Scale Invariant Feature Transform, standard and optimal parameter combinations.

Regarding the set of default parameters normally used, and in comparison with the new parameters found (table 4.4), we were able to minimize the error by 7%.

4.4.0.4 Discussion

In general, better parameter combinations were found for each descriptor used. Despite the fact that the combination of some parameters showed a bigger computational and storing load, the combinations selected did not show these two features at all. Despite the fact that the study focuses on the error rate of the recognition component, it was also important to verify if the new parameters could influence the system in other ways (e.g. storage capacity and processing). After some system verifications it was found that the new parameters established for the descriptors do not influence the system in terms of storage and processing.

4.5 Results

Aiming to evaluate the capacities of the system proposed with a multimodal system, an analysis was done based on the fusion of the information that comes from each descriptor (score level fusion). Equally important, the odds associated with the participant's identification in terms of the rank list are also represented. Finally, the influence of the participants in the system's performance is also analysed.

4.5.1 ROC curves

Through the analysis of the ROC (Receiver operating characteristic) curves, namelly through the calculation of the AUC (Area under Curve), we can observe the real performance in terms of classification by each descriptor and by the fusion at score level. The ROC curves represent the existing relation between true positives and false positives resulting from the system's answer to a testing set.

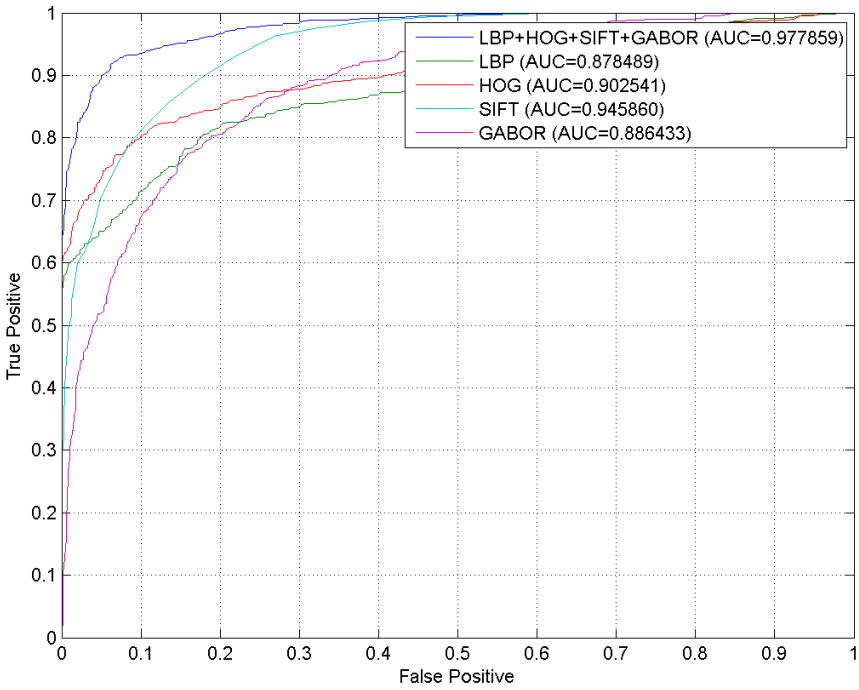


Figure 4.6: Receiver Operating Characteristic Curves observed for each descriptor and fusion of the same.

Considering this scenario, analysed in fig. 4.6, it is remarkable that the SIFT is the descriptor that stands out the most, even though all of the descriptors reveal a considerable success rate. Regarding the information extracted from the iris (Gabor), it reveals a lower success rate in relation to most of the other descriptors (except the LBP). Considering that the iris is a biometric feature stronger than the periocular one, we can conclude that the results are obtained due to less cooperative conditions acquired by the system. Regarding the fusion between all the sources of information, the performance difference regarding a possible particular use of each descriptor is remarkable, leading to a reinforcement of the main objective in this kind of operation.

Since that the system presented is multimodal, it was important to analyse the performance obtained through the fusion with the periocular descriptors and of the periocular with the iris.

Looking at fig. 4.7, it is clear to see that the performance achieved by the fusion of the several periocular descriptors is quite considerable. However, when the iris information is added to the fusion, we can see a slight, yet significant, increase in the system's performance.

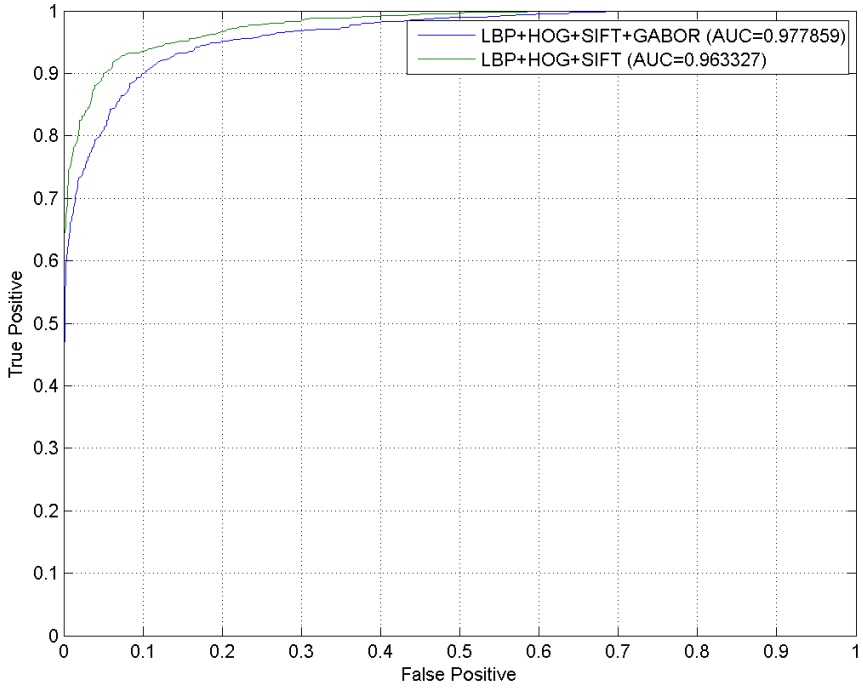


Figure 4.7: Receiver Operating Characteristic Curves observed for periocular descriptors and periocular with the iris fusion.

4.5.2 Cumulative Match Characteristic

More directed to the evaluation of identification systems, in which the output is represented by a list sorted according to the classification assigned to each biometric comparison between individuals. The Cumulative Match Characteristic (CMC) curve is calculated and analysed, which represents the probability of a correct identification of the participant among the first K elements of the final list returned by the system (fig. 4.8). In order to calculate the odds, 93 participants were subject to identification.

The number of times that the correct identification of a determined participant appears among the first K positions of the rank is evaluated. As we can see in fig. 4.8, there is an 82% chance that the correct identification apears at the top of the list. Consequently, there is a 90% chance that the identification appears among the top two positions of the list. In this system, and if the list has 43 positions, it would be possible to claim that the participant's true identity could always be found in the list.

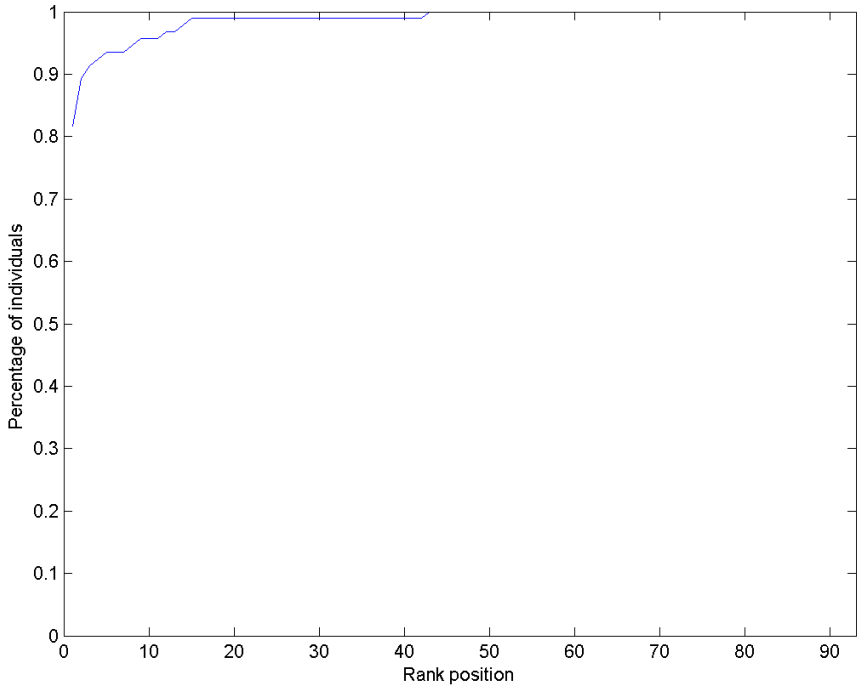


Figure 4.8: Cumulative Match Characteristic observed for the participants in the dataset.

4.5.3 Biometric menagerie

The following evaluation aims to understand the participant's influence in the system's performance. Based on the distribution of the scores obtained by the system for genuine and impostor combinations, it is possible to establish a relation between the two and classify some participants, [55]. That way, the classification is done according to the following classes:

Doves

Represents the kind of participant most favorable to the system's good performance. This class produces low distances for intraclass comparisons and high distance for interclass comparisons.

Chameleons

Represented by how the system easily confuses them with other identities. This class mainly includes high scores, both for interclass and intraclass comparisons, which leads to a considerable false accept rate in comparison with false reject rate.

Phantoms

Unlike the chameleons, this group of subjects normally creates low scores, in both interclass and intraclass comparisons. They are therefore associated with a high false reject rate and with a low false accept rate.

Worms

A class of worms represents the worst kind of participant to which the system can be subject to, considering it makes the system assign high values to both interclass and intraclass

comparisons.

Sheep

This class represents most users of the biometric systems. Normally, these obtain low scores when compared with other people and low scores when compared to themselves.

Particularly for the biometric system and the biometric features of the participants, we obtained the representation in fig. 4.9, with the average scores obtained for genuine and impostor comparisons for each participant.

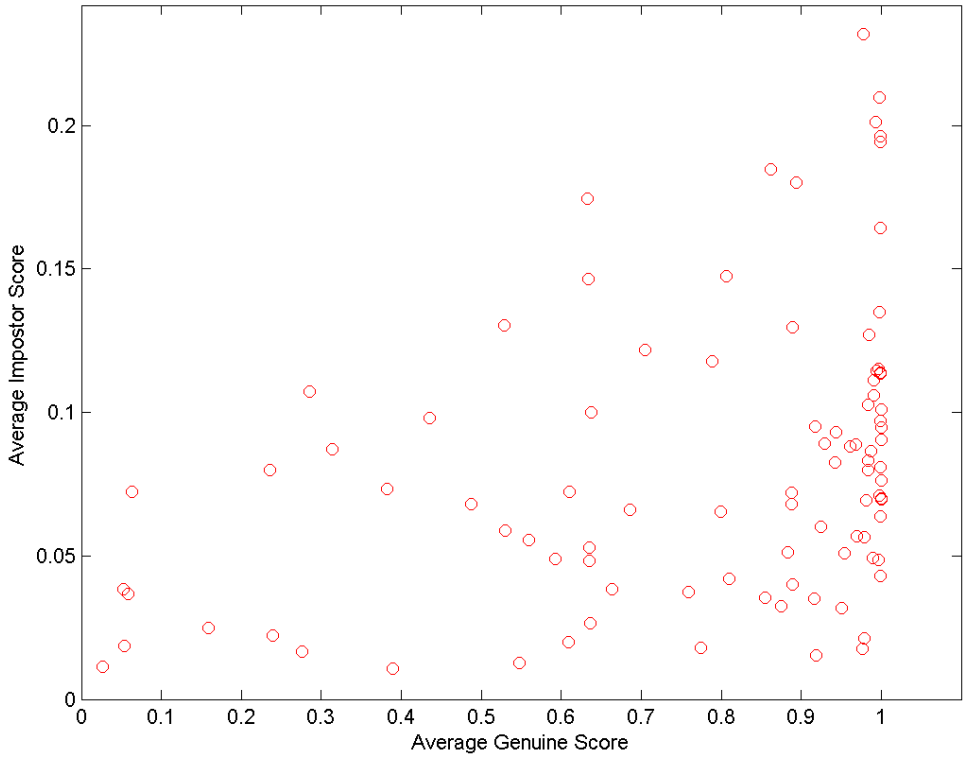


Figure 4.9: Scores distribution in terms of participants.

When analysing and subdividing the distribution's quartiles of the impostor and genuine scores, we were able to count the number of participants included in each class.

Class	Number of subjects
phantoms	10
chameleons	11
worms	2
sheep	49
doves	0

Table 4.5: Participant's classification through biometric menagerie.

Generally, it was not difficult to identify most participants, considering that the sheep class includes most participants. However, there is a considerable number of chameleons and even two worms, which makes it harder for the system to perform its task. Regarding the phantoms,

the existence of participants in this class is as worrying as the appearance of chameleons and worms, considering that phantoms, even when they are not identified, are rarely associated with another participant's identity.

Visualizing fig. 4.10 and fig. 4.11 it is possible establishing a line of reasoning that justifies the classification of the captured data from the participants as sheep or worm. In the first case, (fig. 4.10) is possible to verify that several negative factors as hair occlusion of the periocular region, head tilt and makeup lead to a significant variation between the two acquisition sessions. Bringing to the system a greater difficulty in the recognition phase and classifying this participant as a worm.

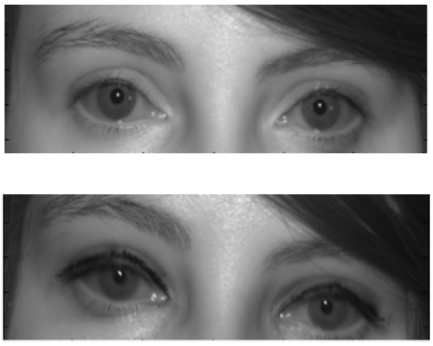


Figure 4.10: Periocular data acquired from two sessions, where the participant was classified as worm.

On the other hand the participants classified as sheep (fig. 4.11) showed no greater variation between the two sessions. Thereby enabling the system to have a larger facility in recognizing this class.

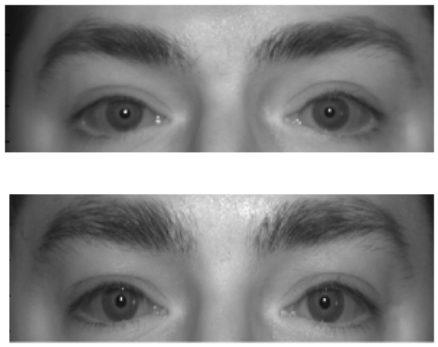


Figure 4.11: Periocular data acquired from two sessions, where the participant was classified as sheep.

It should be noted that this participants classification only makes sense under the same conditions in which the images for the dataset were acquired. The fact that they have been classified in this way before these conditions, does not mean that in future sessions, they can not get another result to the system.

Chapter 5

Conclusion

This project aims to show how it is possible to combine knowledge in biometrics, computational vision and automated learning in order to implement a biometric system.

Given this goal, the proposed system uses as background the state of the art in terms of data capture, pre-processing, feature extraction and recognition. It was possible to draw and implement the Periocular and Iris Recognition System (PAIRS) through the combination of all these components and implemented methods associated with them.

In order to strengthen the quality of the image and the system's automation, quality detection and evaluation modules of the images captured were implemented, thus assuring a decrease of the cooperation required, which is one of the goals to achieve when the system is working properly. The methods associated with the extraction of iris and periocular features followed a line of thought similar to the state of the art found, thus assuring that the system is built on a solid and mature basis. Regarding the knowledge and in order to strengthen the strong points of each component associated with feature extraction, the implemented system uses the fusion of the scores obtained through each descriptor.

When the system's implementation was concluded, the data acquisition that allowed us to underlie and justify the following developing line was required. From this data it was possible to achieve results and make its optimization as presented in Chapter 4. Based on this analysis, it is possible to conclude that the system proposed here includes a set of features needed to answer the desired goals for this project. However, it is important to state that both the optimization and the results achieved are directly related to the dataset used. Nevertheless, the system's ability to achieve such performance levels in different or adverse conditions is not assured in any way.

In the same area of work , especially more upset for performing biometric recognition in mobile devices, an article was published in Pattern Recognition Letters. The published article can be found in the annex of this dissertation.

5.1 Future Work

Considering the goals achieved in this project and with the prospect of continuity the work done so far, some interesting points worth studying in the future are presented. These proposals involve the creation of a new dataset acquired in more defying conditions, the approach of new optimization methods and the use of more advanced processing technology (e.g. GPU-accelerated computing).

Considering that the data capture for the training and testing of the system are a key step to-

wards the system's substantiation, the conditions in which this data is acquired have a big impact on the system's performance. The capture of new data in even less cooperative environments could possibly lead to the birth of new challenges and to notable and higher performance.

Regarding the method optimization, mainly those associated with feature extraction, it would be interesting to approach this question through approaches associated with the field of automated learning (e.g. Genetic algorithms).

In order to increase the computational power, it would be beneficial to use high-level computing (e.g. GPU-accelerated computing), considering there would not be such a high level of concern regarding real-time operation, thus allowing for an increase of computational charge and performance increase.

Bibliography

[1] R. Bolle, *Guide to biometrics*. Springer, 2004. 3

[2] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer, 2007. 3

[3] M. Bromba, "Biometrics faq's," 2014. [Online]. Available: http://www.bromba.com/faq/biofaqe.html 3

[4] J. Renaghan, "Etched in stone," *Zoogoer, August*, 1997. 4

[5] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security & Privacy*, vol. 1, no. 2, pp. 33–42, 2003. 6

[6] A. K. Jain, R. Bolle, and S. Pankanti., *Personal Identification in networked society, 2nd edition*. Kluwer Academic Publisher, 1977. 10

[7] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005. 11

[8] D. T. Meva and C. Kumbharana, "Comparative study of different fusion techniques in multimodal biometric authentication," *International Journal of Computer Applications*, vol. 66, no. 19, 2013. 11

[9] J. R. Matey, R. Broussard, and L. Kennell, "Iris image segmentation and sub-optimal images," *Image and Vision Computing*, vol. 28, no. 2, pp. 215–222, 2010. 16

[10] V. N. Boddeti and B. V. Kumar, "Extended-depth-of-field iris recognition using unrestored wavefront-coded imagery," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 40, no. 3, pp. 495–508, 2010. 16

[11] F. W. Wheeler, A. A. Perera, G. Abramovich, B. Yu, and P. H. Tu, "Stand-off iris recognition system," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*. IEEE, 2008, pp. 1–7. 16, 36

[12] W. Dong, Z. Sun, and T. Tan, "A design of iris recognition system at a distance," in *Pattern Recognition, 2009. CCPR 2009. Chinese Conference on*. IEEE, 2009, pp. 1–5. 16, 36

[13] Y. He, J. Cui, T. Tan, and Y. Wang, "Key techniques and methods for imaging iris in focus," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4. IEEE, 2006, pp. 557–561. 16, 36

[14] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 8, pp. 1529–1535, 2010. 16

[15] J. Daugman, "How iris recognition works," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 21–30, 2004. 16, 21, 29, 30, 31, 47, 48, 56

[16] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997. 16, 17, 21, 22

[17] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal identification based on iris texture analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1519–1533, 2003. 16, 17, 18

[18] C.-l. Tisse, L. Martin, L. Torres, M. Robert *et al.*, "Person identification technique using human iris recognition," in *Proc. Vision Interface*, 2002, pp. 294–299. 16, 17

[19] D. M. Monro, S. Rakshit, and D. Zhang, "Dct-based iris recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 586–595, 2007. 16, 17, 18

[20] H. Proença and L. A. Alexandre, "Iris segmentation methodology for non-cooperative recognition," in *Vision, Image and Signal Processing, IEE Proceedings-*, vol. 153, no. 2. IET, 2006, pp. 199–205. 16, 18

[21] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 9, pp. 1670–1684, 2009. 16, 19, 38, 43, 45

[22] C.-W. Tan and A. Kumar, "Efficient iris segmentation using grow-cut algorithm for remotely acquired iris images," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. IEEE, 2012, pp. 99–104. 17, 20

[23] F. Alonso-Fernandez and J. Bigun, "Iris boundaries segmentation using the generalized structure tensor. a study on the effects of image degradation," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. IEEE, 2012, pp. 426–431. 17, 20

[24] X. Huang, B. Fu, C. Ti, A. Tokuta, and R. Yang, "Robust varying-resolution iris recognition," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. IEEE, 2012, pp. 47–54. 17, 20

[25] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 11, pp. 1148–1161, 1993. 17, 20, 26, 35, 38, 52

[26] J. Daugman, "New methods in iris recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 5, pp. 1167–1175, 2007. 17

[27] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy< i> c</i>-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984. 18

[28] P. Viola and M. Jones, "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, 2001, pp. I–511. 19, 27, 38

[29] W. W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *IEEE transactions on signal processing*, vol. 46, no. 4, pp. 1185–1188, 1998. 21

[30] D. de Martin-Roche, C. Sanchez-Avila, and R. Sanchez-Reillo, "Iris recognition for biometric identification using dyadic wavelet transform zero-crossing," in *Security Technology, 2001 IEEE 35th International Carnahan Conference on*. IEEE, 2001, pp. 272–277. 21

[31] J. Kim, S. Cho, J. Choi, and R. J. Marks Ii, "Iris recognition using wavelet features," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, no. 2, pp. 147–156, 2004. 21

[32] L. Ma, Y. Wang, and T. Tan, "Iris recognition using circular symmetric filters," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 2. IEEE, 2002, pp. 414–417. 21, 22

[33] H. Proença, "A structural pattern analysis approach to iris recognition," in *Computer Recognition Systems 2*. Springer, 2007, pp. 731–738. 21, 23

[34] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–6. 23, 25, 26, 35, 48

[35] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 201–204. 24, 26

[36] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 582–585. 24

[37] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996. 24

[38] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893. 24

[39] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004. 24, 25

[40] P. E. Miller, A. W. Rawls, S. J. Pundlik, and D. L. Woodard, "Personal identification using periocular skin texture," in *Proceedings of the 2010 ACM Symposium on Applied Computing*.

ACM, 2010, pp. 1496–1500. 25, 26

[41] J. Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–8. 25

[42] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," in *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, 2012, pp. 439–445. 25, 26

[43] R. Jillela and A. Ross, "Mitigating effects of plastic surgery: Fusing face and ocular biometrics," in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. IEEE, 2012, pp. 402–411. 25, 26

[44] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Identifying useful features for recognition in near-infrared periocular images," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–8. 25, 26

[45] J. Adams, D. L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn, "Genetic-based type ii feature extraction for periocular biometric recognition: Less is more," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 205–208. 25, 26

[46] T. Beer, "Walsh transforms," *American Journal of Physics*, vol. 49, no. 5, pp. 466–472, 1981. 26

[47] K. I. Laws, "Rapid texture identification," in *24th Annual Technical Symposium*. International Society for Optics and Photonics, 1980, pp. 376–381. 26

[48] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. 100, no. 1, pp. 90–93, 1974. 26

[49] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989. 26

[50] D. J. Hurley, M. S. Nixon, and J. N. Carter, "A new force field transform for ear and face recognition," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1. IEEE, 2000, pp. 25–28. 26

[51] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008. 26

[52] D. A. Clausiyz and M. E. Jernigany, "Towards a novel approach for texture segmentation of sar sea ice imagery," 1996. 26

[53] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *Commu-*

nications, IEEE Transactions on, vol. 31, no. 4, pp. 532–540, 1983. 26

[54] B. J. Kang and K. R. Park, "A study on iris image restoration," in *Audio-and Video-Based Biometric Person Authentication*. Springer, 2005, pp. 31–40. 27, 29, 31, 38, 39

[55] N. Yager and T. Dunstone, "The biometric menagerie," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 2, pp. 220–230, 2010. 63

Appendix A

Anexos

A.1 Fusing iris and periocular information for cross-sensor recognition

Pattern Recognition Letters
journal homepage: www.elsevier.com

Fusing iris and periocular information for cross-sensor recognition

Gil Santos[a,**], Emanuel Grancho[a], Marco V. Bernardo[a,b], Paulo T. Fiadeiro[b]

University of Beira Interior, Covilhã, Portugal

[a] *Department of Computer Science, IT - Instituto de Telecomunicações*
[b] *Department of Physics, Remote Sensing Unit - Optics, Optometry and Vision Sciences Group*

ABSTRACT

Over the last years the usage of mobile devices has substantially grown, along with their capabilities and applications. Extending biometric technologies to such gadgets is quite desirable, as it would represent the ability to perform biometric recognition virtually anytime, anywhere, and by everyone. This paper focus on biometric recognition on mobile environments using the iris and periocular information as main traits, and its main contributions are three-fold: 1) announce the availability of an iris and periocular dataset containing images acquired with 10 different mobile setups, along with the corresponding iris segmentation data. Such dataset allows to evaluate both iris segmentation and recognition methods, as well as periocular recognition techniques; 2) report the outcomes of device-specific calibration techniques that compensate for the different color perception inherent to each setup; 3) propose the application of well-known iris and periocular recognition strategies, based on classical encoding and matching techniques, giving evidence on how they can be fused to overcome the issues associated with mobile environments.

1. Introduction

The evolution of biometric systems over the last years is notorious, with the appearance of new traits and algorithms and the refinement of the existing ones. At the same time that the acquisition constraints are being lowered favoring *in-the-wild* operation, efforts are being put into delivering *off-the-shelf* solutions for everyday consumers, so that biometric systems can run easily on everyday electronics. Mobile devices in particular are preferable targets, as they comprise all the necessary components to carry the whole process, from trait acquisition to the final decision.

From the existing traits, the face and the iris are present in the literature among the most popular (along with the fingerprint) (Bowyer et al., 2008; Zhao et al., 2003). Iris usage as main biometric trait has remained stable despite the evolution of biometrics in the last years. Being a naturally protected organ, visible from the exterior and allowing contact-less acquisition, its circular and planar shape that favors detection and segmentation, and its predominantly randotypic appearance that assures high recognition effectiveness. There are, however, certain scenarios where the iris cannot be properly imaged, and where the complementary use of other ocular information is regarded as a good way to compensate for unreliable iris acquisition – periocular biometrics.

Particular useful on unconstrained scenarios, the periocular region does not require constrained capturing or complex imaging systems, being fairly easy for a mobile user to operate a periocular identification application. The grounds for periocular recognition came from human intrinsic ability to recognize someone just by looking at his/her eyes, which are known to provide substantial amounts of discriminant information whilst remaining relatively stable over large periods of time. Periocular biometrics analyze not only iris structure, but also other surrounding features, such as the shape of eyelids, eyelash distributions, or sclera and skin texture information. At last, both the iris and the periocular region are imaged simultaneously with a single camera.

1.1. Contextualization: Iris biometrics

The commercially deployed iris recognition systems are mainly based on Daugman (1993) pioneering approach, with great effectiveness in relative constrained scenarios, and with data acquired in the near-infrared (NIR) slice of the electromagnetic spectrum (700-900 nm). Even that a few innovations were introduced later on (Daugman, 2007), the process consists in a three stage approach: 1) the segmentation of the iris

** Corresponding author: Tel.: +351-275-242081; Fax : +351-275-319899;
e-mail: gmelfe@ubi.pt (Gil Santos)

boundaries (both pupillary and limbic) followed by the translation into a double dimensionless pseudo-polar coordinate system to achieve invariance to scale and translation; 2) the convolution of this normalized data with a set of Gabor filters at multiple frequencies and orientations and the corresponding output quantized to one of four quadrants, extracting two bits of phase information per convolution; 3) matching of the iris signatures using the fractional Hamming distance, with several comparisons of shifted data to achieve invariance to rotation.

In addition to Daugman's phase-based approach other iris recognition variants were introduced, mainly zero-crossing and texture-analysis methods: Boles and Boashash (1998) computed the zero-crossing representation of a 1D wavelet at different resolutions of concentric circles, and Wildes (1997) proposed the characterization of the iris texture through a Laplacian pyramid with four different levels.

Efforts on "relaxing" the acquisition setup are also registered, being the "iris-on-the-move" project (Matey et al., 2006) a major example on engineering a less intrusive system for subjects: its goal is to acquire near-field NIR iris images as the subjects walk through an access control point.

1.2. Contextualization: Periocular biometrics

The usage of the periocular region as a biometric trait has emerged over the last years (Santos and Proença, 2013). The first relevant studies on periocular biometrics can be traced back to Park et al. (2009) and their pioneering approach, fusing both local and global features from the ocular area. On global feature extraction images were aligned using iris center as anchoring point, and a 7×5 region of interest (ROI) grid defined around it. Scale invariance was achieved using iris radius as side length for the ROI. Those patches were then encoded applying two well known distribution-based descriptors, Local Binary Patterns (LBP) (Ojala et al., 1994) and Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005), quantized into 8-bin histograms. Merging those histograms into a single-dimension array containing both texture and shape information, matching was carried off simply by computing an Euclidean distance. For the local analysis, authors employed Scale-Invariant Feature Transform (SIFT) (Lowe, 2004), allowing sets of key-points to be extracted, encoded with their surroundings, and matched, while providing translation, scaling and rotation invariance. Reported performance was fairly good, showing periocular fitness for recognition purposes, and further analysis was held on noise factors impact on performance (Park et al., 2011).

Inspiring by their work other approaches arose, either by improving Park *et al.* approach, or by introducing new perspectives. Miller et al. (2010) presented an analysis also focused on periocular skin texture, taking advantage of Uniform Local Binary Patterns (ULBP) to achieve "improved rotation invariance with uniform patterns and finer quantization of the angular space" (Ojala et al., 2002). Later on, their work was extended by Adams et al. (2010), who proposed using Genetic & Evolutionary Computing (GEC) to optimize feature set. Juefei-Xu et al. (2010) stressed many local and global feature extraction techniques (Walsh (Beer, 1981) and Laws' masks (Laws, 1980), Discrete Cosine Transform (DCT) (Ahmed et al., 1974), Discrete Wavelet Transform (DWT) (Mallat, 1989), Force Fields (Hurley et al., 2000), Speed-Up Robust Transform (SURF) (Bay et al., 2008), Gabor filters (Clausi and Jernigan, 1996) and Laplacian of Gaussian (LoG)), and on their later work (Juefei-Xu et al., 2011) efforts were made to compensate aging degradation effects on periocular performance. The possibility of score-level fusion with other biometric traits was also addressed (e.g. iris (Woodard et al., 2010)).

Bharadwaj et al. (2010) proposed the fusion of ULBP with five perceptual dimensions, usually applied as scene descriptors: naturallness, openness, roughness, expansion and ruggedness – GIST (Oliva and Torralba, 2001). Images were preprocessed with Fourier transform for local contrast normalization, and then a spatial envelope computed with a set of Gabor filters (4 scales $\times$ 8 orientations). On the final stage, χ^2 distance was used to match the feature arrays, and results fused with a weighted sum. This approach was validated against UBIRIS.v2 data (Proença et al., 2010), simulating realistic unconstrained acquisition setups.

1.3. The Mobile Constraints

When attempting to perform iris or periocular biometrics on mobile environments, several problems arise: the wide variety of camera sensors and lenses mobile phones and tablets come equipped with produce discrepancies in working images, as they are acquired with color distortions, at multiple resolutions, etc.; *on-the-go* acquisition by potentially untrained subjects will result in demanding Pose, Illumination and Expression (PIE) changes, as not all users hold their mobile devices at the same position, resulting in varying acquisition angles and scales, or rotated images; the acquisition environment can have poor or insufficient lighting, and uncontrolled outdoor daylight will most likely produce spectacle reflections over the iris region; etc.

The remainder of this paper is organized as follows: Section 2 describes the Cross-Sensor Iris and Periocular Dataset (CSIP) database, detailing the acquisition conditions, enrolled participants and perceived noise factors; Section 3 presents the proposed methodology, with details on the four main stages: image normalization with device-specific color calibration, iris and periocular feature encoding and matching, and score-level fusion; Section 4 contains a thorough analysis of the results obtained by using the proposed methodology; finally, Section 5 states some final considerations, along with further lines of work.

2. The Cross-Sensor Iris and Periocular Dataset

The main objective of the CSIP database was to gather images from a representative group of participants, acquired over cross-sensor setups and varying acquisition scenarios, thus mimicking the conditions faced on mobile application scenarios. Along with the data acquired with different mobile devices, an iris segmentation mask is also provided, allowing assessing the performance of both iris and periocular segmentation and recognition algorithms on mobile environments.

Table 1: Details of the devices and setups used during the CSIP dataset acquisition.

Device	A		B			C			D	
Manufacturer	Sony Ericsson		Apple			ThL			Huawei	
Model	Xperia Arc S		iPhone 4			W200			U8510	
O.S.	Android 2.3.4		iOS 7.1			Android 4.2.1			Android 4.3.3	
Camera	Rear		Frontal	Rear		Frontal	Rear		Frontal	Rear
Resolution	3264×2448		640×480	2592×1936		2592×1920	3264×2448		640×480	2048×1536
Flash	No	Yes	No	No	Yes	No	No	Yes	No	No
Setup ID	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0

2.1. The Imaging Setup

Considering the heterogeneity of camera sensor/lens setups consumer mobile devices can deliver, a total of 10 different setups were used during the dataset acquisition stage: four different devices, some of them with frontal and rear cameras and LED flash (Table 1). Each participant was imaged at all the considered setups.

Aiming at mimicking the variability of noise factors associated with *on-the-go* recognition, participants were not imaged at a single particular location, but on multiple sites, *as they were*, with artificial, natural and mixed illumination conditions. As we can see from Figure 1, there is a substantial difference between each acquisition setup and surrounding conditions, even when the same setup was used to capture images from different subjects. From visual inspection, eight different noise factors are distinguishable, and can affect the biometric recognition process: multiple scales; chromatic distortions; image rotation; poor lighting; off-angle acquisition; out-of-focus images; deviated gaze; and iris obstructions (including reflexions).

The images were acquired through the standard camera application on mobile phone devices, using default settings for both focus and white-balance. The corresponding files were stored at JPEG format, with the highest possible quality and resolution. A total of 50 participants were enrolled, all Caucasian and most of them male (82%), with ages comprehended between 21 and 62 years old (31.18 ± 9.93). All the participants gave informed consent about the experiment.

2.2. Iris segmentation masks

For each periocular image acquired by the mobile devices, a binary iris segmentation mask is provided with the CSIP dataset. Those masks were automatically obtained using the state-of-the-art iris segmentation approach proposed by Tan et al. (2010). That approach is particularly suitable for uncontrolled acquisition conditions, which has been corroborated by the first place achieved at the Noisy Iris Challenge Evaluation - Part 1 (NICE.I)[1].

At a first stage, a small ROI containing a rough estimate of the iris location is defined. This ROI is determined using a cascade object detector based on Viola and Jones (2001) algorithm, trained for the detection of the right eye using Haar features to encode details (Castrillón et al., 2007). A reflexion removal process is applied, followed by an eight-neighbor connection approach for clustering, and based on the degree of similarity between each pixel and the previously established heterogeneous regions, a region set is established accordingly to the degree of similarity between their elements. In order to label the different clusters, semantic refinements are applied. Several semantic priors like orientation and shape of each region are used to determine the iris correspondent cluster.

Further to that, iris pupillary and limbic boundaries are estimated using an integrodifferential-constellation: based on Daugman (2007) integrodifferential, a constellation is built from several integrodifferential rings of increasing radii, minimizing the initial method's tendency to output local optimal

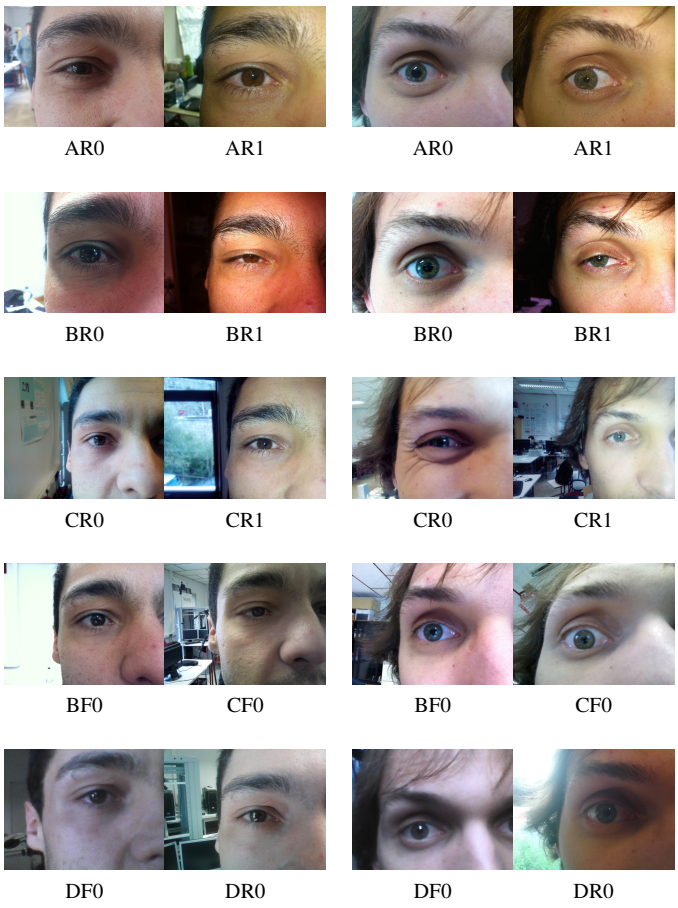


Figure 1: Dataset pictures acquired from two participant at all different setups. Images in the left belong to the first participant, and images in the right belong to the second participant.

[1] http://nice1.di.ubi.pt/

solutions. Possible localization inaccuracies are detected and eliminated based on a threshold, estimated by the intersection of two consecutive annular rings intensity distributions. After iris boundaries are established, eyelid localization and shadow subtraction are performed in order to reduce noise and occlusion. The presence of eyelashes is minimized through 1-D filtering, and edge-detection is applied to find the edge points corresponding to eyelids. Using those edge-points, the localization of the eyelids is estimated using both an upper and lower statistically established curvature model. Ultimately, eyelash and shadow subtraction take into account their darker appearance when compared to the iris itself. The optimization of the classification threshold is obtained from the analysis of the intensity histogram of small homogeneous regions, on iris and shadows noise, and eyelash regions are removed. This technique is explained in more detail at Tan et al. (2010).

2.3. Dataset Availability

The complete CSIP dataset is public and freely available for academic and research purposes[2]. Researchers are granted access to: 1) 2004 images, acquired from 50 subjects at 10 different setups; and 2) the corresponding 2004 binary iris segmentation masks.

3. Proposed Methodology

In this section we describe the four main steps of our approach (Figure 2): the normalization stage, with device-specific color correction and iris boundaries estimation for coordinate conversion and periocular ROI definition; feature encoding, with information from both the iris and the periocular region; feature matching; and score-level fusion.

3.1. Image Normalization

The first stage, image normalization, will allow to compensate for some of the noise factors identified in the dataset: chromatic distortions, varying scales and off-angle acquisition.

3.1.1. Device-Specific Color Correction

Having an uniform calibrated output for each sensor that minimizes the discrepancy to colors as they really appear can be of particular value in mobile scenarios, as a wide range of sensor/lens setups are available.

The access to a reference image captured at a known illuminant allows to estimate the color adaptation matrix that compensate for the inaccurate color representation introduced by each sensor. That adaptation matrix encodes the optimized color channels combination to approximate color information in the acquired image from the ones originally observed in the scene. This section describes the device-specific color correction technique.

A Macbeth ColorChecker® Color Rendition Chart was placed in a dark acquisition scene, illuminated by a standard illuminant produced by a *Barco RLM G5i Performer* (Barco Corporation, Belgium) RGB projector driven by a *Visual Stimulus Generator (VSG2/5)* (Cambridge Research Systems, United Kingdom). In order to mimic standard open-air conditions, the Commission Internationale de l'Eclairage (CIE) D65 illuminant was chosen, as specified by the CIE standard colorimetric observer (2°) (on Illumination, 2004; Smith and Guild, 1931). Illuminants' luminance was regulated at 100 cd/m^2.

Previously, the VSG2/5 generated stimulus were verified and calibrated using a telespectroradiometer (PR-650 *SpectraColorimeter*™- Photo Research, Inc., CA) and a white reference Spectralon® target (Labsphere, Inc., NH). The maximum errors allowed were 0,002 illuminant chromaticities in the CIE 1931 color space and 1 cd/m^2 for luminance. A set of images of the color charts was then captured at all setups (mobile devices), using the standard camera application at default settings.

To obtain the estimate for the color correction matrix we applied the methodology introduced by Wolf (2003), specially designed for digital imaging systems. Knowing the ground-truth red (R), green (G) and blue (B) coordinates for the 24 color samples from the color chart under the D65 illuminant, let us summarize it in a 24×3 matrix O (1). Then, from the color chart photo acquired with the mobile device, we populate a similar matrix P with the RGB coordinates for the same 24 color samples.

$$O = \begin{bmatrix} O\_R_1 & O\_G_1 & O\_B_1 \\ O\_R_2 & O\_G_2 & O\_B_2 \\ \dots & \dots & \dots \\ O\_R_{24} & O\_G_{24} & O\_B_{24} \end{bmatrix} \quad (1)$$

The initial estimate for the adaptation matrix A, that converts the device acquired colors to an approximation $\hat{O}$ of the original

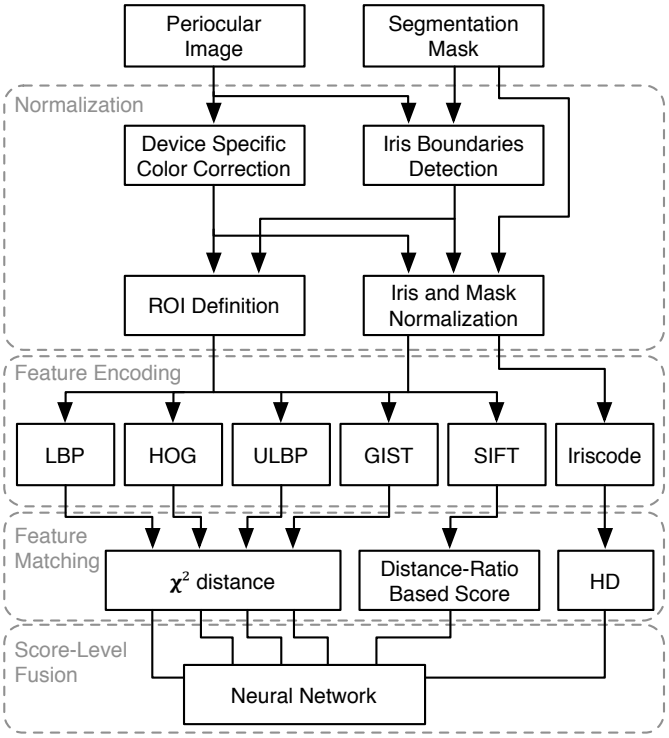


Figure 2: Diagram illustrating the four stages of the proposed methodology.

[2] http://csip.di.ubi.pt

ones, was then found using a least-squares solution (2) where $\underline{1}$ is a 24 positions column vector initialized with ones.

$$\begin{aligned} O \approx \hat{O} = [\underline{1}\ P]A \Leftrightarrow \\ \Leftrightarrow A = ([\underline{1}\ P]^T[\underline{1}\ P])^{-1}[\underline{1}\ P]^T\ O \end{aligned} \tag{2}$$

Further optimization of the adaptation matrix was achieved by applying the following four steps iteratively, until convergence up to the fourth decimal place: 1) compute a cost vector $\underline{C}$ based on the Euclidean distance $\underline{E}$ to the ground-truth color information (3), where ϵ is the relative weight for misfit points; 2) normalize $\underline{C}$ for unity norm, and compute $\underline{C}^2$; 3) generate an empty 24×24 matrix C^2, and populate its diagonal with $\underline{C}^2$; 4) recompute the adaptation matrix, using cost-weighted least-squares fitting (4).

$$\underline{C} = \frac{1}{\underline{E} + \epsilon} \tag{3}$$

$$A = ([\underline{1}\ P]^T C^2 [\underline{1}\ P])^{-1}[\underline{1}\ P]^T C^2 O \tag{4}$$

An adaptation matrix was computed for each acquisition setup (mobile device camera). Prior to the feature extraction stages each one of the dataset images was color corrected, accordingly to its acquisition device and setup, using the corresponding adaptation matrix followed by a non-linear transform.

3.1.2. Iris Boundaries Detection

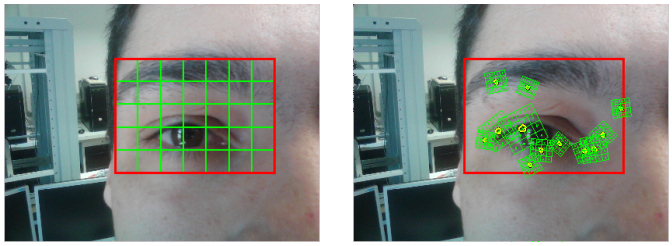
Accurately determining the iris boundaries is a requirement for the following steps, iris and segmentation mask conversion to a pseudo-polar coordinate system and periocular ROI definition), as that will allow to achieve image alignment and scale invariance.

To determine the iris boundaries, the information from both the device acquired image and the binary segmentation mask were combined using a three step approach (Santos and Hoyle, 2012): a) a Hough transform (Ballard and Brown, 1982) is fit to the binary mask boundaries, determining the circle best fitting iris limbic contours; b) a smaller circular ROI is defined of the acquired image, centered in the previously located limbic circle and with 2/3 its radius. Such region is converted to grayscale, its histogram equalized, and an edge map extracted using a Canny edge detector (Canny, 1986); c) a second circle fit to the resulting edge map using another Hough transform, thus approximating the pupillary boundaries.

3.1.3. Iris and Segmentation Mask Normalization

Knowing the iris boundaries, each iris pixel I was assigned to a pair of real coordinates over a double dimensionless pseudopolar coordinate system (5). We followed the rubber-sheet model originally proposed by Daugman (2004) (6), where r and θ are the radius and angle respectively, $x(r,\theta)$ and $y(r,\theta)$ linear combinations of both the set of pupilar boundary points $(x_p(\theta), y_p(\theta))$ and the set of boundary points $(x_s(\theta), y_s(\theta))$ bordering the sclera.

$$I(x(r,\theta), y(r,\theta)) \Rightarrow I(r,\theta) \tag{5}$$



(a) Grid of individual patches (b) SIFT detected features

Figure 3: Illustration of the ROI defined for the global periocular analysis (red), the set of patches used on the distribution-based analysis (a), and SIFT detected features (b).

$$\begin{aligned} x(r,\theta) = (1-r)x_p(\theta) + rx_s(\theta) \\ y(r,\theta) = (1-r)y_p(\theta) + ry_s(\theta) \end{aligned} \tag{6}$$

3.1.4. ROI Definition

To carry on with the periocular analysis, a ROI is defined based on the known iris spatial location (x_i, y_i) and radius (r_i). That ROI is composed by 35 square patches, forming a 7×5 grid, where each patch has an area equivalent to $1.4r_i^2$ (Figure 3).

3.2. Feature Encoding and Matching

At the feature encoding and matching stages, information from two different biometric traits were handled as described below: iris and periocular. On the methods designed to work with single channel images, RGB values were converted to grayscale using a weighted sum (7) prior to feature extraction. The weights in equation (7) are the ones used as standard in National Television System(s) Committee (NTSC) colorspace conversion for computing the effective luminance of a pixel.

$$I(x,y) = 0.2989\mathrm{R}(x,y) + 0.5870\mathrm{G}(x,y) + 0.1140\mathrm{B}(x,y) \tag{7}$$

3.2.1. Periocular Feature Analysis

The periocular analysis here proposed was inspired on the works of Park et al. (2009) and Bharadwaj et al. (2010). In the previously defined ROI, two types of analysis were used: and a distribution-based analysis of every patch, and a global analysis of the whole region.

The distribution-based analysis consists in the computation of three well-known descriptors: HOG, LBP and ULBP. Each descriptor is computed sequentially over each patch and quantized into histograms, forming a global 1-D array where shape and texture information is stored. The HOG descriptor (Dalal and Triggs, 2005), widely applied on computer vision, computes the gradient orientation by filtering the image with two kernels: $[-1, 0, 1]$ and $[-1, 0, 1]^T$. The LBP (Ojala et al., 1994) also works in a quite simple yet efficient fashion: pixel intensity changes from an 8-neighbor region to its central pixel are quantized (8) having the sign of their intensities' difference (9) as reference. $I_{x,y}$ denotes the intensity of the original image at position (x, y), and I_n the intensity of a neighbor pixel.

$$\mathrm{LBP}_{x,y} = \sum_{n=0}^{7} \mathrm{sgn}(I_n - I_{x,y})\, 2^n \tag{8}$$

$$\mathrm{sgn}(I_n - I_{x,y}) = \begin{cases} 1, & \text{if } I_n \geq I_{x,y} \\ 0, & otherwise. \end{cases} \tag{9}$$

The ULBP descriptor differs from the LBP as it achieves "improved rotation invariance with uniform patterns and finer quantization of the angular space" (Ojala et al., 2002). Instead of the 2^n possible binary patterns outputted from the regular LBP over a 8-neighbor region, a uniformity measure U is calculated representing the number of bitwise changes in that same pattern (10). This measure can only assume 59 distinct values.

$$\begin{aligned} U(\mathrm{LBP}_{x,y} = {} & |\mathrm{sgn}(I_7 - I_{x,y}) - \mathrm{sgn}(I_0 - I_{x,y})| + \\ & + \sum_{n=1}^{7} |\mathrm{sgn}(I_n - I_{x,y}) - \mathrm{sgn}(I_{n-1} - I_{x,y})| \end{aligned} \tag{10}$$

At the matching stage, the histogram arrays of size N containing the extracted information were compared through χ^2 distance (11).

$$\chi^2_{(\mathrm{hist}_A, \mathrm{hist}b)} = \frac{1}{2} \sum_{n=1}^{N} \frac{(\mathrm{histA}_n - \mathrm{histB}_n)^2}{\mathrm{histA}_n + \mathrm{histB}_n} \tag{11}$$

On the global analysis, feature extraction techniques were applied not to each individual patch, but to the whole ROI. The applied descriptors were SIFT, and GIST. At first, set of key-points and their surrounding information is extracted using SIFT (Lowe, 2004), known to deliver invariance to translation, scale and rotation. SIFT key-points detection relies on a Difference of Gaussians (DOG) function, and features are extracted for their neighborhood based on gradient magnitude and orientation (Figure 3). At the matching stage, their geometrical alignment is used. Finally, a set of five scene descriptors were used (GIST) as proposed by Oliva and Torralba (2001): *naturalness*, that quantifies vertical and horizontal edge distribution); *openness*, as the presence or lack of reference points; *roughness*, the size of the largest prominent object; *expansion*, the depth of the space gradient; and *ruggedness*, a quantification of contour orientation that assesses the deviation from the horizontal. The GIST descriptor was extracted from each color channel individually, and at the matching stage a χ^2 distance (11) was used upon min-max normalization.

3.2.2. Iris Feature Analysis

The iris information was encoded based on Daugman (1993) approach: iris features were extracted convolving iris data in the pseudopolar coordinate system with a bank of 2-D Gabor wavelets, followed by a quantization stage that produced a binary *iriscode* accordingly to the sign of the 2-D integral. To the purpose of iris identification on mobile environments, we choose to use a very small yet optimized wavelet bank. During filter optimization a smaller representative subset of images was used, and filter parameters cycled through a range of scales, orientations and frequencies, fit for our environment. Chosen configurations were the ones that maximized decidability (13).

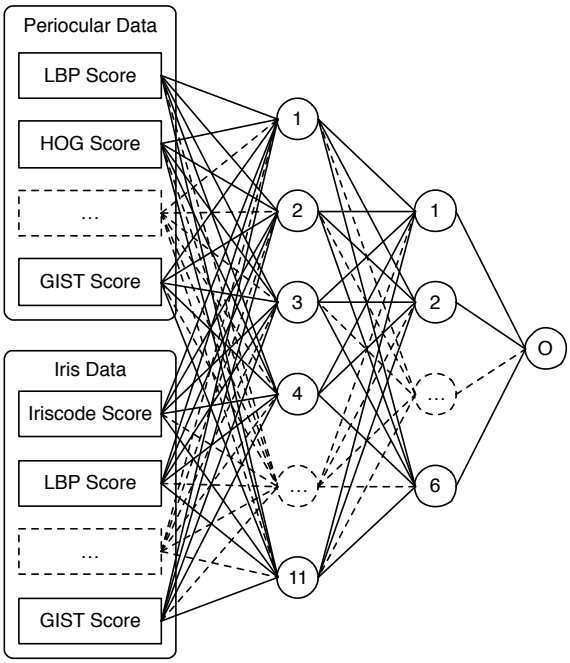


Figure 4: Illustration of the NN architecture used at the score fusion stage. Each circle represents a neuron of the network, and depicted input scores come from the feature matching stage.

At the matching stage, the similarity between two binary codes of size N representing the two irises being compared is assessed through a simple Hamming Distance (HD) (12).

$$\mathrm{HD} = \frac{1}{N} \sum_{n=1}^{N} \mathrm{codeA}_n \otimes \mathrm{codeB}_n \tag{12}$$

Further to that, and as Daugman's technique was developed to deal with iris images captured in controlled settings, the same techniques used to encode periocular data were also applied on the normalized iris region as well.

3.3. Score-level fusion

With several scores resulting from the different encoding/matching methodologies, an Artificial Neural Network (NN) was trained to fuse them into a final recognition score. NN-based methods have been widely applied on classification problems, for their learning abilities and good generalization.

For the purpose of this work a two hidden layers NN was trained with back-propagation (Figure 4). The architecture of the NN was as follows: the first hidden layer had eleven neurons, the same number of scores resulting from the matching stage; the second hidden layer had six neurons; and the final (output) layer with just one, since we are dealing with a binary classification problem. Once again, a smaller data partition was used at the training stage, and was not included on the test phase.

Table 2: Individual performance metrics for each recognition method and trait, along with the ones from iris, periocular and global fusion. Performance metrics are Decidability (DEC), Area Under Curve (AUC) and Equal Error Rate (EER). Top scores are marked *bold*.

	Trait →	*Periocular*						*Iris*							Global
	Method →	LBP	HOG	SIFT	ULBP	GIST	Fusion	Iriscode	LBP	HOG	SIFT	ULBP	GIST	Fusion	Fusion
No color correction	DEC	0.989	0.969	0.716	1.272	1.859	2.164	0.674	0.289	0.515	0.324	0.324	0.320	0.835	2.295
	AUC	0.764	0.751	0.715	0.816	0.915	0.923	0.684	0.588	0.641	0.583	0.589	0.615	0.717	0.932
	EER	0.308	0.315	0.348	0.261	0.166	0.159	0.366	0.443	0.401	0.443	0.440	0.418	0.344	0.148
Histogram equalization	DEC	0.986	0.860	0.668	1.267	1.841	2.101	0.616	0.246	0.371	0.293	0.353	0.199	0.753	2.215
	AUC	0.763	0.725	0.696	0.815	0.910	0.917	0.669	0.581	0.601	0.576	0.600	0.582	0.696	0.925
	EER	0.309	0.340	0.365	0.262	0.172	0.165	0.374	0.446	0.433	0.445	0.433	0.442	0.361	0.155
Device specific correction	DEC	0.989	1.009	0.731	1.270	1.889	2.215	0.639	0.173	0.482	0.347	0.266	0.230	0.809	**2.331**
	AUC	0.766	0.761	0.720	0.817	0.919	0.927	0.675	0.578	0.637	0.590	0.584	0.593	0.711	**0,934**
	EER	0.305	0.308	0.343	0.259	0.163	0.155	0.372	0.450	0.402	0.437	0.444	0.434	0.349	**0,145**

4. Results and Discussion

To assess our method performance, a total of 121.245 random matches were generated, between images from any two acquisition setups, being the inter- to intra-class comparisons ratio 2:1. Three performance measures were used: DEC, AUC and EER. Decidability d' was first introduced by Daugman (1993), and quantifies *intra-* and *inter*-class separability by relating their mean μ and standard deviation σ values.

$$d' = \frac{\|\mu_{\text{inter}} - \mu_{\text{intra}}\|}{\sqrt{\frac{\sigma^2_{\text{inter}} + \sigma^2_{\text{intra}}}{2}}} \tag{13}$$

The Receiver Operating Characteristic (ROC) curve relates the sensitivity, or true positive rate (TPR) with the false positive rate (FPR). Based in that plot, the AUC can be perceived as a quantification of how well pairwise comparisons are performed on a binary classification problem. On the perfect scenario, all positive matches are ranked higher than the negative ones, and the AUC equals one. Finally, setting the operating threshold for the accept/reject decision so that the probability of false acceptance equals the probability of false rejection, we obtain the EER.

The performance registered for every feature encoding and matching technique, *per* trait behavior and global fusion outcome is registered on Table 2. As we can see top results are registered for global fusion, over device-specific color corrected images, with a Decidability of 2.331, and an AUC of 0.934. On quantifying the performance improvement introduced by the sensor-specific color correction technique, we applied our proposed methodology to all matches with three variations in the normalization stage: without performing any color-correction; with the proposed color correction; and, for comparison purposes, with simple histogram equalization over all channels of the working image. As we can see, the proposed color correction technique improves the performance over all the stressed periocular recognition approaches, as well as on the final score resulting from the fusion of all methods. It was a much better approach than the commonly used histogram equalization, whose application actually worsened five of the six periocular approaches, as well as the score-fusion output. Even so, the improvement produced by the sensor-specific correction was not so expressive as initially expected after visually inspecting the color corrected images. A possible explanation is that some of the used feature encoding methods were designed to work over single-channel images, thus not implicitly taking into account some of the chromatic features that could have been lost during grayscale conversion.

Examining *per* trait performance, and paying attention to the values obtained over color-corrected images, we can see how the combined information extracted from the whole periocular region is far more discriminant than the iris, in the mobile application scenario. That is particularly visible in the ROC curves at Figure 5, where we can see that the plot corresponding to the periocular fusion almost overlaps the plot corresponding to the global fusion, being the area between them of only 0.007. In fact, we can't say that periocular analysis does not take iris features into account, as it was not removed nor overlapped prior to the encoding stage.

Reviewing the individual performance of each one of the methods that constitute the proposed periocular analysis (Figure 6a), we can see how GIST descriptors is the approach with

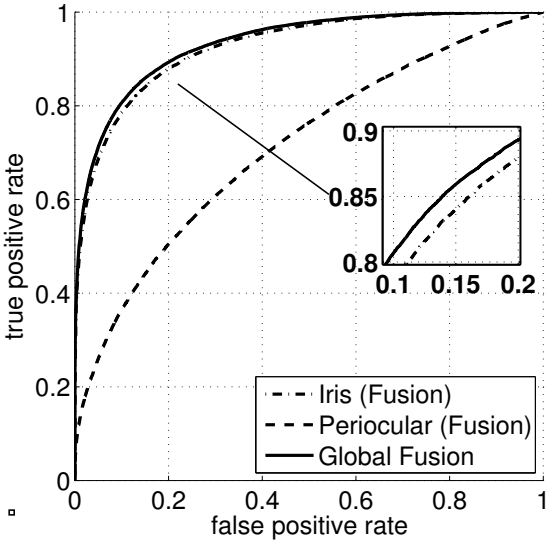
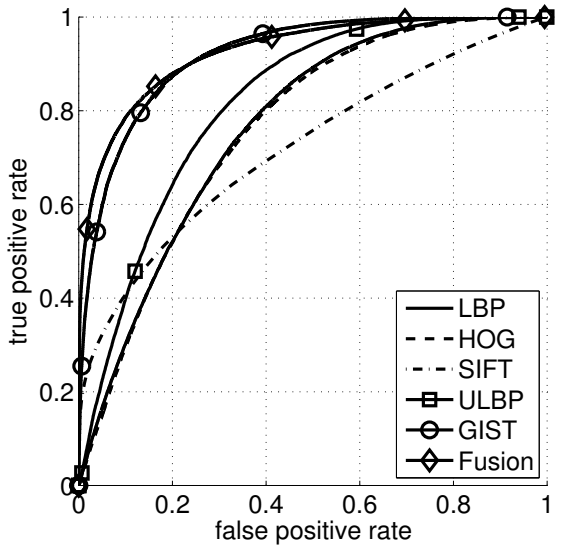


Figure 5: Receiver Operating Characteristic curves for the score-level fusion of the stressed iris recognition methods, the periocular recognition methods, and the global fusion.

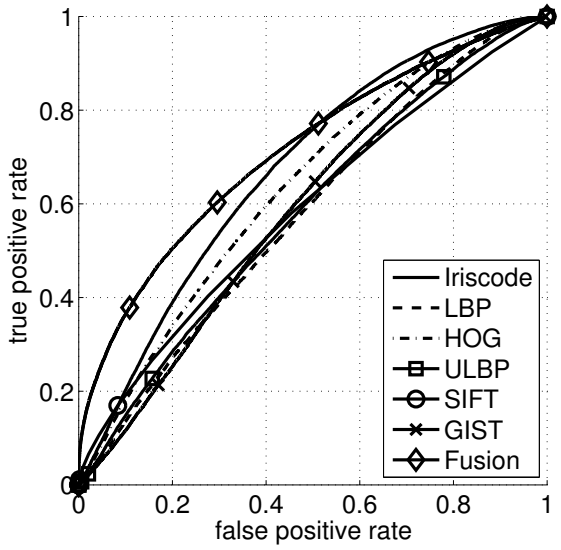
Table 3: Method fusion performance, after color correction, for each acquisition setup.

		AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0
No color correction	DEC	2.481	2.392	2.141	2.456	2.137	2.045	2.153	2.459	2.083	2.423
	AUC	0.940	0.941	0.917	0.940	0.933	0.917	0.922	0.939	0.916	0.938
	EER	0.135	0.135	0.158	0.138	0.149	0.169	0.162	0.138	0.167	0.139
Device specific correction	DEC	2.497	2.446	2.174	2.485	2.164	2.034	2.201	2.561	2.099	2.501
	EER	0.941	0.943	0.921	0.942	0.935	0.917	0.927	0.945	0.917	0.943
	AUC	0.134	0.132	0.156	0.138	0.145	0.165	0.155	0.130	0.166	0.131

highest benefits, followed by ULBP. LBP and HOG have very similar performance, and SIFT was the descriptor with lowest AUC, even though it was able to achieve higher sensitivity with lower FPR for more restrictive thresholds than most of the stressed periocular methods. Since we are aiming at performing biometric recognition in mobile devices, known to have more resources constraints than regular computers, that can be regarded as a good indicator: since SIFT is more computationally expensive than the other tested methods, we can choose not to include it with less impact on the overall performance. Even if only GIST were used, with its five scene descriptors being easily and quickly computed, we could still get an AUC of 0.919. Nonetheless, it is remarkable how such simple feature encoding techniques produce relatively good scores, considering the constraints associated with the mobile working conditions, specially the deterioration of the acquired images. Attending at the same methods' performance over iris data alone (Figure 6b), we can observe that they are not so good at discriminate its features, being the individual method with best performance Daugman's iriscode analysis. We must have in mind that the CSIP acquisition setups didn't favor the capturing of iris details.



(a) Periocular Recognition



(b) Iris Recognition

Figure 6: Receiver Operating Characteristic curves for the stressed periocular (a) and iris (b) recognition methods and their fusion.

Table 3 reports on the recognition performance for when using images acquired at the different imaging setups, with and without performing color correction. Those values were obtained selecting from the total of generated matches the ones where at least one of the images was enrolled at that specific setup. We can see how color correction considerably improves the decidability values on the top performing devices. As we can see, top performances are achieved over images taken with rear cameras, usually without using the device flash. In fact, choosing to use the built-in flash tends to result in performance degradation, even that color correction impact in performance was greater when using images acquired with the flash light on. As for frontal cameras, they do not seem as fit for mobile biometrics as rear ones. That can be particularly tough if the intuit of the application is to verify the phone's user identity, since its fairly more easily obtain a good self-captured image using the frontal camera. Despite the frontal cameras having significantly less resolution that the rear ones, it does not seem to be a relation between that fact and their lack of performance. Device D, for instance, is the device with lower rear sensor resolution and its performance is almost identical to the other setups.

5. Final Considerations

This paper introduces the Cross-Sensor Iris and Periocular Dataset (CSIP) dataset, containing images acquired under ten different mobile setups, with eight visible noise factors. Such particularities make it suitable to evaluate iris and periocular

recognition methods on heterogeneous mobile conditions, and the distribution of each image iris segmentation masks also allow to stress iris segmentation methods on those same conditions.

Further, we identify the chromatic disparity introduced by some devices, proposing the usage of a sensor-specific color correction technique. Results shown that top results were obtained after color correction. Being aware that application scenarios where the biometric recognition process is conducted with images acquired on a single device can deliver better results, we aim at achieving higher cross-sensor performance. That will allow to attain higher confidence on matches between images acquired with very distinct mobile setups, or even on comparisons against a previously stored dataset acquired with other (or multiple) devices.

We proposed the fusion of iris and periocular information to achieve reliable biometric identification in mobile setups, and observed how simple feature encoding techniques deliver considerably good performance. That is particularly convenient when aiming at conduct the whole recognition performance on mobile devices with higher computational constraints, as the top performing methods indeed had considerably low computational cost. Ultimately, and if aiming at reducing even more the computational cost on mobile environments, using only the GIST classifier can be an option.

Finally, results point out that, for the tested setups, high image resolution is not an essential requisite to mobile biometrics, and rear cameras are best suited for periocular recognition, preferable without flash.

5.1. Further Work

At a further stage, authors plan to expand the CSIP dataset with a more significant amount of participants, and a wider range of acquisition setups. We theorize that widening the dataset to further devices and participants, and applying the proposed recognition technique, would emphasize the contribution of both color correction and the usage of iris features.

Another interesting line of work will be to conduct further tests with different iris and periocular recognition methods, specially the ones that explicitly rely on color data, comparing the cross-sensor performance with the performance registered for other existing datasets.

Stressing different color correction techniques could also be interesting, despite the one we applied in this paper having the advantage of being computational efficient and easy to apply, as long as the correction matrix for the camera sensor is known.

Acknowledgments

The authors would like to acknowledge the financial support provided by *FCT - Fundação para a Ciência e Tecnologia* through the research grant SFRH/BD/80182/2011, and the funding from 'FEDER - QREN - Type 4.1 - Formação Avançada', co-founded by the European Social Fund and by national funds through Portuguese 'MEC - *Ministério da Educação e Ciência*'. This work was also supported by the RSU - Remote Sensing Unit through 'PEst-OE-FIS/UI0524/2014' with funds provided by the FCT.

References

Adams, J., Woodard, D., Dozier, G., Miller, P., Bryant, K., Glenn, G., 2010. Genetic-based type ii feature extraction for periocular biometric recognition: Less is more, in: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 205 –208. doi:10.1109/ICPR.2010.59.

Ahmed, N., Natarajan, T., Rao, K., 1974. Discrete cosine transform. Computers, IEEE Transactions on C-23, 90 –93. doi:10.1109/T-C.1974.223784.

Ballard, D., Brown, C., 1982. Computer Vision. Englewood Cliffs, NJ: Prentice-Hall, E.U.A.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (surf). Comput. Vis. Image Underst. 110, 346–359. URL: http://dx.doi.org/10.1016/j.cviu.2007.09.014, doi:10.1016/j.cviu.2007.09.014.

Beer, T., 1981. Walsh transforms. American Journal of Physics 49, 466–472. URL: http://link.aip.org/link/?AJP/49/466/1, doi:10.1119/1.12714.

Bharadwaj, S., Bhatt, H., Vatsa, M., Singh, R., 2010. Periocular biometrics: When iris recognition fails, in: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –6. doi:10.1109/BTAS.2010.5634498.

Boles, W., Boashash, B., 1998. A human identification technique using images of the iris and wavelet transform. IEEE Transactions on Signal Processing 46, 1185–1188.

Bowyer, K., Hollingsworth, K., Flynn, P., 2008. Image understanding for iris biometrics: A survey. Comput. Vis. Image Underst. 110, 281–307. URL: http://dx.doi.org/10.1016/j.cviu.2007.08.005, doi:10.1016/j.cviu.2007.08.005.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence , 679–698.

Castrillón, M., Dénis, O., Guerra, C., Hernández, M., 2007. Encara2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation 18, 130–140.

Clausi, D., Jernigan, M., 1996. Towards a novel approach for texture segmentation of sar sea ice imagery, in: 26th International Symposium on Remote Sensing of Environment and 18th Annual Symposium of the Canadian Remote Sensing Society, Vancouver, BC, Canada. pp. 257–261.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893.

Daugman, J., 1993. High confidence visual recognition of persons by a test of statistical independence. Pattern Analysis and Machine Intelligence, IEEE Transactions on 15, 1148 –1161. doi:10.1109/34.244676.

Daugman, J., 2004. How iris recognition works. Circuits and Systems for Video Technology, IEEE Transactions on 14, 21–30. doi:10.1109/TCSVT.2003.818350.

Daugman, J., 2007. New methods in iris recognition. IEEE Trans. Systems, Man, Cybernetics B 37, 1167–1175.

Hurley, D., Nixon, M., Carter, J., 2000. A new force field transform for ear and face recognition, in: Image Processing, 2000. Proceedings. 2000 International Conference on, pp. 25 –28. doi:10.1109/ICIP.2000.900883.

on Illumination, I.C., 2004. Colorimetry. CIE technical report, Commission internationale de l'Eclairage, CIE Central Bureau.

Juefei-Xu, F., Cha, M., Heyman, J., Venugopalan, S., Abiantun, R., Savvides, M., 2010. Robust local binary pattern feature sets for periocular biometric identification, in: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –8. doi:10.1109/BTAS.2010.5634504.

Juefei-Xu, F., Luu, K., Savvides, M., Bui, T., Suen, C., 2011. Investigating age invariant face recognition based on periocular biometrics, in: Biometrics (IJCB), 2011 International Joint Conference on, pp. 1 –7. doi:10.1109/IJCB.2011.6117600.

Laws, K., 1980. Rapid texture identification, in: Proc. SPIE Conf. Image Processing for Missile Guidance, pp. 376–381. doi:10.1117/12.959169.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94, doi:10.1023/B:VISI.0000029664.99615.94.

Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11, 674 –693. doi:10.1109/34.192463.

Matey, J., Naroditsky, O., Hanna, K., Kolczynski, R., LoIacono, D., Mangru, S., Tinker, M., Zappia, T., Zhao, W., 2006. Iris on the move: Acquisition of

images for iris recognition in less constrained environments, in: Proceedings of the IEEE, pp. 1936–1947. doi:10.1109/JPROC.2006.884091.

Miller, P., Rawls, A., Pundlik, S., Woodard, D., 2010. Personal identification using periocular skin texture, in: Proceedings of the 2010 ACM Symposium on Applied Computing, ACM, New York, NY, USA. pp. 1496–1500. URL: http://doi.acm.org/10.1145/1774088.1774408, doi:10.1145/1774088.1774408.

Ojala, T., Pietikainen, M., Harwood, D., 1994. Performance evaluation of texture measures with classification based on kullback discrimination of distributions, in: Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on, pp. 582 –585 vol.1. doi:10.1109/ICPR.1994.576366.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, 971 –987. doi:10.1109/TPAMI.2002.1017623.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175.

Park, U., Jillela, R., Ross, A., Jain, A., 2011. Periocular biometrics in the visible spectrum. Information Forensics and Security, IEEE Transactions on 6, 96 –106. doi:10.1109/TIFS.2010.2096810.

Park, U., Ross, A., Jain, A., 2009. Periocular biometrics in the visible spectrum: A feasibility study, in: Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on, pp. 1 – 6. doi:10.1109/BTAS.2009.5339068.

Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L., 2010. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, 1529 –1535. doi:10.1109/TPAMI.2009.66.

Santos, G., Hoyle, E., 2012. A fusion approach to unconstrained iris recognition. Pattern Recognition Letters 33, 984 – 990. URL: http://www.sciencedirect.com/science/article/pii/S0167865511002686, doi:10.1016/j.patrec.2011.08.017.

Santos, G., Proença, H., 2013. Periocular biometrics: An emerging technology for unconstrained scenarios, in: Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013, pp. 14–21.

Smith, T., Guild, J., 1931. The C.I.E. colorimetric standards and their use. Transactions of the Optical Society 33, 73–134.

Tan, T., He, Z., Sun, Z., 2010. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image and Vision Computing 28, 223 – 230. doi:10.1016/j.imavis.2009.05.008.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518.

Wildes, R., 1997. Iris recognition: an emerging biometric technology, pp. 1348–1363.

Wolf, S., 2003. Color Correction Matrix for Digital Still and Video Imaging Systems. Technical Report TM-04-406. NTIA - National Telecommunications and Information Administration. USA.

Woodard, D., Pundlik, S., Miller, P., Jillela, R., Ross, A., 2010. On the fusion of periocular and iris biometrics in non-ideal imagery, in: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 201 –204. doi:10.1109/ICPR.2010.58.

Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: A literature survey. ACM Comput. Surv. 35, 399–458. URL: http://doi.acm.org/10.1145/954339.954342, doi:10.1145/954339.954342.