

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Mobile Multimodal Biometric Identification for African Communities

Rui Cardoso Esteves

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

FEUP Supervisor: Jaime dos Santos Cardoso (PhD)

Fraunhofer Portugal Supervisor: Joana Polónia Lobo (MsC)

FEUP Co-Supervisor: João Carlos de Sousa Monteiro (MsC)

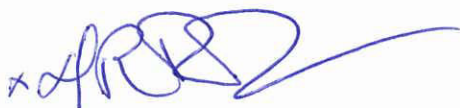
July 22, 2015

A Dissertação intitulada

“Mobile Multimodal Biometric Identification for African Communities”

foi aprovada em provas realizadas em 20-07-2015

o júri



Presidente Professora Doutora Maria do Rosário Marques Fernandes Teixeira de Pinho
Professora Associada do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto



Professor Doutor Hugo Pedro Proença
Professor Associado do Departamento de Engenharia Informática da Universidade da
Beira Interior



Professor Doutor Jaime dos Santos Cardoso
Professor Associado do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.



Autor - Rui Cardoso Esteves

Resumo

A Biometria, reconhecimento de um indivíduo através das suas características físicas ou comportamentais, é utilizada nos países ricos principalmente para fins de segurança como por exemplo, controlo de acesso. Nos países subdesenvolvidos, nomeadamente os países que constituem a África Subsariana, existem indivíduos em pobreza extrema que estão socialmente excluídos por não possuírem nenhum documento oficial de identificação. Na aplicação de técnicas de biometria, pode estar a solução para a identificação destas pessoas.

A situação precária dessas populações e a falta de recursos como instalações e equipamento próprio para realizar o processo de aquisição de imagem são um entrave ao estabelecimento de tecnologias mais avançadas nesses países. Estes factores motivaram o estudo do uso de imagens recolhidas por smartphones na identificação de um indivíduo, proporcionando assim uma alternativa mais económica e robusta.

Neste trabalho é avaliada a capacidade de um dado algoritmo ser capaz de reconhecer indivíduos através de imagens de baixa qualidade, adquiridas em ambientes pouco propícios. Além disso, é também avaliado o impacto da variação dessas condições de aquisição na performance do mesmo algoritmo. Os testes são efectuados recorrendo a uma base de dados já existente que compreende imagens da região periocular de 50 pessoas, captadas em diferentes ambientes e fazendo uso de diferentes telemóveis.

Tendo em vista a aplicação destas soluções em África, foi avaliado o comportamento do algoritmo quando aplicado em indivíduos de origem africana. Para tal, foi construída uma base de dados que junta imagens da mão, da orelha e também da região periocular de 50 indivíduos, onde metade das amostras provêm de sujeitos africanos e outra metade de caucasianos. Optou-se por recolher amostras de diferentes traços para estudar de que forma a multimodalidade poderia melhorar a performance deste tipo de sistemas e também para avaliar a robustez do algoritmo quando aplicado a outros traços.

O algoritmo revelou elevada robustez no reconhecimento de indivíduos, principalmente quando utilizado o GIST para descrever imagens em escala de cinzentos. Os métodos de fusão estudados não trouxeram melhorias significativas ao sistema conferindo-lhe apenas um maior nível de complexidade computacional. Foram ainda identificados vários factores externos, como por exemplo o nível de luminosidade ambiente ou a etnia do indivíduo, que podem ter um elevado impacto do desempenho do sistema.

A metodologia apresentada mostrou-se apta para ser aplicada num contexto real. Contudo e devido à complexidade deste tipo de sistemas, muitas áreas ficaram por explorar, motivando a continuidade deste estudo num trabalho futuro.

Abstract

Biometrics, the automated recognition of individuals based on their behavioral and biological characteristics, is mostly used in developed countries with forensic and security purposes, as well as to regulate interactions between the citizens and the state. In developing countries, namely Sub-Saharan African (SSA) countries, some individuals are not officially identified and since they do not formally exist, they are completely excluded from the citizen life. The solution to this identity gap can lie in the usage of a biometric system.

The underprivileged situation that these populations face and the lack of resources, like facilities and proper equipment to perform a controlled data acquisition, are barriers to the implementation of more advanced techniques. This situation motivates the use of smartphone-based biometric systems to perform individual recognition, providing a more economic and robust alternative.

In this work, an algorithm is evaluated regarding its use on individual recognition, using low quality images acquired on unconstrained environments. The impact of this kind of environments, in the system performance, is also assessed by using an available dataset, with data from 50 individuals, comprising images captured from distinct smartphones.

Having in mind these systems' implementation in SSA, the algorithm was assessed when it is computed on images from African subjects. This was achieved by building a new dataset, that comprises periocular, hand, and ear images from 50 individuals also. However, half of the subjects is African while the other half is Caucasian. The multimodality characteristic of the dataset allow the study of multimodal systems, as well as the evaluation of the algorithm's behaviour when applied to different traits.

The algorithm showed high robustness regarding individual recognition, particularly when grayscale images and GIST descriptor were used. The applied fusion methods brought no significant improvements to the system, contributing just to an increase in its computational complexity. Also, some external factors, such as the environment brightness level or the individual's ethnicity, were identified as having an high impact on the system performance.

The presented methodology proved to be capable of being applied in a real context. However and due to the complexity of such systems, many areas were to explore, encouraging the continuation of this work in the future.

Agradecimentos

Este trabalho representa o culminar da minha aventura de 5 anos como estudante da Faculdade de Engenharia da Universidade do Porto. Assim sendo, e como manda a tradição, não poderia virar esta página sem antes agradecer a quem me guiou até aqui e a todos os que estiveram direta ou indirectamente envolvidos neste trabalho.

Ao professor Jaime Cardoso pela segunda oportunidade. Ao João Monteiro pela dedicação e paciência demonstradas no decorrer deste longo semestre. À Joana Lobo por me ter acolhido tão bem na Fraunhofer Portugal. A todos os que se voluntariaram para a recolha de imagens pela vossa cooperação. Nutro, ainda, um especial carinho pela FEUP por me ter dado a conhecer tantas pessoas fantásticas. Obrigado!

Ao meu amigo Renato e à minha querida Avó por serem a minha inspiração. Aos meus pais por me darem a oportunidade de estudar, pela educação e pela liberdade. Ao meu irmão pelo exemplo que representa. À minha gatona por se preocupar tanto com as minhas férias e por me aturar enquanto a aturo. A toda a minha família por sermos uma Família.

Tenho ainda de agradecer a todos os meus amigos, aos que conheço há quase tantos anos quantos há me conheço a mim mas também aos do GAS Porto, aos da FEUP, em especial ao Luís Moreira, e a todos os outros.

Obrigado a todos por me darem um pouco daquilo que são.

Rui Cardoso Esteves

*“Que se lixe.
Vai assim.”*

Student’s proverb

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objectives	2
1.3	Contributions	2
1.4	Outline	2
2	Literature Review	5
2.1	Context	5
2.2	Introduction to Biometrics	6
2.3	Computer Vision in Biometrics	8
2.3.1	Data Acquisition	8
2.3.2	Pre-processing	9
2.3.3	Feature Extraction	10
2.3.4	Matching	10
2.4	System Performance	11
2.4.1	Biometric Traits	12
2.5	Multimodal Systems	19
2.6	Descriptors	21
2.7	Commercial Solutions	24
2.8	Summary	26
3	MoBIAC Dataset	27
3.1	Dataset Overview	28
3.2	Dataset Description	29
3.3	Acquisition environment	32
4	Periocular Recognition System	33
4.1	Algorithm overview	33
4.1.1	Universal Background Model	34
4.1.2	Hypothesis Modeling	34
4.1.3	Modeling UBM and IDSM	35
4.2	CSIP Dataset	36
4.3	Proposed methodology	36
4.4	Experimental Results	39
4.4.1	k-Mixture GMM	40
4.4.2	Setup comparison	40
4.4.3	Color channel	42
4.4.4	Descriptors comparison	44

4.4.5	Descriptors Fusion	45
4.4.6	Full Dataset	46
4.4.7	Computational Performance	47
4.4.8	Implementation Details	49
5	Preliminary results on MoBIAC dataset	51
5.1	Periocular Region	51
5.2	Contactless Palmprint	53
5.3	Ear	56
5.4	Conclusions	58
6	Conclusions and future work	59
6.1	Future Work	60
A	Results	61
	References	73

List of Figures

2.1	Survey of the Use of Biometrics Technologies for Development, Low middle-Income Countries (2012). [1]	6
2.2	Steps of the enrollment stage [2].	7
2.3	Steps of the verification stage [2].	7
2.4	Steps of the identification stage [2].	8
2.5	A grayscale fingerprint (left) and the corresponding binarized image (right) [3].	9
2.6	Matching result on real-world images with viewpoint change. Green lines are valid matches; red circles indicate unmatched points [4].	10
2.7	The Neyman-Pearson/ROC decision strategy curve [5]	11
2.8	Fingerprint Features [6].	13
2.9	Palmprint [7].	13
2.10	Finger Knuckle Pre-processing [8]	14
2.11	Human Iris [9]	15
2.12	Example of periocular region from right eye	16
2.13	Human ear features [10].	17
2.14	Multimodal biometrics [11].	20
2.15	Gradient magnitude and orientation computation of a single keypoint using SIFT descriptor (adapted from [12]).	22
3.1	Hands from different ethnicities. The top-left photo is a typical SSA individual photo while others are related to different ethnicities. There are obviously marked differences on principal lines contrast.	29
3.2	MoBIAC Hand samples.	30
3.3	MoBIAC Periocular examples.	31
3.4	MoBIAC Ear examples.	32
3.5	Example scheme to illustrate the light variance approach. Region 1 is enlightened by the external light while region 2 is darker because of the counter-light effect.	32
4.1	Details related to each CSIP's acquisition setup [13].	36
4.2	CSIP samples from 2 different subjects representing different setups [13].	37
4.3	Schematic of the proposed methodology.	38
4.4	ROI cropped by Santos et al. method (b) and by this work's approach (b) regarding the same CSIP image.	39
4.5	Average time, in seconds, taken to perform a single projection for each descriptor regarding AR0 setup and for GS colorspace.	47
4.6	Visual comparison of the time taken by each descriptor computation regarding the AR0 for GS colorspace.	48
5.1	MoBIAC Periocular samples.	52

5.2	MoBIAC palmprint samples.	54
5.3	GIST and SIFT likelihood-ratio matrices comparison on GS colorspace. Y and X axis represent the project sample and the identity number, respectively.	56
5.4	MoBIAC Ear samples.	57

List of Tables

2.1	Traits comparison based on evaluation parameters presented on chapter 2.2	18
2.2	Comparison table of some the available commercial solutions	25
3.1	Comparison between some of the available datasets	28
4.1	Feature vector dimension before and after PCA for each one of the 5 descriptors.	38
4.2	Example table to access Rank-1 ratio and decidability performance regarding SETUP images using DESCRIPTOR.	39
4.3	R1 performance comparison between setups for 16-mixture GMM using GS images	41
4.4	DEC performance comparison between setups for 16-mixture GMM using GS images	41
4.5	Comparison between the best color channel and GS for the 5 descriptors, with a 16-mixture GMM, for AR1 and AR0 setups.	42
4.6	Comparison between the worst color channel and GS for the 5 descriptors, with a 16-mixture GMM, for AR1 and AR0 setups.	43
4.7	Comparison between the best color channel plus GS (left column) and RGB for the 5 descriptors and, with a 16-mixture GMM, for AR1 and AR0 setups.	43
4.8	Results performed using the best k for each descriptor, on GS images from AR1 setup.	44
4.9	SIFT performance on non-flash (left) and flash (right) images.	44
4.10	GIST performance on non-Flash (left) and Flash (right) images.	45
4.11	Fusion scenarios performed on rear-camera related setups, for GS images.	46
4.12	Results obtained regarding the full CSIP dataset, for GS colorspace	47
4.13	Time taken by each descriptor computation regarding the AR0 for GS colorspace	48
4.14	Average recognition times for each descriptor regarding a single sample, using 16-mixture GMM for AR0 setup and GS colorspace.	49
5.1	Results obtained regarding periocular region related to the full MoBIAC periocular dataset.	52
5.2	Performance results regarding periocular region related to the African individuals of the MoBIAC dataset	53
5.3	Performance results regarding periocular region related to the Caucasian individuals of the MoBIAC dataset	53
5.4	Results obtained regarding contactless palmprint related to the full MoBIAC hand dataset.	54
5.5	Performance results regarding contactless palmprint related to the African individuals of the MoBIAC dataset.	55
5.6	Performance results regarding contactless palmprint related to the Caucasian individuals of the MoBIAC dataset.	55

5.7	Results obtained regarding the full MoBIAC ear dataset.	57
5.8	Performance results regarding the ear related to the African individuals of the MoBIAC dataset.	58
5.9	Performance results regarding the ear related to the Caucasian individuals of the MoBIAC dataset.	58
A.1	Rank-1 ratio and decidability performance regarding AR1 setup images.	62
A.2	Rank-1 ratio and decidability performance regarding AR0 setup images.	63
A.3	Rank-1 ratio and decidability performance regarding BR1 setup images.	64
A.4	Rank-1 ratio and decidability performance regarding BR0 setup images.	65
A.5	Rank-1 ratio and decidability performance regarding BF0 setup images.	66
A.6	Rank-1 ratio and decidability performance regarding CR0 setup images.	67
A.7	Rank-1 ratio and decidability performance regarding CR1 setup images.	68
A.8	Rank-1 ratio and decidability performance regarding DR0 setup images.	69
A.9	Rank-1 ratio and decidability performance regarding Full CSIP dataset for 16-mixture GMM.	69
A.10	Rank-1 ratio and decidability performance regarding MoBIAC periocular region dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM.	70
A.11	Rank-1 ratio and decidability performance regarding MoBIAC ear dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM. . .	70
A.12	Rank-1 ratio and decidability performance regarding MoBIAC hand dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM. . .	71

Abbreviations

CSIP	Cross Sensor Iris and Periocular
DEC	Decidability
DoG	Difference of Gaussians
EER	Equal Error Rate
FNMR	False Non Match Rate
FNR	False Negative Rate
FPR	False Positive Rate
FTC	Failure To Capture
FTE	Failure To Enroll
GMM	Gaussian Mixture Models
GS	Gray Scale
HOG	Histogram of Oriented Gradients
IDSM	Individual Specific Model
LBP	Local Binary Pattern
MAP	maximum a posteriori
MoBIAC	Mobile Multimodal Biometric Identification for African Communities
PCA	Principal Component Analysis
QA	Quality Assessment
RER	Ranking Error Rate
RGB	Red Green Blue
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SDK	Software Developer Kit
SIFT	Scale Invariant Feature Transform
SSA	Sub-Saharan Africa
UBM	Universal Background Model
ULBP	Uniform Local Binary Pattern

Chapter 1

Introduction

With the marked advance of technology, smartphones are becoming faster and more powerful, with new features and capabilities arriving to the market each year. Among these features built-in sensors, such as high resolution cameras and accurate microphones, can be accounted for. For many, these capabilities are used to gather information about the surrounding world, as well as to record quotidian situations, so they can share them later, as a means of social interaction. However, these mobile devices' potential goes far beyond such entertainment and ordinary applications. This Master's thesis work aims to conceptualize the idea of using smartphones as part of a mobile identification system based on biometric traits to be applied in Sub-Saharan Africa (SSA) developing countries. This development was carried out under the scope of a Master's Thesis project for the Integrated Master in Electrical and Computers Engineering at Faculty of Engineering of University of Porto and in association with Fraunhofer Portugal Research Center for Assistive Information and Communication Solutions (AICOS).

This chapter begins with an introduction to the problem that motivated this work. Then, the objectives for the mobile system solution and its expected innovative contributions are presented. Finally, the various steps to be taken are outlined.

1.1 Motivation

In SSA, the lack of official individual identification [1] promotes poverty and social exclusion for many. Hundreds of millions of poor people have no birth certificates, no addresses and no more than single-word names. With the over-growing of SSA population more and more people will be in total exclusion outside the society's boundaries. This identity gap makes development more difficult and less inclusive. However, there are already many programs that aim to provide individuals a more robust official identity, usually in the context of the delivery of services [1].

Despite the existence of such methods, governments do not answer timely and efficiently to deprived people's needs. Therefore, it is urgent to conceive new methods to quickly answer the population's needs, as well as to cover the large identity gap that is pointed out as one of the biggest obstacles to the development of SSA countries.

1.2 Objectives

During the last decade, biometric industry has “boomed” in the SSA market, mainly for non-security applications. The role of this work is to contribute for that expansion with a multimodal biometric identification solution tailored to be used in unconstrained environments. Multimodal biometric solutions can greatly improve the results in identification over their unimodal counterparts, as they can handle much more information regarding an individual, significantly improving their robustness when facing unfavorable conditions.

In SSA, the majority of the population, with no means to travel, have to be assisted on the field rather than being received within dedicated facilities. With this in mind, the hardware used is restricted to portable devices, which may prove too expensive and sparse to be travelling around from village to village. In order to prevent that, the proposed work is to be developed within a commercial smartphone using only the provided built-in features for data acquisition. Thus, the whole biometric identification process should be prepared to be integrated in such scenarios.

These imposed hardware limitations demand a solid algorithm, so that the developed solution can represent a reliable and plausible alternative to be used in a real-life situation. If the solution fails to improve the identification process, it will further contribute to the social exclusion of even more people.

With all these restrictions, the available biometric algorithms and techniques will be evaluated and studied with the goal of gathering enough information to build a satisfactory solution, adapted to the situation context.

1.3 Contributions

The work developed on the scope of this Master’s Thesis resulted in several contributions such as:

- The construction of an image dataset composed by Ear, Hand and Periocular samples acquired from both Africans and Caucasian individuals under unconstrained scenarios.
- Study of the impact of mobile setups variations on individual recognition using a periocular region-based algorithm.
- Study of five different descriptors and its performance evaluation when used to analyse images acquired in highly unconstrained scenarios.
- Evaluation of a novel algorithm’s behavior when applied to a dataset acquired under unconstrained environments and making use of different biometric traits.

1.4 Outline

After this introduction follows a review of the literature in biometric identification. All the information needed to understand the system regarding its context and design will be presented, as

well as a comprehensive review of already developed methods. Still in Chapter 2, follows a brief comparison between the available solutions in the market. In the 3rd Chapter, work regarding the dataset acquisition setup will be presented. Already in Chapter 4, the studied algorithm will be presented and its performance will be compared with a state-of-the-art algorithms. In order to test the proposed methodology's robustness to varying conditions, in Chapter 5, the algorithm will be applied over different traits, including the ones collected on the dataset described in Chapter 3. In Chapter 6, the main conclusions achieved through this work will be stated and suggestions will be given about possible future work.

Chapter 2

Literature Review

This chapter starts with a discussion regarding biometric identification problems in the specific application scenario of the SSA situation. Then, commercial solutions already available in the market will be presented and analysed. In a more technical topic, typical biometric systems will be described, as well as some research done over different biometric traits.

2.1 Context

During the 1990's, biometrics' market was established almost only on developed countries, due to the improvement of technological solutions verified during those years. However, mostly due to recent advances in digital biometric identification, barriers are being broken, and the main markets have, in recent years, been expanding their products to developing countries, as depicted in Figure 2.1. It is estimated that, in these countries, more than one billion people had their biometrics taken for at least one purpose. As a result, biometrics industry is "booming" with an estimated growth of more than 25% during the last 10 years [1].

Among the fifty-three African countries, thirty-four are among the world's least developed countries. One of the regions where many of these developing countries lie is the SSA. Sub-Saharan Africa gathers all countries located south of the Sahara desert, excluding Sudan. This term is usually used to distinguish these countries from the North-Africa Arab ones since they represent a different reality, with respect to their societies, economics and ethnographic characteristics [1]. Biometric-based implementations in developing countries differ completely from those provided in developed ones. Even if such countries have access to the same technologies, these groups' social context demands different implementations aimed to fulfill distinct objectives. In rich countries, biometrics is used mostly with forensics and security proposes and it is well established to regulate interactions between the citizens and the state (voting, passports) or non-state institution (buying houses and make use of bank accounts). In SSA countries, where many people do not even have any official ID documentation, the purpose of biometrics is to help cover the existing identity gap: civil registers, voter rolls, health records, social transfers, among many others. Individuals that are not officially identified do not formally exist and are completely excluded

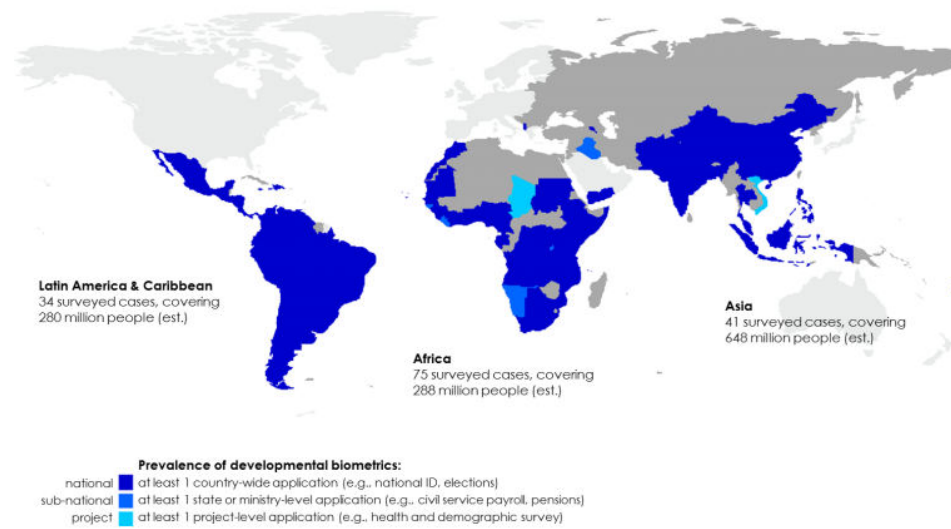


Figure 2.1: Survey of the Use of Biometrics Technologies for Development, Low middle-Income Countries (2012). [1]

from the citizen life. They cannot register property, have bank accounts and receive public transfers or services and they see their opportunities for economic, social and political development limited [1]. In order to implement this kind of solutions, complex systems need to be established within those countries. In the next section, a brief introduction of traditional biometric strategies will be carried out, in order to understand the true needs and challenges when developing such a system.

2.2 Introduction to Biometrics

Biometrics can be defined as the automated recognition of individuals based on their behavioral and biological characteristics [14]. There are several traits that can be used in order to identify an individual, each one with its own distinctive features. In order to qualitatively evaluate a trait [15], some requirements are preferable to be satisfied and should be taken into account:

- Universality. Everyone among the target population should have it;
- Distinctiveness. It should be unique in each individual;
- Permanence. It should be invariant over a given period of time;
- Collectability. Data, from the chosen trait, should be easy to capture.

Usually, three more requirements [15] are considered in real life systems: Performance (the accuracy, speed and resource requirements), Acceptability (harmless to users), and Circumvention (robust to fraudulent methods). Depending on what the context demands, different boundaries to each requirement should be established. For example, if a biometric identification system is to be

applied on infants in a neighbour school, a trait like the hand fulfills the universality requirement. However for retired war veterans identification, the hand cannot be considered a universal trait, since past soldiers could have lost a hand in war.

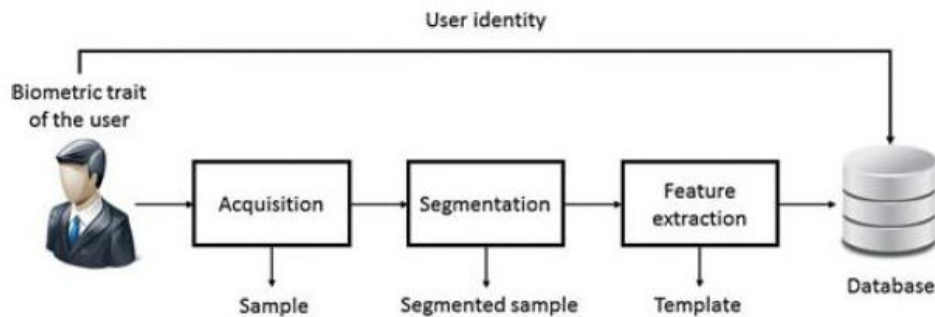


Figure 2.2: Steps of the enrollment stage [2].

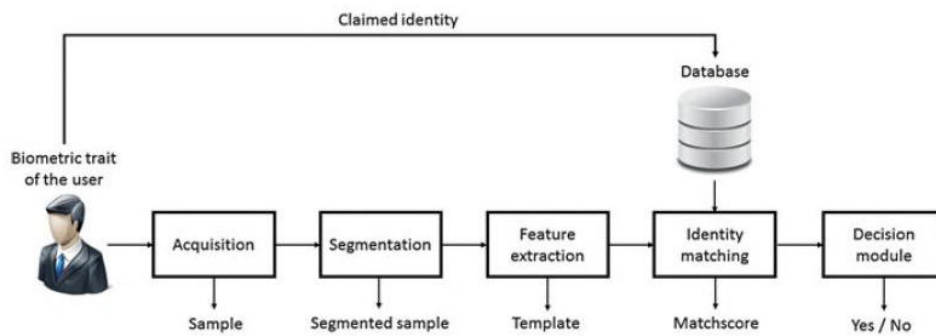


Figure 2.3: Steps of the verification stage [2].

Once the trait is defined, the system has to be planned. A generic biometric system operates in two different stages. The first stage, which is usually called *Enrollment*, Figure 2.2, is where an individual is added into a previously existing system database, by capturing and storing data samples to use as future references. On the second stage, the *Recognition*, new samples are captured and compared with the data stored in the previous stage.

Biometric systems can be grouped in two different types, each one with their purpose. In identification systems, Figure 2.4, the new data is matched with the stored data related to all the subjects enrolled in the dataset, in order to find the most probable identities. In authentication systems, Figure 2.3, the new data is compared only with a single claimed identity, so that the individual identity can be validated. Although differing in their purpose, both systems are structured the same way. Typically, they are composed by four modules [15], each one with defined boundaries and roles:

- *Sensor module*, which acquires the biometric data;

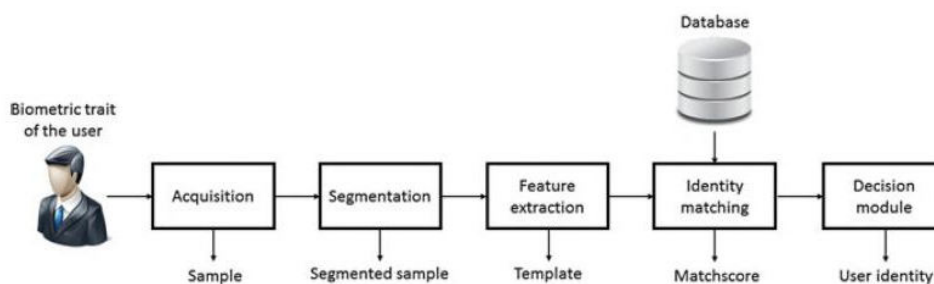


Figure 2.4: Steps of the identification stage [2].

- *Feature extraction module*, responsible for the processing of the acquired data in order to extract the distinctive features and save them in a feature vector;
- *Matching module*, which matches the new data feature vectors against those kept in the database;
- *Decision-making module*, on which the user's identity is established or a claimed identity is accepted or rejected.

Now that the system is overviewed, it is important to understand how automatic image recognition systems work and how it is possible to encode an image, so that it can be seen and analysed by a computer.

2.3 Computer Vision in Biometrics

Computer vision is a field that includes methods for acquiring, processing, analyzing and understanding images [16]. A biometric system based on image analysis makes use of all these methods in order to identify an individual. In this section, follows a brief explanation on those methods and a review on some state of the art algorithms related to image processing, that will prove of interest in further sections.

2.3.1 Data Acquisition

Data acquisition is the first step of a computer vision system. Conditions in which the data is collected have a great influence in the system performance. Each trait demands an architecture that can be chosen from a vast variety of possibilities. As an example for fingerprint identification, one can use a specific contact sensor [17] made for that purpose or simply take a photo of the finger with an ordinary camera [18]. Both methods will provide data with completely different degrees of information and noise. In the context of this research, the main focus is on mobile-phone solutions based on built-in cameras, as detailed in further sections.

2.3.2 Pre-processing

The acquisition of different samples poses challenges in the form of format, size of images, distortion, differences in orientation and variation of gray scale values. During the pre-processing, raw data acquired during the previous stage is processed and filtered in order to minimize any existing noise and highlight the relevant characteristics, promoting and favoring the quality of the features to be extracted. This stage saves time and processing power that, in other way, would be expended in later stages. In biometrics, the region of interest (ROI) is usually the region that contains the trait itself and the noise areas are usually undesirable information like backgrounds, badly illuminated areas, and extra objects (like rings, sunglasses, and earrings) [19]. There is a wide variety of pre-processing methods, each one serving its own purposes, like contrast enhancement, feature highlighting and segmentation.

Segmentation is the process of partitioning a digital image into multiple segments. The goal is to highlight the ROI in order to make it easier to analyse, simplifying the image and establishing boundaries to the region where the main features are contained [20]. To highlight some features one method usually applied on systems that rely on the clarity and detail of ridges (e.g. fingerprint and knuckle), is *Binarization* [3]. As shown in Figure 2.5, *Binarization* differentiates the fingerprint ridges and valleys by simply applying a threshold on the image intensity values, replacing each pixel in an image with a black pixel if the image intensity is less than some fixed constant T , or a white pixel, if the image intensity is greater than that constant. A variation of this method is called *Adaptive Binarization* [21], where the *Binarization* threshold is locally defined in each pixel.

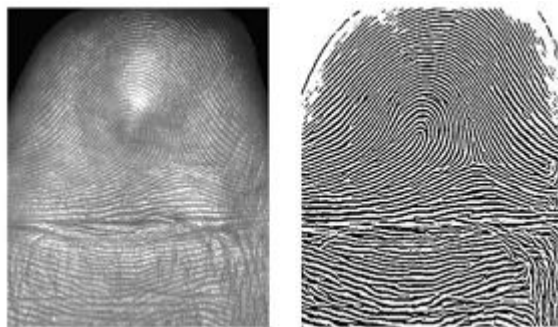


Figure 2.5: A grayscale fingerprint (left) and the corresponding binarized image (right) [3].

For color correction there are some useful tools commonly used in biometrics. For example, *Histogram equalization* [22, 23, 3] algorithms increase the global contrast of images. The *Canny edge detector* [24, 8] is widely used tool to locate sharp intensity changes and object boundaries in an image. This detector classifies a pixel as an edge if the gradient magnitude of that pixel is larger than the neighbour pixels in the direction of maximum intensity change. Another example are *Gabor filters* [24, 8, 25] and their variations like *Log-Gabor filters* that are suitable when feature details are to be acquired from high-frequency areas. Regarding each application and its

expected performance, there are several image processing methods to be used. The difficulty is in adjusting them to the, sometimes wide, variation of samples.

2.3.3 Feature Extraction

Feature extraction is the process where meaningful information is obtained from the processed image. The extracted features contain the relevant information from the input data, so that the matching can be performed by using only this reduced representation. Feature extraction is achieved by means of descriptors which are capable of encoding the image in a more compact representation, making it possible to make an efficient comparison, using only a reduced portion of the original information. Some descriptors will be explored in further detail on Section 2.6, since they are of great importance in this study. Features are related to pixel intensities, as well as their spatial position, shape, orientation, and textures and they are stored within a set of features, commonly called feature vector [26].

2.3.4 Matching

The feature vector, which serves as an image descriptor, is ready to be compared with other feature vectors representing the previously enrolled images, stored in a database. The response of this biometric matching system is typically a matching score, s , that quantifies the similarity between the input and the database template representations. The better the results, the more certain it is that the two samples refer to the same individual. After the new data is matched against a stored template, it is classified given their degree of similarity. On Figure 2.6, a matching result on magazine identification is presented.

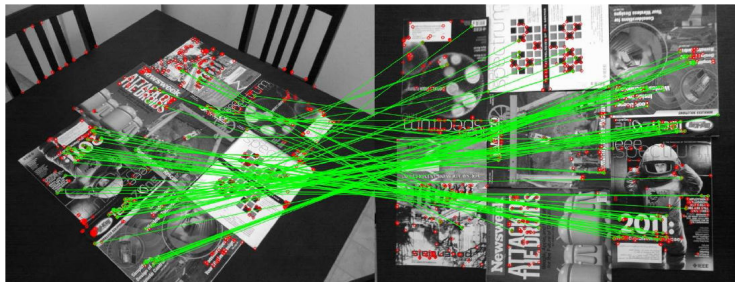


Figure 2.6: Matching result on real-world images with viewpoint change. Green lines are valid matches; red circles indicate unmatched points [4].

In machine learning, the science field responsible for the study of algorithms that can learn from and make predictions on data, the matched sample is classified as belonging to a specific class. A multiclass classification is the method of classifying a sample into one among more than two classes. If, on the other hand, the classification is done between two classes only, it is called binary classification. In biometrics, the multiclass classification is present on identification systems, in which the new data is matched with all the stored data related to all the subjects

enrolled in the dataset. However, in authentication systems binary classification occurs, since the new data is compared with a single claimed identity.

2.4 System Performance

As stated before, biometric systems are grouped in two different types: Identification (1:N) and Verification (1:1). In order to evaluate a verification biometric system, its performance can be measured by the trade-off between the probability that the incorrect individual is falsely recognized, the false positive rate (FPR), and the probability that the correct individual is falsely not recognized, the false negative rate (FNR). Two other error rates can be used to evaluate system performance at a different level: failure to capture (FTC) and failure to enroll (FTE). FTC measures the device failure rate when capturing a sample (e.g. in the presence of noise), and FTE measures whether the user is or is not successfully enrolled in the system [26].

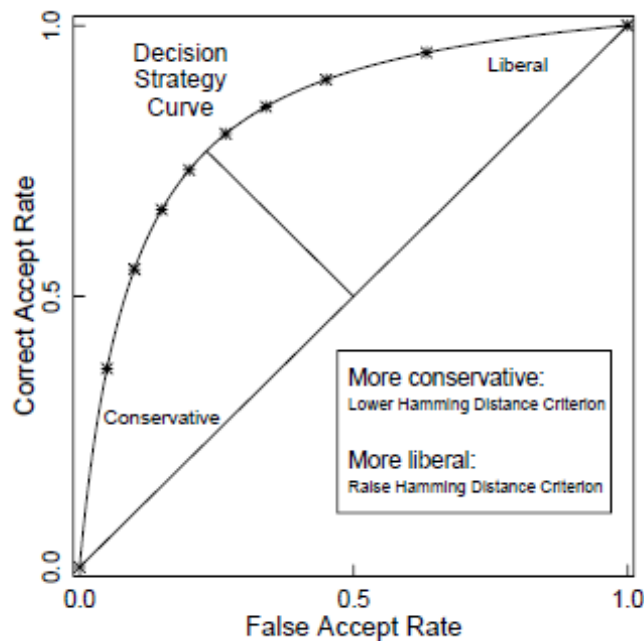


Figure 2.7: The Neyman-Pearson/ROC decision strategy curve [5]

Another common description of the error rate of a verification system is the equal error rate (EER). EER is observed at the decision threshold where false positive and false negative errors occur with equal frequency. The lower the EER the higher the performance of the algorithm. This error relation can be visually analysed through the Receiver Operating Characteristic (ROC) curve (Figure 2.7) [26, 5]. Although they are important, error rates sometimes are not enough to evaluate in detail the performance of an algorithm. A different metric used to evaluate decision-making power is *decidability* (DEC). DEC was first introduced by Daugman [27] and quantifies intra and inter-class separability by relating their mean, μ , and standard deviation, σ , as it follows:

$$DEC = \frac{|\mu_{genuine} - \mu_{impostor}|}{\sqrt{0.5(\sigma_{genuine}^2 + \sigma_{impostor}^2)}}$$

This metric does not depend on the decision threshold used when selecting the trade-off between falsely positive and negative errors. It represents the separability of two distributions, quantifying the separation between genuine and impostor likelihood score distributions.

When it comes to the evaluation of an identification biometric system, different approaches are needed, like the Ranking Error Rate (RER) and Cumulative Match Curves. In identification, the goal is to find the most similar template in a stored database of templates from several individuals. The expected behavior of an ideal similarity function is that highest values are always assigned to genuine comparisons. Ranking is the metric that measures the error rate of select impostors templates as most similar. The Rank-N version of the RER is the ratio of query templates for which genuine comparisons are among the N comparisons with highest similarity value in database [28].

2.4.1 Biometric Traits

The first decision to make before start developing a biometric system is the choice of the trait. Having in mind the requirements presented on Section 2.2 and the purpose of this Master's Thesis, it is necessary to study which traits are most suitable in the SSA context and easier to be acquired by built-in smartphone sensors. Given that, follows a brief review and analysis of the research developed about the most commonly used traits in scenarios similar to the described one.

Fingerprints

The fingerprint is probably the most famous trait because of its use in forensic investigation. Fingerprint based methods consist on analysing the fingertip pattern of ridges and furrows, as seen in Figure 2.5. It is widely used because the fingertip pattern is established during the fetal period, which makes it a lifetime characteristic, and it is expected that all fingers in the world have different patterns [26]. Usually, fingerprint data is collected using an appropriate in-contact sensor. There are some problems related with this kind of approach: the physical contact of a finger with a scanner generates pressure and induces distortions of the fingerprint [3], deteriorating the extracted information and making more difficulting the feature extraction. Fingerprint features [6] are presented on Figure 2.8.

Some studies have explored contactless acquisition techniques. Kumar et al. [25] developed a system where 3D fingerprint surface is reconstructed from 2D images using the shape from shading technique. Then, the 3D minutiae templates, as well as the 2D templates, are both used for identification. The result was an EER of 1.17% from 240 clients' fingerprints using 3600 genuine and 2064960 impostor scores. In the work of Jain et al. [29], images were acquired using two recent mobile models (Iphone 5S and Samsung Galaxy S4) but they had not sufficient quality for feature extraction and matching. Probably, this is the reason why there are not many studies



Figure 2.8: Fingerprint Features [6].

on fingerprint identification systems on mobile phones, despite the already available commercial solutions on the market, which will be described later on.

Palmprint

Palmprint identification systems are similar to the fingerprint ones. However, they can lead to even more distinctive results, mainly due to its larger size and, thus, amount of meaningful information. In addition, palmprint has additional distinctive features such as principal lines and wrinkles. These traces are easily acknowledged and can be captured with a lower resolution scanner providing a cheaper solution [30].

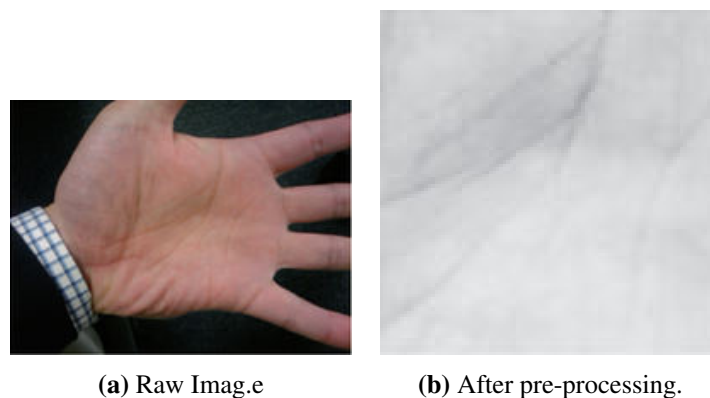


Figure 2.9: Palmprint [7].

Several authors [31, 30, 32] present algorithms based on in-touch palmprint scanners to data acquisition. Others [2, 33] have developed solutions based on contactless scanners to implement complex 3D solutions for palmprint verification systems. Instead of a scanner, a built-in camera of mobile phone can be used for a simple mobile contactless solution as shown in Figure 2.9. The pre-processing algorithm purposed by Aoyama et al. [7], using a HTC Nexus One, consists in applying skin-color thresholding and region growing. The algorithm was compared with previous

ones and better results were achieved. The EER was near 4% when a database obtained using a mobile phone camera was used. Moço et al. [34], developed a native biometric recognition system for the Android platform. The proposed system is based on palmprint features extraction using Orthogonal Line Ordinal Features (OLOF). From ROI extraction technique, the result were images representing a region similar to Figure 2.9 (b).

Hand geometry

Contrary to palmprint, which is similar to fingerprint, hand geometry identification systems are based on the dimensions and shape of fingers, palm and the location of joints [15]. Although they are not known to be very distinctive and should not be able to identify an individual among a large population, they can be used to authenticate and verify a subject [15]. Santos-Sierra et al. [35] implemented a Java-Android mobile phone identification system with EER around 6.8%. Images are taken with an uncontrolled background and captured freely by the user in very unconstrained conditions.

Finger Knuckle

The finger segment joints of human hand generate distinct texture patterns on the finger back surface. Given its uniqueness [36], knuckle features should be considered in biometrics too. In mobile domain, some research has been done with very interesting results like the work done by Cheng and Kumar [8] where the finger knuckle is captured using an Android Smartphone-based system.

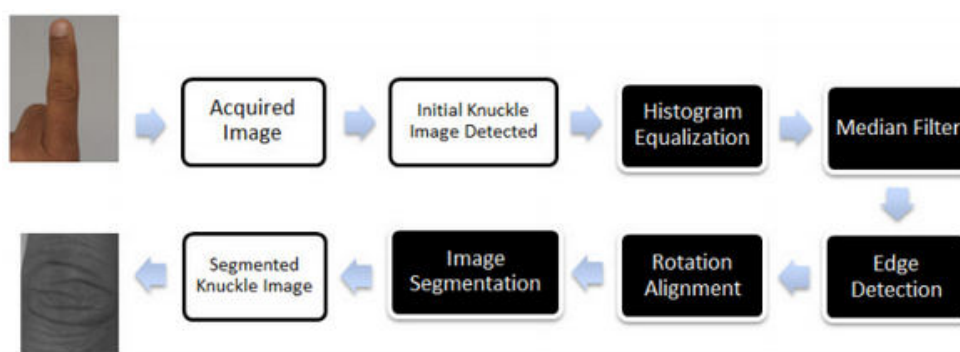


Figure 2.10: Finger Knuckle Pre-processing [8]

The image is captured and then pre-processed as seen in Figure 2.10. Features are, then, extracted by means of Gabor filters, which are highly suitable when the feature details are to be acquired from high-frequency areas [8]. For template matching, the similarity of the two templates is measured using Hamming distance method. This experiment had an EER below 10%, which is acceptable for personal mobile verification. However, for individual identification in a larger population it can contribute to consecutive system failures.

Another research that successfully implemented a contactless identification finger knuckle system was carried out by Zaw and Khaing [24]. Feature extraction is performed by Canny edge detector and Principal Components Analysis (PCA) methods. The extracted features are trained using Artificial Neural Networks (ANN) and the identification is done by the same ANN.

Face

Face is a commonly used biometric trait to perform personal identification. The process can be done by analysing the different facial attributes, such as eyebrows, nose, lips and chin shapes, together with their spatial relationship [26]. Illumination, pose, and age variations are the three major problems among face recognition systems [37]. Face identification mobile systems have been developed, but solely aimed to authentication [38] and face detection [39]. However, face is composed by the anatomical arrangement of several other traits, with potential to be used as singles traits themselves [9, 40]. Some of those traits are iris, periocular region and ear. Even being positioned laterally, ears can be considered as a region of the face, too.

Iris and Pupil

The iris is the annular region of the eye, having the pupil and the sclera (white zone) as its boundaries, Figure 2.11. Its visual texture stabilizes during the first 2 years of life and its complex structure gives very distinctive and useful information for personal identification. A mobile-phone implementation has been developed by Cho et al. [9] for iris and pupil localization only, by using the built-in cellular cameras and software.

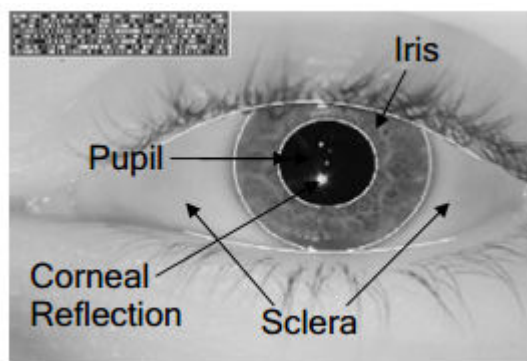


Figure 2.11: Human Iris [9]

First, corneal specular reflection (SR) is used to extract the eye region. Then, with Circular Edge Detection methods, the pupil and iris boundary are found. Cho et al.'s implementation does not cover iris identification. However, Jeong et al. [41] present a method, based on iris code extraction, with EER between 0.08 and 1.5% (depending on Indoor/Outdoor acquisition scenarios). The method is based on Adaptive Gabor Filter (AGF). The kernel size, frequency and amplitude of the filter are determined by the amount of blurring and sunlight in input image,

adaptively. Problems related with iris image capture quality are among this trait cons, creating an high demand for high quality cameras and stable environments to be collected. In order to surpass this, a new trait has become the focus of some studies during the last few years, as it will be explained in the following section.

Periocular

Periocular region is commonly described as the region in the immediate vicinity of the eye [42, 43, 44]. It can be motivated as a middle point between iris and face recognition. This trait has been the target of several research works [44, 42, 45, 13] with different motivations. Monteiro and Cardoso [44] present the periocular region as an alternative to face and iris recognition under unconstrained image acquisition conditions. They propose a novel automatic modeling of Scale Invariant Feature Transform (SIFT) descriptors, using a Gaussian Mixture Model Universal Background Model (UBM). UBM is a common strategy in the field of voice recognition [46]. Another research was carried out by Oh et al. [40], where periocular features, combined with sclera, are used for identity verification. Their approach is based on Local Binary Patterns for template generation. Santos et al. [13] proposed the application of well-known periocular and iris recognition strategies to overcome the issues associated with mobile environments. Although the main goal was to prove the improvements made by fusing both traits, periocular region based recognition showed promising results as a standalone trait. Another relevant study was carried out by Santos and Proença [45] where an overview of the most relevant research works related to periocular recognition is presented.

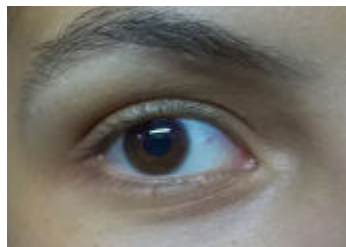


Figure 2.12: Example of periocular region from right eye

A common conclusion on the work developed by several authors is that the periocular region has potential to be an alternative to the most classical traits in unconstrained environments. Its ease of accessibility and high discriminating ability [45] make it a trait to take into account in mobile based applications.

Ear

The increasing attention that ear has received in the literature over the last decade is motivated by the fact that the detailed structure of the ear is not only unique [10], but also presents almost no degradation due to aging effect. It is usually fully grown at the age of eight and stays constant

until around the age of seventy [47]. Ear recognition, like in fingerprint, can be based on the analysis of several features, presented in Figure 2.13, as well as based on its geometry or space relation between its components. Pflug and Busch [10], divide ear recognition approaches into four different subclasses, namely holistic approaches, local approaches, hybrid approaches and statistical approaches. Abaza et al. [47] provide a detailed survey of research carried over on the fields of ear detection and recognition. Hurley et al. [48] present the Force Field Transform in which the ear image is transformed by considering an array of Gaussian attractors, which act as the source of a force field. Yuizono et al. [49] approach is based on Genetic Local Search and presents a recognition rate for the identification of registered persons of approximately 100%, and a rejection rate for unknown samples of 100%. However, the indoor data acquisition was done by using a face-fixed instrument, so that illumination condition are almost the same and the distance between the camera and the trait is almost constant.

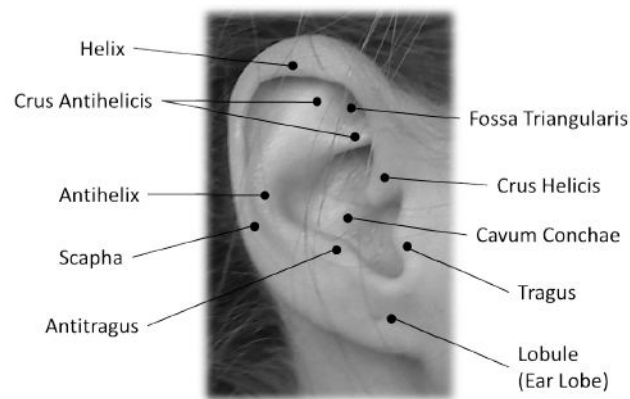


Figure 2.13: Human ear features [10].

Ballprints

Ballprints (the hallucal area under the big toe) possess important properties. If adjusted accordingly, their similarity to fingerprints allows the use of algorithms designed for fingertip biometric systems. However, this parameter is mostly suitable for newborns since the significant advantage is that newborns do not clench their toes contrary to fingers [50].

Voice

Voice is a combination of physical and behavioral biometric characteristics [28] that are used to create a sound. Among the physical characteristics are such like vocal tracts, nasal cavities, mouth and lips. Even if these characteristics were invariant over life-time, the behavioral part changes in time due to age, medical conditions and emotional states [15]. Voice identification systems acquire and match particular voice traits against templates stored previously, in a similar way to face recognition and other related-traits methods. In this work, only image-based systems will be

studied, however it is important to understand that biometrics is not only limited to image-based applications.

Other

The last sections served as a small introduction to the most commonly studied biometric traits. Besides those, there are still some alternatives which are being the target of research studies and even applied in real-life situations, like graphometric signature, finger vein pattern [51] and gait evaluation [52]. However, they are not suitable for this project's context or are still in a very initial stage of research. For example, graphometric signature in the SSA context, where great portion of the population is illiterate, could be hard to implement; Finger vein pattern demands specific hardware and proper light conditions; Identification by gait and other behavioral evaluations are non static methods, taking extra time, as well as demanding complex systems to be acquired and analysed.

Comparison Table

As seen before, no single biometric is the gold standard recognition tool. Some of the aforementioned techniques are compared within Table 2.1, taking into account the parameters presented in section 2.2: Universality (Univ.), Distinctiveness (Dist.), Permanence (Perm.), Collectability (Collec.), Performance (Perf.), Acceptability (Accept.) and Circumvention (Circ.). The evaluation is made by crossing some authors' [53, 23, 15, 26, 16] opinions with conclusions taken from the previously mentioned reviews [16]. Parameters are evaluated with a Good (G), Medium (M) and Bad (B) score that represent the recommendation level of the given trait for the specific parameter. For example, face characteristic is good in terms of collectability, however it has low performance compared with others.

Biometric characteristic	Collec.	Perm.	Dist.	Perf.	Univ.	Circ.	Accept.
Face	G	M	B	B	G	B	G
Periocular	G	G	M	M	G	M	G
Iris	B	G	G	G	G	G	B
Ear	G	G	M	M	M	B	G
Contactless Fingerprint	G	G	M	M	M	M	M
Contactless Palmprint	G	G	M	M	M	M	M
Hand Geometry	G	M	M	M	M	M	M
Finger Knuckle	G	M	M	M	M	M	G
Voice	M	B	B	B	M	B	G

Table 2.1: Traits comparison based on evaluation parameters presented on chapter 2.2

The parameters are presented by order of importance within this work context. Collectiveness is very important, since the data acquisition is made using a mobile phone system sensors that are not designed specifically to biometrics. Permanence is truly important because people are subject to marked physical changes during life. Distinctiveness is relevant when the target population is a

country with millions of people. Performance was considered of medium importance because, if the identification fails, a new data acquisition can be made to double check the results. Universality must be considered as well since the main goal is to develop a mean of integration for SSA people. Circumvention has lower importance within the given context since it is not a security system. Lastly, comes acceptability. This parameter is hard to evaluate and is presented based mostly on the trait's intrusion level during acquisition.

By looking into Table 2.1, voice can be easily excluded. Although it is good in terms of acceptance, it presents the worst scores in several parameters of high relevance. Another trait that is highlighted is the iris. Iris difficulties are inherent to the data acquisition, since it is hard to capture a good quality image with a low budget sensor. Face brings some problems on performance and distinctiveness, but should not be discarded. In order to surpass iris collectiveness and face performance related problems, the periocular region could be chosen due to its medium-high scores over the evaluated parameters. Similarly, the ear is presented as a very balanced trait, too. At last, the four traits related to hand features have an acceptable overall score, with fingerprint and palmprint taking advantage in terms of distinctiveness and permanence. From this overview, it can be concluded that there is no optimal trait. As it is shown in the next section, in order to improve the system performance, multiple traits can be used, in what is called a multimodal system, so that failures of some may be offset by the others.

2.5 Multimodal Systems

Overview

Multimodal biometrics does not only refer to the use of two or more separate biometric traits [11], it means, as presented in Figure 2.14, that the system inputs come from different sources:

- Single trait acquired by multiple Sensors;
- Single trait with multiple classifiers;
- Multiple instances of a single trait;
- Single trait and multiple units;
- Multiple Biometric Traits.

Multimodal biometric strategies are adopted in order to improve matching performance so that some of the unimodal limitations, like universality and spoof problems, can be surpassed.

System architectures

Usually, two main types of multimodal biometric system architectures are considered: serial and parallel. In serial (also known as cascade) architecture the evaluation of the different inputs is carried out sequentially. Therefore, the processing of the second input will be affected by the first

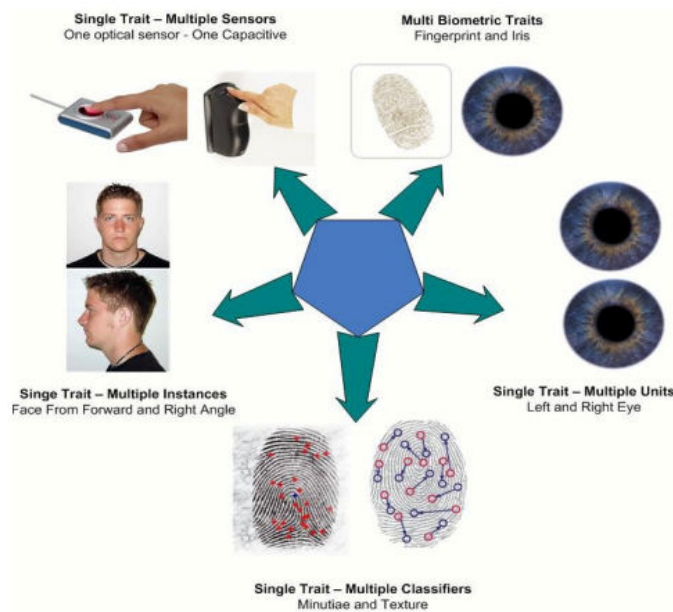


Figure 2.14: Multimodal biometrics [11].

one. In the parallel architecture, the different traits are processed independently and the resulting outputs are then combined [11].

Fusion levels

Depending on the design, the integration of two or more inputs can occur at different levels of the system. These fusion levels are: the sensor module, the feature extraction module, the matching module and the decision module. Due to the ease in accessing and consolidating matching scores, fusion at the matching module is the most popular mechanism [25, 54]. In the matching module, the independently constructed subsystems output scores can be consolidated into a single output. At decision level, the fusion can be seen as a *classification* problem where the feature vector is constructed by mean of the output matching scores by the individual matchers. This vector is then classified as an impostor (if rejected) or a genuine (if accepted). A different approach can be taken by considering fusion as a *combination* problem: matching scores are combined to generate a single score which is then used to make the final decision. Sometimes, a simple sum rule is sufficient to obtain a significant performance improvement, but techniques to incorporate user-specific weights can be integrated to further improve system performance [54].

Related work

Along the years, research has been made to evaluate the combination of different traits to achieve enhanced system performance under less ideal acquisition conditions. With all the available traits, plus new alternatives emerging every year, there are numerous possible combinations deserving

the attention of the biometrics research community. The hand itself can be used in order to extract multiple traits from a subject. For example, the dorsal hand surface is rich in key features like fingernail, vein pattern, knuckles, and hand geometry and it is widely used by researchers targeting multimodal systems. In Kale et al. [36] research, a knuckle and fingernail based identification system is implemented with the analysis of a single hand per subject. Features are extracted using Mel Frequency Cepstral Coefficient (for knuckle prints), wavelet decomposition (for nails) and a multi-Layer perceptron neural network is used as a classifier. The fusion is done at feature level, and the authors refer to some difficulties in implementation. However, the presented results prove the high efficiency in applying fusion algorithms. Other works have also explored the idea of multimodal biometric fusion. For example, Uhl and Wild [23] show that the combination of fingerprints, palmprints and hand geometry from single palm images of the human hand can successfully increase recognition rates. Multimodality can improve the performance of a biometric system, however, the way the information encoded in an image is represented also plays an important role. On the next section four well-known descriptors will be presented to motivate its use in the present work.

2.6 Descriptors

Descriptors encode images so that they can be compared and matched to other images. Global descriptors describe the whole image, while local descriptors describe a defined region or points within the image. Compared with multiple local descriptors, global descriptors are not very robust as a change in a small portion of the image may cause the system to fail as it will affect the resulting descriptor. Another characteristic that distinguish descriptors is its representation. There are binary descriptors like BRIEF [19], BRISK [55], ORB [4], and FREAK [56], and histogram descriptors like the ones presented next.

Scale Invariant Feature Transform

The Scale Invariant Feature Transform (SIFT) local feature descriptor was introduced by Lowe [12]. Although many algorithms have been proposed after Lowe's, SIFT still prevails as one of Computer Vision's algorithms that provides a more stable performance regardless of the image's real-world acquisition conditions. Computation of local feature descriptors and feature detection are usually seen as independent steps. However, Lowe proposes the SIFT descriptor for features detected with the Difference of Gaussians (DoG) operator, allowing a rotation and scale invariance of the resulting representation. Although the scale invariance is guaranteed by the stable set of candidate features points across different scales obtained with DoG, the orientation invariance is provided separately, as an effective method of assigning a repeatable orientation to the feature points is required. To that end, SIFT takes the feature points, also known as keypoints, as well as the detected scales and finds, from the scale-space pyramid, the image L associated to the closest scale to the actual scale of the keypoint. Given a keypoint at location (x,y) and scale s , the

gradient magnitude m and orientation θ are computed using pixel differences, as described by the following equations:

$$M(x,y) = \sqrt{I_x(x,y)^2 + I_y(x,y)^2}$$

$$\Theta(x,y) = \tan^{-1}(I_y(x,y)/I_x(x,y))$$

SIFT's final step is the concatenation of sub-regional histograms of gradients around each keypoint, resulting in a 128-dimensional descriptor that summarizes the local gradient information. This descriptor is built by first computing a set of 8-bin orientation histogram in 4×4 sample regions, as presented in the rightmost image of Figure 2.15, and then concatenating all those histograms into the $4 \times 4 \times 8 = 128$ -dimensional descriptor. As the keypoint dominant orientation is known, an offset can be applied to the histograms in order to secure rotation invariance. Robustness to illumination changes is improved by cropping and renormalizing the resulting vector to unit length.

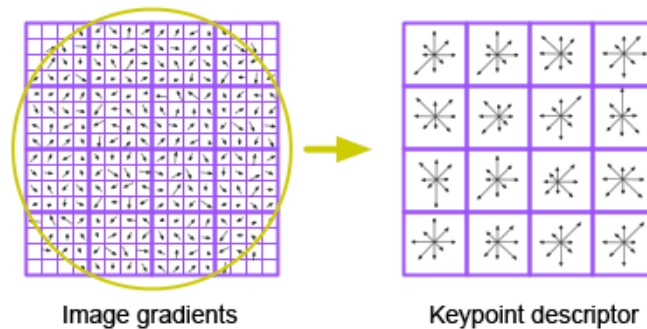


Figure 2.15: Gradient magnitude and orientation computation of a single keypoint using SIFT descriptor (adapted from [12]).

Lately, the SIFT descriptor has also been applied at dense grids, referred as dense SIFT. Experimental evaluations show that better classification results are often obtained by computing the SIFT descriptor over dense grids [57] in the image domain, as opposed to sparse interest points. This approach has been shown to lead to better performance for tasks such as object categorization, texture classification, image alignment, and biometrics.

Local Binary Pattern

The basic local binary pattern (LBP) operator, introduced by Ojala et al. [58], was based on the assumption that local texture has two complementary aspects: a local spatial pattern and its strength. The original LBP labels the pixels of an image by thresholding the 3×3 neighbourhood of each pixel with its center value considering, then, the result as a binary number. This histogram of 2^8 different labels is then used as a texture descriptor. Later, the LBP operator was extended so

that neighbourhoods of different sizes and shapes could be used, for example, applying circular neighborhoods and bilinearly interpolating values. Another variation of LBP is the uniform local binary pattern (ULBP) [59], which results in a more compact version of the feature vector. Such alternative was motivated by the fact that some binary patterns occur more often in texture images than others. Uniform patterns are those where, at most, two bitwise transitions from 0 to 1, or vice-versa, take place. For example, in a 3×3 neighbourhood there are a total of 256 possible patterns, 58 of which are uniform, leading to 59 different labels (58 for uniform patterns plus 1 label for non-uniform pattern). Ojala et al. [59] refer in one of their experiments that nearly 90% of all patterns in the 3×3 neighbourhood and around 70% for a circular neighbourhood, with radius 2 and 16 sample pixels, are uniform. Due to its contribution to the significant progress in texture analysis, LBP methodologies have been widely used and studied. This fact led to the rise of numerous new approaches. For example, the Volume Local Binary Pattern proposed by Zhao and Pietikäinen [60] in which the original LBP operator, once defined to only deal with spatial information, was extended to a spatiotemporal representation for dynamic texture analysis. Another example is the Center-Symmetric Local Binary Pattern, proposed by Heikkilä et al. [61], which integrates a novel descriptor that combines the strengths of both SIFT and LBP. CS-LBP replaced the gradient operator used by SIFT. Mäenpää and Pietikäinen [62] proposed an opponent color LBP by investigating joint and separate use of color and texture in classification. Other approaches, Tan et al. [63] and Wang et al. [64], were taken by combining LBP and Gabor features, the second work focusing on facial age classification. Heikkilä and Pietikäinen [65] were the first to developed a texture-based method for object tracking using background subtraction. The wide range of applications and the research made around LBP have proven its potential and its importance in Computer Vision and in its related fields.

Histogram of Oriented Gradients

Swain and Ballard [66] initiated a research on histogram-based image descriptors by showing that good performance could be achieved by comparing RGB histograms of pictures of objects not having in account the spatial relation between image features at different points. Almost 15 years later, and after several improvements and new approaches on histogram-based descriptors, Dalal and Triggs [67] developed a descriptor defined from a set of gradient orientation histograms. Similar to SIFT, their approach was inspired by the discriminatory property of local position-dependent gradient orientation histograms computed over a grid in the image domain. Compared to SIFT, which is a local image descriptor, Histogram of Oriented Gradients (HOG) descriptor is a regional image descriptor. As the HOG descriptor is not normalized with respect to orientation, it is not rotationally invariant, however, it is normalized with respect to image contrast. Dalal and Triggs showed that this descriptor allows a robust detection of humans in natural environments. HOG divides the image into small connected regions, which are called cells. For each cell, a histogram of gradient directions or edge orientations is computed for the pixels within that cell. Each cell is then discretized into angular bins according to its gradient orientation. The gradient magnitudes of the pixels in the cell are used to vote into the orientation histogram. Adjacent cells

are grouped as a spatial region, which is called a block. Grouping the cells into a block is the basis for normalization of the histograms. A normalized group of histograms represents the block histogram, and the set of these block histograms is called HOG descriptor.

GIST

The global descriptors are used to obtain a basic and superordinate level description of the perceptual dimensions. Some experiments have demonstrated that Humans can integrate enough information about the meaning of a scene in less than 200ms and in fact we recognize its “gist” as quickly and accurately as a single object. In its first appearance [68], GIST was used to describe real world scenes. GIST descriptor encodes the image scenes where the distance between the observer and a fixated point is larger than 5 meters. At that point, 5 perceptual dimensions are used to give a low dimensional and holistic representation of the image [69]. Naturalness, where scenes having a distribution of edges commonly found in natural landscapes would have a high degree of naturalness whereas scenes with edges based toward vertical and horizontal orientations would have a low degree of naturalness. Openness where the existence of a horizon line and the lack of visual references and boundary elements confer to the scene a high degree of Openness. Roughness which is correlated with the fractal dimension of the scene and thus, its complexity. Expansion which describes the depth in the gradient of the space within the image. Ruggedness which gives the deviation from horizontal by assessing the orientation of the contours of the image.

2.7 Commercial Solutions

As previously stated, biometrics is booming in SSA. A lot of new biometric applications are starting to appear in those countries especially for governmental purposes like e-passports, immigration control, and voter registration. In *BiometricUpdate*'s website [70], news can be found about solutions implemented in countries like Nigeria, Gambia, Namibia, Ghana and Uganda since 2014. However, most of them are not smartphone based and they make use of dedicated devices instead, like the ones brought by Suprema [71] and Zete [72]. Among the solutions emerging in SSA, there are enterprises that brought reliable mobile phone related solutions that are already being adopted by some of those countries. Comes next, a review of some available multimodal solutions on the market related to the main objectives of this work.

MorphoTrak

SAFRAN's [73] is one of many worldwide enterprises that provides biometrics technology. It makes use of fingerprint, palm print, unsolved latents, facial, iris, and others, to fight crime, protect identity and enhance security. To match the acquired data, a fast and accurate matching software called MetaMatcher is used. However, these technologies are only supported by a specific hardware, providing an almost ideal acquisition environment. MorphoTrak is already on the SSA

market: in Senegal, it is being used for digital identification and smart transactions, and in South Africa for individual identification.

MegaMatcher

Neurotechnology's [74] software developer kit (SDK) provides technologies that combine fast algorithms with high reliability. MegaMatcher deals with fingerprint, face, eye iris, and voice. These four solutions can be implemented in MegaMatcher SDK and they support mobile android applications, too. The solution is paid and the use of software requires a license. Some solutions based on MegaMather have been already implemented with success in some SSA countries like Republic Democratic of Congo [75], South Africa [76] and Kenya [77]. In Kenya, the voter registration system is based on MegaMatcher multi-biometric technologies.

BioComponents

Aware's [78] BioComponents are modular configurable biometric enrollment software components with a built-in user interface. Each component is independent and self-contained, and each of the modules is specific to a biometric task: data capture and validation, biometric image capture and processing, hardware abstraction, quality assurance, or networking. These APIs are well-established within the Aware SDK and they can be used within either Microsoft C# or Java applications. Aware software is also suitable for supporting multimodal biometric enrollment applications from a single platform. From the same company comes AwareXM Mobile [79]. This SDK includes C libraries and reference applications for performing fingerprint minutiae extraction, template generation, and 1:1 matching on mobile devices running Android, iOS, Blackberry, or Windows Phone operating systems. Fingerprint templates generated on a mobile device can either be matched on-device or transmitted to a server-based application for verification against a central template database.

In Table 2.2, the commercial solutions presented before are compared in terms of: Purpose, if they aim to authentication or identification; Mobility, if it is implemented on a mobile phone; SDK, if there is an available software development kit; Price, what is the cost of the solution or if it is free; SSA Market, if the company has some solution implemented in any of the SSA countries.

Product	Purpose	Mobile	SDK	Price (euros)	SSA Market
BioComponents	Authent.	Yes	Yes	Not-Free	Unknown
MegaMatcher	Identif./Authent.	Yes	Yes	2500+	Yes
MorphoTrak	Identif./Authent.	No	No	Not-Free	Yes

Table 2.2: Comparison table of some the available commercial solutions

Some conclusions can be taken from the analysis of Table 2.2 and the literature review. Companies found a market that offers a lot to explore in terms of biometrics; some of them, are already providing products for verification and identification purposes; companies offer personalized solutions (like SDKs) to let the client shape the solution to his needs; SSA governments are very

receptive to those solutions since these automatic systems greatly improve the society organization when compared with the traditional methods.

2.8 Summary

In order to achieve the optimal performance, a biometric system should be designed within the context. In this project, due to considering that the solution is to be designed for SSA countries, where poverty prevails in several regions, the acquisition conditions are confined to open fields or small vilages, and not to proper facilities with controlled ambiance light and environment. Given that, a lot of noise is expected and a robust solution must be adopted to bring real benefits. To mitigate that noise, a multimodal system must be implemented, making use of more than one trait to gather enough meaningful and discriminative information about a subject. To evaluate the best available traits many parameters can be taken into account. However permanence, as identification should be able to be performed over a life time, distinctiveness, since the system will be implemented for identification purposes, and collectability, given the limitations of mobile phone hardware, are the most relevant characteristics to be taken into consideration. The image codification to optimize comparison and get a more compact representation of the image data is achieved by means of feature descriptors. SIFT, HOG, LBP, and GIST are among the most used histogram descriptors, which already proven their value in biometrics. These descriptors' performance can be evaluated by the decidability for verification and by Rating Error Rate for identification systems. On the next chapter, a dataset, called MoBIAC, acquired to study SSA oriented algorithms, will be presented and described. Then, on the following chapter, the chosen algorithm will be explained and its performance will be tested over CSIP, a periocular dataset. Afterwards, that algorithm will be extrapolated to the MoBIAC dataset to evaluate its performance over different traits.

Chapter 3

MoBIAC Dataset

It should be always taken into account that the most effective way to test an algorithm is to apply it directly in the target population. However, that is not always possible to achieve and, usually, it restrains the development time. In order to surpass that problematic, available datasets can be used to carry system tests in a first phase when the target population is not reachable. Given this work's context, a good SSA representative dataset should contain at least images from African individuals under variable lighting conditions. Next comes a brief review of the available datasets for some of the traits presented in Section 2.4.1.

On Grgic and Delac's *Face Recognition Webpage* [80], more than 70 databases of face images are available with all kinds of characteristics from people around the world: Asian, Middle-East, South-American, etc. Some of them are almost noise-free, while others contain images where the individuals are wearing glasses, for example. However, no African individual is stated on databases' description, except for one that contains 15 African-American faces from 18 to 30 years old. Bosphorus [81] is a database of hand images available for free download with pictures of more than 600 hands. The hand geometry, plus texture data, was acquired using a commercial scanner, with hands placed flat on the glass plate and, according to the description, 180 of the subjects have hand images with time lapses of several months. There are more databases on CVonline [82], some of them with respect to face, fingerprint, voice, and some non-biometric objects. Sequeira et al. [83] proposed a dataset composed by three modalities: iris, face, and voice. Iris and face images from 105 subjects were captured with an ordinary tablet under real unconstrained working conditions. Santos et al. [13] built an iris and periocular dataset of 50 subjects, with images captured from 10 different mobile phone setups under unconstrained pose, illumination, and Expression (PIE) conditions. IIT Delhi ear image database [84] was acquired from 121 different subjects, with at least three ear images per subject. All the images were acquired from a distance (touch-less) using simple imaging setup and the imaging is performed in the indoor environment. A more complete ear dataset is presented by Frejlichowski and Tyszkiewicz [85] with 3000 images from both ears of more than 500 subjects. There is a large variation in image quality in respect to PIE, earrings, and hair covered areas.

In order to compare some of the available free datasets, it is presented a table that takes into

account some important characteristics: chosen traits; if data was acquired through a mobile phone or not (web-cams were considered equal in quality to mobile phones); level of freedom in data acquisition; number of subjects presented in the dataset.

Dataset	Trait	Mobile	Constrain Level	Subj. Number	African
ChokePoint [86]	Face	No	Low	25	Unknown
MOBIO [83]	Face and Iris	Yes	Low	105	No
JAMBDC [78]	Face	Yes	Medium	200	Unknown
BioID [87]	Face	Yes	Low	23	No
JAMBDC [78]	Palmprint	No	High	350	Unknown
CSIP [13]	Periocular	Yes	Low	50	No
IIT Delhi [84]	Ear	Unknown	High	121	No
WPUTEDB [85]	Ear	Unknown	Low	501	No

Table 3.1: Comparison between some of the available datasets

Analysing Table 3.1, it can be concluded that there are a lot of freely available datasets on the Internet provided by research groups. The dataset search was focused on low quality data acquired in mobile-based setups. All the reviewed datasets are missing an important keypoint for this work, as they lack on images acquired from SSA subjects. Here lies a motivation to build a new dataset that allows a performance comparison on biometrics algorithms between different ethnicities. The existing gap on African mobile images datasets can be filled by the data collected within the scope of this work, representing an important contribution for related future works. Some conclusions can be taken from the dataset analysis resumed on Table 3.1: several samples should be taken from the same subject, from different positions, and in different light conditions enhancing the dataset variability. The more subjects, the more relevant is the dataset to a real-life situation context. Regarding the scope of this work, the development of a SSA-based dataset represented a serious challenge, given the amount of time and the difficulty to communicate and coordinate an acquiring process in one SSA country. With such limitations in mind, a different solution was taken. MoBIAC was created by gathering information about both Caucasian and African subjects, to assess the performance of algorithms aiming worldwide applications.

3.1 Dataset Overview

MoBIAC multimodal database construction was motivated, on one hand, by the raising worldwide interest in mobile biometrics and the increasing interest in multimodal systems and, on the other hand, by the lack of biometric information regarding SSA individuals. This motivation led to the creation of a dataset comprising periocular, ear, and hand samples, acquired in unconstrained conditions using a mobile device camera. As stated in Section 2.4.1, hand can be analysed by its geometry or by its texture (palmprint). Both modalities have proven promising results in individual identification, but a study on its performance on SSA countries was never performed. In Figure 3.1, hands from several ethnicities may be observed, with marked differences being easily noticed.



Figure 3.1: Hands from different ethnicities. The top-left photo is a typical SSA individual photo while others are related to different ethnicities. There are obviously marked differences on principal lines contrast.

Periocular region has been target of several studies aiming image analysis of samples acquired in unconstrained conditions using a mobile device, as previously mentioned in Section 2.4.1. Once again, the novelty of MoBIAC concerns its target population. With samples from both African and Caucasian subjects, it is possible to compare the systems performance on those two groups and to evaluate how well an algorithm can perform when applied on a mixed group. To finalize, the ear was chosen to be part of this dataset, on one hand, for its ease of capture and high acceptability by people and, on the other hand, for its increasing relevance in individual identification. The creation of this dataset will be a relevant resource for future research and its value goes beyond its application in this Master's Thesis work.

3.2 Dataset Description

In order to perform a comparison between some biometric recognition algorithms performance on Caucasians and on SS African individuals, the dataset was made half caucasian representative,

half SSA representative, totalling 50 volunteer individuals. The nationality of the 25 Caucasians was Portuguese, with ages between 20 and 30 years old. For the SSA individuals, there were 5 Mozambicans, 3 Angolans, 2 South African and 15 Cape Verdeans, with ages comprising between 20 and 30 years, except for 2 subjects older than 30. There are 4 and 3 females on Caucasians and Africans groups, respectively. From each volunteer, at least 6 images per trait were extracted, performing a total of 1148 images. Only the right ear, hand, and periocular region were considered.

Images were acquired by using the Lumia 830 10MP rear camera, without flash. The flash choice was made based on the idea that the flash itself represents an imposed condition. Photos were acquired with a resolution of 3840×2160 with auto-focus. Blurry images were sometimes acquired by the natural difficulty on the phone to focus on dark environments and sometimes acquired by not giving time to the the lens to focus.

Each volunteer was asked to come near a window, to be exposed to natural light. There were 2 different acquisition scenarios: one at Fraunhofer AICOS and another one at FEUP. However, there was an attempt to provide similar acquisition environments in both places. Samples from the 3 traits were acquired sequentially and were all taken by the same person. The illumination variance was achieved by the rotation of the volunteer in approximately 60 degrees. One photo is granted to be taken in counter-light and other in front-light and there is at least one blurred photo as well. Singularities related to each trait acquisition will be briefly described next.

Hand

For hand acquisition, it was asked of each subject to elevate its forearm in a vertical position. Photos were taken so that all the hand and palm was captured, but it was not said to the volunteers how they should open or join their fingers. It was noticed that each subject had a very personal manner of showing his hand. On Figure 3.2, are presented some examples of observed hand poses.

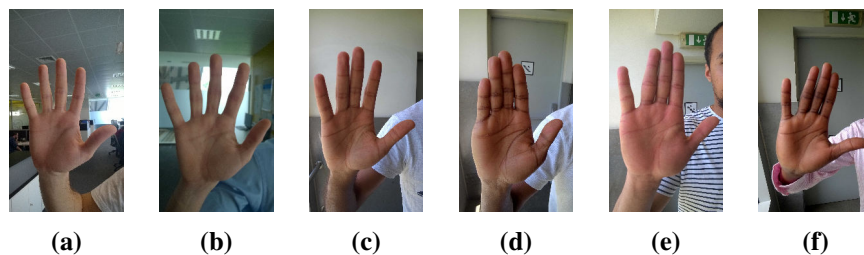


Figure 3.2: MoBIAC Hand samples.

By analysing Figure 3.2, it is possible to distinguish a certain level of pose variation. Even for the same subject, images (c) and (d), there is a difference on finger's position. Image (a) presents an hand slightly turned to the left while image (f) presents an hand with a rotation of approximately 30 degrees. These kind of variations fulfill the objectives of the MoBIAC dataset by imposing a low level of constraints.

Periocular

Periocular region photos captured almost half of the individuals' head. As it is not a well defined region it was captured more information, than could be later needed, instead of later concluding that pertinent information was left out. For the periocular region the only applied constraint was that the phone should capture the whole eye. There were no demands regarding where the eye was looking at. With that in mind, it was observed that, in some of the photos, the eye was closed at the moment of capture. In some photos the eye was looking directly to the camera while on others it was looking left or right. It was noticed that in some photos acquired in counter light, the amount of information related to the eye is almost none. In the following Figure 3.3, some examples of photo variations, observed during the acquisition of periocular images, are presented.



Figure 3.3: MoBIAC Periocular examples.

As it is possible to notice, on Image (a) the individual is almost impossible to be recognised, due to the low luminosity observed on the face region. On Image (b), the subject's gaze is not oriented towards the camera which, for example, can eliminate important information in the recognition process. Images (c) and (d) represent the best-case scenario, with information from the area within the eye and its boundaries properly retrieved. These kinds of variations serve the purpose of this dataset, as a challenge to assess the robustness of a given algorithm.

Ear

Ear photos were captured simply by taking a photo of the right ear with the only imposed constraint being that all the ear should be presented within the image and placed in front of the camera with an inclination of no more than 30 degrees. Some images contain earrings and others are partially covered by hair. In these cases, hair was only removed partially from the image when it was almost impossible to retrieve the ear information. On Figure 3.4, a brief representation of some MoBIAC's ear images is presented.

On Figure 3.4, some characteristic examples of unconstrained scenarios can be observed. On Image (a), a blurry photo from a subject is depicted. On Image (b) a situation where the ear is partially covered by hair and on Image (c) there is a piercing in the top-left region. Image (d) is representative of a captured noise-free image. Different light conditions are, however, still noticeable from photo to photo. This variation was imposed by taking a unique approach regarding the acquisition setup, which will be detailed in the following section.

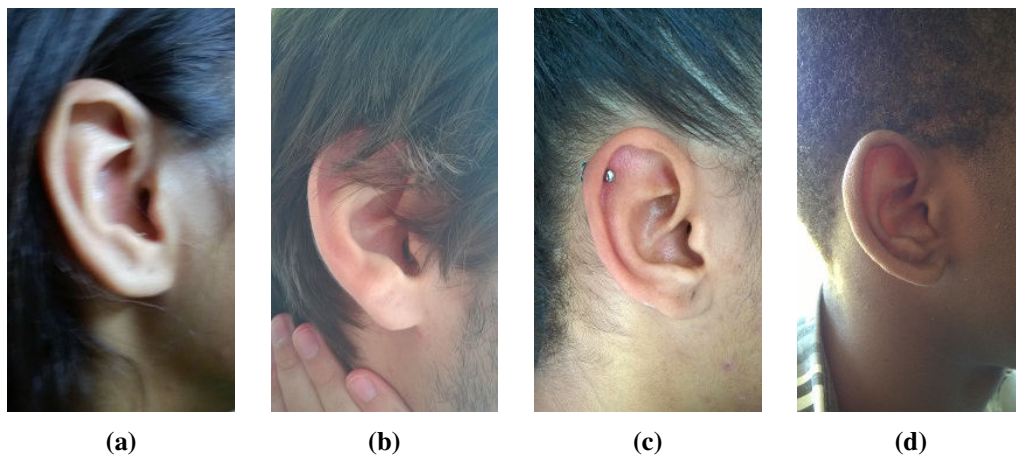


Figure 3.4: MoBIAC Ear examples.

3.3 Acquisition environment

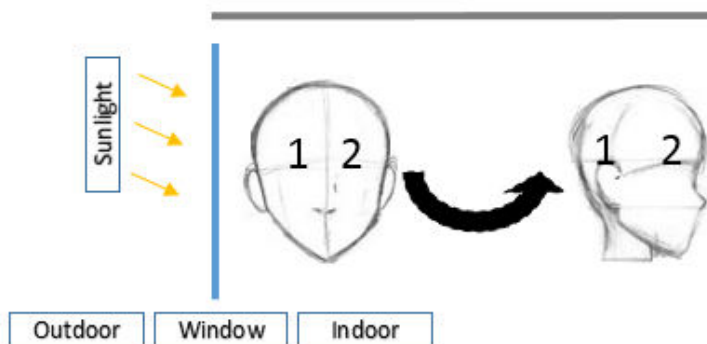


Figure 3.5: Example scheme to illustrate the light variance approach. Region 1 is enlightened by the external light while region 2 is darker because of the counter-light effect.

The MoBIAC dataset was acquired in cooperation with the student community within FEUP and Fraunhofer. The data was collected throughout 5 different days, depending on the availability of each subject. To capture images with different light conditions in a strict and constrained time-period, a unique approach had to be taken to speed-up the collection. The approach to collect data for the MoBIAC was based on varying the position of the subject, with respect to a fixed light source. Each subject was asked to approach a window, wall sized, and then to rotate while the photos were taken. As stated before, at least an image from 6 different positions, from each trait, from each subject, were captured. On Figure 3.5, a scheme of the method to simulate light variance, used, as described above, for image acquisition, may be observed.

Chapter 4

Periocular Recognition System

Periocular recognition under unconstrained scenarios has been the target of several studies over the last few years as stated in Chapter 2.4.1. Some studies focus on understanding how the environment itself can influence the system, by assessing its performance in different conditions, while others focus on developing novel approaches that can handle with most environment variations. What is proposed, on this chapter, is the analysis of a novel algorithm and its application regarding data captured in unconstrained scenarios, but also the evaluation of how those external variations relate to its performance.

The algorithm proposed by Monteiro and Cardoso [44] was chosen to be evaluated for periocular recognition under unconstrained scenarios. Their algorithm is based on Universal Background Model (UBM) using Gaussian Mixture Models (GMM). Although UBM is more commonly applied in the field of voice biometrics [46], the authors proposed its application in periocular recognition, with some interesting results. The algorithm makes use of SIFT keypoint descriptors to perform individual recognition based on periocular region images. Performance was assessed on both MobBIO [83] and UBIRIS.v2 [88] datasets, presenting state-of-the-art performance [44]. In this chapter, an adaptation of the Monteiro and Cardoso's algorithm will be proposed to make use not only of SIFT but also of HOG, ULBP, and GIST descriptors. The comparative performance of such descriptors will be assessed on the Cross Sensor Iris and Periocular (CSIP) database [13]. Furthermore, the results of some score-level fusion methods are presented, evaluating how the system behaves when using combined information from multiple descriptors. This chapter starts with a brief analysis of the proposed algorithm by Monteiro and Cardoso. Then, the CSIP dataset is presented, regarding its main features and detailing its unconstrained acquisition conditions. On the last section, a comparative analysis of the performance using multiple descriptors and fusion scenarios is carried out.

4.1 Algorithm overview

The algorithm presented next follows the typical architecture of biometric systems, as described in Section 2.3.1. It is composed by two main blocks: an enrollment stage followed by, depending

on the functioning mode, an identification or verification stage. During the enrollment stage, individual data is added to a previously existing system database. During the enrollment, a set of N models which describe the unique statistical distribution of biometric features for each individual, is trained by maximum *a posteriori* (MAP) adaptation of a Universal Background Model (UBM). This UBM gathers information from all the individuals, and MAP adaptation adjusts its parameters into a set of individual specific model (IDSM), based on each individual's specific data. This idea of adapting a UBM using MAP was first introduced by Reynolds et al. [89] for the field of speaker recognition. After the training of IDSM and UBM, the recognition phase can be carried out. During this phase, the extracted features from an individual are projected onto both IDSM and UBM. The final recognition score is obtained by the output of the likelihood-ratio between both projections, and identification is then carried out by maximum likelihood-ratio.

4.1.1 Universal Background Model

UBM can be understood, if the problem of biometric verification is interpreted as a basic hypothesis test [44]. Considering the hypothesis, H_0 , of a given sample, Y , belonging to a claimed ID, S , and the alternative hypothesis of not belonging to it, H_1 , the optimal decision can be achieved by a likelihood-ratio test:

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ < \theta & \text{accept } H_1 \end{cases} \quad (4.1)$$

where θ represents the decision threshold for accepting or rejecting H_0 , and $p(Y|H_i)$ is the likelihood of observing sample Y when hypothesis i is considered true. The problem resides, therefore, in computing the likelihood values $p(Y|H_0)$ e $p(Y|H_1)$. Monteiro and Cardoso [44] stated that H_0 should be represented by a model λ_{hyp} that characterises the hypothesized individual, while, on the other hand, a $\lambda_{\overline{hyp}}$ model should be also designed, to characterize all the alternatives, H_1 , to the hypothesized individual.

In order to represent the space of all alternatives to the hypothesized identity, Reynolds et al. [89] proposed the Universal Background Model (UBM). By training the UBM on a large set of data, it is expected to cover a representative user space, including a relevant amount of sources of variability. On the next section, the strategy to efficiently model the UBM and adapt it to individual models is presented.

4.1.2 Hypothesis Modeling

The authors chose Gaussian Mixture Models (GMM) to model both UBM and IDSM. This choice was motivated by its capability of capturing the empirical probability density function (PDF) of a given set of feature vectors. On their approach, GMM are trained using diagonal covariance matrices which are then trained on sets of SIFT keypoints descriptors. Their choice is mainly motivated due to the observation that local descriptors work better than global ones when dealing with data acquired under non-uniform conditions. Monteiro and Cardoso propose performing a

Principal Component Analysis [90] (PCA) on the descriptor. PCA is a standard technique for dimensionality reduction and has been applied to a broad class of computer vision problems. It is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, called principal components. By applying this technique, they reduce the SIFT descriptor from 128 to 32 dimensions, significantly reducing computational complexity of the training phase, as well as improving the distinctiveness and robustness of the extracted feature vector.

4.1.3 Modeling UBM and IDSM

To efficiently train the UBM, a set composed of data from a vast array of individuals should be used, so as to cover a wide range of possibilities in the individual search space (Shinoda and Inoue [91], Monteiro and Cardoso [44]). Therefore, the UBM train is performed by fitting a k -mixture GMM to the set of PCA-reduced feature vectors extracted from all the individuals. As the UBM should represent the whole group of individuals, the IDSM should focus on individual data. This specialization of the UBM is performed by tuning the UBM parameters in a MAP approach, using individual specific data. This IDSM creation method surpasses the challenge of the determination of appropriate initial values of the parameters of a GMM. This step is important because a poor initialization may result in a weak model, especially if the amount of available data is small.

In order to understand the adaptation process of the UBM into the IDSM, proposed by Reynolds et al. [92], a comparison with the traditional Expectation-Maximization algorithm [93] can be made. The first step is similar to the Expectation step, where a set of sufficient statistics are computed from a set of M individuals specific feature vectors, for each mixture of the UBM. Then, considering diagonal covariance matrices, each UBM mixture is adapted using the newly computed sufficient statistics.

When both the IDSM and UBM have been trained, the system is ready to recognise new data from an unknown source. The identity recognition of the new data, represented by a PCA-reduced feature vector, is performed through its projection onto both the UBM and all the IDSM, on identification mode, or on the claimed IDSM, on verification mode. The recognition is computed as the average likelihood-ratio obtained for each tested feature vector, and then the decision is carried out by checking the conditions presented on Section 4.1.1, for verification, or by detecting the maximum likelihood-ratio value for all enrolled IDs, for identification. As some subjects are more likely to generate high likelihood-ratio values, using both UBM and IDSM is a sharp advantage, since the ratio between the IDSM and the UBM probabilities is a more robust decision criterion than relying only on the IDSM probability. Moreover, finding a global optimal value is simplified by using the likelihood-ratio with an universal reference. This works as a normalization step, mapping the likelihood values in accordance to their global optimal projection.

As already stated, the purpose of this chapter is to present a follow up of the described algorithm, by assessing its performance on periocular recognition under unconstrained environments,

but by using different descriptors as alternatives to SIFT. In order to accomplish this goal, a dataset composed of periocular images, obtained with a variety of acquisition setups, is needed. In the next section, the adopted CSIP dataset is detailed.

4.2 CSIP Dataset

Monteiro and Cardoso’s algorithm showed state of the art results when assessed on both UBIRIS.v2 and MobBIO datasets. On this section, the same algorithm’s performance is also assessed, but using the descriptors presented on Section 2.6: LBP, HOG and GIST, as well as SIFT. A different dataset, Cross Sensor Iris and Periocular (CSIP) database [13], was also chosen for its wide variability regarding mobile setups, posing a new challenge to the robustness of the algorithm.

The CSIP [13] database is a dataset of periocular region images, including iris information too. The CSIP dataset is composed of images from 50 participants, acquired regarding several mobile setups and varying acquisition scenarios. Iris segmentation masks are also provided allowing the assessment of performance of both iris and periocular segmentation and recognition algorithms on mobile environments. On Figure 4.1, the different acquisition setups used in the development of the CSIP database are presented.

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
Resolution	3264 × 2448		640 × 480		2592 × 1936	2592 × 1920		3264 × 2448	640 × 480
Flash	No	Yes	No	No	Yes	No	No	Yes	No
Setup ID	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0
									DR0

Figure 4.1: Details related to each CSIP’s acquisition setup [13].

Images were acquired from a total of 10 different setups, using default settings for both focus and white-balance. The dataset comprehends information from 50 participants, all Caucasians, 9 women and 41 men. CSIP dataset takes into account 8 different noise factors: multiple scales, chromatic distortions, image rotation, poor lighting, off-angle acquisition, out-of-focus images, deviated gaze, and iris obstruction, for example by reflection. On Figure 4.2, image samples from the CSIP database regarding each of the 10 different setups are presented.

4.3 Proposed methodology

The work of this Master’s Thesis is based on the algorithm outlined in Section 4.1. On the schematics presented in Figure 4.3, the main steps of the followed methodology are presented. The 4 steps presented on Section 2.3 are represented by the dashed blocks, with the exception of data acquisition, which was previously detailed on the Section 4.2. The first step, pre-processing, comprises two stages: segmentation and resizing. During the segmentation, CSIP periocular region images

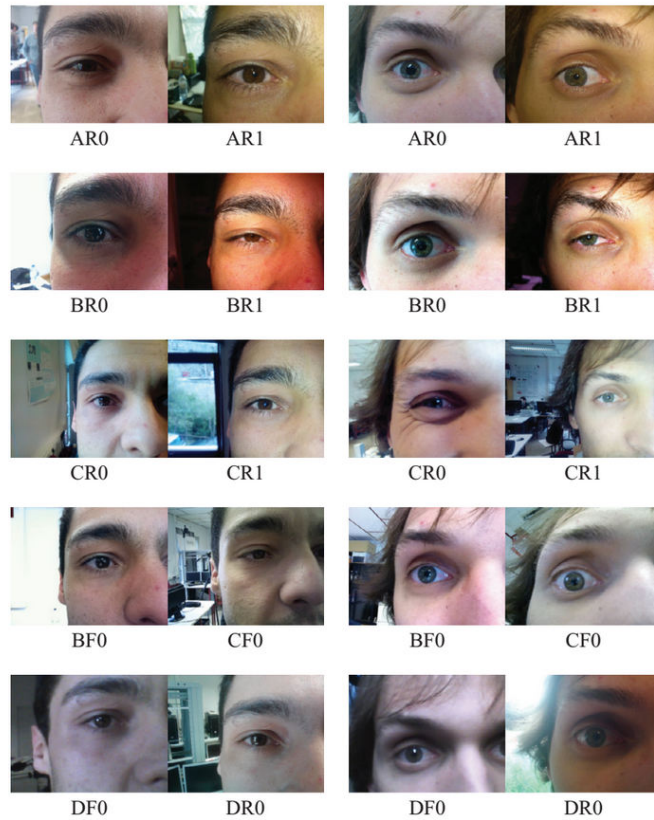


Figure 4.2: CSIP samples from 2 different subjects representing different setups [13].

are cropped using a similar method to the one proposed by Santos et al. [13]. On their work, the ROI is defined as a total of 35 square patches forming a 7×5 grid. The grid is placed accordingly to the iris segmentation mask, and each patch has an area equivalent to $1.4r_i^2$. On this work, and using the provided segmentation mask, the ROI is centered in the eye and the size of each grid patch is defined as a $r \times r$ square region, where r is the maximum measured radius within an iris mask. Resulting ROIs can be observed on Figure 4.4.

After ROI segmentation, each image is sized to 12.5% of its original size. The value of 12.5% was chosen by performing a serie of tests. Using images with 12.5% of their original size resulted in an improvement of 71% on R1 performance. Although excluding a huge part of information when performing a reduction of this size, the resizing step also removes a lot of noise, summarizing the image to its main features. Resizing is also motivated by the significant reduction in the computational complexity, resulting in shorter processing periods.

A great portion of the novelty of this work resides in the feature extraction and description steps. Instead of describing the image using only SIFT, 3 additional descriptors were implemented into the previously existing UBM-MAP framework. Similar to SIFT, the feature vector of LBP, HOG and GIST descriptors also represents an histogram. The challenge in the adaptation of the algorithm lies mostly on understanding how each descriptor works so it is possible to accurately use it. Although the original version of SIFT [12] was implemented, for HOG and LBP some

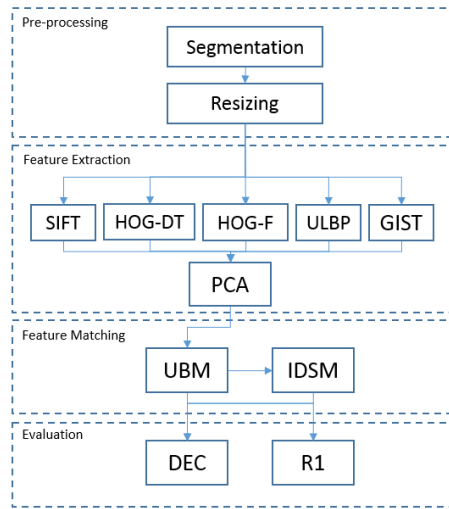


Figure 4.3: Schematic of the proposed methodology .

variations were used. Instead of the original LBP [58], the chosen descriptor was ULBP [59] since it demands less computing power as described in Section 2.6. As far as HOG is concerned, two versions were evaluated. The original one [67], proposed by Dalal and Triggs (HOG-DT), and an alternative version proposed by Felzenszwalb et al. [94] (HOG-F). This variation differs from the original since it computes both directed and undirected gradients, as well as a four dimensional texture-energy feature. At last, GIST was performed as described by its original authors [68]. The 5 descriptors, including SIFT, were processed independently and a PCA reduction was performed on the resulting feature vectors. On Table 4.1, are presented the specific dimensionality reductions applied to each descriptor.

Descriptors	Before PCA	After PCA
SIFT	128	32
Dalal-Triggs HOG	36	18
Felzenszwalb HOG	31	18
ULBP	59	16
GIST	32	32

Table 4.1: Feature vector dimension before and after PCA for each one of the 5 descriptors.

The UBM and IDSM are then created according to Monteiro and Cardoso’s algorithm. The models are trained with half of the dataset regarding each individual. The other half is used to test the algorithm’s performance.

The last step consists in assessing the system performance. For that purpose, two measures were taken into account: Decidability (DEC) and Rank-1. Both measures are described on Section 2.4. These two measures give information regarding performance on both verification (DEC) and identification (R1) modes. DEC represents the separability of two distributions, quantifying the separation between genuine and impostor likelihood score distributions. Higher DEC means



Figure 4.4: ROI cropped by Santos et al. method (a) and by this work's approach (b) regarding the same CSIP image.

higher aptitude to distinguish two IDs. R1 is the ratio of query templates for which genuine comparisons have the highest similarity values in database.

The process explained in the previous paragraphs was repeated for different descriptors, on different acquisition setups, for a variable number of GMM mixtures, $k = 2, 4, 8, 16, 32, 64, 128$, for each image color channel plus gray-scale, resulting in a vast amount of information to analyse. All of the presented results are the product of 10-fold cross-validation.

4.4 Experimental Results

During this section, relevant results related to the algorithm's performance are discussed. However, the full set of all the results obtained during this work may be consulted on Appendix A. Given the large amount to information, this section was divided by objectives, with each subsection concerning conclusions obtained regarding a specific objective. Although tests were taken regarding different aspects like setups, descriptors, and fusion scenarios, the presentation of the results will be carried out in a similar way for each of such aspects. To make it easier to analyse the results presented in Appendix A, an example of a result presentation table is shown next (Table 4.2), as well as a brief explanation on its most relevant information.

a) SETUP	b) DESCRIPTOR											
i) N	d) R G B				e) GS				f) RGB			
	g) R1 %		h) DEC		R1 %		Dec		R1 %		Dec	
c) GMM	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
2												
4												
8												
16												
32												
64												
128												
256												

Table 4.2: Example table to access Rank-1 ratio and decidability performance regarding SETUP images using DESCRIPTOR.

The previous table is similar to the ones gathered in the Appendix A. Each table has, on the top-left corner, an indication on the setup (a) in which a descriptor (b) was used. The leftmost column (c) indicates the number of GMM mixtures used to obtain the results performed in each respective row. The algorithm was tested for each of the color channels (d) (concatenated in this example to fit the page), to grayscale images (e) and by fusing the three color channels (f). For each combination the Rank-1 ratio (g) and Decidability (h) were assessed, and the average result among the N (i) tests is presented as well, as the corresponding standard deviation. Bold values represent top results.

4.4.1 k-Mixture GMM

Before starting the comparison of the chosen descriptors' performance, it is important to analyse the effect that the number of mixtures, k , in a k -mixture GMM has, globally, over the system's performance. By giving an overall look over the results presented in Appendix A, it is possible to understand that the best results are usually achieved on either $k = 8$, $k = 16$ or $k = 32$. At first sight, it could be expected that a larger k would originate better performance, but the results show that when k reaches a certain value, a break in the performance is noticed. This occurrence can be explained mostly by the too high complexity of a k mixture GMM chosen to train both the UBM and IDSM, resulting in a loss of generalization capacity due to *overfitting* of the training data. With respect to GIST with k higher than 16 the algorithm fails to compute. This can be explained by the fact that during GIST computation, the image is described only in 16 data points (one for each sub-image), a number that is smaller than the number, k , of mixtures in the GMM, hindering the training of the models for $k > 16$.

Motivated by the previous conditions, the comparison of the descriptors was carried out regarding the algorithm performance for $k = 16$, except when it becomes pertinent to analyse results obtained for different values of k .

4.4.2 Setup comparison

As presented on Section 4.2, there are 10 different mobile setups on CSIP comprising data acquired under different conditions. In order to evaluate the algorithm's performance on each setup, images were divided in 10 sets of data, each one regarding a setup. For each set, the algorithm previously presented was performed using one of the 5 descriptors at a time. Results obtained on each of these iterations will be discussed from now on.

By looking into the R1 and DEC values presented on Table 4.3 and Table 4.4, it is possible to conclude that AR1 setup performed the best results for both DEC and R1. This is explained by taking into account some aspects. By comparing AR1 and AR0 results, as well as other available mobile-phone related pairs namely $(CR1, CR0)$ and $(BR1, BR0)$, the single variation of using flash nearly doubled both R1 and DEC values. Flash can enhance certain features while uniforming the image lighting, by compensating otherwise shadowed areas. Still, it is important to have in mind

that designing a system limited by the use of a flash light is adding a constraint related to the image acquisition device.

	R1 %				
	ULBP	HUG-F	HOG-DT	SIFT	GIST
AR0	45.3 ±2.3	52.2 ±2.0	45.1 ±2.6	58.8 ±2.9	84.4 ±3.0
BR0	37.0 ±2.6	37.7 ±1.6	42.7 ±1.6	37.3 ±2.9	85.6 ±1.9
CR0	29.9 ±2.4	42.2 ±2.5	28.4 ±2.7	39.6 ±2.6	73.2 ±1.6
DR0	45.6 ±2.3	43.6 ±3.3	35.9 ±3.0	36.4 ±4.2	78.4 ±0.8
AR1	78.1 ±1.8	83.7 ±1.2	80.9 ±2.8	96.6 ±0.8	95.9 ±0.8
BR1	66.4 ±1.4	78.4 ±1.6	67.3 ±2.1	89.4 ±2.1	91.5 ±1.1
CR1	49.5 ±1.2	65.0 ±0.9	58.5 ±2.4	82.6 ±1.3	86.3 ±1.4
BF0	36.6 ±1.3	38.2 ±2.5	34.0 ±1.1	39.5 ±2.3	72.8 ±3.7

Table 4.3: R1 performance comparison between setups for 16-mixture GMM using GS images

	DEC				
	ULBP	HUG-F	HOG-DT	SIFT	GIST
AR0	1.86 ±0.06	2.22 ±0.07	1.87 ±0.10	2.53 ±0.06	3.73 ±0.09
BR0	1.64 ±0.03	1.86 ±0.17	1.94 ±0.04	1.77 ±0.04	3.54 ±0.13
CR0	1.49 ±0.07	1.90 ±0.07	1.63 ±0.03	1.98 ±0.04	2.87 ±0.12
DR0	1.56 ±0.06	2.31 ±0.04	1.98 ±0.03	1.70 ±0.09	2.84 ±0.15
AR1	3.27 ±0.08	3.06 ±0.06	2.91 ±0.10	5.06 ±0.07	5.13 ±0.21
BR1	2.07 ±0.05	2.74 ±0.11	2.40 ±0.11	3.86 ±0.13	3.98 ±0.09
CR1	1.99 ±0.05	2.44 ±0.04	2.35 ±0.09	3.54 ±0.05	3.76 ±0.12
BF0	1.55 ±0.04	1.90 ±0.17	1.63 ±0.10	2.19 ±0.05	3.15 ±0.04

Table 4.4: DEC performance comparison between setups for 16-mixture GMM using GS images

Resolution could be another highly influencing factor over performance. By comparing different setups, it can be concluded that higher resolutions are not directly related to better performances, sometime even contributing to its deterioration. It is true that higher resolution photos present more data regarding individuals, however, a non-negligible part of such information might present only noise to the system. In a large pool of IDs, the descriptor codes a higher amount of information and, thus, a larger amount of noise, resulting in a loss in performance. Given this, images are submitted to a size reduction of 12.5%, reducing the high amount of undesired information. With this lowered resolution, performance showed to be almost independent to the new images' size, exemplified on Table 4.3 by comparing BF0 and CR0 performances, which presented similar results despite their completely different resolutions, 640×480 and 3264×2448 respectively.

The system performance when using rear or frontal cameras was assessed as well. Since the resolution is not the cause of performance loss, it could be assumed that it is mostly related to auto-focus features and by the absence of the flash on frontal cameras. These statements are supported also by the comparison between results from the BF0, frontal cam, with BR0, rear cam, which present similar performance. Still, given the application context there is no special interest

in performing tests regarding frontal cameras and so these setups will not be considered from now on.

4.4.3 Color channel

The description of the image was performed for all the color channels, on the RGB (red, green and blue) color system, plus a fourth alternative obtained by gray-scale (GS) transformation. The RGB values are converted to grayscale by performing a weighted sum of the R, G, and B components:

$$0.2989 \times R + 0.5870 \times G + 0.1140 \times B$$

For setups without flash the R channel prevailed on achieving better results over G, B and even GS. However, for flash setups no direct relation was observed between performance and RGB color channels. The GS images resulted, usually, in good performances, achieving, at least, results near the best color channel, as presented on Table 4.5.

	R1 %				DEC			
	AR0		AR1		AR0		AR1	
	R G B	GS	R G B	GS	R G B	GS	R G B	GS
SIFT	58.5	58.8	98.1	96.6	2.64	2.53	4.79	5.06
HOG-DT	46.4	45.1	80.7	80.9	2.00	1.87	2.96	2.91
HOG-F	55.3	52.2	83.7	83.7	2.23	2.22	3.16	3.06
ULBP	47.8	45.3	75.9	78.1	2.07	1.86	3.10	3.27
GIST	86.6	84.4	97.0	95.9	3.81	3.73	5.22	5.13

Table 4.5: Comparison between the **best** color channel and GS for the 5 descriptors, with a 16-mixture GMM, for AR1 and AR0 setups.

In Table 4.6, a comparison between GS and the worst performing color channel can be observed. GS results are clearly better than the worst performing color channel, but only slightly worse than the best channel. Given this, GS can be a good alternative to the RGB colorspace on cases where there is not an obvious best channel, like in the flash setups. For setups in which images were taken without flash, the R channel prevailed. However, such observation does not lead to the conclusion that the red channel is the best for all the applications. The GS transformation, with results similar to the best channel with both flash and without flash, seems to be an interesting alternative in the design of a more robust recognition system.

	R1 %				DEC			
	AR0		AR1		AR0		AR1	
	R G B	GS	R G B	GS	R G B	GS	R G B	GS
SIFT	51.3	58.8	95.5	96.6	2.29	2.53	4.45	5.06
HOG-DT	41.4	45.1	77.0	80.9	1.95	1.87	2.91	2.91
HOG-F	51.5	52.2	83.2	83.7	1.90	2.22	3.07	3.06
ULBP	37.8	45.3	73.6	78.1	1.64	1.86	3.02	3.27
GIST	83.2	84.4	95.2	95.9	3.55	3.73	4.91	5.13

Table 4.6: Comparison between the **worst** color channel and GS for the 5 descriptors, with a 16-mixture GMM, for AR1 and AR0 setups.

Also, to avoid performing the proposed algorithm for all 3 color channels, choosing the one that performs the best, transforming the image into GS can be a solution to reduce computational processing time. The full set of results can be consulted on Appendix A. Another approach, based on score-level fusion, was also tested in an attempt to further improve the performance in the worst observed setups. The next section will serve to detail such approach.

RGB

As an alternative to using a single channel color or GS images, the performance related to the fusion of the R, G and B channels was also assessed, by averaging the likelihood-ratio scores of the 3 separate channels. Results, exemplified on Table 4.7, suffered a significant improvement in individual identification reaching higher values for both R1 and DEC ratios.

	R1 %				DEC			
	AR0		AR1		AR0		AR1	
	Best	RGB	Best	RGB	Best	RGB	Best	RGB
SIFT	58.8	73.2	98.1	99.7	2.64	3.01	5.06	5.32
HOG-DT	46.4	50.8	80.9	81.7	2.00	2.10	2.96	3.06
HOG-F	55.3	56.8	83.7	85.0	2.23	2.23	3.16	3.22
ULBP	47.8	48.9	78.1	78.7	2.07	1.98	3.27	3.37
GIST	86.6	89.2	97.0	98.4	3.81	3.83	5.22	5.42

Table 4.7: Comparison between the **best** color channel plus GS (left column) and RGB for the 5 descriptors and, with a 16-mixture GMM, for AR1 and AR0 setups.

These results illustrate how much fusion can improve the system performance. Both GS and RGB represent two different fusion scenarios. The first, a feature-level fusion, achieved by summing the three original images and the second, a score-level fusion, based on the recognition scores obtained for each feature (i.e. color channel) individually. GS fusion can also be understood as a method of achieving multimodality, since it results from the fusion of 3 different sources. Without highly increasing the complexity of the system, this simple multimodality approach may lead to an improve in system performance.

4.4.4 Descriptors comparison

On Figure 4.8, the most relevant results obtained for the AR1 setup, with respect to the GS color space, are presented.

	k	R1		Dec	
SIFT	64	99.0	± 0.93	4.90	± 0.11
GIST	16	95.9	± 0.78	5.13	± 0.21
HOG-F	64	87.7	± 1.35	3.17	± 0.06
HOG-DT	32	81.2	± 1.67	3.20	± 0.17
ULBP	16	78.1	± 1.82	3.27	± 0.08

Table 4.8: Results performed using the best k for each descriptor, on GS images from AR1 setup.

There is an obvious gap between GIST/SIFT and ULBP/HOG's results. Both HOG modalities (HOG-F and HOG-DT), ULBP and GIST global descriptors describe the entire image. On the other hand SIFT is a local descriptor, which computes over specific interest points, called keypoints. The enhanced SIFT performance over the other algorithms, except GIST, can be explained by its rotation invariance characteristic. CSIP's photos, taken under unconstrained scenarios, present individual periocular region with variable poses regarding the camera. As a consequence of these pose variations, the distance and relative position between the periocular region elements changes, affecting the global description and clearly impairing both HOG and ULBP's performance. However, the discriminative power of GIST seems able to surpass those limitation, at the cost of a higher computation complexity. GIST performed the best results in terms of DEC among all the descriptors, proving to have the best discriminative power among them. However, an interesting observation can be made while comparing SIFT's performance over different setups.

	R1 %	DEC		R1 %	DEC
AR0	58	2.30	AR1	100	5.32
BR0	73	3.01	BR1	93	4.40
CR0	50	2.09	CR1	93	4.00
mean	61	2.47	mean	95	4.57

Table 4.9: SIFT performance on non-flash (left) and flash (right) images.

On setups in which the flash was used, SIFT's R1 and DEC values are significantly higher than those without flash, as presented in Table 4.9. Flash enhances keypoint detection and serves as a methodology of image illumination normalization, thus influencing SIFT performance. ULBP and both HOG descriptors are also influenced by flash, a fact that can also be explained by the aforementioned image normalization process.

	R1 %	DEC		R1 %	DEC
AR0	89	3.83	AR1	97	5.42
BR0	90	3.73	BR1	94	4.26
CR0	78	2.95	CR1	88	3.78
mean	86	3.50	mean	93	4.49

Table 4.10: GIST performance on non-Flash (left) and Flash (right) images.

Moreover, as presented in Figure 4.10, GIST does not suffer such a marked performance loss as SIFT and other descriptors, for images acquired without flash. This observation further proves the potential value of GIST in biometrics under unconstrained settings.

4.4.5 Descriptors Fusion

Fusion scenarios can contribute, in some complex situations, to an overall improvement of system performance, as seen by the simple averaging of the three RGB individual results presented in Section 4.4.3. On this work, six fusion strategies at score level were evaluated, for three combinations of descriptors: maximum, minimum and median, which take into account only the higher, lower, and median, respectively, values among the descriptors likelihood-ratios; sum, in which the matrices of likelihood-ratios are summed in a final score matrix; multiplication, instead of summing, likelihood-ratios are multiplied, contributing to empathize their differences; and performance. On performance fusion, each likelihood-ratio matrix is multiplied by a weight value, calculated by taking into account the individual descriptor performance. The three combinations evaluated, presented on Table 4.11, were the fusion of all the descriptors, first column, the fusion of all descriptors except GIST (which could not be performed on k 's above 16), second column, and the the fusion between the 2 descriptors which performed significantly better than the others, SIFT and GIST, last column.

Interesting results can be observed on table regarding fusion scenarios. In fact, there is only one case in which the fusion performed a better R1 ratio than the best algorithm, among the fusion, alone: GIST and SIFT performance-base sum-rule fusion, improving in 1% R1 ratio. However, by analysing the DEC results, it seems like the fusion of certain descriptors has a more relevant impact for verification proposes. Another conclusion can be made related to the best fusion method. Top results were performed mainly on performance-base fusion scenarios and the worst DEC resulted in the three combinations by performing a multiplication of the likelihood-ratios.

Two interesting analysis can be performed regarding the fusion strategies including GIST and the ones which do not include such feature. By comparing the top performances, mostly obtained through performance-base sum-rule fusion, with the results observed in Tables 4.3 and 4.4, it is possible to observe that fusion comprising GIST descriptor does not improve the R1 performance except for the CR0 setup. However, there is a general improvement in DEC, possibly given by the fusion of the descriptive power of the 5 algorithms. Hereupon, the implemented fusion strategies seem more relevant in verification than in identification systems, as they allow a better distinction between genuine and impostor samples, rather than identifying an individual among a pool of

		GIST SIFT ULBP HOG's		SIFT ULBP HOG's		GIST SIFT	
		R1 %	DEC	R1 %	DEC	R1 %	DEC
AR0	Max	62.5	2.68	54.6	2.31	79.6	3.66
	Min	64.8	2.68	55.7	2.41	77.3	3.16
	Sum	54.6	2.24	61.4	2.76	81.8	3.77
	Mult	73.9	2.12	61.4	2.19	83.0	3.24
	Median	62.5	3.48	59.1	2.68	81.8	3.77
	Performance	76.1	3.67	62.5	2.81	80.7	3.87
BR0	Max	51.2	2.71	45.1	2.48	56.1	2.73
	Min	52.4	2.41	41.5	1.88	79.3	3.46
	Sum	72.0	3.23	52.4	2.61	78.1	3.42
	Mult	67.1	1.94	53.7	2.06	80.5	3.09
	Median	59.8	2.98	51.2	2.50	78.1	3.42
	Performance	79.3	3.57	53.7	2.61	85.4	3.76
CR0	Max	40.9	2.08	40.9	2.09	37.6	2.11
	Min	72.0	3.13	44.1	1.85	73.1	3.07
	Sum	61.3	2.76	52.7	2.36	67.7	2.87
	Mult	65.6	2.08	52.7	2.09	76.3	2.70
	Median	43.0	2.23	45.2	2.30	67.7	2.87
	Performance	67.7	3.00	51.6	2.41	78.5	3.01
DR0	Max	48.0	2.29	46.7	2.30	45.3	2.13
	Min	72.0	2.92	56.0	2.23	73.3	3.09
	Sum	72.0	3.00	57.3	2.59	72.0	2.88
	Mult	73.3	1.95	57.3	2.04	74.7	2.69
	Median	62.7	2.78	57.3	2.42	72.0	2.88
	Performance	76.0	3.22	57.3	2.61	74.7	3.16

Table 4.11: Fusion scenarios performed on rear-camera related setups, for GS images.

possibilities. Another comparison, still regarding the same Tables, takes into consideration only the combination of SIFT, ULBP and HOG's. The best observed result of the fusion scenarios is always higher, between 4% and 12% regarding R1, than the best results among the single descriptors. This observation demonstrates that fusion can positively contribute to the enhancement of the system's overall performance.

4.4.6 Full Dataset

During the previous section, performance regarding different setups was presented with respect to each descriptor. In a real-system, images captured with different devices can be all mixed-up, so, models may be trained with images from the same subject but captured with devices presenting different characteristics. In order to assess how a system, based on this algorithm will behave in such conditions, all datasets (*AR1 – DR0*) were mixed in a single larger setup, gathering all the available images. The results obtained by using the different descriptors are presented on Table 4.12.

	R1		Dec	
SIFT	38.5	± 0.61	2.15	± 0.01
ULBP	31.8	± 1.32	1.83	± 0.03
HOG-DT	22.7	± 0.82	1.36	± 0.05
HOG-F	25.8	± 0.49	1.40	± 0.04
GIST	72.1	± 0.91	3.25	± 0.05

Table 4.12: Results obtained regarding the full CSIP dataset, for GS colorspace

Images were trained using a scenario similar to the one performed on each setup. Half of each setup was used to train the models, while the other half was kept apart for testing. In order to save some time, given the high complexity related to the amount of images, only 5 tests were made being presented in Table 4.12 the average results.

There is a marked performance drop in all the descriptors, with HOG-DT presenting the worst. Although GIFT suffers from the variability in the computed data, the results are still comparable to the ones obtained with individual setups, performing better than each other descriptor individually. Having in consideration that CSIP comprises data from only 50 individuals, different approaches should be taken to guarantee a good performance in a pool representative of a larger population like in a city or even a country.

4.4.7 Computational Performance

During the previous sections, the algorithm was assessed by comparing the R1 and DEC results. Nevertheless, it is important to consider also, the computing time taken by each descriptor in order to evaluate its efficiency. Since IDSM and UBM have the same amount of points, the time taken by the projection of new data on each model is similar. On Figures 4.5 is presented the time spent, in seconds, for each descriptor, to perform a single projection regarding different k-mixtures GMM. The comparison is not totally fair because each descriptor's feature vector has different length and different dimensions what will directly affect the projection time.

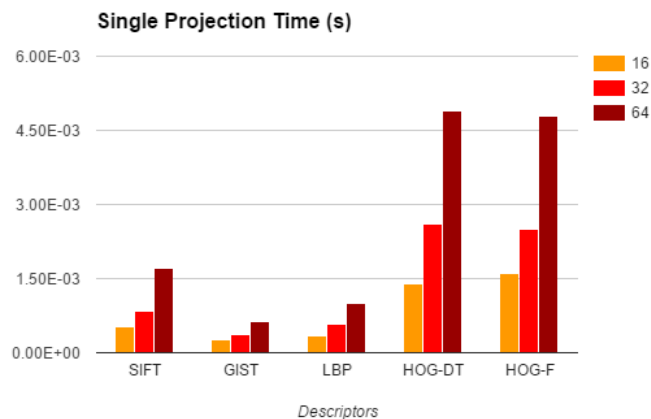


Figure 4.5: Average time, in seconds, taken to perform a single projection for each descriptor regarding AR0 setup and for GS colorspace.

A more relevant comparison can be made by calculating the average time that each descriptor takes to process a single CSIP image. Results are graphically presented on Figure 4.6, serving Table 4.13 to make a more detailed time analysis.

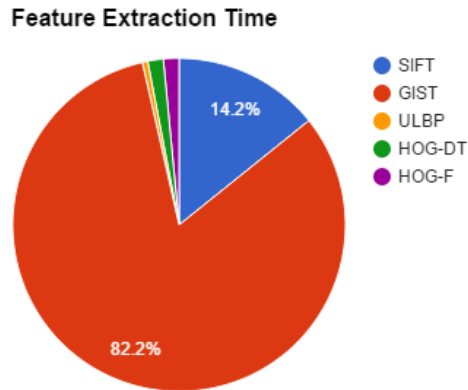


Figure 4.6: Visual comparison of the time taken by each descriptor computation regarding the AR0 for GS colorspace.

	Time (s)	
GIST	1.25E-01	$\pm 2.97E-02$
SIFT	2.16E-02	$\pm 7.90E-03$
HOG-F	2.30E-03	$\pm 1.50E-03$
HOG-DT	2.30E-03	$\pm 1.50E-03$
ULBP	7.83E-04	$\pm 3.64E-04$

Table 4.13: Time taken by each descriptor computation regarding the AR0 for GS colorspace

By looking into both the table and the figure it is possible to conclude that GIST takes longer to describe an image than the others. SIFT also presents a high description time even if significantly smaller than GIST's. These two descriptors presented the best performance results, as seen in the previous sections, however at the cost of more computational power. For ULBP and both HOG's modalities, the description takes only, approximately, twice the time took by a single projection. While for SIFT and GIST the projection time is almost irrelevant when compared to the feature extraction time.

Having both projection time (PT) and feature extraction time (ET), the algorithm's recognition time (RT) can be calculated for a single sample by using the following equation:

$$RT = 2 \times PT + ET$$

The recognition times for each descriptor, calculated based on the previous equation, are presented on Table 4.14.

	RT
GIST	2.50E-01
SIFT	4.36E-02
HOG-F	5.80E-03
HOG-DT	5.70E-03
ULBP	1.86E-03

Table 4.14: Average recognition times for each descriptor regarding a single sample, using 16-mixture GMM for AR0 setup and GS colorspace.

By the table analysis it is possible to conclude that ULBP can contribute to the fastest algorithm architecture while GIST can clearly slow the process.

4.4.8 Implementation Details

The algorithm was performed on MATLAB R2013a and on three different computers with similar characteristics. However, for time comparison all the results were obtained performing on the same computer. To train the GMM's, Netlab toolbox [95] was used. SIFT, HOG's and ULBP description was performed using VLFeat toolbox [96], while GIST was performed using code made available by the authors [68].

In this chapter several characteristics regarding a recognition system were evaluated and the related algorithm behavior was studied in order to gather information about its performance regarding images acquired in unconstrained scenarios. The algorithm proved to be suitable for individual recognition regarding the periocular region achieving the best when SIFT or GIST are used. GIST was also positioned side by side with the most conventional descriptors, presenting outstanding results during the performance assessments. It was also observed that flash can greatly influence image quality and that GS colorspace image transformation can be an interesting approach in the design of a more robust recognition system. Despite GS's performance, RGB channels fusion presented top results at the cost of high computing complexity. System performance regarding descriptors fusion was also evaluated and showed to be of lower relevance, specially when GIST is combined with other descriptors. The results obtained motivated the algorithm extrapolation to different traits, but also to a different dataset comprising periocular images acquired with worse quality. Given this, in Chapter 5, the presented methodology will be replicated to the MoBIAC dataset presented in Chapter 3.

Chapter 5

Preliminary results on MoBIAC dataset

On this chapter, an evaluation of the performance of the algorithm presented during the previous chapter over the MoBIAC dataset will be presented. The goal is to evaluate its robustness if applied in a context similar to the SSA scenario presented in Section 2.1. In each section, two points will be discussed. The first is related to the algorithm's performance when computed on the full dataset. The second is related to the performance comparison when considering Caucasian and African groups independently. This is important to gain insight as to how the algorithm and each descriptor behave regarding ethnicity variations. Furthermore, the MoBIAC images were acquired in an even less constrained environment than CSIP and also contain data regarding handprint and ear in addition to the periocular region.

5.1 Periocular Region

The MoBIAC dataset comprises periocular region images of the right side of the face. However, a lot of background noise is also contained within each photo. Given this, it is important to crop a region that excludes the background while defining the region containing relevant data regarding the individual.

Segmentation

The ROI definition step is important to highlight important information but also to normalize the segmented region. In this work, the periocular region was segmented in a rectangular region centered in the eye with $8r \times 6r$, where r stands for iris radius. The region was automatically cropped by manually signaling in each image the iris center and a point in its boundary. Given the low constrained acquisition scenarios, some eyes are closed while in others the gaze is not pointing towards the camera. In these cases two points were annotated where, theoretically, the iris would be in a normal situation. In Figure 5.1 are presented the results of the cropping method, regarding images acquired on different scenarios. After ROI segmentation, each image is sized to 12.5% of its original size. This value was chosen based on the results, as stated in the Chapter 6.1, regarding CSIP images.

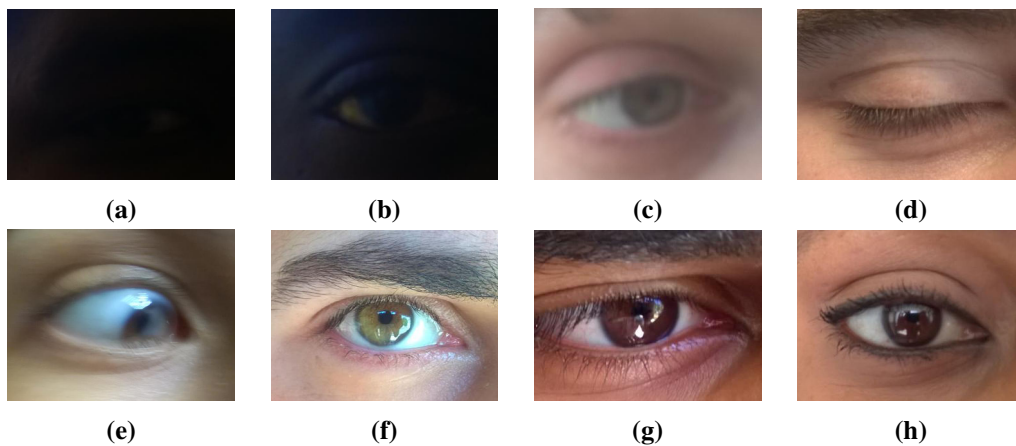


Figure 5.1: MoBIAC Periocular samples.

Full dataset Results

As previously stated, at least one image per subject presents considerable low illumination, as exemplified in images 5.1a, 5.1b, whereas others were acquired under heavily unconstrained environments, which resulted in some images lacking information, as exemplified in images 5.1c, 5.1d and 5.1e. The first presents a blurred nature while the others regard distinct challenges such as eye obstruction and gaze variations. However, these images were all used in the algorithm computation, so as to realistically replicate the greatest number of possible challenging scenarios. The details related to the algorithm computation are similar to the ones applied on CSIP. For each individuals, half of the available images were used to train the model while the other half was kept aside to test its behavior. The results presented on Table 5.1 are related to GS colorspace which, like in the previous Chapter, present a more balanced overall performance than a specific color channel.

	R1	Dec
SIFT	33.9 ±1.73	1.87 ±0.97
GIST	56.5 ±0.89	2.81 ±0.07
HOG-F	25.3 ±1.52	1.50 ±0.02
HOG-DT	16.7 ±1.17	1.13 ±0.03
ULBP	19.6 ±2.79	0.77 ±0.07

Table 5.1: Results obtained regarding periocular region related to the full MoBIAC periocular dataset.

Comparing the results obtained in Section 4.4.6, regarding all the CSIP dataset, with the ones obtained on MoBIAC, the R1's were worse as well as the DEC's. Better results could be probably achieved by filtering the set, removing the images with less quality in a process commonly called, quality assessment [97] (QA). QA of an image measures its degradation during acquisition, compression, transmission, processing, and reproduction. Give the unconstrained acquisition scenario, QA dataset entries could improve the ratio of relevant information presented in all the dataset. A

simple method could be easily implemented by, for example, threshold the image contrast, discarding samples like Images 5.1a and 5.1b.

Ethnicity comparison

Another study was made on the MoBIAC periocular dataset by assessing the algorithm's performance for both Caucasians and Africans sets, separately. Details related to the dataset distribution were similar to the ones performed to the full dataset, with half of the images serving for train the models while the other half to test it. The results presented on Tables 5.2 and 5.3 are related to GS colorspace.

	R1		Dec	
SIFT	30.4	± 1.88	1.47	± 0.02
GIST	56.1	± 0.97	2.57	± 0.06
HOG-F	24.4	± 1.46	1.06	± 0.07
HOG-DT	18.0	± 2.25	0.88	± 0.02
ULBP	20.4	± 1.42	0.59	± 0.09

Table 5.2: Performance results regarding periocular region related to the **African** individuals of the MoBIAC dataset

	R1		Dec	
SIFT	38.9	± 1.53	1.85	± 0.08
GIST	67.1	± 3.81	3.04	± 0.06
HOG-F	38.1	± 2.14	1.63	± 0.04
HOG-DT	25.7	± 2.55	1.30	± 0.03
ULBP	32.9	± 2.63	0.84	± 0.09

Table 5.3: Performance results regarding periocular region related to the **Caucasian** individuals of the MoBIAC dataset

By looking into the Tables it is possible to conclude that the algorithm performed significantly better for the Caucasian group, outperforming, also, the results related to the full dataset. However, for the African group, results were even worse than the ones achieved by the full dataset computation. This interesting behavior shows that full dataset computing was adversely affected by the African set. The weak performance can be related by the fact that African related images acquisition was more affected by the lack of light given the skin-color of the eye boundaries, resulted in less contrasted features. However, to support this explanation, a bigger dataset is needed.

5.2 Contactless Palmprint

MoBIAC dataset also contains images of the right hand. Human hand contains a large amount of information, as seen in Chapter 3, and can be used in several ways to perform individual recognition. For this work, palmprint was chosen for hand analysis, to evaluate how the different features

between African's and Caucasian's hand patterns can affect the system performance. Furthermore, palmprint segmentation is easier to be applied than, for example, hand geometry in which is necessary to highlight the whole hand region.

Segmentation

A simple, manual method, based on Aoyama et al. [7] work was applied to crop the ROI of the MoBIAC hand images. Some resulting images are presented in Figure 5.2.

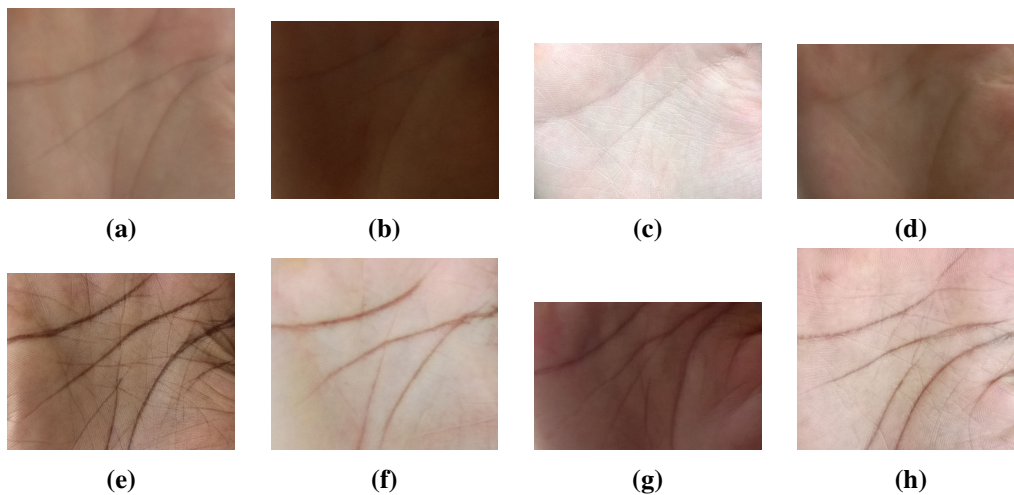


Figure 5.2: MoBIAC palmprint samples.

The segmentation was done manually by cropping a rectangular region with a corner situated on the bottom-middle of the little-finger and the opposite corner situated on the bottom middle of the thumb. Given some hand geometry and rotation variations, the method was not always precisely applied.

Full dataset Results

On Figure 5.2 it is possible to observe how light conditions can vary within the MoBIAC dataset. Light condition is the less constrained variable regarding this hand study. The results performed on MoBIAC hand dataset, regarding the different descriptors, are presented in Table 5.1. The algorithm was computed like previously, with an image reduction of 12.5% with half the dataset serving model training while other half was kept to test. Results are related to GS colorspace.

	R1		Dec	
SIFT	30.8	± 1.39	1.52	± 0.05
GIST	62.5	± 1.58	2.80	± 0.03
HOG-F	16.2	± 0.57	1.32	± 0.04
HOG-DT	21.4	± 2.12	1.33	± 0.05
ULBP	22.9	± 0.94	1.32	± 0.03

Table 5.4: Results obtained regarding contactless palmprint related to the full MoBIAC hand dataset.

Ethnicity comparison

On Figure 5.2 the top row is composed of images regarding Caucasian subjects while the bottom row is composed of images regarding African subjects. There is a noticeable difference, specially in the principal lines, in the hand lines contrast. African hands' lines are natural highlighted by the skin contrast. Results obtained from computing the algorithm on each group of individuals are presented in Tables 5.5 and 5.6.

	R1		Dec	
SIFT	37.7	± 3.69	1.52	± 0.05
GIST	81.3	± 1.46	3.45	± 0.12
HOG-F	24.1	± 2.11	1.37	± 0.07
HOG-DT	36.9	± 5.08	1.56	± 0.07
ULBP	40.5	± 2.33	1.25	± 0.03

Table 5.5: Performance results regarding contactless palmprint related to the **African** individuals of the MoBIAC dataset.

	R1		Dec	
SIFT	31.8	± 4.01	1.23	± 0.08
GIST	58.5	± 4.45	2.42	± 0.03
HOG-F	13.5	± 0.56	0.47	± 0.04
HOG-DT	17.8	± 1.85	0.54	± 0.06
ULBP	16.3	± 0.88	0.84	± 0.02

Table 5.6: Performance results regarding contactless palmprint related to the **Caucasian** individuals of the MoBIAC dataset.

Contrary to periocular, palmprint performed better for African than for Caucasian images, fact that can be explained by that difference in hand lines contrast. For some descriptors, R1 ratio and DEC doubled regarding the African images. As usual, GIST performed the best scores. However, for the African set, SIFT performed similar to the other 3 descriptors and even being outperformed by ULBP regarding R1 results.

An interesting fact can be observed by comparing GIST and SIFT likelihood-ratio matrices. On Figure 5.3 is presented a color representation of those likelihoods.

As can be seen by the existing linear division on Figure 5.3a, SIFT seems to be able to distinguish the two groups among the dataset, presenting difficulties in recognizing a certain identity among those groups separately. However, even presenting a stronger recognition capability and better performance, GIST's results shown that this descriptor makes use of other individual characteristics than ethnicity, during its computation.

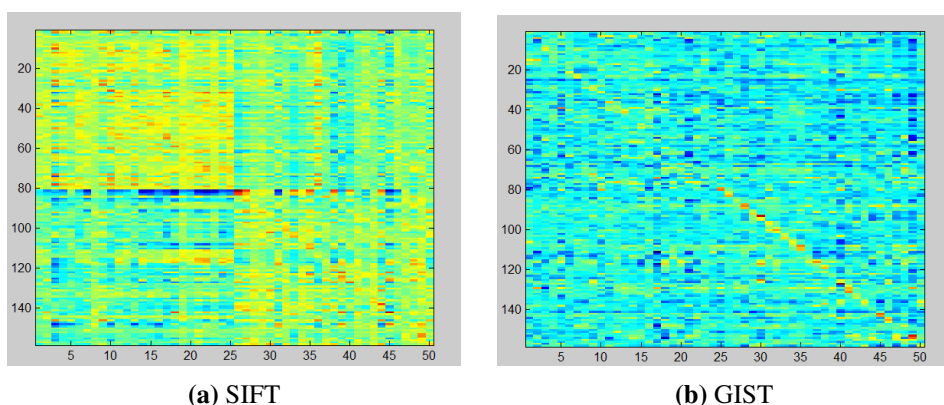


Figure 5.3: GIST and SIFT likelihood-ratio matrices comparison on GS colorspace. Y and X axis represent the project sample and the identity number, respectively.

5.3 Ear

The last trait present in MoBIAC dataset is the ear, more precisely, the right ear. Human ear's unique structure and ease of capturing its data were the motivation to this experiment. The followed methodology is similar to the one used on the previous sets. Half of the dataset samples are used to train the models while the other half is kept aside for test purposes. Also, two scenarios are evaluated. One in which the two groups, Caucasians and Africans, are computed apart and other in which all the dataset is used.

Segmentation

Before proceeding with the algorithm computation it is necessary to crop each MoBIAC ear image in order to highlight the ROI: removing noise and normalizing its data. Images were manually segmented by cropping a rectangular region that contains the ear and a small portion of its neighbourhood. Examples resulting from this method are presented on Figure 5.4.

As can be seen in, ear samples, presented on the dataset, contain different degrees of quality. In some of them, the ear is partially covered by hair while in others, earrings are presented. Furthermore, they suffer from several light variations as well as slightly different poses related to the acquisition method. These variations contribute to a better simulation of a real-life context.

Full dataset Results

After the segmentation, images were scaled to 12.5% of its original size. This value was chosen having in mind computation complexity and the previous results experienced on computing the algorithm over different size samples. Performance related to the full dataset was used to assess how the algorithm behaves and to evaluate if it is adaptable to traits other than the periorcular region. The results presented in the Table 5.7 are referent to GS colorspace.

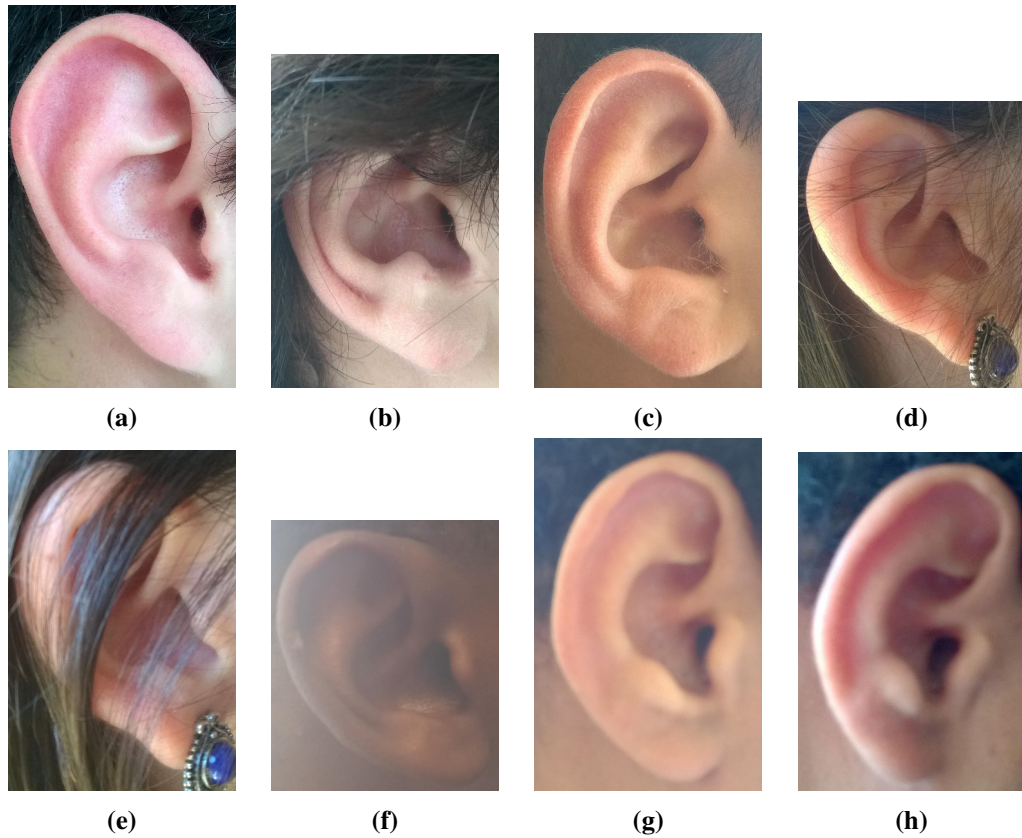


Figure 5.4: MoBIAC Ear samples.

	R1		Dec	
SIFT	38.8	± 1.68	1.84	± 0.05
GIST	85.4	± 0.00	3.96	± 0.10
HOG-F	31.2	± 0.52	1.58	± 0.02
HOG-DT	26.4	± 1.27	1.40	± 0.08
ULBP	22.3	± 0.76	1.21	± 0.03

Table 5.7: Results obtained regarding the full MoBIAC ear dataset.

The direct comparison between different traits is not the most correct way of evaluate the trait, however it can be made to have an idea of the trait unique characteristics. Ear performed the best results between the 3 traits presented in MoBIAC and once again GIST descriptor stood out from the others for its outstanding capabilities.

Ethnicity comparison

In order to evaluate how the algorithm behaves to ethnicities variation the two image groups were computed independently. The results are presented on Tables 5.8 and 5.9 . It is important to have in mind that better results are expected when comparing to the full dataset, since in this experiment, the individuals' pool comprises only 25 identities.

	R1	Dec
SIFT	38.4 \pm 1.16	1.65 \pm 0.05
GIST	82.8 \pm 3.62	3.16 \pm 0.11
HOG-F	25.8 \pm 0.52	1.14 \pm 0.02
HOG-DT	18.6 \pm 2.47	1.00 \pm 0.07
ULBP	25.4 \pm 1.27	0.94 \pm 0.02

Table 5.8: Performance results regarding the ear related to the **African** individuals of the MoBIAC dataset.

	R1	Dec
SIFT	38.4 \pm 1.16	1.65 \pm 0.05
GIST	93.2 \pm 0.98	4.44 \pm 0.11
HOG-F	42.8 \pm 1.58	1.82 \pm 0.04
HOG-DT	38.8 \pm 4.63	1.45 \pm 0.05
ULBP	29.9 \pm 0.64	1.32 \pm 0.05

Table 5.9: Performance results regarding the ear related to the **Caucasian** individuals of the MoBIAC dataset.

5.4 Conclusions

The extrapolation of the proposed methodology to the MoBIAC dataset served to evaluate not only its capability of being applied to different traits but also to assess its performance on a dataset regarding periocular images acquired under heavy unconstrained scenarios. During this chapter it was also observable how each descriptor behaves regarding different conditions like trait, lighting and quality variance. By comparing the three traits performance, it is interesting to observe that the descriptors presented similar results regarding the full sets. Still, when the sets are divided by ethnicities the algorithm behaves differently. While for the periocular region the algorithm performed better for the Caucasian group, for the palmprint, the best performance was achieved on the African group. This proves the importance of testing the system in a scenario as close to the real as possible. While for a Caucasian population, the periocular region can perform better than the palmprint, for an SSA context the hand can provide more accurate results by palmprint recognition.

Chapter 6

Conclusions and future work

This work aimed to evaluate a multimodal biometric recognition system to be applied in SSA context regarding its unconstrained image acquisition scenarios.

With the existing gap in biometric data regarding African individuals, a multimodal database construction was motivated comprising biometric data from Africans and Caucasians. The MoBIAC dataset was built regarding periocular region, ear and hand biometric traits. The dataset presents, for each trait, images with variable quality regarding light, pose and external noise variations.

In order to evaluate the behavior of a biometric system in such unconstrained scenarios, a state of the art algorithm was chosen. This algorithm was presented and its performance was assessed on the CSIP periocular dataset by testing a set of 5 different variations, each one using a different conventional descriptor. Among the descriptors, GIST proved to be the most robust, showing outstanding descriptive power. Some fusion scenarios, regarding score-level, were evaluated but none brought significant improvements to the system, except for the RGB score-level fusion. Performance using GS colorspace was also assessed, presenting results that make it a more reliable alternative rather than relying on a single color channel.

After evaluating the algorithm's performance for periocular recognition on Caucasian population and regarding different variations, its performance was also assessed on a different set of conditions from the MoBIAC dataset. It was possible to achieve, for the 3 traits available, results similar to the ones achieved only on CSIP. This fact, proved the algorithm's robustness to be applied not only in different single trait applications, but also in multimodal systems. Interesting observations were also performed regarding differences in performance between African and Caucasian individuals recognition. Hand, for example, was presented as providing a more distinguishable trait for Africans than for Caucasians, thus further proving the importance of contextualization in the design of biometric systems.

Although, extensive work was carried out around the presented algorithm for unconstrained scenarios, there is still a lot of work to develop regarding other methods to improve its robustness.

6.1 Future Work

Considering the work still to be done about this project, there are several improvements to be evaluated regarding this system:

A quality assessment method must be implemented to evaluate how the removal of images with less relevant information (for examples, closed eyes or severely low or high illumination) can affect system performance.

New fusion methods, at different levels should be evaluated as well, specially regarding fusion at the feature level, by merging feature vectors comprising information regarding different descriptors.

The work developed gave less attention to multimodality as it could be desirable. A method to fuse different traits must be implemented to improve system robustness when information regarding a single trait is not reliable enough.

Regarding the CSIP dataset, the algorithm must be tested and trained with data from different setups, since this is a very common scenario in nowadays, due to the high smartphone variability with which an individual might have to interact routinely during his/her daily life.

To increase MoBIAC's relevance in biometric recognition it would be also interesting to expand it to not only comprise information from different mobile phones, but also to have more information regarding SSA individuals. This could be achieved by creating an online platform where anyone could leave their images. It could be also interesting to study the impact of aging in this kind of systems.

Finally, the algorithm implementation in a more efficient platform, such as C++'s OpenCV computer vision library, must be completed. Its application on a real smartphone is compromised by MATLAB's multiplatform limitations.

Appendix A

Results

ARI	SIFT																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	80.9	0	3.67	0.00	80.9	0	3.79	0.00	71.3	0	3.55	0.00	87.2	0	3.93	0.00	94.7	0	4.39	0.00
4	86.4	0.67	4.03	0.03	88.3	0.87	4.12	0.05	81.1	0.45	4.04	0.03	91.5	2.75	4.15	0.11	97.9	0.5	4.80	0.02
8	93.8	1.65	4.47	0.11	92.6	1.81	4.48	0.10	92.5	1.06	4.26	0.16	93.6	1.81	4.69	0.06	99.3	0.88	5.16	0.07
16	98.1	1.1	4.79	0.16	95.5	1.31	4.74	0.12	96.1	1.51	4.45	0.07	96.6	0.84	5.06	0.07	99.7	0.51	5.32	0.09
32	95.5	1.49	4.75	0.05	97.7	0.67	4.87	0.10	95.4	0.88	4.60	0.05	98.8	0.34	5.17	0.06	100.0	0	5.23	0.06
64	95.1	0.55	4.49	0.06	98.6	0.51	4.80	0.09	95.7	0.71	4.51	0.08	99.0	0.93	4.90	0.11	100.0	0	4.95	0.06
128	96.5	1.01	4.45	0.09	99.2	0.98	4.42	0.07	96.9	1.27	4.32	0.09	98.7	0.98	4.66	0.13	99.7	0.51	4.69	0.05
ARI	HOG-DT																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	59.6	0	2.04	0.00	62.8	0	1.98	0.00	59.6	0	2.10	0.00	62.8	0	1.95	0.00	63.8	0	2.07	0.00
4	66.0	2.84	2.16	0.11	68.5	1.25	2.16	0.01	66.8	2.06	2.28	0.03	64.3	1.68	2.10	0.07	73.3	1.37	2.25	0.04
8	75.2	1.74	2.74	0.12	76.5	1.7	2.55	0.09	76.3	2.3	2.72	0.08	74.6	1.84	2.47	0.07	79.7	1.46	2.79	0.04
16	77.0	1.25	2.96	0.10	80.7	2.03	2.91	0.07	77.7	2.61	2.95	0.05	80.9	2.79	2.91	0.10	81.7	1.21	3.06	0.04
32	80.7	2.9	3.01	0.16	81.1	1.31	2.97	0.22	79.0	1.81	2.90	0.14	81.2	1.67	3.20	0.17	84.4	1.59	3.14	0.12
64	79.0	2.66	3.10	0.10	79.9	2.48	3.03	0.07	77.3	1.33	3.03	0.11	79.7	1.37	3.10	0.11	83.9	1.06	3.24	0.07
128	74.2	2.01	2.65	0.11	77.3	2.7	2.62	0.05	73.4	3.4	2.49	0.08	76.3	1.81	2.61	0.08	80.1	2.3	2.73	0.04
ARI	HOG-F																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	71.3	0	2.35	0.00	73.4	0	2.26	0.00	70.2	0	2.31	0.00	74.5	0	2.25	0.00	75.5	0	2.34	0.00
4	74.5	2.24	2.69	0.09	78.4	0.72	2.57	0.01	81.7	0.67	2.64	0.03	76.6	0	2.72	0.00	79.2	0.55	2.76	0.03
8	79.9	1.7	3.06	0.12	80.9	0	2.87	0.06	80.9	1.74	2.93	0.11	78.9	0.67	2.85	0.06	80.7	1.17	3.04	0.08
16	83.7	1.13	3.11	0.04	83.7	0.72	3.07	0.05	83.2	1.1	3.16	0.05	83.7	1.23	3.06	0.06	85.0	0.34	3.22	0.06
32	85.9	1.74	3.35	0.08	84.7	0.9	3.30	0.12	83.3	0.72	3.44	0.10	85.7	0.9	3.33	0.06	85.4	0.51	3.52	0.05
64	86.8	1.03	3.21	0.08	86.0	1.1	3.11	0.07	83.7	1.13	3.33	0.11	87.7	1.35	3.17	0.06	87.2	1.12	3.34	0.06
128	86.6	1.35	2.95	0.09	86.6	1.25	2.74	0.10	83.6	1.14	2.78	0.05	86.3	1.27	2.86	0.10	88.5	0.67	2.96	0.05
ARI	ULBP																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	55.3	0	2.27	0.00	60.6	0	2.23	0.00	60.6	0	2.40	0.00	59.6	0	2.13	0.00	62.8	0	2.40	0.00
4	63.8	0	2.51	0.00	67.7	0.9	2.43	0.02	65.7	0.45	2.65	0.02	67.5	0.74	2.42	0.04	70.6	0.9	2.69	0.02
8	73.1	1.51	2.90	0.05	71.4	2.48	2.80	0.03	72.1	1.21	3.03	0.04	76.3	0.72	2.79	0.03	76.7	1.27	3.19	0.03
16	75.6	1.62	3.06	0.06	73.6	1.79	3.02	0.08	75.9	1.33	3.10	0.12	78.1	1.82	3.27	0.08	78.7	0.71	3.37	0.05
32	73.9	2.14	2.67	0.06	75.5	1.74	2.73	0.07	78.0	1.01	2.79	0.04	78.0	1.42	3.01	0.06	79.0	1.01	2.99	0.04
64	41.0	2.76	1.86	0.03	42.8	4.67	1.96	0.04	74.3	1.86	2.09	0.07	72.7	2.3	2.43	0.04	55.4	2.38	2.20	0.03
128	9.5	2.58	0.77	0.04	10.9	1.79	0.89	0.03	38.9	1.68	1.32	0.04	46.8	2.51	1.57	0.04	16.9	1.17	1.11	0.03
ARI	GIST																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	89.7	1.01	4.09	0.02	94.7	0	4.29	0.00	95.7	0	4.30	0.00	94.7	0	4.23	0.00	94.8	0.34	4.36	0.00
4	90.9	0.55	4.44	0.01	94.0	0.9	4.71	0.02	93.5	0.34	4.81	0.03	93.8	0.67	4.71	0.06	94.0	0.9	4.86	0.04
8	93.1	1.35	4.84	0.14	94.6	0.6	5.01	0.08	96.6	1.49	5.21	0.17	94.9	0.84	5.02	0.15	94.9	1.31	5.25	0.08
16	95.2	1.03	4.91	0.21	96.2	0.55	5.07	0.25	97.0	1.1	5.22	0.18	95.9	0.78	5.13	0.21	96.7	0.6	5.42	0.17
32	97.6	1.01	5.06	0.26	98.0	0.34	5.18	0.31	98.5	0.74	4.98	0.25	98.3	0.9	5.23	0.28	98.4	0.9	5.53	0.21
64	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
128	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Table A.1: Rank-1 ratio and decidability performance regarding AR1 setup images.

AR0		SIFT																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	34.1	0	2.13	0.00	42.1	0	2.01	0.00	37.5	0	2.01	0.00	36.4	0	2.01	0.00	54.6	0	2.46	0.00	
4	43.1	1.73	2.31	0.13	43.0	2.06	2.20	0.03	43.8	4.52	2.24	0.06	45.1	1.08	2.21	0.03	63.0	2.02	2.80	0.05	
8	46.7	2.65	2.38	0.07	45.8	3.39	2.36	0.04	46.9	4.15	2.30	0.07	50.5	2.16	2.24	0.09	67.6	2.29	2.93	0.05	
16	58.5	2.69	2.54	0.08	52.5	3.78	2.64	0.08	51.3	2.7	2.29	0.05	58.8	2.94	2.53	0.06	73.2	1.95	3.01	0.08	
32	62.6	2.3	2.61	0.08	57.3	2.35	2.50	0.06	54.0	2.85	2.31	0.06	61.7	3.22	2.49	0.04	75.5	0.96	2.92	0.05	
64	63.0	1.87	2.46	0.07	59.3	4.99	2.53	0.08	55.1	3.04	2.32	0.07	60.9	3.52	2.30	0.06	76.9	2.84	2.80	0.08	
128	66.4	1.44	2.42	0.08	60.5	3.82	2.25	0.09	52.4	4.06	2.08	0.05	62.5	2.21	2.24	0.09	76.5	2.08	2.57	0.05	
AR0		HOG-DT																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	31.8	0	1.47	0.00	28.4	0	1.42	0.00	30.7	0	1.49	0.00	28.4	0	1.40	0.00	30.7	0	1.51	0.00	
4	38.4	1.17	1.63	0.07	33.0	1.31	1.60	0.03	33.6	0.59	1.64	0.03	33.2	1.84	1.55	0.03	35.9	1.87	1.68	0.02	
8	40.0	1.92	1.82	0.05	42.3	2.44	1.80	0.06	36.4	1.42	1.80	0.11	40.6	3.3	1.73	0.08	43.8	2.29	1.91	0.05	
16	45.3	2.54	1.99	0.06	46.4	2.26	2.00	0.16	41.4	1.71	1.95	0.06	45.1	2.63	1.87	0.10	50.8	2.28	2.10	0.08	
32	44.3	3.21	2.06	0.11	42.8	2.01	2.07	0.10	42.7	1.62	2.04	0.05	46.5	1.46	2.12	0.10	49.4	1.87	2.21	0.06	
64	39.0	2.08	1.99	0.02	44.0	2.63	1.96	0.08	40.1	2.4	1.84	0.06	43.4	1.68	1.98	0.05	48.2	0.96	2.08	0.04	
128	66.4	1.44	2.42	0.08	60.5	3.82	2.25	0.09	52.4	4.06	2.08	0.05	62.5	2.21	2.24	0.09	76.5	2.08	2.57	0.05	
AR0		HOG-F																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	44.3	0	1.94	0.00	39.8	0	1.82	0.00	33.0	0	1.83	0.00	43.2	0	1.84	0.00	44.3	0	1.92	0.00	
4	45.2	1.17	1.98	0.00	46.4	3.38	1.98	0.08	44.3	4.29	1.98	0.06	48.0	2.2	1.92	0.06	50.5	1.71	2.08	0.04	
8	52.3	0	2.31	0.00	50.1	3.1	2.04	0.09	44.3	2.07	1.97	0.05	50.5	1.22	2.05	0.09	54.6	2.21	2.22	0.04	
16	55.3	1.78	2.23	0.08	51.6	1.33	2.18	0.05	51.5	1.32	1.90	0.06	52.2	1.96	2.22	0.07	56.8	1.31	2.23	0.03	
32	52.5	1.68	2.15	0.02	52.1	2.26	2.29	0.15	52.6	2.01	2.12	0.14	55.1	2.74	2.37	0.17	56.8	2.34	2.32	0.09	
64	50.5	2.41	2.16	0.06	54.0	1.44	2.03	0.06	49.6	1.71	1.92	0.03	56.0	1.7	2.09	0.05	56.9	1.81	2.17	0.02	
128	49.4	2.02	1.84	0.05	49.2	1.7	1.66	0.04	44.1	1.4	1.48	0.02	50.9	2.87	1.75	0.04	55.5	1.68	1.76	0.02	
AR0		ULBP																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	30.7	0	1.31	0.00	25.0	0	1.15	0.00	30.7	0	1.24	0.00	22.7	0	1.13	0.00	28.4	0	1.28	0.00	
4	31.8	0	1.61	0.00	26.1	0	1.44	0.00	30.7	1.2	1.45	0.09	27.4	0.36	1.39	0.00	30.1	0.6	1.60	0.03	
8	40.5	0.59	1.90	0.03	33.3	0.94	1.70	0.05	37.8	1.32	1.73	0.03	33.0	0	1.63	0.03	39.0	1.2	1.90	0.02	
16	47.8	2.17	2.07	0.02	43.1	2.36	1.85	0.04	37.8	2.01	1.64	0.04	45.3	2.3	1.86	0.06	48.9	1.93	1.98	0.01	
32	48.2	2.94	1.77	0.02	43.1	2.86	1.64	0.07	38.1	1.72	1.37	0.02	44.1	1.59	1.66	0.04	45.5	0.54	1.69	0.03	
64	34.9	3.17	1.29	0.03	22.5	1.4	1.12	0.05	23.8	2.81	0.95	0.03	29.1	1.87	1.23	0.05	28.2	2.13	1.20	0.02	
128	14.7	1.96	0.70	0.06	12.8	1.52	0.56	0.05	12.2	0.94	0.38	0.02	14.8	2.07	0.72	0.03	12.6	1.56	0.59	0.01	
AR0		GIST																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	75.0	1.86	3.76	0.02	78.5	1.13	3.77	0.00	72.7	0	3.42	0.00	79.7	0.36	3.81	0.06	81.0	0.55	3.84	0.01	
4	72.7	0	3.72	0.00	78.4	0	3.82	0.00	73.9	0	3.49	0.00	78.4	0	3.82	0.00	81.8	0	3.86	0.00	
8	79.9	0.94	3.71	0.06	83.8	1.08	3.85	0.07	79.9	1.86	3.56	0.09	82.7	0.48	3.88	0.13	86.6	0.72	3.91	0.04	
16	83.2	3.16	3.55	0.16	86.6	3.16	3.81	0.06	83.5	1.44	3.60	0.07	84.4	3.03	3.73	0.09	89.2	2.02	3.83	0.07	
32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
64	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
128	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	

Table A.2: Rank-1 ratio and decidability performance regarding AR0 setup images.

BR1		SIFT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	64.7	0	3.36	0.00	70.6	0	3.61	0.00	64.1	1.86	3.27	0.12	75.3	0	3.64	0.00	82.4	1.24	4.54	0.01
4	77.2	1.49	3.52	0.03	80.8	2.15	3.84	0.03	74.4	0.5	3.64	0.06	81.4	2.76	3.86	0.07	88.4	1.17	4.65	0.04
8	80.1	2.18	3.70	0.07	85.9	1.57	4.04	0.10	85.3	1.39	3.92	0.03	88.5	1.91	4.14	0.20	91.4	0.79	4.75	0.05
16	81.3	1.29	3.64	0.08	90.7	2.51	4.04	0.12	90.5	1.7	3.86	0.05	89.4	2.08	3.86	0.13	93.1	1.03	4.40	0.03
32	84.6	2.38	3.44	0.05	89.3	1.29	3.76	0.07	91.7	1.29	3.67	0.07	90.2	1.76	3.76	0.08	94.1	0.78	4.02	0.04
64	83.3	2.34	3.05	0.06	90.0	1.69	3.50	0.08	90.4	1.65	3.46	0.06	90.8	0.93	3.60	0.04	93.4	1.14	3.63	0.03
128	83.4	0.87	2.90	0.07	90.1	0.99	3.39	0.07	88.6	1.57	3.25	0.08	89.8	1.67	3.52	0.05	92.6	1.93	3.43	0.05
BR1		HOG-DT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	49.4	0	1.91	0.00	54.1	0	1.64	0.00	51.8	0	1.75	0.00	54.1	0	1.77	0.00	56.5	0	1.82	0.00
4	56.5	0	2.03	0.00	54.1	0	1.81	0.00	52.9	0	1.88	0.00	52.9	0	1.92	0.00	58.8	0	1.97	0.00
8	65.5	4.04	2.28	0.05	64.6	0.87	2.20	0.04	62.8	2.61	2.26	0.07	66.9	1.51	2.33	0.03	69.4	1.11	2.36	0.03
16	65.4	2.61	2.32	0.08	68.9	1.77	2.26	0.18	66.9	2.03	2.39	0.04	67.3	2.13	2.40	0.11	73.2	1.45	2.45	0.09
32	63.7	2.25	2.35	0.14	71.1	2.37	2.42	0.07	66.9	1.79	2.12	0.07	72.6	4.26	2.51	0.11	73.3	2.42	2.49	0.07
64	64.2	1.68	2.37	0.04	72.0	2.87	2.42	0.09	69.7	1.45	2.33	0.09	77.1	2.9	2.54	0.06	77.9	1.45	2.56	0.06
128	60.9	2.4	1.92	0.06	62.8	2.95	1.97	0.08	56.7	2.47	1.80	0.08	63.7	3.3	2.12	0.10	69.1	2.54	2.05	0.04
BR1		HOG-F																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	60.0	0	2.23	0.00	61.2	0	2.09	0.00	58.8	0	2.18	0.00	62.4	0	2.25	0.00	62.4	0	2.26	0.00
4	65.9	0	2.35	0.00	70.6	0	2.26	0.00	71.8	0	2.51	0.00	69.4	2.48	2.42	0.04	74.1	0	2.51	0.00
8	65.3	2.24	2.43	0.07	73.1	1.7	2.50	0.04	72.0	0.74	2.55	0.06	72.5	1.14	2.57	0.04	76.8	0.97	2.65	0.05
16	72.9	1.84	2.61	0.09	78.4	1.68	2.53	0.22	76.1	1.25	2.67	0.04	78.4	1.59	2.74	0.11	81.4	1.45	2.75	0.11
32	80.7	2.55	3.05	0.18	81.2	1.24	3.00	0.08	73.8	2.15	2.67	0.07	80.9	1.82	3.03	0.14	83.9	1.12	3.11	0.08
64	78.9	3.66	2.89	0.06	81.2	2	2.81	0.06	75.2	3.57	2.57	0.04	81.8	1.86	2.83	0.05	85.8	1.17	2.93	0.03
128	71.4	3.68	2.20	0.06	76.6	1.79	2.35	0.06	69.7	2.59	2.11	0.03	78.2	2.02	2.37	0.04	81.1	1.61	2.37	0.02
BR1		ULBP																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	44.7	0	1.95	0.00	48.2	0	1.57	0.30	49.4	0	1.72	0.00	47.1	0	1.83	0.00	46.8	1.82	1.88	0.14
4	52.9	0	2.09	0.00	56.5	3.04	2.07	0.00	51.8	0	2.01	0.00	56.6	1.7	2.11	0.04	59.8	1.22	2.22	0.01
8	57.9	1.08	2.07	0.07	60.7	2.16	2.17	0.03	55.3	2.29	2.18	0.08	64.1	2.62	2.25	0.05	64.8	1.96	2.34	0.05
16	60.2	2.06	2.12	0.11	59.5	2.09	1.89	0.04	58.7	2.03	2.09	0.05	66.4	1.38	2.07	0.05	63.8	1.98	2.26	0.04
32	52.0	2.53	2.03	0.08	50.0	4.97	1.38	0.11	51.5	2.4	1.53	0.03	59.3	3.65	1.56	0.06	59.9	3.01	1.78	0.06
64	34.1	2.35	1.30	0.04	29.7	3.36	0.82	0.03	31.2	4.34	0.92	0.07	34.4	3.54	0.95	0.04	35.3	1.84	1.08	0.04
128	22.9	1.78	0.64	0.07	17.2	3.69	0.34	0.04	20.4	3.64	0.41	0.03	21.1	2.91	0.44	0.06	24.1	2.17	0.50	0.03
BR1		GIST																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	82.4	0	3.66	0.00	89.4	0	4.20	0.07	87.1	0	4.18	0.00	88.2	0	4.03	0.00	89.4	0	4.23	0.03
4	89.3	0.37	3.98	0.00	89.3	0.37	4.34	0.04	90.6	0	4.12	0.01	89.3	0.37	4.25	0.05	91.7	0.37	4.29	0.02
8	86.8	0.93	4.03	0.03	89.4	0.55	4.48	0.11	90.8	1.55	4.09	0.12	88.2	1.11	4.15	0.15	91.5	0.74	4.37	0.05
16	89.1	1.36	3.90	0.09	91.2	1.49	4.14	0.09	91.9	1.17	4.08	0.11	91.5	1.08	3.98	0.09	93.8	1.08	4.26	0.07
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.3: Rank-1 ratio and decidability performance regarding BR1 setup images.

BR0		SIFT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	28.1	0	1.43	0.00	25.6	0	1.39	0.00	17.1	0	1.15	0.00	14.6	0	1.34	0.00	34.2	0	1.65	0.00
4	28.8	1.54	1.58	0.03	27.0	0.39	1.61	0.00	15.1	1.03	1.19	0.00	23.7	2.31	1.50	0.00	45.0	1.86	1.84	0.02
8	33.7	2.94	1.66	0.03	32.9	2.94	1.70	0.07	28.7	4.31	1.50	0.05	33.7	2.58	1.62	0.09	48.7	2.19	2.06	0.05
16	38.7	3.3	1.77	0.07	35.1	1.61	1.78	0.08	28.1	2.23	1.54	0.07	37.3	2.89	1.77	0.04	50.1	1.46	2.09	0.05
32	45.2	1.86	1.86	0.05	39.4	2.3	1.73	0.09	28.3	2.06	1.47	0.07	40.5	3.44	1.76	0.06	49.5	1.18	2.11	0.06
64	46.6	3.58	1.82	0.10	41.3	2.72	1.77	0.07	29.0	3.76	1.33	0.08	43.8	3.22	1.83	0.07	51.1	2.96	2.02	0.06
128	46.7	2.76	1.87	0.09	38.2	2.82	1.66	0.11	32.0	3.44	1.47	0.08	43.5	2.7	1.83	0.05	54.3	3.51	2.04	0.08
BR0		HOG-DT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	34.2	0	1.58	0.00	29.3	0	1.45	0.00	28.1	0	1.28	0.00	30.5	0	1.49	0.00	40.2	0	1.49	0.00
4	41.5	0	1.85	0.00	36.7	1.77	1.68	0.02	30.6	0.39	1.49	0.01	39.8	1.03	1.73	0.01	40.1	1.07	1.76	0.01
8	43.7	0.77	1.81	0.04	38.5	2.17	1.84	0.04	31.3	2.93	1.65	0.02	39.3	1.5	1.89	0.04	44.6	1.54	1.90	0.02
16	42.7	2.15	2.05	0.04	43.1	2.51	2.04	0.07	38.3	1.65	1.80	0.06	42.7	1.63	1.94	0.04	47.1	1.92	2.14	0.05
32	42.3	2.88	1.98	0.09	41.7	2.92	1.93	0.07	36.0	4.54	1.77	0.08	39.4	3.2	1.80	0.05	47.3	1.8	2.08	0.05
64	38.7	3.15	1.67	0.09	41.0	3.11	1.62	0.05	33.5	2.77	1.47	0.06	40.4	2.96	1.57	0.06	46.8	1.65	1.75	0.04
128	33.8	2.88	1.24	0.05	32.7	2.92	1.23	0.06	28.8	3.36	1.09	0.05	32.7	1.5	1.08	0.07	37.8	2.44	1.33	0.03
BR0		HOG-F																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	36.6	0	1.53	0.00	28.1	0	1.48	0.00	24.4	0	1.37	0.00	29.3	0	1.50	0.00	28.1	0	1.51	0.00
4	40.2	0	1.64	0.00	32.9	0	1.53	0.00	25.5	0.39	1.38	0.01	30.5	0	1.55	0.01	36.5	0.39	1.57	0.00
8	42.6	0.39	1.72	0.00	34.6	1.18	1.62	0.02	31.0	1.54	1.49	0.04	38.8	0.77	1.62	0.01	40.2	1	1.67	0.01
16	43.2	2.77	1.76	0.05	34.0	2.9	1.71	0.18	33.3	1.73	1.64	0.14	37.7	1.57	1.86	0.17	41.1	1.91	1.77	0.09
32	44.8	2.44	2.08	0.08	38.7	3.1	1.93	0.05	32.7	2.14	1.76	0.03	39.5	2.01	2.02	0.03	41.1	1.41	2.01	0.03
64	41.7	3.19	1.95	0.03	42.0	2.58	1.86	0.03	36.3	2.14	1.71	0.04	42.9	3.24	1.94	0.04	44.6	2.17	1.93	0.03
128	41.2	4.37	1.95	0.04	42.1	2.52	1.83	0.02	42.7	2.07	1.68	0.03	40.9	2.46	1.88	0.03	48.9	1.34	1.90	0.02
BR0		ULBP																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	26.8	0	1.64	0.00	22.0	0	1.26	0.00	17.1	0	1.22		20.7	0	1.34		26.8	0	1.44	0.00
4	30.5	0	1.74	0.00	24.6	3.09	1.40	0.06	19.0	2.24	1.34	0.04	27.4	1.32	1.48	0.01	30.1	0.59	1.60	0.03
8	29.2	0.9	1.82	0.04	29.3	2.51	1.52	0.05	21.2	1.18	1.32	0.02	28.2	0.39	1.57	0.00	34.3	1.34	1.68	0.03
16	40.7	2.83	1.95	0.06	34.3	2.78	1.62	0.05	25.4	2.49	1.38	0.07	37.0	2.57	1.64	0.03	38.3	2.45	1.81	0.04
32	40.9	2.39	1.69	0.10	36.0	1.65	1.42	0.04	22.9	2.36	1.09	0.04	41.5	2.23	1.48	0.04	41.0	3.21	1.55	0.04
64	24.2	2.21	1.08	0.06	21.5	2.01	0.86	0.04	10.9	2.66	0.50	0.06	27.3	2.65	0.99	0.06	21.5	2.01	0.91	0.04
128	12.8	3.46	0.46	0.10	12.2	1.82	0.35	0.08	4.8	2.54	0.06	0.05	16.2	4.15	0.44	0.08	12.0	2.62	0.32	0.05
BR0		GIST																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	78.2	1.77	3.34	0.00	73.7	3.05	3.49	0.02	67.8	1.18	3.40	0.02	76.6	1.26	3.52	0.00	81.3	0.82	3.62	0.01
4	76.8	0	3.33	0.00	76.8	0.81	3.46	0.00	69.4	0.39	3.31	0.00	77.9	0.39	3.45	0.00	78.1	0	3.51	0.00
8	83.5	2.32	3.51	0.06	87.1	1.92	3.79	0.03	82.1	1.41	3.56	0.05	86.5	1.95	3.75	0.05	89.3	1.26	3.79	0.03
16	85.5	1.86	3.59	0.11	87.6	2.06	3.59	0.04	83.9	1.5	3.41	0.13	85.6	1.89	3.54	0.13	90.0	2.21	3.73	0.08
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.4: Rank-1 ratio and decidability performance regarding BR0 setup images.

BF0		SIFT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	30.5	0	1.78	0.00	32.9	0	1.88	0.00	29.3	0	1.81	0.00	25.6	0	1.96	0.00	42.7	0	2.06	0.00
4	40.1	0.39	2.05	0.00	37.8	0	2.06	0.00	31.1	0.64	1.88	0.04	31.7	0	2.14	0.00	50.7	0.63	2.32	0.02
8	36.1	2.65	2.10	0.04	35.6	3.14	2.09	0.01	33.5	3.31	1.84	0.01	34.3	1.21	2.11	0.05	53.5	2.96	2.30	0.04
16	45.6	3.21	2.27	0.12	41.2	2.14	2.16	0.07	37.1	2.77	1.96	0.04	39.5	2.31	2.19	0.05	60.1	3.4	2.42	0.07
32	49.2	2.64	2.15	0.05	51.6	4.15	2.30	0.05	39.6	1.93	1.88	0.04	49.5	1.92	2.20	0.04	63.1	1.91	2.39	0.05
64	48.5	2.14	1.93	0.06	50.2	4.02	2.02	0.07	39.2	3.07	1.82	0.08	50.4	2.08	2.06	0.09	55.6	4.07	2.17	0.05
128	45.2	3.37	1.72	0.05	46.2	2.78	1.84	0.07	36.2	2.37	1.54	0.06	40.5	2.43	1.79	0.09	50.2	2.98	1.95	0.04
BF0		HOG-DT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	18.3	0	1.30	0.00	19.5	0	1.25	0.00	14.6	0	1.18	0.00	17.1	0	1.26	0.00	17.1	0	1.27	0.00
4	26.8	2.15	1.34	0.07	30.9	1.41	1.36	0.02	20.0	3.78	1.31	0.03	28.7	3.6	1.36	0.02	30.4	1.34	1.38	0.02
8	30.5	3.81	1.42	0.13	29.4	1.86	1.40	0.06	25.4	2.14	1.38	0.12	29.6	2.44	1.41	0.08	31.2	1.54	1.45	0.06
16	33.9	1.89	1.61	0.11	30.7	1.89	1.55	0.10	30.4	2.78	1.52	0.08	34.0	1.07	1.63	0.10	33.8	1.82	1.63	0.04
32	38.1	2.75	1.77	0.09	32.0	2.14	1.70	0.06	31.1	2.01	1.59	0.06	34.8	2.17	1.72	0.06	35.6	0.96	1.75	0.04
64	34.8	1.75	1.85	0.04	32.3	2.17	1.74	0.03	30.2	2.49	1.67	0.06	33.9	2.06	1.84	0.04	36.8	1.38	1.84	0.03
128	31.0	1.18	1.80	0.06	32.9	2.44	1.63	0.05	28.1	2.3	1.57	0.04	32.2	2.31	1.73	0.05	34.8	1.65	1.74	0.04
BF0		HOG-F																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	36.6	0	1.53	0.00	28.1	0	1.48	0.00	24.4	0	1.37	0.00	29.3	0	1.50	0.00	28.1	0	1.51	0.00
4	40.0	0.77	1.65	0.03	33.3	1.16	1.53	0.00	26.5	2.7	1.39	0.03	30.5	0	1.56	0.00	37.2	1.55	1.58	0.01
8	42.5	0.39	1.72	0.00	34.8	1.19	1.61	0.03	31.7	0	1.51	0.00	38.9	0.9	1.62	0.02	40.5	0.77	1.67	0.01
16	40.5	4.37	1.78	0.08	36.2	3.04	1.88	0.12	34.6	3.31	1.62	0.15	38.2	2.51	1.90	0.17	39.9	1.53	1.83	0.05
32	43.9	1.82	2.09	0.05	37.2	1.75	1.94	0.03	33.2	2.21	1.77	0.03	40.0	1.8	2.02	0.04	39.9	1.29	2.02	0.02
64	41.8	2.99	1.96	0.04	41.7	1.5	1.85	0.03	36.1	2.09	1.71	0.04	42.4	2.75	1.95	0.04	43.9	2.15	1.93	0.02
128	40.4	2.11	1.96	0.03	42.6	2.84	1.84	0.03	41.7	2.21	1.67	0.03	41.8	2.99	1.88	0.05	47.7	1.46	1.91	0.01
BF0		ULBP																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	20.7	0	1.14	0.00	18.3	0	1.04	0.00	19.5	0	1.04	0.00	17.1	0	1.06	0.00	22.0	0	1.08	0.00
4	31.7	0	1.28	0.00	23.2	0	1.20	0.00	29.3	0	1.25	0.00	26.8	0	1.21	0.00	31.7	0	1.27	0.00
8	28.4	4.74	1.28	0.07	29.8	2.38	1.30	0.02	33.7	1.18	1.28	0.01	30.2	2.06	1.36	0.00	35.6	1.89	1.33	0.02
16	41.8	1.29	1.70	0.06	33.1	1.07	1.55	0.03	33.4	2.17	1.50	0.03	36.6	1.29	1.55	0.04	38.9	1.77	1.64	0.02
32	38.1	2.29	1.59	0.03	34.6	1.74	1.48	0.02	31.8	1.67	1.41	0.03	37.4	1.16	1.47	0.03	37.7	1.07	1.55	0.01
64	30.7	2.06	1.47	0.05	32.8	1.46	1.26	0.02	27.7	2.08	1.12	0.04	34.4	1.61	1.28	0.02	32.2	1.03	1.34	0.03
128	12.8	3.51	0.81	0.03	13.1	1.41	0.60	0.04	8.7	3.01	0.50	0.05	11.8	2.51	0.65	0.03	12.6	1.41	0.67	0.03
BF0		GIST																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	54.9	0	2.90	0.00	64.6	0	3.11	0.00	53.7	0	2.71	0.00	65.9	0	3.13	0.00	58.5	0	3.13	0.00
4	57.9	1.19	2.96	0.05	64.8	3.12	2.89	0.16	49.3	1.18	2.49	0.05	64.4	1.97	2.91	0.06	60.7	0.77	3.03	0.09
8	56.2	1.46	3.01	0.04	63.3	1.21	3.35	0.04	60.8	2.14	2.95	0.08	65.2	1.04	3.27	0.03	67.2	2.03	3.37	0.04
16	63.5	3.12	2.81	0.03	72.3	2.88	3.21	0.03	67.8	4.31	2.96	0.04	72.8	3.68	3.15	0.04	77.6	1.92	3.24	0.02
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.5: Rank-1 ratio and decidability performance regarding BF0 setup images.

CR0		SIFT																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	24.7	0	1.69	0.00	18.3	0	1.39	0.00	19.4	0	1.47	0.00	25.8	0	1.58	0.00	34.4	0	1.87	0.00	
4	25.8	2.87	1.75	0.01	28.9	2.24	1.61	0.03	24.1	4.31	1.66	0.05	32.5	1.42	1.72	0.02	41.1	1.81	2.08	0.03	
8	35.5	3.12	1.84	0.04	32.5	3.28	1.80	0.04	25.1	2.33	1.80	0.05	35.6	3.3	1.84	0.06	47.3	2.87	2.25	0.03	
16	37.9	3.58	1.96	0.08	41.1	3.58	1.88	0.06	35.6	3.02	1.86	0.03	39.6	2.62	1.98	0.04	58.4	2.33	2.30	0.07	
32	49.9	2.22	2.19	0.08	42.0	3.45	1.95	0.04	39.8	3.96	1.75	0.07	45.6	3.8	2.02	0.05	61.2	1.47	2.37	0.05	
64	52.6	3.49	2.12	0.06	43.6	2.79	1.84	0.05	42.0	2.75	1.73	0.05	46.9	2.49	1.94	0.07	63.8	2.96	2.28	0.06	
128	52.7	2.32	2.04	0.07	45.6	3.56	1.83	0.06	40.2	1.98	1.63	0.07	49.9	2.74	1.89	0.07	64.7	2.37	2.18	0.06	
CR0		HOG-DT																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	21.5	0	1.41	0.00	19.4	0	1.35	0.00	21.5	0	1.27	0.00	16.1	0	1.36	0.00	20.4	0	1.40	0.00	
4	24.8	1.38	1.59	0.04	19.4	1.6	1.49	0.02	20.7	1.42	1.40	0.00	20.5	2.4	1.50	0.03	21.8	1.35	1.57	0.01	
8	30.3	2.42	1.71	0.10	21.3	2.2	1.59	0.02	24.4	1.35	1.46	0.06	25.3	2.65	1.63	0.06	28.5	1.36	1.69	0.04	
16	32.4	2.7	1.78	0.06	24.7	2.09	1.63	0.04	25.4	1.04	1.43	0.04	28.4	2.74	1.63	0.03	30.8	1.91	1.73	0.03	
32	33.0	1.44	1.82	0.08	27.3	3.09	1.80	0.09	26.0	3.39	1.47	0.07	29.5	1.84	1.80	0.05	32.0	1.32	1.85	0.05	
64	35.8	1.76	1.72	0.06	31.9	2.73	1.69	0.03	25.5	2.38	1.38	0.05	31.2	1.76	1.64	0.06	37.2	1.84	1.75	0.04	
128	30.0	2.7	1.33	0.05	30.8	2.74	1.30	0.07	24.3	2.54	0.97	0.05	27.5	3.66	1.35	0.04	32.9	1.91	1.32	0.03	
CR0		HOG-F																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	28.0	0	1.62	0.00	27.7	2.14	1.65	0.01	19.4	0	1.53	0.00	28.3	1.02	1.56	0.00	30.4	0.73	1.69	0.01	
4	34.4	4.39	1.90	0.08	36.5	1.93	1.88	0.03	25.5	1.61	1.66	0.09	33.9	1.27	1.85	0.05	36.1	1.91	1.93	0.04	
8	35.8	1.9	1.95	0.06	35.0	1.04	1.90	0.03	28.7	2.21	1.67	0.03	35.3	1.5	1.77	0.03	38.5	1.95	1.97	0.03	
16	41.3	2.1	2.06	0.05	40.7	2.47	1.96	0.07	33.0	3.33	1.71	0.05	42.2	2.47	1.90	0.07	44.2	2.47	2.04	0.04	
32	44.6	2.34	2.03	0.10	43.1	2.29	1.90	0.06	34.8	3.41	1.65	0.12	44.8	2.09	1.84	0.08	46.0	1.81	2.00	0.06	
64	45.1	3.14	1.90	0.05	42.9	1.79	1.74	0.05	36.5	1.93	1.57	0.05	45.1	2.06	1.75	0.04	49.4	1.38	1.89	0.02	
128	40.8	2.61	1.41	0.03	38.5	2.52	1.36	0.05	33.1	2.31	1.21	0.03	40.0	2.08	1.47	0.04	44.2	2.22	1.44	0.02	
CR0		ULBP																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	24.7	0	1.01	0.00	21.4	0.34	0.99	0.00	20.4	0	0.94	0.00	18.3	0	0.92	0.00	21.5	0	1.04	0.00	
4	29.7	0.56	1.27	0.11	18.9	0.91	1.03	0.06	17.5	0.52	1.08	0.01	20.4	0	1.34	0.00	24.2	0.76	1.20	0.06	
8	31.6	0.75	1.52	0.19	24.7	0	1.33	0.03	20.7	0.45	1.20	0.06	25.2	1.54	1.28	0.22	26.9	1.13	1.45	0.06	
16	36.5	2.24	1.46	0.08	30.2	2.98	1.38	0.08	25.4	1.91	1.30	0.06	29.9	2.37	1.49	0.07	31.9	1.35	1.48	0.05	
32	37.2	1.26	1.31	0.03	27.1	1.81	1.30	0.03	24.5	1.81	1.12	0.03	30.0	2.06	1.26	0.03	30.7	1.62	1.35	0.03	
64	22.3	7.14	0.80	0.29	22.2	8.22	0.86	0.33	14.3	5.81	0.69	0.29	19.7	8.74	0.80	0.36	20.7	7.38	0.87	0.33	
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CR0		GIST																			
10	R				G				B				Gray				RGB				
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		
2	66.7	2.27	2.70	0.02	67.9	1.56	2.76	0.00	66.8	0	2.85	0.00	68.8	0	2.74	0.00	70.1	0.45	2.87	0.01	
4	63.4	0	2.80	0.00	65.6	0	2.85	0.00	68.8	0	2.97	0.00	68.6	0.45	2.82	0.00	68.8	0	2.96	0.00	
8	74.7	1.04	2.91	0.07	74.4	0.85	3.11	0.11	77.2	1.5	3.14	0.07	75.8	2.39	3.06	0.13	77.1	1.44	3.15	0.06	
16	71.9	1.93	2.77	0.05	71.8	2.01	2.81	0.10	76.3	1.68	2.95	0.08	73.2	1.64	2.87	0.12	77.9	1.26	2.95	0.07	
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Table A.6: Rank-1 ratio and decidability performance regarding CR0 setup images.

CR1	SIFT																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	61.3	0	3.01	0.00	59.1	0	2.78	0.00	43.0	0	2.36	0.00	54.8	0	2.62	0.00	77.4	0	3.41	0.00
4	69.0	3.58	3.23	0.06	75.9	5.44	3.25	0.17	51.5	2.18	2.62	0.01	73.7	1.7	3.04	0.08	84.1	2.08	3.75	0.07
8	79.5	2.4	3.58	0.05	78.7	2.2	3.33	0.12	63.0	3.52	2.86	0.06	74.5	1.9	3.25	0.06	88.8	1.84	3.82	0.07
16	87.4	1.61	4.01	0.07	84.4	1.91	3.51	0.06	71.3	0.89	3.05	0.05	82.6	1.32	3.54	0.05	92.9	1.26	4.00	0.06
32	88.4	1.32	3.93	0.11	88.5	1.44	3.57	0.07	72.3	2.52	3.15	0.08	86.7	1.62	3.61	0.10	93.9	1.25	3.94	0.04
64	88.3	1.99	3.53	0.08	88.1	1.99	3.34	0.10	75.9	2.92	2.98	0.07	85.9	1.79	3.33	0.07	93.0	1.98	3.59	0.06
128	87.9	2.03	3.30	0.10	86.7	2.16	3.08	0.04	77.2	1.59	2.86	0.05	86.3	3.13	3.09	0.06	93.6	1.43	3.33	0.04
CR1	HOG-DT																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	45.2	0	1.81	0.00	43.0	0	1.72	0.00	34.4	0	1.74	0.00	43.0	0	1.71	0.00	48.4	0	1.82	0.00
4	50.5	0	1.97	0.00	41.9	0	1.94	0.00	45.6	2.79	2.05	0.06	51.1	1.77	1.91	0.04	55.7	1.74	2.07	0.02
8	59.3	1.29	2.36	0.02	57.2	0.68	2.27	0.03	52.4	1.44	2.36	0.13	56.1	1.42	2.21	0.05	62.3	1.64	2.46	0.04
16	57.1	2.56	2.36	0.04	63.3	3.34	2.40	0.04	58.1	1.9	2.44	0.06	58.5	2.44	2.35	0.09	65.5	1.56	2.56	0.03
32	60.4	1.88	2.53	0.08	64.0	2.6	2.51	0.05	61.0	2.49	2.49	0.04	61.9	2.44	2.48	0.08	68.4	1.91	2.71	0.05
64	60.9	2.39	2.35	0.05	64.1	2.16	2.33	0.05	58.5	1.7	2.27	0.04	61.4	2.24	2.32	0.07	69.1	1.35	2.49	0.03
128	58.5	3.59	1.96	0.04	58.8	2.59	1.92	0.05	57.9	3.35	1.80	0.06	56.8	2.37	1.94	0.03	68.4	2.33	2.05	0.03
CR1	HOG-F																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	48.4	0	2.03	0.00	48.4	0	1.99	0.00	45.2	0	2.05	0.00	46.2	0	1.93	0.00	55.9	0	2.10	0.00
4	59.6	2.39	2.63	0.07	59.7	1.7	2.41	0.01	57.1	0.34	2.50	0.00	55.0	0.34	2.40	0.08	66.1	0.91	2.56	0.03
8	65.6	0	2.53	0.00	66.2	0.91	2.59	0.04	62.2	0.68	2.60	0.06	64.7	0.68	2.50	0.02	68.6	0.45	2.71	0.03
16	67.1	2.28	2.54	0.03	64.3	1.67	2.62	0.02	65.1	1.16	2.77	0.05	65.0	0.91	2.44	0.04	71.1	1.47	2.79	0.02
32	64.6	1.38	2.46	0.07	67.2	2.04	2.55	0.04	64.6	1.72	2.53	0.03	66.2	2.39	2.46	0.04	72.6	1.27	2.65	0.04
64	69.0	2.52	2.26	0.04	67.6	1.72	2.24	0.04	63.9	1.7	2.19	0.07	68.4	1.91	2.23	0.03	73.1	1.43	2.34	0.03
128	67.5	3.39	2.02	0.05	64.3	2.86	1.90	0.04	58.2	2.45	1.69	0.03	69.4	2.28	2.02	0.05	71.2	1.95	1.96	0.03
CR1	ULBP																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	37.6	0	1.65	0.00	34.4	0	1.48	0.00	37.6	0	1.52	0.00	35.5	0	1.45	0.00	41.9	0	1.60	0.00
4	38.7	2.78	1.83	0.02	37.6	0	1.91	0.10	40.9	0	2.03	0.00	37.6	0	1.63	0.02	42.2	0.45	2.03	0.06
8	46.3	0.34	2.03	0.01	47.0	3.69	1.94	0.04	41.8	1.29	1.85	0.03	44.2	1.29	1.86	0.04	52.4	1.44	2.05	0.02
16	56.0	1.56	2.10	0.03	51.2	1.98	2.04	0.02	49.7	2.14	1.91	0.05	49.5	1.24	1.99	0.05	54.7	1.38	2.18	0.03
32	55.8	3.91	1.86	0.04	51.2	2.44	1.83	0.06	48.5	2.66	1.65	0.03	52.6	2.51	1.80	0.06	55.8	1.38	1.94	0.03
64	42.8	3.32	1.34	0.03	36.3	2.95	1.29	0.05	32.2	2.24	1.12	0.04	43.3	1.97	1.32	0.03	39.3	2.28	1.37	0.02
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CR1	GIST																			
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	71.5	0.57	3.17	0.00	75.3	0	3.42	0.00	77.9	0.56	3.52	0.00	71.0	0	3.40	0.00	75.3	0	3.44	0.00
4	71.0	0	3.32	0.01	74.6	1.26	3.41	0.07	81.8	1.29	3.58	0.01	74.2	1.43	3.48	0.06	77.4	0	3.54	0.02
8	83.4	0.91	3.74	0.06	82.5	1.67	3.91	0.07	82.6	1.67	3.87	0.04	83.2	3.17	3.86	0.04	85.7	0.73	4.00	0.03
16	86.1	1.72	3.60	0.07	85.5	1.27	3.73	0.15	86.6	1.36	3.56	0.13	86.3	1.35	3.76	0.12	87.5	0.75	3.78	0.06
32	87.1	2.53	3.20	0.11	87.5	1.36	3.43	0.14	88.4	1.22	3.38	0.08	88.1	1.64	3.41	0.10	89.1	1.56	3.48	0.09
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.7: Rank-1 ratio and decidability performance regarding CR1 setup images.

DR0		SIFT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	24.0	0	1.26	0.00	20.0	0	1.33	0.00	12.0	0	1.14	0.00	24.0	0	1.31	0.00	29.3	0	1.57	0.00
4	27.9	2.03	1.53	0.06	28.0	2.95	1.49	0.05	17.5	2.03	1.26	0.04	29.6	0.56	1.53	0.04	38.3	2.18	1.84	0.04
8	31.9	4.19	1.62	0.07	33.6	3.65	1.60	0.09	21.7	3.21	1.38	0.08	34.1	3.45	1.60	0.06	42.7	3.01	1.94	0.06
16	40.9	3.08	1.70	0.06	34.8	1.72	1.66	0.08	26.4	3.31	1.43	0.07	36.4	4.17	1.70	0.09	46.3	2.6	2.01	0.05
32	39.7	3.49	1.79	0.07	37.3	2.95	1.70	0.07	26.8	2.31	1.41	0.09	38.9	4.74	1.78	0.07	45.2	2.39	2.01	0.06
64	38.5	2.47	1.79	0.11	36.4	2.36	1.61	0.12	27.9	5.35	1.29	0.16	39.7	4.35	1.69	0.09	46.9	3.54	1.93	0.09
128	35.9	4.42	1.70	0.09	36.1	2.84	1.58	0.09	25.9	3.34	1.23	0.09	38.0	4.41	1.64	0.07	45.9	2.68	1.85	0.05
DR0		HOG-DT																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	25.3	0	1.75	0.00	24.0	0	1.79	0.00	18.7	0	1.73	0.00	24.0	0	1.82	0.00	29.3	0	1.86	0.00
4	38.7	0	1.92	0.00	28.0	0	1.99	0.00	28.0	0	1.91	0.00	30.7	0	2.03	0.00	37.3	0	2.12	0.00
8	40.0	0.63	1.93	0.04	25.3	0.63	1.97	0.04	28.7	1.81	2.02	0.05	34.4	2.5	2.00	0.06	36.8	1.57	2.16	0.03
16	45.1	3.07	2.07	0.05	32.3	2.16	1.98	0.06	29.9	1.91	2.00	0.04	35.9	2.98	1.98	0.03	40.9	2.36	2.26	0.03
32	42.1	3.88	1.79	0.08	35.3	3.1	1.81	0.09	30.9	3.31	1.76	0.07	39.9	3.74	1.89	0.06	47.1	2.09	2.03	0.06
64	37.9	3.22	1.37	0.03	31.5	3.68	1.40	0.07	28.0	2.74	1.47	0.07	32.8	3.16	1.40	0.06	44.4	3.45	1.62	0.01
128	22.9	2.65	0.86	0.04	17.2	2.77	0.80	0.07	17.2	2.31	0.86	0.07	17.3	2.08	0.92	0.06	24.5	1.57	0.99	0.04
DR0		HOG-F																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	40.0	0	2.21	0.00	42.7	0	2.10		34.7	0	1.95	0.00	41.6	0	2.16	0.02	48.0	0.84	2.24	0.00
4	45.3	2.81	2.34	0.06	41.1	2.87	2.23	0.10	36.9	3.98	2.03	0.02	46.4	1.05	2.24	0.07	51.3	1.69	2.39	0.04
8	46.8	8.01	2.34	0.05	40.0	2.95	2.15	0.12	41.2	0.42	2.02	0.02	43.3	3.22	2.18	0.08	51.2	0.93	2.35	0.08
16	49.1	1.38	2.29	0.03	42.0	4.88	2.21	0.04	42.7	3.82	2.22	0.04	43.6	3.27	2.31	0.04	53.1	2.25	2.44	0.04
32	52.4	2.44	1.99	0.03	44.3	1.97	2.15	0.05	42.4	3.54	2.01	0.05	52.7	2.69	2.25	0.05	52.9	1.78	2.27	0.04
64	43.1	3.39	1.58	0.04	42.7	3.5	1.69	0.06	32.4	2.6	1.38	0.06	44.0	2.27	1.65	0.06	52.8	2.53	1.75	0.04
128	34.1	3.03	1.01	0.04	25.7	2.44	0.98	0.07	18.4	1.86	0.80	0.06	30.1	3.73	1.08	0.06	32.5	3.94	1.07	0.04
DR0		ULBP																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	33.3	0	1.75		25.3	0	1.58		22.7	0	1.36		38.7	0	1.63	0.00	33.3	0	1.65	0.00
4	33.6	1.38	1.78	0.00	31.3	3.22	1.61	0.03	25.5	1.6	1.41	0.02	34.7	0	1.65	0.00	35.2	2.28	1.70	0.02
8	43.1	2.67	1.89	0.05	41.5	2.7	1.69	0.05	31.1	1.1	1.47	0.04	46.5	0.42	1.74	0.01	47.2	2.1	1.81	0.02
16	45.5	2.31	1.61	0.07	43.9	2.7	1.57	0.07	34.5	4.33	1.36	0.09	45.6	2.33	1.56	0.06	51.7	3.65	1.67	0.04
32	26.8	4.38	1.13	0.06	22.9	2.42	1.06	0.04	19.2	3.28	0.92	0.05	35.3	3.34	1.19	0.06	23.7	2.8	1.18	0.04
64	13.1	4.61	0.63	0.10	13.6	2.65	0.66	0.10	12.8	3.03	0.52	0.09	17.7	3.27	0.77	0.10	14.9	3.76	0.73	0.08
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DR0		GIST																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
2	65.3	0	2.61	0.00	70.7	0	2.80	0.00	61.3	0	2.86	0.00	68.0	0	2.75	0.00	74.7	0	2.86	0.00
4	75.6	0.64	2.50	0.00	76.0	0	2.82	0.00	69.3	0	2.72	0.01	73.3	0	2.72	0.00	80.0	0	2.74	0.00
8	78.1	2.2	2.56	0.09	77.5	0.76	2.80	0.04	73.9	1.43	2.80	0.02	78.4	1.51	2.77	0.07	81.6	1.97	2.78	0.05
16	75.2	1.12	2.82	0.21	77.9	2.01	2.88	0.18	72.1	1.47	2.85	0.12	78.4	0.84	2.84	0.15	79.9	1.47	2.98	0.09
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.8: Rank-1 ratio and decidability performance regarding DR0 setup images.

k = 16		Full CSIP Dataset																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	36.2	0.94	2.15	0.02	35.2	1.29	2.13	0.01	32.4	0.43	2.02	0.02	38.5	0.61	2.15	0.01	46.7	0.51	2.47	0.01
GIST	70.4	0.31	3.18	0.04	71.8	0.8	3.25	0.05	70.8	0.14	3.21	0.05	72.1	0.91	3.25	0.05	73.1	0.33	3.34	0.04
ULBP	36.1	0.54	1.97	0.03	30.1	1.33	1.73	0.07	30.8	0.59	1.82	0.03	31.8	1.32	1.83	0.03	36.4	0.37	1.99	0.01
HOG-DT	24.8	1.21	1.42	0.02	21.8	0.46	1.38	0.05	21.2	0.84	1.36	0.04	22.7	0.82	1.36	0.05	24.9	0.65	1.45	0.01
HOG-F	29.8	0.33	1.51	0.03	24.1	0.6	1.39	0.03	24.4	0.58	1.41	0.01	25.8	0.49	1.40	0.04	28.7	0.27	1.49	0.01

Table A.9: Rank-1 ratio and decidability performance regarding Full CSIP dataset for 16-mixture GMM.

k = 16		MoBIAC - Periocular Region - Full Dataset																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	32.5	16.81	1.92	0.99	29.4	15.26	1.83	0.95	20.9	10.85	1.54	0.80	33.9	17.52	1.87	0.97	38.9	20.11	2.22	1.15
ULBP	24.3	1.32	0.86	0.04	18.8	1.16	0.77	0.05	15.9	1.57	0.69	0.05	19.6	2.79	0.77	0.07	27.0	1.29	0.92	0.02
HOG-DT	18.5	1.89	1.12	0.04	19.1	0.93	1.13	0.03	14.4	1.68	1.07	0.01	16.7	1.17	1.13	0.03	22.9	1.63	1.31	0.03
HOG-F	30.3	1.03	1.42	0.02	28.4	1.29	1.50	0.01	21.6	0.74	1.26	0.03	25.3	1.52	1.50	0.02	32.0	0.86	1.60	0.01
GIST	59.7	1.3	2.71	0.03	50.9	1.44	2.53	0.07	45.8	1.87	2.24	0.09	56.5	0.89	2.81	0.07	57.5	1.76	2.70	0.03
k = 16		MoBIAC - Periocular Region - African Samples																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	30.4	1.88	1.55	0.09	25.0	3.26	1.35	0.11	19.9	3.14	1.09	0.02	30.4	1.88	1.47	0.02	37.7	3.32	1.74	0.08
ULBP	22.8	3.26	0.73	0.09	25.7	2.83	0.56	0.05	12.4	2.5	0.49	0.12	20.4	1.42	0.59	0.09	28.5	3.8	0.72	0.03
HOG-DT	25.7	2.83	0.87	0.16	19.1	2.94	0.89	0.03	18.9	1.97	0.80	0.10	18.0	2.25	0.88	0.02	25.4	2.94	1.03	0.10
HOG-F	30.4	2.31	1.09	0.10	21.5	3.3	0.99	0.05	20.7	4.92	0.96	0.05	24.4	1.46	1.06	0.07	34.4	1.97	1.17	0.06
GIST	57.0	2.38	2.32	0.08	51.3	1.42	2.25	0.10	42.2	2.09	1.86	0.08	56.1	0.97	2.57	0.06	56.7	1.19	2.38	0.03
k = 16		MoBIAC - Periocular Region - African Samples																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	41.0	2.88	1.96	0.07	34.1	3.89	1.76	0.08	26.4	2.3	1.49	0.05	38.9	1.53	1.85	0.08	42.0	2.38	2.18	0.06
ULBP	29.7	5.29	0.88	0.10	26.6	3.59	0.80	0.11	23.8	4.02	0.71	0.08	32.9	2.63	0.84	0.09	35.8	1.78	0.99	0.02
HOG-DT	32.0	4.11	1.57	0.05	28.0	1.53	1.29	0.06	21.2	1.86	0.96	0.03	25.7	2.55	1.30	0.03	34.6	1.97	1.47	0.04
HOG-F	34.1	1.44	1.69	0.01	38.1	2.14	1.61	0.05	28.0	2.41	1.14	0.03	38.1	2.14	1.63	0.04	36.2	3.57	1.65	0.01
GIST	69.7	1.29	3.05	0.05	62.8	4.68	2.82	0.06	55.5	3.16	2.54	0.08	67.1	3.81	3.04	0.06	66.6	0.64	3.06	0.01

Table A.10: Rank-1 ratio and decidability performance regarding MoBIAC periocular region dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM.

k = 16		MoBIAC - Ear - Full Dataset																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	42.6	1.33	2.03	0.04	37.5	1.27	1.82	0.05	31.7	1.4	1.47	0.05	38.8	1.68	1.84	0.05	48.4	0.76	2.06	0.05
ULBP	27.1	0.67	1.50	0.03	23.2	1.68	1.25	0.07	19.8	2	1.07	0.05	22.3	0.76	1.21	0.03	27.8	0.98	1.33	0.03
HOG-DT	36.6	1.68	1.66	0.07	27.1	1.68	1.44	0.05	25.6	1.05	1.30	0.03	26.4	1.27	1.40	0.08	33.2	0.96	1.53	0.04
HOG-F	36.3	1.17	1.87	0.03	33.1	1.28	1.69	0.01	26.6	2.09	1.56	0.05	31.2	0.52	1.58	0.02	36.1	1.12	1.78	0.03
GIST	86.4	0.76	3.97	0.15	85.6	1.21	4.03	0.10	77.1	1.51	3.69	0.10	85.4	0	3.96	0.10	84.8	0.58	4.09	0.03
k = 16		MoBIAC - Ear - African Samples																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	44.0	3.01	1.85	0.07	40.5	1.91	1.52	0.04	32.6	2.73	1.13	0.09	38.4	1.16	1.65	0.05	46.7	1.52	1.71	0.05
ULBP	28.1	1.52	1.20	0.03	21.6	0.64	0.93	0.03	17.0	2.41	0.79	0.04	25.4	1.27	0.94	0.02	24.0	1.76	1.01	0.02
HOG-DT	28.6	1.76	1.25	0.05	21.9	2.24	1.06	0.03	17.2	0.52	0.82	0.02	18.6	2.47	1.00	0.07	23.5	2.24	1.08	0.02
HOG-F	33.3	2.8	1.44	0.02	27.2	1.33	1.19	0.04	20.7	2.08	0.91	0.05	25.8	0.52	1.14	0.02	28.4	2.92	1.23	0.03
GIST	84.4	1.04	3.25	0.11	81.6	1.91	3.15	0.06	67.9	1.76	2.87	0.04	82.8	3.62	3.16	0.11	79.5	2.41	3.24	0.02
k = 16		MoBIAC - Ear - Caucasian Samples																		
10	R				G				B				Gray				RGB			
	R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT	46.1	3.76	1.99	0.06	45.4	3.18	1.99	0.06	36.2	3.26	1.63	0.06	44.2	3.39	2.03	0.05	56.9	2.71	2.22	0.05
ULBP	30.4	0.98	1.44	0.03	26.1	0.53	1.23	0.05	27.8	2.44	1.05	0.02	29.9	0.64	1.32	0.05	28.9	0.64	1.27	0.01
HOG-DT	50.8	2.55	1.79	0.06	37.4	4.11	1.60	0.07	33.7	3.87	1.46	0.03	38.8	4.63	1.45	0.05	44.9	3.05	1.70	0.03
HOG-F	51.3	1.78	2.05	0.03	45.9	2.2	1.98	0.03	41.9	2.44	1.73	0.04	42.8	1.58	1.82	0.04	53.2	2.26	1.99	0.02
GIST	95.3	0	4.71	0.12	93.4	0.64	4.58	0.11	93.2	0.53	4.42	0.08	93.2	0.98	4.44	0.11	93.9	0.53	4.87	0.05

Table A.11: Rank-1 ratio and decidability performance regarding MoBIAC ear dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM.

k = 16		MoBIAC - Hand - Full Dataset																			
10		R				G				B				Gray				RGB			
		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT		29.8	1.95	1.43	0.05	30.0	1.82	1.52	0.13	25.1	1.59	1.37	0.06	30.8	1.39	1.52	0.05	37.6	0.96	1.69	0.04
ULBP		22.5	1.93	1.30	0.08	22.0	1.76	1.24	0.04	22.0	1.04	1.24	0.06	22.9	0.94	1.32	0.03	24.1	0.63	1.40	0.04
HOG-DT		23.3	1.04	1.25	0.03	25.1	1.82	1.39	0.04	18.9	1.7	1.28	0.06	21.4	2.12	1.33	0.05	25.2	1.13	1.41	0.03
HOG-F		20.8	0.53	1.27	0.07	18.1	1.46	1.34	0.07	12.8	0.53	1.24	0.07	16.2	0.57	1.32	0.04	16.6	0.53	1.36	0.03
GIST		56.8	2.26	2.42	0.03	64.4	2.26	2.81	0.05	59.1	1.59	2.64	0.02	62.5	1.58	2.80	0.03	61.4	1.27	2.74	0.01
k = 16		MoBIAC - Hand - African Samples																			
10		R				G				B				Gray				RGB			
		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT		37.4	3.78	1.53	0.08	39.5	5.97	1.61	0.11	29.0	3.8	1.37	0.05	37.7	3.69	1.52	0.05	51.3	3.95	1.77	0.08
ULBP		32.6	2.5	1.35	0.09	40.3	1.15	1.29	0.02	36.4	1.15	1.15	0.09	40.5	2.33	1.25	0.03	42.1	1.07	1.38	0.02
HOG-DT		32.8	3.91	1.47	0.13	38.5	2.03	1.54	0.05	29.7	1.4	1.38	0.07	36.9	5.08	1.56	0.07	38.7	1.4	1.57	0.03
HOG-F		25.4	2.78	1.15	0.07	23.9	1.46	1.29	0.07	17.2	1.46	1.10	0.02	24.1	2.11	1.37	0.07	24.1	1.07	1.26	0.04
GIST		73.9	0.7	3.16	0.07	82.6	1.94	3.58	0.10	79.2	1.67	3.46	0.10	81.3	1.46	3.45	0.12	80.8	0.91	3.57	0.07
k = 16		MoBIAC - Hand - Caucasian Samples																			
10		R				G				B				Gray				RGB			
		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec		R1 %		Dec	
SIFT		27.3	2.05	1.04	0.05	28.5	2.24	1.12	0.05	27.3	1.85	1.04	0.08	31.8	4.01	1.23	0.08	34.3	1.9	1.31	0.07
ULBP		16.5	1.05	0.89	0.04	14.3	1.43	0.85	0.07	13.0	1.9	0.74	0.06	16.3	0.88	0.84	0.02	15.3	1.37	0.94	0.04
HOG-DT		17.5	1.25	0.57	0.07	17.3	1.05	0.62	0.07	15.8	1.68	0.46	0.04	17.8	1.85	0.54	0.06	21.8	1.12	0.59	0.03
HOG-F		18.0	0.68	0.47	0.02	10.3	0.56	0.44	0.03	9.3	0.68	0.28	0.00	13.5	0.56	0.47	0.04	11.5	1.37	0.42	0.01
GIST		46.3	2.34	1.95	0.03	62.0	2.74	2.50	0.08	55.0	1.53	2.26	0.07	58.5	4.45	2.42	0.03	56.3	1.98	2.44	0.05

Table A.12: Rank-1 ratio and decidability performance regarding MoBIAC hand dataset with respect to the full dataset and the two ethnicities groups for 16-mixture GMM.

References

- [1] A. Gelb and J. Clack. Identification for development: The biometrics revolution, 2013.
- [2] A. Genovese, V. Piuri, and F. Scotti. *Touchless Palmprint Recognition Systems*. Advances in Information Security. Springer International Publishing, 2014. URL: <https://books.google.it/books?id=CP6JBAAAQBAJ>.
- [3] V. Oliyil Kunnil M. Baier S. Mil'shtein, A. Pillai and P. Bustos. Recent Application in Biometrics. *Applications of Contactless Fingerprinting*, 2011. URL: <http://www.intechopen.com/books/recent-application-in-biometrics/applications-of-contactless-fingerprinting>.
- [4] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, pages 2564–2571, Washington, DC, USA, 2011. IEEE Computer Society. URL: <http://dx.doi.org/10.1109/ICCV.2011.6126544>, doi: 10.1109/ICCV.2011.6126544.
- [5] John Daugman. Biometric decision landscapes, 2000.
- [6] P.S.P. Wang. *Pattern Recognition, Machine Intelligence and Biometrics*. Springer Berlin Heidelberg, 2012. URL: <https://books.google.pt/books?id=GnwyMm0QH5UC>.
- [7] Shoichiro Aoyama, Koichi Ito, Takafumi Aoki, and Haruki Ota. A contactless palmprint recognition algorithm for mobile phones. In *International Workshop on Advanced Image Technology*, 2013.
- [8] KamYuen Cheng and A. Kumar. Contactless finger knuckle identification using smartphones. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*, pages 1–6, Sept 2012.
- [9] Dal-ho Cho, Kang Ryoung Park, Dae Woong Rhee, Yanggon Kim, and Jonghoon Yang. Pupil and iris localization for iris recognition in mobile phones. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2006. SNPD 2006. Seventh ACIS International Conference on*, pages 197–201, June 2006. doi:10.1109/SNPD-SAWN.2006.58.
- [10] A. Pflug and C. Busch. Ear biometrics: a survey of detection, feature extraction and recognition methods. *Biometrics, IET*, 1(2):114–129, June 2012. doi:10.1049/iet-bmt.2011.0003.

- [11] Madhavi Gudavalli, S. Viswanadha Raju, A. Vinaya Babu, and D. Srinivasa Kumar. Multimodal biometrics – sources, architecture and fusion techniques: An overview. *Biometrics and Security Technologies, International Symposium on*, 0:27–34, 2012. doi:<http://doi.ieeecomputersociety.org/10.1109/ISBAST.2012.24>.
- [12] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [13] Gil Melfe Mateus Santos, Emanuel Grancho, Marco V. Bernardo, and Paulo Torrão Fiadeiro. Fusing iris and periocular information for cross-sensor recognition. *Pattern Recognition Letters*, 57:52–59, 2015. URL: <http://dx.doi.org/10.1016/j.patrec.2014.09.012>, doi:10.1016/j.patrec.2014.09.012.
- [14] *Biometric Recognition: Challenges and Opportunities*. The National Academies Press, 2010. URL: <http://www.nap.edu/catalog/12720/biometric-recognition-challenges-and-opportunities>.
- [15] K. Delac and M. Grgic. A survey of biometric recognition methods. In *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*, pages 184–193, June 2004.
- [16] Ben Taylor. How the smartphone defeated the point-and-shoot digital camera, 2014. URL: <http://www.pcworld.com/article/2466500/how-the-smartphone-defeated-the-point-and-shoot-digital-camera.html>.
- [17] R. Klette. *Concise Computer Vision: An Introduction into Theory and Algorithms*. Undergraduate Topics in Computer Science. Springer Science and Business, 2014. URL: <https://books.google.pt/books?id=ZCu8BAAAQBAJ>.
- [18] N.D. Lane, E. Miluzzo, Hong Lu, D. Peebles, T. Choudhury, and A.T. Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, Sept 2010. doi:10.1109/MCOM.2010.5560598.
- [19] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision: Part IV, ECCV’10*, pages 778–792, Berlin, Heidelberg, 2010. Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=1888089.1888148>.
- [20] Linda G. Shapiro and George C. Stockman. *Computer Vision*. Prentice Hall, 2001.
- [21] Mengling Feng and Yap-Peng Tan. Adaptive binarization method for document image analysis. In *ICME*, pages 339–342. IEEE, 2004. URL: <http://dblp.uni-trier.de/db/conf/icmcs/icme2004.html#FengT04>.
- [22] A. Kumar and Y. Zhou. Human identification using knucklecodes. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS ’09. IEEE 3rd International Conference on*, pages 1–6, Sept 2009. doi:10.1109/BTAS.2009.5339021.
- [23] A. Uhl and P. Wild. Comparing verification performance of kids and adults for fingerprint, palmprint, hand-geometry and digitprint biometrics. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS ’09. IEEE 3rd International Conference on*, pages 1–6, Sept 2009. doi:10.1109/BTAS.2009.5339069.

- [24] K. P. Zaw and A. S. Khaing. Implementation of Contactless Finger Knuckle Identification System. *International Journal of Science, Engineering and Technology Research*, 3, June 2014.
- [25] A. Kumar and C. Kwong. Towards contactless, low-cost and accurate 3d fingerprint identification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3438–3443, June 2013. doi:10.1109/CVPR.2013.441.
- [26] Anil Jain, Lin Hong, and Sharath Pankanti. When it comes to working biometric identification technologies, it's not only our fingerprints that do the talking. now, our eyes, hands, signature, speech, and even facial temperature can id us. *COMMUNICATIONS OF THE ACM*, 43(2):91–98, 2000.
- [27] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1148–1161, November 1993. URL: <http://dx.doi.org/10.1109/34.244676>, doi:10.1109/34.244676.
- [28] A.A. Ross, K. Nandakumar, and A.K. Jain. *Handbook of Multibiometrics*. International Series on Biometrics. Springer, 2006. URL: <https://books.google.pt/books?id=JpUdlJnuE2MC>.
- [29] Anil K. Jain, Kai Cao, and Sunpreet S. Arora. Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–8, Sept 2014. doi:10.1109/BTAS.2014.6996252.
- [30] G. Karthick Allen George. Palmprint recognition using ridge features. *International Journal of Advanced Research in Computer and Communication Engineering*, 2:23–26, 2013.
- [31] S.Yagasudha. User verification by hierarchical palmprint matching. *International Journal of Computer Applications® (IJCA)*, pages 23–26, 2014.
- [32] Jifeng Dai, Jianjiang Feng, and Jie Zhou. Robust and efficient ridge-based palmprint matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(8):1618–1632, 2012. URL: <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.237>, doi:10.1109/TPAMI.2011.237.
- [33] W. Li, L. Zhang, D. Zhang, G.M. Lu, and J.Q. Yan. Efficient joint 2d and 3d palmprint matching with alignment refinement. In *CVPR10*, pages 795–801, 2010.
- [34] N. Moço, P.L. Correia, and L. D. Soares. Smartphone-based palmprint recognition system. In *International Conf. on Telecommunications - ICT*, volume ., pages .–., May 2014.
- [35] Alberto de Santos-Sierra, Carmen Sanchez-Avila, Javier Guerra-Casanova, and Aitor Mendaza-Ormaza. Hand biometrics in mobile devices. In Girija Chetty, editor, *Advanced Biometric Technologies*. InTech, 2011. URL: <http://www.intechopen.com/books/advanced-biometric-technologies/hand-biometrics-in-mobile-devices1>.
- [36] K.V. Kale, Y.S. Rode, M.M. Kazi, S.B. Dabhade, and S.V. Chavan. Multimodal biometric system using fingernail and finger knuckle. In *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*, pages 279–283, Aug 2013. doi:10.1109/ISCBI.2013.63.

- [37] Rabia Jafri and Hamid R. Arabnia. A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(2):41–68, 2009. doi:10.3745/jips.2009.5.2.041.
- [38] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J.-F. Bonastre, P. Tresadern, and T. Cootes. Bi-modal person recognition on a mobile phone: Using mobile phone data. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 635–640, July 2012. doi:10.1109/ICMEW.2012.116.
- [39] Marc Davis, Michael Smith, John Canny, Nathan Good, Simon King, and Rajkumar Janakiraman. Towards context-aware face recognition. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, pages 483–486, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1101149.1101257>, doi:10.1145/1101149.1101257.
- [40] Kangrok Oh, Beom-Seok Oh, Kar-Ann Toh, Wei-Yun Yau, and How-Lung Eng. Combining sclera and periocular features for multi-modal identity verification. *Neurocomput.*, 128:185–198, March 2014. URL: <http://dx.doi.org/10.1016/j.neucom.2013.01.066>, doi:10.1016/j.neucom.2013.01.066.
- [41] DaeSik Jeong, Hyun-Ae Park, KangRyoung Park, and Jaihie Kim. Iris recognition in mobile phone based on adaptive gabor filter. In David Zhang and AnilK. Jain, editors, *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 457–463. Springer Berlin Heidelberg, 2005. URL: http://dx.doi.org/10.1007/11608288_61, doi:10.1007/11608288_61.
- [42] C.N. Padole and H. Proenca. Periocular recognition: Analysis of performance degradation factors. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 439–445, March 2012. doi:10.1109/ICB.2012.6199790.
- [43] Jonathon M. Smerka and B. V. K. Vijaya Kumar. What is a "good" periocular region for recognition? In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW '13*, pages 117–124, Washington, DC, USA, 2013. IEEE Computer Society. URL: <http://dx.doi.org/10.1109/CVPRW.2013.25>, doi:10.1109/CVPRW.2013.25.
- [44] J. C. Monteiro and J. S. Cardoso. Periocular recognition under unconstrained settings with universal background models.
- [45] *2013 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, CIBIM 2013, Singapore, April 16-19, 2013*. IEEE, 2013. URL: <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6595340>.
- [46] Daniel Povey, Lukáš Burget, Mohit Agarwal, Pinar Akyazi, Feng Kai, Arnab Ghoshal, Ondřej Glembek, Nagendra Goel, Martin Karafiát, Ariya Rastrow, Richard C. Rose, Petr Schwarz, and Samuel Thomas. The subspace gaussian mixture model—a structured model for speech recognition. *Comput. Speech Lang.*, 25(2):404–439, April 2011. URL: <http://dx.doi.org/10.1016/j.csl.2010.06.003>, doi:10.1016/j.csl.2010.06.003.
- [47] EmguCV. .Net wrapper to the OpenCV image processing library. URL: http://www.emgu.com/wiki/index.php/Main_Page.

- [48] D.J. Hurley, M.S. Nixon, and J.N. Carter. Force field energy functionals for image feature extraction. In *Proc. BMVC*, pages 60.1–60.10, 1999. doi:10.5244/C.13.60.
- [49] Kiminori SATOH Takaya YUIZONO, Yu WANG and Shigeru NAKAYAMA. Study on individual recognition for ear images by using genetic local search. 2002.
- [50] J. Kotzerke, A. Arakala, S. Davis, K. Horadam, and J. McVernon. Ballprints as an infant biometric: A first approach. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on*, pages 36–43, Oct 2014. doi:10.1109/BIOMS.2014.6951533.
- [51] Lingyu Wang, Graham Leedham, and David Siu-Yeung Cho. Minutiae feature analysis for infrared hand vein pattern biometrics. *Pattern Recogn.*, 41(3):920–929, March 2008. URL: <http://dx.doi.org/10.1016/j.patcog.2007.07.012>, doi:10.1016/j.patcog.2007.07.012.
- [52] Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. Biometric gait authentication using accelerometer sensor. *Journal of Computers*, 1(7), 2006. URL: <http://www.ojs.academypublisher.com/index.php/jcp/article/view/01075159>.
- [53] H. Hollien, R. Huntley Bahr, and J. D. Harnsberger. Issues in forensic voice. *J Voice*, 28(2):170–84, 2014. Hollien, Harry Huntley Bahr, Ruth Harnsberger, James D eng Review 2013/11/02 06:00 J Voice. 2014 Mar;28(2):170-84. doi: 10.1016/j.jvoice.2013.06.011. Epub 2013 Oct 28. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24176301>, doi:10.1016/j.jvoice.2013.06.011.
- [54] Arun Ross and Anil Jain. Multimodal biometrics: An overview. In *Proc. of 12th European Signal Processing Conference (EUSIPCO)*, pages 1221–1224, 2004.
- [55] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, pages 2548–2555, Washington, DC, USA, 2011. IEEE Computer Society. URL: <http://dx.doi.org/10.1109/ICCV.2011.6126542>, doi:10.1109/ICCV.2011.6126542.
- [56] Raphael Ortiz. Freak: Fast retina keypoint. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 510–517, Washington, DC, USA, 2012. IEEE Computer Society. URL: <http://dl.acm.org/citation.cfm?id=2354409.2354903>.
- [57] Jian-Gang Wang, Jun Li, Wei-Yun Yau, and Eric Sung. Boosting dense sift descriptors and shape contexts of face images for gender recognition. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 96–102. IEEE, 2010.
- [58] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, January 1996. URL: [http://dx.doi.org/10.1016/0031-3203\(95\)00067-4](http://dx.doi.org/10.1016/0031-3203(95)00067-4), doi:10.1016/0031-3203(95)00067-4.
- [59] Ojala T., Pietikäinen M, and Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. 2002. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7):971 - 987.

- [60] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):915–928, June 2007. URL: <http://dx.doi.org/10.1109/TPAMI.2007.1110>, doi:10.1109/TPAMI.2007.1110.
- [61] Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with center-symmetric local binary patterns.
- [62] Topi Maenpää and Matti Pietikainen. Classification with color and texture: jointly or separately?, 2004.
- [63] Xiaoyang Tan and Bill Triggs. Fusing gabor and lbp feature sets for kernel-based face recognition. In *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures, AMFG'07*, pages 235–249, Berlin, Heidelberg, 2007. Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=1775256.1775278>.
- [64] *IEEE Workshop on Applications of Computer Vision (WACV 2009), 7-8 December, 2009, Snowbird, UT, USA*. IEEE Computer Society, 2009. URL: <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5398686>.
- [65] Marko Heikkilä, Matti Pietikainen, and Senior Member. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Anal. Machine Intell.*, 28:2006, 2006.
- [66] Michael J. Swain and Dana H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1):11–32, November 1991. URL: <http://dx.doi.org/10.1007/BF00130487>, doi:10.1007/BF00130487.
- [67] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005. URL: <http://lear.inrialpes.fr/pubs/2005/DT05>.
- [68] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145–175, 2001.
- [69] S. Bharadwaj, H.S. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–6, Sept 2010. doi:10.1109/BTAS.2010.5634498.
- [70] BiometricUpdate. News about biometrics. URL: <http://www.biometricupdate.com/tag/africa>.
- [71] Suprema. biometric technologies don't just meet, but well exceed customers' expectations. URL: <http://www.suprema.co.za/>.
- [72] Zetes. European expert in supply chain, government, and mobility solutions and services. URL: <http://www.suprema.co.za/>.
- [73] Morphotrak. SAFRAN Morphotrak - Delivering innovative multi-biometric solutions. URL: http://www.morphotrak.com/MorphoTrak/MorphoTrak/mt_multi-biometrics.html.

- [74] MegaMatcher. Biometric and Artificial Intelligent technologies. URL: <http://www.neurotechnology.com/megamatcher.html>.
- [75] data6tems. Software Development & Biometric Solutions. URL: <http://www.data6tems.com>.
- [76] Fingerprint-I.T. Fingerprint Identification Technology. URL: <http://www.fingerprint-it.com/Pages/Products-Neuro-nosidebar.aspx>.
- [77] NEUROTECHNOLOGY. Kenya's new electronic voter registration system. 2011.
- [78] BioComponents. AWARE Biometrics Software. URL: <http://www.aware.com/biometrics/index.html>.
- [79] AwareXM Mobile. AwareXM Mobile Biometrics Software. URL: <http://www.aware.com/biometrics/awarexmmobile.html>.
- [80] VCL. Face recognition Researcher Group Homepage. URL: <http://www.face-rec.org/general-info/>.
- [81] Bosphorus. BOSPHORUS Hand Database. URL: <http://bosphorus.ee.boun.edu.tr/hand/Home.aspx>.
- [82] CVonline. CVonline: Image Databases. URL: <http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm>.
- [83] Sebastiano Battiato and José Braz, editors. *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 3, Lisbon, Portugal, 5-8 January, 2014*. SciTePress, 2014.
- [84] Ajay Kumar and Chenye Wu. Automated human identification using ear imaging. *Pattern Recognition*, 45(3):956 – 968, 2012. URL: <http://www.sciencedirect.com/science/article/pii/S0031320311002706>, doi:<http://dx.doi.org/10.1016/j.patcog.2011.06.005>.
- [85] Dariusz Frejlichowski and Natalia Tyszkiewicz. The west pomeranian university of technology ear database - a tool for testing biometric algorithms. In Aurélio C. Campilho and Mohamed S. Kamel, editors, *ICIAR (2)*, volume 6112 of *Lecture Notes in Computer Science*, pages 227–234. Springer, 2010. URL: <http://dblp.uni-trier.de/db/conf/iciar/iciar2010-2.html#FrejlichowskiT10>.
- [86] Yongkang Wong, Shaokang Chen, Sandra Mau, Conrad Sanderson, and Brian C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. *CoRR*, abs/1304.0869, 2013. URL: <http://arxiv.org/abs/1304.0869>.
- [87] Oliver Jesorsky, Klaus J. Kirchberg, and Robert Frischholz. Robust face detection using the hausdorff distance. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA '01*, pages 90–95, London, UK, UK, 2001. Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=646073.677460>.

- [88] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L.A. Alexandre. The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance. *IEEE Trans. PAMI*, 32(8):1529–1535, August 2010. doi:[10.1109/TPAMI.2009.66](https://doi.org/10.1109/TPAMI.2009.66).
- [89] D.A. Reynolds. An overview of automatic speaker recognition technology. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–4072–IV–4075, May 2002. doi:[10.1109/ICASSP.2002.5745552](https://doi.org/10.1109/ICASSP.2002.5745552).
- [90] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987.
- [91] K. Shinoda and N. Inoue. Reusing speech techniques for video semantic indexing [applications corner]. *Signal Processing Magazine, IEEE*, 30(2):118–122, March 2013. doi:[10.1109/MSP.2012.2230520](https://doi.org/10.1109/MSP.2012.2230520).
- [92] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1–3):19 – 41, 2000. URL: <http://www.sciencedirect.com/science/article/pii/S1051200499903615>, doi:<http://dx.doi.org/10.1006/dspr.1999.0361>.
- [93] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.
- [94] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1627–1645, September 2010. URL: <http://dx.doi.org/10.1109/TPAMI.2009.167>, doi:[10.1109/TPAMI.2009.167](https://doi.org/10.1109/TPAMI.2009.167).
- [95] I. T. Nabney. *NETLAB: Algorithms for Pattern Recognition*. Advances in Pattern Recognition. Springer, 1st edition, 2002.
- [96] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.
- [97] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP Journal on Image and Video Processing*, 2014(1), 2014. URL: <http://dx.doi.org/10.1186/1687-5281-2014-34>, doi:[10.1186/1687-5281-2014-34](https://doi.org/10.1186/1687-5281-2014-34).