UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR
DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA Y DE LAS COMUNICACIONES

DEALING WITH VARIABILITY FACTORS AND ITS APPLICATION TO BIOMETRICS AT A DISTANCE

–TESIS DOCTORAL–

TRATAMIENTO DE FACTORES DE VARIABILIDAD Y SU APLICACIÓN EN BIOMETRÍA A DISTANCIA

Author: Pedro Tomé González
(Ingeniero de Telecomunicación, Universidad Autónoma de Madrid)

A Thesis submitted for the degree of:

Doctor of Philosophy

Madrid, November 2013

Colophon

This book was typeset by the author using LaTeX2e. The main body of the text was set using a 11-points Computer Modern Roman font. All graphics and images were included formatted as Encapsuled Postscript (TM Adobe Systems Incorporated). The final postscript output was converted to Portable Document Format (PDF) and printed.

This Thesis was printed with the financial support from EPS-UAM and the Biometric Recognition Group-ATVS.

Department:	Tecnología Electrónica y de las Comunicaciones Escuela Politécnica Superior Universidad Autónoma de Madrid (UAM), SPAIN
PhD Thesis:	Dealing with variability factors and its applications to biometrics at a distance
Author:	**Pedro Tomé González** Ingeniero de Telecomunicación (Universidad Autónoma de Madrid)
Advisor:	**Julián Fiérrez Aguilar** Doctor Ingeniero de Telecomunicación (Universidad Politécnica de Madrid) Universidad Autónoma de Madrid, SPAIN
Year:	2013
Committee:	President: **Javier Ortega García** Universidad Autónoma de Madrid, SPAIN
	Secretary: **José M. Martínez Sánchez** Universidad Autónoma de Madrid, SPAIN
	Vocal 1: **Mark S. Nixon** University of Southampton, UNITED KINGDOM
	Vocal 2: **José Luis Alba Castro** Universidad de Vigo, SPAIN
	Vocal 3: **Enrique Cabello Pardos** Universidad Rey Juan Carlos, SPAIN

The research described in this Thesis was carried out within the Biometric Recognition Group – ATVS at the Dept. of Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid (from 2009 to 2013). The project was partially funded by a PhD scholarship from Universidad Autónoma de Madrid.

The author was awarded with a PhD scholarship from Universidad Autónoma de Madrid between 2009 and 2013 which supported the research summarized in this Dissertation.

The author was awarded with a COST STSM - Short-Term Scientific Mission from the European project COST 2101 Biometric For Identity Documents and Smart Cards, which supported his research stay carried out at University of Kent, Canterbury, England from October 2009 to December 2009.

The author was awarded with a travel and fees grant to the 7th International Summer School on Biometrics European Commission, IAPR, MORPHO, IEEE Biometrics Council at June 2010.

Abstract

THIS THESIS IS FOCUSED ON dealing with the variability factors in biometric recognition and applications of biometrics at a distance. In particular, this PhD Thesis explores the problem of variability factors assessment and how to deal with them by the incorporation of soft biometrics information in order to improve person recognition systems working at a distance. The proposed methods supported by experimental results show the benefits of adapting the system considering the variability of the sample at hand.

Although being relatively young compared to other mature and long-used security technologies, biometrics have emerged in the last decade as a pushing alternative for applications where automatic recognition of people is needed. Certainly, biometrics are very attractive and useful for video surveillance systems at a distance, widely distributed in our lifes, and for the final user: forget about PINs and passwords, you are your own key. However, we cannot forget that as any technology aimed to provide a security service, biometric systems should ensure a reliable performance in any scenario. Thus, it is of special relevance to understand and analyse the variability factors to which they are subjected in order to ensure a suitable performance and increase their benefits for the users.

In this context, the present PhD Thesis gives an insight into the difficult problem of variability factors evaluation through the systematic study of biometric scenarios at a distance and the analysis of effective compensation methodologies that can minimize the effects of them. Pursuing the aim to increase the performance of the remote person recognition in this thriving technology. This way, the experimental studies presented in this Dissertation can help to further develop the ongoing variability compensation efforts, and may be used as guidelines to adapt the existing systems in biometric at a distance and make them more secure and stable.

The problem of variability compensation in biometric systems had already been addressed in some previous works, but in most cases not using the acquisition distance related with the variability factors in order to identify and define scenarios. In this Dissertation, after summarizing and classifying the most relevant works related to the Thesis and defining what we understand as scenario at a distance, we describe methods applied throughout the experimental chapters. These experimental chapters are dedicated first to the study of variability factors (scenario analysis), and then to the application of the proposed techniques to deal with them (soft biometrics and adaptive fusion). All experiments are conducted using standard biometric data and benchmarks.

The experimental part of the Thesis starts with a scenario evaluation of the variability factors found in face recognition systems. We evaluate, between others, the relationship between variability factors and the acquisition distance in this kind of systems, the variability of facial landmarks in mugshot and CCTV images, and the performance variability of different facial regions of the human face on various forensic scenarios at a distance. In addition to be useful

background information that can guide and help experts to interpret and evaluate face evidences, these findings can have a significant impact on the design of face recognition algorithms.

We then study various types of soft biometric information available in biometrics at a distance suitable for video surveillance and forensics applications. These soft labels can be visually identified at a distance by humans (or an automatic system) and their discriminative information will vary depending on the distance. It is worth noting that this relation between scenarios at a distance and the performance of soft biometrics for person recognition has not been studied in this way before. Moreover, the largest set of morphological facial soft biometric features extracted following forensic protocols is also introduced and evaluated. The experimental results using this set of features show that a system that is completely based on facial soft biometrics features for forensics is feasible.

Finally, we study experimentally various types of adaptive fusion exploiting soft biometrics. In particular, we study: scenario-based, soft biometrics-based, facial regions-based, and color facial regions-based schemes of score–level fusion and their benefits in systems at a distance. The proposed adaptive fusion schemes achieve notable improvements demonstrating their utility in biometrics at a distance.

The research work described in this Dissertation has led to novel contributions which include the development of two new methods to deal with variability factors in biometrics systems at a distance, namely: *i*) soft biometrics suitable for video surveillance and forensics, and *ii*) adaptive fusion schemes at score–level based on scenario acquisition, soft biometrics, facial regions, and color facial regions. Moreover, different original experimental studies have been carried out during the development of the Thesis (e.g., relation between scenarios at a distance and variability factors). Besides, the research work completed throughout the Thesis includes the generation of various literature reviews and the generation of new biometric resources.

A MI FAMILIA.

A MI PADRE.

(One can only see what one observes,
and one observes only things which are already in the mind)
(Se puede ver sólo lo que se observa
y se observa sólo lo que ya está en la mente)

Alphonse Bertillon, (Francia 1853, Suiza 1914).

Acknowledgements

THIS PHD THESIS summarizes the work carried out during my Ph.D. studies with the Biometric Recognition Group - ATVS since 2009. This research group was established in 1994 at the Dept. of Ingeniería Audiovisual y Comunicaciones (DIAC) of the Universidad Politécnica de Madrid (UPM) and since 2004 is affiliated to the Dept. of Tecnología Electrónica y de las Comunicaciones of the Universidad Autónoma de Madrid (UAM). This Thesis has been mainly supported by a Ph.D. scholarship granted to the author by Universidad Autónoma de Madrid, which covered the period between October 2009 and October 2013, and also by various Spanish and European projects referenced in related publications.

Foremost, I would like to thank to Prof. Javier Ortega-García for his guidance and support over these years, who opened his door for me to his group (Biometric Recognition Group - ATVS) for a challenging project when I was pursuing my Master degree and who gave me the opportunity of know the research world. I really appreciate the confidence he has always shown in me and has definitely fostered my motivation. During these years I have benefited from his courage, self-mastery, and intelligent effort.

I would like to give an special mention to my advisor Dr. Julián Fiérrez (*the motivator*), who has motivated me all the time with his wise and intelligent advices. Thank for his guidance and support over the past four years. I really appreciate to have worked together, his help and support because thanks to it I have benefited from his future vision, discipline, hard work, and intelligent effort. I don't know how but in his prophetic jigsaw puzzle all the pieces interlock with each other and it always has solution in his brain.

During the journey of my Ph.D. degree I have been fortunate to meet many excellent professionals and colleagues. I am especially grateful to Dr. Rubén Vera-Rodríguez, one fellow researcher at ATVS who has actively contributed to some of the research lines that have finally shaped the Thesis, for his continuous support, for his valuable advice, for his reviews and comments on the Thesis. I can't either forget to acknowledge to my first and Master degree advisor, Dr. Fernando Alonso-Fernández with whom I started my research career and always had a smile and time for me, thanks Fer!.

In the framework of the ATVS research group, I have also received support from Prof. Joaquín González-Rodríguez and Dr. Doroteo Torre-Toledano, feeling fortunate to be their student during my BSc and MSc degree and learn from their wise advice as student at the beginning and as work colleges in the last years.

From my early years as an undergraduate student I would like also to thank the professors who motivated me to endeavor a research career in signal processing and pattern recognition, specially Profs. José M. Martínez, Jesús Bescós-Cano, Jorge A. Ruíz-Cruz, José Luis Masa Campos, Bazil Taha Ahmed, Alejandro Sierra Urrecho, Javier Garrido Salas, F. Javier Gómez Arribas, José Colás Pasamontes, Eduardo Boemo, J.E. López de Vergara, etc.

The opportunity of visiting foreign institutions during my Ph.D. studies was a real luck. I especially remember my first research stay at the University of Kent, Canterbury, UK. I had the fortune to meet Prof. Mike Fairhurst there, who kindly host me at the Image and Information Engineering Research Group where I met a fantastic group of people. I would like to especially mention Dr. Márjory Da Costa-Abreu whose extraordinary care during my visit during those three months. Also, they were partly responsible of my happiness during the stay, with their attention and help, thank you Teresa (my mafiosa) and Lina (my teacher), I will never forget you, and yes, I still have to visit Sicily and Jordan.

My second stay was at Information: Signals, Image and Systems Research Group (ISIS) at the University of Southampton, Southampton, UK, with Prof. Mark Nixon. During those three months I received an enormously motivating support and confidence from Prof. Mark S. Nixon, and I could benefit from his mastery in the field of computer vision, which led to some contributions contained in this Thesis. I had also the opportunity to meet a fantastic group of people among others, Dr. Darko Matovski, whose kindness and patience as well as the friendly conversations helped me in my research and made me feel at home in the daily work.

My last two-month research stay was conducted with Prof. Brian Lovell at the University of Queensland, Brisbane, Australia, where the Dissertation was written. I really appreciate his support in spite of his multiple commitments as director of the Security and Surveillance Research Group (SAS). There I also had the fortune to meet a fantastic group of people supported me in my Thesis writing and made me feel at home in the daily work. I can't forget to mention here to Jan and Irina (lovely couple), my flatmates during my stay in Brisbane and who I had great time together, I will always to remember your families and I wish you the best together.

Ya llegando a este punto permítanme explayarme en mi lengua materna, en la cuál, aún no siendo un Cervantes o un Lope de Vega quiero dejar constancia y agradecer a tantas personas que seguro injustamente olvidaré a muchas de ellas.

Dicen que un gran hombre se rodea de grandes personas, yo no se si seré tal, pero una cosa tengo clara, decir grande se quedaría pequeño como adjetivo para toda la gente que me rodea.

Empezando por la gente que me ha acompañado durante estos años en el ATVS, mis compañeros de trabajo y amigos, con quiénes he pasado grandes momentos. Desde esos con los que empecé en el laboratorio, hoy día reputados profesores, Julián Fiérrez (Julk), Daniel Ramos (thief of smiles), Javier Galbally (er cantabrú), Javier González (esto no es bambi) no podéis imaginar las innumerables cosas que he aprendido y sigo aprendiendo de cada uno de vosotros. Continuando con mis compañeros de sala, Javi Franco (artisten), Rubén Vera (pixxxaa er sevillano), Marta Gómez (malita sijuan), aquellos que ya se fueron Manuel Freire (kedro para mi), Ignacio López (niño google), Miriam Moreno (barbie morennnnnn, I miss you everyday), María Puertas (Popiiiiiinnsssss), y aquellos que llegaron al final Ruifang Wang (One, no two), Ester González (Canaria que no Cubana), Alicia Lozano (cumar in love), Álvaro Diegüez (timofonico boy), Sara Antequera (Secretaria), Fernando Espinoza (currante nato), Chicas Cát (Sandra, Fátima, María y Sara) y Ram (the Indian guy). Mención especial para ese visitante, ese loco pero gran tipo Rudolf Haraksim (alias hello kitty), remember that you have a friend in Spain.

Ya terminado no debo olvidar a aquellos o aquellas que ya no están pero que he tenido a mi lado en algún momento en todos estos años, Virginia Ruiz y Almudena Gilperez (mis babies), Alicia Beiner (la rubia del lugar),... Y ya para acabar a aquellos a los que han tenido que aguantarme como tutor del PFC, Andrei, Pablo, Luis, Maya, con especial mención de mi pupilo y amigo Javier Eslava. Gracias ateuveianos, por los mil y un recuerdos que compartimos, por esas aventuras, por esas mañanas, tardes y noches de fiesta. ¡A tod@s GRACIAS!.

Otra mención especial se merecen esas personas con las que he compartido mi licenciatura, grandes todos, Rodrigo López (business man), Verónica Peña (ni niña siempre, cuídala Javi), Elena Ortiz (Eleee), Sonsoles Herrero (sonssssssoo), Fernando García (delegado), Alberto Harriero (chinitaaaaa), Alejandro Abejón (abejito), Juan Manuel Rubio (one backpack, 15 days), David García (be water my friend), Fernando Alonso (finesssss o finlandés), Ismael Mateos (the big one), Ignacio Izquierdo (space man),... Señores os espero en nuevas aventuras a lo largo del mundo.

Durante estos años he conocido a mucha gente pero tengo que dedicaros unas líneas a vosotros *zumbao* (que no tumbao) salsa students, los originales dónde los haya: Gloria (vaya tela...), Patri (no me rio de ti sino contigo), Julia (¡quiero el divorcio!), Aurora (bachata teacher), José (grandullón), Ángel (esa etiqueta la pusiste tú allí), Miguel (deja de doparte) y por supuesto Robin (Master), os agradezco esas tardes de risas, noches de bailes y demás momentos que tanto hemos reído juntos. Que nos quiten los bailao chicos y vosotras me debéis un baile.

Siguiendo por la línea de aficiones no puedo olvidar a mi gente del Jiujitsu, Jorge, Maestro y amigo tus enseñanzas van más allá del Tatami, por otros tantos Danes a tu lado, Germán (quién nos ha visto y quién nos ve), Isabel, Mari Carmen, David, mi José, Ana, Eva, Marí Mar,... gracias familia, ¡Tai otoshis para todos!

Ya llegando al final me centro en ese grupo con denominación de origen *MorataYork*, barrio, donde he conocido a los mejores compañeros de toda la vida, amigos desde la infancia de lo que puedo presumir con gran cariño hacia todos, que pase lo que pase, estemos donde estemos, seguimos juntos. Ingenieros, doctores, jueces, informáticos, lingüistas, actrices, farmacéuticas, deportistas, periodistas, educadoras, bailarinas, artistas,... a vosotros a los cuales considero mi familia os dedico este espacio: A Pitu, Nacho, Belén, Samantha, Laura, Irene, Lidia, Yessi, Lucía, Angela, y a ti Lenita (Alemana con corazón de Española), sin duda alguna sois grandes personas. A Hugo (su señoria), la razón y la cordura pero controlada por un corazón fiel y alocado, a Miguel (follow your dreams), tu pasión, tu amistad, tu forma de ver las cosas, gracias por todos esos momentos vividos y por los que nos quedan. Y que decir de vosotras mis “Galgas”, Sara (awesome dancer) e Itziar (mi cantautora preferida, famous one day) os quiero pequeñas, sois lo mejor, estáis colgadas pero sois lo mejor de lo mejor, afortunado el que os tenga. Y sobre todo a ti, Isma por esas tardes de retiro madrileño corriendo, por esas noches de cinco minutos convertidos en horas de conversaciones apasionantes, mi sancho sin panza, mi escudero en un sin fin de aventuras con tantas experiencias compartidas, mi hermano, desde el amor al odio brindo por todas aquellas que aún nos quedan por lidiar juntos. A todos gracias, gracias y más gracias por estar a mi lado durante todo este tiempo y por todo lo que nos queda por compartir.

Podría decir que ya no queda nada pero mentiría, no puedo olvidar mencionar entre líneas o simplemente agradecer a esa persona que recuperó mi palpito de vida, para ti, psicóloga vasca (más madrileña cada día), atractora de locos, sonriente y positiva donde las haya, soñadora en vida y simplemente grandísima persona, eskerrik asko. Y a ti Koala, creías que no me acordaría, no puedo terminar sin darte las gracias por esas largas conversaciones, por tu locura retenida, por tu siempre pensar y pensar, por tus amenazas, tu madurez, y sobre todo por amenizar mi paso por el otro lado del mundo, veremos que depara googlemaps.

La vida siempre puede sonreírte, todo depende de cómo la miremos, en mi caso no puedo dejar de hacerlo a través de dos ñajos, mi presumida ahijada, Patricia y mi siempre sonriente enano, Nico, no os podéis imaginar lo que os quiero pitufos. Y como todos fuimos niños alguna vez yo también lo fui, es por ello debo agradeceros a vosotros hermanos, Pablo y Susana, me honra el haber compartido una infancia llena de sonrisas, peleas, juegos y demás venturas a vuestro lado. Y por supuesto a ti Pilar, mi segunda hermana, gracias por formar parte de nuestra familia y hacer feliz mi hermano. ¡Os quiero!.

Es evidente que todo principio tiene un final, a veces ese final no es como soñamos pero depende de nosotros levantarnos y continuar. Durante este periodo de mi vida han sido muchas las vivencias pero el recuerdo de aquellos quienes se fueron al final, con lágrima en ristre, es inevitable llegado a este punto. Mi padre ya no está y me gustaría dedicarle esta prosa, tantas enseñanzas, tantas virtudes adquiridas, tantas cosas que siquiera él puede imaginar que he aprendido. Gracias con una lágrima en los ojos, pero no de tristeza sino de felicidad. Al lado de esa humilde señora, mi madre, a la que otro tanto tengo que agradecer, conseguisteis si cabe que haya llegado a ser quién soy hoy día. A vosotros padres, por vuestra inmensa labor, mi más sincero agradecimiento.

Gracias.

Pedro Tomé González
Madrid, November 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

THIS PHD THESIS IS FOCUSED ON dealing with the variability factors in biometric recognition and applications of biometrics at a distance. In particular, this PhD Thesis explores the variability factors present in practical applications of biometrics at a distance, and then studies how soft biometrics information can help in such scenarios.

Nowadays, due to the expansion of the networked society, there is an increasing need for reliable personal identification by automatic means. Establishing the identity of individuals is recognized as fundamental not only in numerous governmental, legal or forensic operations, but also in a large number of civilian applications. This has resulted in the establishment of a new technological area known as biometric recognition, or simply *biometrics* [Jain *et al.*, 2006]. The basic aim of biometrics is to discriminate automatically between subjects in a reliable way and according to some target application based on one or more signals derived from physical or behavioral traits, such as face, fingerprint, iris, voice, hand, signature, etc. These personal traits are commonly denoted as *biometric modalities* or also as *biometrics*.

The difficulties associated with person identification and individualization were already highlighted by the pioneers of forensic sciences. Alphonse Bertillon developed in the eighteenth century an anthropometric identification approach, based on the measure of physical characteristics of a person [Bertillon, 1893]. Automatic person authentication has been a subject of study for more than thirty five years [Atal, 1976; Kanade, 1973], although it has not been until the last decade when biometric research has been established as an specific research area. This is evidenced by recent reference texts [Jain *et al.*, 2008, 2011b; Ratha and Govindaraju, 2008; Ross *et al.*, 2006; Tistareli *et al.*, 2009], specific conferences [Bowyer *et al.*, 2008a; Fierrez *et al.*, 2013; Lee and Li, 2007; Tistarelli and Maltoni, 2007; Vijaya-Kumar *et al.*, 2008], common benchmark tools and evaluations [Beveridge *et al.*, 2013; Phillips *et al.*, 2000a; Phillips, 2006; Phillips *et al.*, 2011, 2009a,b; Przybocki and Martin, 2004; Yeung *et al.*, 2004], cooperative international projects [BBfor2, 2010; BioSec, 2004; Biosecure, 2004; COST, 2007; MTIT, 2009; Tabula Rasa, 2010], international consortia dedicated specifically to biometric recognition [BC, 2005; BF, 2009; BI, 2009; EBF, 2009], standardization efforts [ANSI/NIST, 2009; BioAPI, 2002; ISO/IEC JTC 1/SC 27 , 2009; SC37, 2005], and increasing attention both from government

[BWG, 2009; DoD, 2005] and industry [IBIA, 2009; International Biometric Group, 2006].

This introductory chapter presents the basics of biometric systems, including properties, systems and biometric traits. We also outline the topic of variability factors in biometrics at a distance, from which the motivation of this Thesis is also derived. We finish the chapter by stating the Thesis, giving an outline of the Dissertation, and summarizing the research contributions originated from this work.

Although no special background is required for this chapter, the reader will benefit from introductory readings in biometrics such as Jain *et al.* [2008, 2006, 2004d]. A deeper reference is Jain *et al.* [2011b].

1.1. Biometric Systems

A biometric system is essentially a pattern recognition system that makes use of biometric traits to recognize individuals. The objective is to establish an identity based on *who you are or what you produce*, rather than by *what you possess* or *what you know*. This new paradigm not only provides enhanced security but also avoids, in authentication applications, the need to remember multiple passwords and maintain multiple authentication tokens. Who you are refers to *physiological* characteristics[1] such as face, iris or fingerprint. What you produce refers to *behavioral* patterns which entail a learning process and that characterize your identity such as the gait, voice or the written signature.

The digital representation of the characteristics or features of a biometric trait is known as *template*. Templates are stored in the system database through the *enrollment* or *training* process, which is depicted in Figure 1.1 (top). The database can either be centralized (this is the case of most biometric systems working at the moment), or distributed (as in Match-on-Card systems where each user carries the only copy of his template in a personal card [Bergman, 2008]). Once the users have been enrolled to the system, the recognition process can be performed in two modes [Jain *et al.*, 2011b]:

- **Identification**. In this mode, the question posed to the system is: is this person in the database?, the answer might be No (the person is unknown to the system), or any of the registered identities in the database. In order to give the answer the system has to perform a one-to-many matching process, as it has to compare the input biometric to all the stored templates (Fig. 1.1, center).

  In most practical cases, under the identification operation mode, the system usually returns, in a ranked manner, those identities that are more likely to be the searched person in a previously created database (i.e., those that have produced a higher similarity score), and then a human expert decides whether the subject is or not within that reduced group

[1]Although the term *physiological characteristic* is commonly used when describing biometrics, the purpose is to refer to the morphology of parts of the human body, therefore the proper term is *morphological characteristic*.

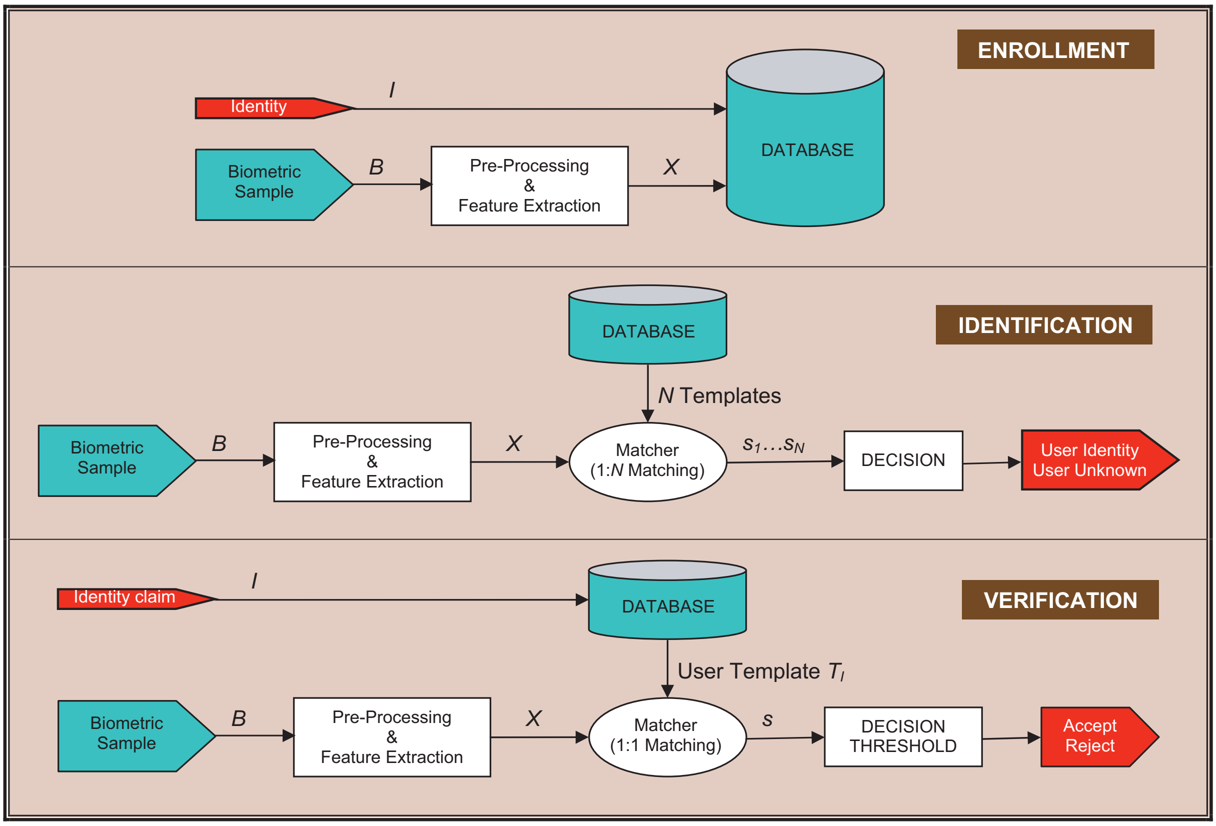


Figure 1.1: *Diagrams of the typical modes of operation in a biometric system.*

of people. Typical identification applications include Automated Fingerprint Identification Systems [Komarinski, 2005].

- **Verification**. In this case what we want to know if a person is really who she claims to be. This way, under the verification mode (Fig. 1.1, bottom), the system performs a one-to-one matching process where the submitted biometric trait is compared to the enrolled pattern associated with the claimed identity, in order to determine if the subject is a *client* (the identity claim is *accepted*), or an *impostor* (the identity claim is *rejected*). Typical verification applications include network logon, ATMs, physical access control, credit-card purchases, etc.

This Thesis is focused on the evaluation of biometric systems working under the verification mode (also known as *authentication*). In this mode, the *clients* or *targets* are known to the system (through the enrollment process), whereas the *impostors* can potentially be the world population. The result of the comparison between the feature vector X (extracted from the biometric sample B provided by the user) and the template T_I corresponding to his/her claimed identity I is a similarity score s which is compared to a decision threshold. If the score is higher than the decision threshold, then the claim is accepted (client), otherwise the claim is rejected (impostor).

1.1.1. Biometric Modalities

A number of different biometrics have been proposed and are used in various applications [Jain *et al.*, 2011b]. As mentioned before, biometric traits can be classified into *physiological* biometrics (also known as *anatomical* or *morphological*) which include images of the ear, face, hand geometry, iris, retina, palmprint or fingerprint, and *behavioral* biometrics including voice, written signature, gait or keystroking. This classification is just indicative, as some of the traits are not easy to categorize in any of the groups. The voice, for instance, is commonly accepted to be a behavioral biometric (as the voice is something that we *learn* to *produce*), however its distinctiveness largely depends on physiological characteristics (e.g., vocal tracts, mouth, nasal cavities, or lips). On the other hand, other very distinctive human feature, the DNA, is usually not considered a biometric modality as recognition systems based on it still require manual operation and cannot be used in (pseudo) real-time. Example images from various biometrics are given in Fig. 1.2.

In theory, any human characteristic can be used as a biometric identifier as long as it satisfies these requirements:

- **Universality**, which indicates to what extent a biometric is present in the world population.
- **Distinctiveness**, which means that two persons should have sufficiently different biometrics.
- **Permanence**, which indicates that the biometric should have a compact representation invariant over a sufficiently large period of time.
- **Collectability**, which refers to the easiness of the acquisition process and to the ability to measure the biometric quantitatively.

Other criteria required for practical applications include:

- **Performance**, which refers to the efficiency, accuracy, speed, robustness and resource requirements of particular implementations based on the biometric.
- **Acceptability**, which refers to which people are willing to use the biometric and in which terms.
- **Circumvention**, which reflects the difficulty to fool a system based on a given biometric by fraudulent methods.
- **Exception handling**, which has to do with the possibility to complete a manual matching process for those people that cannot interact in a normal way with the system (e.g., impossibility to perform the feature extraction process due to an excessive degradation of the trait).

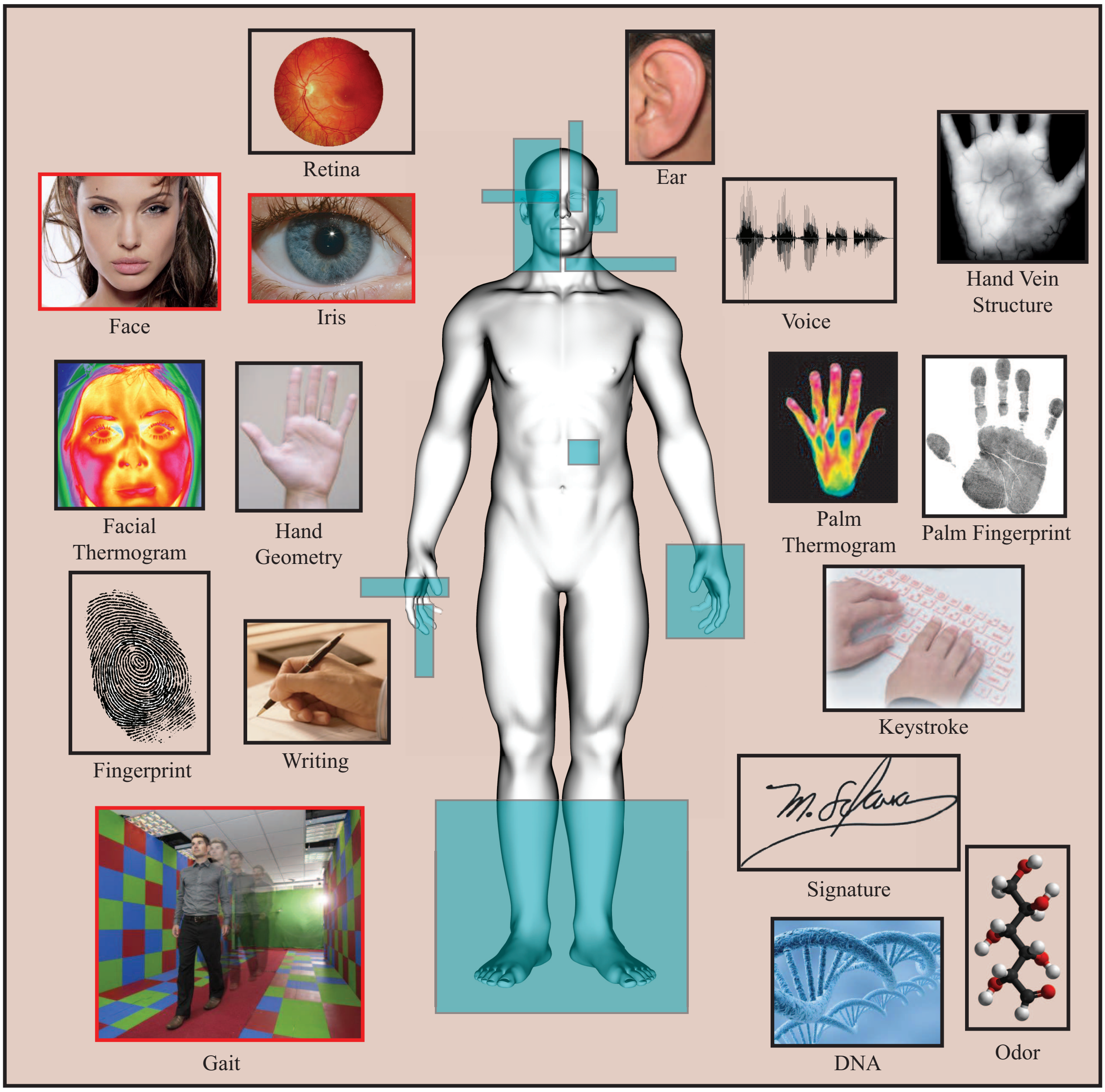


Figure 1.2: *Examples of common biometrics. Red bounding box indicates suitable biometrics at a distance.*

- **Cost**, which refers to all the costs that would be necessary to introduce the system in a real-world scenario.

An ideal biometric system should meet all these requirements; unfortunately, no single biometric trait satisfies all the above mentioned properties. While some biometrics have easy and friendly collectability (e.g. face or voice), their distinctiveness is low. Other biometrics with high distinctiveness are not easy to acquire (e.g. iris or fingerprint).

1.2. Variability Factors in Biometric Systems at a Distance

First of all, it is important to remember that absolute variability compensation in image and video processing does not exist: there are countless variability sources in uncontrolled and unconstrained systems at a distance. The objective of the research community is usually to develop applications in which the variability sources are under certain margins in order to guarantee a desired recognition accuracy.

In the next sections a number of variability related issues are discussed in order to clarify the perspective followed during the development of the Thesis, and to define our position within the complex field of variability research in biometrics at a distance.

1.2.1. Variability vs Distance

Nowadays, biometric devices use, between others, the face, the iris, and even the gait in order to recognize the identity of a person. These technologies are still far away to be mature systems and do not answer all the necessities of the wide number of potential applications. In particular, there is an increasing interest in acquiring biometric information in a non intrusive way such as with people on the move or at a distance.

Biometric recognition on the move or at a distance leads to the necessity to properly consider the scenario at hand in order to have the variability factors under certain margins. In this kind of heterogeneous scenarios, the selection of the best recognition strategy strongly depends on the scenario, therefore systems ideally should be able to automatically identify and classify each scenario by the different variability factors affecting in each case.

The concept of estimating the acquisition distance in order to define different scenarios has not been traditionally used in person recognition at a distance. This will be exploited afterwards in this Thesis.

Throughout the Dissertation different variability factors that may affect biometric systems at a distance are pointed out, systematically evaluated, and compensated based on different acquisition distances between the subject and the camera.

1.2.2. Soft Biometrics vs Variability

The first personal identification system developed by Bertillon [1896] for identification of criminals was based on three sets of features: *i*) body measurements (anthropometrics) like height and length of the arm, *ii*) morphological description of the appearance and shape of the body like eye color and anomalies of the fingers, and *iii*) peculiar marks observed on the body like moles, scars, and tattoos. Although the Bertillon system was very useful in tracking criminals, it had an unacceptably high rate of false identification. This was due to two reasons. Firstly, several individuals can have the same set of values for these measurements (*inter-user variability*). Secondly, for the same individual, these values can change over time (*intra-user variability*). In other words, these characteristics do not have the distinctiveness and permanence to uniquely identify an individual over a period of time and hence we refer them as **soft**

biometric traits. Soft biometric traits are those characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals.

Soft biometric traits can either be continuous or discrete. Traits such as gender, eye color, ethnicity, etc. are discrete in nature. On the other hand, traits like height and weight are continuous variables. In principle, a system that is completely based on soft biometric traits cannot provide enough accuracy in the recognition of individuals. However, soft biometric traits can be used to improve the performance of a traditional biometric system (e.g., face, gait, etc.) in many ways. One of these ways that will be explored in the Dissertation is the increased robustness that can be achieved in highly variable scenarios when properly integrating soft biometrics to primary traditional biometric systems. This way, throughout the Dissertation different soft biometric information that may be extracted from biometric systems at a distance are pointed out, systematically evaluated, and incorporated through adaptive fusion to person recognition systems working at a distance.

1.3. Motivation of the Thesis

Provided that the performance of a biometric system at a distance is heavily affected by the variability factors of multiple sources, this Thesis is focused on the identification and classification of variability factors in biometrics at a distance, and then presents methods to deal with them (soft biometrics and adaptive fusion). Note that we aim to be comprehensive in our study of variability factors, but a full report of variability compensation methods is out of the scope of the Thesis. Here we only provide two methods to deal with the challenging factors found in biometrics at a distance. The research carried out in this area has been mainly motivated by five observations from the state-of-the-art.

Automatic face recognition technology is still an open problem, particularly when working with video surveillance imagery. Such progress for face recognition is one of the goals of the FBI's Next Generation Identification program [Next Generation Identification]. Face recognition in video surveillance scenarios is a very challenging task due to the variability that can be present. In this sense, there are several studies Li and Jain [2011]; Tome *et al.* [2010b, 2012]; Zhang and Gao [2009] based on realistic scenarios trying to understand the effect of the different variability factors in this field. However, in most of those valuable research contributions, a complex question remains unanswered: *how variability factors affect the face systems at a distance?*.

The second observation is strongly related to the first one. In the existing publications in face recognition at a distance, experimental results are obtained and reported considering fixed and isolated variability sources. In practice the actual variability sources are multiple and unknown.

The third observation comes from the different initiatives that are currently trying to assist the development of face and person recognition algorithms [Beveridge *et al.*, 2013; Phillips *et al.*, 2011, 2009a]. These evaluations are designed by [National Institute of Standards and Technology (NIST)] to provide the research community and law enforcement agencies with information to assist them in determining where and how facial recognition technology can best be deployed. These initiatives are focusing their interest in the last years on biometrics systems at a distance.

The fourth observation that has motivated this Thesis is the constant need for high accuracy in person recognition applications (and in this particular case, in biometric systems at a distance), in order to make them reliable in challenge scenarios and motivate the industry.

The last observation is that the development of new variability compensation approaches for the studied biometric systems at a distance is currently a research challenge. Although different efforts have been carried out in this direction [Cardinaux *et al.*, 2006; Lucey and Chen, 2004; McCool and Marcel, 2009; McCool *et al.*, 2013; Sanderson and Lovell, 2009], there is still no definitive solution for some of the variability factors focused in important applications of biometrics such as surveillance and forensics.

1.4. The Thesis and Main Contributions

The Thesis developed in this Dissertation can be stated as follows:

> *The incorporation of soft biometrics information through adaptive fusion to person recognition systems working at a distance can provide significant benefits in these very challenging scenarios. In particular, the variability factors found in practical biometrics applications working on the move or at a distance can be compensated to some extent exploiting this idea.*

The approach we follow to develop this PhD Thesis is in two steps: *i*) understanding the variability factors associated with specific scenarios of practical importance (e.g. surveillance and forensics), and *ii*) proposing and studying new methods in soft biometrics and adaptive fusion.

The **main contributions** in these two steps are summarized as follow:

- *First step.* The research contributions are methodological based on how variability is studied in an unique way, we also provide experimental evidences, and finally we contribute with new biometric data made public for the research community.

- *Second step.* The contributions are new algorithms for soft biometrics and adaptive fusion supported by experimental results on realistic scenarios at a distance (in video surveillance and forensics).

1.5. Outline of the Dissertation

The main objectives of the PhD Thesis are as follows: 1) reviewing and studying the problem of variability factors associated with realistic scenarios at a distance in order to identify and evaluate the variability sources and the suitable biometrics; 2) devising practical compensation methods based on soft biometrics and adaptive fusion to deal with variability factors in order to enhance the robustness of biometric systems at a distance; and 3) applying the proposed techniques and methodologies to practical scenarios, systems, and databases widely available for the biometrics research community, with emphasis on face verification systems.

The Dissertation is structured according to a traditional complex type with background theory, practical methods, and three independent experimental studies in which the methods are applied [Paltridge, 2002]. The chapter structure is as follows:

- Chapter 1 introduces the topic of variability factors in biometric systems at a distance and gives the motivation, outline and contributions of this PhD Thesis.
- Chapter 2 summarizes related works which have motivated this Thesis.
- Chapter 3 introduces two novel methods proposed in the framework of this Thesis and that are later used in the experimental part of the Dissertation. These methods are: *i*) soft biometric information suitable for video surveillance and forensic applications, and *ii*) some adaptive fusion schemes using ancillary information and distance estimation (which presents the advantage over previously proposed schemes of using the distance to identify the scenario and apply the best solution).
- Chapter 4 considers the issue of performance evaluation in biometric systems and presents the methodology followed in the Dissertation for evaluation of biometric systems at a distance. The biometric databases used in this Dissertation are also introduced.
- Chapter 5 studies the variability in practical scenarios at a distance at different acquisition distances.
- Chapter 6 studies the variability and discrimination power of the soft biometric proposed in Chapter 3, in video surveillance and forensics applications.
- Chapter 7 studies the application of the adaptive score fusion schemes proposed in Chapter 3 to biometrics at a distance in different scenarios.
- Chapter 8 concludes the Dissertation summarizing the main results obtained and outlining future research lines.

The dependence among the chapters is illustrated in Fig. 1.3. For example, before reading any of the experimental Chapters 5, 6 and 7 (shaded in Fig. 1.3), one should read first Chapters 4 and 3. Before Chapter 4 one should start with the introduction in Chapter 1, and the recommendation of reading Chapter 2. Following the guidelines given in Fig. 1.3 and assuming a background in biometrics [Jain *et al.*, 2011b], one should read the experimental Chapter 5 before the Chapters 6 and 7.

The methods developed in this PhD Thesis are strongly based on popular approaches from the pattern recognition literature. The reader is referred to standard texts for a background on

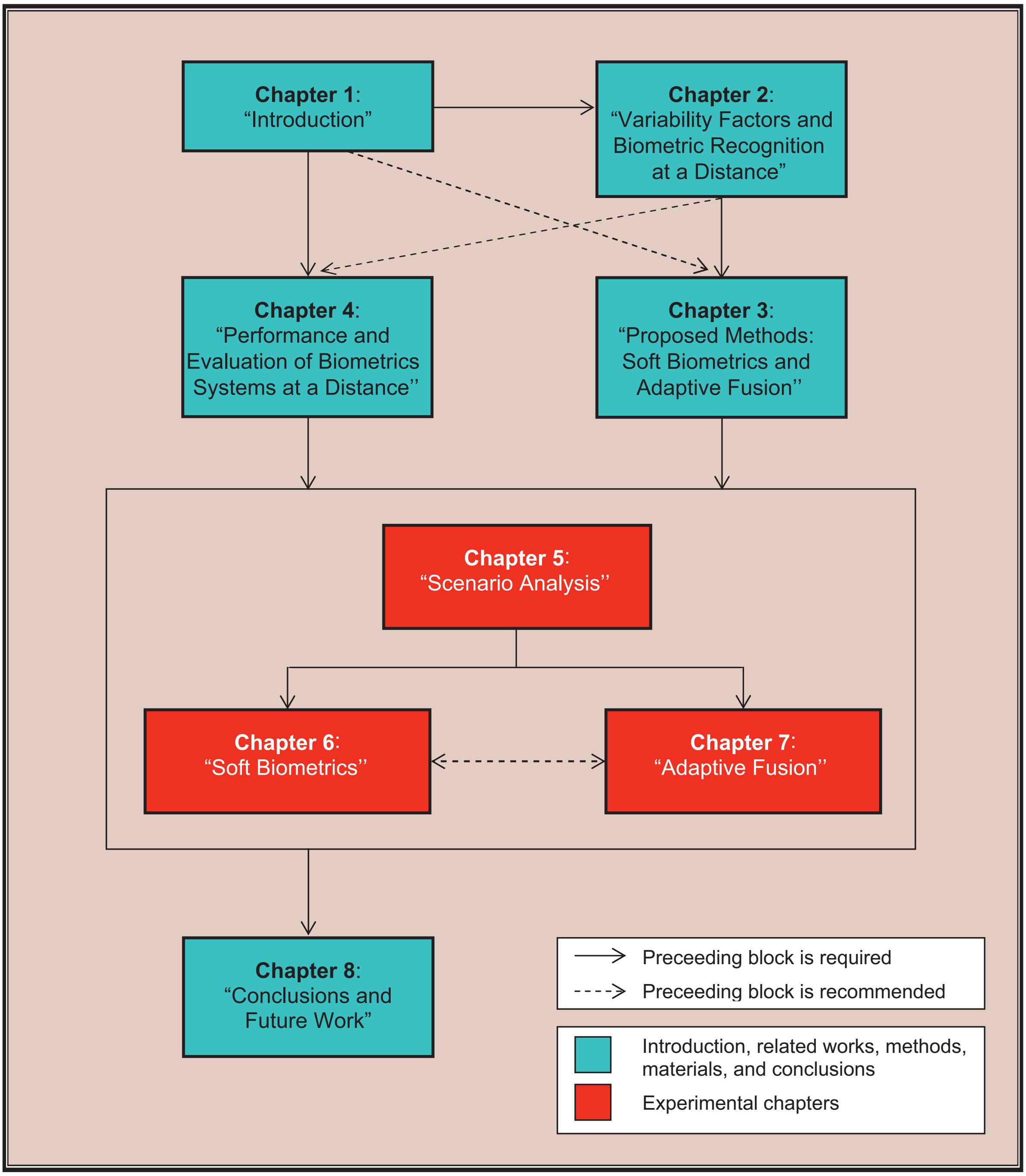


Figure 1.3: *Dependence among Dissertation chapters.*

the topic [Duda *et al.*, 2001; Theodoridis and Koutroumbas, 2008]. This is especially useful for dealing with Chapter 3. Chapters 3 and 5 assume a knowledge of the fundamentals of image processing [Gonzalez and Woods, 2006], and pattern recognition [Bigun, 2006].

1.6. Detailed Research Contributions

The research contributions of this PhD Thesis are the following (some publications appear in several items of the list):

- NOVEL METHODS.

  1. Novel methods for incorporating soft biometrics information to biometric systems at a distance.

     - P. Tome, J. Fierrez, R. Vera-Rodriguez and D. Ramos. "Identification using Face Regions: Application and Assessment in Forensic Scenarios", *Forensic Science International (FSI)*, n. 233, pages 75 - 83, 2013e.
     - P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia. "Scenario-based score fusion for face recognition at a distance", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 67 - 73, June 2010a.

  2. Novel methods for extracting and analysing facial regions in biometric systems at a distance.

     - P. Tome, L. Blazquez, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, N. Exposito, and P. Leston. "Understanding the discrimination power of facial regions in forensic casework", in *Proc. International Workshop on Biometrics and Forensics*, pages 1 - 4, Lisboa, Portugal, April 2013a.
     - P. Tome, J. Fierrez, R. Vera-Rodriguez and D. Ramos. "Identification using Face Regions: Application and Assessment in Forensic Scenarios", *Forensic Science International (FSI)*, n. 233, pages 75 - 83, 2013e.

  3. Novel methods for combining different facial regions for face verification.

     - P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia. "Combination of face regions in forensic scenarios", *Rapid Communications in Forensic Science International (FSI)*, 2013c. submitted.
     - P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. "Fusion of facial regions using color information in a forensic scenario", in *Proc. 18th Iberoamerican Congress on Pattern Recognition (CIARP2013)*, LNCS 8259, pages 399 - 406, Habana, Cuba, November 2013f.

- NEW BIOMETRIC SYSTEMS.

  1. A new face verification system based on Sparse Representation.

     - P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. "Variability compensation using NAP for unconstrained face recognition", in *Proc. 10th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS12)*, volume 151, pages 129 - 139. Springer, March 2012.

2. A new Soft biometric verification system.

• P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon. “Soft biometrics and their application in person recognition at a distance”, *IEEE Transaction on Information Forensics and Security*, 2013b. submitted. • P. Tome, J. Fierrez, R. Vera-Rodriguez, and Javier Ortega-Garcia. “Facial soft biometrics features for forensic face verification”, *Pattern Recognition*, 2013d. submitted.

- NEW BIOMETRIC DATA.

1. A mugshot faces biometric database (ATVS Forensic DB) including three different views from 50 subjects acquired in the framework of this PhD Thesis.

• R. Vera-Rodriguez, P. Tome, J. Fierrez, N. Exposito, and F. J. Vega. “Analysis of the variability of facial landmarks in a forensic scenario”, in *Proc. International Workshop on Biometrics and Forensics*, pages 1 - 4, Lisboa, Portugal, April 2013a.

2. A set of manually labelled facial landmarks from the ATVS Forensic DB including 21 facial landmarks of 50 subjects with 8 frontal face images each.

• R. Vera-Rodriguez, P. Tome, J. Fierrez, N. Exposito, and F. J. Vega. “Analysis of the variability of facial landmarks in a forensic scenario”, in *Proc. International Workshop on Biometrics and Forensics*, pages 1 - 4, Lisboa, Portugal, April 2013a.

3. A set of manually labelled facial landmarks from SCface DB and the MORPH DB including 21 facial landmarks of a subset of 130 subjects with 6 frontal face images each.

• P. Tome, J. Fierrez, R. Vera-Rodriguez and D. Ramos. “Identification using Face Regions: Application and Assessment in Forensic Scenarios”, *Forensic Science International (FSI)*, n. 233, pages 75 - 83, 2013e.

4. A large sets of facial soft biometric features from the ATVS Forensic DB and the MORPH DB generated in the framework of this PhD Thesis, developed jointly with Binetskaya [2013].

• P. Tome, J. Fierrez, R. Vera-Rodriguez, and Javier Ortega-Garcia. “Facial soft biometrics features for forensic face verification”, *Pattern Recognition*, 2013d. submitted.

- NEW EXPERIMENTAL STUDIES

1. Acquisition scenario analysis for face recognition at a distance

• P. Tome, J. Fierrez, M. C. Fairhurst, and J. Ortega-Garcia. “Acquisition scenario analysis for face recognition at a distance”, in *Proc. 6th International Symposium on Visual Computing (ISVC)*, LNCS 6453, pages 461 - 468, November-December 2010b. • R. Vera-Rodriguez, J. Fierrez, P. Tome and J. Ortega-Garcia, “Face Recognition at a Distance: Scenario Analysis and Applications”, in *Proceedings 7th International Symposium on Distributed Computing and Artificial Intelligence*, Springer Advances in Intelligent and Soft Computing-79, pages 341 - 348, 2010.

2. Incorporating acquisition distance information to improve face recognition systems at a distance.

- P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia. "Scenario-based score fusion for face recognition at a distance", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 67 - 73, June 2010a.

3. Analysis of the variability of facial landmarks.

- P. Tome, J. Fierrez, R. Vera-Rodriguez and D. Ramos. "Identification using Face Regions: Application and Assessment in Forensic Scenarios", *Forensic Science International (FSI)*, n. 233, pages 75 - 83, 2013e.
- R. Vera-Rodriguez, P. Tome, J. Fierrez, N. Exposito, and F. J. Vega. "Analysis of the variability of facial landmarks in a forensic scenario", in *Proc. International Workshop on Biometrics and Forensics*, Lisboa, Portugal, April 2013a.
- R. Vera-Rodriguez, P. Tome, J. Fierrez, and J. Ortega-Garcia. "Comparative analysis of the variability of facial landmarks for forensics using cctv images", in *Proc. 6th Pacific Rim Symposium on Image and Video Technology (PSIVT)*, 2013b.

4. Variability compensation using Nuisance Attribute Projection for face verification.

- P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. "Variability compensation using NAP for unconstrained face recognition", in *Proc. 10th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS12)*, volume 151, pages 129 - 139. Springer, March 2012.

5. Study and evaluation of the discrimination power of different facial regions for face verification.

- T. Ali, P. Tome, J. Fierrez, R. Vera-Rodriguez, L. Spreeuwers and R. Veldhuis, "A study of identification performance of facial regions from CCTV images", in *Proc. 5th International Workshop on Computational Forensics (IWCF2012)*, Tsukuba, Japan, November 2012.
- P. Tome, L. Blazquez, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, N. Exposito, and P. Leston. "Understanding the discrimination power of facial regions in forensic casework", in *Proc. International Workshop on Biometrics and Forensics*, pages 1 - 4, Lisboa, Portugal, April 2013a.
- P. Tome, J. Fierrez, R. Vera-Rodriguez and D. Ramos. "Identification using Face Regions: Application and Assessment in Forensic Scenarios", *Forensic Science International (FSI)*, n. 233, pages 75 - 83, 2013e.

6. Evaluation of the performance of soft biometrics for video surveillance systems in terms of the proposed adaptive fusion aimed to improve face recognition performance.

- P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon. "Soft biometrics and their application in person recognition at a distance", *IEEE Transaction on Information Forensics and Security*, 2013b. submitted.

7. Study of combination of different facial regions for face verification.

- P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia. "Combination of face regions in forensic scenarios", *Rapid Communications in Forensic Science International (FSI)*, 2013c. submitted.

8. Study of combination of color information with different facial regions for face verification.

- P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. "Fusion of facial regions using color information in a forensic scenario", in *Proc. 18th Iberoamerican Congress on Pattern Recognition (CIARP2013)*, LNCS 8259, pages 399 - 406, Habana, Cuba, November 2013f.

Other contributions so far related to the problem developed in this Thesis but not presented in this Dissertation include:

- NEW BIOMETRIC SYSTEMS.

  1. An iris verification system based on Gabor features.

     - [Tome, 2008]

  2. An iris verification system based on SIFT features, developed jointly with Alonso-Fernandez *et al.* [2009].

     - F. Alonso-Fernandez, P. Tome, V. Ruiz-Albacete and J. Ortega-Garcia, "Iris Recognition Based on SIFT Features", in *IEEE Proc. International Conference on Biometrics, Identity and Security, BIDS*, September 2009.

  3. An automatic facial landmarks error detector, developed jointly with Blazquez [2012].

     - [Blazquez, 2012]

  4. A face verification system based on LBP and PCA features, developed jointly with Eslava-Rios [2013].

     - [Eslava-Rios, 2013]

  5. A forensic face verification system based on morphological features, developed jointly with Binetskaya [2013].

     - [Binetskaya, 2013]

- NEW BIOMETRIC DATA.

  1. A new database (BIOGIGA) composed of simulated images of 50 people at 94GHz (within the millimeter wave band - MMW).

     - [Moreno-Moreno *et al.*, 2011]

- NEW BIOMETRIC APPLICATIONS.

  1. Application of face verification in real time together with an interactive interface, developed jointly with Eslava-Rios [2013].

     - [Eslava-Rios, 2013]

- NEW EXPERIMENTAL STUDIES.

1. A study of a scenario for biometric recognition at a distance adapted to acquisition of MMW images.

    - [Moreno-Moreno *et al.*, 2010]

2. A study of different approaches for human action recognition in real videos, developed jointly with Herranz [2010].

    - [Herranz, 2010]

3. A study of different face and iris detectors for face verification, developed jointly with Dragolici [2010].

    - [Dragolici, 2010]

4. Study of effects of time variability in iris recognition.

    - P. Tome, F. Alonso-Fernandez and J. Ortega-Garcia. "On the Effects of Time Variability in Iris Recognition", in *Proc. IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1 - 6, Washington DC, USA, September 2008.

5. Analysis of direct attacks to iris verification systems using high quality printed images and software attack based on generic algorithm to iris recognition systems.

    - V. Ruiz-Albacete, P. Tome, F. Alonso-Fernandez, J. Galbally, J. Fierrez and J. Ortega-Garcia. "Direct attacks using fake images in iris verification", in *Proc. COST 2101 Workshop on Biometrics and Identity Management, BIOID*, Springer LNCS 5372, pages 181 - 190, Roskilde, Denmark, May 2008.
    - M. Gomez-Barrero, J. Galbally, P. Tome and J. Fierrez, "On the Vulnerability of Iris-based Systems to a Software Attack based on a Genetic Algorithm", in *Proc. Iberoamerican Conference on Pattern Recognition, CIARP*, Springer LNCS 7441, pages 114 - 121, Buenos Aires, Argentina, September 2012.

6. Comparative evaluation of gait recognition systems on lower part of the human body and with limited data information.

    - S. Gabriel-Sanz, R. Vera-Rodriguez, P. Tome and J. Fierrez, "Assessment of gait recognition based on the lower part of the human body", in *Proc. International Workshop on Biometrics and Forensics (IWBF'13)*, pages 1 - 4, 2013.
    - R. Vera-Rodriguez, S. Gabriel-Sanz, J. Fierrez, P. Tome and a. J. Ortega-Garcia, "Analysis of Gait Recognition on Constrained Scenarios with Limited Data Information", in *Proc. 11th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS-13)*, Springer Berlin Heidelberg Communications in Computer and Information Science-365, pages 231 - 239, May 2013.

7. Combination of footsteps and face recognition systems on uncontrolled environment.

    - R. Vera-Rodriguez, P. Tome, J. Fierrez and J. Ortega-Garcia, "Fusion of Footsteps and Face Biometrics on an Unsupervised and Uncontrolled Environment", in *Proc. International Conferencie SPIE Defense Security and Sensing, Biometric Technology for Human Identification IX (SPIE'12)*, Baltimore, USA, April 2012.

Chapter 2

Variability Factors and Biometric Recognition at a Distance

Biometric variability measurement is an operationally important step in systems at a distance that is nevertheless under-researched in comparison to the primary feature extraction and pattern recognition task. Recently, variability factors measurement has emerged in the biometric community as an important concern after the poor performance observed in biometric systems at a distance. There are a number of variability factors that can affect the performance of biometric systems.

Independent evaluations of commercial and research biometric systems conducted during the last decade included in each edition new scenarios and conditions that are progressively more difficult in nature. We can observe that, in many cases, this has resulted in a performance worsening, and it is not until subsequent editions that the algorithms show progress under the new challenging conditions. For instance, in the 1996, 2002 and 2006 editions of NIST: Face Challenges, the face samples used were acquired in controlled scenarios, resulting in an increment of verification rates (see Fig. 2.1). However, in the 2010 edition, these challenges started to change, face samples were intentionally corrupted or acquired in uncontrolled conditions, focusing towards biometrics at a distance scenarios. The result was that the verification rates of the best systems are much worse (an order of magnitude) than those of previous editions, although the technology improvement for acquisition sensors. The last editions of these challenges are totally focused on person recognition at a distance and the study of the variability problem which degrades the system performance. This result shows the significant impact that the variability factors can have on the recognition performance, and highlights the importance of measuring and dealing with them in biometric systems at a distance.

There are at least two reasons for this trend; the first is the wide range of commercial and law enforcement applications in these challenging scenarios and the second is the availability of feasible technologies after decades of research. In addition, the problem of machine recognition of human faces continues to attract researchers from disciplines such as image processing, pattern

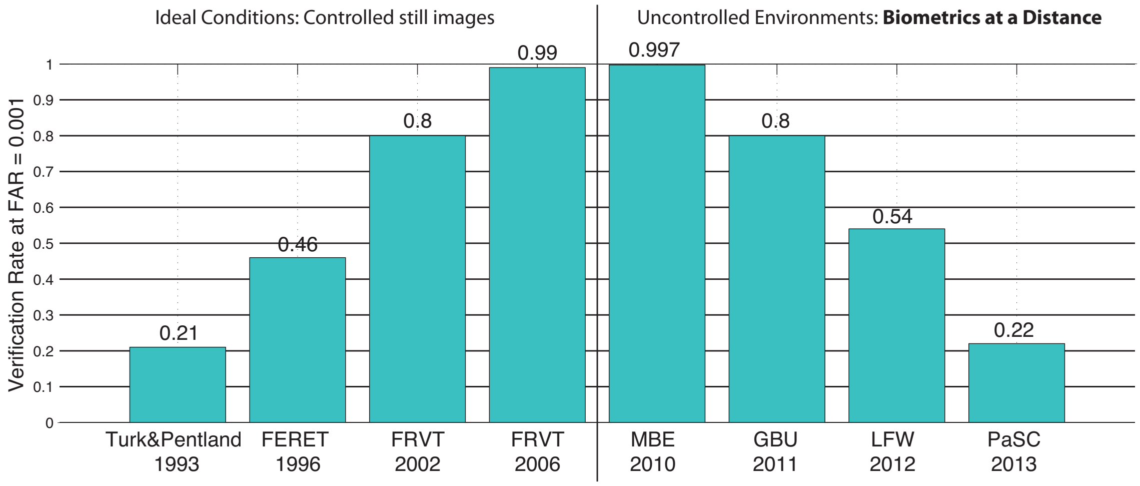


Figure 2.1: *Results in terms of Verification Rate (VR) at FAR = 0.001 of the best performing algorithm in all the databases of the NIST competitions based on face recognition [NIST: Face Challenges]. Performance progressively drops when shifting from controlled scenarios to uncontrolled conditions.*

recognition, neural networks, computer vision, computer graphics, and psychology.

This chapter summarizes the state-of-the-art in the biometric variability assessment problem in systems at a distance, giving an overall framework of the different factors related to it. It is structured as follows. We first define what variability is from the point of view of biometric systems at a distance. Next, we present the factors influencing biometric variability and the possible variability sources of acquired biometric samples at a distance. Next, we define what we consider to be a biometric system At a Distance (AD). The relationship between subjects and the acquisition distance, as well as the role of the camera to subject distance within biometric systems at a distance is then analysed.

Original contributions in this chapter include a taxonomy of variability factors affecting biometric performance, and a taxonomy of roles of the acquisition distance in the context of biometric systems at a distance.

2.1. Definition of Biometric Variability Factors

Broadly speaking, a variability factor is anything that degrades the similarity between a biometric sample and its source in terms of people recognition. Using the standard [ISO/IEC 29794-1, 2006], based on biometric quality, the variability factors in systems at a distance can be considered from two different points of view, see Fig. 2.2: *i*) *stability*, which refers to the intra-variability attributable to inherent biometric sample features of the subject; and *ii*) *degradation*, which is the degree of variability between a biometric sample and its source, attributable to each step through which the sample is processed. The *stability* of the sample source and the *degradation* of the processed sample contribute to, or similarly detract from, the *utility* of the sample, which is the impact of the individual biometric sample on the overall performance of a biometric system.

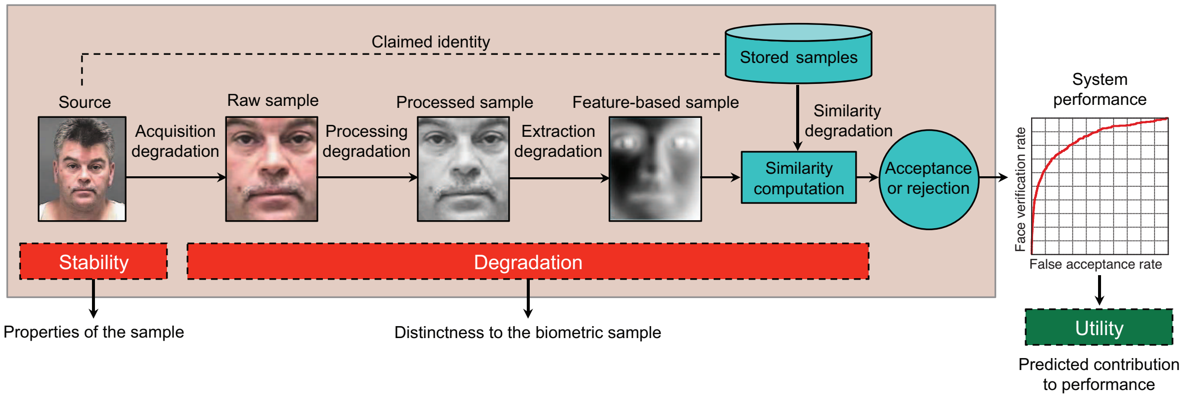


Figure 2.2: *Defining biometric variability factors from two different points of view: stability and degradation. The stability and degradation contribute to or detract from the sample's utility.*

It is generally accepted that a compensation of variability factors should most importantly be focused in maximizing the *utility* of the sample, so that samples with high distinctiveness and reduced variability lead to better identification of individuals. An adequate variability compensation technique will be largely dependent on the type of variability considered in each scenario.

2.2. Sources and Classification of Biometric Variability Factors

There are a number of factors affecting the variability of biometric signals in systems at a distance. Unfortunately, most of them cannot be controlled. We summarize in Fig. 2.3 the different variability factors that can affect the performance of biometric systems. They are classified depending on their relationship with the different parts of the system. We can distinguish four different classes: factors related entirely to the user, factors that have to do with the user-sensor interaction process, factors related to the acquisition device, and factors related with the processing system.

As can be seen in Fig. 2.3, the user-related factors affect the *stability* of the biometric sample, that is, the intra-variability attributable to the inherent sample features. In this sense, the control we have on these factors is low, as the inherent features of a person are difficult or impossible to modify. The remaining factors affect the *degradation*, or in other words, the difference between a biometric sample and its source. These factors can be better controlled than user-related factors.

2.2.1. User Factors

The user-related factors are classified as ***physiological*** and ***behavioral*** factors. As they have to do entirely with the "user side", they are the most difficult to control. We give a summary of the most important ones in Fig. 2.3 (top-right). Notice that most *physiological* factors cannot be controlled (e.g. age, gender, race, etc.) A number of them do not necessarily

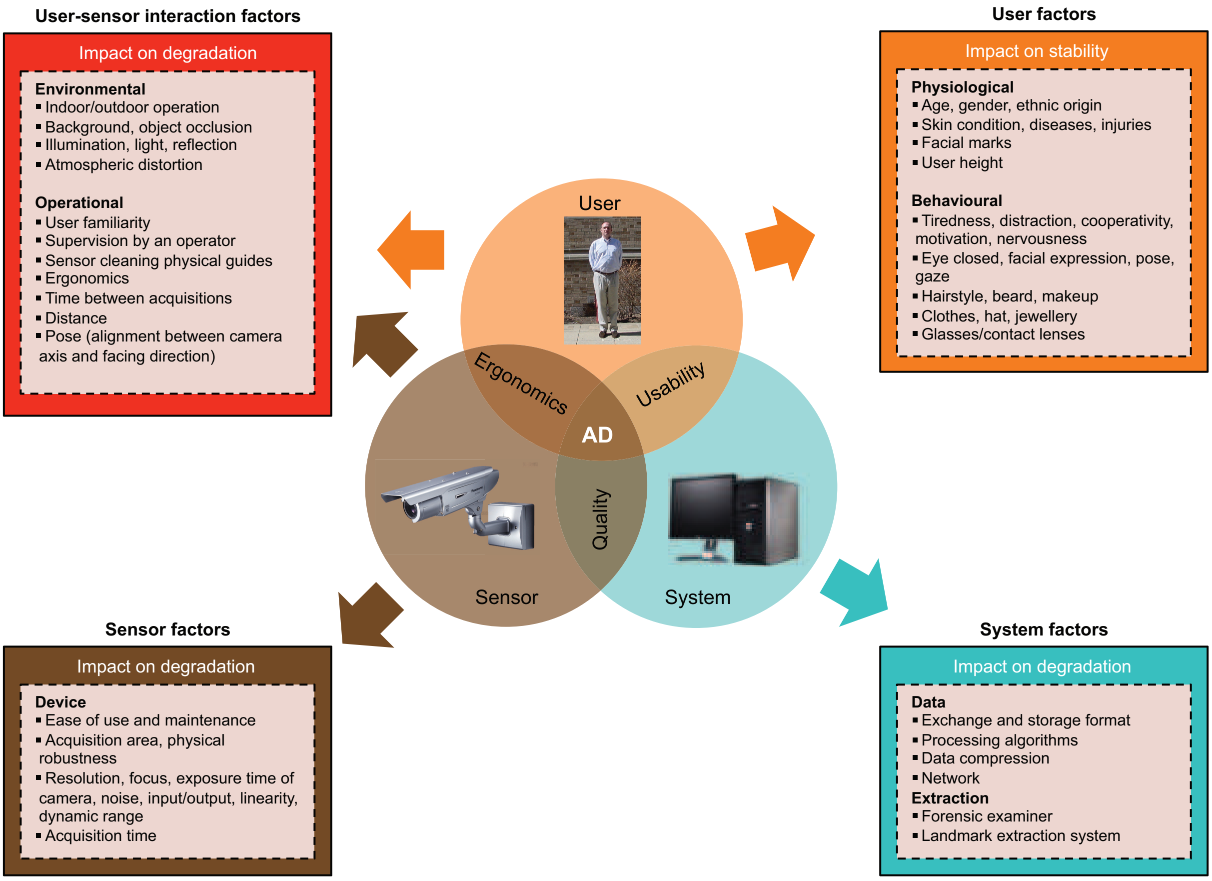


Figure 2.3: *Variability sources that can affect the biometric performance of systems at a distance (AD).*

produce degradation on the biometric data, but additional biometric intra-variability (e.g. face or speech characteristics are different in males and females, faces change as we grow up, etc.). These additional variability factors, if not properly considered by the recognition algorithm, may lead to degraded performance. Other factors, like diseases or injuries, may alter a part of our body, our skin, our voice, our ability to walk, etc., resulting in invalid data. In some cases, the alteration may be irreversible, making the affected biometric trait infeasible for recognition. On the contrary, *behavioral* factors are easier to control than *physiological* ones, although it is not always possible or convenient, as we would have to *modify* the people's behavior or habits.

The acquisition process is usually uncontrolled in systems at a distance, hence people on the move their biometric data. Also, the people may be tired, distracted or nervous. Note that when dealing with some user factors, one solution is just to recapture after taking corrective actions (e.g. "put off your hat/coat/ring/glasses" or "keep a frontal pose"), but this is not always possible or appropriate.

2.2.2. User-Sensor Interaction Factors

There are two types of factors related to the interaction between the user and sensor: ***environmental*** and ***operational***, which we summarize in Fig. 2.3 (top-left). In principle, they are easier to control than user-related factors, although users still play a part in them. For instance, impact of *environmental* factors will be low if we can control the environment. The variability of face and gait images or videos depends on illumination, background, object occlusion, etc., and also face images are affected by modifications of the properties of the skin and the reflections. The illumination and light reflections have great impact on iris images at a distance due to the reflective properties of the eye, whereas the variability of face and gait are highly dependent of the subject pose. Outdoor operation is specially problematic, as we can lose control of many factors affecting not only the biometric trait but also the sensor itself: temperature, humidity, weather, noise, illumination, etc. Outdoor operation demands additional actions to us regarding sensor conditions and its maintenance. Unfortunately, in certain applications, we cannot control the environment, as in the case of modern applications that make use of handheld devices with acquisition capabilities of biometric samples (e.g. webcams, laptops, smartphones, etc.)

As in the case of *environmental* factors, *operational* ones (Fig. 2.3 (top-left)) can also be controlled to some extent. Again, if the acquisition is not done physically in our premises, we will not be able to provide help or supervision to the user, we will not know if the sensor is cleaned periodically, or we will not be able to guarantee the ergonomics of the acquisition setup. An important factor that has to do with the operation of the system is the time passed between acquisitions, also known as ageing. There is an intrinsic variability in biometric data characteristics as time passes, not only in the long-term (e.g. changes of our face, voice, etc. or differences in the way we interact with the system) but also in the short-term (e.g. clothes, temporary diseases). The most important consequence is that biometric data acquired from an individual at two different moments may be very different. This affects any biometric trait, although some of them are more sensitive than others [Jain *et al.*, 2011b], as it is the case of gait and face at a distance. Another *operational* factor that we should consider is if the user receives feedback of the acquired data via display or similar, which leads to better acquired samples. But these kind of factors are discarded in systems at a distance where the subject move freely and the acquisition is totally uncontrolled.

2.2.3. Sensor Factors

Although the acquisition sensor is physically a part of the biometric system, it is the only point of interaction with users and even in some cases, people interact with the system using their own devices (e.g. mobile telephones). For these reasons, a number of sensor factors can affect the variability of acquired biometric data: its ease of use and maintenance, the size of its acquisition area, the resolution or the acquisition noise, its reliability and physical robustness, its dynamic range or the time it needs to acquire a sample. It is important that these factors are compliant with existing standards, so we will be able to replace the sensor if needed without degrading the

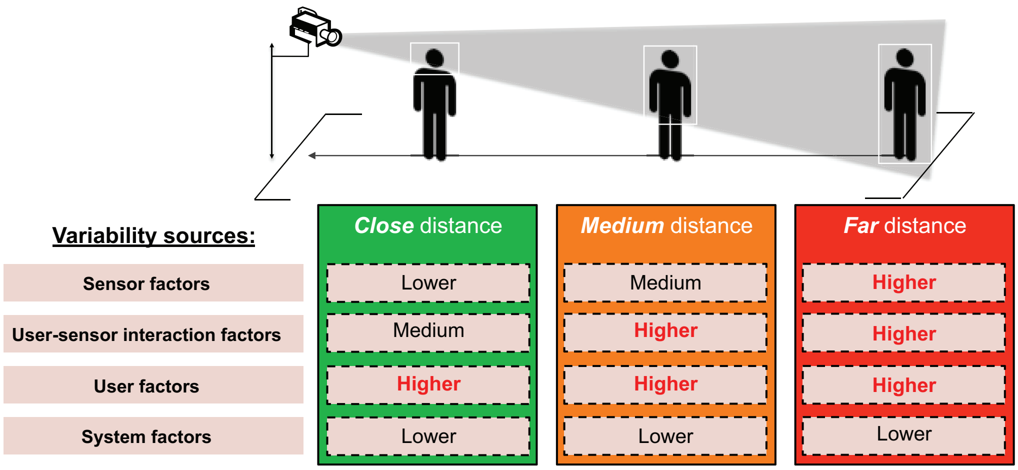


Figure 2.4: Classification of variability factors depending on the acquisition distance, together with their impact in degrading the system performance.

reliability of the acquisition process. This is specially important, because replacing the sensor is very common in operational situations when it is damaged or newer designs appear. Standards compliance also guarantees that we can use different sensors to interact with the system.

2.2.4. System Factors

Here we find the factors that are easiest to control, which are related to how we process a biometric sample once it has been acquired by the sensor. Factors affecting here are the data format we use for exchange or storage and the algorithms we apply for data processing. If there are storage or exchange speed constraints, we may need to use data compression techniques, which may degrade the sample or even introduce variability in the template.

2.2.5. Graduation of Variability Factors in Systems At a Distance

As previous section explained there are a number of uncontrolled variability factors affecting the biometric signals in systems at a distance. These variability factors vary with the acquisition distance, therefore an important aspect in a biometric system at a distance is to *understand the scenario*, i.e., *i*) how the variability factors are affecting the system depending of the acquisition distance, and *ii*) delimiting their range of variability. Fig. 2.4 shows a classification of the degradation degree produced by the main variability sources in the different acquisition distances. The sensor-related factors have high impact on the system at far distances compared to close distances (a low resolution webcam is enough at close distance). On the other hand the user-related factors (such as height, gender, ethnicity, etc.) have a high impact on the system in all the distances.

Additionally, each variability factor summarized in Fig. 2.3 has an individual graduation of degradation depending the acquisition distance between *slight* to *severe*, e.g., illumination outdoors affect in a slight level if the object is close to the camera and in a severe level if the

Figure 2.5: *Face Recognition At a Distance (FRAD) example on an high quality image from a real scenario extracted from http://avigilon.com.*

object is far away from the camera. Therefore, measurement or approximation of the acquisition distance based on the object of interest in the scene is useful to identify the variability factors present in the scene.

2.3. Definition of Systems AD (At a Distance)

There is no formal closed definition of biometric system at a distance in the literature. As previously explained the variability factors vary with the acquisition distance between the user and the camera. In terms of this distance from user to the camera, biometric recognition systems can be categorized into *close* distance (often used in cooperative applications), *medium* distance, and *far* distance. In this PhD Thesis, we consider biometric systems at a distance as systems where a biometric signal is captured at a distance in a controlled or uncontrolled environment, with or without user cooperation, and influenced by several known or unknown variability factors.

The most common biometric trait visually available considered in recognition systems at a distance is the human face. It is both visible and readily imaged from a distance. For security or covert applications, facial imaging can be achieved without the knowledge of the subject. Fig. 2.5 shows an example of a biometric system at a distance based on face recognition (referred to also in the literature as FRAD) where some variability factors affecting in terms of acquisition distance from user to the camera can be seen. There is also great interest in iris at a distance [Matey *et al.*, 2006], however it is doubtful that iris will outperform face with a comparable system complexity and cost. Gait information can also be acquired over large distances, but crowded places and multi-person recognition make the face a more discriminating identifier.

In real scenarios at a distance, subjects may be sparsely distributed and standing or walking along predictable trajectories (e.g., airport passport control, building access, etc.), or they may

Areas	Specific applications
Entertainment	Video game, virtual reality, training programs Human-robot-interaction, human-computer-interaction
Smart cards	Drivers' licenses, entitlement programs Immigration, national ID, passports, voter registration Welfare fraud
Information security	TV Parental control, personal device login, desktop login, event login Application security, database security, file encryption Intranet security, internet access, medical records Secure trading terminals, marketing
Law enforcement and surveillance	Watch-list, white-list, on-line recognition Advanced video surveillance, CCTV control Portal control, post-event analysis, access control Shoplifting, suspect tracking and investigation

Table 2.1: *The most important applications of recognition systems at a distance.*

be in a crowd, moving in a chaotic manner, and occluding each other (e.g., surveillance cameras). Therefore, the nature of the activity of subjects and the size of the surveillance area can vary considerably with the scenario of application and this is closely related with the degree of difficulty and the variability factors affecting to the system.

The two main difficulties faced by systems at a distance are: *i*) *acquisition* of biometric signals (images) from a distance, and *ii*) *recognition* of the person in spite of imperfections and variability factors in the captured data.

There are a lot of advantages/disadvantages that biometric systems at a distance have and a lot of applications as we can see in Table 2.1, which lists some of the most important applications.

2.3.1. Acquisition Distance Levels

The acquisition distance is a very important factor to be studied in biometric systems at a distance. The concept of estimating the acquisition distance in order to define different scenarios has not been traditionally used in person recognition at a distance. It is important to emphasize that in biometrics at a distance this scenario estimation based on the acquisition distance is very useful, as different scenarios will usually have different variability factors and thus they may be processed differently.

In order to develop this idea, we first propose a classification of acquisition distances, which will help to study the variability factors and their degree of influence in systems at a distance. The acquisition distances of biometric systems at a distance can be categorized into three levels summarized in Fig. 2.6 with an example based on the human face.

- *Level 1 distances* (real distances), which are the actual real distances between camera and subject. The acquisition setup is configured (resolution, focus, etc.) for working at a fixed distance. This is considered a controlled acquisition where variability factors generated by the capture sensor are minimized. The camera configuration is assumed to be correct for the scene at hand and therefore the captured image has correct resolution, focus, and

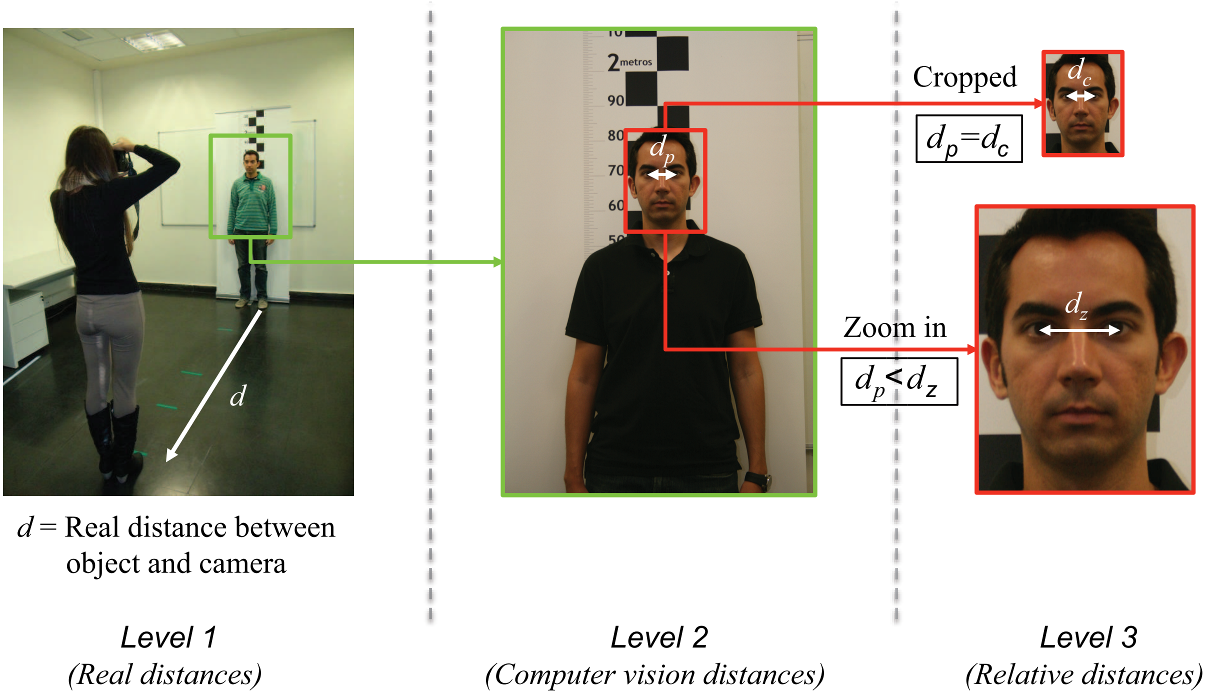


Figure 2.6: *Distance levels example for systems at a distance based on human face.* d_p*,* d_c*, and* d_z*, represent the Interpupillarity Pixel Distance (IPD) for original, cropped, and zoomed image, respectively.*

other factors attributable to the acquisition sensor, but can be influenced by the user-sensor interaction variability such as the environmental factors (e.g., lighting in outdoor scenarios).

Level 1 distances are commonly used to design acquisition scenarios such as border control, restricted walking zone, static capture, building access, etc.

- *Level 2 distances* (computer vision distances), which are distances between camera to object calculated using camera calibration approaches. Camera calibration is a common step in computer vision. Although some information from a scene can be obtained by using uncalibrated cameras, calibration is essential when metric information is required. The use of precisely calibrated cameras makes the measurement of distances in the real world from their projections on the image plane possible [Dang *et al.*, 2009]. These distances can be estimated using camera parameters. There are two kinds of parameters to be considered for the calibration: *i)* intrinsic parameter set, which models the internal geometry and optical characteristics of the image sensor, and *ii)* extrinsic parameters that measure the position and orientation of the camera with respect to a world coordinate system.

  Level 2 distances are mainly used in computer vision when metric information from the scene is required.

- *Level 3 distances* (relative distances) are estimated or relative distances, which are measured with respect to the image plane. In this level, we do not have access to information

Figure 2.7: *Examples of sensors and scenarios of recognition at a distance in the real life.*

about the camera configuration, therefore we take the image plane as a reference. Fig. 2.6 shows an example, where the videos or images are post-processed (cropped, zoom in, ...) or modified in some way, hence the configuration metadata is lost.

One example Level 3 distance is the segmented face area with respect to the full image area. As we will see in the next chapters, such a simple computation is strongly correlated with the actual acquisition distance, and therefore it will be very useful.

2.4. Biometric Traits At a Distance

The biometric data sources in scenarios at a distance are usually based on video surveillance cameras [Li and Jain, 2011; Tistareli *et al.*, 2009], also called Closed-Circuit Television Video (CCTV) cameras, which can produce images or recordings for surveillance purposes, and can be either video cameras, or digital stills cameras. Fig. 2.7 shows some examples of sensor technology and real scenarios at a distance.

As this Fig. 2.7 shown the demand for human identification at a distance has gained considerable attraction, particularly due to the need for covertly recognizing individuals in unconstrained environments with uncooperative subjects. In such environments, the person of interest may not be interacting with the biometric system in a concerted manner. Further, the individual might be moving in this environment characterized by variable illumination and a non-uniform background. Biometric modalities such as fingerprint and static iris cannot be easily acquired at large stand-off distances. On the contrary, the face, gait, and iris on the move modalities

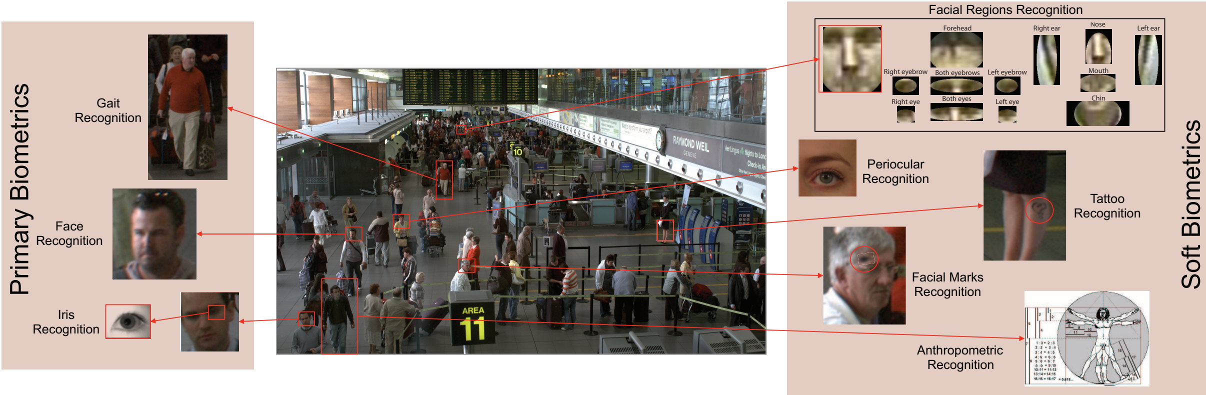


Figure 2.8: *Biometrics traits suitable to be used in systems at a distance. Center image extracted from http://avigilon.com.*

can easily be acquired at a distance (see an example in Fig. 2.8), although the smaller spatial resolution of the face at long distances can degrade the accuracy of face recognition systems. In the next sections the suitable *primary biometrics* traits to be extracted and used in systems at a distance are discussed.

There are many situations where *primary biometric* traits (i.e., gait, face, and iris) are either corrupted or unavailable, and *soft biometric* information is the only available clue for person recognition. For example, while a surveillance video may not capture the complete face of a suspect, the face image in the video may reveal anthropometric information such as the suspect's gender, complexity, ethnicity, etc., or the presence of a mark or tattoo may provide additional valuable clues.

In the next sections we will also discuss possible uses of *soft biometric* traits in biometrics at a distance (i.e., facial regions, periocular, facial marks, tattoos and anthropometric information, see Fig. 2.8). The periocular biometrics are gaining increasing attention since they offer a trade-off between using the entire face image and the iris portion only. Facial marks and tattoos are also gaining widespread attention since they offer complementary information that can be exploited along with *primary biometric* traits.

2.4.1. Primary Biometrics

- **Gait Recognition**.

  Gait-based human recognition is a suitable technology for biometric recognition at a distance [Nixon and Carter, 2006]. Gait is defined as the pattern of locomotion in animals and humans, i.e., it is the manner in which people walk. The formal definition of gait recognition refers to human motion but practical approaches include both dynamic and static features (such as body shape) of the moving human body.

  Therefore, gait recognition is perceived as an attractive solution for distance-based identification that shows some advantages in this kind of scenarios: *i*) the gait biometric can be acquired passively in a not intrusive way and, therefore, explicit subject cooperation is not required for data acquisition, and *ii*) this biometric trait can be extracted in low resolution images or videos, this means that common camera devices such as low resolution CCTV cameras may be used for acquisition. On the other hand this technology also present some disadvantages in systems at a distance that are mainly the occlusions due to the clothes (e.g., a person wearing an attire such as a trench coat) or other bodies (e.g., crowded scenarios). Gait recognition has also limitations when people are walking on a bumpy surface, downhill, uphill, etc.

- **Face Recognition.**

  The face modality has several advantages that make it preferable and the most suitable in biometric systems at a distance [Zhao *et al.*, 2003]. Firstly, unlike other biometrics such as fingerprints, face can be acquired at a distance using non-contact sensors. This trait is generally available in challenging scenarios (e.g., crowded scenarios) in contrast to others like gait. Also, the face has several interesting information sources in addition to the identity such as the emotions of a person (e.g., happiness, anger, etc.) as well as biographic information (e.g., gender, ethnicity, and age). Finally, this is a biometric trait very accepted in the population and people are generally willing to share it in the public domain (e.g., social media applications such as Facebook).

  While humans seem to be adept in determining the similarity between two face images acquired under diverse conditions, the process of automated face recognition is beset with several challenges in systems at a distance: *i*) The main disadvantage of this technology is the variation of a face image due to variability factors such as the age, pose, illumination, and facial expressions. *ii*) The occlusions are also a very important disadvantage in face recognition systems at a distance that produce changes in appearance due to make-up, facial hair, or accessories (e.g., hat, sunglasses, etc.) *iii*) Moreover, there may be similarities between the face images of different persons, especially if they are genetically related (e.g., identical twins, father and son, etc.) Such inter-class similarities further compound the difficulty of recognizing people based on their faces. Despite these challenges, significant progress has been made in the field of automated face recognition at a distance over the past two decades.

- **Iris Recognition.**

  Iris recognition is one of the most powerful techniques for biometric identification [Bowyer *et al.*, 2008b]. In the beginning this technology was conceived for controlled acquisition systems at close distance where users had to cooperate in the acquisition process. Therefore, the main problem in such systems were the constraints on position and motion but there are new advances based on high-resolution cameras, video synchronized strobed illumination, and specularity based image segmentation which try to deal with these problems. Last research advances open the opportunity to this technology to be suitable to biometric recognition at a distance. The Iris on the Move (IOM) system [Matey *et al.*, 2006] is the first system to enable the capture of iris images of sufficient quality at a distance while the subject is moving at a normal walking pace through a minimally confining portal. These advances make this technology to be considered for future applications in biometrics at a distance.

2.4.2. Soft Biometrics

- **Facial Regions Recognition.**

  The concept of facial region recognition is based on the use of facial regions such as the nose, mouth, eyes, eyelids, etc. as independent biometric traits in order to recognize people [Tome *et al.*, 2013e].

  This approach has some remarkable benefits in systems at a distance: *i*) facial regions can be extracted from a low resolution images (e.g., CCTV scenarios), *ii*) allow investigators to work only with particular regions of the face, and *iii*) prevent that incomplete, noisy, and missing regions degrade the recognition accuracy. In the same way that the field of cognitive science continues to investigate the precise roles of facial regions and holistic processing in human face perception [Gold *et al.*, 2012], automatic face recognition algorithms also need to explore the role that facial regions processing could have in improving the performance of systems at a distance.

- **Periocular Recognition**

  Periocular recognition is based on person identification using the region around the eyes (i.e., the skin, eyebrow, and eye) [Park *et al.*, 2011]. The use of the periocular region as a biometric cue represents a good trade-off between using the entire face region or using only the iris for recognition. In biometric systems at a distance, the periocular biometric has an interesting role such as soft biometric information that can be useful for identification when good quality images are available and as a complementary information with low resolution images.

- **Face Marks Recognition.**

  Advances in sensing technology have made it easy to capture high resolution face images. From these high resolution face images, it is possible to extract details of skin irregularities, also known as facial marks. This has opened new possibilities in face representation and matching schemes [Jain *et al.*, 2011a]. These skin details are mostly ignored and considered as noise in a typical face recognition system. However, facial marks can be used to *i*) supplement existing facial matchers to improve the identification accuracy, *ii*) facilitate fast face image retrieval, *iii*) enable matching or retrieval with partial or off-frontal face images, and *iv*) provide more descriptive evidence about the similarity or dissimilarity between face images, which can be used as evidence in legal proceedings.

- **Tattoo Recognition**

  The use of tattoos imprinted on the human body in suspect identification started with the Bertillon system [Bertillon, 1896]. Since then, images of tattoos on the human body have been routinely collected and used by law enforcement agencies to assist in suspect and victim identification. When the primary biometric traits are unavailable or corrupted, tattoos can be used to identify victims or suspects as demonstrated by Jain *et al.* [2012a]. Tattoos provide more discriminative information than the traditional demographic indicators such as age, height, race, and gender for person identification. In most of the cases these tattoos are visually available and may be distinguished at a distance.

- **Anthropometric Soft Biometrics.**

  The anthropometric soft biometrics are based on intrinsic characteristics of the subjects (i.e., height, arm length, complexity, hair colour, etc.) This kind of soft biometrics features can be recognized at a distance in most of cases and can be classified as:

  - ***Global***. Features such as age, ethnicity and sex. The demographic information as the gender and ethnicity of a person does not typically change over the lifetime, so it can be used to filter the database to narrow down the number of candidates. On the other hand, age is easily estimated by physical traits at a distance and it can also be used to filter suspects.
  - ***Body***. Features that describe the target's perceived somatotype [Macrae and Bodenhausen, 2000] (height, weight, etc.) These traits have a close correlation between the style and kind of clothes that the subject is wearing in the annotation process. For example, tight clothes will allow to obtain more stable labels than loose clothes.
  - ***Head***. This is, an area of the body humans pay great attention to if it is visible [Hewig *et al.*, 2008] (hair colour, beards, etc.) These are very interesting soft biometrics to be fused with face recognition systems.
  - ***Facial***. Features that describe the facial regions based on the morphological analysis, i.e., kind of eyelids, eyebrows, length and width of eyebrows, eyes shape, etc.
  - ***Context information***. Information of the scene where the object of interest is immersed (e.g., we want to find a person on a beach, so the context information of the beach is going to be very useful.) These features are very useful at information retrieval applications.

2.5. Chapter Summary and Conclusions

Since the establishment of biometric research as an specific research area, the biometric community has focused its efforts in the development of accurate recognition algorithms in controlled scenarios. Nowadays, biometric recognition is a mature technology that is used in many applications (e.g. e-Passports, ID cards or border control [US-VISIT Program of the U.S. Department of Homeland Security]). More recently, the biometric community is focusing on biometrics at a distance in unconstrained and uncontrolled scenarios. We can notice in recent studies that the performance of biometric systems at a distance is heavily affected by different variability factors depending on the acquisition distance. The problem of dealing with these variability factors is a current research challenge within the biometric community [NIST: Face Challenges], and the main topic of the present Dissertation.

In this chapter, we present an overall overview of the different components and issues that conform the biometric variability assessment problem in systems at a distance. Issues like the variability factors influencing biometric samples, the strategies and approaches to dealing with these variability factors, the definition of scenarios and biometric systems At a Distance (AD), or the role of distance measures within this kind of biometric systems are addressed here. We also present a framework for graduation and evaluation of the variability factors depending on the acquisition distance, as well as existing primary and soft biometric traits suitable for biometric scenarios at a distance.

This chapter includes novel contributions regarding the taxonomy of variability factors affecting biometric samples.

Chapter 3

Proposed Methods: Soft Biometrics and Adaptive Fusion

In this chapter we present the two methods that have been proposed during the development of the Thesis, and which will be used to deal with variability factors in the experimental part of the Dissertation (Chapters 5, 6, and 7.) There is a large amount of literature describing compensation methods to deal with general and specific variability but it is not the purpose of this Thesis to organize and summarize this growing area. Based on the comprehensive treatment of variability factor of Chapter 2, in the present Chapter we just focus on presenting two new methods that can provide big benefits in highly variable scenarios.

The proposed methods are: *i*) soft biometrics based on morphological information that can be applied in a straight forward manner to different matchers and biometric traits, and *ii*) adaptive fusion schemes using ancillary information and person to camera distance estimation (which present the advantage over previously proposed schemes of using the distance to identify the scenario and apply the best solution). Both methods have been validated on realistic databases following systematic and replicable protocols, reaching remarkable results.

The soft biometrics considered will be extensively analysed in Chapter 6, where they will be applied to video surveillance and forensics. The adaptive fusion schemes proposed will be applied in Chapter 7 as a method to deal with variability factors fusing ancillary information (e.g. soft biometrics and facial regions). The distance estimation between the subject and the camera (studied in Chapter 5) is used to carry out various combinations of face recognition systems in Chapter 7.

This chapter is structured as follows. The soft biometrics are presented in Sect. 3.1, and the adaptive fusion approaches in Sect. 3.2. The chapter summary and conclusions are given in Sect. 3.3.

This chapter assumes a basic understanding of the fundamentals of pattern recognition and classification [Duda *et al.*, 2001; Jain *et al.*, 2000; Theodoridis and Koutroumbas, 2008].

This chapter is based on the publications: Tome *et al.* [2013b,c,e, 2012].

3.1. Soft Biometrics

The first system in the history that attempted to describe people for identification based on the morphological and physiological traits was the anthropometric system developed by Bertillon [1896]. This system was based on features such as body measurements (anthropometry), morphological description of the appearance and shape of the body, and peculiar marks observed on the body. This system was very useful in tracking criminals in the beginning but it had an unacceptably high rate of false identification. This was due to that these characteristics did not have the distinctiveness and permanence to uniquely identify an individual over a period of time. Therefore, soft biometric traits are defined as those characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals (see Fig. 2.8 for examples of soft biometric traits).

Soft biometric traits can either be continuous or discrete. Traits such as gender, eye color, ethnicity, etc. are discrete in nature. On the other hand, traits like height and weight are continuous variables. In principle a system that is completely based on soft biometric traits cannot provide the required accuracy in the recognition of individuals. However, soft biometric traits can be used to improve the performance of a traditional biometric system (e.g., gait, face, etc.) in many ways.

The first works in soft biometrics [Heckathorn *et al.*, 1997; Jain *et al.*, 2004a,b] tried to use demographic information (e.g., gender and ethnicity) and soft attributes like eye color, height, weight and other visible marks like scars [Jain and Park, 2009; Park and Jain, 2010] and tattoos [Lee *et al.*, 2008] as ancillary information to improve the performance of biometric systems. They showed that soft biometrics can complement the traditional (primary) biometric identifiers (like face recognition) and can also be useful as a source of evidence in courts of law because they are more descriptive than the numerical matching scores generated by a traditional face matcher. But in most cases, this ancillary information by itself is not sufficient to recognize a user. In contrast, this PhD Thesis involves the application of an extensive set of labels that can be visually described by humans at a distance.

More recently, Kumar *et al.* [2009] explored comparative facial attributes in the LFW Face Database [Huang *et al.*, 2007] for face verification. Gupta and Davis [2008] and Siddiquie and Gupta [2010] used prepositions and adjectives to relate objects (e.g., persons) to each other for more effective contextual modelling and active learning, respectively. Recent works such as Reid and Nixon [2013] introduce the use of comparative human descriptions for facial identification. They use twenty-seven comparative traits to accurately describe facial features, which are determined by the Elo rating system from multiple comparative descriptions. These facial features are extracted from mugshot images.

The use of soft biometric traits in automated human recognition systems has several benefits. It is, therefore, essential to carefully investigate issues related to its extraction and recognition capacity. Surveillance footage is generally of inferior quality and so traditional forms of identification at a distance cannot be easily used. Soft biometrics offer a solution in this regard but lack the distinctiveness that is expected of biometric traits.

This technology is ideal for applications with smaller populations, such as tracking people within a camera network or identifying people who are known to be located within a certain area. In these applications the view invariance is a key aspect when working with surveillance footage, a successful technique must identify soft biometric features from any view of the subject. Emphasis must be placed on finding practical ways to obtain view of invariant features (similar to [Denman *et al.*, 2009]) or developing methods to predict hidden features based on what can be observed.

In the next sections we present two sets of soft biometrics: *i*) **soft biometrics for video surveillance**, which are extracted visually from a subject at a distance (e.g., gender, height, hair length, etc.), and *ii*) **soft biometrics for forensics**, which are extracted at close distance (e.g., eyebrows form, nose height, mouth length, etc).

3.1.1. Soft Biometrics for Video Surveillance

As presented in previous Chapter 2, the two most popular traits for identification at a distance are face [Zhao *et al.*, 2003] and gait [Nixon and Carter, 2006]. These can suffer from the poor sensor quality of most CCTV cameras. Low resolution can seriously impair facial recognition, and low frame rates (sometimes even time-lapse cameras) obscure the motion of the human body required for gait recognition. In contrast, several soft traits can often be obtained from very poor quality video or images. This has huge potential for immediate real world use without upgrading the vast surveillance infrastructure already deployed.

In this PhD Thesis a set of soft biometrics features have been used, whose main value is that they are discernible by humans at a distance. These physical trait labels were proposed by Samangooei [2010] and are available at the Southampton Multibiometric Tunnel Database (TunnelDB) [Seely *et al.*, 2008].

This soft biometric information was annotated against recordings taken of the individuals in laboratory conditions [Seely, 2010]. A range of discrete values is given to each trait label, e.g. "Arm length" marked as 1 (very short), 2 (short), 3 (average), 4 (long), and 5 (very long). The annotation process of each label is described in detail in [Samangooei, 2010]. A summary of these trait labels and their associated discrete semantic terms is provided in Table 3.1. These labels were designed based on which traits humans are able to consistently and accurately use when describing people at a distance. The traits were grouped in 3 classes, namely:

Body features that describe the target's perceived somatotype [Macrae and Bodenhausen, 2000] (height, weight, etc.) These traits have a close correlation between the style and kind of clothes that the subject is wearing in the annotation process. For example, tight clothes will allow to obtain more stable labels than loose clothes.

Global traits (age, ethnicity and sex). The demographic information as the gender and ethnicity of a person does not typically change over the lifetime, so it can be used to filter

Body

Trait	**Range of Values**
1. Arm Length	Very Short, Short, Average, Long and Very Long
2. Arm Thickness	Very Thin, Thin, Average, Thick and Very Thick
3. Chest	Very Slim, Slim, Average, Large and Very Large
4. Figure	Very Small, Small, Average, Large and Very Large
5. Height	Very Short, Short, Average, Tall and Very Tall
6. Hips	Very Narrow, Narrow, Average, Broad, and Very Broad
7. Leg Length	Very Short, Short, Average, Long and Very Long
8. Leg Direction	Very Bowed, Bowed, Straight, Knock Kneed, and Very Knock Kneed
9. Leg Thickness	Very Thin, Thin, Average, Thick and Very Thick
10. Muscle Build	Very Lean, Lean, Average, Muscly, and Very Muscly
11. Proportions	Average and Unusual
12. Shoulder Shape	Very Rounded, Rounded, Average, Square and Very Square
13. Weight	Very Thin, Thin, Average, Big and Very Big

Global

Trait	**Range of Values**
14. Age	Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior
15. Ethnicity	European, Middle Eastern, Indian/Pakistan, Far Eastern, Black, Mixed, Other
16. Sex	Female, Male

Head

Trait	**Range of Values**
17. Skin Colour	White, Tanned, Oriental and Black
18. Facial Hair Colour	None, Black, Brown, Red, Blond and Grey
19. Facial Hair Length	None, Stubble, Moustache, Goatee and Full Beard
20. Hair Colour	Black, Brown, Red, Blond, Grey and Dyed
21. Hair Length	None, Shaven, Short, Medium and Long
22. Neck Length	Very Short, Short, Medium and Long
23. Neck Thickness	Very Thin, Thin, Average, Thick and Very Thick

Table 3.1: *Soft biometrics for surveillance. Extracted from [Samangooei, 2010].*

the database to narrow down the number of candidates. On the other hand, age is easily estimated by physical traits at a distance and it can also be used to filter suspects.

Head features, an area of the body humans pay great attention to if it is visible [Hewig *et al.*, 2008] (hair colour, beards, etc.) These are very interesting soft biometrics to be fused with face recognition systems.

Following the definitions in Chapter 11 of [Theodoridis and Koutroumbas, 2008], we can see that some of the features are *nominal*, i.e., their values can not be ordered meaningfully (e.g., ethnicity (15), sex (16), skin (17), facial hair (18) and hair colour (20)) whereas others are *ordinal*, i.e., their values can be meaningfully ordered (e.g., arm length (1), arm thickness (2), height (4), weight (13), and hair length(21)).

We assume that these soft biometrics for surveillance scenarios are discernible by humans at a distance and are extracted manually by an annotator or automatically by an automatic system. The other assumption in biometrics at a distance is that the quantity of soft biometric

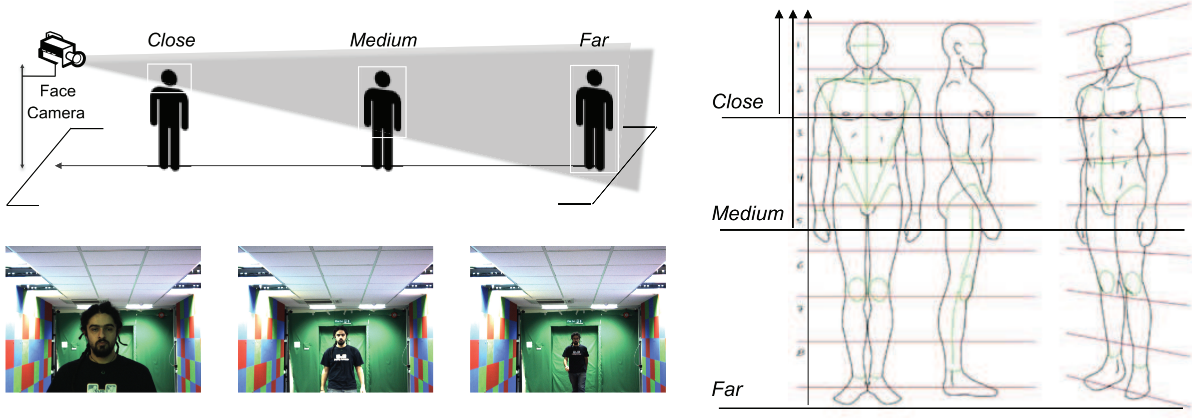


Figure 3.1: *Body region visible at the three distances considered. A person walking frontal to the camera is captured by a high-resolution video camera (10 fps and resolution of* 1600×1200*) and soft labels available visually in each scenario are extracted.*

features in the scene is variable with the distance as it shown in Fig. 3.1.

Definition 3.1.1. Given a population of S subjects and a soft biometric feature X, the problem of extraction of this feature can be carried out following two configurations: i) manually, by a human annotator a, or ii) automatically, by an automatic system a. Each soft biometric feature can be extracted by several annotators or different automatic systems, therefore $a = \{1, \ldots, A\}$, where A is the total number of annotators. We can also formally define a set of soft biometric features as X^k, where kth feature is $k = \{1, \ldots, K\}$, and K is the total number of soft biometric features.

Based on previous assumptions, in this PhD Thesis we develop a general methodology to understand the behaviour of soft biometric labels and their best application to biometrics at a distance based on the use of just the available soft biometric information in the scene. For this purpose, three general scenarios varying the distance between camera and subject are defined and used in our experiments. The three scenarios are defined as follows (see Fig. 3.1):

- ***Close*** distance. Includes both the face and the shoulders.
- ***Medium*** distance. Includes the upper half of the body.
- ***Far*** distance. Includes the full body.

The rationale behind the proposed methodology is the fact that depending on the particular scenario, some labels may not be visually present and others may be occluded. As a result, the discriminative information of the soft biometrics will vary depending on the distance. Table 3.2 shows the soft labels available for each of the scenarios defined.

The detailed description and evaluation of these soft biometrics features for video surveillance can be found in Chapter 6.

		Close	Medium	Far
Body	1. Arm Length		X	X
	2. Arm Thickness		X	X
	3. Chest		X	X
	4. Figure		X	X
	5. Height			X
	6. Hips			X
	7. Leg Length			X
	8. Leg Direction			X
	9. Leg Thickness			X
	10. Muscle Build		X	X
	11. Proportions	X	X	X
	12. Shoulder Shape	X	X	X
	13. Weight		X	X
Global	14. Age	X	X	X
	15. Ethnicity	X	X	X
	16. Sex	X	X	X
Head	17. Skin Colour	X	X	X
	18. Facial Hair Colour	X	X	X
	19. Facial Hair Length	X	X	X
	20. Hair Colour	X	X	X
	21. Hair Length	X	X	X
	22. Neck Length	X	X	X
	23. Neck Thickness	X	X	X

Table 3.2: *Soft biometrics features available (marked with X) visually in each scenario at a distance.*

3.1.2. Soft Biometrics for Forensics

Most forensic laboratories follow methodologies based on Bertillon's approach [Bertillon, 1896] such as the police sketch used by the Spanish Guardia Civil (DGGC) or Netherlands Forensic Institute (NFI) in the identification of criminals and the standards defined by Facial Identification Science Working Group (FISWG). This sketch consists of a verbal description of specific facial traits following a precise and well defined procedure. In particular, the morphological facial features of a subject are classified in three groups:

- Morphological
- Chromatic
- Complementary

where *morphological* features describe the form, magnitude and direction of the facial traits, the *chromatic* features are focused on the different colouration that a face has, and the *complementary* features refer to other concepts that can not be analysed by the other two previous types.

In this PhD Thesis we present a set of regions extracted from a frontal human face based on these procedures. The face regions proposed are the next:

- Forehead
- Eyebrows
- Eyeball and Orbit
- Nose
- Mouth and Lips
- Chin
- Ear
- Contours

Based on the facial regions, we have defined, also based on procedures from DGGC and NFI, the set of soft biometrics shown in Table 3.3. These attributes are:

Continuous features which take continuous values, generally distances in facial traits (e.g. eyebrows length, nose height and width, mouth length, etc.)

Discrete features that take a finite number of categories. For example, eyebrow form that can be arched, rectilinear and sinuous. This group of features needs a training set in order to establish the thresholds between the range of values.

Similarly to previous Definition 3.1.1 we can also formally define here a set of facial soft biometric features as X^k, where $k = \{1, ..., K\}$, and $K = 56$ is the total number of soft biometric features, 32 continuous and 24 discrete.

The detailed description and evaluation of these facial soft biometrics features can be found in Chapter 6 of the present Dissertation.

3.2. Adaptive Fusion

This section describes the adaptive score fusion schemes proposed in this Thesis to deal with the variability factors. These schemes are divided into three classes: 1) scenario-based, 2) soft biometrics-based, and 3) regions-based. The three classes are introduced sequentially in order to facilitate the description.

We use the following nomenclature and conventions throughout the rest of the chapter. Given a multimodal biometric verification system consisting of M different unimodal systems $j = 1, \ldots, M$, each one computes a similarity score s between an input biometric pattern B and the enrolled pattern or model of the given claimant u. The similarity scores s are normalized to $\hat{s}$. Let the normalized similarity scores provided by the different unimodal systems be combined into a multimodal score $\hat{\mathbf{s}} = [\hat{s}_1, \ldots, \hat{s}_M]^T$, where $[\cdot]^T$ denotes transpose. The design of a fusion scheme consists in the definition of a function $f : \mathbb{R}^M \to \mathbb{R}$, so as to maximize the separability of client $\{f(\hat{\mathbf{s}})|\text{client attempt}\}$ and impostor $\{f(\hat{\mathbf{s}})|\text{impostor attempt}\}$ fused score distributions. This function may be trained by using labelled training scores $(\hat{\mathbf{s}}_i, z_i)$, where $z_i = \{0 = \text{impostor attempt}, 1 = \text{client attempt}\}$, and $i = 1, \ldots, N$. The rest of the chapter deals with different schemes for constructing this function adapted both to the acquisition distance, soft biometrics information, and/or the facial regions of the input biometric signals according to different criteria. In Fig. 3.2 we depict the general system model including all the notations defined above.

Facial Trait	Continuous	Discrete
Forehead	1. Height	1. Height (Short, Average, and Long)
	2. Width	2. Width (Small, Average, and Large)
Eyebrows	3. Separation (Distance between eyebrows)	3. Separation (Near and Distant)
	4. Elevation$_L$ inner (Distance eyebrow and eye)	4. Elevation (Low, Average, High, and Asymmetric)
	5. Elevation$_L$ outer (Distance eyebrow and eye)	
	6. Elevation$_R$ inner (Distance eyebrow and eye)	
	7. Elevation$_R$ outer (Distance eyebrow and eye)	
	8. Length$_L$	5. Length$_L$ (Short and Long)
	9. Length$_R$	6. Length$_R$ (Short and Long)
	10. Average Width$_L$	7. Width$_L$ (Narrow, Linear, and Wide)
	11. Average Width$_R$	8. Width$_R$ (Narrow, Linear, and Wide)
	12. Angles between corners$_L$	9. Direction$_L$ (Horizontal, Oblique Internal, and Oblique External)
	13. Angles between corners$_R$	
		10. Direction$_R$ (Horizontal, Oblique Internal, and Oblique External)
		11. Form$_L$ (Arched, Rectilinear, and Sinuous)
		12. Form$_R$ (Arched, Rectilinear, and Sinuous)
Eyeball and Orbit	14. Horizontal Opening$_L$	13. Horizontal Opening$_L$ (Small and Large)
	15. Horizontal Opening$_R$	14. Horizontal Opening$_R$ (Small and Large)
	16. Interocular Distance (inner corners)	15. Interocular Distance (Small, Normal, and Large)
	17. Angles$_L$ between corners	
	18. Angles$_R$ between corners	
Nose	19. Width	16. Width (Small, Average, and Large)
	20. Height	17. Height (Short, Average, and Long)
	21. Nose Root Width	18. Nose Root Width (Narrow, Average, and Wide)
	22. Naso-Labial Height	19. Naso-Labial Height (Short, Average, and Long)
Mouth	23. Length	20. Length (Small, Average, and Large)
	24. Average Height	21. Orientation (Oblique$_L$, Neutral, and Oblique$_R$)
	25. Angles between corners	22. Particularities (Heart Form)
Chin	26. Width	23. Width (Small and Large)
	27. Height	24. Height (Short, Average, and Long)
Ears	28. Length$_L$	
	29. Length$_R$	
	30. Angle$_L$ between corners	
	31. Angle$_R$ between corners	
Contours	32. Average Line Length	

Table 3.3: *Facial soft biometric features and their associated semantic terms grouped in continuous and discrete values.*

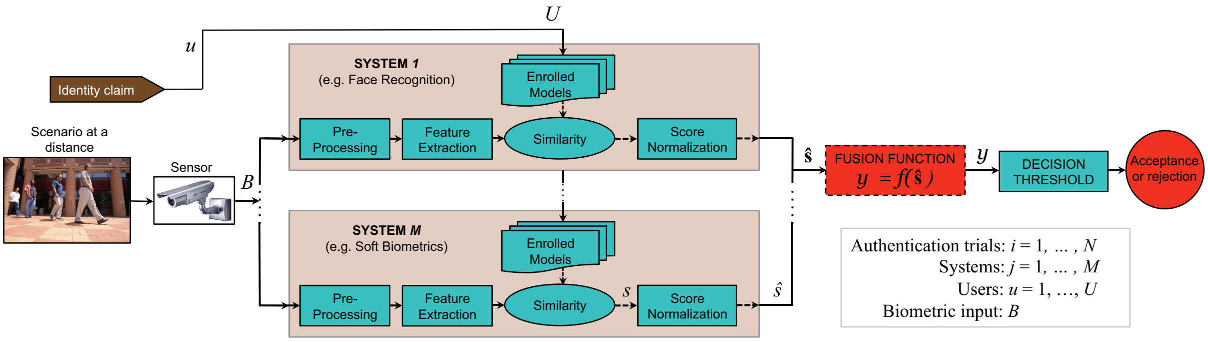


Figure 3.2: *General system model of multimodal biometric authentication using score level fusion including name conventions.*

To carry out the fusion stage of the biometric modalities, scores s of the different systems were first normalized $\hat{s}$ to the $[0,1]$ range using the tanh-estimators described by Jain *et al.* [2005]:

$$\hat{s}_i = \frac{1}{2}\left\{\tanh\left(C \cdot \frac{s_i - \mu_{SD}}{\sigma_{SD}}\right) + 1\right\}, \tag{3.1}$$

where s_i is the ith score, $\hat{s}_i$ denotes the normalized score, C is a constant, and μ_{SD} and σ_{SD} are respectively the estimated mean and standard derivation of the score distribution.

The tanh-estimators introduced by [Hampel *et al.*, 2005] are robust and highly efficient. This method reduces the influence of the points at the tails of the distribution during the estimation of the location and scale parameters. Hence, this method is robust against outliers.

3.2.1. Scenario-based Fusion

Distance measures between the camera and the person to be recognized can be used for adapting the different modules of a multimodal authentication system at a distance. Although any processing module is subject to this adaptation based on the person to camera distance, only scenario-based score fusion is considered in this Thesis. In Sect. 8.2 we provide some points of ongoing efforts and future works using camera to person distance measures for adapting other modules. The system model of scenario-based score fusion proposed in this work is shown in Fig. 3.3.

One straightforward way to incorporate the acquisition distance to the score fusion approach is by including weights in simple combination approaches [Fierrez, 2006]. This can be achieved by using the following scenario-based score fusion function

$$y_i = \sum_{j=1}^{M} g^j(d_i)\hat{s}_i^j, \tag{3.2}$$

where d_i is an acquisition distance measure corresponding to the score $\hat{s}_i^j$. The function $g^j(d_i)$ takes a camera to person distance estimation and outputs a confidence measure of the system

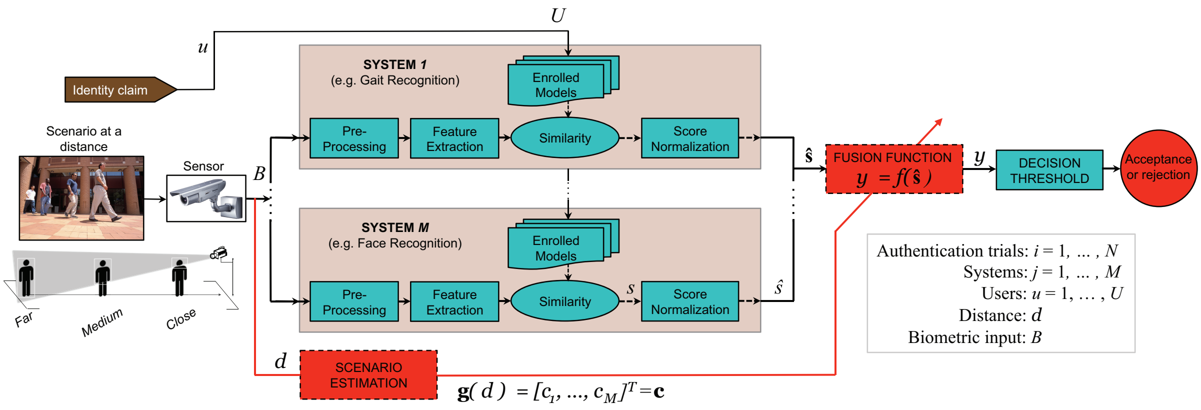


Figure 3.3: *System model of biometric authentication with scenario-based score fusion.*

j in providing a reliable matching score for the particular biometric signal being tested i, with $sum_{j=1}^{M} g^j(d_i) = 1$. For a particular biometric input i, as shown in Fig. 3.3, we can collect all confidence measures corresponding to the M different fused systems in vector $\mathbf{c} = [c_1, \ldots, c_M]^T$.

The concept of measuring or estimating the acquisition distance in order to define different scenarios has not been traditionally used in person recognition at a distance. This automatic scenario estimation based on the acquisition distance gives us knowledge about the variability level that affects the system (i.e., different scenarios usually present different variability factors) and therefore is a valuable tool for system adaptation. It is important to emphasize that in biometrics at a distance the scenario estimation is an important challenge because as the person is moving away from the acquisition device variability factors lead to a change of scenario. Thus, the variability factors can affect in different levels depending on the distance to the capturing device.

In the experimental Chapter 7, we apply this methodology with an example camera to person distance estimator based on the face size in the image plane.

The detailed description and evaluation of this example acquisition distance index and scenario-based fusion can be found in Chapter 7 of the present Dissertation.

3.2.2. Soft Biometrics-based Fusion

Processing soft biometrics typically require less computation and input data quality compared to other forms of identification at a distance, making them cheap and non-intrusive. Niinuma *et al.* [2010] clearly demonstrate the suitability of soft biometric traits for continuous user authentication. Determining applications adequate for this form of biometric identification is essential for advancing the field.

The objective in this part of the Dissertation is stressing the importance of soft biometrics at a distance using fusion techniques. By agglomerating multiple soft biometric features using fusion techniques, the recognition performance can be enhanced significantly in very challenging and realistic scenarios.

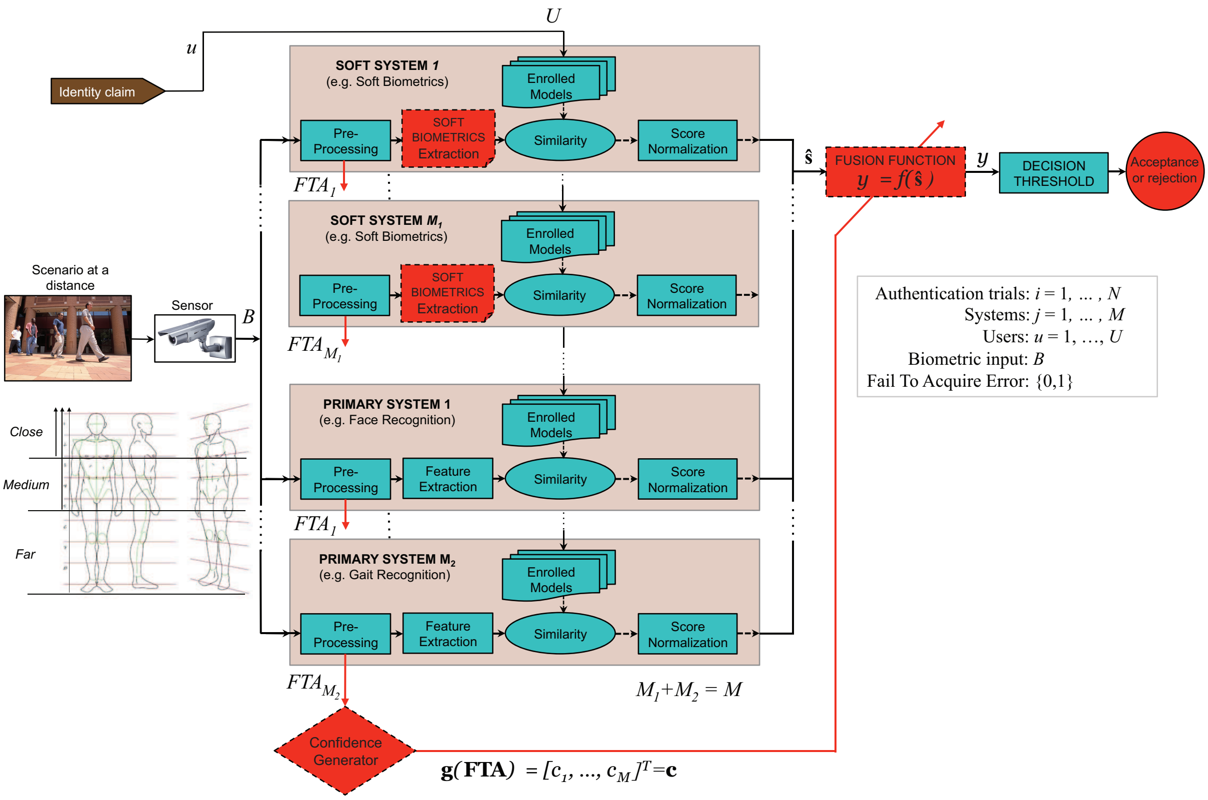


Figure 3.4: *System model of biometric authentication with soft biometrics-based score fusion.*

Soft biometrics offer several benefits over other forms of identification at a distance as they can be acquired from low resolution and low frame rate videos, and from an arbitrary viewpoint of the subject. This allows for the use of soft biometrics when primary biometric identifiers cannot be obtained or when only a description of the person is available. The system model of soft biometrics-based score fusion proposed in this work is shown in Fig. 3.4.

One straightforward way to incorporate soft biometrics to the score fusion approach is by considering Failure to Acquire (FTA) errors in simple fusion approaches [Fierrez, 2006]. This can be achieved similarly to Eq. (3.2) by the following fusion function:

$$y_i = \sum_{j=1}^{M} g^j(\mathrm{FTA}_i^j)\hat{s}_i^j. \tag{3.3}$$

In this case $g^j(\mathrm{FTA}_i^j)$ takes as input the binary features $\mathrm{FTA}_i^j = \{0, 1\}$ corresponding to FTA events in each system j for a particular input i, and outputs, similarly as in Eq. (3.2), a confidence measure of the system j in providing a reliable matching score for the particular biometric signal being tested i.

FTA is the Fail To Acquire error produced when there is a biometric trait in the image, but it is not detected. Other important error to be considered is FTD, Fail To Detect error produced when the biometric trait detector finds an object in the image, but it is not a biometric trait.

As it will be further developed in the experimental Chapter 7, this general approach can

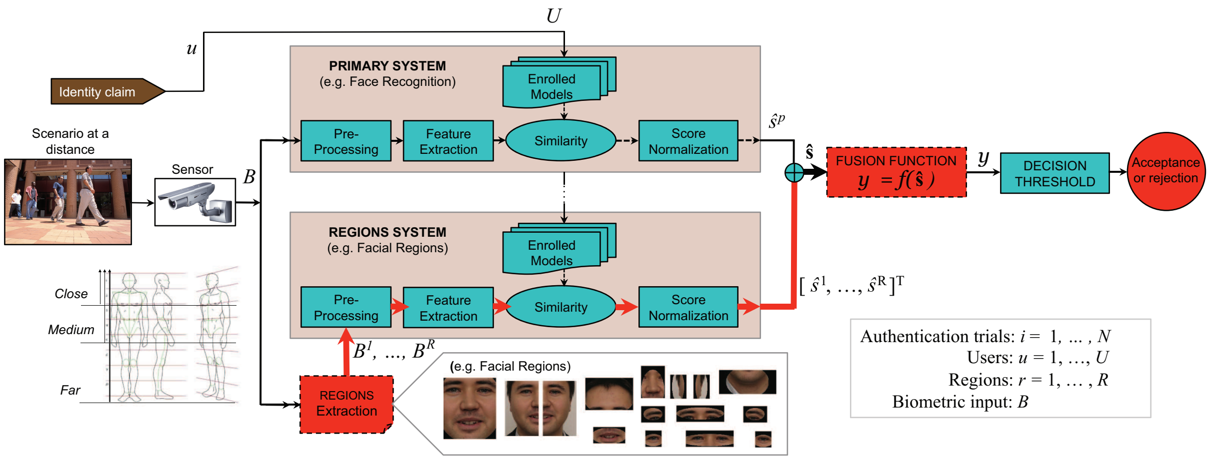


Figure 3.5: *System model of biometric authentication with regions-based score fusion.*

be applied, for example, in a switch fashion [Fronthaler *et al.*, 2008] to consider soft biometrics in cases where primary biometrics are detected, and weighted sum fusion of both primary and soft biometrics when both scores are available. As it will be also shown experimentally, this helps in realistic challenging scenarios dealing with low resolution images such as surveillance and forensics.

The detailed description and evaluation of this soft biometrics-based fusion can be found in Chapter 7 of the present Dissertation.

3.2.3. Regions-based Fusion

As discussed in Chapter 2, the two most popular traits for identification at a distance are face [Zhao *et al.*, 2003] and gait [Nixon and Carter, 2006]. Automatic person recognition systems are generally designed to match images of full faces or bodies. However, in practice, the full trait is not always available, e.g., due to occlusions and other variability factors. On the other hand, in forensics, the examiners usually carry out a manual inspection of the face images, focussing their attention not only on the full face but also on face regions. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc.

Understanding how different human facial and body regions are combined on different application scenarios has some remarkable benefits, for example: *i*) allowing to work only with particular regions, or *ii*) preventing that incomplete, noisy, and missing regions degrade the recognition accuracy. Further, a better understanding of the combination of facial and body regions should facilitate the study of regions-based person recognition. Therefore a fusion scheme based on different regions is proposed to deal with variability factors and improve the performance of biometric systems at a distance.

The system model of regions-based score fusion proposed in this work is shown in Fig. 3.5. One straightforward way to incorporate the regions to the score fusion approach is by including various combinations of regions. Various selections of such regions in practical applications will be studied in Chapter 7.

Once the regions to fuse are determined, a similar fusion approach to the one defined in Eq. (3.2) can be also applied here:

$$y_i = \sum_{r=1}^{M} g^r(B_i)\hat{s}_i^r. \tag{3.4}$$

In this case, the weight $g^r(B_i)$ is an estimation of the confidence in obtaining a reliable recognition using only the region r of the face or body for the input B_i. In the experimental chapter we will obtain $g^r(B_i)$ based on the recognition errors on a development database using standard recognizers.

A key element in the proposed region-based fusion approach is to properly segment the face or body regions. This will be studied, respectively in Chapters 5 and 7 of the present Dissertation, where we will describe and evaluate new face segmentation methods with application to surveillance and forensics.

3.3. Chapter Summary and Conclusions

In this chapter we have summarized the main contributions of this PhD Thesis, namely: soft biometrics and adaptive fusion for challenging biometric scenarios.

The presented algorithms include a soft biometrics system based on face regions and morphological information that can be applied in a straight forward manner to different matchers and biometric traits, and adaptive fusion schemes using ancillary information and camera to person distance estimation. All the methods will be validated on realistic databases following systematic and replicable protocols, reaching remarkable results as will be shown in the next chapters of the Dissertation.

All the methods proposed in this chapter are original contributions.

Chapter 4

Performance Evaluation of Biometric Systems at a Distance

This chapter summarizes the common practices in performance testing of biometric systems and presents the evaluation methodology followed in the Thesis for the variability assessment of automatic recognition systems. The main biometric databases at a distance used for both types of evaluations (performance and variability) are also described.

The chapter is organized as follows. First we summarize the guidelines for performance evaluation used in this Dissertation (Sect. 4.1). Finally we give an overview of the main existing biometric databases at a distance (Sect. 4.2) and we thoroughly describe the most important ones used in this Thesis.

4.1. Performance Evaluation of Biometric Systems

The practice in first research works on biometrics starting over three decades ago was to report experimental results using biometric data specifically acquired for the experiment at hand [Atal, 1976; Kanade, 1973; Nagel and Rosenfeld, 1977]. This approach made very difficult the fair comparison of different recognition strategies, as the biometric data was not made publicly available.

With the popularity of biometric systems and the creation of new research groups working in the same topics, the need for common performance benchmarks was recognized early in the past decade [Jain *et al.*, 2004d; Phillips *et al.*, 2000c]. In this environment, the first series of international competitions for person authentication based on different biometric traits were organized. In these competitions, biometric data along with specific experimental protocols were established and made publicly available. Some examples include the following campaigns: NIST Facial Recognition Technology Evaluations (FERET), starting in 1994 [Phillips *et al.*, 2005, 2000c]; NIST Speaker Recognition Evaluations (SRE), held yearly since 1996 [Przybocki and Martin, 2004]; NIST Iris Challenge Evaluations (ICE), first organized in 2005 [Phillips, 2006];

Fingerprint Verification Competitions (FVC), held biannually since 2000 [Cappelli *et al.*, 2006]; the Signature Verification Competition (SVC), organized in 2004 [Yeung *et al.*, 2004]; and the BioSecure Multimodal Evaluation Campaign held in 2007 [Mayoue *et al.*, 2009]. Comparative evaluations of commercial biometric technologies can also be found nowadays by standards institutions like NIST [Grother *et al.*, 2003; Wilson *et al.*, 2004] and CESG [Mansfield *et al.*, 2001], or consulting firms like the International Biometric Group [2006]. This is also at least one laboratory exclusively focused in the performance evaluation of biometric systems (Biometric Services International [BSI, 2009], a non-profit company working under the National Biometric Security Project [NBSP, 2009]) of the ISO/IEC 17025:2005 accreditation for testing [ISO/IEC 17025, 2005].

In this environment, and as a result of the experience gained in biometric performance evaluation, the UK Biometrics Working Group generated a set of best practices for testing and reporting performance results of biometrics systems [Mansfield and Wayman, 2002], to which we adhere in this PhD Thesis.

Performance evaluation of biometric recognition systems can be carried out at three different levels [Phillips *et al.*, 2000b]: technology, scenario, and operational.

The goal of a technology evaluation is to compare competing algorithms thus identifying the most promising recognition approaches and tracking the state-of-the-art. Testing of all algorithms is carried out on a standardized database. Performance with this database will depend upon both the environment and the population from which the data are collected. Because the database is fixed, the results of technology tests are repeatable. Some important aspects of a given database are: 1) Number of users, 2) number of recording sessions, and 3) number of different samples per session. Most standardized benchmarks in biometrics are technology evaluations conducted by independent groups or standards institutions [Maio *et al.*, 2004; Petrovska-Delacretaz *et al.*, 2009; Phillips *et al.*, 2000c; Przybocki and Martin, 2004; Yeung *et al.*, 2004].

The goal of scenario evaluations is to measure overall system performance for a prototype scenario that models an application domain. Scenario evaluations are conducted under conditions that model real-world applications [Bone and Blackburn, 2002; Mansfield *et al.*, 2001]. Because each system has its own data acquisition sensor, each system is tested with slightly different data, and thus scenario tests are not repeatable. An operational evaluation is similar to a scenario evaluation. While a scenario test evaluates a class of applications, an operational test measures performance of a specific algorithm for a specific application [Bone and Crumbacker, 2001].

In this Thesis we carry out the performance evaluation experiments as technology evaluations of different systems at a distance working in the *verification* mode.

4.1.1. Performance Measures of Authentication Systems

Biometric technologies can be ranked according to several criteria, including [Jain *et al.*, 2004d]: universality, distinctiveness, permanence, collectability, performance, acceptability and

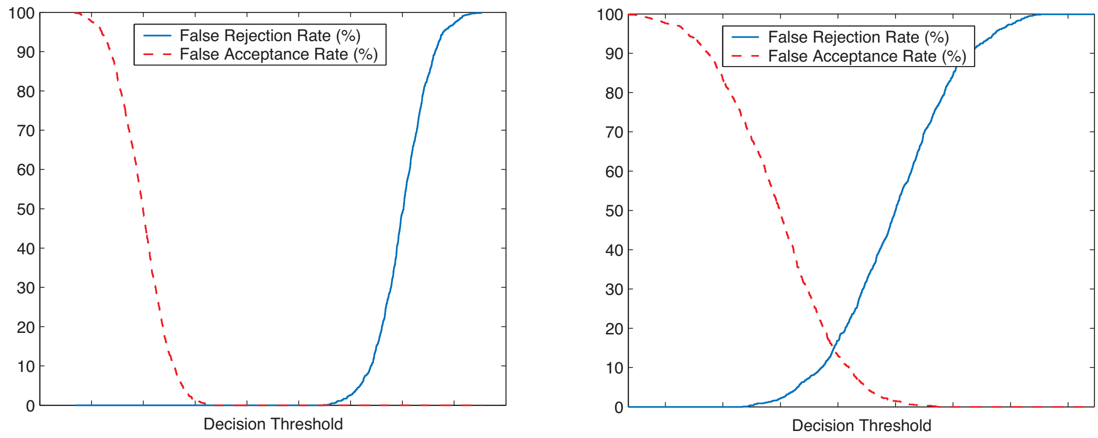


Figure 4.1: *FA and FR curves for an ideal (left) and real (right) authentication systems.*

circumvention, as it was mentioned in Sect. 1.1.1. In the experiments of this Thesis we concentrate on performance indicators to compare different methods, and more specifically on authentication error rates.

Biometric authentication can be considered as a detection task, involving a tradeoff between two types of errors [Ortega-Garcia *et al.*, 2004]: 1) False Rejection (FR), occurring when a client, target, genuine, or authorized user is rejected by the system, and 2) False Acceptance (FA), taking place when an unauthorized or impostor user is accepted as being a true user. Although each type of error can be computed for a given decision threshold, a single performance level is inadequate to represent the full capabilities of the system. Therefore the performance capabilities of authentication systems have been traditionally shown in the form of FA and FR Rates versus the decision threshold, as depicted in Fig. 4.1 for an ideal system (a), and a real system (b). Another commonly used graphical representation of the capabilities of an authentication system, specially useful when comparing multiple systems, is the ROC (Receiver -or also Relative- Operating Characteristic) plot, in which FA Rate (FAR) versus FR Rate (FRR) is depicted for variable decision threshold. A variant of the ROC curve, the so-called DET (Detection Error Tradeoff) plot [Martin *et al.*, 1997] uses a non-linear scale and makes the comparison of competing systems easier. A comparison between ROC and DET curves for two hypothetical competing authentication systems A and B is given in Fig. 4.2.

A specific point is attained when FAR and FRR coincide, the so-called EER (Equal Error Rate). The global EER of a system can be easily detected by the intersection between the DET curve of the system and the diagonal line $y = x$. Nevertheless, and because of the discrete nature of FAR and FRR plots, EER calculation may be ambiguous according to the above-mentioned definition, so an operational procedure for computing the EER must be followed. In the present contribution, the procedure for computing the EER described by Maio *et al.* [2002] has been applied.

In face recognition systems, the common graphical representation used is the VR-ROC plot, in which FAR versus Verification Rate, VR = (1 - FRR) is depicted to analyse the capabilities

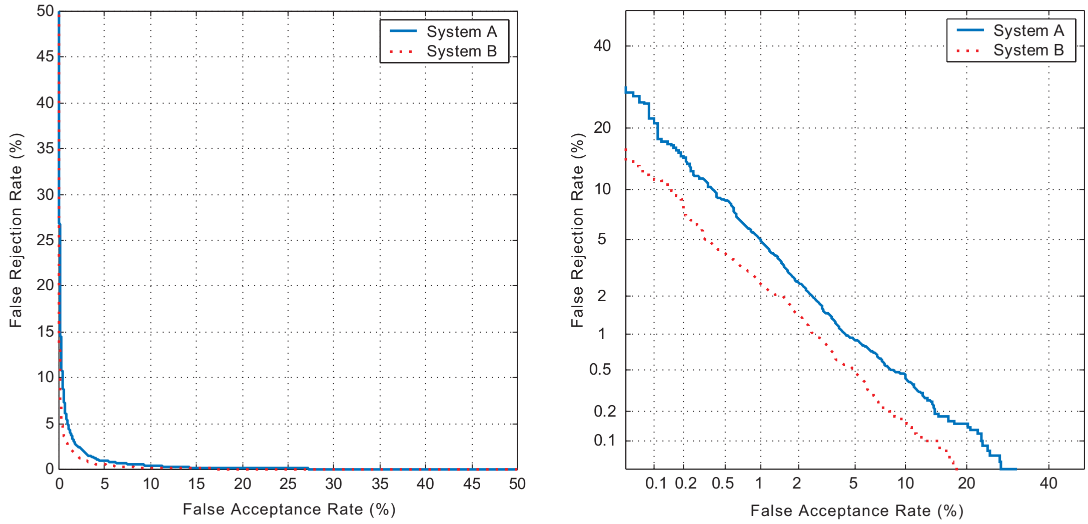


Figure 4.2: *Example of verification performance with ROC (left) and DET curves (right).*

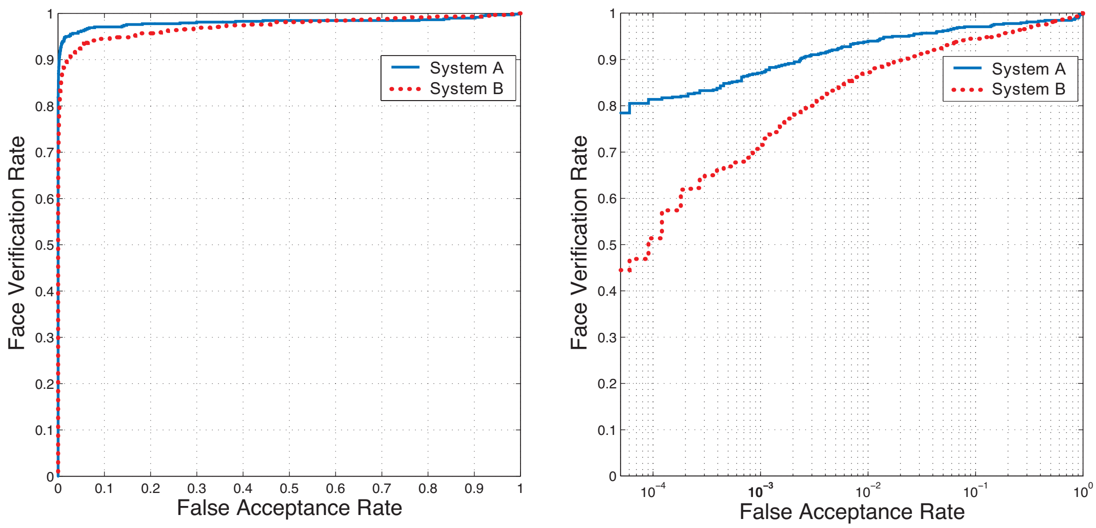


Figure 4.3: *Example of face verification performance with VR-ROC (left) and VR-DET curves (right).*

of an authentication system. The VR-ROC and a variant of the VR-ROC curve, the so-called VR-DET plot, are used in this Thesis [Phillips *et al.*, 2005]. In this case, the use of a non-linear scale (logarithmic scale at FAR axis) makes the comparison of competing systems easier at different FAR points. The state-of-the-art in face recognition is commonly working at FAR = 0.001 (10^{-3}). A comparison between VR-ROC and VR-DET curves for two hypothetical competing authentication systems A and B is given in Fig. 4.3.

4.2. Biometric Databases at a Distance

One key element for performance evaluation of biometric systems is the availability of biometric databases. Some relevant efforts in this regard have been directed to the acquisition of large multimodal (i.e., comprising different biometric traits of the same users) datasets [Fierrez *et al.*, 2009, 2007; Ortega-Garcia *et al.*, 2009]. Multimodal databases have the clear advantage over unimodal corpora of permitting to carry out research studies using individual or different combined traits (i.e., multibiometrics) [Fierrez-Aguilar *et al.*, 2005; Ross *et al.*, 2006]. This kind of databases are in general acquired in controlled conditions, however the advances in the last years are focused in real uncontrolled scenarios using biometrics at a distance and on the move [Phillips *et al.*, 2009a]. In this sense, the acquisition of biometric databases at a distance involves a low cost of resources (a webcam camera, is enough) making the database collection an easy process, in which a low degree of cooperation of the donors is needed. Additionally, acquisition at a distance involves the undesirable presence of variability factors (i.e., sensor, sensor-user interaction, user and system effects), making very challenging the authentication task. On the other hand, the legal issues regarding data protection are delicate [Flynn, 2007; Wayman *et al.*, 2005]. For these reasons, nowadays, the number of existing public biometric databases at a distance is quite limited.

The databases at a distance currently available have resulted from collaborative efforts in recent research projects. Examples of these joint efforts include European projects like M2VTS [Messer *et al.*, 1999] or BANCA [Bailly-Bailliere *et al.*, 2003], and other initiatives by NIST [National Institute of Standards and Technology (NIST)], with their series of biometrics challenges.

In the following sections we provide an overview of existing biometric databases at a distance, provide some information of current efforts in the acquisition of new corpora, and finally an extended description of the databases at a distance used in this PhD Thesis.

4.2.1. Existing Databases at a Distance

Most test databases for face recognition contain images or video captured at close range with cooperative subjects. They are thus best suited for training and testing face recognition systems for access control applications. However, there are a few datasets that are more suitable for face recognition at a distance development and evaluation.

- **UTD** [O'Toole *et al.*, 2005]. The database collected at the University of Texas at Dallas for the DARPA Human ID program includes close-up still images and video of subjects and also video of persons walking toward a still camera from distances of up to 13.6 m and video of persons talking and gesturing from approximately 8 m. The collection was performed indoors, but in a large open area with one wall made entirely of glass, approximating outdoor lighting conditions. A fairly low zoom factor was used in this collection. This is a database of static images and video clips of human faces and people that is useful for testing algorithms for face and person recognition, head/eye tracking, and computer graphics modelling of natural human motions. For each person there are nine static "facial mugshots" and a series of video streams. The videos include a "moving facial mugshot", a facial speech clip, one or more dynamic facial expression clips, two gait videos, and a conversation video taken at a moderate distance from the camera. Complete data sets are available for 284 subjects and duplicate data sets, taken subsequent to the original set, are available for 229 subjects.

- **UTK-LRHM** [Yao *et al.*, 2008]. The face video database UTK-LRHM was acquired from long distances and with high magnifications. Both indoor and outdoor sequences were collected under uncontrolled surveillance conditions. This was the first database to provide face images from long distances (indoor: 10-16 m and outdoor: 50-300 m). The corresponding system magnifications range from 3× to 20× for indoor and up to 284× for outdoor conditions. This database has applications in experimentations with human identification and authentication in long range surveillance and wide area monitoring.

- **GBU** [Phillips *et al.*, 2011]. The Good, the Bad, and the Ugly challenge consists of three frontal still face partitions. The paritions were designed to encourage the development of face recognition algorithms that excel at matching "hard" face pairs, but not at the expense of performance on "easy" face pairs. The images in this challenge problem are frontal face stills taken under uncontrolled illumination, both indoors and outdoors. The three partitions were constructed by analyzing results from the FRVT 2006. The Good set consisted of face pairs that had above average performance, the Bad set consisted of face pairs that had average performance, and the Ugly set consisted of face pairs that had below average performance. There are 437 subjects in the data set. All three partitions have the same 437 subjects. All three paritions have 1085 images in both the target and query sets.

- **ND-QO-Flip** [Barr *et al.*, 2011]. ND-QO-Flip Crowd Video Database contains 14 between 25-59 second crowd video clips of 90 subjects, five of whom appear in multiple videos and 85 of whom appear in one video. These videos were acquired between November 2009 and May 2010 (over a period of seven months) in different locations around the University of Notre Dame campus. In each clip, the camera pans and zooms over a crowd of four to 12 people. Most people are seen from a nearly frontal viewpoint because the observers tend to face toward the camera or focus on an object behind it. The crowd members were allowed to exhibit any facial expression they chose. The video set contains 12 outdoor videos, including six that were acquired under overcast conditions, six that were recorded when the sun was visible, three with snow cover and one with falling snow. The videos thus contain extensive variations in illumination and facial expression along with partial face occlusions caused by the way the crowds formed. The acquisition process is carried out using a Cisco Flip hand-held camcorder. All videos were compressed with the H.264 codec, have a 640 × 480 resolution and a frame rate of 30 frames per second.

- **LDHF-DB** [Maeng *et al.*, 2013]. Long Distance Heterogeneous Face Database (LDHF-DB) contains both visible (VIS) and near-infrared (NIR) face images at distances of 60 m, 100 m and 150 m outdoors and at a 1 m distance indoors. Face images of 100 subjects (70 males and 30 females) were captured; for each subject one image was captured at each distance in daytime and nighttime. All the images of individual subjects are frontal faces without glasses and collected in a single sitting.

- **PaSC** [Beveridge *et al.*, 2013]. Inexpensive point-and-shoot camera technology has combined with social network technology to give the general population a motivation to use face recognition technology. The Point-and-Shoot Face Recognition Challenge (PaSC) is a challenge which includes $9,376$ still images of 293 people balanced with respect to distance to the camera, alternative sensors, frontal versus non-frontal views, and varying location. There are also $2,802$ videos for 265 people: a subset of the 293. Verification results are also presented together with the database for public baseline algorithms and a commercial algorithm for three cases: comparing still images to still images, videos to videos, and still images to videos.

4.2.2. MBGC DB

The Multiple Biometric Grand Challenge (MBGC) [Phillips *et al.*, 2009a] is being conducted in two parts with two versions of the challenge problems. First version 1 introduced the participants to the challenge problems and the MBGC protocol. And the version 2 encouraged the development of algorithms that can handle large datasets.

The goal of the MBGC is to improve the performance of face and iris recognition technology from biometric samples acquired under unconstrained conditions. The MBGC is organized into three challenge problems. Each challenge problem relaxes the acquisition constraints in different directions. In the *Portal Challenge Problem*, the goal is to recognize people from near-infrared (NIR) and high definition (HD) video as they walk through a portal. Iris recognition can be performed from the NIR video and face recognition from the HD video. The availability of NIR and HD modalities allows for the development of fusion algorithms. The *Still Face Challenge Problem* has two primary goals. The first is to improve recognition performance from frontal and off angle still face images taken under uncontrolled indoor and outdoor lighting. The second is to improve recognition performance on still frontal face images that have been resized and compressed, as is required for electronic passports. In the *Video Challenge Problem*, the goal is to recognize people from video in unconstrained environments. The video is unconstrained in pose, illumination, and camera angle. All three challenge problems include a large data set, experiment descriptions, ground truth, and scoring code.

The Still Face Challenge Problem builds on the Face Recognition Grand Challenge (FRGC) adding non-frontal face images to the task and increasing the size and scope of experiments with images acquired under uncontrolled illumination. A survey of work addressing face recognition from unconstrained environments can be found elsewhere [Zhao *et al.*, 2003] and provides some

Figure 4.4: *MBGC database example images of indoor and outdoor conditions.*

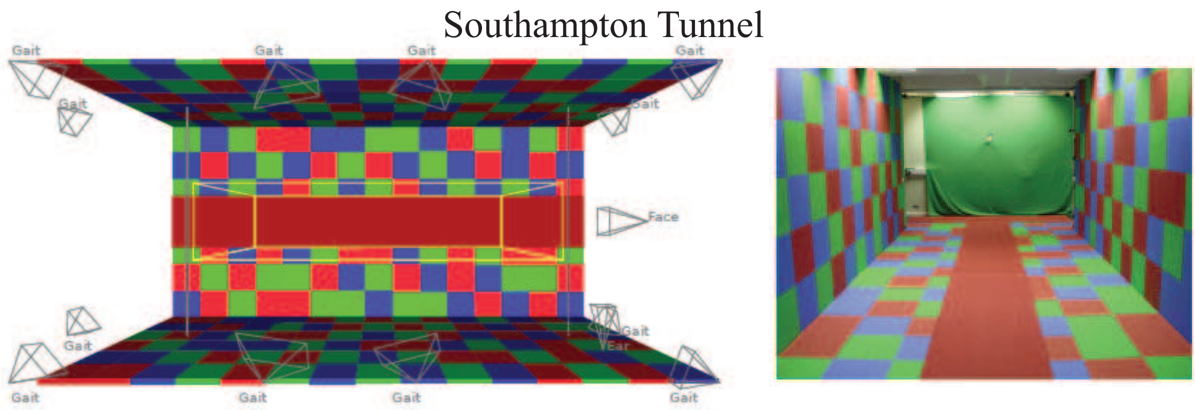


Figure 4.5: *Tunnel database setup. There are eight cameras acquiring gait signal and one high-resolution camera acquiring frontal people walking.*

baseline estimates of the state-of-the-art for this problem. The increase in the scope of the Still Face Challenge Problem includes recognition from highly compressed images.

The data for the Still Face Challenge Problem were collected with high resolution digital cameras, 4 and 6 mega-pixels. Images were collected with both controlled and uncontrolled illumination. The images with controlled illumination were collected in a studio environment with controlled lighting. The images with uncontrolled illumination were collected in hallways and outdoors. Fig. 4.4 shows an example of face images collected.

The Video Challenge Problem is the first NIST challenge problem to address face recognition from unconstrained video. It makes available information that is commonly included in surveillance video and so provides an estimate of the recognition performance that can be achieved in security applications where video is available.

4.2.3. Tunnel DB

The new Southampton Multibiometric Tunnel Database (TunnelDB) [Seely *et al.*, 2008] contains biometric samples of 227 subjects for which 10 gait sample videos from between 8 to 12 viewpoints are taken simultaneously and stored to extract 3D gait information. TunnelDB also contains high resolution frontal videos to extract face information and high resolution side face images taken to extract ear biometrics. There are roughly 10 such sets of information gathered for each subject in TunnelDB.

Of the 227 subjects, 67% were male; the majority were aged between 18-28 years old and 70% were of European origin. These biases in the demographic of the dataset were expected, as this closely represents the student population.

The acquisition process is shown in Figs. 4.5 and 4.6, where subjects were collected walking through an entry beam on a straight red path towards the exit beam and therefore towards a face camera. During a single walk (a sample), the subject was simultaneously captured by the gait cameras and the face camera. Upon reaching the exit beam, a single flash camera was used to photograph the right ear.

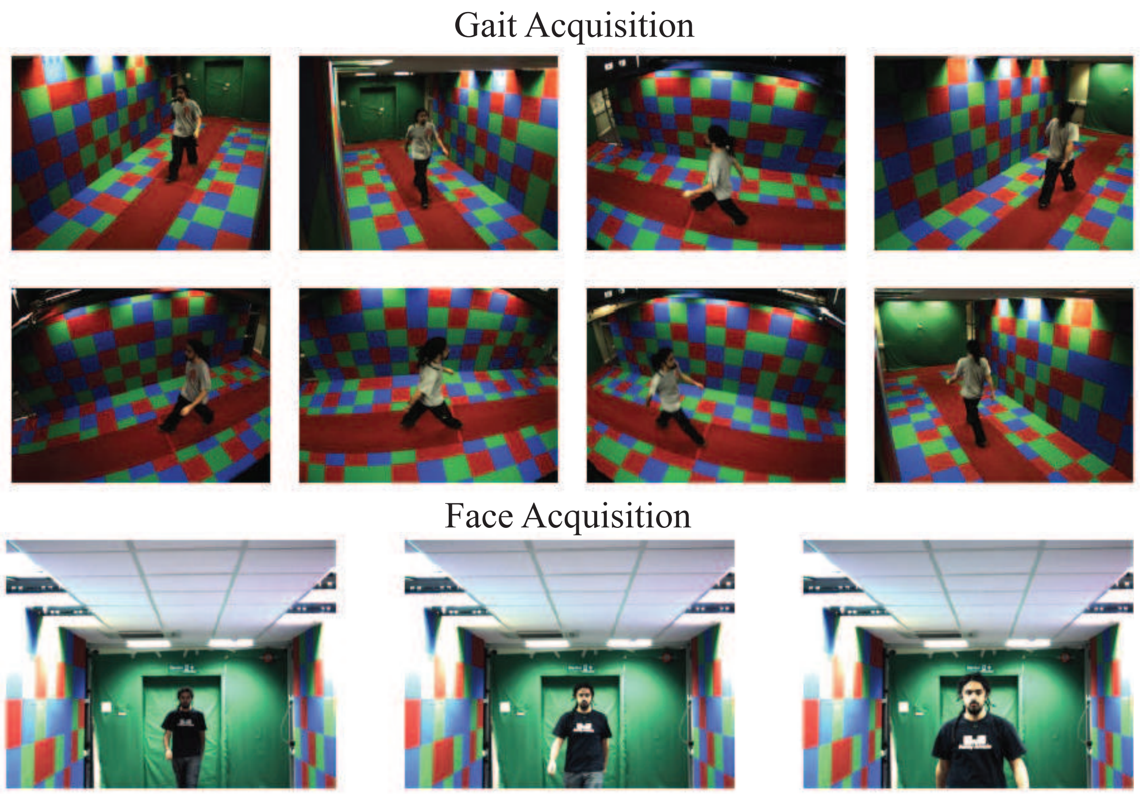


Figure 4.6: *Tunnel database samples.*

The gait information is recorded by 12 cameras, and while gait images were taken, a single higher definition camera at the end of the tunnel captures a 1600×1200 high resolution face images at 27 frames per second. In the tunnel scenario, direction of gaze was guaranteed by instructing the subjects as well as by their walking direction. Lighting and other environmental variables were also controlled.

Figs. 4.5 and 4.6 shows the acquisition setup, together with an example of each biometric trait captured per subject.

4.2.4. SCface DB

Surveillance Cameras Face Database (SCface) [Grgic *et al.*, 2011] is a database of static images of human faces. Images were taken in uncontrolled indoor environment using five video surveillance cameras of various qualities. The database contains 4160 static images (in visible and infrared spectrum) of 130 subjects. Images from different quality cameras mimic the real-world conditions and enable robust face recognition algorithms testing, emphasizing different law enforcement and surveillance use case scenarios. SCface database is freely available to the research community. Subjects' images were taken at three distinct distances from the cameras with the outdoor (sun) light as the only source of illumination. All images were collected over a 5 day period.

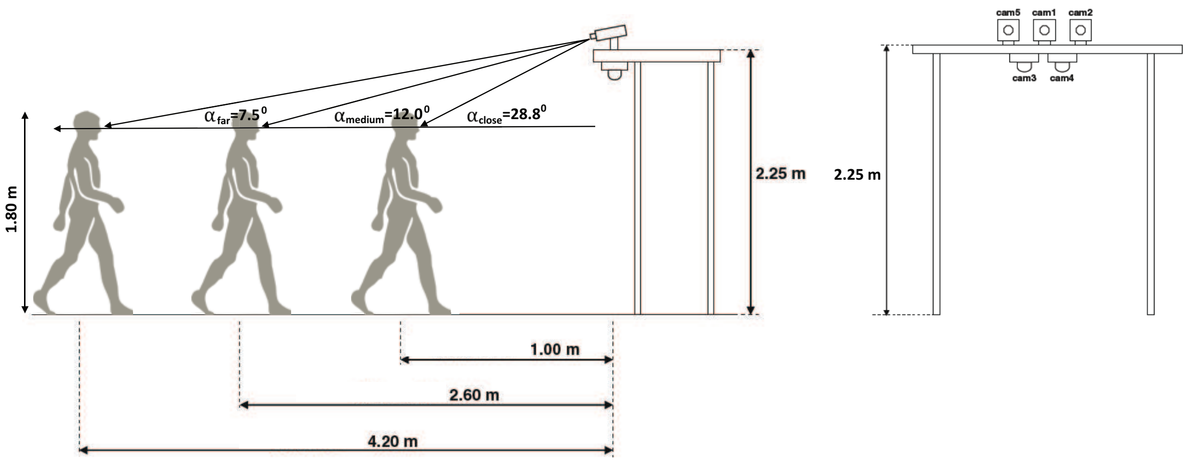


Figure 4.7: *SCface database. There are three different acquisitions distances: close, medium and far. Acquisition angle of each distance calculated for a subject with mean height of 1.80 meters.*

Here is a short summary of what makes this dataset interesting for the face recognition research community:

1. Different quality and resolution cameras were used.
2. Images were taken under uncontrolled illumination conditions
3. Images were taken from various distances
4. Head pose variation in surveillance images is typical for a commercial surveillance system, i.e. the camera is placed slightly above the subject's head, making the recognition even more demanding. Besides, during the surveillance camera recordings the individuals were not looking to a fixed point.
5. Database contains nine different poses suitable for head pose modelling and/or estimation.
6. Database contains images of 130 subjects, enough to eliminate performance results obtained by pure coincidence.
7. Both identification and verification scenarios are possible, but the main idea is for it to be used in difficult real-world identification experiments.

This database used only as source of illumination the outdoor light, which came through a window on one side. Two (out of five) surveillance cameras were able to record in IR night vision mode as well. The sixth camera was installed in a separate, darkened room for capturing IR mug shots. The high-quality photo camera for capturing visible light mugshots was installed the same way as the infrared camera but in a separate room with the standard indoor lighting and it was equipped with adequate flash. Mugshot imaging conditions are exactly the same as would be expected for any law enforcement or national security use (passport images or any other personal identification document). All six cameras (five surveillance and one IR mug shot)

Figure 4.8: *SCface image samples of each dataset for CCTV, mugshot, and IR scenarios.*

were connected to a professional digital video surveillance recorder, which was recording all six video streams simultaneously all the time on internal hard disk.

They used 5 surveillance cameras named cam1, cam2, cam3, cam4 and cam5. Cam1 and cam5 are also able to work in IR night vision mode. They decided to name the images taken by them in the IR night vision mode as cam6 (actually cam1 in night vision mode) and cam7 (actually cam5 in night vision mode). Camera for taking IR mugshots was named cam8. All cameras (surveillance and photo) were installed and fixed to the same positions and were not moved during the whole capturing process.

Cameras 1-5 are visible light cameras and generate images of different qualities. There are three images per subject for each camera, taken at three discrete distances (4.20, 2.60 and 1.00 m), see Fig. 4.7. This gives a total of 15 images per subject in this set (1950 in total). This set was designed to test the face recognition algorithms in real-world surveillance setup. As can be seen from the Fig. 4.8, the images differ substantially in quality and resolution. IR night vision mug shots were taken in a separate dark room with a resolution of 426×320 pixels, grayscale. There is one image per subject in this set, yielding a total of 130 of those images in the database.

All participants in this project passed through the following procedure. First they had to walk in front of the surveillance cameras in the dark and after that they had to do the same in uncontrolled indoor lighting. During their walk in front of the cameras they had to stop at three previously marked positions. This way 21 images per subject were taken (cam1-7 at distances of 4.20, 2.60 and 1.00 meters). In the end, subjects went into the dark room where the high quality IR night vision surveillance camera was installed for capturing IR mug shots at close range. In overall that gives 32 images per subject in the database.

The participants in this project were students, professors or employees at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. From the total of 130 volunteers, 115 were males and 15 females. All participants were Caucasians, between the ages of 20 and 75.

4.2.5. ATVS Forensic DB

The ATVS Forensic Database [Vera-Rodriguez *et al.*, 2013a] was acquired in collaboration with the Spanish Guardia Civil (DGGC) by the Universidad Autonoma de Madrid (UAM). The main objective of the project was the acquisition of a realistic mugshot forensic database. Its main characteristics are as follows:

1. **Number of subjects**: a total of 50 users (32 men and 18 women) were acquired.

2. **Number of samples**: 4 acquisition sets capturing 5 different mugshot images in each set.

3. **Number of sessions**: 2 sessions distributed in a 5 month time span.

In Table 4.1 the most relevant statistics of the ATVS Forensic database are shown.

	ATVS Forensic DB. 50 subjects
Gender Distribution	32 (Male) / 18 (Female)
Age Distribution	(18–35)
Vision Aids	100% (None)

Table 4.1: *Statistics of the ATVS Forensic database.*

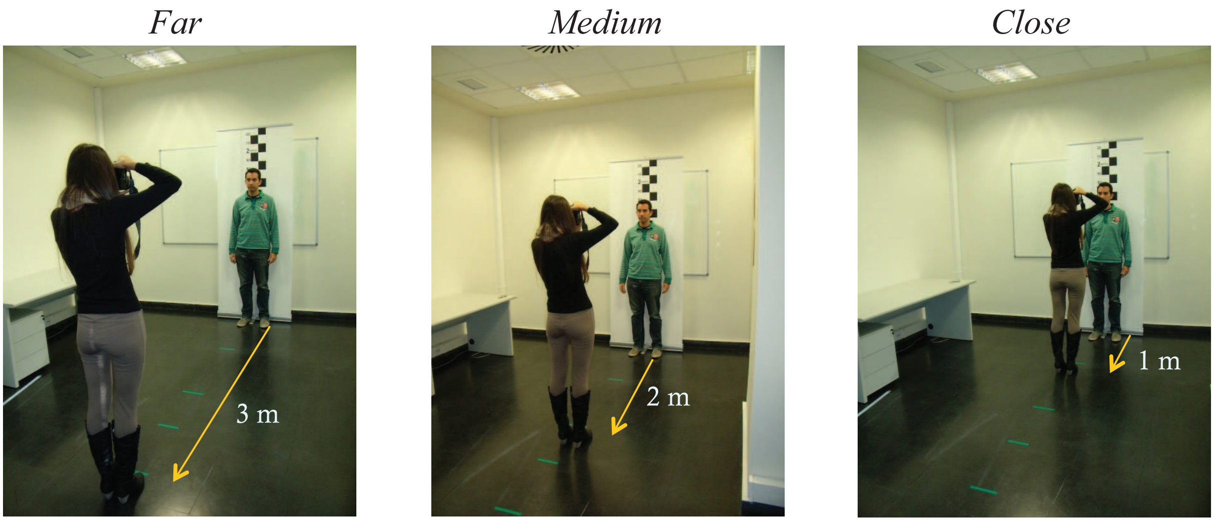


Figure 4.9: *Example setup used and the process followed in the acquisition of the ATVS Forensic database.*

4.2.5.1. Acquisition Environment

The 3 acquisition distances follow the same very general indications about the environmental conditions, regarding illumination (neutral lighting with no preponderant focuses), noise (indoor conditions with no excessive background noise), and pose of the contributor (stand up close to a vertical reference scale). These relaxed environmental conditions allow a desirable variability between the samples acquired (e.g., background in facial images) which simulates the changing working conditions of a real-world biometric application. In Fig. 4.9 we show the acquisition setup prepared, together with an example of the capture process followed in the acquisition.

During the acquisition procedure a human operator gave the necessary instructions to the contributors so that the acquisition protocol was followed. In order to ensure that the ATVS Forensic database complies with the acquisition protocol, all biometric samples were manually verified by a human expert who either corrected or discarded non-valid data.

The camera used in the database collection was a Canon EOS 400D and, as a can be seen in the Fig. 4.9, a vertical scale was used in the background, as it is done in current DGGC forensic practice.

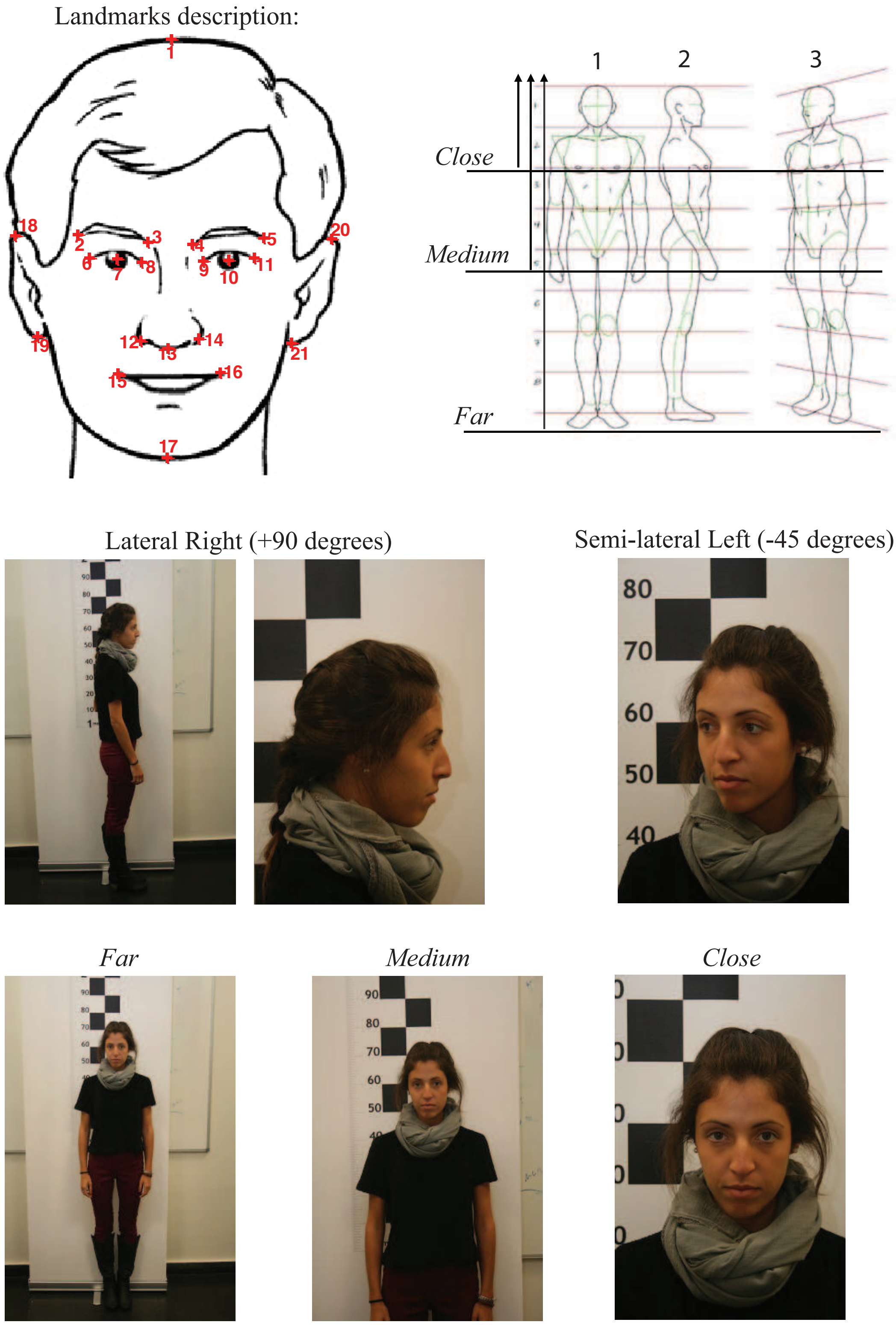


Figure 4.10: *ATVS Forensic database image samples of each dataset for mugshot* close, medium, *and* far *distances, lateral right (+90 degrees), and semi-lateral left (−45 degrees) images. Facial landmarks provided together with the mugshot frontal images are also shown on the top.*

4.2.5.2. Acquisition Protocol

The acquisition protocol followed for the ATVS Forensic database imitates current practice by the Spanish Guardia Civil. Six pictures of each person were taken at every mugshot photo session:

1. Full body. 2 pictures: front and lateral right view. Approximately three meters distance from camera to person.

2. Upper body. 1 picture: front view. Approximately two meters distance from camera to person.

3. Face. 3 pictures: front, lateral right (+90 degrees) and semi-lateral left (-45 degrees). Approximately one meter distance from camera to person.

In this Thesis, only the three frontal images (full body, upper body and face) have been used in the experimental Chapter 5 of this Dissertation. In Fig. 4.10 we summarize the data samples captured for every user.

A manual facial landmark tagging of the different users in the database is also provided. The first step after database collection was to define a set of facial landmarks to include in this database. A set of 21 facial landmarks was defined following recommendations from the Spanish Guardia Civil (DGGC), Netherlands Forensic Institute (NFI) and ENFSI [European Network of Forensic Science Institutes], including the irises (2 landmarks) , inner and outer eye corners (4), eyebrow ends (4), mouth corners (2), nose corners (2), center of the nose (1), chin (1), upper and lower ears ends (4) and highest point on the head (1). Figure 3 shows the 21 facial landmarks considered in this database (see Fig. 4.10 top). Therefore, a manual tagging of the 21 landmarks for the whole database was carried out by the same person, imitating the work of a forensic examiner. This manual tagging was performed over the acquired mugshot images.

The database comprises data from 50 persons (32 men and 18 women) acquired in two different sessions. The collection process took place from July to November 2012. The sessions were collected in different days for the same persons. In each session the procedure was repeated four times. Therefore, obtaining a total of 2400 mugshot images (50 persons $\times$ 2 sessions $\times$ 4 times $\times$ 6 images).

4.2.5.3. Potential Uses of the Database

Several potential uses of the database have already been pointed out throughout this Thesis. In this section some of the research lines that can be further developed upon this data set are summarized. It has to be emphasized that due to its characteristics in terms of acquisition environment (reproducing real forensic protocols) and demographic distribution, the ATVS Forensic database represents a good benchmark for the studied the variability factors presents in this kind of scenarios and their possible solutions.

Figure 4.11: *MORPH database image samples of the subset European.*

4.3. Other Databases

This section describes other database used in this PhD Thesis, which is classified only in the close distance acquisition.

4.3.1. MORPH DB

The Craniofacial Longitudinal Morphological Face Database (MORPH) [Ricanek and Tesafaye, 2006] contains 55.000 frontal face images from more than 13.000 subjects, acquired from 2003 to late 2007. The distribution of ages ranges from 16 to 77 with an average age of 33. The average number of images per individual is 4 and the average time between pictures is 164 days, with the minimum being 1 day and the maximum being 1.681 days. The MORPH database is divided in 5 subsets named: *i*) African, *ii*) European, *iii*) Asian, *iv*) Hispanic and *v*) Other.

The subset "European" comprises 2.704 subjects (2.070 males plus 634 females) and has been selected for the experiments in this Thesis. Fig. 4.11 shows an example of this European subset. The detailed description of the evaluation of these databases can be found, respectively, in Chapters 5, 6 and 7 of the present Dissertation.

4.4. Chapter Summary and Conclusions

In this chapter we have outlined some best practices for performance evaluation in biometric authentication. We have also provided a description of the evaluation protocol followed in this Thesis, which can serve as guideline to carry out systematic and replicable studies of biometrics at a distance. Finally we have given an overview of the main existing biometric databases at a distance and we have described the most important ones used in this Thesis: MBGC, TunnelDB, SCface, the developed ATVS Forensic DB, and MORPH database.

This chapter includes novel contributions in the survey of the most relevant biometric databases at a distance, and in the description of the new corpus ATVS Forensic DB.

Chapter 5

Scenario Analysis

THIS CHAPTER studies the variability in scenarios at a distance and its effect on the system performance at different acquisition distances, with special focus in surveillance and forensic scenarios. The variability of facial landmarks is analysed in forensic scenarios. The discriminative power of different facial regions of the human face on various forensic scenarios is also evaluated.

As indicated in Chapter 2, the variability factors of a biometric system at a distance can be broadly divided into *user*-related, *sensor*-related, *user-sensor interaction*, and *system* factors, being the first the ones most difficult to control, and the *user-sensor interaction* factors the ones that affect the most in systems at a distance. It would be desirable for a compensation variability method to identify the acquisition distance between the subject and the camera in order to compensate the variability factors that are severely affecting. The first step towards this purpose is a better understanding of the scenario at hand. After a first general study in this regard, with application to surveillance and forensics, the rest of the chapter will be focused in forensic face recognition.

Large amounts of research are being carried out trying to compensate variability sources (such as illumination, pose, facial expressions, occlusions, etc.) that affect significantly reducing the performance of the face recognition systems. In a forensic scenario, these variability factors are crucial, because forensic examiners have to deal with face images extracted from CCTV cameras and other low quality sources, which make the task really difficult. Among the tasks carried out by forensic examiners, in an anthropometric analysis they extract manually a set of facial landmarks, then compute some distances between them, which can be used as features in their analysis. The variability of these facial landmarks extracted by forensic experts (*system*-related variability factors) has not been widely studied and must be considered in real caseworks.

Additionally, automatic face recognition systems are generally designed to match images of full faces. However, in practice, forensic examiners carry out a manual inspection of the face images, focussing their attention not only on the full face but also on individual traits. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc. In this sense, a study of the discriminative power of different facial regions individually finding the most discriminative areas of the face for recognition on different acquisition scenarios will have remarkable benefits.

The chapter is structured as follows. One section is dedicated to each of the three different studies: *i*) quantitative study of scenario at a distance, analysing the information content in segmented faces and the effect of the different acquisition distances on the performance of face verification (Sect. 5.1), *ii*) analysis of the variability of facial landmarks in a forensic scenario over a database of mugshots and CCTV images (Sect. 5.2), and *iii*) exhaustive analysis of the discriminative power of the different regions of the human face on various forensic scenarios (Sect. 5.3). Each of these sections share a common structure where the scenario to analyse is first described, then the database and experimental protocol, followed by the recognition systems being evaluated are presented, and finally the results of the evaluation are given and analysed. Finally, the chapter summary and conclusions are given in Sect. 5.4.

This chapter is based on the publications: Tome *et al.* [2013a, 2010b, 2013e]; Vera-Rodriguez *et al.* [2013a,b].

5.1. Scenario Analysis for Face Recognition at a Distance

A new research line growing in popularity is focused on using biometrics for person recognition in less constrained scenarios in a non-intrusive way, including acquisition "On the Move" and "At a Distance". The biometric technologies are still in their infancy when considering uncontrolled scenarios, and much research and development is needed in order to achieve the levels of precision and performance that these challenging conditions require. In this Section, an experimental analysis of three acquisition scenarios for face recognition at a distance is reported, namely: close, medium, and far distance between camera and query face, the three of them considering templates enrolled in controlled conditions. These three representative scenarios are studied using data from the NIST - Multiple Biometric Grand Challenge [MBGC], as the first step in order to understand the main variability factors that affect face recognition at a distance based on realistic yet workable and widely available data.

The scenario analysis is conducted quantitatively in two ways. First, we analyse the information content in segmented faces in the different scenarios. Second, the effect of different acquisition distances on the performance of face verification is studied. In particular, we evaluate two standard face recognition approaches using popular features (PCA and DCT) and matchers (SVM and GMM) under variation in the acquisition distance. The DCT-GMM-based system is found to be more robust to acquisition distance degradation than the PCA-SVM-based system.

5.1.1. Database and Scenario Definition

The scenarios under study are extracted from the NIST Multiple Biometric Grand Challenge [MBGC], which is focused on biometric recognition at a distance using iris and face. In particular, we use a subset of this benchmark dataset consisting of images from a total of 147

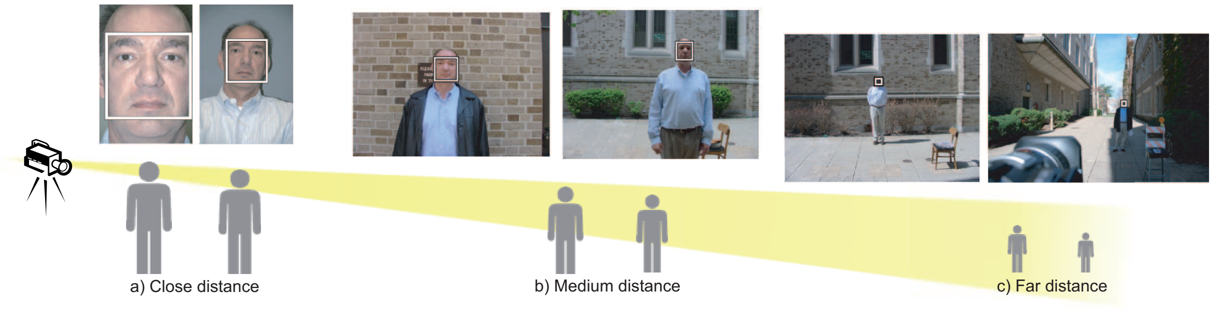


Figure 5.1: *Example images of the three scenarios: a) close distance, b) medium distance, and c) far distance.*

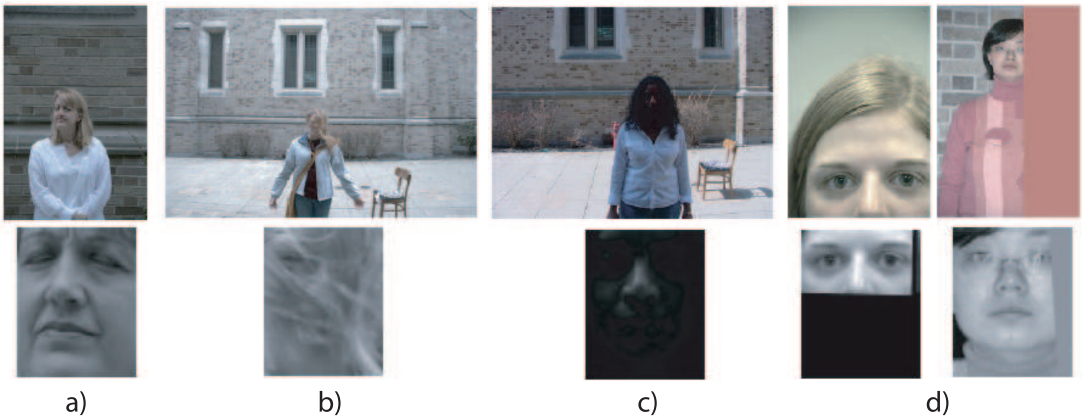


Figure 5.2: *Example images of the different cases of missing values: a) eyes closed, b) face occluded, c) low illumination and d) missing parts of the face.*

subjects and 3482 images acquired at different distances and varying conditions regarding illumination, pose/angle of head, and facial expression, some of them acquired in controlled conditions and others in uncontrolled environments.

Fig. 5.1 shows some examples of different kinds of face images in this dataset where we can identify three big groups:

1. *Close* distance, in which the shoulders may be present (controlled conditions).
2. *Medium* distance, including the upper body (uncontrolled conditions).
3. *Far* distance, including the full body (uncontrolled conditions).

Using these three general definitions we tagged manually all the 3482 face images from the 147 subjects present in the dataset NIST MBGC v2.0 Face Stills [MBGC]. Furthermore, although this information is not used in the present Thesis, all the images were marked as indoor or outdoor in order to carry out future studies of the effects in indoor versus outdoor conditions. Sect. 4.2.2 describes in detail the MBGC database and Fig. 4.4 shows some samples of the scenarios available in dataset.

Figure 5.3: *Distribution of samples per user of the three scenarios defined.*

In this first stage during the manual marking we found some images unusable, which we marked as *missing values*. A portion of the dataset was discarded (360 images from 89 subjects), following the next four conditions: a) closed eyes completely; b) hair or other artefacts occluded the face; c) the illumination completely degraded the face; and d) faces with unknown parts or not present in the image. Some examples of these problems are shown in Fig. 5.2.

After this marking stage, the final result is shown in Fig. 5.3 where we can see the number of images per user for each subset. As we can observe the dataset is composed by a bigger number of close distance images and more or less the same quantity of images in the two other scenarios. In both cases images are equally distributed among users.

Observing Fig. 5.3 we have defined two different partitions of the subset in order to carry out two different studies: *i*) case study 1, represents the situation where the subject carries out a controlled enrollment and is recognized in controlled/uncontrolled environments; *ii*) case study 2, studies uncontrolled enrollment and controlled/uncontrolled recognition.

5.1.1.1. Case Study 1: Controlled Enrollment

In this case study we have tried to simulate a real situation where the subject is enrolled in the system in a controlled manner (close images, relatively controlled illumination, neutral pose, uniform background, etc.) with good quality images. After the subject is enrolled in the system, we have defined situations of recognition with different distances emulating e.g. a control access system (close distance), an easy surveillance system (medium distance) and a remote surveillance system (far distance) where variability factors are uncontrolled.

Fig. 5.4 summarizes the users chosen and discarded after the preprocessing stage that will be used for the statistical analysis and experiments. In order to enable verification experiments considering enrollment at close distance and testing at close, medium, and far distance scenarios, we kept only the subjects with at least 2 images in close and at least 1 image in both of the

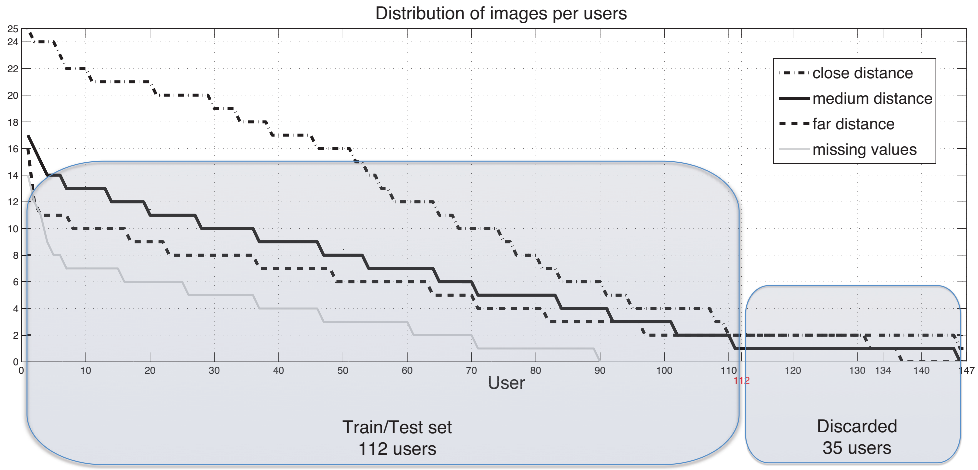


Figure 5.4: *Distribution of images per users for the three scenarios defined with the division carried out for case study 1.*

Num. users	Close distance	Medium distance	Far distance	Discarded images	Total
147	1539	870	713	360	**3482**
	At least 2 images per user	*At least 1 image per user*			
112	1468	836	660		**2964**

Table 5.1: *Number of images of each scenario constructed from NIST MBGC v2.0 Face Visible Stills for case study 1.*

two other scenarios. The data selection process is summarized in Table 5.1, where we can see that the three considered scenarios result in 112 subjects and 2964 face images (1468 close, 836 medium and 660 far distance images).

5.1.1.2. Case Study 2: Uncontrolled Enrollment

In this second case study, we have focused our efforts in the analysis of behaviour of face recognition systems in conditions where the enrollment is uncontrolled, e.g., in cases where the enrollment has to be performed at a distance. We want to study the effects on the system performance in this kind of situations.

We have also defined a new partition called *mix* comprised of the sum of the three acquisition distance datasets. This is designed to study the effects of combining several kinds of information (training with different acquisition distances).

The Fig. 5.5 summarizes the users chosen for development and evaluation and discarded users. In this case only subjects with at least 4 images were kept in each scenario considered. As before, a portion of the dataset was discarded: missing values (360 images from 89 subjects), and a reduced number of subjects (13) were completely discarded (those that had less than 4 image per scenario) discarding a total 403 images of the whole dataset. The data selection process

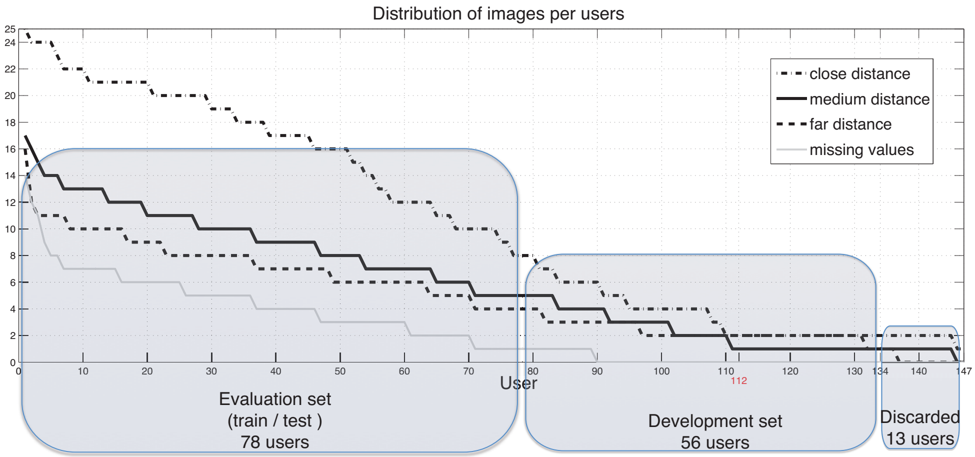


Figure 5.5: *Distribution of images per user of the three scenarios defined with the division carried out for case study 2.*

	Development	Evaluation	Discarded	Total
# users	56	78	13	147
Condition: at least 4 images per scenario: 2 train and 2 test.				
# images	484	2595	403	3482

Table 5.2: *Number of users and images of NIST MBGC v2.0 Face Stills dataset used.*

Development Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	222	132	130	484

Evaluation Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	1290	727	578	2595
Train	661	386	304	1351
Test	629	341	274	1244

Table 5.3: *Configuration of the datasets (close, medium, far and mix combination of all of them) for each acquisition scenario.*

is summarized in Tables 5.2 and 5.3, where we can see that the two considered sub-corpora result in 134 subjects, using 484 images of 56 subjects for the development of the systems and 2595 images of 78 subjects for the evaluation.

5.1.2. Scenario Analysis

This section describes the analysis, from a statistical and an applied point of view, of the datasets trying to study the differences among then related to the acquisition at a distance. Our study is divided in three levels: segmentation, quality and content information aspects.

Type	Image sizes (Pixels)	Resolutions (Mpx)	Num. Images
Close distance	1000 × 1504	1.4688	88
	3039 × 2014	5.9771	425
	2592 × 3872	9.8010	88
	3904 × 2616	9.9735	867
			1468 images
Medium distance	1200 × 1600	1.8750	474
	2592 × 3872	5.9771	4
	2616 × 3904	9.8010	345
	3039 × 2014	9.9735	13
			836 images
Far distance	1200 × 1600	1.8750	315
	2592 × 3872	9.8010	177
	2616 × 3904	9.9735	168
			660 images

Table 5.4: *Sub-corpus description of each kind of images and resolutions available in the database.*

	Close distance	Medium distance	Far distance	Discarded	Total
Num. Images	1468	836	660	360	**3324**
Errors	21	151	545		**848**
Errors(%)	1.43%	18.06%	**82.57%**		

Table 5.5: *Segmentation results based on errors produced by the face extractor of VeriLook SDK.*

First of all, we analyse the distribution of segmented face sizes per set constructed. As we can see in Table 5.4 the resolutions available for each set are variable: {1.5, 2, 6 and 10 Mpx.} except for far distance set where we only have two of these. It is important to note that the close distance set has high quality/resolution images (most of them 10 Mpx.) and the far distance set is balanced between 2 and 10 Mpx.

5.1.2.1. Face Segmentation and Resolution

We first segmented and localized the faces (square areas) in the three acquisition scenarios using VeriLook SDK as discussed in Sect. 5.1.3. Segmentation results are shown in Table 5.5, where the segmentation errors increase significantly across scenarios, from only 1.43% in close distance to 82.57% in far distance. Segmentation errors here mean that VeriLook software could not find a face in the image. For all the faces detected by VeriLook, we conducted a visual check, where we observed 3 and 10 segmentation errors for medium and far distance, respectively.

All the segmentation errors were then manually corrected by manually marking the eyes. The face area was then estimated based on the marked distance between eyes.

The resulting sizes of the segmented faces are shown in Fig. 5.6, where we observe to what extent the face size decreases with the acquisition distance. In particular, the average face size in pixels for each scenario is: 988 × 988 for close, 261 × 261 for medium, and 78 × 78 for far distance.

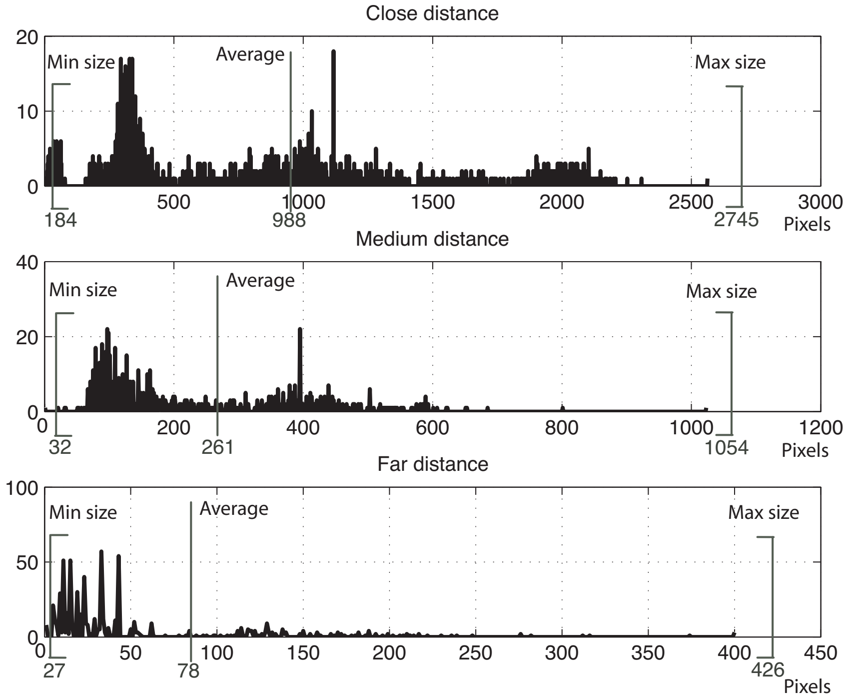


Figure 5.6: *Histograms of face sizes for each scenario (side of the square area in pixels).*

5.1.2.2. Quality

Another data statistic we computed for the three scenarios is the average face quality index provided by VeriLook (0 = lowest, 100 = highest): 73.93 for close, 68.77 for medium, and 66.50 for far distance (see Fig. 5.7, computed only for the faces correctly segmented by VeriLook). As stated by VeriLook providers, this quality index considers factors such as lightning, pose, and expression.

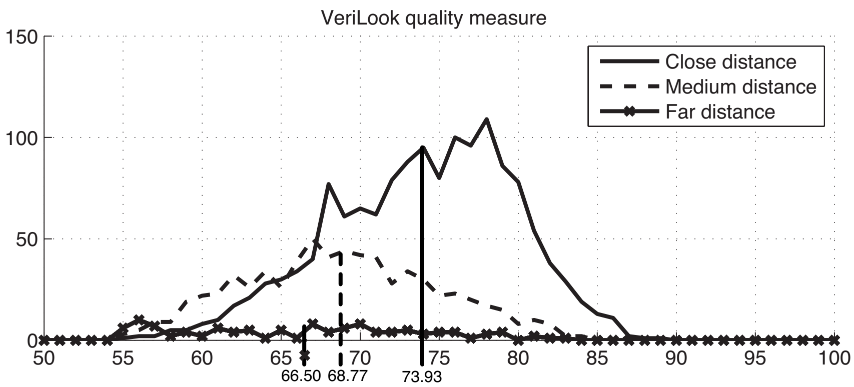


Figure 5.7: *Histogram of face quality measures produced by VeriLook SDK.*

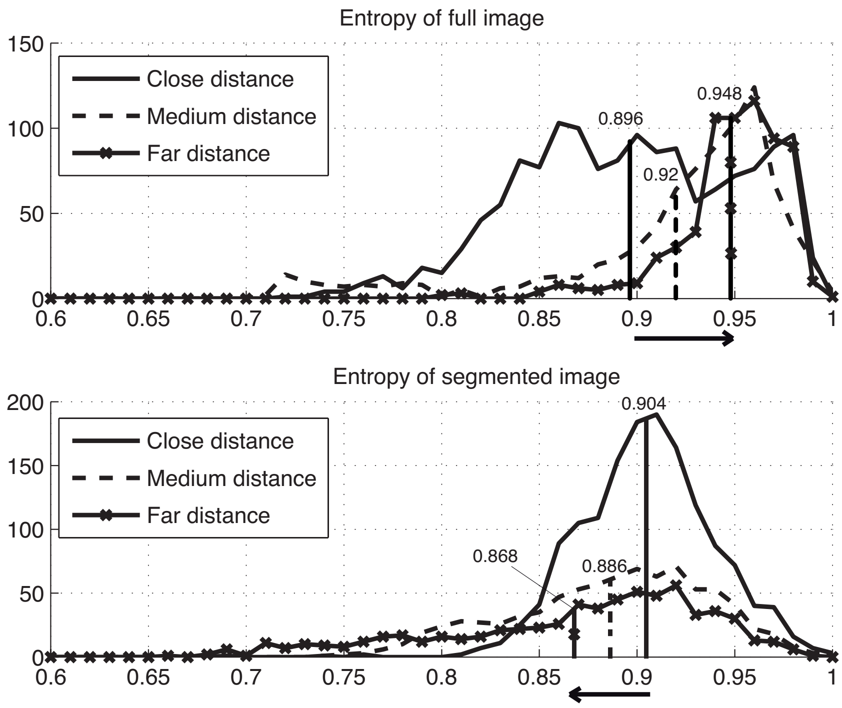


Figure 5.8: *Histograms of entropy for full images (top) and segmented faces (bottom) for the three scenarios with their corresponding average value.*

5.1.2.3. Information Content

In information theory, entropy is the measure of the amount of information that is missing before reception and is sometimes referred to as Shannon entropy. The definition of the information entropy is, however, quite general and is expressed in terms of a discrete set of probabilities p_i as we can see in Eq. (5.1):

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i) \tag{5.1}$$

Applying this concept to image processing, the entropy of the face images in the different acquisition scenarios represents a quantitative assessment of the information content in the gray levels of the images. In principle, an image acquired in controlled conditions (illumination, clean background, neutral pose, ...) would have less entropy than other image acquired at a distance in uncontrolled conditions.

In Fig. 5.8 (top), this effect is patent: the farther the distance the higher the entropy. When considering only the information within the segmented faces, as shown in Fig. 5.8 (bottom), the opposite occurs: the farther the distance the lower the entropy.

These two measures (increment in entropy of the full image, and decrement in entropy of the segmented faces), can therefore be seen, respectively, as a quantitative measure of how the scenario complexity increases due to background effects, and the reduction in information within the region of interest due to the change on the acquisition scenario.

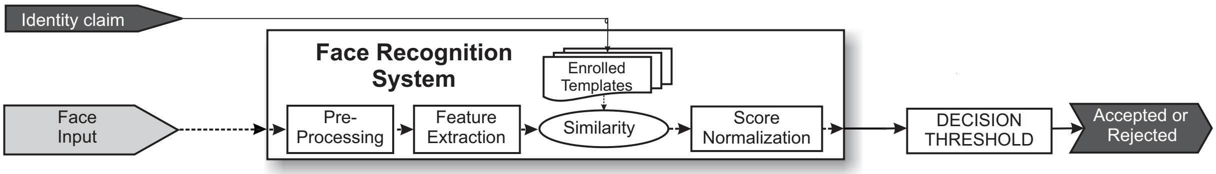


Figure 5.9: *Generic scheme of a face recognition system.*

5.1.3. Face Verification Systems

The presented systems are: *i*) a commercial system *VeriLook*, System Development Kit provided by Neurotechnology which will be used as reference system of the state of the art, *ii*) *PCA-SVM*, a system based on Principal Component Analysis (PCA) [Turk and Pentland, 1991b] and Support Vectors Machines (SVMs) [Cortes and Vapnik, 1995], and *iii*) *DCT-GMM*, a system based on Discrete Cosine Transforms (DCT) in combination with Gaussian Mixture Models (GMM) with a part-based representation of the face [Cardinaux *et al.*, 2003]. The three the systems follow the processing stages depicted in Fig. 5.9.

5.1.3.1. VeriLook SDK

VeriLook facial identification technology is intended for biometric systems developers and integrators. This is the commercial face recognition system provided by Neurotechnology [1]. This technology assures system performance and reliability with live face detection, simultaneous multiple face recognition and fast face matching in 1-to-1 and 1-to-many modes.

This system is available as a software development kit (SDK) that allows development of PC- and Web-based solutions on Microsoft Windows, Linux and Mac OS X platforms. It provides a toolkit with extractor and matcher module for face segmentation and verification. The system also provides a quality measure based on pose, expression, lighting and changes of the face. The extractors can be used with face images from cameras and/or files. The use of this system gives us an idea of the performance and robustness of commercial applications.

5.1.3.2. PCA-SVM System

This second face verification system used for performance evaluation is based on the well known eigenfaces technique introduced by Turk and Pentland [1991a]. This algorithm applies eigen-decomposition to the covariance matrix of a set of M vectorised training images $\overline{x_i}$. In statistical pattern recognition this technique is referred to as PCA [Fukunaga, 1990]. This method has become a *de facto* standard for face verification and was used to present initial results for the Face Recognition Grand Challenge evaluation [Phillips *et al.*, 2005].

The first similarity measure used to compare PCA-based features was the Euclidean distance, however several other similarity measures have been later proposed and studied [Perlibakas, 2004].

[1] http://www.neurotechnology.com/

The evaluated system uses normalized and cropped face images of size 64 $\times$ 80 (width $\times$ height), to train a PCA vector space where 96% of the variance is retained. This leads to a system where the original image space of 5120 dimensions is reduced to 249 dimensions ($K = 249$). Similarity scores are computed in this PCA vector space using a SVM classifier with linear kernel [Cortes and Vapnik, 1995].

5.1.3.3. DCT-GMM System

The GMM parts-based system used in the evaluation divides the 64 $\times$ 80 face images into 8 $\times$ 8 blocks with horizontal and vertical overlap of 4 pixels. This tessellation process results in 285 blocks and from each block a feature vector is obtained by applying the Discrete Cosine Transform (DCT); from the possible 64 DCT coefficients only the first 15 coefficients are retained ($N = 15$). The blocks are used to derive a world GMM Ω_w and a client GMM Ω_c [Cardinaux *et al.*, 2003]. From previous experiments we obtained that using ($M = 1024$) mixture components per GMM gave the best results.

When performing a query, or match, the averages score of the 285 blocks from the input image are used. The DCT feature vector from each block v_i (where $i = 1...285$) is matched to both Ω_w and Ω_c to produce a log-likelihood score. These scores are then combined using the log-likelihood ratio, $S_{llr,j} = \log\left[P\left(v_j|\Omega_c\right)\right] - \log\left[P\left(v_j|\Omega_w\right)\right]$, and the average of these scores is used as the final score, $S_{GMM} = \frac{1}{285}\sum_{j=1}^{285} S_{llr,j}$. This means that the query template can be considered to be a feature matrix formed by 285 fifteen dimensional vectors (representing each of the blocks in the image).

5.1.4. Experimental Protocol

The experimental framework followed is covered in next sections. Here, we have studied the effect of training and testing with images acquired at different distances using a face commercial system (VeriLook SDK) and two classical face recognition approaches (DCT-GMM- and PCA-SVM- based systems).

5.1.4.1. Case Study 1: Controlled Enrollment

For the first experiments, we have used the available users (112) that satisfy the condition of having at least 2 images in close and at least 1 image in both of the two other scenarios. The data selection process is summarized in Table 5.6, where we can see that the three considered scenarios result in 112 subjects and 2964 face images.

Three main experiments are defined for the verification performance assessment across scenarios:

- *Close2close* protocol. This will give us an idea about the performance of the systems in ideal conditions (both enrollment and testing using close distance images). About half of the close distance subcorpus (754 images) is used for development (training the PCA subspace, SVM, etc.), and the rest (714 images) is used for testing the performance.

<table>
<tr><td rowspan="2">Train</td><td rowspan="2">Close</td><td rowspan="2">1468</td><td>754</td><td>Close</td><td>Train</td></tr>
<tr><td>714</td><td>Close</td><td rowspan="3">Test</td></tr>
<tr><td rowspan="2">Test</td><td>Medium</td><td colspan="2">836</td><td>Medium</td></tr>
<tr><td>Far</td><td colspan="2">660</td><td>Far</td></tr>
</table>

Table 5.6: *Configuration of datasets for the experiments of case study 1.*

Development Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	222	132	**130**	484

Evaluation Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	1290	727	578	2595
Train	661	386	304	1351
Test	629	341	274	1244

Table 5.7: *Configuration of the datasets (close, medium, far and mix combination of all of them) of each acquisition scenario for the case study 2.*

- *Close2medium*, and *close2far* protocol. These two other protocols use as training set the whole close distance dataset (1468 face images). For testing the performance of the systems, we use the two other datasets: 836 medium distance images for *close2medium*, and 660 far distance images for *close2far*.

5.1.4.2. Case Study 2: Uncontrolled Enrollment

For the experiments in this section we have divided the data in 56 subjects as development for tuning the systems and the remaining 78 subjects as evaluation (see Fig. 5.5).

The dataset was then divided according to the three acquisition distance scenarios defined in Sect. 5.1.1. The resulting subsets are shown in Table 5.7. The development set is used to train a PCA subspace and GMM world model per scenario (close, medium, far and mix). Here it is important to note that we have tuned the systems with an equal number of images (130 images, given by the smaller scenario, i.e. the "far" one).

On the other hand, the evaluation set was equally divided into a training and a test set, the first one for training the models of SVM and GMM per user and the other to test the system performance. Table 5.7 shows the different divisions of data in the three scenarios defined. It is possible to appreciate that the number of images is not perfectly distributed between these two sets (train and test) due to an imbalance in the number of samples per user.

Four main experiments are defined for verification performance assessment across scenarios:

- *close2x*. This is designed to obtain the performance of the systems in situations where only high quality controlled images are used to train the system. This will be considered as the

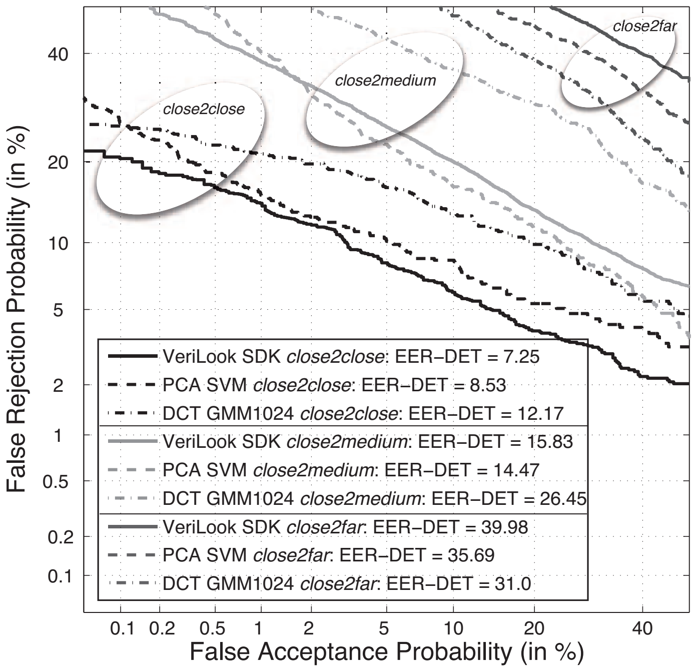


Figure 5.10: *Verification performance results for the three scenarios and three systems considered.*

Baseline system. In this case, only the 661 images of the close train set are used to train the GMM and SVM classifiers.

- *medium2x*, This protocol uses 386 images as training set from the medium distance dataset.
- *far2x* protocol. This protocol uses 304 images as training set from the far distance dataset.
- *mix2x*. This is designed to study the effects of combining several kinds of information (training with different acquisition distances). The train dataset is comprised of the sum of the three acquisition distance datasets (1351 images).

5.1.5. Results

5.1.5.1. Case study 1

In Fig. 5.10 we show the verification performance for the three considered scenarios: *close2close*, *close2medium*, and *close2far*. We first observe that VeriLook is the best of the three systems in *close2close* with an EER around 7%. At the same time, this commercial system is the most degraded in uncontrolled conditions, with an EER close to 40% in *close2far*, worse than the other two systems. This result corroborates the importance of analysing and properly dealing with variability factors arising in biometrics at a distance.

We also observe in Fig. 5.10 that the GMM-based system works better in far distance conditions than the other systems, although being the less accurate in *close2close* and *close2medium*. This result demonstrates the greater generalization power of this recognition approach, and its robustness against uncontrolled acquisition conditions.

EER	M Gaussians								
N Coeff.	*close2close*			*close2medium*			*close2far*		
DCT	1024	128	8	1024	128	8	1024	128	8
15	**12.17**	14.62	20.06	26.45	29.06	36.19	31.01	32.52	38.74
10	13.22	15.97	19.62	**26.09**	28.72	34.90	**29.80**	32.83	38.58
5	17.66	19.80	22.15	31.72	34.60	35.43	33.46	37.07	39.37

Table 5.8: *Verification performance of the DCT-GMM system for different configurations.*

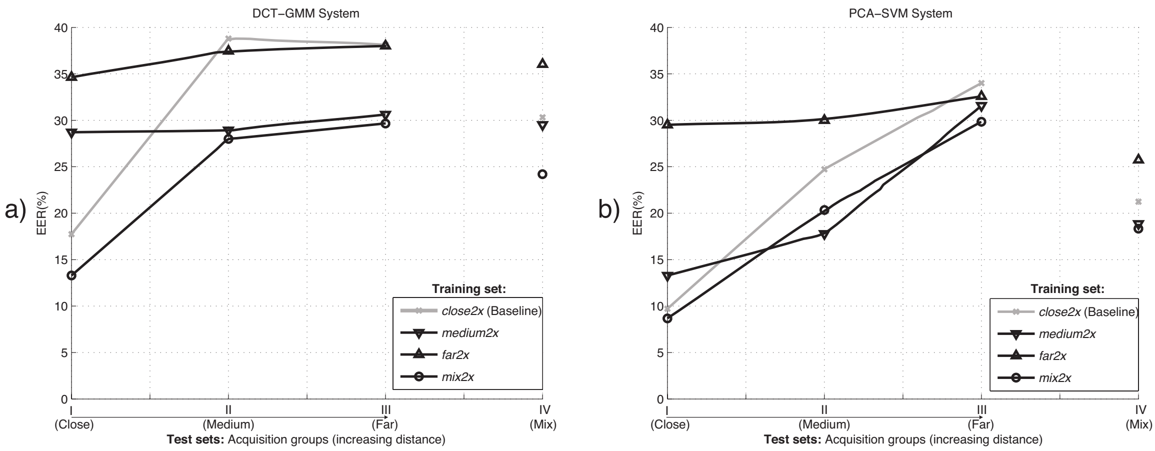


Figure 5.11: *Verification performance of the individual matchers (DCT-GMM- and PCA-SVM- based) and their work in different conditions of training and test sets with different acquisition scenarios.*

Based on this observation, we finally conducted a last experiment simplifying the DCT-GMM complexity in order to enhance its generalization power, seeking for a maximum of performance in the challenging *close2far* scenario. The verification performance results are given in Table 5.8 as EER for decreasing DCT-GMM complexity (N = DCT coefficients, M = Gaussian components per GMM). The results indicate in this case that decreasing the recognition complexity (i.e., improving the generalization power) of this simple recognition method does not help in improving its robustness against uncontrolled conditions. In other words, the DCT-GMM recognition complexity initially considered ($N = 15, M = 1024$), is the most adequate for the *close2far* scenario studied here.

5.1.5.2. Case study 2

Verification performance results are given in Fig. 5.11 for the individual matchers: a) DCT-GMM-based and b) PCA-SVM-based. This figure shows all the possible combinations between training and test sets. The four curves represent the Equal Error Rate (EER) of each **train set** defined (*close2x*, *medium2x*, *far2x*, and *mix2x*) matched with each **test set** (Close, Medium, Far, and Mix). As the person is going away the acquisition device the face information available decreases. This effect is appreciated on the system performance in Fig. 5.11 where both systems degrade their performance when the acquisition distance and the variability increases. We can appreciate how the increment of acquisition distance produces an increment of variability.

By analysing these curves, it can be seen that the DCT-GMM-based matcher is more robust against an increasing acquisition distance [Tome *et al.*, 2010b], as it was shown in previous section, especially when training with the medium distance dataset. Conversely, although being much better in ideal conditions (*close2close*), the PCA-SVM-based matcher degrades quick when increasing the acquisition distance.

When the system is trained with the highest quantity of information possible (*mix2x* protocol), we obtain better performance in general but we must be careful in the comparison because in this case we are training with a higher number of images compared to other scenarios.

5.2. Facial Landmarks Variability

Automatic face recognition over forensic caseworks is still a challenge for the research community. Large amounts of research are being carried out trying to compensate variability sources (such as illumination, pose, facial expressions, occlusions, etc.) that affect significantly reducing the performance of the face recognition systems. In a forensic scenario, these variability factors are crucial, because forensic examiners have to deal with face images extracted from CCTV cameras and other low quality sources, which makes the task really difficult.

Many different techniques have been developed to automatically tag facial landmarks on a face [Arca *et al.*, 2004, 2006; Beumer *et al.*, 2006; Gupta *et al.*, 2010]. These techniques achieve good results over good quality and frontal faces, but are still not good enough for the cases of having high variability and low quality images. Actually, in practice there is still no automatic system that can achieve such a detailed level compared to humans. On the other hand, humans are subjective and do not work as systematically as computers. For this reason, in practice forensic examiners make use of semiautomatic systems, which can help in the suspects identification tasks [Jain *et al.*, 2012b].

Among the tasks carried out by forensic examiners, they perform identification tasks of a given probe image in a set of gallery images (with known identity). In an anthropometric analysis they extract manually a set of facial landmarks, then compute some distances between them, which can be used as features in their analysis. Figure 5.12 shows a diagram of this procedure.

This section focuses on an analysis of the variability of facial landmarks in a forensic scenario over a database of mugshots and CCTV images. This variability between landmarks is affected by two factors, on the one hand the accuracy of the process of landmark tagging, which can be done manually or automatically and can vary significantly due to the quality of the images, and on the other hand it is also affected by the intrinsic variation of the landmarks, due to changes in pose, expression or occlusions among others.

For mugshot images scenario, the ATVS Forensic database comprised of mugshot images at three different distances (1, 2, and 3 meters) from 50 persons has been used. 21 facial landmarks were defined and 1200 images have been manually tagged imitating the work performed by a forensic examiner.

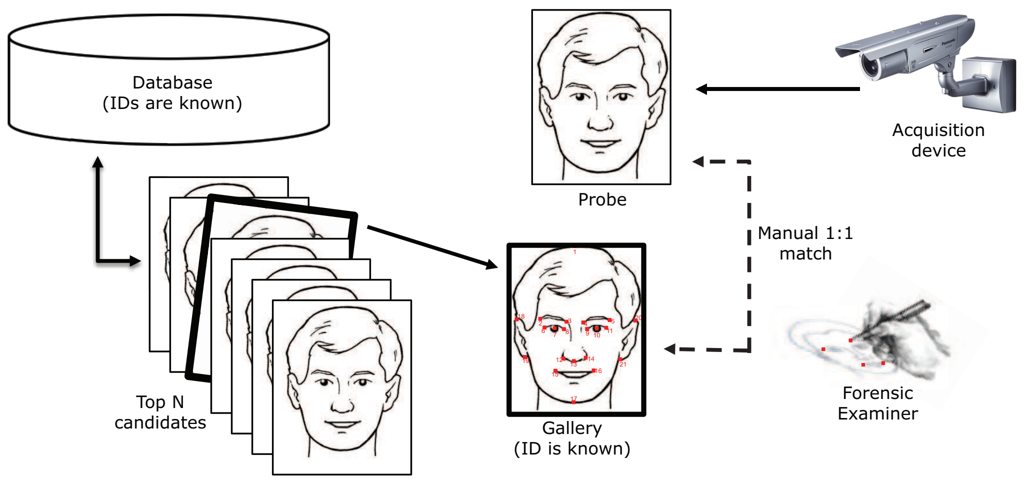


Figure 5.12: *General procedure followed by a forensic examiner to compare two face images.*

Following the same criterion in the CCTV images scenario, we carry out the study using SCface database, which is comprised of images at three different distances (1, 2.6, and 4.20 meters) between the camera and the persons. In this PhD Thesis we analyse two effects: *i*) the effect of the distance between the subject and the camera in the landmark variability, and *ii*) the comparison between the variability of an automatic system compared to a manual landmark tagging imitating the work of a forensic examiner.

Some of the findings of this study are that in general facial landmarks located in the outer part of the face (highest point on the head, ears and chin) have a high level of variability, due possibly to hair occlusions. For mugshot images regarding the distances between the camera and the persons, the variability increases gradually with the distance but not very much. On the other hand in CCTV images the variability increases gradually with the distance. And surprisingly in this CCTV scenario, very similar results are achieved for both manual and automatic approaches, although not all the landmark points were able to be tagged by the automatic system. The findings of this Section could be useful information for forensic examiners.

5.2.1. Database and Experimental Protocol

The experiments for the analysis of the variability of facial landmarks are divided in two complementary studies and are carried out in two different databases: ATVS Forensic DB and SCface DB, which are broadly described in Chapter 4.

The ATVS Forensic database is comprised of mugshot images at three different distances (1, 2, and 3 meters) from 50 persons. An example of the three frontal images captured is shown Fig. 5.13 (top). This database is going to represent the mugshot scenario because the collection process replicates the real practice of forensic laboratories. The database comprises data from 50 persons (32 men and 18 women) acquired in two different sessions. The sessions were collected in different days for the same persons and in each session the procedure was repeated four times. Therefore, obtaining a total of 1200 face images (50 persons $\times$ 2 sessions $\times$ 4 times $\times$ 3 images).

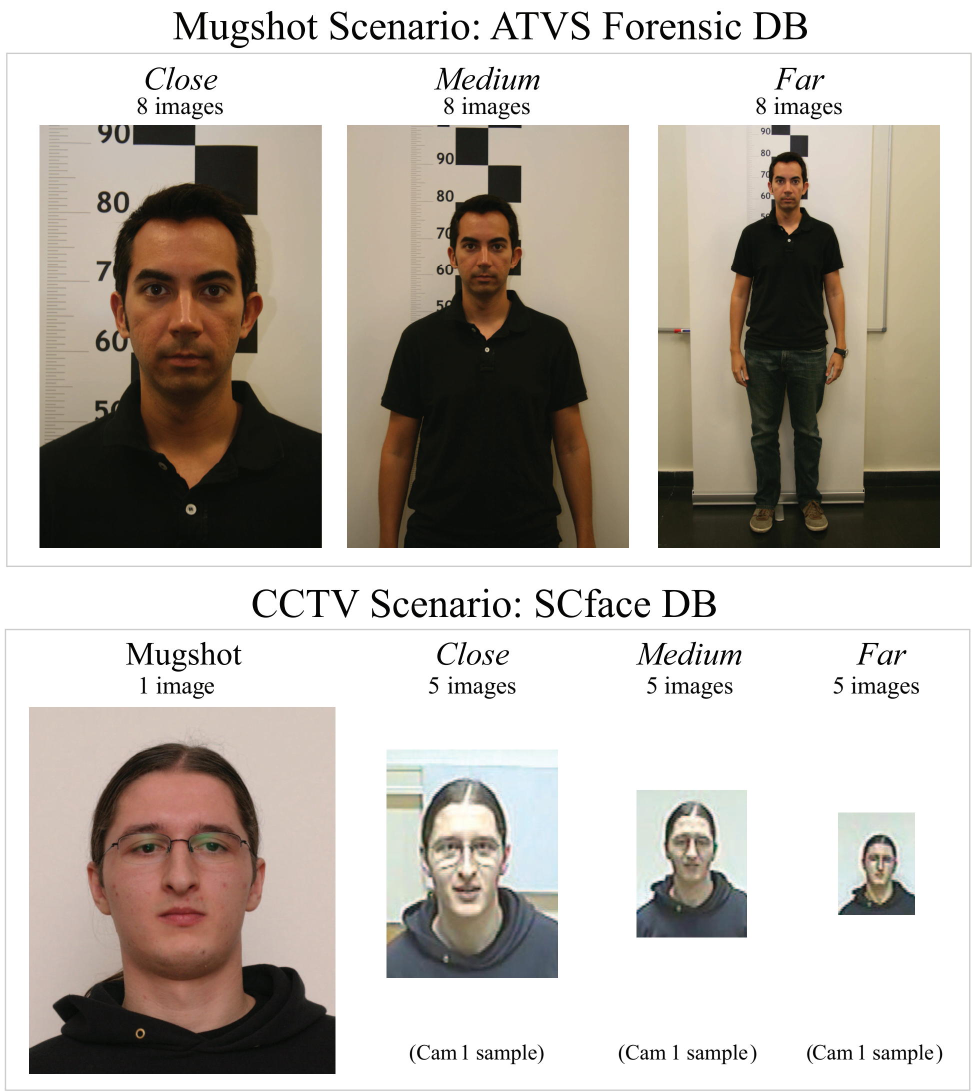


Figure 5.13: *On the top, examples from ATVS Forensic DB of front images acquired at a mugshot session considering three distances between the person and the camera. On the bottom, image samples from SCface database. High quality mugshot image, and 3 CCTV images acquired at three distances: close, medium and far, for one of the five CCTV cameras.*

The second one, SCface DB is a database of static images of human faces with 4.160 images (in visible and infra-red spectrum) of 130 subjects (115 men and 15 women). The dataset is divided into 6 different subsets: *i*) mugshot images, which are high resolution frontal images, and *ii*) five visible video surveillance cameras. Each of these subsets contains 130 images, one per subject at three different distances: 1.0 m (Close), 2.6 m (Medium) and 4.2 m (Far) respectively, and acquired while the subject walked towards the cameras. Fig. 5.13 shows an example of a mugshot image, and the images acquired by one of the surveillance cameras.

As can be seen there is a considerable scenario variation in terms of quality, pose and illumination. The effect of the pose is specially important due to the different angles between the person and the cameras.

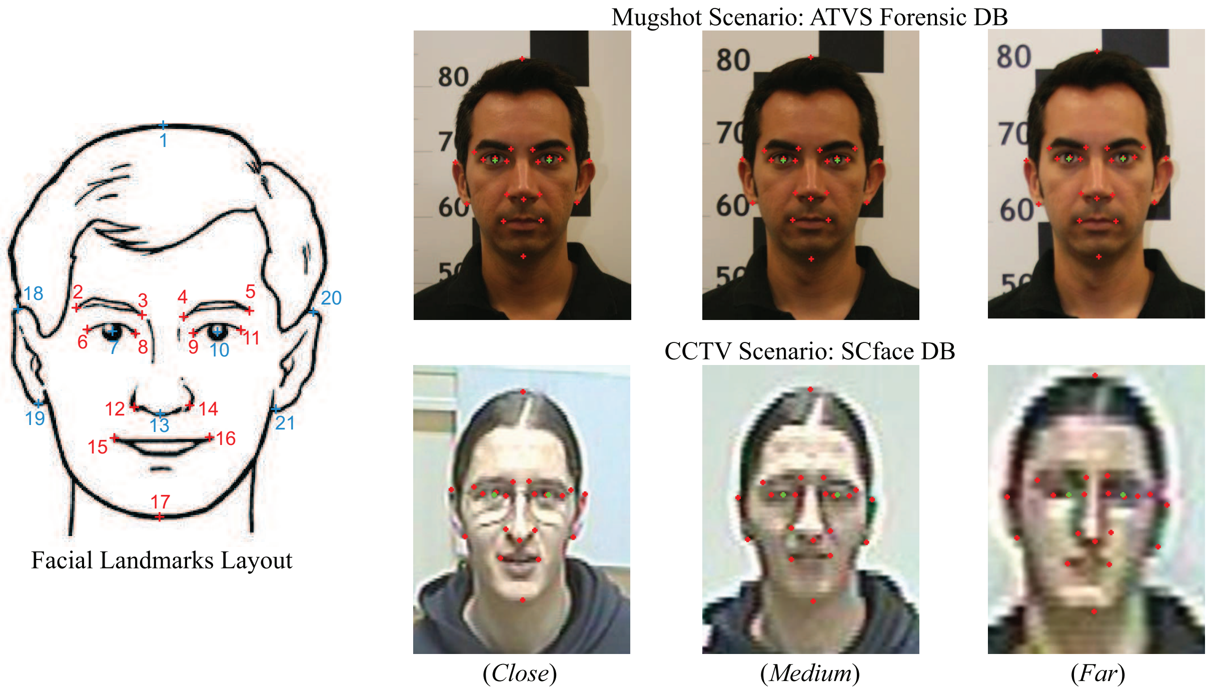


Figure 5.14: On the left, the set of 21 facial landmarks defined (in red are the landmarks considered for automatic tagging). On the right, same example as in Fig. 5.13 (for SCface only for the CCTV images) but normalizing the faces with 75 pixels between the center of the eyes. Also, the 21 manual facial landmarks are shown (red), plus the center of the eyes (green).

This second database is of particular interest from a forensic point of view because images were acquired using commercially available surveillance equipment, under realistic conditions. One of its drawbacks is that it is just comprised of one mugshot session, so it is not possible to study the landmark variability for the mugshot images, as several pictures per person are needed. For this study, we use the 5 available images per person and per distance to analyse the variability of the facial landmarks (1950 images in total, 3 distances $\times$ 5 cameras $\times$ 130 persons). Also, we carry out this study both in a manual way imitating the work that a forensic examiner would perform, and using an automatic system to detect the facial landmarks.

5.2.2. Facial Landmarks Extraction

This section describes the process of facial landmark tagging and image processing in order to analyse the variability of facial landmarks. The first step is to define a set of facial landmarks to include in this study. A set of 21 facial landmarks was defined following recommendations from the Spanish Guardia Civil (DGGC), Netherlands Forensic Institute (NFI) and European Network of Forensic Science Institutes, including the irises (2 landmarks), inner and outer eye corners (4), eyebrow ends (4), mouth corners (2), nose corners (2), center of the nose (1), chin (1), upper and lower ears ends (4) and highest point on the head (1). Fig. 5.14 (left) shows the 21 facial landmarks considered in this study.

For ATVS Forensic database the process was carried out manually, while for SCface database this process of facial landmark tagging was carried out both manually and automatically. The manual landmark tagging was carried out by the same person, imitating the work of a forensic examiner. In this case the set of 21 landmarks was tagged for the whole database. On the other hand for CCTV scenario, it is interesting to compare this experimental work with an automatic landmark tagging system. For this case, Luxand Face SDK was used, which is a high performance face recognition commercial software based on facial landmarks features. In this case, a set of 13 facial landmarks (in red in Fig. 5.14 (left)) were considered, as the automatic system was not able to locate most of the other 8 remaining landmarks.

A second stage of image processing was carried out in order to normalise the facial images to the same size and position. Thus, the midpoint between the eye corners (midpoint between points 6 and 8, and midpoint between 9 and 11) was computed and used instead of the irises positions to align the faces, because the position of the irises can vary if the person does not look at the camera directly. The positions of these two points were fixed having 75 horizontal pixel between them following the recommendation from the ISO standard [ISO/IEC 19794-5:2011, 2011]. Therefore, translation, rotation and scaling of the original images was carried out to normalize the database. This was done in the same way for images collected at different distances between the camera and person. Fig. 5.14 (right) shows an example of the three mugshot and CCTV face images shown in Fig. 5.13 but size normalised, and showing the positions of the 21 facial landmarks in red and the positions of the center of the eyes in green. As can be seen, this is a challenging scenario for both manual and automatic landmark tagging due to the low quality of the images to analyse.

5.2.3. Results

5.2.3.1. Person Specific Landmark Variability

In this experiment person specific landmarking variability (LV) was studied. Thus, the 8 and 5 available facial images per person and per distance for mugshot and CCTV scenarios were considered, respectively. The mean and standard deviation for each facial landmark were computed for the two (x,y) spatial dimensions ($\sigma_{x,i}$, $\sigma_{y,i}$, with *i=1,...,13* or *21* depending on the automatic or manual landmark tagging process), assuming following a Gaussian distribution (see Fig. 5.15). Fig. 5.16 shows some example face images superimposing for each facial landmark the result of tagging the available images in each database. An ellipse around each facial landmark is computed using as the radios ($2\sigma_{x,i}$, $2\sigma_{y,i}$). Throughout this Section the variability of the different facial landmarks is shown quantitatively together with these ellipses as $2\times$mean($\sigma_{x,i}$,$\sigma_{y,i}$) in pixels (for normalized images with 75 pixels between eye positions).. For example, in the image shown in Fig. 5.16 the landmark for the highest point on the head shows a variability of $\pm$ 9.8 pixels, which covers 95.44% of the assumed Gaussian distribution.

This procedure was followed for the 50 and 130 persons comprising the two databases, and it was found that in both scenarios the variability of the facial landmarks, specially for the

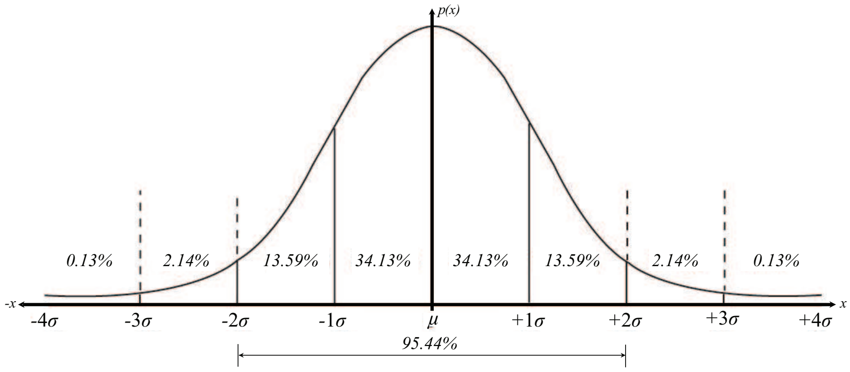


Figure 5.15: *Gaussian distribution showing the range* $[\mu - 2\sigma, \mu + 2\sigma]$*, covering the* 95.44% *of the distribution.*

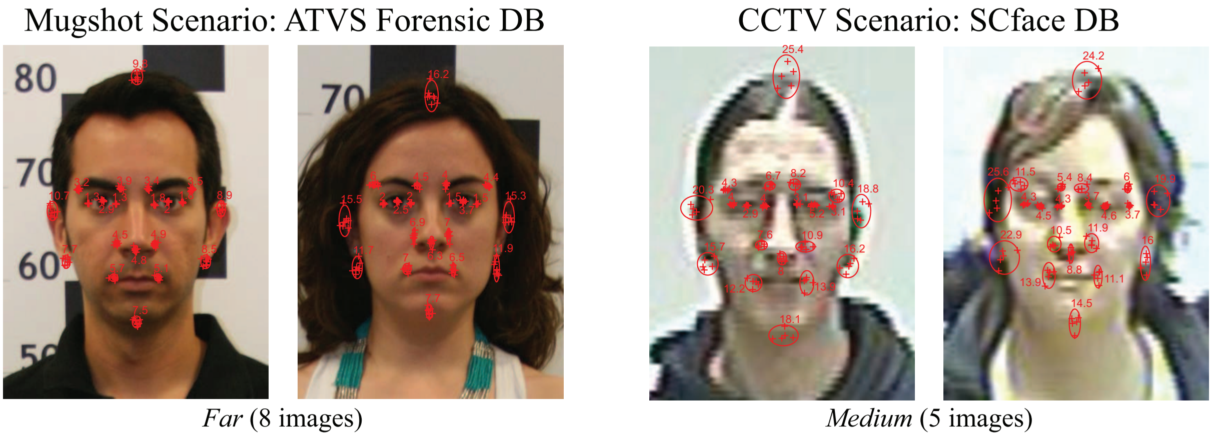


Figure 5.16: *On the left, examples of the landmarking variability for two persons present in ATVS Forensic database taken at (*far*) 3 meters distance between the person and the camera. On the right, examples of the manual landmarking variability for two persons present in the SCface database for images taken at (*medium*) 2.60 meters distance between the person and the camera.*

outer ones varies significantly from person to person. In the examples shown in Fig. 5.16, the variability of these landmarks on the outer part of the face (highest point on the head, chin and ears) is very dependent on hair occlusions, more frequent in women than men for the population considered.

This effect is very interesting and therefore, the second experiment is designed to study the variability of facial landmarks across gender on the mugshot scenario.

5.2.3.2. Gender Specific Landmark Variability

This section describes the experimental results of the variability of the facial landmarking comparing results achieved for males and females contained in the mugshot scenario using ATVS Forensic database. This gender specific landmarking variability experiment was not carried out on CCTV scenario due to the low quantity of female subjects, 15 against 115 males.

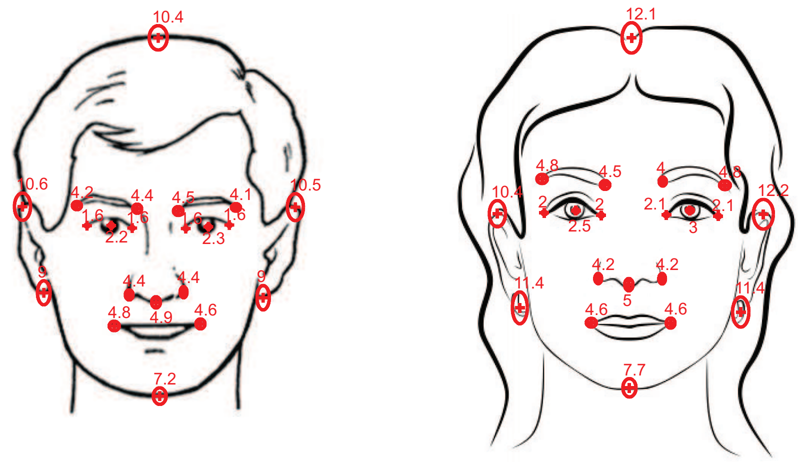


Figure 5.17: *Results of the landmarking variability for male and female for pictures taken at 3 meters distance between the subjects and the camera.*

In order to compute a global landmarking variability (LV) for males and females, the mean of the different individual values of the variability of each facial landmark is computed, following the equation:

$$LV_{M,i} = \frac{1}{N_M} \sum_{j=1}^{N_M} (\sigma_{x,i,j} + \sigma_{y,i,j}) \tag{5.2}$$

where $i = 1, \ldots, 21$ are the landmarks and $j = 1, \ldots, N_M$, being N_M the number of males. Similarly we compute $LV_{F,i}$ for the N_F females present in our database. Fig. 5.17 shows the results achieved for male and female respectively for the case of the 3 meter distance mugshot. As can be seen, in general the landmarking variability is larger for female compared to male (16 landmarks out of the 21), mostly for the landmarks placed in the outer part of the face (i.e., highest point on the head and ears), where there can be more the hair occlusions. Still the difference in absolute number of pixels having normalised the images with 75 pixels between the eyes is not very significant.

5.2.3.3. Distance Specific Landmark Variability

This experiment reports the experimental results achieved for the global landmarking variability (i.e., an average of the individual results and without distinction for males and females) considering the effect of the 3 distances between the camera and the persons for each database. In order to compute a global landmarking variability (LV), the mean of the different individual values of the variability of each facial landmark is computed, following the previous equation 5.2, where $i = 1, \ldots, 13$ or 21 (for automatic or manual tagging respectively) are the landmarks and $j = 1, \ldots, N$, being $N = N_M + N_F$ the maxima number of persons in the database, 50 or 130 in each case depending the scenario analysed. This procedure is followed for each of the three distances considered.

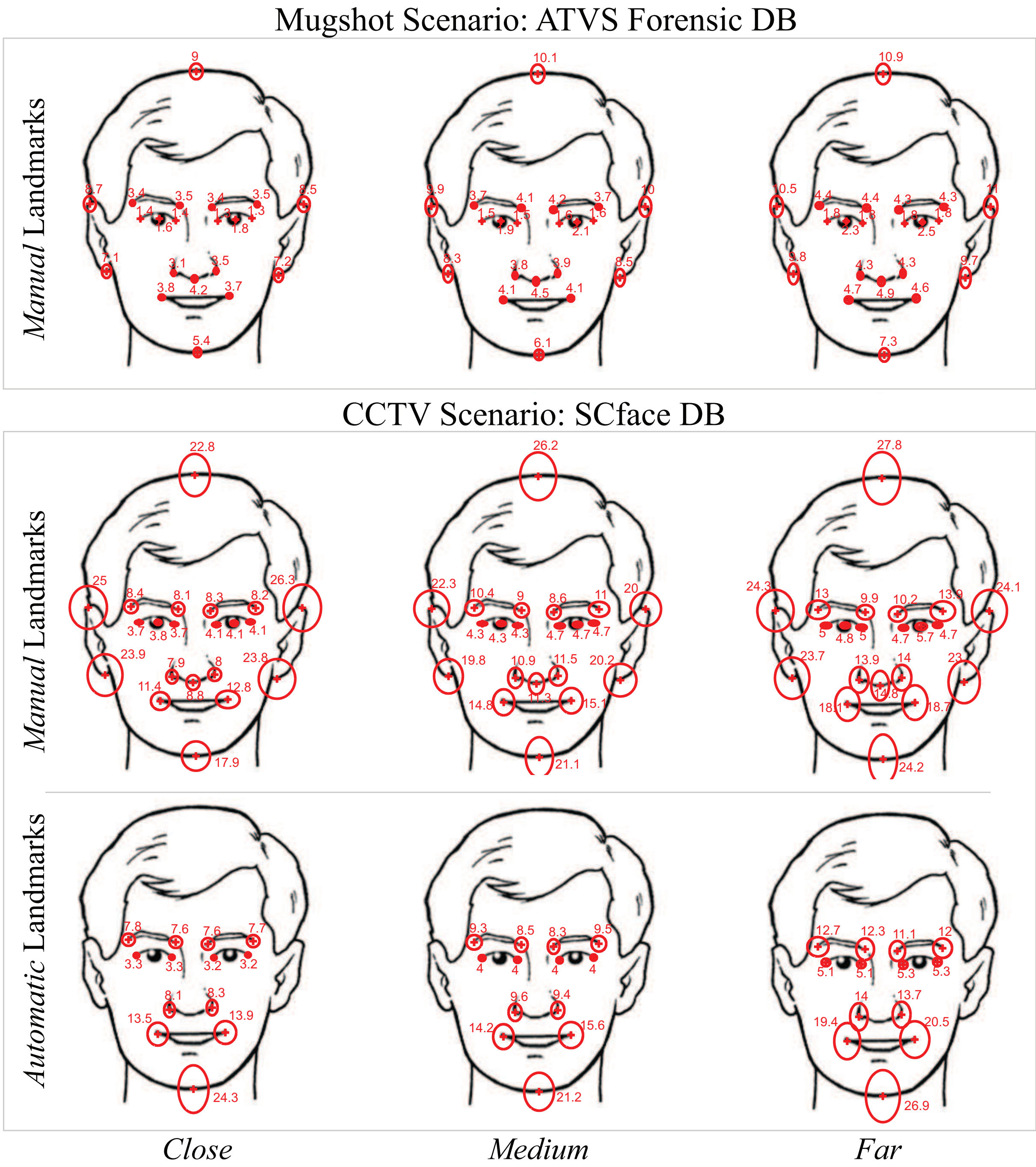


Figure 5.18: *On the top, results of the landmarking variability for the three distances considered between the persons and the camera: close (1 m), medium (2 m), and far (3 m). On the bottom, results of the landmarking variability for the three distances considered between the persons and the camera: close (1.0 m), medium (2.6 m), and far (4.2 m).*

Fig. 5.18 (top) shows the results for the mugshot scenario, while Fig. 5.18 (middle and bottom) shows the results for CCTV scenario achieved for the case of manual and automatic landmark tagging, respectively. Here we focus on the analysis of the distance for the manual case, as next section compares the case of manual vs. automatic landmark tagging. As can be seen, there is a clear increment of the landmark variability regarding the acquisition distance between the subject and the camera for all the facial landmarks considered for both scenarios (except for the ears in CCTV scenario). Again, the outer facial landmarks (highest point on the head, ears and chin) present the highest variability, then the mouth and nose areas, and the parts with the least variability are the eyes and eyebrows. It is worth noting that the normalization of the faces was done using the center of the eyes, so it is also natural that these parts present less variability than the rest.

It is also worth noting that as the landmark tagging is carried out over the original size images, in this case close images have a bigger size compared to far images, as can be seen in Fig. 5.13. Therefore, the process of landmark tagging can be done with more precision for the close images and therefore reducing the landmark variability.

For completeness, results achieved in the comparison between both scenarios CCTV and mugshot are shown in Fig. 5.18. It is worth noting that in the mugshot scenario, the images were of a much higher quality as the images from SCface database considered in the CCTV scenario. As can be seen, the landmark variability in mugshots is much lower in all three distances compared to the results achieved in CCTV scenario. This significant difference of the variability is mainly due to the quality of the images considered. It is also worth mentioning that SCface database was acquired in uncontrolled conditions while the mugshot database was acquired in a controlled scenario.

5.2.3.4. Manual Vs. Automatic Landmarking Variability

Fig. 5.18 shows the landmark variability for both manual (middle) and automatic (bottom) procedures. The number of facial landmarks tagged is different in both cases, 21 for manual and 13 for automatic tagging, as described in Sect. 5.2.2. The results show that the landmark variability is very similar for the set of common landmarks between the automatic and the manual procedures. Specifically, for the close images, the landmarks located in the ocular region present lower variability for the automatic system compared to the manual case, while the landmarks located in the mouth region present a higher variability for the automatic system.

For the medium distance, both ocular and mouth region present in general a lower variability for the automatic system, but in the far distance where the quality of the images is very low the manual procedure achieves lower landmark variability. It is also worth noting that the automatic system only considers 13 facial landmarks as it was not able to locate correctly most of the remaining 8 landmarks (mainly the outer ones), but in general it achieves better results than expected a priori.

5.3. Facial Regions

Automatic face recognition systems are generally designed to match images of full faces. However, in practice, forensic examiners carry out a manual inspection of the face images, focussing their attention not only on the full face but also on individual traits. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc.

There are some previous works where region-based face recognition is studied [Bonnen *et al.*, 2013; Gupta *et al.*, 2010; Heisele *et al.*, 2007; Jain *et al.*, 2012b; Kim *et al.*, 2005; Li *et al.*, 2008; Ocegueda *et al.*, 2011; Sadr *et al.*, 2003; Tistarelli, 1995] but non of them focus their attention in the regions normally considered by forensic experts. In this work, we have extracted facial regions (as detailed in Sect. 3.2.3 of Chapter 3) following forensic protocols from law enforcement agencies, allowing us to study the discriminative power of different facial regions individually. In particular, we address in this section the problem of finding the most discriminative areas of the face for recognition on different acquisition scenarios.

Studying the discrimination power of different facial regions on a wide population has some remarkable benefits, for example: *i*) allowing investigators to work only with particular regions of the face, *ii*) preventing that incomplete, noisy, and missing regions degrade the recognition accuracy. Further, a better understanding of the individuality of facial regions should facilitate the study of facial regions-based face recognition. In the same way that the field of cognitive science continues to investigate the precise roles of facial regions and holistic processing in human face perception [Gold *et al.*, 2012], automatic face recognition algorithms also need to explore the role that facial regions processing could have improving their performance.

The present Section reports an exhaustive analysis of the discriminative power of the different regions of the human face on various forensic scenarios. In this scenario it is very important to know based on scientific methods to what extent each facial region can help in identifying a person. This knowledge is obtained using quantitative and statistical methods on given populations can then be used by the examiner to support his observations. In order to generate such scientific knowledge useful for the expert, several methodologies are compared, such as manual and automatic facial landmarks extraction, different facial regions extractors described in Sect. 3.2.3, and various distances between the subject and the acquisition camera. Also, three scenarios of interest for forensics are considered comparing mugshot and CCTV face images using MORPH and SCface databases. One of the findings is that depending of the acquisition distances, the discriminative power of the facial regions change, having in some cases better performance than the full face. Fig. 5.19 shows a diagram of the methodology followed in this Section.

5.3.1. Facial Regions Extraction and Representation

The algorithm for extracting facial regions, as the rest of extraction approaches, is an iterative method that takes advantage of the facial landmarks tagged by a human examiner or an

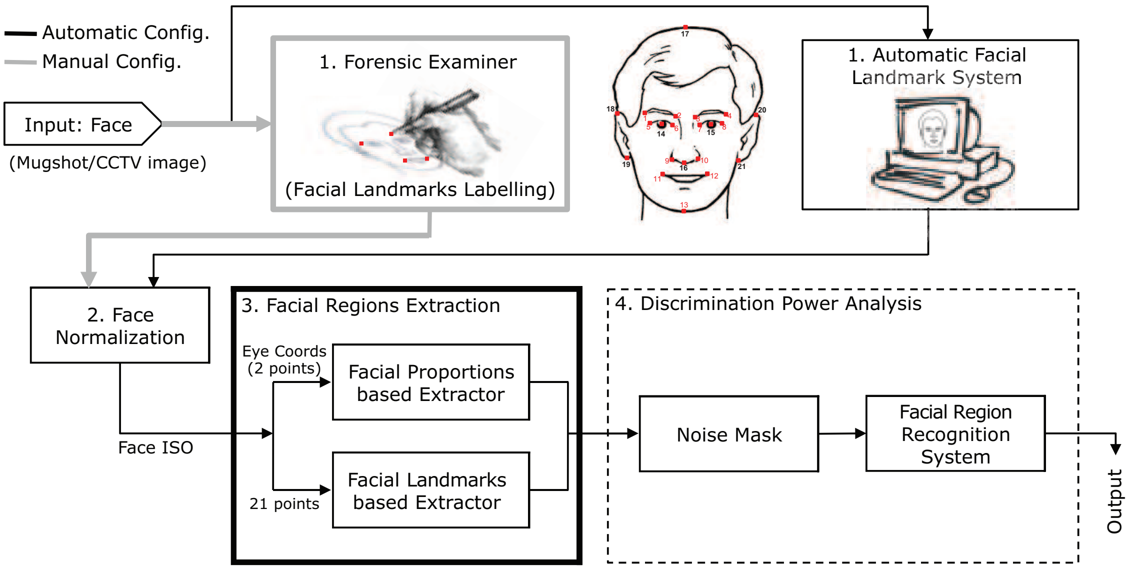


Figure 5.19: *Experimental framework followed to study the discrimination power of the 15 facial regions.*

automatic system to extract a number of $R = 15$ facial regions based on the forensic practice. The main difference with other extraction techniques is that in this case the extraction can be done on controlled and uncontrolled images thanks to the use of facial landmarks and proportions. This fact allows the algorithm to be suitable to be used in face biometric systems at a distance working with facial landmarks easily tagged with an automatic system (regardless of the tagged systems of facial landmarks, or the type of matcher being used). Fig. 5.20 shows a diagram of the methodology followed. The extraction is defined by two main parameters: *i*) interpupilarity pixel distance (IPD), which indicates the number of pixels between the eye centres of the subject, and *ii*) L, which defines the number of facial landmarks tagged. $L = 2$ for the extractor based on facial proportions and $L = 13$ or $L = 21$ for the automatic or manual extraction based on facial landmarks.

The experimental framework implemented based on these two extractors allows the extraction of 15 different facial regions as can be seen in Fig. 5.22. The election of these 15 regions is based on protocols from international forensic laboratories [Netherlands Forensic Institute (NFI); Spanish Guardia Civil (DGGC)]. The extraction procedure follows three steps:

1. Detection of facial landmarks.
2. Face normalization.
3. Facial region extraction.

Facial Landmark Detection. The first step is to extract a predefined set of anthropometric landmarks. This step has two different configurations: automatic and manual in order to find the facial landmarks.

Given the variability of facial appearances, as well as the variability caused by pose and expression changes, the extraction of facial landmarks is often a difficult task to be performed automatically. When considering challenging scenarios at a distance, the low quality of the

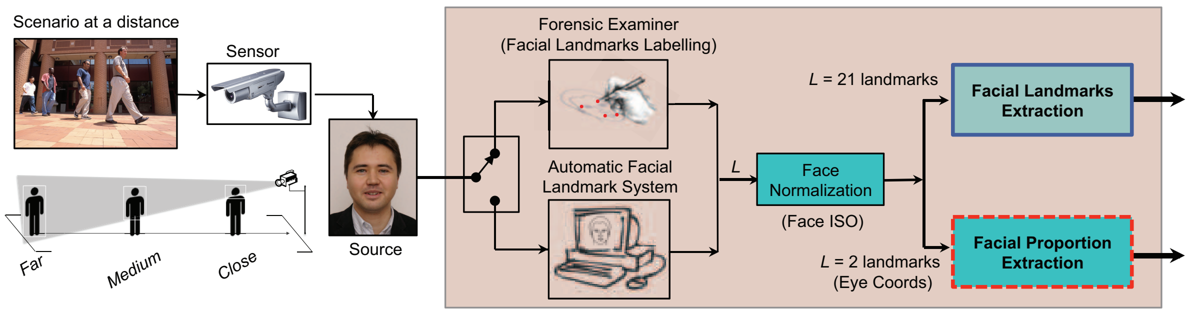


Figure 5.20: *Experimental framework followed to extract the facial regions.*

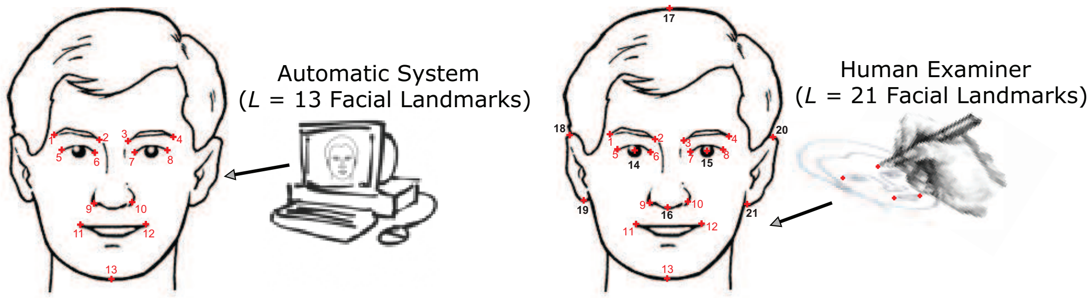


Figure 5.21: *Facial landmarks selected for the automatic and manual configurations.*

images introduces another important factor that makes even more difficult the detection task.

On the other hand, the common practice of forensic examiners is mainly based on manual and individual skills using some general image processing tools. This approach permits to have reliable landmark information even in lower quality images but may introduce a subjective bias in the process. In this PhD Thesis we have studied both automatic and manual facial landmark detection.

For the automatic approach, the commercial SDK Luxand FaceSDK 4.0 [Luxand Face SDK], was first used to automatically detect 65 facial landmarks. Next, the landmarks of each facial region (eyebrows, eyes, nose, mouth and chin) were automatically selected and the rest were removed. The result of this step is an initial placement of facial landmarks where just 13 of them are considered as Fig. 5.21 (left) shows. These 13 facial landmarks have been selected following forensic face recognition protocols by Spanish Guardia Civil (DGGC) and Netherlands Forensic Institute (NFI) and they indicate the terminations of each trait in a human face.

For the manual approach for landmark detection, a human manually tagged 21 facial landmarks imitating the procedure of a forensic examiner, as shown in Fig. 5.21 (right). As can be seen the 13 automatic facial landmarks are included as a subset of the 21 marked with the manual approach. In the manual approach, the ears and the upper end of the head are also marked.

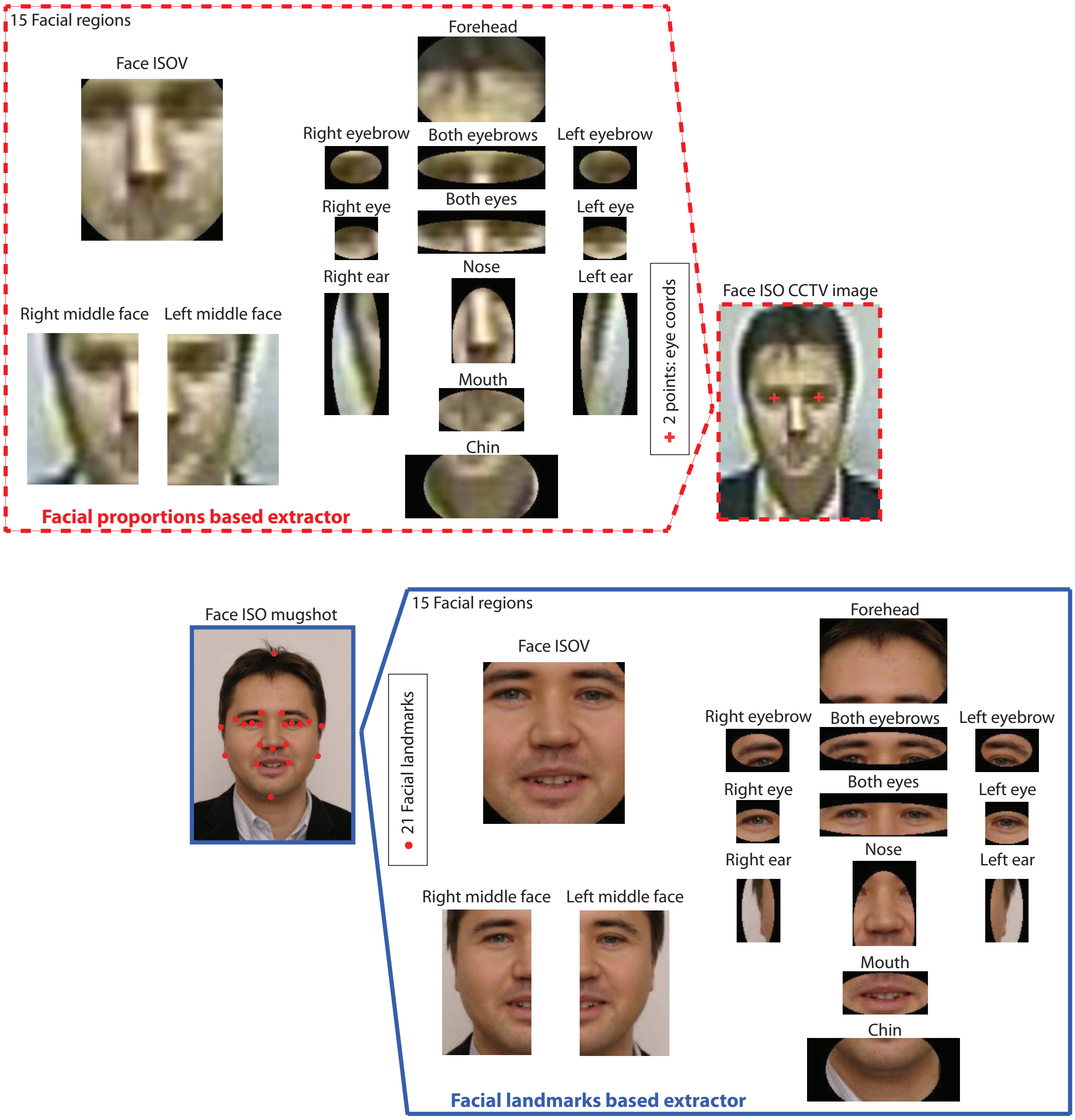


Figure 5.22: *Facial regions extraction. On the top side, with dashed line, the extractor based on facial proportions and on the bottom side, with solid line, the extractor based on facial landmarks.*

Face Normalization. Once the facial landmarks have been detected, the next step is the extraction of the facial regions. This is performed following two approaches: *i*) based on human face proportions, and *ii*) based on facial landmarks.

Before extracting the facial regions all the faces were normalised following the ISO standard [ISO/IEC 19794-5:2011, 2011] with an interpupillary pixel distance (IPD) of 75 pixels. This step eliminates variations in translation, scale and rotation in horizontal plane, and provides a normalized face in order to compare it or extract facial regions with a standard size for all faces considered.

Extractor based on Facial Proportions. The extractor based on facial proportions uses the proportionality relationships in a human face. These relationships divide the human face in several horizontal and vertical areas with the same size as shown in Fig. 5.23 (top). There are previous works where facial proportions of a human face were studied [Gunes and Piccardi, 2006; Jefferson, 2004; Shiang, 1999]. Based on these works an automatic facial region extractor system following these proportions rules has been developed. This extractor applies facial proportions rules using the eye centers as reference point. Fig. 5.23 (top) shows the proportions calculated based on these reference points (lines A-M) used to extract the 15 facial regions described in Fig. 5.22 (top).

Using just the two eyes coordinates, following single facial proportions rules considering the IPD distance, 15 facial regions (eyebrows, eyes, nose, mouth, etc.) can be extracted from a frontal face. The main drawback of this approach is the low precision, which can produce small misalignments of the region for the different face images.

On the other hand, this extractor would be of interest in challenging uncontrolled scenarios where landmarks are very difficult to be extracted automatically, but an automatic face recognition system can locate the eyes coordinates easily or they can be tagged manually. An example of this extraction can be seen in Fig. 5.22 (top), which shows the 15 regions considered based on protocols from international forensic laboratories [Netherlands Forensic Institute (NFI); Spanish Guardia Civil (DGGC)].

Extractor Based on Facial Landmarks. The second extractor, is based on anthropometric facial landmarks allowing us to extract the facial regions with higher precision. In this case, a facial region is extracted by estimating the center between each one of two facial landmarks per facial trait and by applying a vertical and horizontal offset to generate a bounding box that contains the facial region, as can be seen in Fig. 5.23 (bottom). This procedure is followed automatically for the extraction of the 15 forensic facial regions as shown in Fig. 5.22 (bottom).

The main drawback of this approach is that the precision of the extraction depends on the correct manual or automatic localization of the facial landmarks. On the other hand, this method provides a good alignment allowing us to compare facial regions keeping their relationships of shape and size.

There are previous techniques [Kim *et al.*, 2005; Pan *et al.*, 2007] which have used pre-defined cropping boundaries, and a more recent work [Bonnen *et al.*, 2013] uses alignment approaches such as procrustes analysis [Gower, 1975]. In our case, the ISO normalization step previously applied, together with the central point estimation step allows us to solve alignment problems in the extraction process.

Table 5.9 shows the size of the 15 facial regions for the two extractors. As can be seen, the extractor based on proportions needs a bigger bounding box than the extractor based on facial landmarks.

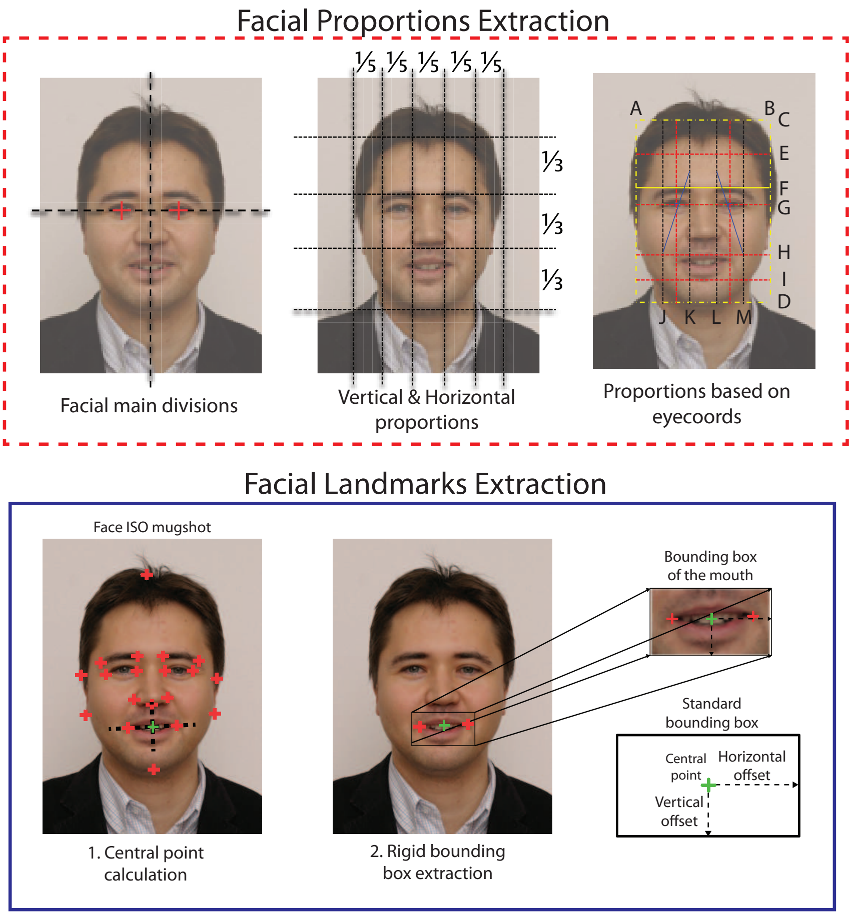


Figure 5.23: *(Top) Facial proportions: main facial divisions, horizontal, vertical and proportions based on eyecoords. (Bottom) Extraction procedure of the mouth region using the extractor based on facial landmarks.*

Once each facial region has been extracted, histogram equalization is applied to each grayscale facial region. In order to avoid external noise in each region, a noise mask is applied (see black areas in Fig 5.22). Then, eigen-region (PCA) [Turk and Pentland, 1991a] is applied to each facial region over the training set considering the first 200 principal components for each region. Similarity scores are computed in this PCA vector space using a Support Vector Machine (SVM) classifier with a linear kernel [Cortes and Vapnik, 1995].

5.3.2. Database and Experimental Protocol

The experimental work described in this section has been carried out using a collection of mugshot and CCTV face images of 130 subjects from two different databases: SCface [Grgic *et al.*, 2011] and MOPRH [Ricanek and Tesafaye, 2006].

SCface is a database of static images of human faces with 4.160 images (in visible and infrared spectrum) of 130 subjects. The database is divided into 6 different subsets: *i)* mugshot images, which are high resolution frontal images, and *ii)* five visible video surveillance cameras.

Id Num.	Facial Region	Prop. based Extractor	Landmarks based Extractor
1	Chin	55x188	75x181
2	Left ear	145x76	75x51
3	Right ear	145x76	75x51
4	Left eyebrow	48x57	51x75
5	Right eyebrow	48x57	51x75
6	Both eyebrows	48x132	51x151
7	Left eye	48x57	51x51
8	Right eye	48x57	51x51
9	Both eyes	48x132	51x151
10	Face ISOV	192x168	192x168
11	Forehead	71x132	101x151
12	Left middle face	180x132	173x106
13	Right middle face	180x132	173x106
14	Mouth	57x113	51x101
15	Nose	112x76	101x75

Table 5.9: *Facial regions sizes for both extractors based on proportions and facial landmarks (height $\times$ width in pixels).*

The database is broadly described in previous Sect. 4.2.4. The effect of the pose is specially important due to the different angles between the person and the cameras as shown in Fig. 4.7. This database is of particular interest from a forensic point of view because images were acquired using commercially available surveillance equipment, under realistic conditions.

One of its drawbacks is that it is just comprised of one mugshot session, so the comparison of mugshots versus mugshots cannot be carried out only being able to compare mugshots vs. CCTV images and CCTV vs. CCTV. In order to solve this limitation a second dataset for our experiments was used: the MORPH Non-Commercial Release database [Ricanek and Tesafaye, 2006] in order to study the mugshot vs. mugshot scenario. This database is described in Sect. 4.3.1. For the experiments in this Thesis, we generated a new dataset comprised of 780 mugshot images for 130 subjects from the subset "European" with 6 sessions per subject and with a time lapse between sessions around one year. As a result, a similar structure compared to SCface DB is obtained, which facilities their comparison. Fig. 5.24 shows an example of the images available for a person in both databases.

It is important to note that SCface was collected in 5 days while the time lapse in MORPH database is around one year. This is an important difference that will be considered in the experimental results and findings.

The experimental protocol followed in these experiments is similar to the one proposed in [Wallace *et al.*, 2011][1]. The database was divided into 3 subsets based on the subject ID: development (1-43), SVM training (44-87), and test (88-130).

[1] http://scface.org/

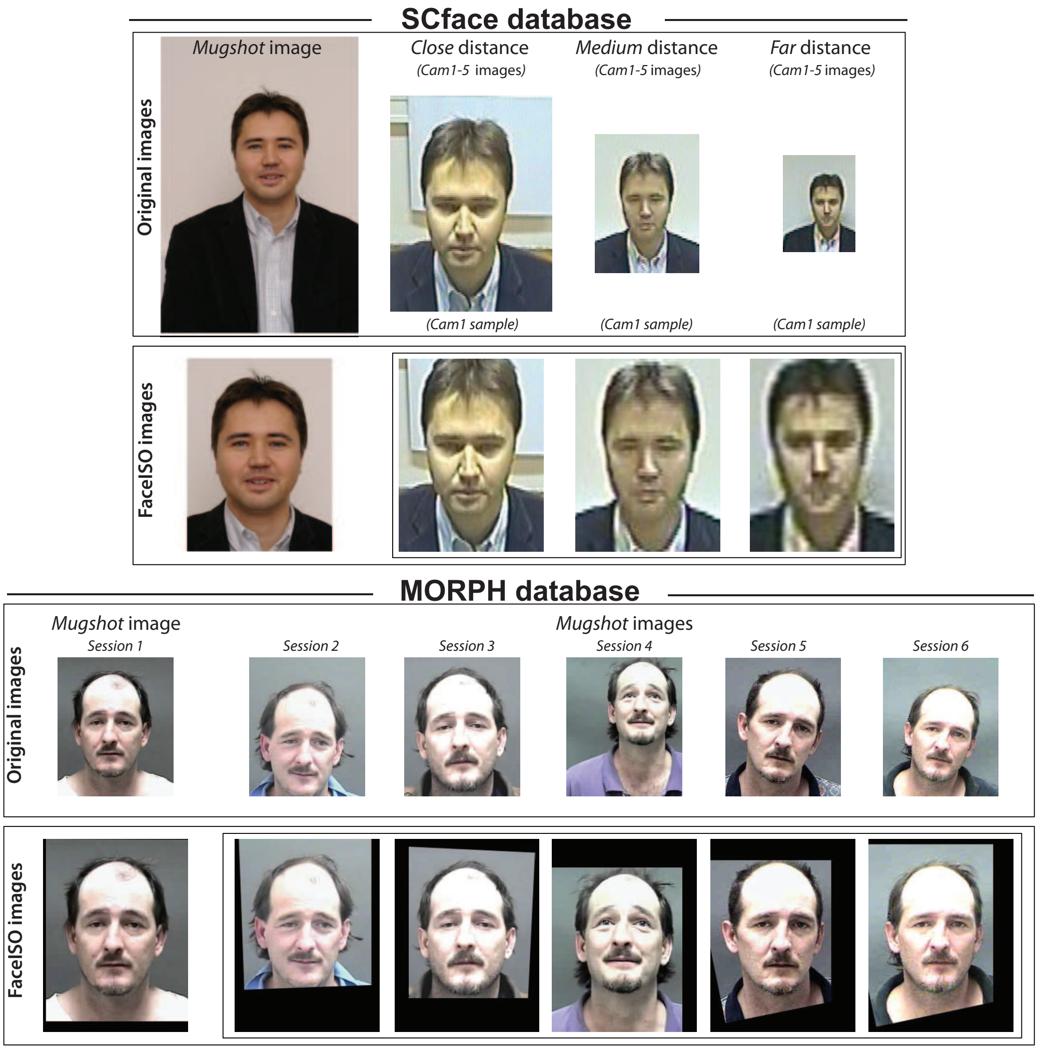


Figure 5.24: *(Top) SCface image samples of each dataset for mugshot and Cam*1 *images, and their corresponding normalized face ISO for the close, medium, and far distance. (Bottom) MORPH image samples (*200×240*) of each session and their corresponding normalized face (*300×400*).*

In this work three different protocols are defined considering the different cases that a forensic examiner can find in practice:

1. *Mugshot vs mugshot* protocol

2. *Mugshot vs CCTV* protocol

3. *CCTV vs CCTV* protocol

These three protocols are considered to extract conclusions that can be helpful in the forensic practice or for improving the traditional face recognition systems in these challenging scenarios.

In addition, three distances between subject and camera are analysed: *close*, *medium* and *far* distance. The analysis of these 3 configuration is also of great interest for forensics and face biometrics.

5.3.2.1. Mugshot vs Mugshot Protocol

This protocol has been defined in order to study the performance of different facial regions in a controlled scenario. For this particular case the subset of the MORPH database previously described is used.

This protocol compares good quality mugshot images against the same kind of images. The development set consists of the 6 available images per subject (1 image $\times$ 6 sessions) for 43 subjects (218 images in total, subjects 1 to 43). This set is used to train the PCA subspace.

Each subject model in the Test set (subjects 88 to 130) is then trained using the first session (s1), as Client data for SVM Training and all images from subjects 44 to 87 as Impostor data. As test images we consider the other 5 sessions (s2-s6) in the Test set. This dataset partitioning is summarized in Table 5.10.

5.3.2.2. Mugshot vs CCTV Protocol

This scenario is common in forensic laboratories, and it is very challenging because the difficulty in finding reliable similarities between doubted CCTV images and undoubted mugshot images from police records. For this reason, the results obtained in this scenario are specially helpful for the forensic practice.

In this case each subject model is trained using a single mugshot image (SVM Training Clients). Then, test images are taken from the 5 surveillance cameras at 3 different distances: *close*, *medium* and *far* (Test set). The Development and SVM Training sets are similar to the previous protocols as can be seen in Table 5.11.

5.3.2.3. CCTV vs CCTV Protocol

A third protocol was designed to compare CCTV against CCTV images. In this case the same variability factors (low resolution, pose, illumination, etc.) affect both train and test images (see *Cam1* images in Fig. 5.24 (top)). This protocol was defined in order to understand the performance when the training set is influenced by the same variability factors present in test images.

As shown in Table 5.12, the partitioning of the SCface DB into Development, SVM Training, and Testing is similar to the previous protocols, only considering in this case the information from the 5 surveillance cameras, and using the first one for modelling each subject (through SVM Training).

In this scenario the system is trained with images with *close* distance and compared with images from the three distances: *close*, *medium*, and *far*.

5.3.3. Results

This section describes the experimental results and findings achieved following the protocols described in Sect. 5.3.2. The main goal of the experiments is to study the discrimination power of the different facial regions.

	MORPH DB (130 Subjects) - Mugshot vs Mugshot protocol		
Subsets	1...43 Subject (43 Subjects)	44...87 Subject (44 Subjects)	88...130 Subject (43 Subjects)
*s*1	Development set (PCA subspace)	SVM Training (Impostors)	SVM Training (Clients)
*s*2			Test (Clients/Impostors)
*s*3			
*s*4			
*s*5			
*s*6			

Table 5.10: *Partitioning of the MORPH DB according to the Mugshot vs Mugshot images evaluation protocol.*

	SCface DB (130 Subjects) - Mugshot vs CCTV protocol		
Subsets	1...43 Subject (43 Subjects)	44...87 Subject (44 Subjects)	88...130 Subject (43 Subjects)
Mugshot	Development set (PCA subspace)	SVM Training (Impostors)	SVM Training (Clients)
Cam 1			Test (Clients/Impostors)
Cam 2			
Cam 3			
Cam 4			
Cam 5			

Table 5.11: *Partitioning of the SCface DB according to the Mugshot vs CCTV images evaluation protocol.*

	SCface DB (130 Subjects) - CCTV vs CCTV protocol		
Subsets	1...43 Subject (43 Subjects)	44...87 Subject (44 Subjects)	88...130 Subject (43 Subjects)
Cam 1	Development set (PCA subspace)	SVM Training (Impostors)	SVM Training (Clients)
Cam 2			Test (Clients/Impostors)
Cam 3			
Cam 4			
Cam 5			

Table 5.12: *Partitioning of the SCface DB according to the CCTV vs CCTV images evaluation protocol.*

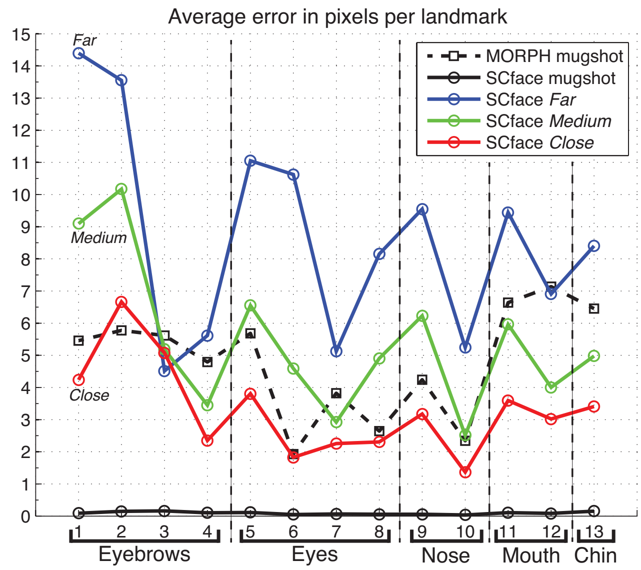


Figure 5.25: Comparative error analysis between the automatic facial landmark system and a manual examiner (ground truth) based on Euclidean distance in the different scenarios analysed. Pixel values are normalised to 240×200 *image size. Legend of landmark's number can be seen in Fig. 5.21.*

5.3.3.1. Comparison of Manual and Automatic Facial Landmark Detection

This experiment analyses the error introduced in the process of automatic facial landmark tagging with respect to manual tagging (ground truth) in order to understand if it can affect the performance. Fig. 5.25 shows the normalised average error in number of pixels for the 13 facial landmarks considered. Results are computed for the different datasets considered in this section.

As can be seen there is a notable difference between the mugshot subsets for the MORPH and SCface databases (i.e., error is much higher in MORPH mugshot). This difference is due to the original image size (resolution), MORPH images are 240×200 pixels, and SCface mugshot images are 3072×2048. Facial landmark tagging over high resolution images can be much more accurate compared to lower resolution images as concluded in previous Sect. 5.2.

CCTV images have a higher error compared to SCface mugshots. We have to note that the image size for the *close* (224×168), *medium* (144×108), and *far* (100×75) scenarios is different, which may be the main reason for the increasing error between *close* and *far*.

It is interesting to note that the landmarks for CCTV on the right part of the face image present a higher error compared the left side. This effect can be due to illumination or pose artifacts. The mugshot images do not present the previous effect.

As a result, we observe an increasing error between mugshot and CCTV, and between *close* and *far* distances for the automatic landmark detection compared to the labelling done by a human expert used as ground truth.

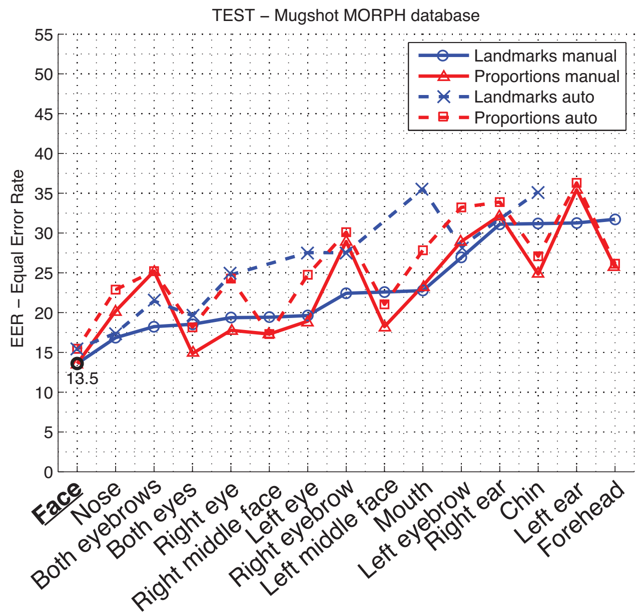


Figure 5.26: *EER values for the different facial regions extracted for the mugshot vs mugshot images scenario. Curves are ordered by the manual landmarks results.*

5.3.3.2. Mugshot vs Mugshot

This section presents the results for the mugshot versus mugshot scenario using the MORPH database. Results for both manual and automatic landmark tagging together with the two facial region extractors are presented and compared.

Better results are obtained with the manual landmark tagging in both extractors as shown in Fig. 5.26. This graph presents the Equal Error Rate (EER) of the whole *face* region compared with the rest of the facial regions extracted. The 15 facial regions are ordered from lower to higher EER (left to right).

The *face* region achieves the best recognition performance. Inner traits of the face such as the *nose*, *both eyebrows*, *both eyes*, etc., have better performance in this mugshot controlled scenario compared to the outer traits of the face such as the *ears*, *chin*, and *forehead*. This is in concordance with previous works [Faltemier *et al.*, 2008; Ocegueda *et al.*, 2011; Tome *et al.*, 2013a; Vera-Rodriguez *et al.*, 2013a].

Regarding the two facial region extractors, it is interesting to see that some regions achieve a better performance for the extractor based on the proportions (*both eyes, middle faces, chin*, and *forehead*), and some others achieve a better performance using the extractor based on the landmarks (*nose, both eyebrows, mouth,* and *ears*).

5.3.3.3. Mugshot vs CCTV

As discussed before, this is probably the most interesting and challenging scenario for forensic examiners. Results are shown in Fig. 5.27, where EER achieved for each facial region extracted

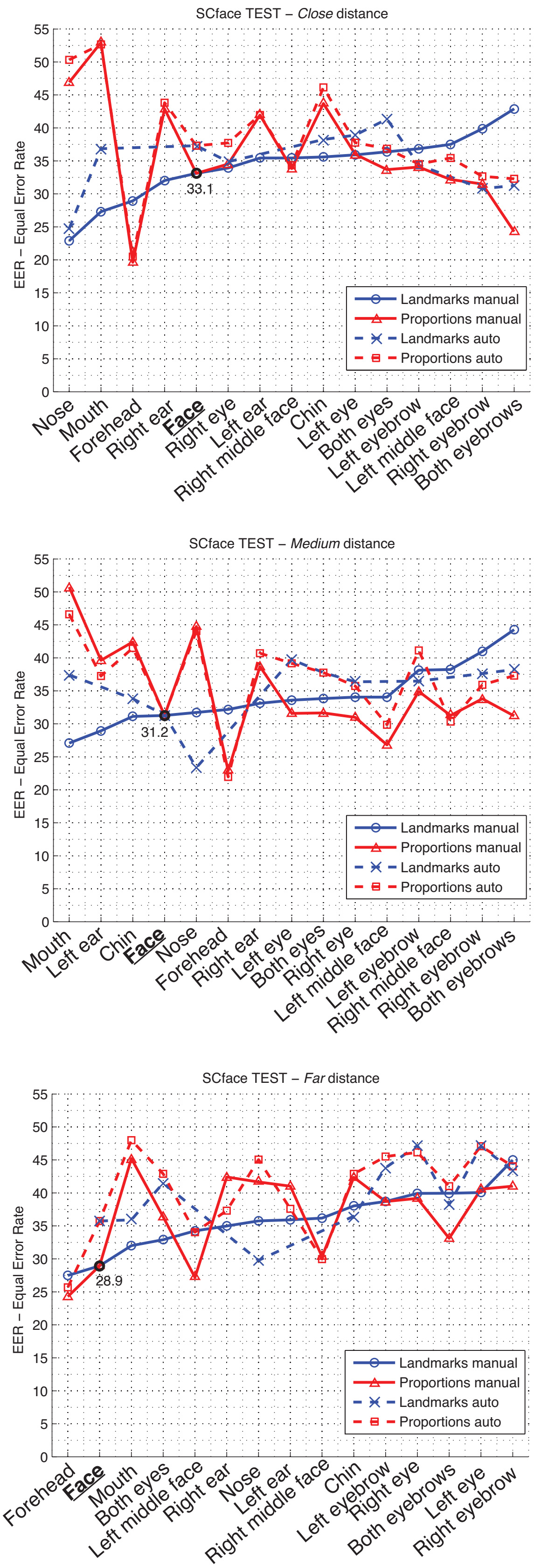


Figure 5.27: *EER values for the different facial regions extracted for the three different distances: close, medium and far for the mugshot vs CCTV images scenario.*

is represented over the three scenarios at a distance: *close*, *medium* and *far* for the SCface database.

Fig. 5.27 presents a very interesting experimental finding comparing the EER of the *face* region with the rest of the facial regions extracted. As can be seen, the recognition performance considering the whole face improves when the distance increases (31.1% to 28.9% EER for *close* and *far* distance, respectively). We believe this is mainly due to the varying acquisition angle. As can be seen in Fig. 4.7 this angle is smaller in the *far* scenario, and therefore the pose is more similar to the mugshot image. This can also explain that the best facial region performance is achieved for the *nose, mouth* and *forehead* for the three distances (*close*, *medium*, and *far*) respectively.

Similar to the previous scenario, here better results are obtained in all scenarios at a distance by the manual landmarks tagging for both extractors, as could be expected.

Another very interesting finding is the discriminative power achieved by the *forehead* region in the *far* distance which is better than the full *face*. These results could be due to the system used where the features based on PCA may be not representing well the full potential of mugshot and surveillance camera images.

From a global point of view, both extractors based on proportions and facial landmarks experiment different performances for the different regions, but it is interesting to see how they follow the same trend for the three distances considered (i.e., manual are usually better than automatic landmarks, and proportions-based is usually better than landmark-based region extraction). This effect could be due to that the extractor based on proportions estimates the facial regions approximately, which even using the rigid noise masks, can include more information and improves the EER.

These results suggest that for low quality images at a distance, facial regions could be extracted just using two points (eye coordinates), which an automatic face recognition system can locate easily and the recognition result of each region would be similar to the one obtained using a more sophisticated facial landmark extractor.

5.3.3.4. CCTV vs CCTV

This scenario presents better performances in all distances than the analysed before on the SCface database. This improvement could be because the training and testing sets include more or less the same environmental variability (see *Cam1* images in Fig. 5.24 (top)), and the effect of the acquisition angle (pose) is not so important considering that all cameras are in the same static position and never totally frontal as the acquisition of mugshot images. Here, the system is trained with images from the *close* scenario and compared with images from the three scenarios: *close*, *medium*, and *far*.

It is important to emphasize the unexpected by good performance achieved in this scenario, which is better than the one in the *mugshot vs mugshot* scenario. This result can be explained by the much larger time lapse between training and testing data for MORPH DB (mugshot vs mugshot) compared to SCface (CCTV vs CCTV).

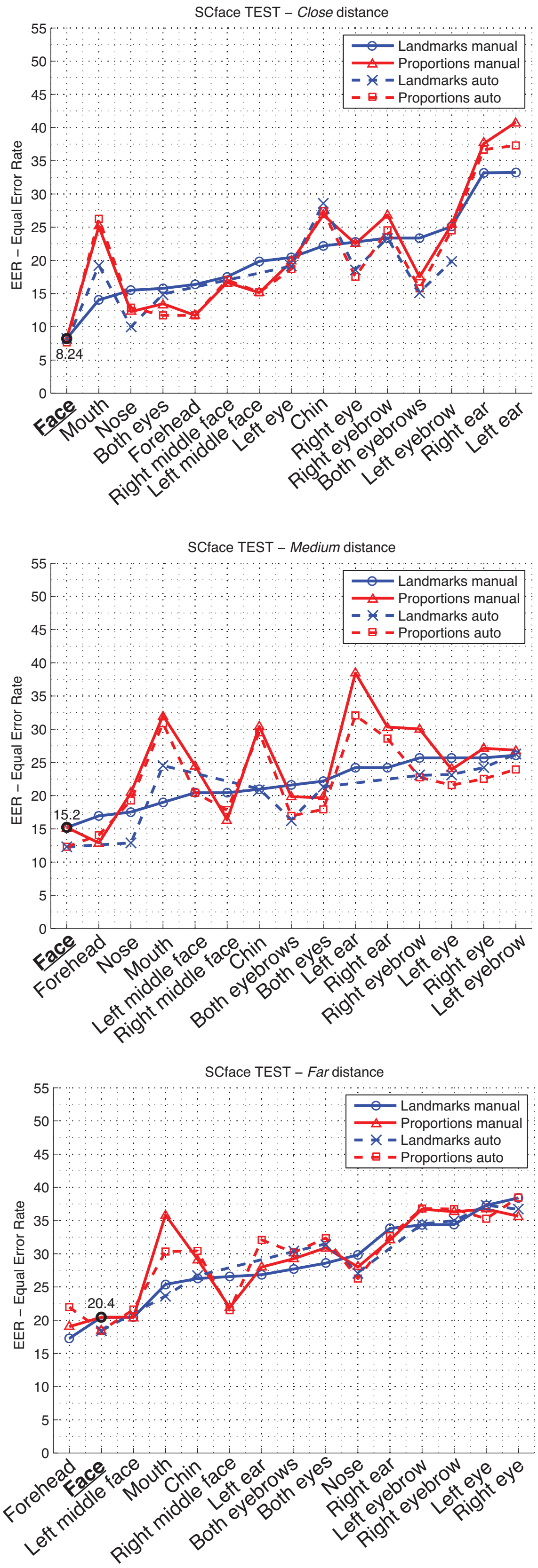


Figure 5.28: *EER values for the different facial regions extracted for the three different distances: close, medium and far for the CCTV vs CCTV scenario.*

Results are shown in Fig. 5.28 where a similar tendency between both extractors under study is observed. In particular it is very remarkable the similar performance for the *far* scenario. This demonstrates that, the proposed simple region extractor based on only eye coordinates and face proportions could be very useful for unconstrained scenarios at a distance.

In general terms there is a decrement of performance when the distance increases. In this case, as expected because training and testing conditions are similar here, the performance of the *face* region decreases with the distance between the subject and the camera.

The best performances are achieved for the *face, forehead, nose* and *mouth* in the three distances. Again the *forehead* region reaches the best performance in the *far* scenario and the second position in the *medium* distance scenario. This could be because 115 subjects of the database are male and just 15 are female. Male subjects usually have short hair and therefore the forehead is free of occlusions. Female subjects on the other hand, usually have long hair and more occlusions which may lead to decreased performance in this region. The *forehead region* reaches an important role in this uncontrolled scenario in comparison with the previous *mugshot versus mugshot* scenario. While in controlled scenarios the *forehead* region achieved the worst results, here this facial region outperforms the other facial regions. Even in the *medium* and *far* scenarios this facial trait is one of the most discriminative.

5.4. Chapter Summary and Conclusions

In this chapter we have performed three studies related to the biometric variability in surveillance and forensic scenarios.

The first study analyses the variability factors in face recognition at a distance. In particular, we have conducted a data-driven analysis of three realistic acquisition scenarios at different distances (close, medium, and far), as a first step towards devising adequate recognition methods capable to work in less constrained scenarios. This data-driven analysis has been made for a subset of the benchmark dataset NIST MBGC v2.0 Face Stills. Our analysis has been focused on: *i*) data statistics (segmented face sizes, resolutions, quality and entropy measures), and *ii*) verification performance of three systems. The results showed that the considered systems degrade significantly in the far distance scenario, being more robust to uncontrolled conditions the simplest approach based on DCT-GMM. Noteworthy, the scenarios considered in the present Dissertation differ not only in the distance factor, but also in illumination and pose (being the illumination variability much higher in far distance than in close distance).

The second study presented in Sect. 5.2 reports an study of the variability of facial landmarks over two mugshot and CCTV databases, in controlled and uncontrolled scenarios with low quality images and large range of variability factors. Mugshot images are taken with three distances between the persons and the camera (1, 2, and 3 meters) showing the face, the upper body and the full body respectively. CCTV face images are also taken with three distances between the persons and the 5 CCTV cameras (1, 2.60 and 4.20 meters). 21 facial landmarks were defined and the database was manually tagged imitating the procedure followed by a forensic examiner.

Also, an automatic system was used to tag 13 out of the 21 landmarks defined. The main conclusions are that the landmarks located in the outer part of the face have a much higher variability compared to the landmarks placed near the eyes. A reason for this is mainly that these areas can have more hair occlusions than other areas. This is more accentuated for the females.

Regarding the distances between the camera and the person, the landmarking variability increases with the distance. In mugshot scenario this effect is not very significant. In the case considered here, images taken at 3 meters have around 100 pixels between the center of the eyes, which is not a very small size. In case of having mugshots taken for farther distances we believe that the landmarking variability would be much higher. Comparing the two manual and automatic tagging approaches, the results show that the landmark variability is very similar for the set of common landmarks; having in some cases lower variability for the automatic system.

A final comparison of both scenarios mugshot and CCTV scenarios shows that the CCTV images present significantly higher landmark variability, which is mainly due to lower quality of the images making very difficult to tag the facial landmarks with high precision.

Finally, last Sect. 5.3 reports an exhaustive analysis of the discriminative power of the different regions of the human face on various forensic scenarios. We first described an experimental framework to extract 15 different facial regions of a human face following forensic protocols with 4 variants: manual or automatic landmark detection, and then region extraction based either on the full set of landmarks (13 and 21 for automatic and manual landmark detection, respectively) or only the eye coordinates and general face proportions.

The comparison of the two region extractors resulted in better recognition performance for the outer facial regions when using the extractor based on proportions. In contrast better recognition performance is achieved for the inner facial regions when using the extractor based on landmarks. As a result, we obtain that the extractor based on proportions can be very useful in scenarios at a distance, where obtaining reliable landmark information with automatic systems is very difficult. Also interestingly, similar performance is obtained with both extractors in the *far* scenario where images are degraded. This means that for low quality images at a distance, facial regions could be extracted just using two points (eye coordinates), and the recognition result of each region would be similar to a facial landmark extractor.

Furthermore, we studied three scenarios with different distance between subject and camera common in forensic casework. In all cases, we obtain that the recognition performance of facial regions depends on the acquisition distance. The best three facial regions with high discrimination power in the *close* distance are the *face, nose*, and *forehead*. However in *far* distance, the best performance is achieved by the *forehead*. This facial region acquires an important role on scenarios at a distance such as CCTV versus CCTV. It was noted that this result could be due to the great majority of short hair males, as with females that region may be much more reliable.

In the most common forensic scenario (*mugshot vs CCTV* images), variability factors have a high importance and produce a decrement of recognition performance with respect to the more controlled *mugshot vs mugshot* scenario. In addition to be a useful background information that can guide and help experts to interpret and evaluate face evidences, these findings can have a significant impact on the design of face recognition algorithms.

This chapter includes novel contributions in the consistent and replicable methodology used, the face region extraction methods developed, the scenario at a distance analysis, the facial landmarks variability related findings, and in the findings about the discriminative power of different facial regions extracted from a human face.

Chapter 6

Soft Biometrics

This chapter studies the variability and discrimination power of soft biometric information useful for forensics and video surveillance applications.

As introduced in Chapter 3, the proposed soft biometrics are either continuous or discrete. Traits such as gender, eye color, ethnicity, etc. are discrete in nature. On the other hand, traits like height and weight are continuous variables. Usually a system that is completely based on soft biometric traits cannot provide the same accuracy level in the recognition of individuals compared to more distinctive traits such as fingerprints or iris. However, soft biometric traits can be used to improve the performance of a traditional biometric system (e.g., gait, face, etc.) in many ways.

As described in Chapter 3, soft biometric information provides to biometric systems at a distance a complementary information to improve the identification and recognition rates.

The use of soft biometric traits in automated human recognition systems has several benefits. It is, therefore, essential to carefully investigate issues related to its extraction and recognition capacity. Surveillance footage is generally of inferior quality and so traditional forms of identification at a distance cannot be easily used. Soft biometrics for video surveillance offers a solution in this regard but lacks the distinctiveness that is expected from biometric traits.

In forensic practice, examiners carry out a extensive morphological manual inspection of the face images, focussing their attention not only on the full face but also on individual traits. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc. In this sense facial features have been already analysed in the literature, but the differences of the facial traits in a human face are not yet well understood. In the second part of this chapter, we propose two large sets of continuous and discrete facial soft biometric features based on the morphological analysis of forensic laboratories such as Spanish Guardia Civil (DGGC) or Netherlands Forensic Institute (NFI). The benefits of these facial features, among others, are: 1) they are extracted automatically from the facial landmarks of a frontal face image, allowing us to extract population statistics from large databases, 2) they are suitable for person recognition, and 3) they can be fused with face information in order to improve the recognition performance.

The main contributions of this chapter are:

- An experimental study of the benefits of soft biometric labels as ancillary information for challenging person recognition scenarios at a distance.

- Evaluation of the proposed facial soft biometric features based on morphological face features.

In particular, we provide experimental evidence on how the soft labels of individuals witnessed at a distance can be used to improve their identification and help to reduce the effect of variability factors in these scenarios. The value of facial soft features as a unique source of information for person recognition is also provided.

The chapter is structured as follows. One section is dedicated to each of the studied soft biometric feature sets for video surveillance and forensics (Sects. 6.1 and 6.2, respectively). Both sections share a common structure: first, an introduction and description of the soft features is given, followed by a detailed features analysis (correlation, stability, and discrimination power of each soft feature used). Then the systems used in the evaluation are presented. The database and experimental protocol are described in another subsection, and finally we analyse and discuss the results, where training set size, individual and grouped features, and feature selection are evaluated. The summary and conclusions of the chapter appear in the final section (Sect. 6.3).

This chapter is based on the publications: Tome *et al.* [2013b,d].

6.1. Soft Biometrics for Video Surveillance

A wide variety of biometric systems have been developed for automatic recognition of individuals based on their physiological/behavioural characteristics. These systems make use of a single or a combination of traits like face, gait, iris, etc., for recognizing a person. On the other hand, the use of other ancillary information based on the description of human physical features for face recognition [Park and Jain, 2010] has not been explored in much depth.

Biometric systems at a distance have an outstanding advantage: they can be used when images are acquired non-intrusively at a distance and other biometric modes such as fingerprint cannot be acquired properly. Given such situations, some biometrics can experiment low performance due to variability factors caused by the acquisition at a distance but they can still be perceived semantically using human vision. In this section we analyse how these semantic annotations (labels) are used as soft biometric signatures, useful for identification tasks.

A research line growing in popularity is focused on using this ancillary information (soft biometrics) in less constrained scenarios in a non-intrusive way, including acquisition "on the move" and "at a distance" [Li *et al.*, 2009]. These kinds of scenarios are still in their infancy, and much research and development is needed in order to achieve the levels of precision and performance that certain applications require. As a result of the interest in these biometric applications at a distance, there is a growing number of research works studying how to compensate for the main degradations found in uncontrolled scenarios [Robust2008, 2008].

The main contribution of this section is an experimental study of the benefits of soft biometric labels as ancillary information for challenging person recognition scenarios at a distance. In particular, we provide experimental evidence on how the soft labels of individuals at a distance can be used to improve their identification and help to reduce the effect of variability factors in these scenarios.

In order to do so, the largest and most comprehensive set of soft biometrics available in the literature is first described. These soft biometrics labels (called from now on **soft labels**) are manually annotated by several experts. These soft labels have been grouped considering three physical categories: global, body and head. The stability of the annotations of the different experts and their discriminative power are also studied and analysed.

Finally, the available soft biometric information in scenarios of varying distance between camera and subject (*close*, *medium* and *far*) has been analysed. The rationale behind this study is that depending on the particular scenario, some labels may not be visually present and others may be occluded. As a result, the discriminant information of soft labels will vary depending on the distance.

The experimental framework used in this PhD Thesis is detailed in Sect. 3.1. This section explains how from a video at a distance of a person walking, soft labels and faces from a subject are extracted. In this case, human annotators extract soft labels manually because this process is still far from being implemented by an automatic system.

This section includes novel contributions in the experimental findings about the relation between the distance and the performance of soft biometrics for recognition at a distance.

6.1.1. Soft Biometrics Data Analysis

In this section a set of soft biometrics has been used, whose main value is that it is discernible by humans at a distance. These are detailed in Sect. 3.1.1.

These physical trait labels are obtained from the Southampton Multibiometric Tunnel Database (TunnelDB), previously described in Chapter 4, which contains biometric samples from 227 subjects for which 10 gait sample videos from between 8 to 12 viewpoints are taken simultaneously. The TunnelDB database also contains high-resolution frontal videos to extract face information and high-resolution still images taken to extract ear biometrics. There are roughly 10 of such sets of information gathered for each subject.

The TunnelDB datasets were annotated against recordings taken of the individuals in laboratory conditions (see Sect. 4.2.3). A range of discrete values is given to each trait label, e.g. "Arm length" marked as 1 (very short), 2 (short), 3 (average), 4 (long), and 5 (very long). The annotation process of each label is described in detail in [Samangooei, 2010]. A summary of these trait labels and their associated discrete semantic terms is provided in Table 3.1 of Chapter 3.

As is explained there, these labels were designed based on which traits humans are able to consistently and accurately use when describing people at a distance. The traits were grouped in 3 classes, namely: *Global*, *Body*, and *Head*.

To understand the role of soft labels and their application to biometrics at a distance, the internal correlation, the stability, and the discrimination power of the different labels with semantic annotations is studied and analysed in the next Section. In the next experiments, a total of 13.340 labels from 58 subjects annotated by 10 different experts are used. The remaining subjects in TunnelDB were annotated only by just 1 or 2 different experts and were rejected for this analysis.

6.1.1.1. Correlation Between Labels

This section reports an analysis of the correlation between the labels defined. For this purpose the correlation between all pair of labels of the three groups defined (*global*, *body* and *head*) is computed using the Pearson's correlation coefficient:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \tag{6.1}$$

where σ_{XY} represents the covariance of the two variables X and Y divided by the product of their standard deviations σ_X and σ_Y. The variables X and Y represent numerical values associated to the pairs of semantic terms at hand. Here each semantic term was converted to numerical values in the range 1 to 5 if the annotation contains the semantic term (e.g. very short, short, average, long and very long) and 0 if the annotation was left empty by the annotator (he was not sure what to annotate). X_i and Y_i are the label values across all individuals and annotators, therefore $N = 580$ annotations (58 subjects $\times$ 10 annotators). The value r provides the correlation coefficient which ranges from -1.0 to 1.0. A value of 1.0 implies that a linear equation perfectly describes the relationship between X and Y, with all data points lying on a line for which Y increases as X increases. A value of -1.0 implies that all data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

The correlation matrix containing the correlation between all labels is represented graphically in Fig. 6.1. Colours in the red range represent correlation coefficients close to 1.0 and thus a positive correlation, while colours in the blue range represent correlation coefficients close to -1.0 and thus a negative correlation. Pale green represents no correlation between labels.

Focusing our attention in the *global* labels, very small correlation between these 3 features and all the remaining ones is observed in the graph as could be expected. On the other hand, some *body* labels are very correlated between them mainly due to the proportion relationships of the human body (e.g., the larger the arms the larger the legs). This means that physical characteristics like the chest (3), and the figure (4) are very correlated. Therefore if we try to recognize people just by using these correlated features the success rate won't be very high.

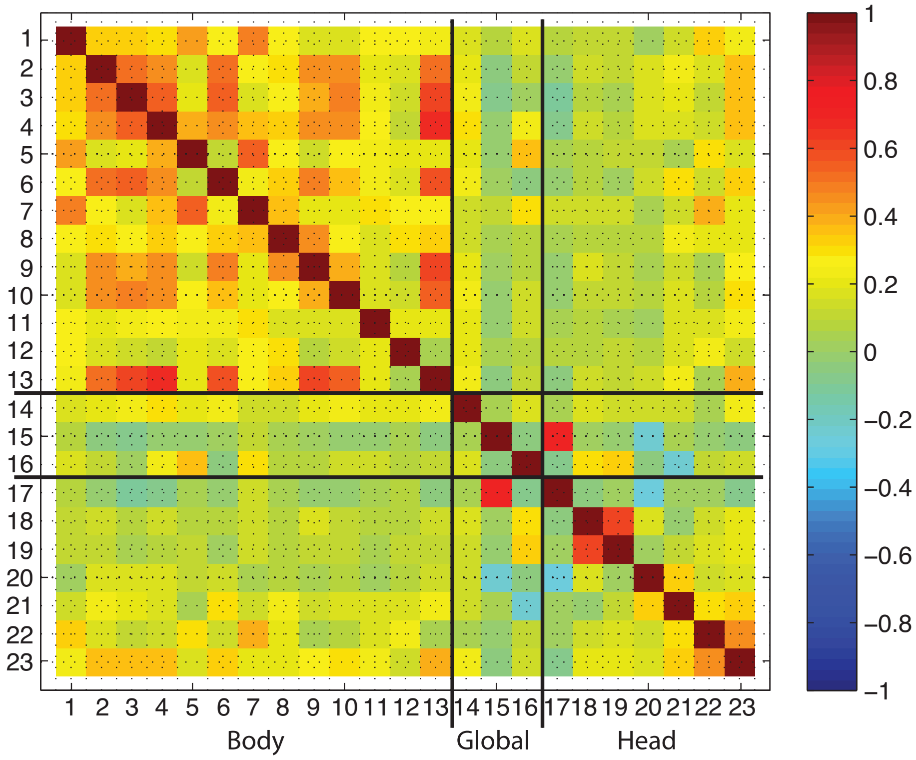


Figure 6.1: *Correlation between labels based on Pearson's coefficient r (see Eq. (6.1)). The numbering of soft labels is detailed in Table 3.1.*

Head features do not present the same correlation between them compared to *body* traits (except e.g. facial hair colour (18) and facial hair length (19) or neck length (22) and neck thickness (23) which are highly correlated).

Fig. 6.1 also shows some strong relationships between demographic traits such as ethnicity (15) and skin colour (17), or hair colour (20), as was expected.

6.1.1.2. Stability Analysis of Annotations

This section reports an analysis of the stability of the human annotations for all soft labels. This is done by calculating the stability coefficient, defined for label X as:

$$\text{Stability}_X = 1 - \frac{1}{SA}\sum_{i=1}^{S}\sum_{a=1}^{A}|X_{ia} - \text{mode}_a(X_{ia})| \qquad (6.2)$$

where X_{ia} is the annotated value for subject i by annotator a, $A = 10$ is the total number of annotators, $S = 58$ is the total number of subjects, and $\text{mode}_a(X_{ia})$ is the statistical mode across annotators (i.e., the value most often annotated for subject i).

The resulting stability coefficients for all labels are depicted in Fig. 6.2. Using the definitions in chapter 11 of [Theodoridis and Koutroumbas, 2008], we can see that some of the features are *nominal*, i.e., their values can not be ordered meaningfully (e.g., ethnicity (15), sex (16), skin (17), facial hair (18) and hair colour (20)) whereas others are *ordinal*, i.e., their values can be

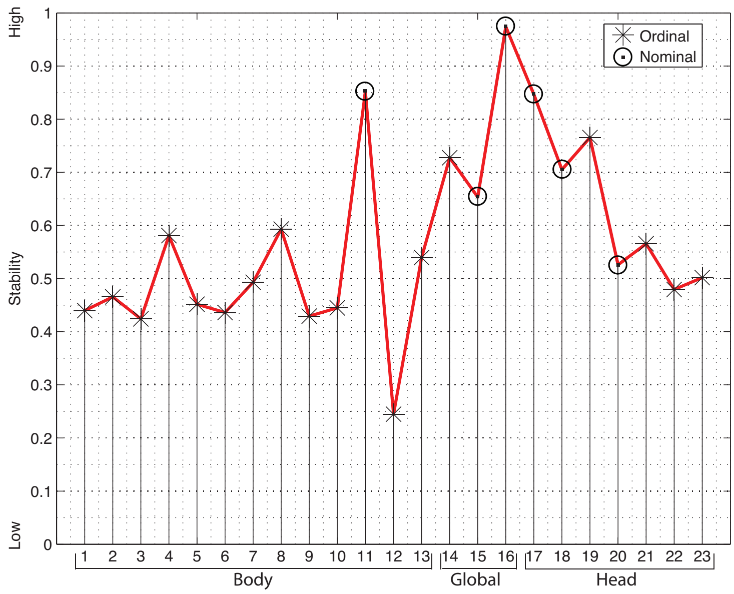


Figure 6.2: *Annotators' stability for the* 23 *soft labels considered (see Table 3.1).*

meaningfully ordered (e.g., arm length (1), arm thickness (2), height (4), weight (13), and hair length(21)).

In Fig. 6.2 we can see that sex (16) (a *nominal* label that has just two terms, male and female), is the most stable label due to the low variability. Other *nominal* features such as body proportions (11) and skin colour (17) have also high stability. On the other hand, the stability of *ordinal* features such as arm length (1), height (5), hips (6), or shoulder shape (12) is lower due to the high variability and the different point of view of the annotators.

Although these two types of features (*nominal* and *ordinal*) may be processed differently (e.g., using different similarity measures), here in this PhD Thesis we have processed them in the same way as an initial approach.

6.1.1.3. Discriminative Power Analysis

In order to evaluate the discriminative power of the soft label X, we compute for it the ratio between the inter-subject variability, and the intra-subject variability as follows:

$$\text{Discrimination}_X = \frac{\frac{1}{S(S-1)} \sum_{i=1, i \neq j}^{S} \sum_{j=1}^{S} |\mu_i - \mu_j|}{\sigma} \tag{6.3}$$

$$\mu_i = \underset{a}{\text{mean}}(X_{ia}), \;\; \mu_j = \underset{a}{\text{mean}}(X_{ja}), \;\; \sigma = \frac{1}{S}\sum_{i=1}^{S} \sigma_i, \;\; \sigma_i = \underset{a}{\text{std}}(X_{ia}) \tag{6.4}$$

where i and j index subjects, and a indexes annotators.

The discrimination coefficient for the X^k labels ($k = \{1, ..., K = 23\}$) is depicted in Fig. 6.3. There we can see that the *body* features (IDs 1-13) are less discriminant than the *global* (IDs 14-16) and *head* (IDs 17-23) features. The least discriminant features are the arm length (1)

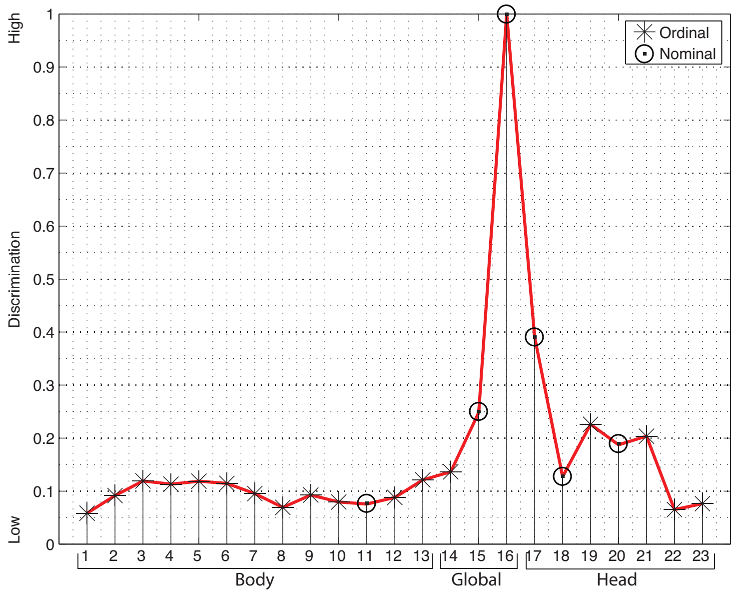


Figure 6.3: *Discrimination power of the* 23 *soft labels considered (see Table 3.1).*

and neck length (22) followed by leg direction (8) and neck thickness (23). These are *ordinal* features and therefore the majority of the subjects share similar annotations.

Better results are reached for the *nominal* features such as ethnicity (15), or skin color (17), and the most discriminative is the sex (16) due to the clear identification by the human annotators in the TunnelDB database. Consequently, we can predict that *global* and *head* features will provide better person recognition results than *body* features.

6.1.2. Verification Based on Soft Biometrics

This section describes a person verification system based only on soft biometrics. First, each label in numeric form (see Sect. 6.1.1) is normalised to the range [0,1] using the tanh-estimators described in [Jain *et al.*, 2005]:

$$\hat{X}^k = \frac{1}{2}\left\{\tanh\left(C\left(\frac{X^k - \mu_{X^k}}{\sigma_{X^k}}\right)\right) + 1\right\} \tag{6.5}$$

where $C = 0.01$, X^k is the $k = \{1, ..., K\}$ soft label ($K = 23$), $\hat{X}^k$ denotes the normalized label, and μ_{X^k} and σ_{X^k} are respectively the estimated mean and standard deviation of the label under consideration (see Table 3.1 for the list of the labels).

Note that, depending on the scenario considered (*close*, *medium*, and *far*), there are $K = 12$, 17, or 23 labels, respectively (see Table 3.2).

Similarity scores $s(\mathbf{x}, \mathcal{C})$ are computed using the Mahalanobis distance Theodoridis and Koutroumbas [2008] between the test vector with K labels $\mathbf{x} = [\hat{X}^1, ..., \hat{X}^K]^T$ and a statistical model $\mathcal{C}$ of the client, obtained using a number of training labels (9 examples per label in

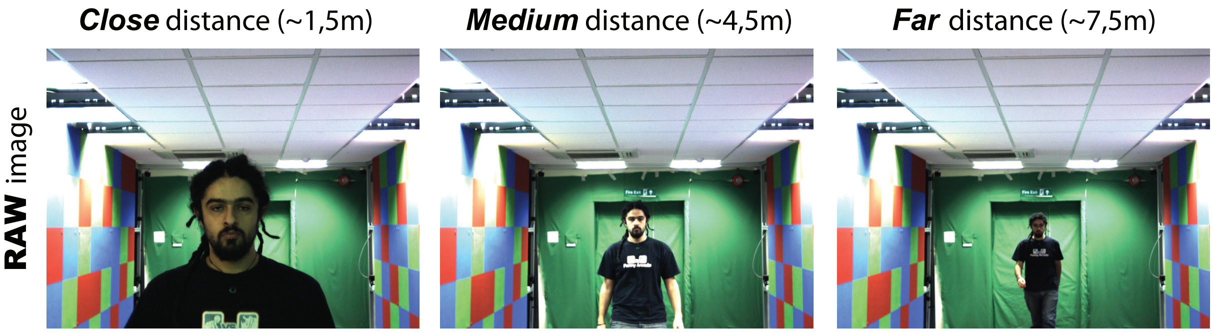


Figure 6.4: *Scenario defined based on the TunnelDB Seely et al. [2008]: close, medium and far distance images used in the experimental work.*

our experiments), as follows:

$$s(\mathbf{x}, \mathcal{C}) = \frac{1}{\left((\mathbf{x} - \boldsymbol{\mu}^{\mathcal{C}})^T \, (\boldsymbol{\Sigma}^{\mathcal{C}})^{-1} \, (\mathbf{x} - \boldsymbol{\mu}^{\mathcal{C}}) \right)^{1/2}} \tag{6.6}$$

where $\boldsymbol{\mu}^{\mathcal{C}}$ and $\boldsymbol{\Sigma}^{\mathcal{C}}$ are respectively the mean vector and covariance matrix obtained from the training labels, which form the statistical model of the client $\mathcal{C} = \{\boldsymbol{\mu}^{\mathcal{C}}, \boldsymbol{\Sigma}^{\mathcal{C}}\}$.

6.1.3. Database and Experimental Protocol

The physical trait labels were obtained from the Southampton Multibiometric Tunnel Database (TunnelDB) [Seely *et al.*, 2008] as described previously in Chapter 4. Three different challenging scenarios varying the distance between camera and subject have been defined and used in our experiments in order to understand the behaviour of soft biometric labels and their best application to biometrics at a distance. For this purpose, high resolution frontal face sample videos from the TunnelDB database (see an example in Fig. 6.4) have been used together with their corresponding physical soft labels analysed in the previous sections. This process is detailed in Sect. 3.1.1 of the Chapter 3. The face recognition results will be presented in Chapter. 7.

Three different scenarios are defined at three different distances: *i*) *close*, including both the face and the shoulders, *ii*) *medium*, including the upper half of the body, and *iii*) *far*, including the full body. The rationale behind this study is the fact that depending on the particular scenario, some labels may not be visually present and others may be occluded. As a result, the discriminative information of the soft biometrics will vary depending on the distance. Table 3.2 in Chapter 3 shows the soft labels available for each of the scenarios defined.

The experimental protocol followed is as follows. As was introduced in Sect. 6.1.1, the dataset selected for the soft labels from the TunnelDB was comprised of 58 subjects annotated by 10 annotators. Therefore, each user has 10 sessions, so 580 soft biometric templates per scenario from the soft labels detailed in Table 3.1 have been used, having in total 1740 soft biometrics templates (58 subjects $\times$ 10 sessions $\times$ 3 distances).

The database was divided into training and testing sets. For each subject a number of face

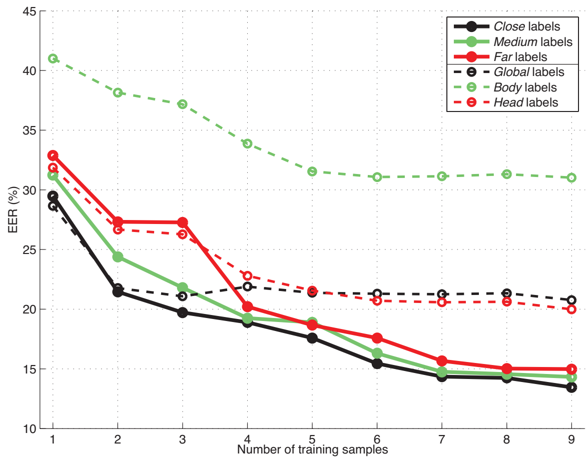


Figure 6.5: *EER (%) obtained when varying the number of training samples.*

images and sets of soft labels (ranging from 1 to 9 samples) were used for the training and one of the remaining samples was used for testing following a leave-one-out approach [Theodoridis and Koutroumbas, 2008] generating this way 580 similarity target scores and 33640 similarity non-target scores in all tests.

6.1.4. Results

This section describes the experimental analysis of the discrimination power of individual and grouped soft labels and the performance of them in the three scenarios defined. Results are reported using ROC curves, with EERs and verification rates (VR) working at different FAR points (FAR = 0.1%, 1%, and 10%).

6.1.4.1. Analysis of Training Set Size for Soft Labels

An important parameter to be consider in soft labels systems is the size of the training set. For this purpose, we have evaluated the system with different number of training samples (varying between 1 to 9 samples) following a leave-one-out methodology (i.e., rotating 10 times the selected training set with one sample in each rotation left out for testing).

Fig. 6.5 shows the different configurations analysed for the six sets of soft labels defined in the previous section. As can be seen all soft label sets follow the same trend, the system recognition performance (EER) improves significantly when more samples are used in the training stage. For *global*, *body*, and *head* sets using more than 5 training samples the system performance saturates. On the other hand, for *close*, *medium*, and *far* sets, the performance saturates for more than 7 training samples.

It is also interesting to note that the more features are included in the set (e.g., for *far* labels which include all 23 labels) the larger the performance improvement for increasing training samples until saturation. The relative performance improvement before the saturation for small datasets (e.g., *global* with only 3 labels) is much smaller.

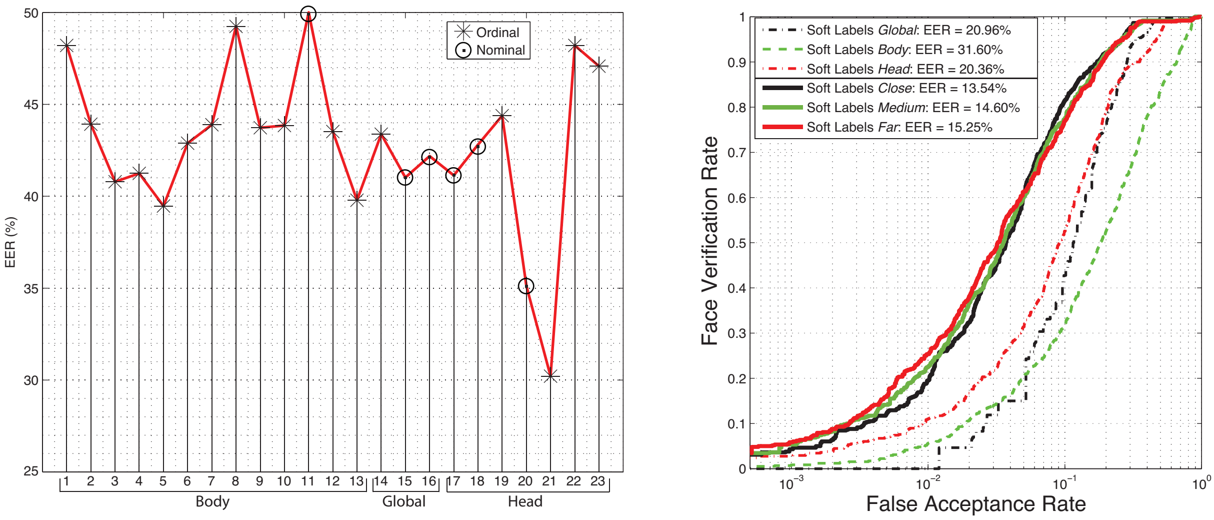


Figure 6.6: *On the left, EER (%) obtained for each individual soft label defined in Table 3.1. On the right, ROC curves obtained for the physical labels sets (global, body and head) and for the three defined scenarios (close, medium and far).*

6.1.4.2. Analysis of Individual Soft Labels

This section presents the discrimination power of each individual soft label following the leave-one-out experimental protocol described in Sect. 6.1.3. As shown in Fig. 6.6 (left), hair length (21) achieves the best results (EER = 30.27%) but it is worth noting that this was not the most discriminative feature regarding the initial experiments shown in Fig. 6.3. Another relevant label with a high performance and discrimination power is hair color (20) with an EER = 35.11%. The rest of soft labels achieve similar performance, with better results in general for *head* labels compared to *body* labels, as anticipated in Section 3.3. As can be seen, individual labels are not very discriminative on their own.

6.1.4.3. Analysis of Grouped Soft Labels

The aim of this experiment is to study the discrimination power of the three groups of soft labels considered in the different scenarios at a distance defined in Section 4.1. Fig. 6.6 (right) shows the performance of each set of labels considered. Here, dashed lines represent the sets: *global*, *body* and *head*, while solid lines represent all the available labels in each scenario at a distance as defined in Table 3.2.

There is a significant difference between *global*, *head* and *body* regarding the performance as can be observed. The performance of *body* labels is clearly lower compared to *global* and *head* sets as predicted in Sections 3.2 and 3.3 through the stability and discrimination analysis.

Regarding the other 3 groups of labels that take into account the labels visible at the 3 distances defined the difference of performance is not that significant as can be seen in Fig. 6.6 (right). *Far* scenario is comprised of all available labels including *body* labels, therefore it experiments a decrement of EER performance compared to the other scenarios in some regions of the plot (e.g., around FAR = 10% = 0.1). On the other hand, the other two scenarios have a lower number of soft labels available but result in better EER performance.

It is important to note that although soft labels provide low recognition performance when used as a stand alone system, they can help to improve hard biometric systems as will be shown in Chapter 7.

6.2. Soft Biometrics for Forensics

The first system in the history that attempted to describe people for identification based on the morphological and physiological traits was the anthropometric system developed by Bertillon [1896]. This system was based on three sets of features: *i*) body measurements (anthropometry) like height and length of the arm, *ii*) morphological description of the appearance and shape of the body like eye color and anomalies of the fingers, and *iii*) peculiar marks observed on the body like moles, scars, and tattoos.

More recently, most of the forensic laboratories follow methodologies based on Bertillon's system such as the police sketch used by Spanish Guardia Civil or NFI laboratories in the identification of criminals. This document consists of a verbal description of specific facial traits using the information given by a person that observed the subject/criminal.

Traits such as eyebrows height and width, interocular distance, naso-labial height, etc. are continuous variables in nature. These continuous traits can be converted to discrete values using thresholds in order to improve their classification and compute population statistics.

The main contribution of this section is an experimental study of the benefits of facial soft biometric features extracted automatically from a face image for forensics. In particular, we provide experimental evidence on how the facial soft labels of individuals can be used to reduce the effect of variability factors in this challenging scenario.

In order to do so, the set of facial soft biometric features proposed in Chapter 3 is analysed. These facial soft biometrics labels are automatically extracted by an automatic system based on facial landmarks following the procedure of the morphological analysis of forensic laboratories. These facial soft features have been grouped considering two categories: continuous and discrete values. The stability of such features and their discriminative power are also studied and analysed. Finally, we exploit the automatic extraction of these features to generate population statistics over large databases. These statistics are important for forensic practitioners and towards a better understanding of the facial information content.

The experimental framework used in this Thesis is introduced in Sect. 3.1.2 and detailed in Fig. 6.7. This Section describes how from a frontal mugshot image of a subject, facial soft biometrics features are extracted. In this case, a human examiner extracts facial landmarks manually in order to discard the variability introduced by the system in the analysis, but the process can be made by an automatic system.

This experimental framework implemented is based on protocols from international forensic laboratories [Netherlands Forensic Institute (NFI); Spanish Guardia Civil (DGGC)], and allows

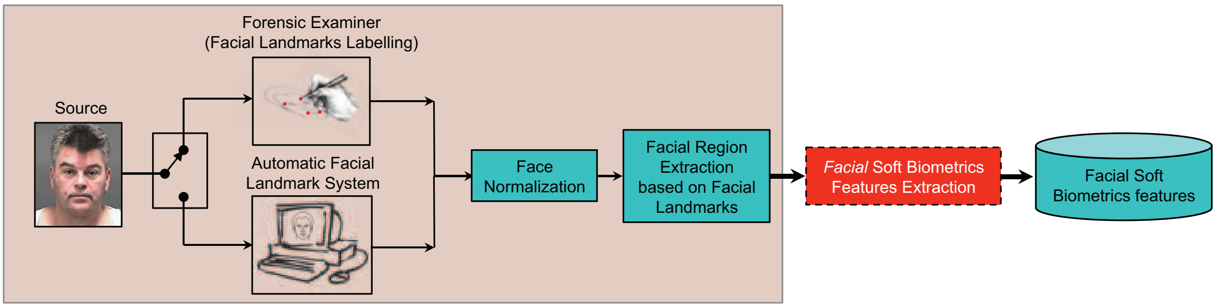


Figure 6.7: *Experimental framework followed to extract facial soft biometrics features. The system has two configurations: manual or automatic for facial landmark extraction.*

us the extraction of facial soft biometrics features. This procedure (summarized in Fig. 6.7) uses the facial landmarks extracted from a human face together with an extractor based on facial regions in order to extract all the facial soft biometric features proposed.

6.2.1. Soft Biometrics Data Analysis

In this section the proposed sets of facial soft biometrics features, described is detailed in Sect. 3.1.2 and summarized in Table 3.3, are analysed. The proposed facial soft biometric features are divided in *continuous* measures and *discrete* labels extracted from a face image of a subject.

These physical trait labels are obtained from two mugshot databases ATVS Forensic DB [Vera-Rodriguez *et al.*, 2013a] and a subset of MORPH DB [Ricanek and Tesafaye, 2006], which contain frontal face biometrics samples from 50 and 130 subjects, respectively. The first database provides high resolution images and the second MORPH database is comprised of low resolution frontal images.

In the next Section we study the application of the proposed facial soft features to forensics, their internal correlation, their stability, and their discrimination power.

6.2.1.1. Population statistics

The extraction of discrete features allows us the analysis of the population from a statistical point of view. This means that we can automatically analyse the physical traits of the human face across a population, which is very interesting for forensics. In forensics, the examiners usually carry out a manual inspection of the face images, focussing their attention not only on the full face but also on individual traits. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc. Population statistics of such traits give the examiners a very useful information towards their decisions.

This section analyses the statistics of two populations from two databases with high and low resolution images. For this purpose the distribution of each facial trait assigned to the discrete features across all subjects in the databases are calculated.

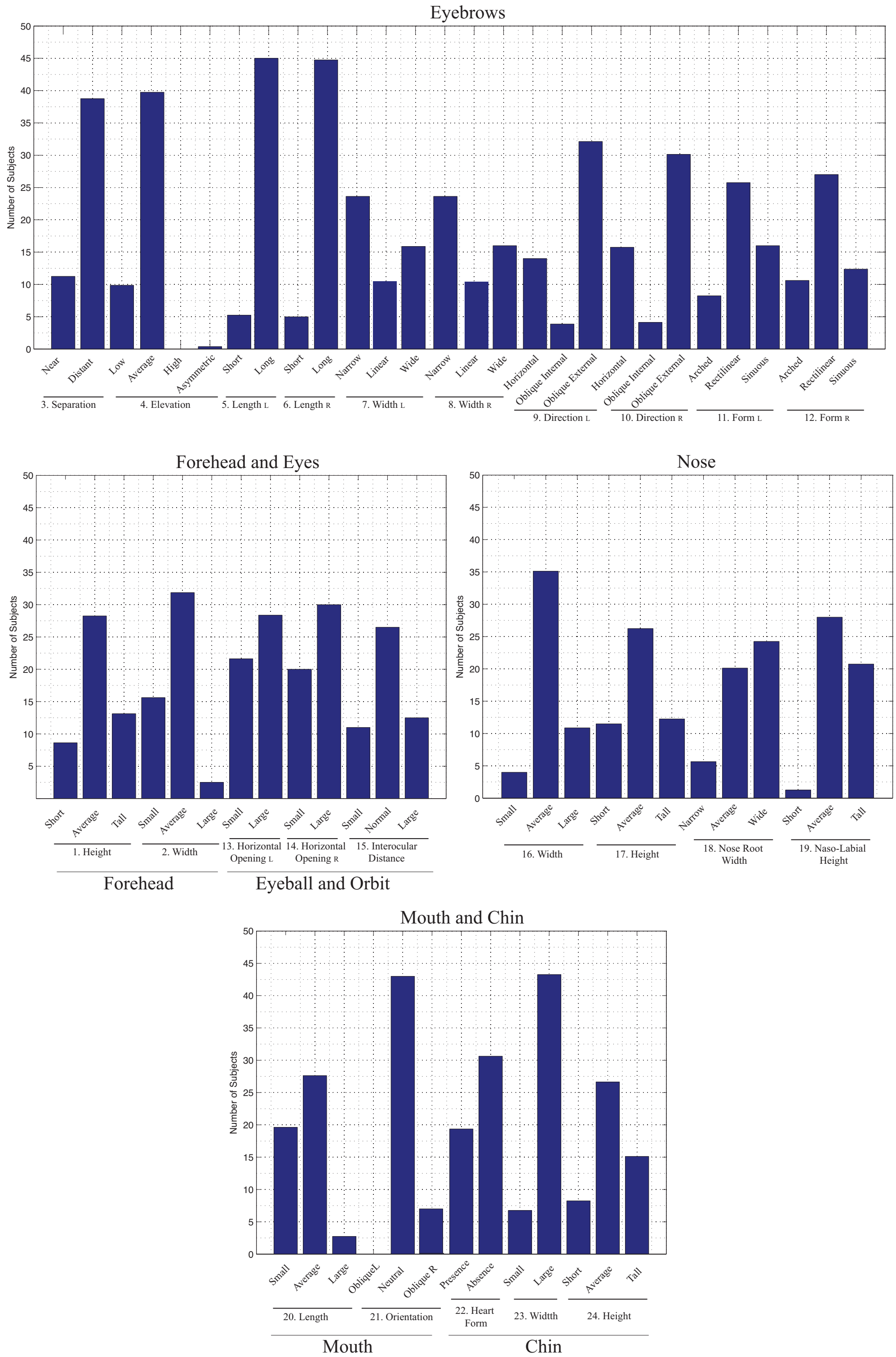


Figure 6.8: *Population statistics from ATVS Forensic DB based on the discrete facial soft biometrics features detailed in Table 3.3.*

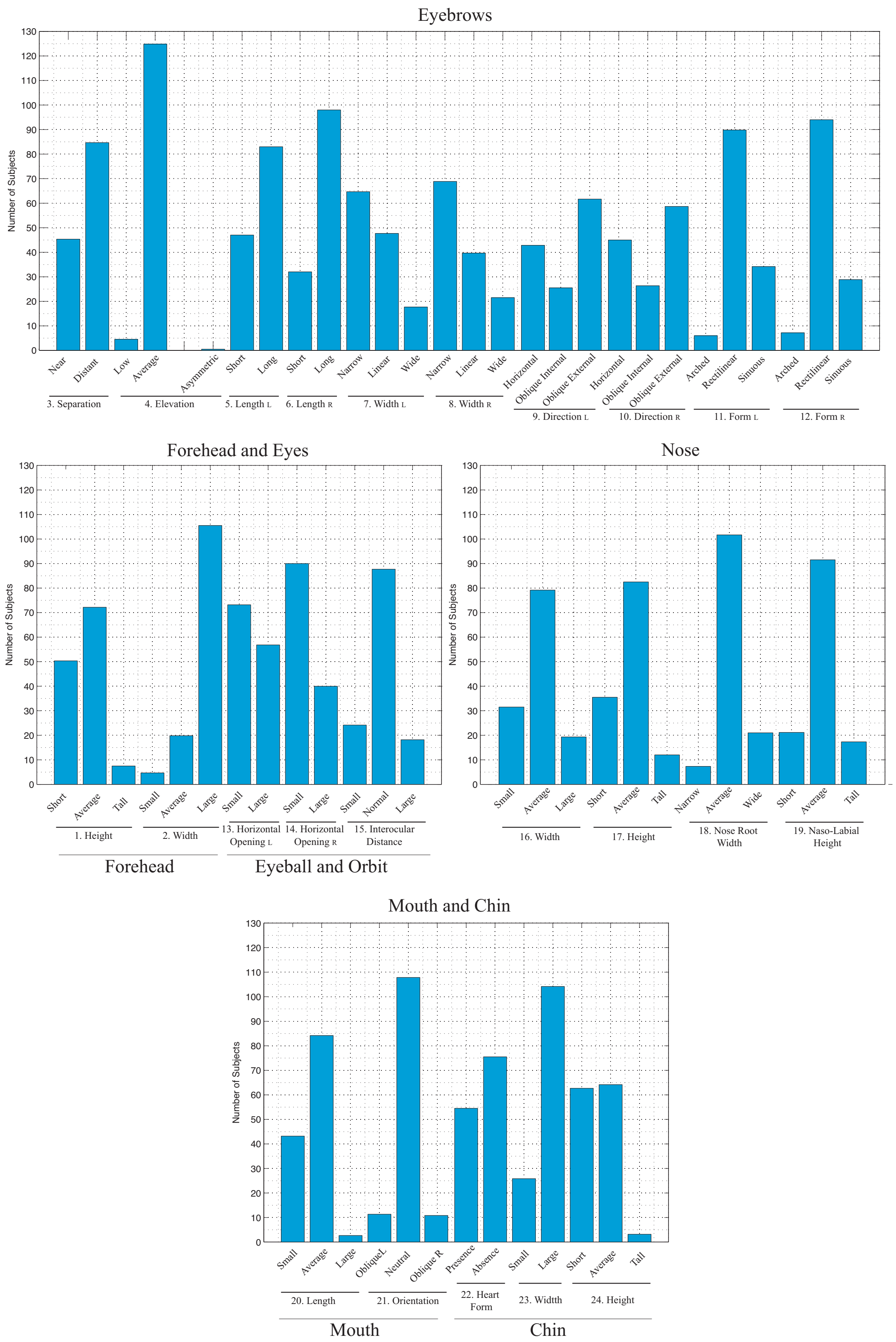


Figure 6.9: *Population statistics from MORPH DB based on the discrete facial soft biometrics features detailed in Table 3.3.*

Fig. 6.8 and 6.9 show the population statistics for the 24 discrete values proposed in both ATVS Forensic and MORPH databases. The first one, ATVS Forensic DB, represents a population of 50 European subjects (32 male and 18 female) as previously detailed in Sect. 4.2.5. The age range in the first database is between 19–45 years captured in an academic environment while this range in the MORPH database is between 16–60 years captured in a criminal environment.

As shown in the graphs of both databases the main differences are on forehead width (2), eyebrows width (7, 8), horizontal opening of eyes (13, 14), nose root width (18), naso-labial height (19), and chin height (24). The other facial traits have approximately the same distribution. The resolution differences between both databases (see Figs. 4.10 and 4.11) can explain some of these statistical differences.

It is important to note that these population differences can be useful to improve the identification tasks in forensics.

6.2.1.2. Correlation Between Labels

This section reports an analysis of the correlation between the soft features defined. For this purpose the correlation between all pairs of labels of the two groups defined (*continuous* and *discrete*) is computed using the Pearson's correlation coefficient defined in previous Eq. (6.1):

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \tag{6.7}$$

where σ_{XY} represents the covariance of the two variables X and Y divided by the product of their standard deviations σ_X and σ_Y. The variables X and Y represent numerical values associated to the pair of *continuous* or *discrete* terms at hand. For *discrete* features each semantic term was converted to numerical values in the range 1 to 3 (e.g. short, average, and long) using thresholds trained by a subset of each database. X_i and Y_i are the label values across all individuals and images, therefore $N = 400$ images (50 subjects $\times$ 8 images per subject) in ATVS Forensic DB and $N = 780$ in MORPH DB. The value r provides the correlation coefficient which ranges from -1.0 to 1.0. A value of 1.0 implies that a linear equation perfectly describes the relationship between X and Y, with all data points lying on a line for which Y increases as X increases. A value of -1.0 implies that all data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

Results of correlation between *continuous* features are presented in Fig. 6.10, where both databases are compared. Some features such as forehead height (1) and average line length (32), eyebrows length (8 and 9), eyebrows angles (12 and 13), horizontal opening of eyes (14 and 15), ears length (28 and 29), and ears angle (30 and 31) are clearly positive correlated in both databases as was expected. This means for example that when the forehead height increases the average line of the face also increases. Note that there exits a notable difference for the ears trait due to the different face image resolution of both databases.

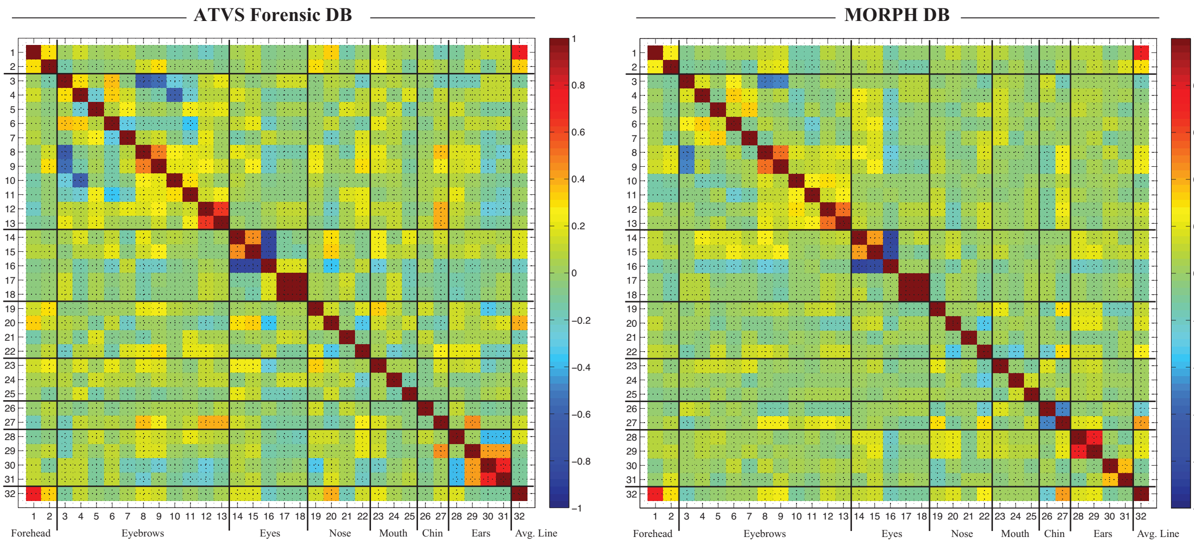


Figure 6.10: *Correlation between continuous labels based on Pearson's coefficient r (see Eq. (6.7)) for ATVS Forensic DB (left) and MORPH DB (right). Numbering of facial soft biometrics is detailed in Table 3.3.*

On the other hand a negative correlation exits in traits such as eyebrows separation (3) and length (8,9), eyebrows elevation (4) and width (10), and a remarkable negative correlation between horizontal opening of both eyes (14,15) and interocular distance (16). This means that for example when the eyebrows separation increases their length decreases. Note an important inverse correlation between chin height (26) and width (27) in MORPH DB that denotes a singularity in this population analysed.

In the same way, the correlation between *discrete* soft biometric features have been analysed and is shown in Fig. 6.11. Again a positive correlation between some facial features such as eyebrows length (5 and 6), eyebrows direction (8 and 9), eyebrows form (11, 12), horizontal opening of eyes (13 and 14) and the nose height (17) can be observed.

The most stressed negative correlations are between horizontal opening of the eyes (13,14) and interocular distance (15), mouth orientation (21) and mouth heart form (22). Note that in MORPH database the difference between chin width (23) and height (24) is also presented in discrete facial features. Another interesting negative correlation in ATVS DB is between interocular distance (15) and nose height (17), i.e., as the interocular distance increases the height or the nose is reduced.

6.2.1.3. Stability Analysis of Numerical Translation

This section reports an analysis of the stability of the *continuous* and *discrete* features for all facial soft biometrics features. This is done by calculating the stability coefficient, defined

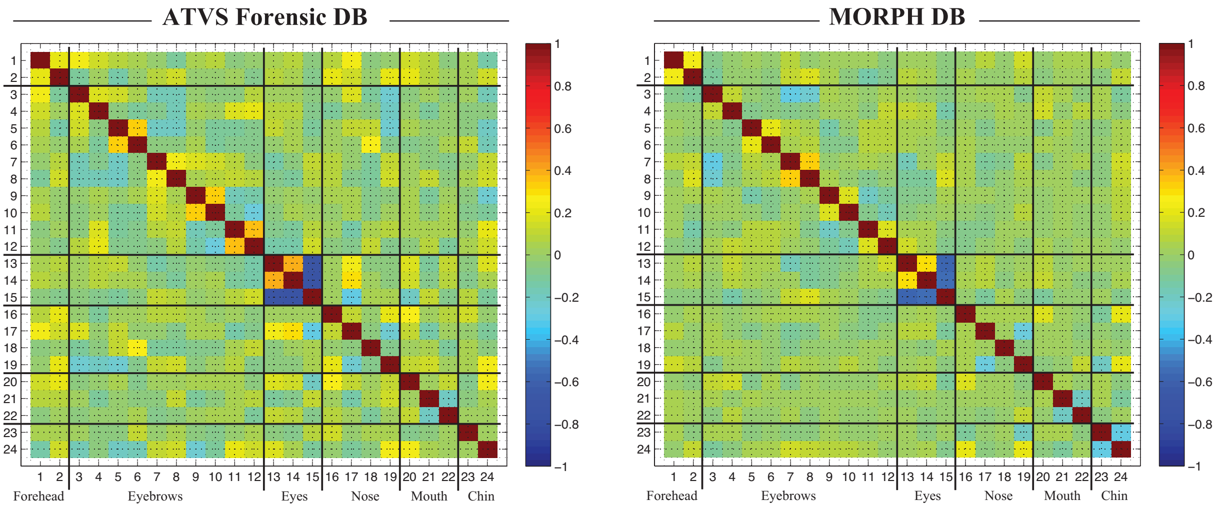


Figure 6.11: *Correlation between discrete labels based on Pearson's coefficient r (see Eq. (6.7)) for ATVS Forensic DB (left) and MORPH DB (right). Numbering of facial soft biometrics features is detailed in Table 3.3.*

similarly to Eq. (6.2) for label X as:

$$\text{Stability}_X = 1 - \frac{1}{SM} \sum_{i=1}^{S} \sum_{m=1}^{M} |X_{im} - \text{mode}_m(X_{im})| \tag{6.8}$$

where X_{im} is the extracted value for subject i from its sample image m, $M = 8$ or $M = 6$ is the total number of sample images for ATVS Forensic DB and MORPH DB, respectively, and, $S = 50$ or $S = 130$ is the total number of subjects, also respectively, and $\text{mode}_m(X_{im})$ is the statistical mode across the discrete values (i.e., the value most often value selected for subject i). In the case of continuous features the statistical mean instead of the mode is calculated $\text{mean}_m(X_{im})$ across m values (i.e., the mean value for a given subject i).

The resulting stability coefficients for all facial soft biometrics features are depicted in Fig. 6.12. In this figure (left) the *continuous* features are shown and we can see that for ATVS Forensic DB the nose (21), forehead width (2), and chin width (27) are the most stable labels, this is due to the normalization process. Faces are normalized based on the distance between eye centers therefore the real horizontal width of some traits is lost. On the other hand, MORPH DB presents stable labels such as ears labels (28-31) as a consequence of the low resolution of the images.

The stability results of *discrete* features are shown in Fig. 6.12 (right). Again small differences between both databases are observed due to the difference in quality. In general, forehead height (1), eyebrows traits (3-12), and nose width (16) and height (17) in ATVS Forensic DB have better stability than in MORPH DB.

As observed in Fig. 6.12, both databases (high and low resolution) present approximately the same stability in the facial soft biometric features extracted. This demonstrates the value of the proposed features.

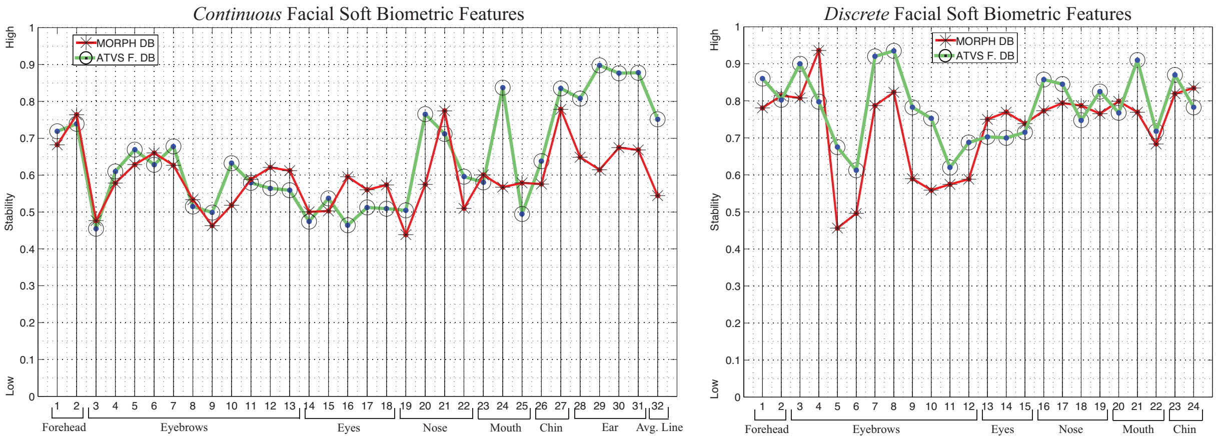


Figure 6.12: *Features' stability for the* 32 *continuous and* 24 *discrete facial soft biometrics features considered for both databases (see Table 3.3).*

In this PhD Thesis these two types of features (*continuous* and *discrete*) will be processed differently (e.g., using different similarity measures), and together in order to study the potential of each of them.

6.2.1.4. Discriminative Power Analysis

In order to evaluate the discriminative power of the facial soft biometric feature X, we compute for it the ratio between the inter-subject variability, and the intra-subject variability using the previous Eq. (6.3), where the annotators are changed by the number of images per subject as follows:

$$\text{Discrimination}_X = \frac{\frac{1}{S(S-1)} \sum_{i=1, i \neq j}^{S} \sum_{j=1}^{S} |\mu_i - \mu_j|}{\sigma} \tag{6.9}$$

$$\mu_i = \underset{m}{\text{mean}}(X_{im}),\ \ \mu_j = \underset{m}{\text{mean}}(X_{jm}),\ \ \sigma = \frac{1}{S}\sum_{i=1}^{S}\sigma_i,\ \ \sigma_i = \underset{a}{\text{std}}(X_{ia}) \tag{6.10}$$

where i and j index subjects, and m indexes images for a given subject.

The discrimination coefficient for the X^k features ($k = \{1, ..., K\}$, $K = 32$ or $K = 24$, for *continuous* or *discrete* values) is depicted in Fig. 6.13. There we can see that for *continuous* and *discrete* features the eyebrows and eyes traits are less discriminant than the nose, and forehead traits.

The least discriminant *continuous* facial soft features are the right eyebrow outer elevation (7) and the chin width (26) in ATVS Forensic DB, while the eyes angles between corners (17-18) and mouth angles (25) are in MORPH DB. In contrast the least discriminant in *discrete* features are mouth heart form (22) and chin width (23), and the eyebrows direction (9-10), respectively in ATVS Forensic DB and MORPH DB.

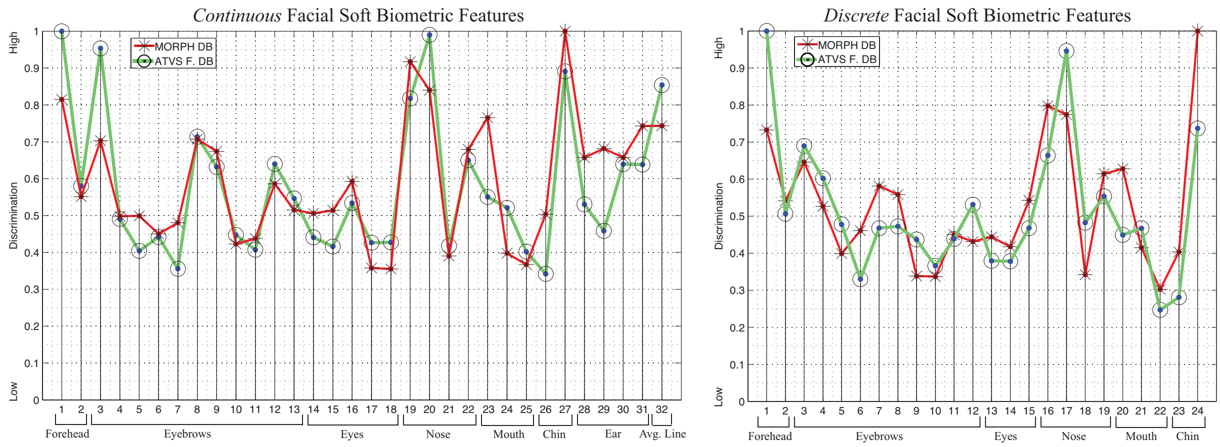


Figure 6.13: *Discrimination power of the* 32 *continuous and* 24 *discrete facial soft biometrics features considered for both databases (see Table 3.3).*

6.2.2. Verification Based on Facial Soft Biometrics

This section describes a person verification system based only on facial soft biometric features. First, each *continuous* or *discrete* feature in numeric form (see Sect. 3.1.2) is normalised to the range [0,1] using the tanh-estimators described in [Jain *et al.*, 2005]:

$$\hat{X}^k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{X^k - \mu_{X^k}}{\sigma_{X^k}}\right)\right) + 1\right\} \tag{6.11}$$

where X^k is the $k = \{1, ..., K\}$ soft label, $\hat{X}^k$ denotes the normalized label, and μ_{X^k} and σ_{X^k} are respectively the estimated mean and standard deviation of the label under consideration (see Table 3.3 for the list of the labels). Note that, depending on the features considered (*continuous* or *discrete*), there are $K = 32$ or 24 facial labels, respectively (see Table 3.3).

In this Section three different similarity measures based on various distances Theodoridis and Koutroumbas [2008] are compared: *i*) Mahalanobis, *ii*) Euclidean, and *iii*) Hamming.

Similarity scores based on the Mahalanobis distance between the test vector with K features $\mathbf{x} = [\hat{X}^1, ..., \hat{X}^k]^T$ and a statistical model $\mathcal{C}$ of the client are computed as follows:

$$s_M(\mathbf{x}, \mathcal{C}) = \frac{1}{\left((\mathbf{x} - \boldsymbol{\mu}^{\mathcal{C}})^T (\boldsymbol{\Sigma}^{\mathcal{C}})^{-1} (\mathbf{x} - \boldsymbol{\mu}^{\mathcal{C}})\right)^{1/2}} \tag{6.12}$$

where $\boldsymbol{\mu}^{\mathcal{C}}$ and $\boldsymbol{\Sigma}^{\mathcal{C}}$ are respectively the mean vector and covariance matrix obtained from the training labels ($M = 8$ and 6 training samples for ATVS Forensic DB and MORPH DB, respectively), which form the statistical model of the client $\mathcal{C} = \{\boldsymbol{\mu}^{\mathcal{C}}, \boldsymbol{\Sigma}^{\mathcal{C}}\}$.

Similarity scores based on the Euclidean distance are computed as follows:

$$s_E(\mathbf{x}, \mathcal{C}) = -\frac{1}{M}\sum_{i=1}^{M}\left((\mathbf{x} - \mathbf{y}_i)^T (\mathbf{x} - \mathbf{y}_i)\right)^{1/2} \tag{6.13}$$

where $\mathbf{y}_i$ are the M training vectors corresponding to subject $\mathcal{C}$.

The similarity measure based on Hamming distance is computed as:

$$s_H(\mathbf{x}, \mathcal{C}) = -\frac{1}{MK}\sum_{i=1}^{M} \#_k\{\hat{X}^k \neq \hat{Y}^k\} \tag{6.14}$$

where $\mathbf{x} = [\hat{X}^1, \ldots, \hat{X}^K]^T$, $\mathbf{y}_i = [\hat{Y}_i^{\,1}, \ldots, \hat{Y}_i^{\,K}]^T$ are the M training vectors corresponding to subject $\mathcal{C}$, and $\#_k\{condition\}$ indicates the number of cases across k where the *condition* holds.

In summary, in this section two different sets of labels (*continuous* and *discrete*) have been evaluated using three different similarity measures (Mahalanobis, Euclidean, and Hamming distance). It is important to note that the Hamming distance only makes sense with *discrete* features and it will not be applied to *continuous* features. Results will be presented in the next sections.

6.2.3. Database and Experimental Protocol

The facial soft features were obtained from two databases: ATVS Forensic DB and MORPH DB as was described previously. Two different sets of facial soft biometric features based on *continuous* and *discrete* values have been defined and used in our experiments in order to understand the behaviour of facial soft biometric features and their best application to forensics and face recognition. For this purpose, high and low resolution frontal face samples from these two databases (see an example in Fig. 4.10 and 4.11) have been used together with their corresponding physical facial soft biometric labels analysed in the previous sections. A general description of this process is detailed in Sect. 3.1.2 of the Chapter 3.

Two different sets of features are defined at close distance: *i*) *continuous* features, including distances of different traits and *ii*) *discrete* features, derived from the previous continuous values using a set of subjects to train the thresholds. Table 3.3 in Chapter 3 shows the soft labels available for each of the sets defined.

The experimental protocol followed is based on cross-validation (leave-one-out approach) due to the low quantity of subjects in the ATVS Forensic DB (50 subjects).

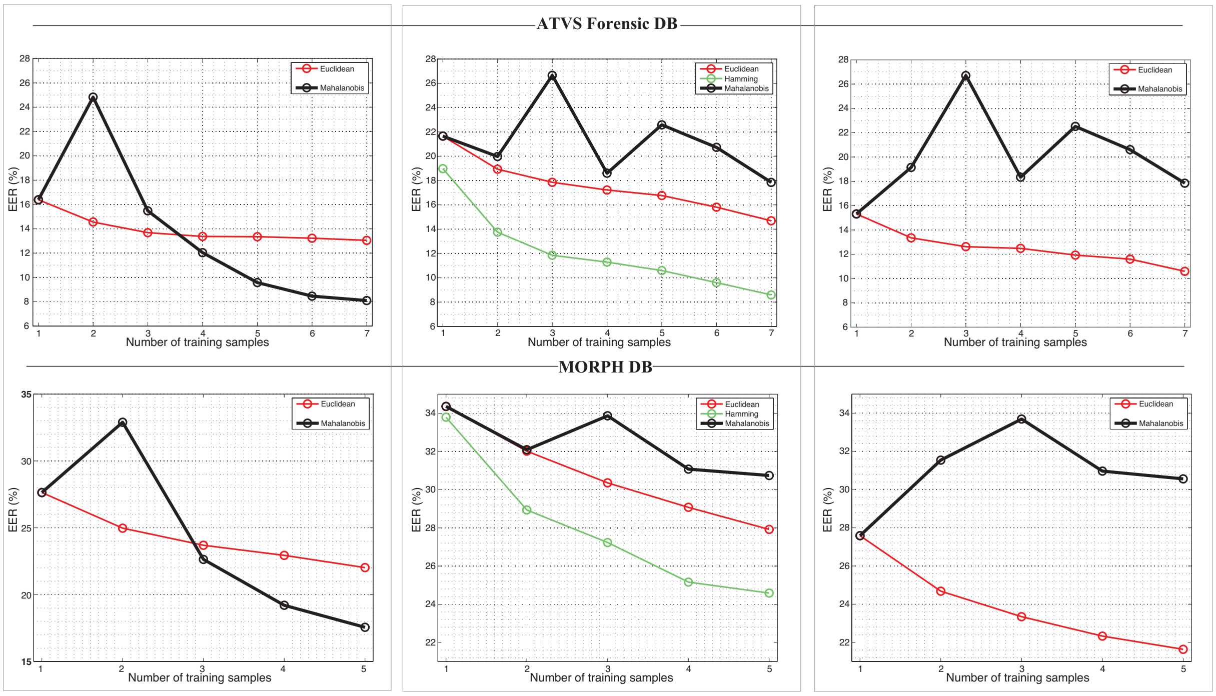
The leave-one-out approach that we have implemented first divides the data using a varying number of training samples and one of the remaining samples not used for training is left out for testing. We then iterate by rotating the selected training samples a number of times equal to the total quantity of samples ($M = 8$ in ATVS DB and $M = 6$ in MORPH DB in our experiments).

6.2.4. Results

6.2.4.1. Analysis of Training Set Size for Soft Labels

An important parameter to be considered in soft biometric systems is the size of the training set. For this purpose, we have evaluated the system with different number of training samples following the leave-one-out methodology explained in Sect. 6.2.3.

Fig. 6.14 shows the different configurations analysed for the three following sets of soft labels: *i*) *continuous*, *ii*) *discrete*, and *iii*) *mixed* features, which analyses both of them all together.



Continuous Facial Soft Biometric Features ***Discrete*** Facial Soft Biometric Features ***Mixed*** Facial Soft Biometric Features

Figure 6.14: *EER (%) obtained when varying the number of training samples for the three set of features considered: continuous, discrete, and mixed. On the top results from ATVS database are presented, and on the bottom results from MORPH database. Note the different range of EER in the axes for different plots.*

The figure shows results for the three different similarity distances defined (Euclidean, Hamming, and Mahalanobis).

As can be seen all soft label sets follow the same trend, the system recognition performance (EER) improves significantly when more samples are used in the training stage. For Euclidean distance in *continuous* set using more than 3 training samples the system performance saturates in both databases. However, in case of the Mahalanobis distance, the system improves when more samples are used in the training stage as was expected.

In contrast, the Hamming distance achieves the best results on *discrete* features in both databases. This Hamming distance achieves a relative improvement of 12-24% and 60-70% for MORPH and ATVS Forensic databases compared to the Euclidean and Mahalanobis distances. Therefore the Hamming distance is suitable for the *discrete* features.

In the *mixed*-features scenario both *continuous* and *discrete* features are analysed simultaneously. In this case the literature [Theodoridis and Koutroumbas, 2008] recommends the use of continuous similarity measures such as Euclidean distance. This is confirmed in Fig. 6.14 (right) in both databases the Euclidean distance achieved better performance results than the Mahalanobis distance in contrast to the *continuous*-features scenario. The Mahalanobis distance achieved the worst performance results in the discrete scenario, as was previously observed.

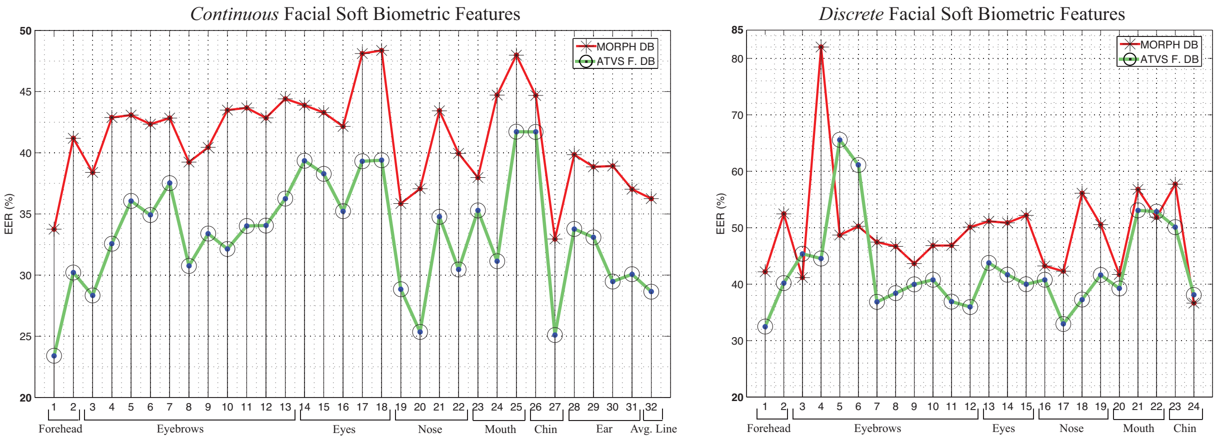


Figure 6.15: *Average EER (%) obtained for each individual facial soft biometric features (32 continuous and 24 discrete) defined in Table 3.3. Average EER calculated between the three difference distances considered: mahalanobis, hamming, and euclidean. The hamming distance is not considered to compute the results of the continuous features.*

6.2.4.2. Analysis of Individual Soft Labels

This section presents the discrimination power of each individual facial soft label following the leave-one-out experimental protocol described in Sect. 6.2.3. As shown in Fig. 6.15 (left), the continuous set of soft labels in both databases follow the same trend but the system performance of ATVS Forensic DB is slightly worse in all features. This is mostly due to the difference of resolution between both databases.

The forehead height (1) followed by nose height (20) and chin height (27) achieve the best results ($EER < 25\%$ and $EER < 35\%$ in both databases) and it is worth noting that these were the most discriminative features regarding the initial experiments shown in Fig. 6.13. Another relevant label with a high performance and discrimination power is nose width (19) with an EER = 28.84% and EER = 35.85%, respectively in ATVS Forensic DB and MORPH DB.

Discrete facial features achieve similar performance results than continuous features. Again individual facial features of ATVS Forensic DB achieved better performance than features of MORPH DB as was expected. The heights of forehead (1) and nose (17) continue having a relevant role by obtaining the best performance results. The worst peformance results are obtained by eyebrows elevation (4) in MORPH DB and eyebrow left length (5) in the case of ATVS Forensic DB, results previously predicted in Fig. 6.13.

The remaining facial soft labels in both feature sets achieve higher performance, with better results in general for forehead and nose labels compared to eyebrows or eye labels, as anticipated in Sect. 6.2.1. As can be seen, individual facial labels are not very discriminative on their own.

6.2.4.3. Analysis of Grouped Soft Labels

The aim of the following experiment is to study the discrimination power of the three groups of soft labels considered using the 3 different similarity measures defined in Sect. 6.2.3. Fig. 6.16

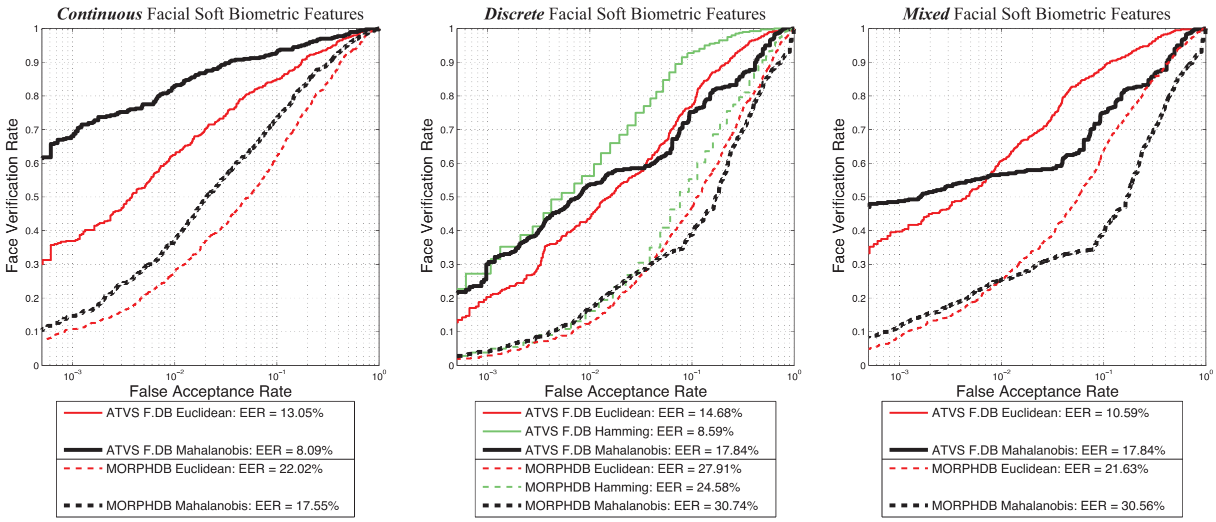


Figure 6.16: *ROC curves obtained for the facial soft biometric features sets: continuous, discrete, and mixed.*

shows the performance of each set of labels considered. Here, dashed lines represent the sets (*continuous*, *discrete*, and *mixed*) on MORPH DB, while solid lines represent their system performance on ATVS Forensic DB.

We can observe a significant difference between the *continuous*, *discrete* and *mixed* facial soft biometric features regarding the performance of the different similarity measures. The performance of *continuous* labels is clearly lower compared to *discrete* set as predicted in previous sections. An important experimental finding is that the Mahalanobis and Hamming distances computed in both databases achieved by far the best system performance results in the *continuous* and *discrete* set of features, respectively.

Regarding the other set of *mixed* labels, the difference of performance is not that significant as can be seen in Fig. 6.16 (right). *Mixed* set is comprised of all available facial labels including *continuous* and *discrete* labels, therefore it experiments a decrement of EER performance compared to the other individual sets in some regions of the plot (e.g., around FAR $= 0.1 = 10\%$). On the other hand, the other two features sets have a lower number of facial soft labels considered but result in better EER performance with the appropriated similarity measure. These results are confirmed by the literature [Theodoridis and Koutroumbas, 2008] where we can see how different types of features may be processed differently (e.g., using different similarity measures), in order to obtain better performance.

It is important to note that although soft labels provide low recognition performance when used as a stand alone system, they can help to improve hard biometric systems.

6.2.4.4. Analysis of Feature Selection: SFFS

In order to find the most discriminative set of facial soft biometrics-based features, and therefore increase the performance of the biometric system, feature selection is performed. In

Database	Distance	SFFS Feature Selection	# Selected Features	EER (%)
ATVS F. DB	Euclidean	(**27**,**20**,**3**,31,19,9,30,16,22,23,32,12,25,13)	14	7.38
	Mahalanobis	(**27**,**3**,31,**20**,30,19,8,23,9,32,12,28,13,25)	14	3.70
MORPH DB	Euclidean	(**19**,**20**,**27**,3,1,31,23,9,30,22,28,4,8,16,32,13)	16	16.50
	Mahalanobis	(**27**,1,**19**,31,**20**,23,29,30,3,22,9,15,12,32,28,6)	16	14.01

Table 6.1: *SFFS selected* ***continuous*** *features (defined in Table 3.3) for each system analysed. The three most discriminative features in Fig. 6.13 (left) are bold for each database.*

Database	Distance	SFFS Feature Selection	# Selected Features	EER (%)
ATVS F. DB	Euclidean	(**17**,**1**,**24**,18,19,12,3,20,22,16,7,13,4,14)	14	10.34
	Hamming	(**17**,**1**,15,8,18,9,3,11,16,**24**,20,12,4,6,14,21,7,22,19,2,23,13,10)	23	7.84
	Mahalanobis	(**1**,**17**,**24**,16,19)	5	13.34
MORPH DB	Euclidean	(**24**,**17**,20,1,**16**,3,7,5,13,15,8)	11	21.78
	Hamming	(**24**,**17**,3,20,**16**,9,1,10,13,5,18,4,8,7,22)	15	20.89
	Mahalanobis	(**24**,**16**,**17**,1)	4	24.95

Table 6.2: *SFFS selected* ***discrete*** *features (defined in Table 3.3) for each system analysed. The three most discriminative features in Fig. 6.13 (right) are bold for each database.*

addition, the reduction of the number of features decreases the computational cost too.

Among the different feature selection algorithms [Molina *et al.*, 2002], the one employed in this work is the Sequential Floating Forward Selection (SFFS) [Pudil *et al.*, 1994]. This suboptimal searching technique is an iterative process in which, in each iteration, a new set of features (whose choice is based on the results of previous subsets) is used to compute a certain criterion. This is done until the criterion does not improve. For more details see [Molina *et al.*, 2002; Pudil *et al.*, 1994; Theodoridis and Koutroumbas, 2008]. In our case the criterion is related to the performance of the system, in particular, is to minimize the value of the EER (Equal Error Rate).

Once the features are selected, the feature vector has, depending on the experiment, between 4 and 23 components. The SFFS algorithm is able to provide the most discriminative set of features with a dimension specified by the user or with the dimension that gives the best value for the criterion (in that case the dimension is not specified). The latter approach was performed in our system. Tables 6.1 and 6.2 summarize the results after applying the SFFS algorithm to each system for both databases.

As shown in Table 6.1 in boldface, the features most frequently selected across databases are: eyebrows separation (3), nose width (19), nose height (20), and chin height (27) , which correspond to the features most discriminative individually (see Fig. 6.13 (left)). In the same way, similar results can be seen for discrete features in bold in Table 6.2 and Fig. 6.13 (right): the features most discriminant individually are always in the SFFS selected feature sets: forehead height (1), nose width (16), nose height (17), and chin height (24).

Database		EER (%) Indiv. Results			EER (%) SFFS Results		
ATVS F. DB	Individual	8.09 (32 Feat. s_M) 8.59 (24 Feat. s_H)			3.70 (14 Feat. SFFS s_M) 7.84 (23 Feat. SFFS s_H)		
	Fusion	F. Sum	F. Prod	F. Weight	F. Sum	F. Prod	F. Weight
		5.59	5.59	**4.90**	3.95	4.00	**3.06**
MORPH DB	Individual	17.55 (32 Feat. s_M) 24.58 (24 Feat. s_H)			14.01 (16 Feat. SFFS s_M) 20.89 (15 Feat. SFFS s_H)		
	Fusion	F. Sum	F. Prod	F. Weight	F. Sum	F. Prod	F. Weight
		16.63	16.62	**15.86**	14.20	14.20	**12.27**

Table 6.3: *Fusion results of the best systems in Fig. 6.16 and SFFS results in Tables 6.1 and 6.2 for the continuous ($s_{Mahalanobis}$) and discrete ($s_{Hamming}$) features for ATVS Forensic and MORPH databases.*

Another interesting result is the number of selected features in each experiment. For *continuous* features the selection results for both similarity measures in (14 or 16 features) depending the database. On the other hand the number of selected *discrete* features is variable. For both ATVS Forensic and MORPH database the minimum number of regions is always selected for the Mahalanobis-based system, followed by the Euclidean-based system, and finally by the Hamming-based system, which achieved the best EER performance results.

6.2.4.5. Fusion of *Continuous* and *Discrete* Features

This section describes the fusion of both *continuous* and *discrete* facial soft biometric features in order to increase the system recognition performance. For these fusions we have selected the best system in each set of features, Mahalanobis-based and Hamming-based system, respectively for continuous and discrete sets. The comparison between all features performance and the most discriminative features selected by SFFS is also presented.

Three different fusion rules have been evaluated: (i) sum, (ii) product, and (iii) weighted sum fusion. The weighted sum fusion gives more weight to the most robust system, which is the continuous system based on the EER of the systems to be fused. For the experiments, weights of 70% and 30% have been used.

Table 6.3 shows the fusion results of these three different fusion rules. As we can see for all the fusions the best individual system is improved, thus this demonstrates how different similarity measures applied to different features can improve the system performance. It is also interesting to note that all the fusions improve the system performance of the feature fusion analysed in Fig. 6.16 (right), where *continuous* and *discrete* features are analysed together in a *mixed* set.

The best results are achieved using a weighted fusion in both databases using the most discriminant features obtained by SFFS in the previous section. Therefore the potential of these proposed facial soft biometric features is confirmed.

6.3. Chapter Summary and Conclusions

In this chapter we have performed two studies related to the proposed soft biometric information in scenarios at a distance and forensics. For this purpose soft biometric information suitable for video surveillance and facial soft biometric information adequate for forensics have been studied and evaluated. It is important to emphasize that the use of this ancillary information is very interesting in scenarios suffering from very high variability conditions. These soft labels can be visually identified at a distance by humans (or an automatic system) and fused with hard biometrics (as e.g., face recognition). It is important also to note that this kind of soft information is still a developing field in relation to its automatic extraction.

First, the stability and discriminative power of the largest and most comprehensive set of soft labels for video surveillance available from the literature, has been studied and analysed. The discriminative information of these labels grouped by physical categories (*body*, *global* and *head*) has also been studied.

Moreover, the available soft biometric information in scenarios of varying distance between camera and subject (*close*, *medium* and *far*) has been analysed. The rationale behind this study is that depending on the scenario, some labels may not be visually present and others may be occluded. Thus, the discriminative information of soft biometrics will vary depending on the distance. It is worth noting that this relation between scenarios at a distance and the performance of soft biometrics for person recognition has not been studied in this way before.

The experimental results have shown that a system that is completely based on soft biometrics traits for video surveillance results in moderate accuracy in the recognition of individuals, which will not be usually enough for demanding real-world applications. However, these soft biometric traits can be used to improve the performance of a traditional biometric system (e.g., gait, face, etc.) in many ways. These approaches will be studied in the next Chapter.

A wide set of facial soft biometrics for forensics has been also introduced and evaluated in this chapter. These features are extracted following forensic protocols based on the forensic morphological analysis. The facial soft biometric traits can either be continuous or discrete. Traits such as eyebrows height and width, interocular distance, naso-labial height, etc. are continuous variables in nature. On the other hand, these traits can be converted to discrete values using thresholds in order to simplify their classification and to compute population statistics.

The correlation, stability, and discriminative power of the proposed facial soft features have been broadly studied and evaluated. The experimental results have shown that a system that is completely based on facial soft biometric features for forensics can provide good accuracy in person recognition tasks. Additionally, these facial soft biometric traits can be used to improve the performance of face recognition systems as it will be studied in Chapter 7.

This chapter includes novel contributions in the experimental findings comparing different approaches and scenarios, the consistent and replicable methodology used, and the proposed facial soft biometric features for surveillance and forensics.

Chapter 7

Adaptive Fusion

THIS CHAPTER describes the application of the proposed adaptive score fusion schemes introduced in Chapter 3 to biometric authentication at a distance. The proposed adaptive fusion schemes are: *i*) scenario-based, where the acquisition distance between the subject and the camera is used to adapt the system, *ii*) soft biometric-based that introduces how to combine the soft biometric information with primary biometric systems, and *iii*) facial regions-based, approach that uses the different facial regions extracted from a human face (including color information).

For scenario-based fusion, the score-level combination of two standard approaches are evaluated under variation in the acquisition distance. As was confirmed in Chapter 5, where we studied both approaches individually (DCT-GMM and PCA-SVM), the DCT-GMM system is found to be more robust against degradations due to the acquisition distance compared to the PCA-SVM system. In the present Chapter we exploit this fact by introducing an adaptive score fusion scheme based on automatic scenario estimation which is shown to improve our system in uncontrolled environments.

In Chapter 6 we studied various soft biometrics extracted from a human body (e.g., height, gender, skin color, hair color, etc.) that can easily distinguished at a distance and observed that those features are not fully distinctive by themselves in recognition tasks. However, this soft information can be fused with biometric recognition systems to improve the overall recognition when confronting high variability conditions. One significant example is visual surveillance, where face images are usually captured in poor quality conditions with high variability and automatic face recognition systems do not work properly. This chapter also presents an experimental study of the benefits of soft biometric labels as ancillary information based on the description of human physical features to improve challenging person recognition scenarios at a distance.

On the other hand, the combination of different regions of the human face on various forensic scenarios is also presented in this chapter. In order to generate scientific knowledge useful for the forensic experts three scenarios of interest are considered comparing mugshot and CCTV face images. One of the findings achieved in Chapter 5 was that depending of the acquisition distance the discriminative power of the facial regions changes, having in some cases better performance than the full face. This effect can be exploited by facial regions for face recognition, which results in a very significant improvement of the discriminative performance compared to just using the face.

In the same sense, an analysis of the benefits of using color information on a region-based face recognition system is reported. Three different color spaces are analysed (RGB, YC_bC_r, $l\alpha\beta$) in a very challenging scenario matching good quality mugshot images against video surveillance images. This scenario is of special interest for forensics, where examiners carry out a comparison of two face images using the global information of the faces and where the variability is very high. As discussed in Chapter 5 this a very complicated task where automatic systems usually work using only grayscale images. Here we demonstrate the usefulness of considering also color information.

The chapter is structured as follows. One section is dedicated to each of the proposed adaptive score fusion schemes, with the scenario-based score fusion scheme being presented in Sect. 7.1, the soft biometrics-based fusion approach in Sect. 7.2, the regions-based fusion scheme in Sect. 7.3, and the color regions based-fusion approach in Sect. 7.4. These four sections share a common structure, with a brief introduction to the problem, the description of the databases and experimental protocol, the verification system used, and finally the fusion experiments, results and discussion. The chapter summary and conclusions are presented in Sect. 7.5.

This chapter is based on the publications: Tome *et al.* [2010a, 2013b,c,f].

7.1. Scenario-based Fusion

The first objective in this section is to investigate the effects of acquisition distance variation on the performance of automatic face recognition systems. This is motivated by the analysis of the results from the past Multiple Biometric Grand Challenge (MBGC 2009) [Phillips *et al.*, 2009a] and the Face Recognition Vendor Test (FRVT 2006) [Phillips *et al.*, 2009b], which show that a large amount of research is still needed to overcome this problem. As a result, the National Institute of Standards and Technology (NIST) proposed a new challenge called "the Good, the Bad and the Ugly" [Phillips *et al.*, 2011] which makes use of three partitions of the MBGC Still Face dataset of frontal images [Phillips *et al.*, 2009a]. This challenge was designed by NIST to develop new face algorithms capable to match correctly difficult face pairs. In this sense, we have studied the degradation effects in three different scenarios defined by the acquisition distance between subject and camera, namely *close*, *medium* and *far* distance as previously described in Sect. 5.1.

Li *et al.* [Li *et al.*, 2009] consider the problem of Biometrics at a Distance as having no restrictions over conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background, or photographic quality. Many solutions have been proposed in the literature to deal with these factors individually but a suitable solution to the global problem of unconstrained environments has not been developed yet.

The effect of training and testing with images acquired at different distances using two

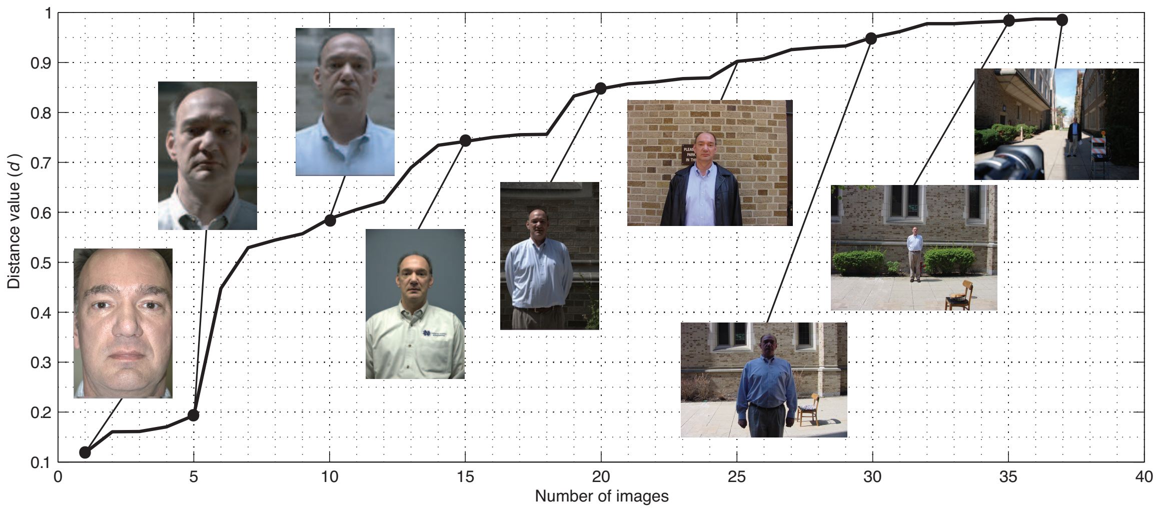


Figure 7.1: *Example of the estimated acquisition distance d for an example subject from MBGC database.*

classical face recognition approaches (DCT-GMM and PCA-SVM systems) is first studied. We also investigate experimentally the effects of acquisition distance variation on a multi-algorithm approach [Jain and Ross, 2004] based on these matchers. Then we propose a novel scenario estimator that enables system adaptation depending on the predicted acquisition conditions. Finally, we evaluated the proposed scenario-based fusion approach presented in Sect. 3.2.1 that exploits this scenario estimator.

7.1.1. Acquisition Distance Estimation

The concept of estimating the acquisition distance in order to define different scenarios has not been traditionally used in face recognition. Automatic scenario estimation gives us knowledge about the variability level that affects the system (i.e., different scenarios usually present different variability factors) and therefore is a valuable tool for system adaptation.

Face localization (eyes coordinates) is the first stage in face recognition systems and after this process, images are compensated in rotation and normalized to the same width (W_I). We define the estimated acquisition distance d as:

$$d = 1 - \frac{\text{IPD}}{W_I}, \tag{7.1}$$

where IPD and W_I are respectively the interpupillary pixel distance and the image width captured. Therefore d is a function of the distance between eyes (IPD), which will be strongly correlated to the acquisition distance d. The minimum possible value of d will tend to 0, when the segmented face occupies the whole image. As the person goes away from the camera, d increases, until it reaches a maximum value of 1. Fig. 7.1 shows the estimated acquisition distance (d) from an example subject under different acquisition conditions.

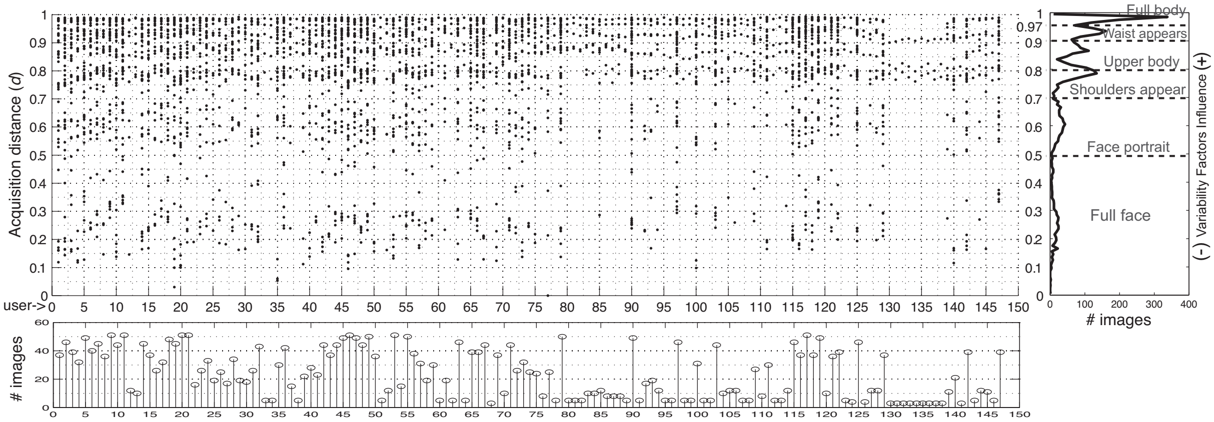


Figure 7.2: *Histogram of the estimated acquisition distance (d) from MBGC database described by Eq. (7.1).*

Fig. 7.2 shows the distribution of the proposed acquisition distance d in the database used in these experiments, where it is possible to appreciate the differences between the three scenarios defined (see Sect. 5.1.1) by the acquisition distance.

We can note in Fig. 7.1 an important correlation of the distance estimator d with respect to the actual acquisition distance. As we have observed in previous chapters, if the acquisition distance increases, variability factors increase and degrade the system performance, therefore we can use this estimator d as a predictor of the variability present in our system in each case.

7.1.2. Database and Experimental Protocol

The database used for the experimental work presented in this section is a subcorpus called "Face Stills dataset" of the NIST Multiple Evaluation Grand Challenge (MBGC) v2.0 [Phillips *et al.*, 2009a] described in previous Sect. 4.2.2. The database is comprised of 3842 face images from 147 subjects acquired at different distances. We further classify all the face images into three acquisition distance groups as follows. We consider three different scenarios: 1) *close* distance, in which the shoulders may be present; 2) *medium* distance, including the upper body; and 3) *far* distance, including the full body.

Only subjects with at least 4 images were kept in each scenario considered. A portion of the dataset was discarded (360 images from 89 subjects), because the face was occluded or the illumination completely degraded the face. A reduced number of subjects ($S = 13$) were completely discarded (less than 4 image per scenario) discarding a total 403 images of the whole dataset. The data selection process is detailed in Sect. 5.1.1.2 and summarized in Table 5.2, where we can see that the two considered subcorpora result in 134 subjects, using 484 images of 56 subjects for the development of the systems and 2595 images of 78 subjects for the evaluation.

The experimental protocol followed in the next experiments considers a number of 56 subjects as development for tuning the systems and the remaining 78 subjects as evaluation (see Table 5.2).

The dataset was then divided according to the three acquisition distance scenarios defined in Sect. 5.1.1. The resulting subsets are shown in Table 5.3. The development set is used to train a PCA subspace and GMM world model per scenario (close, medium, far and mix). Here it is important to note that we have tuned the systems with an equal number of images (130 images, given by the smaller scenario, i.e. the *far* one).

On the other hand, the evaluation set was equally divided into a train and a test set, the first one for training the SVM and GMM models per user and the other to test the system performance. Table 5.3 shows the different divisions of data in the three scenarios defined. It is possible to appreciate that the number of images is not perfectly distributed between these two sets (train and test) due to an imbalance in the number of samples per user. Four main experiments are defined for verification performance assessment across scenarios:

- *close2x*. This is designed to obtain the performance of the systems in situations where only high quality controlled images are used to train the system. This will be considered as the Baseline system. In this case, only the 661 images of the close training set are used to train the GMM and SVM classifiers.
- *medium2x*, This protocol uses 386 images as a training set from the medium distance dataset.
- *far2x* protocol. This protocol uses 304 images as a training set from the far distance dataset.
- *mix2x*. This is designed to study the effects of combining several kinds of information (training with different acquisition distances). The training set consists of the three acquisition distance datasets (1351 images).

7.1.3. Face Verification Systems

The architecture of the face recognition system used is shown in previous Fig. 3.3 of the Chapter 3. The preprocessing stage is divided into: *i*) automatic localization of the face, *ii*) segmentation, *iii*) size normalization to a constant size (64×80 in our experiments), and *iv*) pose and illumination compensation. The preprocessing stage was executed using VeriLook SDK v2.0 (commercial system) and the few produced errors were manually corrected as described in previous Sect. 5.1.3.

Two approaches are used for face verification. These two matchers receive a normalized face from the preprocessing stage:

- **PCA-SVM system**. This verification system uses Principal Component Analysis (PCA). The evaluated system uses normalized and cropped face images of size 64×80 pixels (width $\times$ height) to train a PCA vector space where 96% of the variance is retained. This leads to a system where the original image space of 5120 dimensions is reduced to 249 dimensions. Similarity scores are computed in this PCA vector space using a SVM classifier with linear kernel.

- **DCT-GMM system**. This verification system divides the 64×80 face image into 8×8 blocks with horizontal and vertical overlap of 4 pixels. This process results in 285 blocks per segmented face. From each block a feature vector is obtained by applying the Discrete Cosine Transform (DCT) from which only the first 15 coefficients ($N = 15$) are retained. The blocks are used to derive a world GMM Ω_w and a client GMM Ω_c Galbally *et al.* [2010]. From previous experiments we obtained that using $M = 1024$ mixture components per GMM gave the best results. The DCT feature vector from each block is matched to both Ω_w and Ω_c to produce a log-likelihood score.

To carry out the fusion stage, scores of the two systems are first normalized to the $[0, 1]$ range using the tanh-estimators described in Sect. 3.2. The constant $C = 0.4$ in Eq. (3.1) is used for the experiments carried out in this section.

7.1.4. Fusion Results

The fusion method used is based on the combination of the two systems at the score-level following the sum rule approach [Kittler *et al.*, 1998]. Our basic assumption for the adaptive scenario-based fusion approach implemented here is that the verification performance of one of the algorithms drops significantly compared to the other one for increasing acquisition distance. This fact is exploited with the adaptive distance-based fusion strategy described in Sect. 3.2.1 by considering from Eq. (3.2) that the function $g^j(d_i)$ takes the camera to person distance estimation presented in Eq. 7.1 of Sect. 7.1.1, and outputs a confidence $c_i^j = d_i/2$ for the system that degrades the most with the acquisition distance (i.e., $j = PCA - SVM$), and $c_i^j = 1 - d_i/2$ for the other one (i.e., $j = DCT - GMM$).

In the following experiments we are going to analyse the particular case of considering $M = 2$ systems to fuse (i.e., $\hat{s}_i^{PCA-SVM}$ and $\hat{s}_i^{DCT-GMM}$) for a particular biometric input i, obtaining confidence measures c^j corresponding to the two systems following the fusion model presented in Sect. 3.2.1. Fig. 3.3 represents the fusion scheme proposed where the proposed person to camera distance estimator (see Sect. 7.1.1) is used in order to adjust the scenario in each situation.

The combination of these systems through the sum fusion rule, and the proposed scenario-based weighted sum for different face acquisition distance groups is presented in Fig. 7.3. As can be seen, the fixed fusion strategy based on the sum rule only leads to improved performance over the best individual system in *medium2x* and *mix2x* scenarios, shown in Fig. 7.3 b) and d). The proposed adaptive fusion approach results in improved performance for all the acquisition distance groups, outperforming the standard sum rule approach, especially in *medium2close* testing conditions in Fig. 7.3 b), where the performance of the individual matchers are very different.

As shown in Fig. 7.3, the best results against increased acquisition distance are obtained when the system is trained with medium distance images and the mix of acquisition distance groups (*medium2x* and *mix2x* protocols). The baseline scenario (close distance training images) shows less robustness, with great degradation as the acquisition distance increases.

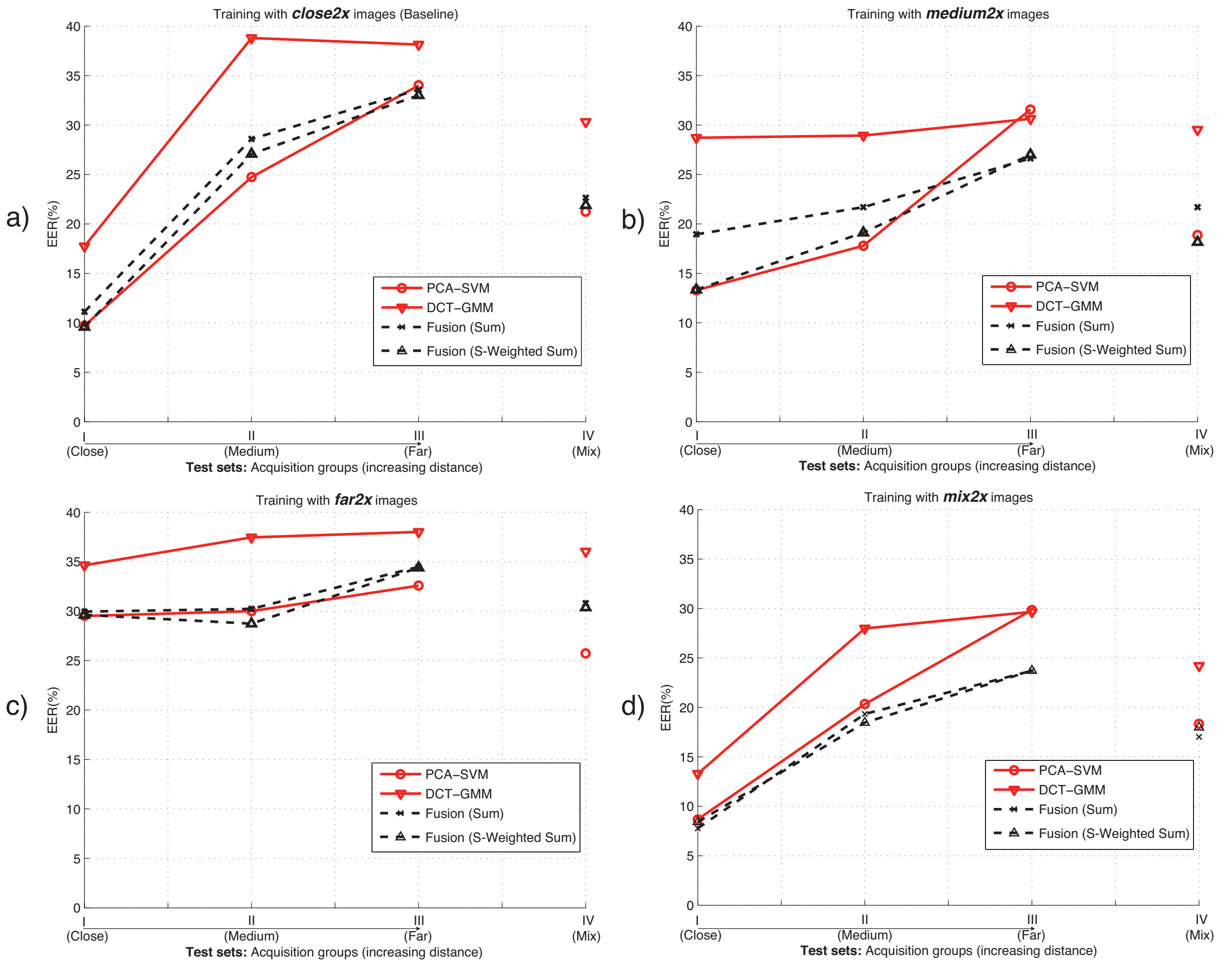


Figure 7.3: *Verification performance of the individual matchers (DCT-GMM- and PCA-SVM- based), their combination through the sum fusion rule, and the proposed distance/scenario-based weighted sum for increasing the system performance at a distance. The results are displayed in the different acquisition scenarios under study.*

Training with medium distance images is a good way to control the performance degradation due to varying distance. The DCT-GMM system generates low performance but stable results and the PCA-SVM system provides better performance but deteriorates quickly with the distance. Here the proposed fusion provides better results for far distance where both systems have a similar performance, and the new adaptive fusion is capable to equal the best system in closed distance testing.

The best performance is obtained with the *mix2* protocol where we are using the whole information of different acquisition distances in the training stage. As can be observed the fusion has an important role, increasing the system performance of the best individual system in all the cases. Also worth noting, in far distance conditions the fusion schemas improve the performance remarkably.

7.2. Soft Biometrics For Video Surveillance

Soft biometric information extracted from a human body (e.g., height, gender, skin color, hair color, etc.) is ancillary information easily distinguished at a distance but it is not usually distinctive by itself in recognition tasks. However, this soft information can be explicitly fused with biometric recognition systems to improve the overall recognition when confronting high variability conditions. One significant example is visual surveillance, where face images are usually captured in poor quality conditions with high variability and automatic face recognition systems do not work properly. In this scenario, soft biometric information can provide very valuable information for person recognition. This section presents an experimental study of the benefits of soft biometric labels as ancillary information based on the description of human physical features to improve challenging person recognition scenarios at a distance.

Experimental results based on the Southampton Multibiometric Tunnel Database (TunnelDB) described in Sect. 4.2.3, show that the use of soft biometric traits is able to improve the performance of face recognition on real and ideal scenarios by adaptive fusion rules.

The main contribution of this section is a new adaptive method for incorporating soft biometrics information to this kind of challenging scenarios considering face recognition. In order to do so, the largest and most comprehensive set of soft biometrics available in the literature was previously described and analysed in Chapter 6.

These soft labels have been grouped considering the available soft biometric information in scenarios of varying distance between camera and subject as (*close*, *medium* and *far*). The rationale behind this study is that depending on the particular scenario, some labels may not be visually present and others may be occluded. The process is broadly described in Sect. 6.1.

The experimental framework used in this section, shown in Fig. 3.4, describes how from a video at a distance of a person walking, soft labels and faces from a subject are extracted. The experiments study two configurations: *i*) when the primary biometric is missing (face in this case), and only the soft biometric information considered, and *ii*) when both hard (face recognition) and soft (labels) information are available and can be fused.

7.2.1. Database and Experimental Protocol

The same dataset and protocol selected for the soft labels from the TunnelDB (detailed in Sect. 6.1) was used for the face recognition system. Each user has 10 sessions, so 580 images per scenario from high-resolution frontal face sample videos have been used. For each of the 10 sessions of a subject, the first frame (*close* distance), the middle frame (*medium*) and the last frame (*far* distance) from the frontal videos have been selected to generate the image samples used in the experiments, having in total 1740 images (58 subjects $\times$ 10 sessions $\times$ 3 distances).

The database was divided into training and testing sets. For each subject 9 face images and 9 sets of soft labels were used for the training and the remaining session was used for testing following a leave-one-out approach Theodoridis and Koutroumbas [2008] generating this way 580 similarity target scores and 33.640 similarity non-target scores.

7.2.2. Face Verification Systems

For the face recognition experiments, two different systems have been used and compared (one commercial and one proprietary): *i*) Luxand FaceSDK 4.0, and two face recognition systems based on SRC [Wright *et al.*, 2009], *ii*) VJ-SRC, using automatic face detection based on Viola Jones [Viola and Jones, 2004], and *iii*) ID-SRC using ideal face detection marked manually.

FaceSDK by Luxand[1] is a high-performance and multi-platform face recognition solution based on facial fiducial feature recognition.

A proprietary **VJ-SRC** face recognition system based on Viola Jones to detect faces and using a matcher based on SRC [Huang and Aviyente, 2006; Wright *et al.*, 2009] is also used. Face segmentation and location of the eyes are two of the main problems in face recognition systems at a distance. For our experiments, we have also manually tagged the eyes' coordinates which allows us to consider an ideal case of face detection in the **ID-SRC** face recognition system. This way, we can compare the behaviour of soft labels when fused with face images on real (VJ-SRC) and ideal (ID-SRC) scenarios at a distance free of segmentation errors.

The SRC matcher is a state-of-the-art system based on recent works in sparse representation for classification purposes. Essentially, this kind of systems span a face subspace using all known training face images, and for an unknown face image they try to reconstruct the image sparsely. The motivation of this model is that given sufficient training samples of each person, any new test sample for this same person will approximately lie in the linear span of the training samples associated with the person.

7.2.2.1. Analysis of Face Detection Errors

This section presents an analysis of the three scenarios considered: *close*, *medium*, and *far*. Two face detection systems have been evaluated: *i*) proprietary based on Viola Jones, and *ii*) a commercial system (FaceSDK) based on facial landmarks.

Two different detection errors have been defined and analysed:

- Fail To Acquire (FTA): when there is a face in the image, but it is not detected.
- Fail To Detect (FTD): when the face detector finds an object in the image, but it is not a face.

The first error FTA will be a feedback report for the systems but the second error FTD has to be analysed manually by an operator or automatically by an error detector system. In this PhD Thesis the FTD error was evaluated manually observing the faces detected by both systems.

Table 7.1 shows the detection errors for the two systems evaluated. Firstly, Viola Jones approach achieves less FTA errors than FaceSDK system, but introduces a high number of FTD errors, which will affect the system recognition performance. The FTA errors in *close* scenario are due to short people whose middle part of the face is outside of the vision plane of the camera.

[1] http://www.luxand.com/facesdk/

	Scenario	FTA	FTD	Total_E=FTA+FTD
Viola Jones	*Close*	11 (1.89%)	11 (1.89%)	22 (3.79%)
	Medium	2 (0.34%)	116 (20.00%)	118 (20.34%)
	Far	96 (16.55%)	304 (52.40%)	400 (68.96%)
FaceSDK	*Close*	5 (0.86%)	20 (3.44%)	25 (4.31%)
	Medium	365 (62.93%)	66 (11.37%)	431 (74.31%)
	Far	571 (98.44%)	9 (1.55%)	580 (100%)

Table 7.1: *Face detection errors in the three scenarios at a distance for Viola Jones and FaceSDK systems. FTA and FTD error percentages are calculated for the total number of face images (N=580).*

As can be seen, the scenarios at a distance analysed are very challenging. Analysing the results both systems work poorly at *medium* and *far* distances due to the high variability and the low quality of face images. The Viola Jones approach achieves a reasonable FTA error in these distances but a large number of detections are not faces (FTD error is very high). On the other hand, the FaceSDK system has a higher FTA with lower FTD. The total error is so large for FaceSDK (74.31% and 100% for *medium* and *far*, respectively)) that was discarded for the following experiments.

7.2.2.2. Analysis of Face Recognition Systems

The results achieved for VJ-SRC and ID-SRC systems with automatic and manual (FTA = 0% and FTD = 0%) face detection are presented in Fig. 7.4. As can be seen in the manual face detection (ID-SRC system, solid lines), the database analysed is very challenging and the system performance decreases quickly when the acquisition distance increases. On the other hand, poor results are achieved for the case of using the automatic Viola Jones face detector (VJ-SRC) due to the high number of FTD errors but also because in this case there is no pose compensation and normalization regarding the position of the eyes as in the ideal case.

Therefore, a large improvement in the EER is achieved for all distances by considering manual face detection compared to Viola Jones in the SRC system. On the other hand, the system performance with automatic face detection is very poor in a FAR = 0.001 = 0.1% with VR lower than 5%. It is important to note that for *far* scenario with ideal face detection (ID-SRC system) the VR is lower than 30%, which shows the complexity of the database analysed.

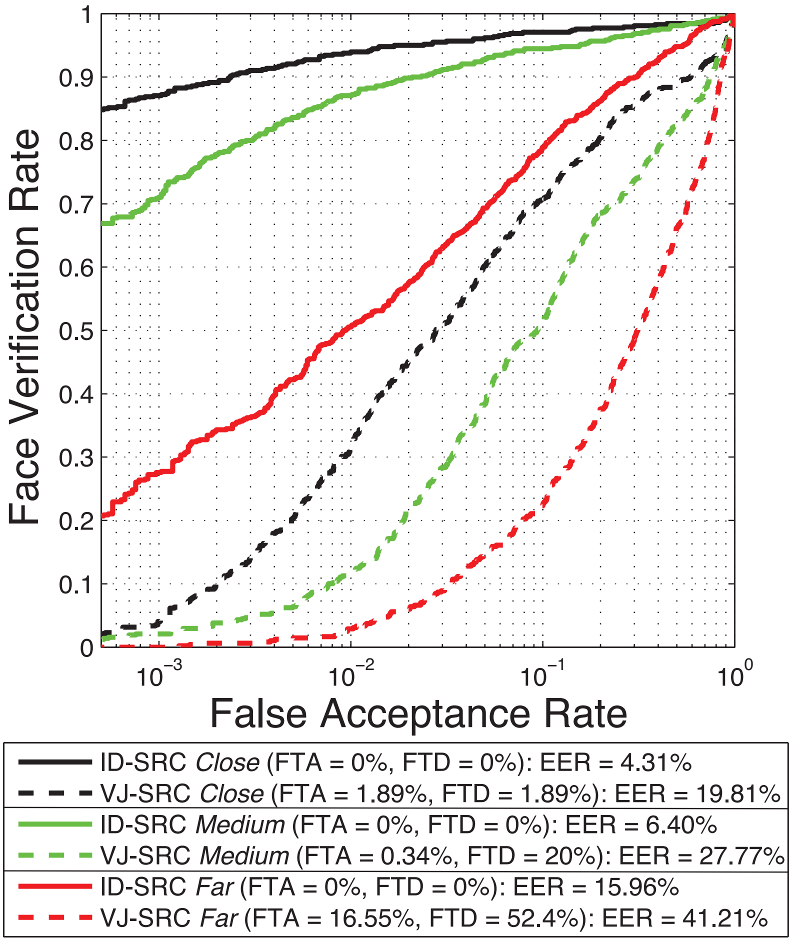


Figure 7.4: *ROC curves of SRC systems obtained using two configurations: automatic (VJ-SRC, dashed lines) and manual (ID-SRC, solid lines, FTA = 0%, FTD = 0%).*

7.2.3. Fusion Results

Soft biometrics offers several benefits over other forms of identification at a distance as they can be acquired from low resolution and low frame rate videos, and from an arbitrary viewpoint of the subject. This allows for the use of soft biometrics when primary biometric identifiers cannot be obtained or when only a description of the person is available.

This section analyses how soft labels can improve the face recognition system performance through the fusion of both biometric systems. The fusion method used is based on the combination of the systems at the score-level described in Sect. 3.2.2. In particular, we compare here: *i*) the sum rule, *ii*) an adaptive switch fusion rule, and *iii*) a weighted fusion rule. We obtain a switch fusion from the general fusion model presented in Sect. 3.2.2 by considering from Eq. (3.3) that $g^{face}(FTA^{face}) = 0$ when $FTA^{face} = 1$ and $g^{face}(FTA^{face}) > 0$ when $FTA^{face} = 0$, i.e., considering only the soft biometrics score if the face is not detected, and both face and soft biometrics scores if the face is detected.

In the following experiments $FTA^{soft} = 0$ always (we only use the soft biometrics available in each scenario) and $g^{face}(FTA^{face}) = g^{soft}(FTA^{soft}) = c = 0.5$ when the face is detected. As discussed in the next sections, other confidences $0 < c^j < 1$ for $j = \{face, soft\}$ and $\sum_j c^j = 1$ are also tried, which result in weighted sum approaches.

To carry out the fusion stage of the two biometric modalities, scores of the different systems were first normalized to the $[0, 1]$ range using the tanh-estimators described in previous Sect. 3.2.

Experiments are carried out by fusing the soft labels with VJ-SRC and ID-SRC face recognition systems over the three acquisition distances: *close*, *medium* and *far*. First, we consider the case of the fusion of soft labels with automatic face detection, and then the case of their fusion with an ideal face recognition using manual face detection.

7.2.3.1. Fusion with Automatic Face Detection Errors

This experiment studies the fusion of soft labels with the VJ-SRC system with automatic face detection carried out using a switch fusion. In case the face recognition system fails to acquire (FTA) a face due to variability factors, soft labels can help to improve the system performance.

Fig. 7.5 shows 4 ROC profiles in each graph: the VJ-SRC face recognition system, the soft labels system and two (sum and switch sum) fusions. This graph (bottom) also shows the VR and EER based on weight distribution for the weighted and switch weighted fusion. The first fusion applies a sum rule of the scores from the two systems only if both of them are available, otherwise it emits a FTA. As a result using this sum fusion FTA is non-zero. On the other hand, the switch fusion always results in an output score as described above, reducing the FTA error to 0 in this case. Detection errors showed in Table 7.1 correspond to the cases in which the switch fusion selects only the soft labels for the three scenarios defined.

The sum fusion of the two systems achieves relative improvements of 50.05%, 53.33%, and 59.88% of EER for *close*, *medium*, and *far* scenarios, respectively compared to the VJ-SRC face recognition system. As it is shown, soft labels improve the system performance and allow to keep the system robust in a *far* scenario. The same conclusion is confirmed for the switch fusion of the systems, which achieves relative improvements of 45.05%, 54.95%, and 60.03% of EER for *close*, *medium*, and *far* scenarios, respectively, compared to the VJ-SRC face recognition system.

As can be seen, the EERs for sum or weighted and switch fusion are similar, with the advantage of switch fusion of eliminating all FTA errors.

In these scenarios various weighted sum fusion functions have been also evaluated. As shown in Fig. 7.5 (bottom) the results are very similar compared to the sum and switch sum fusion and therefore those results are not present in the top graphs.

In conclusion, as the results show, a real face recognition system, which do not have a good performance due to the variability factors derived from acquisition at a distance, could be improved using soft biometric labels visually available in the scene, both increasing its verification performance and reducing its FTA errors.

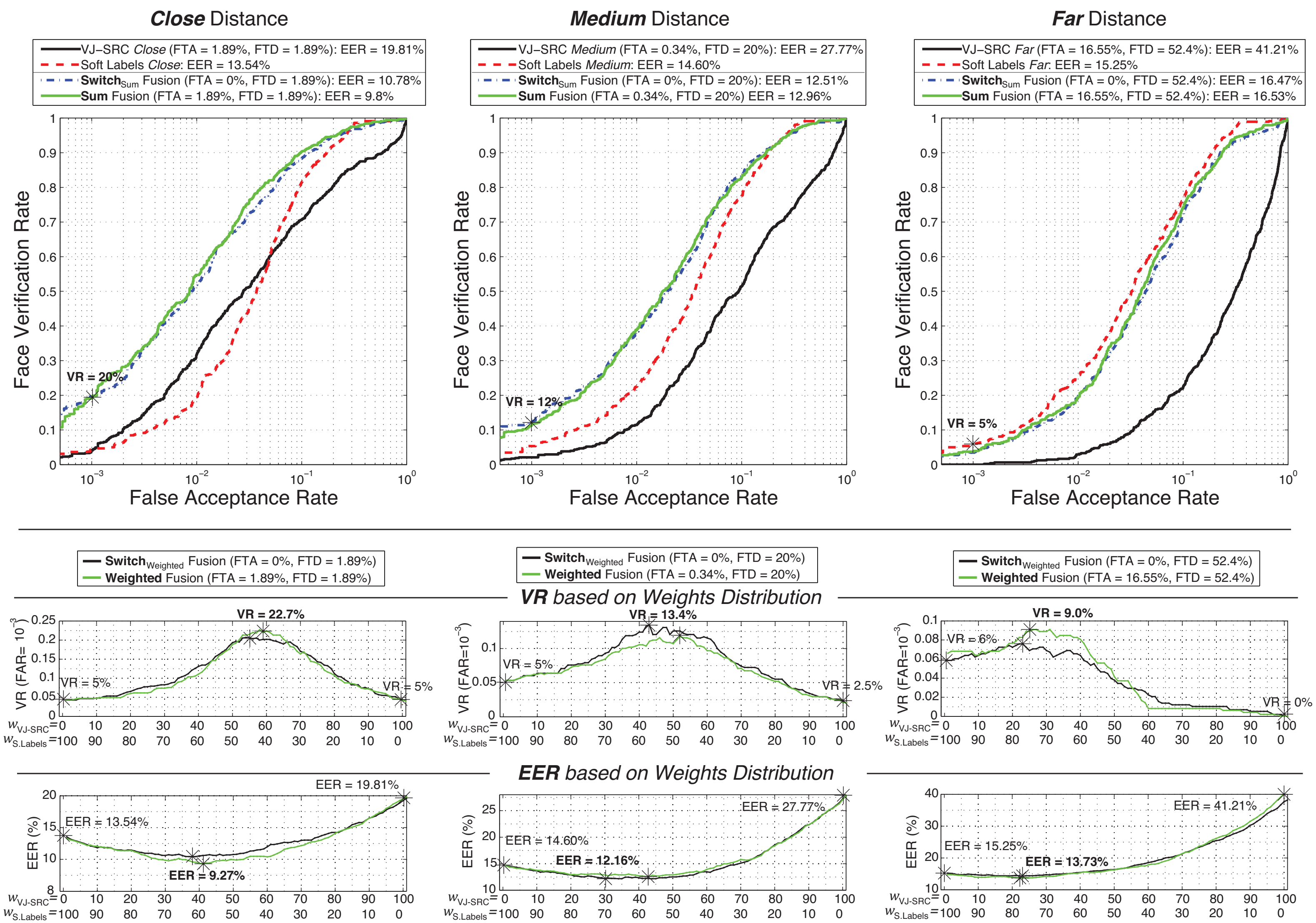


Figure 7.5: *ROC curves for the VJ-SRC system (automatic face detection errors) together with the corresponding improvement by sum and switch fusion for the three scenarios defined: close (left), medium (center), and far (right). On the bottom, the VR and EER based on weight distribution.*

7.2.3.2. Fusion with Manual Face Detection

This experiment focuses on use of the soft labels in order to improve the ID-SRC system with ideal face detection (FTA = 0% and FTD = 0%). Fig. 7.6 shows the ROC curves of both systems and two fusions (sum and weighted fusion rules) for different FAR points.

In this case the incorporation of soft labels improves the face recognition system performance. The sum fusion achieves significant relative improvements of 30.16%, 33.90%, and 49.87% in the EER for *close*, *medium*, and *far* scenarios respectively. On the other hand analysing the Verification Rate (VR) in a high security point such as FAR = 0.001 (0.1%), the system performance deteriorates. A relative decrement of about 10% in the VR for *close* and *medium* scenarios is obtained but in *far* scenario the VR increases moderately. These results are confirming that the soft biometric labels have a poor performance in a high security working point.

A weighted fusion has been proposed in order to solve the problem of the VR deterioration. The fusion gives more weight to the most robust system which is the face recognition system in FAR = 0.1%. Different weights have been tuned for the 3 distances based on the EER performance of the systems. Fig. 7.6 (bottom) shows the VR and EER based on weight distribution. In particular, we have used $c^{face} = 0.8$ and $c^{soft} = 0.2$ for *close* and *medium* distance, and finally $c^{face} = 0.7$ and $c^{soft} = 0.3$ for *far* distance. Using this configuration we achieved a significant increment in VR of 92.4%, 80%, and 45%, for *close*, *medium*, and *far* scenarios, respectively.

Therefore, the usage of soft labels can still help to improve the systems in these challenging high security conditions. The face detection stage is a key factor in order to achieve good results in scenarios at a distance. Consequently a single weighted fusion rule combining soft biometrics allows to improve the system performance where the primary biometrics are not working due to variability factors in the scenarios at a distance.

7.3. Facial Regions-based Fusion

Automatic face recognition systems are generally designed to match images of full faces. However, in practice, the full face is not always available, e.g., due to occlusions and other variability factors. This is one of the reasons why forensic examiners carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc.

This section is focused on the regions normally considered by forensic experts. In this PhD Thesis facial regions have been extracted following forensic protocols from law enforcement laboratories, allowing us to study individually the different facial regions from a human face as was presented in Sect. 5.3. In particular, we address in this section the problem of combining the most discriminative areas of the face for recognition on different acquisition scenarios following the scheme proposed in Sect. 3.2.3.

Understanding how different facial regions are combined on different forensic scenarios has some remarkable benefits, for example: *i*) allowing investigators to work only with particular regions of the face, or *ii*) preventing that incomplete, noisy, and missing regions degrade the

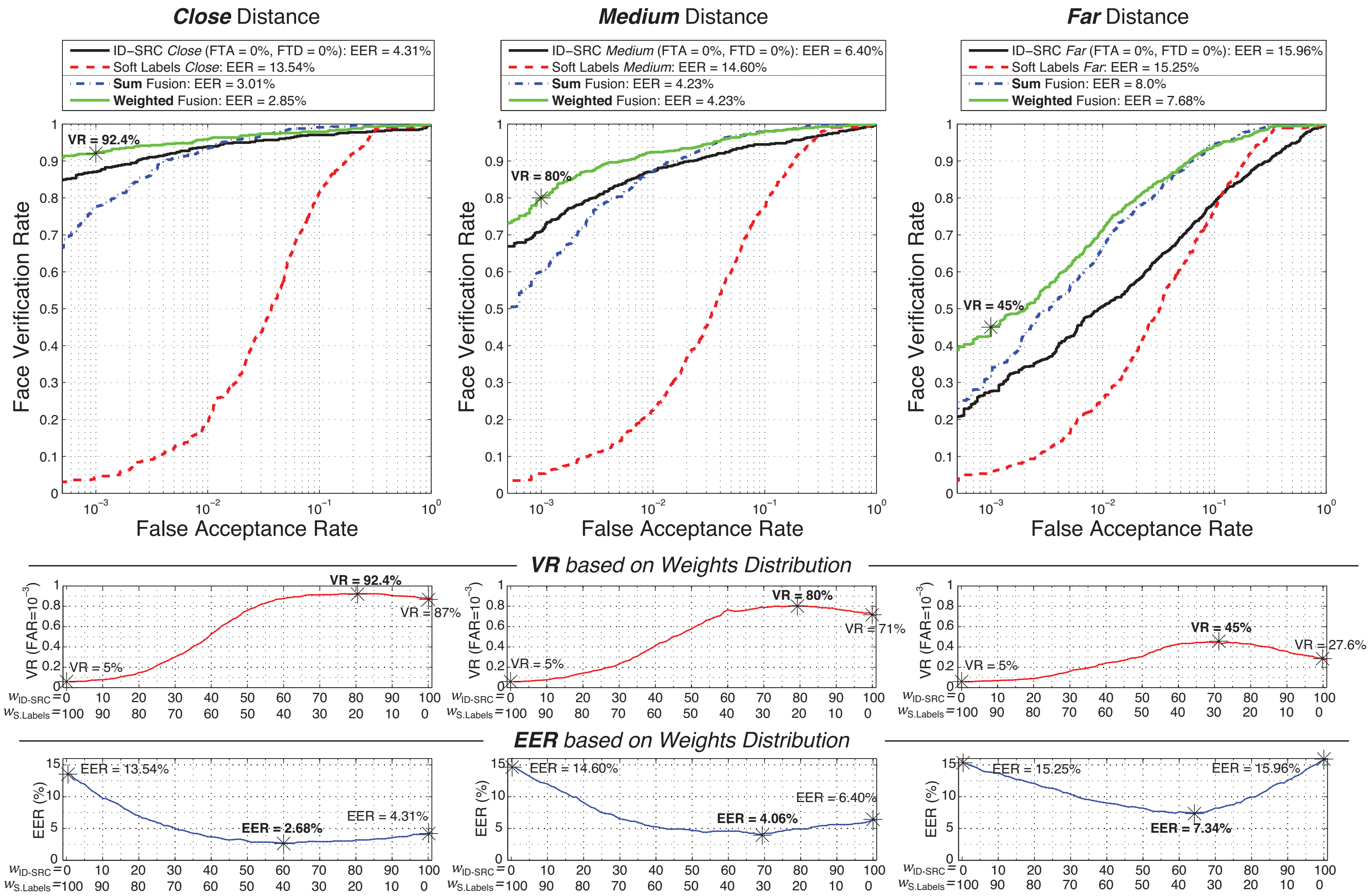


Figure 7.6: *ROC curves for the ID-SRC system (manual face detection) and its corresponding improvement by sum and weighted fusion rule for the three scenarios defined. On the bottom, the VR and EER based on weight distribution.*

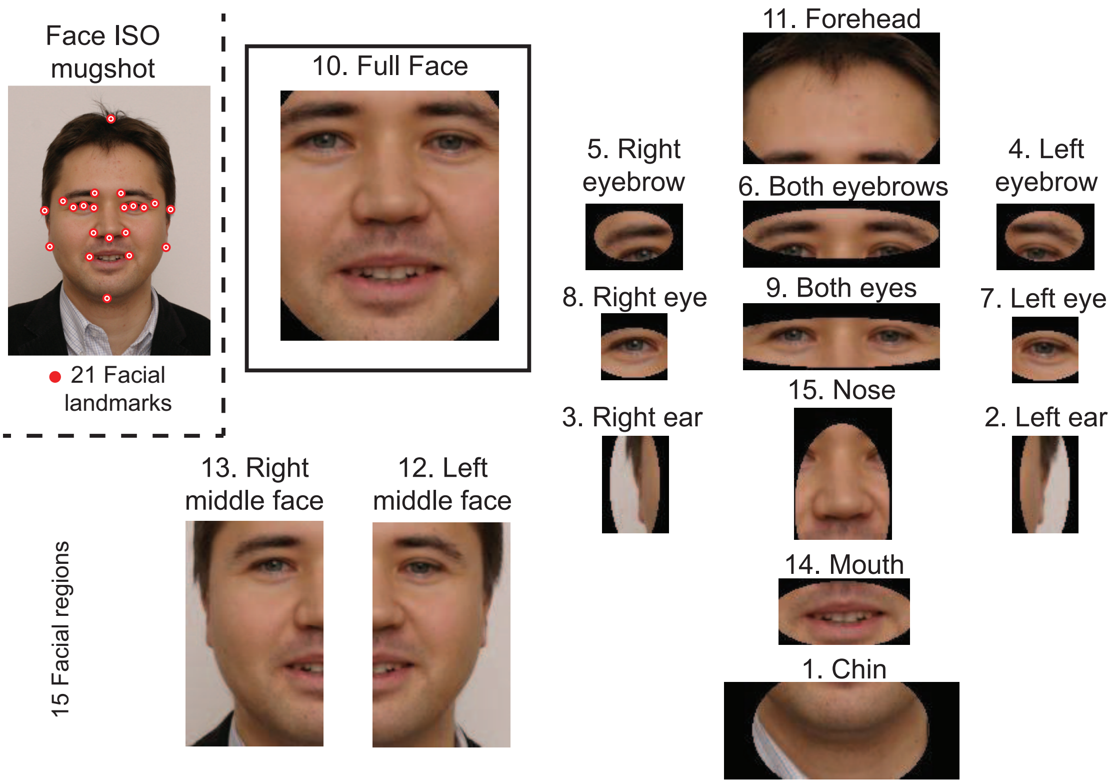


Figure 7.7: *The 15 facial regions obtained with the extractor based on facial landmarks (red dots).*

recognition accuracy. Further, a better understanding of the combination of facial regions should facilitate the study of facial regions-based face recognition. Therefore the fusion of the different facial regions is performed achieving significant improvements of performance compared to a traditional face recognition system based only on the face as a whole.

7.3.1. Facial Regions Extraction

The proposed facial regions extraction framework is described in detail in Sect. 5.3. In this framework, two kinds of region extraction are defined: *i*) based on human facial proportions, and *ii*) based on facial landmarks. The first one extracts the considered facial areas of interest of the face (eyebrows, eyes, nose, mouth, etc.) using as input information just the two eye coordinates, and simple facial proportions rules. The mentioned extractor would be of interest in challenging uncontrolled scenarios where landmarks are very difficult to be extracted automatically. On the other hand, the second extractor, based on facial landmarks correctly located (either manually or automatically), allows to extract the facial regions with high precision. The final region extraction result is the set of 15 facial regions based on forensic laboratories protocols (Spanish Guardia Civil DGGC) and Netherlands Forensic Institute (NFI)) as shown in Fig. 7.7.

Both extractors have two different configurations in order to find the initial facial landmarks: *i*) automatic (in our experiments we used Luxand FaceSDK 4.0), and *ii*) manual, carried out by a human examiner.

7.3.2. Databases and Experimental Protocol

The experimental work described in this section has been carried out using a collection of mugshot and CCTV face images of 130 subjects from two different databases: SCface [Grgic *et al.*, 2011] and MOPRH [Ricanek and Tesafaye, 2006] described and studied in Chapters 4 and 5.

Once each facial region has been extracted, eigen-regions (Principal Component Analysis, PCA) from each facial region are computed. Then, similarity scores are computed in this PCA vector space (dimension 200, retaining 98% of the energy of the original eigen-region space) using a Support Vector Machine (SVM) classifier with a linear kernel. The experimental protocol followed is described with more detail in Sect. 5.3.2. Both databases used in our experiments (SCface and MORPH described in Sect. 4.2.4 and 4.3.1), were divided into 3 subsets based on the subject ID: development (1-43), SVM training (44-87), and test (88-130). These three subsets were used for training the PCA features, as impostors in the training of SVMs, and for testing the final system performance, respectively.

In this work three different scenarios are studied considering the different cases that a forensic examiner can find in practice: *i*) *mugshot vs mugshot*, *ii*) *mugshot vs CCTV*, and *iii*) *CCTV vs CCTV*. In addition, three distances between subject and camera typical in practical applications are analysed: *close*, *medium* and *far* distance. This process is detailed in Sect. 5.3.2.

7.3.3. Fusion Results

The fusion of the $R = 15$ forensic facial regions in comparison with the performance of the whole *face* region is performed. The fusion is carried out at the score-level for various combinations of regions obtained via sequential search (detailed in Sect. 3.2.3). In particular, the R facial regions are fused using a parallel fusion (sequential search) approach based on the sum rule [Fierrez, 2006], starting from the most discriminative, then fusing this trait with the rest and keeping the best fusion of two regions, and continuing this process until all the regions are fused (i.e., using the fusion model in Sect. 3.2.3 and Eq. (3.4) where $g^r(B_i) = 1/R$ for all the R selected regions to fuse). This methodology can also be used by other fusion schemes such as: product, max-min, and weight. The fusion results are reported only for the case of manual landmark tagging with an extractor based on facial landmarks. Results with similar trends are obtained for the other configurations described.

Before carrying out the fusion, scores of the different facial regions are first normalized to the $[0, 1]$ range using the tanh-estimators described in Sect. 3.2 in Eq. (3.1), with $C = 0.01$, and μ_{SD} and σ_{SD} are respectively the estimated mean and standard deviation of the genuine score distribution using the development and SVM training sets.

Fig. 7.8 shows the results of the fusion approach for the three scenarios analysed, which will be studied in the next sections.

7.3.3.1. Mugshot vs Mugshot

This experiment presents the fusion results in the *mugshot versus mugshot* scenario using the MORPH database [Ricanek and Tesafaye, 2006]. Results of this fusion process can be seen in Table 7.2.

The best fusion is reached using the full *face* and the following 6 facial regions: inner facial traits (*both eyebrows*, *nose*, and *left/right eye*) and the two *middle faces*. A relative improvement of 51.5% in the EER is obtained with the fusion (from 13.5% to 6.61% EER) compared to only using the *face* region.

Similarly in previous Sect. 5.3, using the inner facial traits provides good performance in the *mugshot versus mugshot* scenario. Hence, the fusion of the whole face with the inner facial regions produces the best recognition performance.

7.3.3.2. Mugshot vs CCTV

This scenario is analysed using the SCface database [Grgic *et al.*, 2011]. The fusion results obtained for the three distances are shown in Table 7.2. Similar to the previous case the system performance improves fusing several facial regions compared to just using the full *face* region.

Close and *medium* distance scenarios combine 7 facial regions to achieve the best result, but the *far* scenario needs to combine a total of 10 facial regions to obtain it.

It is interesting to note that in the *close* scenario the best result is obtained with the fusion of inner and outer facial traits together with the full *face* (relative improvement of 56.7% in the EER with respect to using only the full face).

Similarly, in the two other distances considered, the best fusion includes inner and outer parts of the face, and relative improvements of over 40% in the EER are obtained with the fusion of regions compared to using only the full face.

As can be seen in the fusion EER, this scenario results in significantly worse performance compared to the previous and following scenarios. This is mainly due to the differences between gallery (mugshot) and probe images (CCTV).

7.3.3.3. CCTV vs CCTV

Table 7.2 shows the fusion results obtained for the three distances analysed for the *CCTV vs CCTV* scenario (using the SCface database [Grgic *et al.*, 2011]). As can be seen, when the acquisition distance increases more facial regions need to be fused with the full *face* region in order to achieve the best performance. Thus, an increment of variability and complexity involves more information to be fused, as could be expected.

A combination of inner (*mouth*, *nose*, and *right eyebrow*) and outer (*forehead*, *chin*, and *right ear*) facial regions are the best combination in this case. *Close* and *medium* scenarios just need 7 facial regions to achieve the best performance. On the other hand, the *far* scenario again needs a bigger number of facial regions to reach the best fusion result as also happened in the previous section in *far* distance.

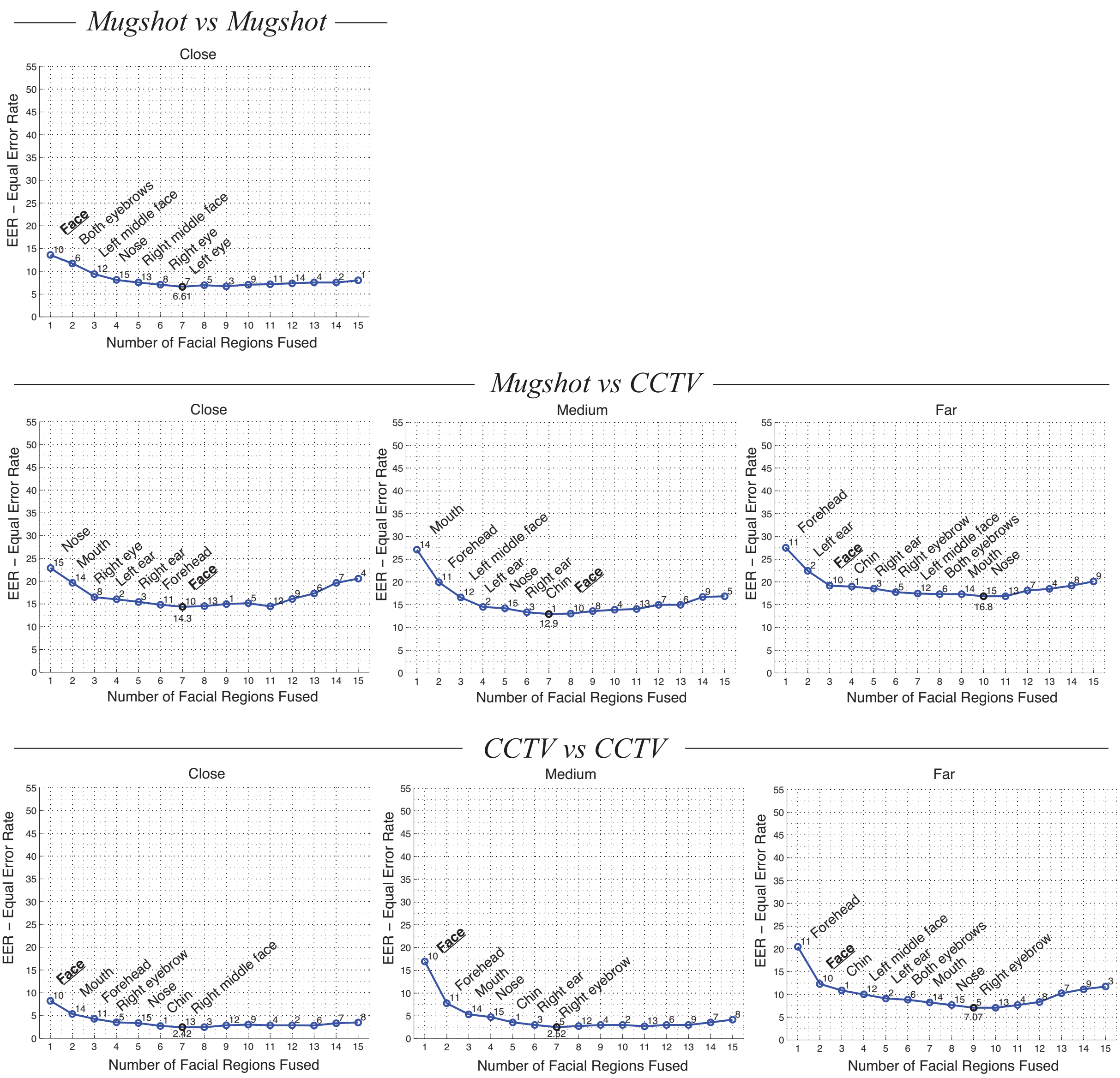


Figure 7.8: *EER for sequential sum fusion of the best combination of different facial regions for the three scenarios: mugshot versus mugshot, mugshot versus CCTV, and CCTV versus CCTV. For the last two scenarios the three distance are represented: close, medium and far.*

In this case, relative improvements of 70.6%, 83.4%, and 65.3% in the EER for the *close*, *medium*, and *far* scenarios are achieved respectively for the proposed fusion of regions compared to only using the full face for recognition.

Therefore, the combination of different facial regions can help to improve the system performance in challenging forensics scenarios and at a distance. The variability considered in train and test sets is a key factor in order to achieve good results in these conditions. Consequently a single parallel fusion rule combining some of these facial regions allows improving the system performance, where the traditional face recognition systems are not working with a desirable performance.

Scenarios		Full Face EER	Best Individual EER (Region Id)	Facial Regions Fused via Sequential Search (Best Combination)	(Best Combination) # Regions Fused	Fusion EER	Relative Improvement Over Full Face
Mugshot vs Mugshot		13.50%	13.50 % (10)	(**10**,6,12,15,13,8,7),5,3,9,11,14,4,2,1	7	6.61%	51.5%
Mugshot vs CCTV	*1. Close*	33.10%	22.89% (15)	(15,14,8,2,3,11,**10**),13,1,5,12,9,6,7,4	7	14.30%	56.7%
	2. Medium	31.20%	27.08% (14)	(14,11,12,2,15,3,1),**10**,8,4,13,7,6,9,5	7	12.90%	58.6%
	3. Far	28.90%	27.49% (11)	(11,2,**10**,1,3,5,12,6,14,15),13,7,4,8,9	10	16.80%	41.8%
CCTV vs CCTV	*1. Close*	8.24%	8.24% (10)	(**10**,14,11,5,15,1,13),3,12,9,4,2,6,7,8	7	2.42%	70.6%
	2. Medium	15.20%	15.20% (10)	(**10**,11,14,15,1,3,5),12,4,2,13,6,9,7,8	7	2.52%	83.4%
	3. Far	20.40%	17.25% (11)	(11,**10**,1,12,2,6,14,15,5),13,4,8,7,9,3	9	7.07%	65.3%

Table 7.2: *Overview of EER results obtained for the full face, the best individual facial region, and the proposed fusion. This is given for the three scenarios considered: Mugshot versus Mugshot, Mugshot versus CCTV, and CCTV versus CCTV scenarios. Fig. 7.7 shows the facial regions with their corresponding id number (e.g. the id numbers: 10, 6, 12, correspond to full face, both eyebrows, and left middle face, respectively).*

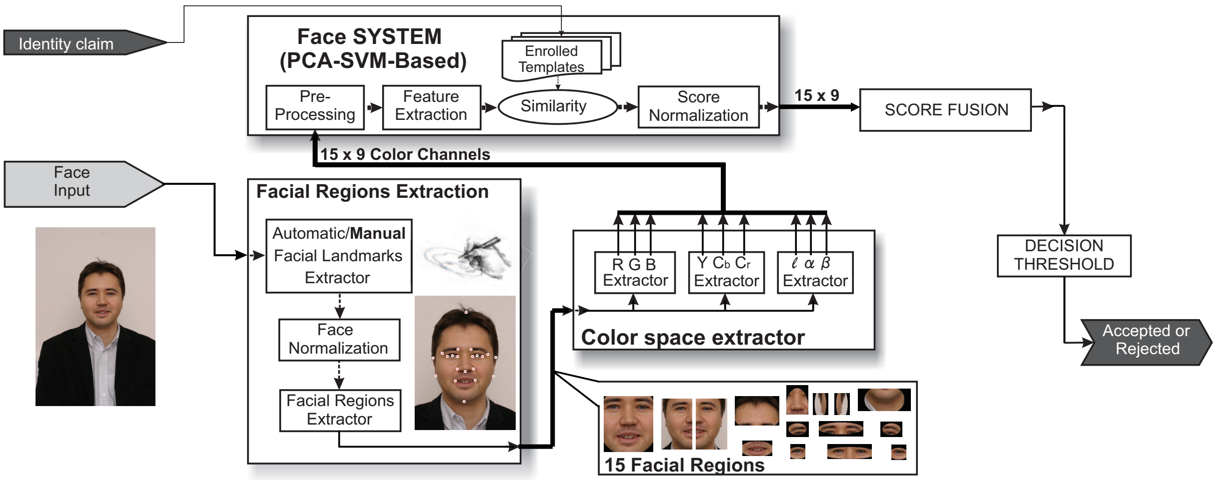


Figure 7.9: *Experimental framework diagram description for facial region fusion considering color information.*

7.4. Facial Regions-based Fusion using Color Information

Automatic face recognition systems are generally designed to match grayscale images of full faces. In most cases color information is discarded to decrease the computational cost of the algorithms and therefore additional discriminative information may be lost.

There are some previous works where grayscale facial region-based recognition is studied [Bonnen *et al.*, 2013; Ocegueda *et al.*, 2011; Tome *et al.*, 2013e] but non of them focus their attention in the color regions normally considered by forensic experts. In this work, we have considered the facial regions proposed in previous Sect. 5.3 following forensic protocols from law enforcement laboratories, allowing us to study individually the different facial regions normally considered in current practice of forensic examiners. In particular, we address in this section the problem of combining the most discriminative areas of the face for recognition using the available color information on a very challenging video surveillance scenario.

In contrast to traditional grayscale systems presented in previous sections, this section studies the discriminative power of each facial region using three color spaces: RGB, YC_bC_r, and $l\alpha\beta$. Fig. 7.9 summarizes the experimental framework followed.

The main objective here is therefore to understand to what extent the color information can help in region-based face recognition.

7.4.1. Extraction and Color Methodology

The proposed facial regions extraction framework is described in detail in previous Sect. 7.3. In this framework, two kinds of regions extraction are defined: *i*) based on human facial proportions, and *ii*) based on facial landmarks. For this section, the second extractor based on facial landmarks has been adopted. This extractor, based on facial landmarks manually located, allows to extract the facial regions with high precision. The final region extraction result is the

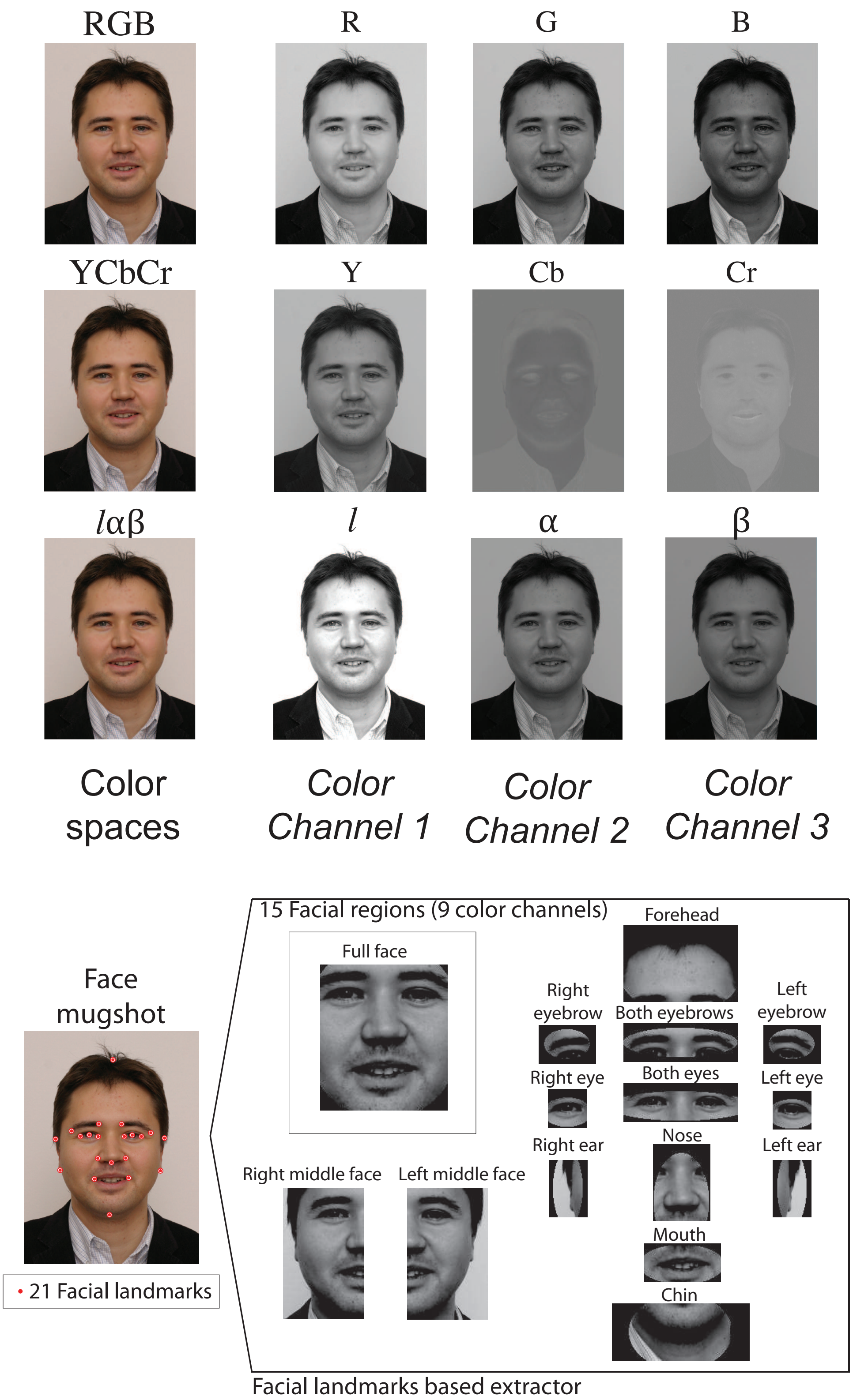


Figure 7.10: *(Top) Grayscale intensity values of faces for each color space analysed. (Bottom) Facial regions extraction based on facial landmarks extractor. The regions are extracted for the 9 color channels considered here.*

Color Channel 1 Id Num.	Color Channel 2 Id Num.	Color Channel 3 Id Num.	Facial Region	Facial Region Size (h × w)
1	16	31	Chin	75x181
2	17	32	Left ear	75x51
3	18	33	Right ear	75x51
4	19	34	Left eyebrow	51x75
5	20	35	Right eyebrow	51x75
6	21	36	Both eyebrows	51x151
7	22	37	Left eye	51x51
8	23	38	Right eye	51x51
9	24	39	Both eyes	51x151
10	**25**	**40**	**Full face**	**192x168**
11	26	41	Forehead	101x151
12	27	42	Left middle face	173x106
13	28	43	Right middle face	173x106
14	29	44	Mouth	51x101
15	30	45	Nose	101x75

Table 7.3: *Facial regions id for each color channel and their sizes for extractor based on facial landmarks (height × width in pixels).*

set of 15 facial regions (see Table 7.3) based on forensic laboratories protocols such as Spanish Guardia Civil (DGGC) or Netherlands Forensic Institute (NFI) as shown in Fig. 7.10.

There are some previous works where color spaces such as RGB or YC_bC_r have been used for face recognition [de Dios and Garcia, 2004; Liu and Liu, 2009; Singh *et al.*, 2003]. But, to the best of our knowledge, this is the first work where color information is used for face recognition using 15 facial regions.

When dealing with color images, the RGB color space is commonly used. This color space is composed by three channels (red, green, and blue), which are correlated among them. The components that form the second color space considered YC_bC_r are as follows: Y, luminance component, C_b, blue component $(B-Y)$, and C_r, red component $(R-Y)$ [Gonzalez and Woods, 2006].

Both RGB and YC_bC_r color spaces have correlated color channels among them. We also consider the $l\alpha\beta$ color space [Ruderman *et al.*, 1998], which minimizes the perceptual correlation among the channels of an image. The parameter l represents the luminance or brightness of the image and α and β represent the chromatic content, i.e., the color information. Fig. 7.10 (top) shows an example of each color channel for these three color spaces considered in the experiments.

7.4.2. Database and Experimental Protocol

The database used in our experiments SCface [Grgic *et al.*, 2011] (see Sect. 4.2.4). The experimental procedure followed corresponds to the *mugshot versus CCTV* scenario studied in Sect. 5.3.2, in this case also considering the color information of the facial regions.

7.4.3. Fusion Results

This section describes the fusion of the 15 forensic facial regions extracted from a human face in comparison with the performance of the whole *face* region normally used in face recognition systems. The fusion is carried out at score–level combining the facial regions for the color channels considered here.

Before carrying out the fusion, scores of the different facial regions are first normalized to the $[0, 1]$ range using the tanh-estimators described in Sect. 3.2, and then they are combined using sum fusion (i.e., using the fusion model in Sect. 3.2 and Eq. (3.1) where $g^r(B_i) = 1/R$ for all the R selected regions to fuse).

For this Thesis three different experiments were defined in order to analyse the potential of color information in a face recognition system: *i) Exp.1* Grayscale baseline system, where the grayscale facial regions are fused as the traditional face recognition systems. *ii) Exp.2* Fusion of color channels from each color space, (e.g. for RGB color space, the channels $\{R, G, B\}$ are fused for each facial region considered). *iii) Exp.3* Fusion of all color channels, where all 9 available color channels are fused for each face region.

7.4.3.1. *Exp.1*: Grayscale (Baseline System)

The baseline system is described in previous Sect. 7.3 where the fusion is carried out at the score–level for various combinations of grayscale regions.

The fusion results obtained for the three distances are summarized in Table 7.4 (*Exp.1*). As can be seen the system performance improves fusing several facial regions compared to just using the full *face* region.

Close and *medium* distance scenarios combine 7 facial regions to achieve the best result, but the *far* scenario needs to combine a total of 10 facial regions to obtain it. It is interesting to note that in the *close* scenario the best result is obtained with the fusion of inner and outer facial traits together with the full *face* (relative improvement of 56.7% in the EER with respect to using only the full face).

Similarly, in the two other distances considered, the best fusion includes inner and outer parts of the face, and relative improvements of over 40% in the EER are obtained with the fusion of regions compared to using only the full face.

7.4.3.2. *Exp.2*: Fusion of Three Color Channels

For the *Exp.2*, the score–level fusion is carried out fusing the three channels in a color space, i.e., $15 \times 3 = 45$ facial regions (as Table 7.3 shows) using a parallel fusion approach as in the previous experiment.

Table 7.4 (*Exp.2*) shows the fusion results for the three distances analysed. Fig. 7.11 shows the sequential fusion results obtained for the three distances and their corresponding color space with best performance ($l\alpha\beta$ for *close* and *far* distance, and RGB for *medium* distance). Similar to the previous case the system performance improves fusing several facial regions compared

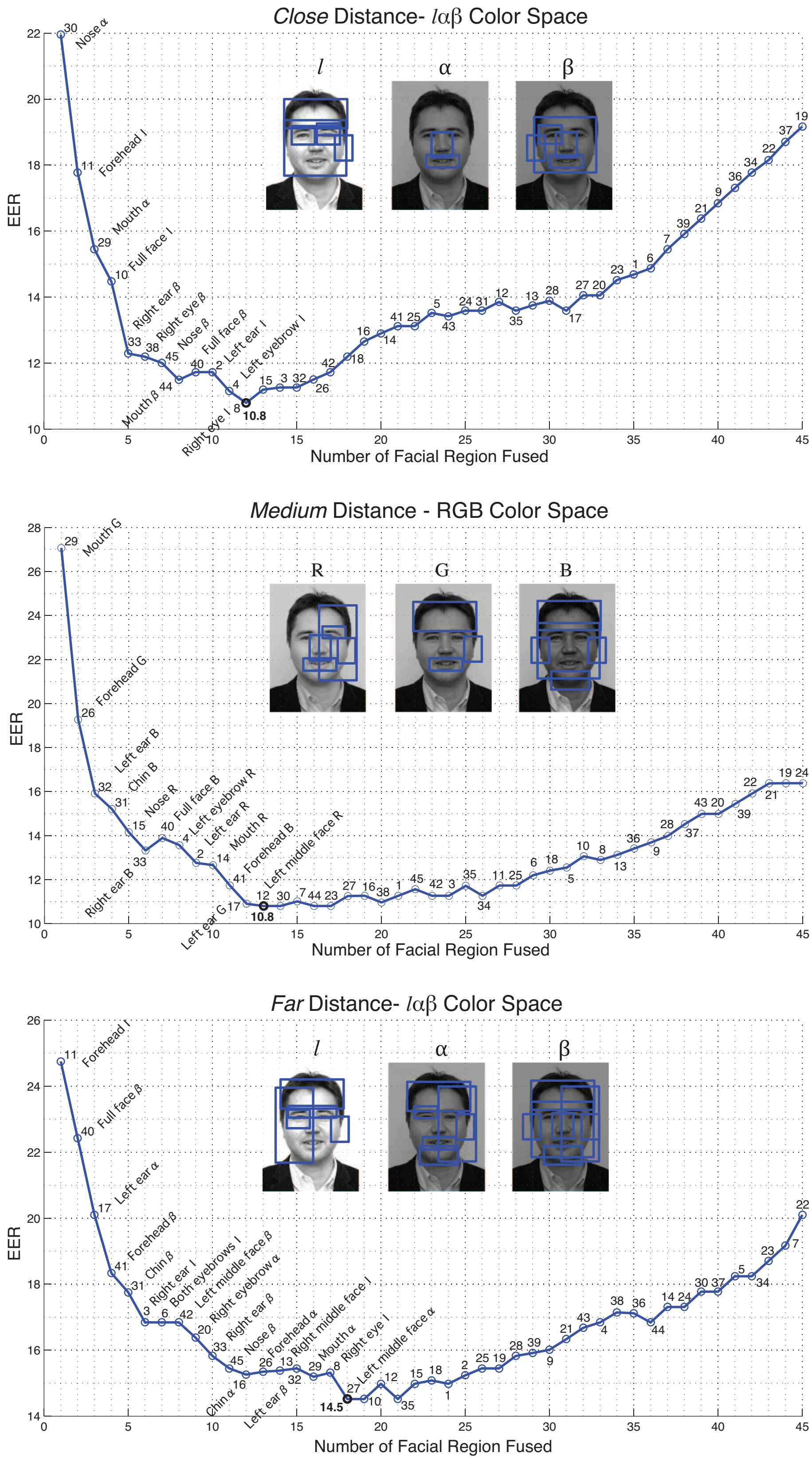


Figure 7.11: *EER for sum sequential fusion of the best combination of different facial regions for the best individual color space in each distance scenario: close (lαβ), medium (RGB) and far (lαβ).*

	Color Space	*Close Distance* Fusion (# Regions) – Full face	*Medium Distance* Fusion (# Regions) – Full face	*Far Distance* Fusion (# Regions) – Full face
Exp.1	*Grayscale*	14.30% (7) – 33.10%	12.90% (7) – 31.20%	16.80% (10) – 28.90%
Exp.2	*RGB*	11.58% (12) – 32.19%	**10.79% (13)** – 30.21%	14.61% (15) – 29.96%
	YC_bC_r	12.89% (16) – 29.50%	12.65% (8) – 33.35%	16.37% (21) – 31.72%
	$l\ \alpha\ \beta$	**10.79% (12)** – 31.82%	11.20% (16) – 31.09%	**14.50% (18)** – 28.93%
Exp.3	*ALL*	**9.03% (27)** – 29.96%	**10.33% (22)** – 30.33%	**13.12% (39)** – 28.93%

Table 7.4: *EER results for the score–level fusion obtained for sequential region fusion and the full face for the color channels of the three color spaces. In brackets we indicate the number of regions fused.*

to just using the *full face* region. It is interesting to note that the number of regions fused to obtain the best performance increases with the distance between the subject and the camera.

Comparing the fusion results with the baseline system based on grayscale facial regions, relative improvements of performance of 24.5%, 16.3%, and 13.7% for *close*, *medium* and *far* distance, are achieved respectively. These results support the utility of color information using facial regions to improve the performance of traditional face recognition systems.

7.4.3.3. *Exp.3*: Fusion of All Color Channels

In this case, all facial regions from all color channels are combined following the same fusion methodology. In this case, we combine the 3 sets of 45 facial regions considered in the previous experiment, i.e., 135 facial regions in total.

Table 7.4 (*Exp.3*) shows the fusion results for this experiment. As can be seen this experiment achieves the best EER results for the three distances compared to the previous experiment. However this case needs to fuse more facial regions to achieve the best performance (approximately double than *Exp.2*), and just around 1% EER of improvement is achieved compared to *Exp.2*. Again, the increment of the acquisition distance increases the number of facial regions to be combined to achieve the best performance.

Similarly, in the three distances considered, the best fusion includes inner and outer parts of the face, and relative improvements of over 66% in the EER are obtained with the regions fusion compared to only using the *full face*.

7.5. Chapter Summary and Conclusions

In this chapter we have evaluated the adaptive fusion schemes presented in Chapter 3. This study has compared scenario-based, soft biometrics-based, facial regions-based, and an extension of the last one using color facial regions-based schemes of score–level fusion and has studied their benefits in systems at a distance.

Regarding the scenario-based fusion, the effects of face acquisition distance on the performance of two common approaches for face verification have been studied using a new scenario estimator presented in Sect. 3.2.1. It has been found that the approach based on PCA subspace information and SVM classifier outperforms the DCT-GMM-based approach in close acquisition distance conditions but the approach based on DCT and GMM classifier is more robust to increasing acquisition distance.

We have also shown how the proposed acquisition distance estimator can be used in an adaptive score-level fusion approach to control the degradation observed in scenarios of varying acquisition distance. The proposed scheme leads to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of face acquisition distances.

Regarding the usage of soft biometrics, this chapter also reports an study of how the usage of soft labels can help to improve a biometric system for challenging person recognition scenarios at a distance. It is important to emphasize that the use of this ancillary information is very interesting in scenarios suffering from very high variability conditions. These soft labels can be visually identified at a distance by humans (or an automatic system) and fused with hard biometrics (as e.g., face recognition).

A soft biometric-based fusion has been proposed and studied to incorporate soft biometrics to these kinds of challenging scenarios at a distance considering a state-of-the-art face recognition system. Experiments are carried out considering both automatic and manual face detection. Results have shown the benefits of the soft biometrics information keeping robust the face recognition performance and also improving the performance on a high security level.

Regarding the facial regions-based fusion, this chapter reports an study of the combination of 15 human facial regions on various forensic scenarios. The best fused performance of facial regions is compared with the full *face* region, which is the normal case in face recognition. Results show that a combination of a set of facial regions can significantly improve the system performance by total average improvements of 51.5%, 52.3%, and 73.1% in the three scenarios considered, namely: *mugshot vs mugshot*, *mugshot vs CCTV*, and *CCTV vs CCTV*. Facial region-based fusion on these scenarios has been demonstrated to significantly improve a traditional full face recognition. In addition to be a useful background information that can guide and help experts to interpret and evaluate face evidences, these findings can have a significant impact on the design of face recognition algorithms. In particular, the approach followed for combining the information provided by the different regions can be significantly improved using more sophisticated fusion approaches (e.g., quality-based, user-dependent), and using more robust facial features descriptors.

Finally, the previous study was extended considering also color information by considering the 15 human facial regions previously extracted in three different color spaces. The best fused performance of facial regions is compared with the *full face* region, which is the normal case in face recognition. Experimetnal results show that a combination of a set of facial regions in different color spaces can significantly improve the system performance by a relative average improvement of over 66% for the three distances considered.

This chapter includes novel contributions in the application of adaptive fusion schemes (based on scenario, soft biometrics, face regions, and color information) to various scenarios at a distance but not in the individual systems used.

Chapter 8

Conclusions and Future Work

THIS THESIS has considered the problem of dealing with the variability factors affecting biometric systems at a distance through the use of soft biometric information and adaptive fusion. After a summary of the state-of-the-art in variability assessment, we have then defined what we understand by the scenario at a distance, and explained the evaluation methodology followed in the Thesis. These procedural guidelines for the systematic and objective evaluation of variability factors have been applied in the experimental studies described in the last chapters of the Dissertation to systems at a distance in three blocks: 1) first we have studied various variability sources in surveillance and forensic scenarios, 2) we have then studied the application of stand-alone soft biometrics to these scenarios, and 3) finally we have applied soft biometrics in such scenarios in combination to face recognition using various types of adaptive fusion.

8.1. Conclusions

Chapter 1 introduced the basics of biometric systems, biometric modalities, our perspective of the variability assessment problem in biometrics at a distance, the motivation of the Thesis, and the research contributions originated from this Thesis. Chapter 2 defined what we understand by scenario at a distance and summarized the most relevant works related to the different research lines developed in the Dissertation. The proposed methods were presented in Chapter 3, which are later studied in the experimental chapters. These new methods proposed and studied for overcoming the degradation found in biometric systems under uncontrolled variability factors are: 1) soft biometrics with application to video surveillance and forensics, and 2) adaptive fusion schemes based on the acquisition distance, soft biometric information, and facial regions. The first part of the Dissertation concluded with the description of the evaluation methodology followed in the Thesis, which also described the state-of-the-art in biometric databases at a distance and the most relevant datasets used in the Thesis.

The experimental part of the Thesis started in Chapter 5 studying the variability factors related to biometric scenarios at a distance. First, an exhaustive data-driven analysis was conducted on three realistic acquisition scenarios at different distances (*close*, *medium*, and *far*), as a first step towards devising adequate recognition methods capable to work in less constrained scenarios. Then, the effects of the face acquisition distance on the performance of two common approaches for face verification were studied using the new scenario estimator proposed in Chapter 3. The results demonstrated that the variability present in scenarios at a distance can be used in the training stage in order to stabilize the system performance degradation occurring in varying acquisition conditions. Moreover, a study of the variability of facial landmarks over two mugshot and CCTV databases have been reported, analysing both controlled and uncontrolled scenarios with low quality images and a large range of variability factors, finding that the landmarking variability increases with the distance, as expected. Comparing the two manual and automatic tagging approaches, the results show that the landmark variability is very similar for the set of common landmarks. The chapter concludes reporting an exhaustive analysis of the discriminative power of the different facial regions of the human face on various forensic scenarios. The comparison is carried out using two different region extractors based on facial landmarks and proportions, which are evaluated with automatic and manual landmarks. In all cases, we obtained that the recognition performance of facial regions depends on the acquisition distance. The best three facial regions with high discrimination power in the *close* distance are the *face, nose*, and *forehead*. However, in *far* distance, the best performance is achieved by the *forehead*. This facial region acquires an important role on scenarios at a distance such as CCTV versus CCTV. In addition to be a useful background information that can guide and help experts to interpret and evaluate face evidences, these findings can have a significant impact on the design of face recognition algorithms for these challenging scenarios at a distance.

Chapter 6 proposed and studied various types of soft biometric information suitable for video surveillance and forensics applications. It is important to emphasize that the use of this ancillary information is very interesting in scenarios suffering from very high variability conditions. These soft labels can be visually identified at a distance by humans (or automatic systems) and fused with hard biometrics (as e.g., face recognition). It is important to note that this kind of soft information is still a developing field in relation to its automatic extraction. A novelty in our study of soft biometrics is the treatment we have carried out depending on the acquisition distance. For that, we have defined and used three scenarios with different acquisition distance: *close*, *medium*, and *far*. The rationale behind this study is that depending on the scenario, some labels may not be visually present and others may be occluded. Thus, the discriminative information of soft biometrics will vary depending on the distance. It is worth noting that this relation between scenarios at a distance and the performance of soft biometrics for person recognition has not been studied in this way before. Regarding the soft biometrics proposed for forensics, we have followed forensic protocols based on the forensic morphological analysis. The resulting facial soft biometric traits can be either continuous or discrete. Traits such as the eyebrows height and width, interocular distance, naso-labial height, etc. are continuous variables in nature. On the other hand, these traits can be converted to discrete values using thresholds in order to simplify their classification and to compute population statistics. The experimental results have shown that a system that is completely based on facial soft biometrics features for forensics can provide good accuracy in person recognition tasks.

Chapter 7 evaluated the adaptive fusion schemes presented in Chapter 3. This study compared scenario-based, soft biometrics-based, facial regions-based, and color facial regions-based schemes of score–level fusion and studied their benefits in systems at a distance. As demonstrated, the variability present in scenarios at a distance can be used in the training stage in order to stabilize the system performance degradation occurring in varying acquisition conditions. In particular, we have shown how the proposed distance estimator can be used in an adaptive score-level fusion approach to control this degradation. The proposed scenario-based scheme leads to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of face acquisition distances. This chapter also presented a soft biometric-based fusion and studied how to incorporate soft biometrics to these kinds of challenging scenarios at a distance considering a state-of-the-art face recognition system. Experiments were carried out considering both automatic and manual face detection. Results have shown the benefits of the soft biometrics information improving the performance on a high security level. Moreover, this chapter reports an study of the combination of 15 human facial regions on various forensic scenarios. The best fused performance of facial regions is compared with the full *face* region, which is the normal case in face recognition. Preliminary results show that a combination of a set of facial regions can significantly improve the system performance by a total average improvement of 51.5%, 52.3%, and 73.1% in the three scenarios considered, namely: *mugshot vs mugshot*, *mugshot vs CCTV*, and *CCTV vs CCTV*. Facial region-based fusion on these scenarios has been demonstrated to significantly improve a traditional full face recognition. In addition, the combination of facial regions with color information allows to improve the system performance with a relative improvement of over 20% comparing with the traditional face recognition systems using only grayscale information.

In summary, the main results and contributions obtained from this Thesis are:

- The evaluation methodology of biometric systems at a distance followed throughout the Dissertation
- The relationship between the acquisition distance and the variability factors in order to define different scenarios, each of which can be analysed and processed differently.
- The new algorithms developed and used for dealing with variability factors in biometric systems at a distance: 1) soft biometrics for video surveillance and forensics, and 2) scenario-based and region-based fusion.
- The individual facial regions extractors developed, suitable for video surveillance systems.
- The landmarks and mugshot biometric data acquired, which is now available for research purposes.
- The experimental evidence and findings of the incorporation of soft biometrics information through adaptive fusion to person recognition systems working at a distance.

8.2. Future Work

A number of research lines arise from the work carried out in this Thesis. We consider of special interest the following ones:

- One promising direction for future research is the idea of using an "enhanced" enrollment including multiple images with lighting, pose, and expression variations for applications that allow it (e.g. some access control applications and some surveillance camera applications). The motivation for this approach comes from research on human face recognition (in other words the ability of people to recognize faces). One of the findings of this research is that people are much better at recognizing faces of familiar people. It appears that people build very good models of familiar faces and so can recognize such faces well even from very low resolution images. A computer algorithm could similarly take advantage of a good model of each face built at enrollment time. A single query image may then be compared to the detailed model for each enrolled face.

- Several works have already been published where the authors study the effects of different variability factors individually, such as [Phillips *et al.*, 2009a, 2005] or [Lui *et al.*, 2009]. The development of evaluation guidelines to analyse these effects jointly would help to build a better understanding about the real magnitude of the actual variability factors in operational face recognition systems at a distance.

- Li *et al.* [2009] consider realistic biometric at a distance environments without restrictions over environmental conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background, or photographic quality. Many algorithms do deal with these factors individually have been proposed in the literature [Chen *et al.*, 2006; Gross *et al.*, 2004; Lee *et al.*, 2005; Li *et al.*, 2007; Wang *et al.*, 2003; Zhou *et al.*, 2007]. Following the methodology used in this Thesis (based on scenario definition and characterization), one can group the various variability factors present in real-world application-oriented scenarios, and deal with them as a whole. This is source for future research.

- Searching for new variability compensation approaches for face recognition systems. For instance, Factor Analysis [McCool *et al.*, 2013; Prince *et al.*, 2008], a statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called factors. This compensation method would try to model the observed variables as a linear combination of potential factors in order to compensate the challenging variability factors produced by the increment of the acquisition distance in face recognition systems at a distance or on the move.

- Searching for new face recognition approaches for biometric systems at a distance. For instance, the Scale-Invariant Feature Transform (or SIFT) [Lowe, 1999], is an algorithm used in computer vision to detect and describe local features in images. Despite being an approach with initial target different to biometric recognition, several works have been published where the authors use this approach for carrying out biometric recognition systems on different traits, such as face [Križaj *et al.*, 2010], iris [Alonso-Fernandez *et al.*, 2009] or fingerprint [Park *et al.*, 2008]. This would give an interesting approach for biometric recognition at a distance, alternative to the actual methods, most of them not independent to scale.

- Combine the proposed facial regions with other existing face recognition approaches suitable for video surveillance such as [Sanderson and Lovell, 2009], where the feature extraction process in based on dividing the face on nine fixed regions.

- Searching for new methodologies to incorporate soft biometrics to compensate the variability in biometric systems at a distance. For instance, the development of reliable automatic systems able to extract soft biometrics information from the object of interest and from the scene where it is immersed. These systems would be very useful in surveillance and forensics helping the examiners to give informed decisions and the research community the opportunity to improve the systems recognition performance.

- Evaluating the robustness of other biometric traits suitable at a distance. Face recognition [Zhao *et al.*, 2003] is the more suitable trait used in biometric at a distance but there are important advances in sensor development capable to extract iris at a distance [Matey *et al.*, 2006]. The study of these techniques in combination with face or gait (when is available) may lead to enhanced multimodal approaches capable to work in demanding applications beyond the current state-of-the-art.

Apéndice A

Resumen Extendido de la Tesis

Tratamiento de Factores de Variabilidad y su Aplicación en Biometría a Distancia

Se denominA *reconocimiento biométrico* al proceso que permite asociar una identidad con un individuo de forma automática, mediante el uso de alguna característica personal que le sea inherente [Jain *et al.*, 2011b]. Aunque en el ámbito forense (judicial, policial y pericial), el análisis científico de evidencias biométricas se ha venido usando desde hace más de un siglo, el reconocimiento biométrico como medio automático de autenticación personal en aplicaciones comerciales o civiles es un área de investigación y desarrollo reciente.

Hoy en día el reconocimiento biométrico se puede considerar como un campo de investigación asentado, con libros de referencia [Jain *et al.*, 2008, 2011b; Ratha and Govindaraju, 2008; Ross *et al.*, 2006; Tistareli *et al.*, 2009], conferencias específicas en el área [Bowyer *et al.*, 2008a; Fierrez *et al.*, 2013; Tistarelli and Maltoni, 2007; Vijaya-Kumar *et al.*, 2008], evaluaciones y pruebas comparativas [Beveridge *et al.*, 2013; Phillips, 2006; Phillips *et al.*, 2011, 2009a,b], proyectos internacionales [BBfor2, 2010; BioSec, 2004; Biosecure, 2004; COST, 2007; MTIT, 2009; Tabula Rasa, 2010], consorcios específicos dedicados al reconocimiento biométrico [BC, 2005; BF, 2009; BI, 2009; EBF, 2009], esfuerzos de estandarización [ANSI/NIST, 2009; BioAPI, 2002; ISO/IEC JTC 1/SC 27 , 2009; SC37, 2005], y un creciente interés tanto por parte de gobiernos [BWG, 2009; DoD, 2005] como del sector comercial [IBIA, 2009; International Biometric Group, 2006].

Pese a la madurez de este campo de investigación, con trabajos que se remontan más de tres décadas en el tiempo [Atal, 1976; Bertillon, 1896; Kanade, 1973], el reconocimiento biométrico sigue siendo un área muy activa de investigación, con numerosos problemas prácticos aún por solucionar [Jain *et al.*, 2004c]. Estos problemas prácticos han hecho que, pese al interés de las aplicaciones biométricas, la integración en el mercado de estas nuevas tecnologías sea más lenta de lo esperado.

A.1. Resumen

Esta Tesis se centra en el tratamiento de los factores de variabilidad que afectan a sistemas de reconocimiento biométrico y aplicaciones biométricas a distancia. En particular, esta Tesis Doctoral explora el problema de la evaluación de los factores de variabilidad y cómo lidiar con ellos mediante la incorporación de información biométrica complementaria (del inglés "*soft biometrics*") con el fin de mejorar los sistemas de reconocimiento de personas a distancia. Los métodos propuestos apoyados por los resultados experimentales muestran los beneficios de la adaptación del sistema teniendo en cuenta la variabilidad de la muestra considerada.

A pesar de ser relativamente joven en comparación con otras tecnologías de seguridad maduras y ampliamente utilizadas, el reconocimiento biométrico ha surgido en la última década como una alternativa para aplicaciones donde se necesita el reconocimiento automático de personas. Ciertamente, el reconocimiento biométrico es muy atractivo y útil para los sistemas de vídeo vigilancia a distancia, ampliamente distribuidos en nuestro entorno, y para el usuario final (olvídese de PINs y contraseñas, usted es su propia llave). Sin embargo, no podemos olvidar que, como cualquier tecnología destinada a proporcionar un servicio de seguridad, los sistemas biométricos deben garantizar un rendimiento fiable en cualquier situación. Por lo tanto, es de especial relevancia comprender y analizar los factores de variabilidad a los que están sometidos dichos sistemas con el fin de asegurar su adecuado funcionamiento y aumentar sus beneficios para los usuarios.

En este contexto, la presente Tesis Doctoral da una idea del difícil problema de la evaluación de los factores de variabilidad a través del estudio sistemático de los escenarios biométricos a distancia y el análisis de las metodologías de compensación eficaces que pueden reducir al mínimo los efectos de los mismos. Por lo tanto, se persigue el objetivo de aumentar el rendimiento del reconocimiento de personas a distancia en esta próspera tecnología. De esta manera, los estudios experimentales presentados en esta Tesis Doctoral pueden ayudar a desarrollar aún más los esfuerzos tecnológicos de compensación variabilidad en curso, y pueden ser utilizados como guía para adaptar los sistemas existentes en reconocimiento biométrico a distancia para hacerlos más seguros y estables.

El problema de la compensación de la variabilidad en los sistemas biométricos ya había sido tratado en algunos trabajos anteriores, pero en la mayoría de los casos no se utiliza la relación de la distancia de adquisición con los factores de variabilidad, a fin de identificar y definir los escenarios de aplicación. En este trabajo, después de resumir y clasificar las obras más relevantes de la Tesis y definir lo que entendemos como escenario a distancia, se describen los métodos propuestos y se evalúan a lo largo de los capítulos experimentales. Estos capítulos experimentales se dedican primero al estudio de los factores de variabilidad (análisis de escenarios), y después a la aplicación de las técnicas propuestas para compensar los mismos (*soft biometrics* y fusión adaptativa). Todos los experimentos se llevaron a cabo utilizando bases de datos biométricas estándar de facto.

La parte experimental de la Tesis Doctoral comienza con la evaluación de los factores de variabilidad que se encuentran en los escenarios de los sistemas de reconocimiento facial. Evaluamos, entre otros: *i*) la relación entre los factores de variabilidad y la distancia en la adquisición de este tipo de sistemas, *ii*) la variabilidad de los puntos de referencia faciales sobre imágenes de la ficha policial y de circuito cerrado de televisión (CCTV), y *iii*) la variabilidad del rendimiento de diferentes regiones faciales del rostro humano en varios escenarios forenses a distancia. Dichos hallazgos pueden tener un impacto significativo en el diseño de algoritmos de reconocimiento facial además de ser una información útil que puede guiar y ayudar a los expertos en la interpretación y evaluación de evidencias faciales.

A continuación, estudiamos varios tipos de información complementaria (*soft biometrics*) disponible en el reconocimiento biométrico a distancia y adecuada para aplicaciones de vídeo vigilancia y forenses. Estas etiquetas *soft* pueden ser identificadas visualmente a distancia por los seres humanos (o un sistema automático) y su información discriminante puede variar dependiendo de la distancia. Es de interés señalar que esta relación entre los escenarios a distancia y el rendimiento de los sistemas *soft biometrics* para el reconocimiento de personas no se ha estudiado de esta manera antes. Por otra parte, también se introduce y evalúa un gran conjunto de características morfológicas biométricas faciales *soft*, extraídas siguiendo protocolos forenses. Los resultados experimentales que utilizan este conjunto de características demuestran que un sistema que se basa totalmente en las características faciales biométricas *soft* es factible para el análisis forense.

Por último, se estudian experimentalmente varios esquemas de fusión adaptativa que hacen uso de los sistemas *soft biometrics*. En particular, se estudian esquemas de fusión a nivel de puntuación basados en: identificación de escenario, *soft biometrics*, regiones faciales, y en regiones faciales combinadas en diferentes espacios de color. Los esquemas de fusión de adaptación propuestos logran mejoras notables que demuestran su utilidad en el reconocimiento biométrico a distancia.

El trabajo de investigación descrito en esta Tesis Doctoral ha dado lugar a aportaciones novedosas, que incluyen el desarrollo de dos nuevos métodos para hacer frente a los factores de variabilidad en los sistemas de reconocimiento biométrico a distancia, denominados: *i*) *soft biometrics* adecuados para vídeo vigilancia y el análisis forense, y *ii*) esquemas de fusión adaptativa a nivel de puntuación basados en el escenario de adquisición, *soft biometrics*, regiones faciales, y las regiones faciales usando información de color. Por otra parte, diferentes estudios experimentales originales se han llevado a cabo durante el desarrollo de la Tesis (por ejemplo, relación entre los escenarios a distancia y los factores de variabilidad). Además, el trabajo de investigación realizado a lo largo de la Tesis incluye la generación de diversas revisiones de la literatura y la generación de nuevos recursos biométricos.

A.2. Conclusiones

ESTA TESIS DOCTORAL ha considerado el problema del tratamiento de los factores de variabilidad que afectan a sistemas biométricos a distancia a través del uso de información biométrica complementaria ("*soft biometrics*") y fusion adaptativa. Después de resumir el estado del arte en evaluación de la variabilidad, hemos entonces definido lo que entendemos por un escenario a distancia, y explicado la metodología de evaluación utilizada a lo largo de la Tesis. Estas directrices de procedimiento para la evaluación sistemática y objetiva de los factores de variabilidad se han aplicado en los estudios experimentales, descritos en los últimos capítulos de la Disertación, a los sistemas a distancia en tres bloques: 1) primero hemos estudiado diversas fuentes de variabilidad en escenarios de vídeo vigilancia y forenses, 2) a continuación, se ha estudiado la aplicación del *soft biometrics* de forma independientes a estos escenarios, y 3) finalmente hemos aplicado esta biometría complementaria (*soft biometrics*) en dichos escenarios en combinación con el reconocimiento facial utilizando diversos tipos de fusión adaptativa.

El Capítulo 1 introdujo los conceptos básicos de los sistemas biométricos, las modalidades biométricas, nuestra perspectiva del problema de la evaluación de la variabilidad en la biométrica a distancia, la motivación de la Tesis, y las contribuciones de investigación originadas a partir de esta Tesis Doctoral. En el Capítulo 2 se definió lo que entendemos por escenario a una distancia y se resumieron los trabajos más relevantes relacionados con las diferentes líneas de investigación desarrolladas en la Tesis. Los métodos propuestos fueron presentados en el Capítulo 3, los cuales fueron estudiados posteriormente en los capítulos experimentales. Los métodos propuestos y estudiados para hacer frente a la degradación encontrada en sistemas biométricos con factores de variabilidad no controlados son los siguientes: 1) *soft biometrics* con aplicación en vídeo vigilancia y análisis forense, y 2) esquemas de fusión adaptativa basados en la distancia de adquisición, la información complementaria (*soft*), y las regiones faciales. La primera parte de la Tesis concluye con la descripción de la metodología de evaluación utilizada en la Tesis, la cual también describe el estado del arte en bases de datos biométricas a distancia y los conjuntos de datos de mayor relevancia utilizadas en esta Tesis Doctoral.

La parte experimental de la Tesis comenzó en el Capítulo 5 estudiando los factores de variabilidad relacionadas con escenarios biométricos a distancia. En primer lugar se llevó a cabo un análisis exhaustivo de datos en tres escenarios de adquisición realistas a diferentes distancias (*cerca*, *medio*, y *lejos*), como un primer paso hacia la elaboración de los métodos de reconocimiento adecuados capaces de trabajar en escenarios menos controlados. A continuación, se estudiaron los efectos de la distancia adquisición en los sitemas de reconocimiento facial cara en base al rendimiento de los sistemas mediante el nuevo estimador escenario propuesto en el Capítulo 3. Los resultados demostraron que la variabilidad presente en escenarios a distancia se puede utilizar en la etapa de entrenamiento con el fin de estabilizar la degradación del rendimiento del sistema en diferentes condiciones de adquisición.

Además, se ha realizado un estudio de la variabilidad de los puntos de referencia faciales sobre dos bases de datos con imágenes mugshot y de circuito cerrado de televisión (CCTV).

Este análisis se ha llevado a cabo en escenarios tanto controlados y no controlados con imágenes de baja calidad y una amplia gama de factores de variabilidad, encontrando que la variabilidad en el marcado aumenta con la distancia, como se esperaba. Los resultados muestran que la variabilidad punto de referencia es muy similar para el conjunto de puntos de referencia comunes comparando enfoques manuales y automáticas de marcado.

El capítulo concluye con el análisis exhaustivo del poder discriminante de las distintas regiones faciales del rostro humano en varios escenarios forenses. La comparación se realiza mediante dos extractores regiones diferentes basados en puntos de referencia y en proporciones faciales, que son evaluados con los puntos de referencia extraídos de forma automática y manual. En todos los casos, se obtuvo que el rendimiento del reconocimiento de las regiones faciales depende de la distancia de adquisición. Los tres mejores regiones faciales mejor poder de discriminación en la distancia *cerca* son la *cara*, la *nariz*, y la *frente*. Sin embargo, en la distancia *media*, el mejor rendimiento se logra por medio de la *frente*. Esta región facial adquiere un papel importante en escenarios a distancia tal como CCTV frente a CCTV. Además de ser una información de base útil que puede guiar y ayudar a los expertos para la interpretación y evaluación de las evidencias faciales, estos hallazgos pueden tener un impacto significativo en el diseño de algoritmos de reconocimiento facial en escenarios complejos como son los escenarios a ditancia.

En el Capítulo 6 se propuso y estudió diversos tipos de información biométrica complementaria decuado para vídeo vigilancia y aplicaciones forenses. Es importante destacar en que el uso de esta información auxiliar es muy interesante en escenarios que sufren condiciones de alta variabilidad. Estas etiquetas *soft* pueden ser identificadas visualmente a distancia por los seres humanos (o mediante sistemas automáticos) y pueden ser fusionadas con los sistemas tradicionales de reconocimiento biométrico (como por ejemplo, el reconocimiento facial). Es importante señalar que la extracción automática de este tipo de información complemetaria es todavía un campo en desarrollo. Una novedad en nuestro estudio de *soft biometrics* es el tratamiento que hemos llevado a cabo en función de la distancia de adquisición. Para ello, se han definido y utilizado tres escenarios con diferentes distancias de adquisición: *cerca*, *medio*, y *lejos*. La lógica detrás de este estudio es que, dependiendo del escenario de aplicación, algunas etiquetas *soft* no se encuentran visualmente presentes y otras pueden estar ocluidas. Por lo tanto, la información discriminate de los sistemas de *soft biometrics* variará dependiendo de la distancia. Vale la pena señalar que esta relación entre los escenarios a distancia y el rendimiento de los sistemas de *soft biometrics* para el reconocimiento personas no se ha estudiado de esta manera antes. Por otro lado se han propuesto y analizado un conjunto de datos faciales biométricos *soft* extraídos siguiendo protocolos forenses basadas en el análisis morfológico. Los rasgos biométricos faciales *soft* resultantes son dos conjuntos de datos: 1) continuos (como por ejemplo, la altura y la anchura cejas, distancia interocular, altura naso-labial, etc) y 2) discretos. Dichos rasgos continuos han sido utilizados para extraer otro conjunto de valores discretos utilizando umbrales con el fin de simplificar su clasificación y para calcular estadísticas de población. Los resultados experimentales han demostrado que un sistema basado totalmente en las características faciales biométricas *soft* propuestas para análisis forense puede proporcionar una buena precisión en tareas de reconocimiento de persona.

Por último en el Capítulo 7 se ha evaluado los esquemas de fusión adaptativa que se presentarón en el Capítulo 3. Dichos esquemas de fusión a nivel de puntuación fueron los basados en: distancia de acquisición, *soft biometrics*, regiones faciales y regiones faciales utilizando información de color, para los cuales se estudió los beneficios que aportaban a los sistemas a distancia. Como se demostró el estimador de distancia de adquisición puede ser utilizado en el esquema propuesto de fusión basada en escenario como enfoque para controlar la degradación de los sistemas a distancia. Dicho esquema basado en escenario conduce a un rendimiento mejorado del sistema mateniendo su robustez en un amplio rango de distancias de adquisición de la cara. Este capítulo también analizó la fusión basada en *soft biometrics* y estudió la manera de incorporar la información complementaria a escenarios complejos a distancia considerando sistemas de reconocimiento facial del estado del arte. Los experimentos se llevaron a cabo teniendo en cuenta detección facial automática y manual. Los resultados han demostrado los beneficios de la información biométrica complementaria mejorando el rendimiento de los sistemas en un punto de trabajo de seguridad.

Además, este capítulo informa de un estudio de la combinación de 15 regiones faciales del rostro humano en diversos escenarios forenses. El mejor rendimiento de la combinación de las distintas regiones faciales se compara con la región completa de la *cara*, región habitualmente utilizada en los sistemas de reconocimiento facial. Los resultados muestran que la combinación de un conjunto de regiones faciales puede mejorar significativamente el rendimiento del sistema obteniendo una mejora promedio de 51.5%, 52.3%, and 73.1% en los tres escenarios considerados, denominados: *mugshot vs mugshot*, *mugshot vs CCTV*, and *CCTV vs CCTV*. Demostrando por tanto que la fusión basada en regiones faciales en este tipo de escenarios mejora significativamente los sistemas tradicionales de reconocimiento facial que utilizan la cara completa. Adicionalmente, se ha demostrado que la combinación de las regiones faciales con información de color permite mejorar el rendimiento del sistema con una mejora relativa de más del 20% en comparación con los sistemas de reconocimiento facial tradicionales que trabajan sólo con imágenes en escala de grises.

En resumen, los principales resultados y contribuciones obtenidos en esta Tesis Doctoral son:

- La evaluación metodológica de los sistemas biométricos a distancia seguida a lo largo de toda esta disertación.

- La relación entre la distancia de adquisición y los factores de variabilidad con el fin de definir escenarios, de forma que cada uno de los cuales pueda ser analizado y procesado de forma diferente.

- El desarrollo y uso de nuevos algoritmos para el tratamiento con los factores de variabilidad en sistemas de reconocimiento biométrico a distancia: 1) información complementaria (*soft biometrics*) aplicada a vídeo vigilancia y el análisis forense, y 2) fusión basada en escenario y en regiones faciales.

- Los extractores de regiones faciales individuales desarrollados, adecuados para sistemas de vídeo vigilancia.
- Los puntos de referencia faciales etiquetados sobre imágenes mugshot y CCTV de bases de datos del estado del arte, los cuales ahora están disponibles para propósitos de investigación.
- Las evidencias y hallazgos experimentales de la incorporación de información complementaria a través de fusión adaptativa para los sistemas de reconocimiento de personas a distancia.

A.3. Líneas de Trabajo Futuro

Se proponen las siguientes líneas de trabajo futuro relacionadas con el trabajo desarrollado en esta Tesis Doctoral:

- Una interesante línea de investigación futura es la idea de usar un registro "mejorado", incluyendo varias imágenes con variaciones de iluminación, pose, y expresión para las aplicaciones que lo permiten (aplicaciones de control de acceso y vídeo vigilancia). La motivación de este enfoque proviene de la investigación sobre como el ser humano es capaz de reconocer rostros. Uno de los hallazgos de esta investigación es que las personas son mucho mejores reconociendo rostros de personas conocidas, es decir, construimos modelos robustos con gran cantidad de caras conocidas aumentando así la precisión, incluso a partir de imágenes de muy baja resolución. Por lo tanto, un algoritmo informático podría aprovechar dicha ventaja construyendo un buen modelo de cada cara en el momento del registro. De forma que una simple imagen pueda ser comparada con un modelo robusto y detallado para cada cara registrada en el sistema.
- Varios trabajos han sido publicados [Phillips *et al.*, 2009a, 2005] or [Lui *et al.*, 2009] donde los autores estudian los efectos de los diferentes factores de variabilidad de forma individual. El desarrollo de pautas de evaluación para analizar estos efectos de forma conjunta ayudaría a construir un mejor entendimiento acerca de la influencia real de dichos factores de variabilidad en sistemas de reconocimiento facial a distancia.
- Li *et al.* [2009] consideran el reconocimiento biométrico real en entornos a distancia sin restricciones sobre las condiciones ambientales como la escala, la pose, la iluminación, el enfoque, la resolución, la expresión facial, los accesorios, el maquillaje, las oclusiones, fondo, o la calidad fotográfica. Muchos algoritmos han sido propuestos en la literatura para hacer frente a estos factores de forma individual [Chen *et al.*, 2006; Gross *et al.*, 2004; Lee *et al.*, 2005; Li *et al.*, 2007; Wang *et al.*, 2003; Zhou *et al.*, 2007]. Siguiendo la metodología usada en esta Tesis (basada en la definición de escenarios y caracterización), se pueden agrupar los diversos factores de variabilidad presentes en los escenarios de aplicación orientados al

mundo real, y tratar con ellos como un conjunto no individual. Esto es una interesante línea de investigación futura.

- La búsqueda de nuevos enfoques de compensación variabilidad para los sistemas de reconocimiento facial. Por ejemplo, Factor Analysis [McCool *et al.*, 2013; Prince *et al.*, 2008], un método estadístico utilizado para describir la variabilidad entre las variables observadas en términos de un número potencialmente menor de variables no observadas llamadas factores. Este método de compensación trataría de modelar las variables observadas como una combinación lineal de los factores potenciales con el fin de compensar los factores de variabilidad desafiantes producidos por el incremento de la distancia de la adquisición en sistemas de reconocimiento facial a distancia o en el movimiento.

- La búsqueda de nuevos métodos de reconocimiento facial para sistemas biométricos a distancia. Por ejemplo, Scale-Invariant Feature Transform (or SIFT) [Lowe, 1999], es un algoritmo utilizado en la visión artificial para detectar y describir las características locales de las imágenes. A pesar de ser un algoritmo con un enfoque inicial diferente al reconocimiento biométrico, podemos encontrar varios trabajos publicados donde los autores utilizan este enfoque para la realización de sistemas de reconocimiento biométrico utilizando diferentes características, tales como la cara [Križaj *et al.*, 2010], el iris [Alonso-Fernandez *et al.*, 2009] o la huella dactilar [Park *et al.*, 2008]. Este podría ser un enfoque interesante para el reconocimiento biométrico a distancia, alternativa a los métodos actuales donde la mayoría de ellos no son independientes a escala.

- Combinar las regiones faciales propuestas con otros enfoques de reconocimiento facial existentes adecuados para la vídeo vigilancia, como [Sanderson and Lovell, 2009], donde el proceso de extracción de características se realiza en base a la división de la cara en nueve regiones fijas.

- Búsqueda de nuevas metodologías para incorporar datos biométricos *soft* para compensar la variabilidad en los sistemas de reconocimiento biométricos a distancia. Por ejemplo, el desarrollo de sistemas automáticos fiables capaces de extraer información biométrica complementaria del objeto de interés y de la escena en la que éste está inmerso. Estos sistemas serían muy útiles en la vídeo vigilancia y el análisis forense ayudando a los expertos a emitir decisiones y dando a los investigadores la oportunidad de mejorar el rendimiento de los sistemas de reconocimiento.

- La evaluación de la robustez de otros rasgos biométricos adecuados para sistemas a distancia. El reconocimiento facial [Zhao *et al.*, 2003] es el rasgo más adecuado usado en el reconocimiento biométrico a distancia, pero hay importantes avances en el desarrollo de sensores capaces de extraer rasgos tales como el iris a distancia [Matey *et al.*, 2006]. El estudio de estas técnicas en combinación con la cara o la forma de andar (cuando esté disponible) puede conducir a enfoques mejorados multimodales capaces de trabajar en las aplicaciones más exigentes más allá del actual estado del arte.

References

F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia. Iris recognition based on sift features. In *IEEE Proc. Intl. Conf. on Biometrics, Identity and Security, BIDS*, September 2009. 14, 165, 174

ANSI/NIST. NIST ITL american national standards for biometrics, 2009. http://fingerprint.nist.gov/standard/. 1, 167

S. Arca, P. Campadelli, and R. Lanzarotti. An efficient method to detect facial fiducial points for face recognition. In *17th International Conference on Pattern Recognition (ICPR'04)*, pages 532–535, Washington, DC, USA, 2004. IEEE Computer Society. 79

S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432–443, Mar. 2006. 79

B. S. Atal. Automatic recognition of speakers from their voices. 64:460–475, 1976. 1, 47, 167

E. Bailly-Bailliere *et al.* The BANCA database and evaluation protocol. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 625–638. Springer LNCS-2688, 2003. 51

J. R. Barr, K. W. Bowyer, and P. J. Flynn. Detecting questionable observers using face track clustering. In *Workshop on Applications of Computer Vision*, 2011. 52

BBfor2, 2010. BBfor2: Bayesian Biometrics for Forensics, FP7-ITN-2008-238803. (http://www.bbfor2.net/). 1, 167

BC, 2005. Biometrics Consortium. (http://www.biometrics.org/). 1, 167

C. Bergman. *Advances in biometrics*, chapter Match-on-card for secure and scalable biometric authentication, pages 407–421. Springer, 2008. 2

A. Bertillon. Identification anthropométrique et instructions signalétiques. *Melun: Imprimerie administrative*, 1893. 1

A. Bertillon. *Signaletic Instructions including the theory and practice of Anthropometrical Identification*. R.W. McClaughry Translation,The Werner Company, 1896. 6, 30, 34, 38, 117, 167

G. M. Beumer, Q. Tao, A. M. Bazen, and R. N. J. Veldhuis. A landmark paper in face recognition. In *7th International Conference on Automatic Face and Gesture Recognition*, FGR '06, pages 73–78, Washington, DC, USA, 2006. IEEE Computer Society. 79

R. Beveridge, J. Phillips, D. Bolme, B. Draper, G. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. Bowyer, P. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras (pasc). In *Biometrics Theory, Applications and Systems (BTAS)*, 2013. 1, 7, 53, 167

BF. The biometric foundation, 2009. (http://www.biometricfoundation.org/). 1, 167

BI. Biometrics institute, 2009. (http://www.biometricsinstitute.org/). 1, 167

J. Bigun. *Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision*. Springer, 2006. 11

M. Binetskaya. Reconocimiento facial en el ámbito forense. Master's thesis, Universidad Autonoma de Madrid, September 2013. 12, 14

BioAPI, 2002. ANSI INCITS 358-2002 - Information Technology - BioAPI Specification (Version 1.1). 1, 167

BioSec, 2004. Biometrics and Security, FP6 IP IST-2002-001766. (http://www.biosec.org/). 1, 167

Biosecure, 2004. Biometrics for Secure Authentication, FP6 NoE IST-2002-507634. (http://www.biosecure.info/). 1, 167

L. Blazquez. Reconocimiento facial basado en puntos característicos de la cara en entornos no controlados. Master's thesis, Universidad Autonoma de Madrid, February 2012. 14

M. Bone and D. Blackburn. Face recognition at a chokepoint. Technical report, DoD Counterdrug Technology Development Program Office, November 2002. 48

M. Bone and C. Crumbacker. Facial recognition: Assessing its viability in the corrections environment. *Corrections Today Magazine*, pages 62–64, July 2001. 48

K. Bonnen, B. Klare, and A. K. Jain. Component-based representation in automated face recognition. *IEEE Transactions on Information Forensics and Security*, 8(1):239–253, 2013. 88, 92, 153

K. W. Bowyer, V. Govindaraju, M. Nixon, and N. Ratha, editors. *Proc. of Second International Conference Biometrics: Theory, advances and systems (BTAS)*, 2008a. IEEE Press. 1, 167

K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Comput. Vis. Image Underst.*, 110(2):281–307, 2008b. ISSN 1077-3142. 28

BSI. Biometric services international, 2009. (http://www.biometricsinternational.org/). 48

BWG. Communications-electronics security group (CESG) – biometric working group (BWG) (UK government), 2009. http://www.cesg.gov.uk/policy_technologies/biometrics/index.shtml. 2, 167

R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance evaluation of fingerprint verification systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(1):3–18, 2006. 48

F. Cardinaux, C. Sanderson, and S. Bengio. User authentication via adapted statistical models of face images. *Transaction on Signal Processing*, 54(1):361–373, 2006. 8

F. Cardinaux, C. Sanderson, and S. Marcel. Comparison of mlp and gmm classifiers for face verification on xm2vts. In *IARP International Conference on Audio- and Video-based Biometric Person Authentication*, pages 911–920, 2003. 74, 75

T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1519–1524, 2006. ISSN 0162-8828. 164, 173

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. ISSN 0885-6125. 74, 75, 93

COST. COST 2101: Biometrics for identity documents and smart cards, 2007. http://cost2101.org/. 1, 167

T. Dang, C. Hoffmann, and C. Stiller. Continuous stereo self-calibration by camera parameter tracking. *Trans. Img. Proc.*, 18(7):1536–1550, 2009. ISSN 1057-7149. 25

J. J. de Dios and N. Garcia. Fast face segmentation in component color space. In *Proc. of IEEE International Conference on Image Processing*, pages 191–194, 2004. 155

S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan. Soft-biometrics: Unconstrained authentication in a surveillance environment. In *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, pages 196–203, 2009. 35

DoD, 2005. Biometrics Management Office, Department of Defense, USA. (http://www.biometrics.dod.mil/). 2, 167

A. Dragolici. Detección, localización e identificación biométrica de caras en imágenes: métodos y evaluación en el marco nist-mbgg. Master's thesis, Universidad Autonoma de Madrid, June 2010. 15

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001. 11, 33

EBF, 2009. European Biometrics Forum. (http://www.eubiometricforum.com/). 1, 167

J. Eslava-Rios. Reconocimiento facial en tiempo real. Master's thesis, Universidad Autonoma de Madrid, June 2013. 14

European Network of Forensic Science Institutes. http://enfsi.eu/. 62, 82

Facial Identification Science Working Group (FISWG). https://www.fiswg.org. 38

T. Faltemier, K. Bowyer, and P. Flynn. A region ensemble for 3-d face recognition. *Information Forensics and Security, IEEE Transactions on*, 3(1):62–73, 2008. ISSN 1556-6013. 99

J. Fierrez. *Adapted Fusion Schemes for Multimodal Biometric Authentication*. PhD thesis, Universidad Politecnica de Madrid, 2006. 41, 43, 149

J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano, G. G. de Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardeñoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Uruñuela, F. Martinez-Contreras, and J. J. Gracia-Roche. BiosecurID: a multimodal biometric database. *Pattern Analysis and Applications*, 13:235–246, 2009. To appear. 51

J. Fierrez, A. Kumar, M. Vatsa, R. Veldhuis, and J. Ortega-Garcia, editors. *Proc. of 6th International Conference on Biometrics (ICB)*, 2013. IEEE. 1, 167

J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez. BioSec baseline corpus: A multimodal biometric database. *Pattern Recognition*, 40:1389–1392, 2007. 51

J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition*, 38(5):777–779, 2005. 51

P. J. Flynn. *Handbook of Biometrics*, chapter Biometric Databases. Springer, 2007. 51

H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Fingerprint image quality estimation and its application to multi-algorithm verification. *Trans. on Information Forensics and Security*, 3(2):331–338, 2008. 44

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd ed. edition, October 1990. 74

J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43(3):1027–1038, March 2010. 138

J. Gold, P. Mundy, and B. Tjan. The perception of a face is no more than the sum of its parts. *Psychol Science*, 23(4):427–34, 2012. 29, 88

R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006. ISBN 013168728X. 11, 155

J. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975. ISSN 0033-3123. 92

M. Grgic, K. Delac, and S. Grgic. Scface - surveillance cameras face database. *Multimedia Tools Appl.*, 51(3): 863–879, Feb. 2011. ISSN 1380-7501. 56, 93, 149, 150, 155

R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(4):449–465, 2004. ISSN 0162-8828. 164, 173

P. J. Grother, R. J. Micheals, and P. J. Phillips. Face recognition vendor test 2002 performance metrics. In *Proc. IAPR Audio- and Video-based Person Authentication (AVBPA)*, pages 937–945. Springer LNCS-2688, 2003. 48

H. Gunes and M. Piccardi. Assessing facial beauty through proportion analysis by image processing and supervised learning. *International Journal of Man-Machine Studies*, 64(12):1184–1199, 2006. 92

A. Gupta and L. S. Davis. Beyond nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers. In *Proc. of ECCV*, pages 16–29. LNCS-5302, 2008. 34

S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3d face recognition. *Int. J. Comput. Vision*, 90(3): 331–349, Dec. 2010. ISSN 0920-5691. 79, 88

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions (Wiley Series in Probability and Statistics)*. Wiley-Interscience, New York, revised edition, April 2005. ISBN 0471735779. 41

D. Heckathorn, R. Broadhead, and B. Sergeyev. A methodology for reducing respondent duplication and impersonation in samples of hidden populations. In *Proc. of Annual Meeting of the American Sociological Association*, pages 543–564, Toronto, Canada, 1997. 34

B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *Int. J. Computer Vision*, 74(2):167–181, Aug. 2007. ISSN 0920-5691. 88

P. Herranz. Reconocimiento de actividades en video basado en ejemplos. Master's thesis, Universidad Autonoma de Madrid, May 2010. 15

J. Hewig, R. H. Trippe, H. Hecht, T. Straube, and W. H. R. Miltner. Gender differences for specific body regions when looking at men and women. *Journal of Nonverbal Behavior*, 32(2):67–78, 2008. Hewig, Johannes Trippe, Ralf H. Hecht, Holger Straube, Thomas Miltner, Wolfgang H. R. 30, 36

G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 34

K. Huang and S. Aviyente. Sparse representation for signal classification. pages 609–616, 2006. 141

IBIA. International biometric industry association, 2009. (http://www.ibia.org/association/). 2, 167

International Biometric Group. Biometrics market and industry report 2006-2010, 2006. (http://www.biometricgroup.com/). 2, 48, 167

ISO/IEC 17025. ISO/IEC 17025:2005, general requirements for the competence of testing and calibration laboratories., 2005. 48

ISO/IEC 19794-5:2011. Information technology – biometric data interchange formats – part 5: Face image data. In *International Organization for Standardization*, 2011. 83, 91

ISO/IEC 29794-1. ISO/IEC 29794-1:2006 biometric quality framework standard, 2006. 18

ISO/IEC JTC 1/SC 27 . IT security techniques, 2009. http://www.jtc1.org/sc27/. 1, 167

A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005. ISSN 00313203. 41, 113, 125

A. K. Jain, S. C. Dass, K. Nandakumar, and K. N. Soft biometric traits for personal recognition systems. In *Proceedings of International Conference on Biometric Authentication, Hong Kong*, pages 731–738, 2004a. 34

A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. 22(1):4–37, 2000. 33

A. K. Jain, R. Jin, and J.-E. Lee. Tattoo image matching and retrieval. *Computer*, 45(5):93–96, 2012a. 30

A. K. Jain, B. Klare, and U. Park. Face recognition: Some challenges in forensics. In *FG*, pages 726–733. IEEE, 2011a. 29

A. K. Jain, B. Klare, and U. Park. Face matching and retrieval in forensics applications. *IEEE MultiMedia*, 19 (1):20, 2012b. 79, 88

A. K. Jain, K. N, X. Lu, and U. Park. Integrating faces, fingerprints and soft biometric traits for user recognition. In *In Proceedings of Biometric Authentication Workshop, LNCS 3087*, pages 259–269. Springer, 2004b. 34

A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman. Biometrics: A grand challenge. In *in Proc. IAPR Int. Conf. on Pattern Recognition (ICPR)*, volume 2, pages 935–942, 2004c. 167

A. K. Jain and U. Park. Facial marks: Soft biometric for face recognition. In *IEEE International Conference on Image Processing*, pages 37–40, nov 2009. 34

A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM*, 47(1):34–40, 2004. 135

A. K. Jain, A. Ross, and P. Flynn, editors. *Handbook of biometrics*. Springer, 2008. 1, 2, 167

A. K. Jain, A. Ross, and S. Pankanti. Biometrics: A tool for information security. *IEEE Trans. on Information Forensics and Security*, 1(2):125–143, 2006. 1, 2

A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1):4–20, 2004d. 2, 47, 48

A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer Science+Business Media, LLC, 2011b. 1, 2, 4, 9, 21, 167

Y. Jefferson. Facial beauty-establishing a universal standard. *IJO*, 15(1):9, 2004. 92

T. Kanade. *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973. 1, 47, 167

T.-K. Kim, H. Kim, W. Hwang, and J. Kittler. Component-based lda face description for image retrieval and mpeg-7 standardisation. *Image Vision Comput.*, 23(7):631–642, Jul 2005. ISSN 0262-8856. 88, 92

J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998. ISSN 0162-8828. 138

P. Komarinski. *Automated Fingerprint Identification Systems (AFIS)*. Elsevier, 2005. 3

J. Križaj, V. Štruc, and N. Pavešić. Adaptation of sift features for robust face recognition. In *Proceedings of the 7th International Conference on Image Analysis and Recognition (ICIAR 2010)*, pages 394–404, Povoa de Varzim, Portugal, June 2010. 165, 174

N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *In IEEE International Conference on Computer Vision (ICCV*, 2009. 34

J. Lee, A. K. Jain, and R. Jin. Scars, marks and tattoos (smt): Soft biometric for suspect and victim identification. In *In Proc. Biometric Symposium, Biometric Consortium Conference*, pages 1–8, 2008. 34

K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):684–698, 2005. ISSN 0162-8828. 164, 173

S. Lee and S. Li, editors. *Proc. of Second International Conference on Biometrics (ICB)*, 2007. Springer LNCS-4642. 1

F. Li, H. Wechsler, and M. Tistarelli. Robust fusion using boosting and transduction for component-based face recognition. In *10th International Conference on Control, Automation, Robotics and Vision, ICARCV 2008, Hanoi, Vietnam, 17-20 December 2008, Proceedings*, pages 434–439. IEEE, 2008. 88

S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):627–639, 2007. ISSN 0162-8828. 164, 173

S. Z. Li and A. K. Jain, editors. *Handbook of Face Recognition*. 2nd Edition. Springer, 2011. 7, 26

S. Z. Li, B. Schouten, and M. Tistarelli. *Handbook of Remote Biometrics for Surveillance and Security*, chapter Biometrics at a Distance: Issues, Challenges, and Prospects, pages 3–21. Springer, 2009. 108, 134, 164, 173

Z. Liu and C. Liu. Robust face recognition using color information. In M. Tistarelli and M. Nixon, editors, *Advances in Biometrics*, volume 5558 of *Lecture Notes in Computer Science*, pages 122–131. Springer Berlin Heidelberg, 2009. 155

D. G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150–1157 vol.2. IEEE Computer Society, August 1999. ISBN 0-7695-0164-8. 164, 174

S. Lucey and T. Chen. A gmm parts based face representation for improved verification through relevance adaptation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 855–861, 2004. 8

Y. M. Lui, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens, and P. J. Phillips. A meta-analysis of face recognition covariates. In *BTAS'09: Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, pages 139–146, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-5019-0. 164, 173

Luxand Face SDK. http://www.luxand.com. 83, 90

C. N. Macrae and G. V. Bodenhausen. Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51(1):93–120, 2000. 30, 35

H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. Nighttime face recognition at long distance: cross-distance and cross-spectral matching. In *11th Asian conference on Computer Vision - Volume Part II*, ACCV'12, pages 708–721, 2013. 53

D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. FVC2000: Fingerprint Verification Competition. 24(3):402–412, 2002. 49

D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. FVC2004: Third Fingerprint Verification Competition. In D. Zhang and A. K. Jain, editors, *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 1–7. Springer LNCS-3072, 2004. 48

A. Mansfield, G. Kelly, D. Chandler, and J. Kane. Biometric product testing final report. Technical report, CESG Biometrics Working Group, March 2001. (http://www.cesg.gov.uk/site/ast/biometrics/media/BiometricTestReportpt1.pdf). 48

A. Mansfield and J. Wayman. Best practices in testing and reporting performance of biometric devices. Technical report, CESG Biometrics Working Group, August 2002. (http://www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf). 48

A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET curve in assessment of decision task performance. In *Proc. of ESCA Eur. Conf. on Speech Comm. and Tech., EUROSPEECH*, pages 1895–1898, 1997. 49

J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W.-Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, 2006. 23, 29, 165, 174

A. Mayoue, B. Dorizzi, L. Allano, G. Chollet, J. Hennebert, D. Petrovska-Delacretaz, and F. Verdet. *Guide to biometric reference systems and performance evaluation*, chapter BioSecure multimodal evaluation campaign 2007 (BMEC 2007), pages 327–372. Springer, 2009. 48

MBGC. Multiple Biometric Grand Challenge. *NIST - National Institute of Standard and Technology*, http://face.nist.gov/mbgc/ . xxiii, xxiii, 66, 67, 135, 136

C. McCool and S. Marcel. Parts-based face verification using local frequency bands. In *Third International Conference on Advances in Biometrics, ICB*, pages 259–268, 2009. 8

C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel. Session variability modelling for face authentication. *IET Biometrics*, 2013. 8, 164, 174

K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. IAPR Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 72–77. Springer LNCS-2091, 1999. 51

L. C. Molina, L. Belanche, and A. Nebot. Feature selection algorithms: A survey and experimental evaluation. In *IEEE International Conference on Data Mining*, pages 306–312. IEEE Computer Society, 2002. 130

M. Moreno-Moreno, J. Fierrez, P. Tome, and J. Ortega-Garcia. An‡lisis de escenarios para reconocimiento biomŽtrico a distancia adecuados para la adquisici—n de im‡genes mmw. In *Proceedings of XXV Simposium Nacional de la Union Cientifica Internacional de Radio, URSI 2010*, September 2010. 15

M. Moreno-Moreno, J. Fierrez, P. Tome, R. Vera-Rodriguez, J. Parron, and J. Ortega-Garcia. Biogiga: Base de datos de imagenes sinteticas de personas a 94 ghz con fines biometricos. In *Proc. of XXVI Simposium Nacional de Union Cientifica Internacional de Radio, URSI 2011*, September 2011. 14

MTIT. Minutiae Template Interoperability Testing, FP6-2004-IST-4, 2009. http://www.mtitproject.com/index.html. 1, 167

R. Nagel and A. Rosenfeld. Computer detection of freehand forgeries. *IEEE Trans. on Computers*, 26(9):895–905, 1977. 47

National Institute of Standards and Technology (NIST). http://www.nist.gov. 7, 51

NBSP. National Biometric Security Project, 2009. http://www.nationalbiometric.org/. 48

Netherlands Forensic Institute (NFI). http://www.forensicinstitute.nl. 38, 62, 82, 89, 90, 92, 107, 117, 155

Next Generation Identification. http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi. 7

K. Niinuma, U. Park, and A. K. Jain. Soft biometric traits for continuous user authentication. *Transaction on Information, Forensics and Securisty*, 5(4):771–780, Dec. 2010. ISSN 1556-6013. 42

NIST: Face Challenges. http://face.nist.gov. xix, 17, 18, 31

M. S. Nixon and J. N. Carter. Automatic Recognition by Gait. *Proceedings of the IEEE*, 94(11):2013–2024, 2006. 27, 35, 44

O. Ocegueda, S. K. Shah, and I. A. Kakadiaris. Which parts of the face give out your identity? In *Proccedings of CVPR*, pages 641–648. IEEE, 2011. 88, 99, 153

J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. Gonzalez-Rodriguez. Authentication gets personal with biometrics. *IEEE Signal Processing Magazine*, 21(2):50–62, 2004. 49

J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. W. R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, and A. Savran. The multi-scenario multi-environment BioSecure multimodal database (BMDB). *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009. to appear. 51

A. J. O'Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, and H. Abdi. A video database of moving faces and people. *IEEE Transaction on Pattern Analaysis and Machine Intelligence*, 27(5):812–816, May 2005. 51

B. Paltridge. Thesis and dissertation writing: An examination of published advice and actual practice. *English for Scientific Purposes*, 21:125–143, 2002. 9

K. Pan, S. Liao, Z. Zhang, S. Li, and P. Zhang. Part-based face recognition using near infrared images. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6, 2007. 92

U. Park and A. K. Jain. Face matching and retrieval using soft biometrics. *Trans. Info. For. Sec.*, 5(3):406–415, Sept. 2010. ISSN 1556-6013. 34, 108

U. Park, R. Jillela, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, 2011. 29

U. Park, S. Pankanti, and A. K. Jain. Fingerprint verification using sift features. volume 6944, page 69440K. SPIE, 2008. 165, 174

V. Perlibakas. Distance measures for pca-based face recognition. *Pattern Recognition Letters*, 25(6):711–724, 2004. ISSN 0167-8655. 74

D. Petrovska-Delacretaz, A. Mayoue, and B. Dorizzi. *Guide to biometric reference systems and performance evaluation*, chapter The BioSecure benchmarking methodology for biometric performance evaluation, pages 11–25. Springer, 2009. 48

J. P. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000a. 1

P. Phillips, A. Martin, C. Wilson, and M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, 33(2):56–63, 2000b. 48

P. J. Phillips. Face and iris evaluations at NIST. In *CardTech/SecurTech*, May 2006. 1, 47, 167

P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer. An introduction to the good, the bad, amp; the ugly face recognition challenge problem. In *Int. Conf. on Automatic Face Gesture Recognition and Workshops (FG 2011)*, pages 346 –353, march 2011. 1, 7, 52, 134, 167

P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, Iii, and S. Weimer. Overview of the multiple biometrics grand challenge. In *Third International Conference on Advances in Biometrics (ICB'09)*, pages 705–714, Berlin, Heidelberg, 2009a. Springer-Verlag. ISBN 978-3-642-01792-6. 1, 7, 51, 53, 134, 136, 164, 167, 173

P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 947–954, Washington, DC, USA, 2005. 47, 50, 74, 164, 173

P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000c. 47, 48

P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, 2009b. 1, 134, 167

S. J. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti. Tied factor analysis for face recognition across large pose differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:970–984, 2008. ISSN 0162-8828. 164, 174

M. Przybocki and A. Martin. NIST Speaker Recognition Evaluation chronicles. In J. Ortega-Garcia *et al.*, editors, *ISCA Workshop on Speaker and Language Recognition, ODYSSEY*, pages 15–22, 2004. 1, 47, 48

P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994. 130

N. K. Ratha and V. Govindaraju, editors. *Advances in biometrics: Sensors, algorithms and systems*. Springer, 2008. 1, 167

D. Reid and M. Nixon. Human identification using facial comparative descriptions. In *IAPR International Conference on Biometrics, ICB*, June 2013. 34

K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341 –345, april 2006. 63, 93, 94, 118, 149, 150

Robust2008. Robust biometrics: Understanding science & technology. http://biometrics.cylab.cmu.edu/ROBUST2008, 2008. 109

A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer, 2006. 1, 51, 167

D. L. Ruderman, T. W. Cronin, and C. C. Chiao. Statistics of coneresponses to natural images: implications for visual coding. *Journal Optical Society Of America*, 15:2036–2045, 1998. 155

J. Sadr, I. Jarudi, and P. Sinha. The role of eyebrows in face recognition. *Perception*, 32(3):285–293, 2003. 88

S. Samangooei. *Semantic Biometrics*. PhD thesis, School of Electronics and Computer Science, University of Southampton, UK, May 2010. xxv, 35, 36, 109

C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *Third International Conference on Advances in Biometrics, ICB*, pages 199–208, 2009. 8, 165, 174

SC37, 2005. ISO/IEC JTC 1/SC 37 . (http://www.jtc1.org/sc37/). 1, 167

R. D. Seely. *On a three-dimensional gait recognition system*. PhD thesis, School of Electronics and Computer Science, University of Southampton, UK, 2010. 35

R. D. Seely, S. Samangooei, L. Middleton, J. Carter, and M. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *Proc. of IEEE Biometrics: Theory, Applications and Systems (BTAS'08)*, pages 1–6, September 2008. xxii, 35, 55, 114

T. Shiang. A statistical approach to data analysis and 3-d geometric description of the human head and face. *Proc Natl Sci Counc Repub China B*, 23(1):19–26, 1999. 92

B. Siddiquie and A. Gupta. Beyond active noun tagging: Modeling contextual interactions for multi-class active learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2979–2986, 2010. 34

S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh. A robust skin color based face detection algorithm, tamkang. *Journal of Science and Engineering*, 6:227–234, 2003. 155

Spanish Guardia Civil (DGGC). http://www.guardiacivil.es/. 38, 59, 62, 82, 89, 90, 92, 107, 117, 155

Tabula Rasa, 2010. TABULA RASA: Trusted Biometrics under Spoofing Attacks, FP7-ICT-257289. (http://www.tabularasa-euproject.org). 1, 167

S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 4th edition, 2008. ISBN 1597492728, 9781597492720. 11, 33, 36, 111, 113, 115, 125, 127, 129, 130, 140

M. Tistareli, S. Z. Li, and R. Chellappa, editors. *Handbook of Remote Biometrics for Surveillance and Security*. Springer, 2009. 1, 26, 167

M. Tistarelli. Active/space-variant object recognition. *Image and Vision Computing*, 13(3):215–226, 1995. 88

M. Tistarelli and D. Maltoni, editors. *Proc. of Second IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, 2007. IEEE Press. 1, 167

P. Tome. Reconocimiento automático de patrones de iris. Master's thesis, Universidad Autonoma de Madrid, June 2008. 14

P. Tome, L. Blazquez, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, N. Exposito, and P. Leston. Understanding the discrimination power of facial regions in forensic casework. In *International Workshop on Biometrics and Forensics*, pages 1–4, Lisboa, Portugal, April 2013a. 66, 99

P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia. Scenario-based score fusion for face recognition at a distance. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 67 – 73, June 2010a. 134

P. Tome, J. Fierrez, M. C. Fairhurst, and J. Ortega-Garcia. Acquisition scenario analysis for face recognition at a distance. In *6th International Symposium on Visual Computing (ISVC)*, pages 461–468. LNCS 6453, November-December 2010b. 7, 66, 79

P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon. Soft biometrics and their application in person recognition at a distance. *IEEE Transaction on Information Forensics and Security*, 2013b. submitted. 33, 108, 134

P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia. Combination of face regions in forensic scenarios. *Rapid Communications in Forensic Science International (FSI)*, 2013c. submitted. 33, 134

P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia. Facial soft biometrics features for forensic face verification. *Pattern Recognition*, 2013d. submitted. 108

P. Tome, J. Fierrez, R. Vera-Rodriguez, and D. Ramos. Identification using face regions: Application and assessment in forensic scenarios. *Forensic Science International (FSI)*, 233:75–83, 2013e. 29, 33, 66, 153

P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Variability compensation using NAP for unconstrained face recognition. In *Proc. 10th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS12)*, volume 151, pages 129–139. Springer, March 2012. 7, 33

P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Fusion of facial regions using color information in a forensic scenario. In *18th Iberoamerican Congress on Pattern Recognition (CIARP2013)*, LNCS 8259, pages 399–406, Habana, Cuba, November 2013f. 134

M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991a. 74, 93

M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (ICCVPR)*, pages 586–591, 1991b. 74

US-VISIT Program of the U.S. Department of Homeland Security. 31

R. Vera-Rodriguez, P. Tome, J. Fierrez, N. Exposito, and F. J. Vega. Analysis of the variability of facial landmarks in a forensic scenario. In *International Workshop on Biometrics and Forensics*, Lisboa, Portugal, April 2013a. 59, 66, 99, 118

R. Vera-Rodriguez, P. Tome, J. Fierrez, and J. Ortega-Garcia. Comparative analysis of the variability of facial landmarks for forensics using cctv images. In *6th Pacific Rim Symposium on Image and Video Technology (PSIVT)*, 2013b. 66

B. Vijaya-Kumar, S. Prabhakar, and A. Ross, editors. *Proc. of Fifth Conference on Biometric Technology for Human Identification (BTHI V)*, 2008. SPIE. 1, 167

P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2): 137–154, May 2004. 141

R. Wallace, M. McLaren, C. McCool, and S. Marcel. Inter-session variability modelling and joint factor analysis for face authentication. In *Proceedings of the 2011 International Joint Conference on Biometrics*, IJCB '11, pages 1–8, Washington, DC, USA, 2011. IEEE Computer Society. 94

H. Wang, S. Z. Li, Y. Wang, and W. Zhang. Illumination modeling and normalization for face recognition. In *AMFG '03: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, page 104, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-2010-3. 164, 173

J. Wayman, A. Jain, D. Maltoni, and D. Maio. *Biometric systems. Technology, design and performance evaluation*. Springer, 2005. 51

C. Wilson *et al.* FpVTE2003: Fingerprint Vendor Technology Evaluation 2003, June 2004. NIST Research Report NISTIR 7123 (http://fpvte.nist.gov/). 48

J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2 2009. ISSN 0162-8828. 141

Y. Yao, B. R. Abidi, N. D. Kalka, N. A. Schmid, and M. A. Abidi. Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111–125, 2008. ISSN 1077-3142. 52

D. Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First International Signature Verification Competition. In D. Zhang and A. K. Jain, editors, *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 16–22. Springer LNCS-3072, 2004. 1, 48

X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, pages 2876–2896, 2009. 7

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003. ISSN 0360-0300. 28, 35, 44, 53, 165, 174

S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):230–245, 2007. ISSN 0162-8828. 164, 173