UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES

ADAPTED FUSION SCHEMES FOR MULTIMODAL BIOMETRIC AUTHENTICATION

–TESIS DOCTORAL–

ESQUEMAS ADAPTADOS DE FUSIÓN PARA AUTENTICACIÓN BIOMÉTRICA MULTIMODAL

Author: Julián Fiérrez Aguilar
(Ingeniero de Telecomunicación, UPM)

A Thesis submitted for the degree of:

Doctor of Philosophy

Madrid, May 2006

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.

ISBN 84-689-8479-5

This Thesis was printed with the financial support from ATVS-UAM.

Department: Señales, Sistemas y Radiocomunicaciones
Esc. Técnica Superior de Ing. de Telecomunicación
Universidad Politécnica de Madrid (UPM), SPAIN

PhD Thesis: Adapted Fusion Schemes for
Multimodal Biometric Authentication

Author: **Julián Fiérrez Aguilar**
Ingeniero de Telecomunicación (UPM)

Advisor: **Javier Ortega García**
Doctor Ingeniero de Telecomunicación (UPM)
Universidad Autónoma de Madrid, SPAIN

Year: 2006

Committee: **Narciso García Santos**
Universidad Politécnica de Madrid, SPAIN

Joaquín González Rodríguez
Universidad Autónoma de Madrid, SPAIN

Luis Torres Urgell
Universidad Politécnica de Catalunya, SPAIN

Josef Bigün
Halmstad University, SWEDEN

Luis A. Hernández Gómez
Universidad Politécnica de Madrid, SPAIN

The research described in this Thesis was carried out within the Biometrics Research Lab.–ATVS at the Dept. of Ingeniería Audiovisual y Comunicaciones, Escuela Universitaria de Ingeniería Técnica de Telecomunicación, Universidad Politécnica de Madrid (from 2002 to 2004); and at the Dept. of Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid (from 2004 to 2006). The project was partially funded by a PhD scholarship from Comunidad de Madrid and Fondo Social Europeo.

The author was awarded with a PhD scholarship from Consejería de Educación de la Comunidad de Madrid and Fondo Social Europeo between 2002 and 2006 which supported the research summarized in this Dissertation.

The author has been awarded with the Best Poster Paper Award in the IAPR International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA 2003) for one publication originated from this Dissertation: J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "A Comparative Evaluation of Fusion Strategies for Multimodal Biometric Verification", Springer Lecture Notes in Computer Science, Vol. 2688, pp. 830-837, June 2003.

The author has been awarded with the Motorola Best Student Paper Award in the IAPR International Conference on Biometrics (ICB 2006) for one publication originated from this Dissertation: J. Fierrez-Aguilar, Yi Chen, Javier Ortega-Garcia, and Anil K. Jain, "Incorporating Image Quality in Multi-Algorithm Fingerprint Verification", Springer Lecture Notes in Computer Science, Vol. 3832, pp. 213-220, January 2006.

This work has been awarded with the 2004/05 Rosina Ribalta First Prize for the Best Doctoral Thesis Project within the areas of Information Technologies and Communications by the Fundación Epson Ibérica.

Abstract[1]

THIS THESIS IS FOCUSED ON the combination of multiple biometric traits for automatic person authentication, in what is called a *multimodal biometric system*. More generally, any type of biometric information can be combined in what is called a *multibiometric system*. The information sources in multibiometrics include not only multiple biometric traits but also multiple sensors, multiple biometric instances (e.g., different fingers in fingerprint verification), repeated instances, and multiple algorithms. Most of the approaches found in the literature for combining these various information sources are based on the combination of the matching scores provided by individual systems built on the different biometric evidences. The combination schemes following this architecture are typically based on combination rules or trained pattern classifiers, and most of them assume that the score level fusion function is fixed at verification time. This Thesis considers the problem of adapting the score fusion functions in multimodal biometric authentication, with application also to other multibiometric scenarios.

The term *adapted* in this Thesis refers to fusion approaches that are trained using background information, for example a pool of users, and then adjusted considering input information such as user-dependent scores or test-dependent quality measures. In this regard, the user-dependent score fusion methods found in the literature are not adapted to the users, but trained either on the pool of users or on the particular user being tested. On the other hand, the idea of adapted fusion from quality information was already embedded in some previous works, but not in an explicit way as developed in this Dissertation.

After a summary of the state-of-the-art in fusion strategies for multimodal biometrics, a number of novel adapted fusion schemes are proposed. These approaches adapt either to individual users through a reduced number of user-specific matching scores or to the input biometric quality. User-dependent fusion methods are further classified into three groups: 1) user-dependent score normalization plus simple fusion, 2) user-dependent score fusion, and 3) user-dependent decision. For most of the proposed approaches, two implementations are given, one based on statistical assumptions and the other one based on discriminative criteria using Support Vector Machines.

We then consider the issue of performance evaluation in multimodal biometric authentication systems, and introduce the experimental framework and the biometric databases used in this Dissertation. This is followed by the application of the proposed methods to competitive multi-algorithm approaches for three individual biometrics, namely: signature, voice, and fingerprint; using standard biometric data and experimental benchmarks.

The experimental part of the Thesis starts with a study of user-dependent score normalization and decision in multi-algorithm on-line signature verification. For this study we introduce two new systems based on local and global information, respectively. The local system is also used to

[1]Un resumen extenso de la Tesis en español se incluye en el Apéndice A.

study various practical aspects of system development including feature extraction and modeling, and to demonstrate the benefits of incorporating user-dependent score normalization. We finally combine the local and global systems using simple score level fusion rules, demonstrating both the complementarity of the two approaches and the benefits of incorporating user-dependent decision thresholds.

We then study the application of adapted user-dependent fusion to multi-algorithm speaker verification using third party systems. We compare user-independent, user-dependent, and adapted user-dependent versions of score level fusion. It is shown that the proposed adapted approach outperforms both user-independent and user-dependent traditional fusion schemes.

After that, we study the effects of image quality on the performance of two common approaches for fingerprint verification. It is observed that the approach based on ridge information outperforms the minutiae-based approach in low image quality conditions. This is exploited by an adapted score-level fusion approach using quality measures estimated in the frequency domain. The proposed scheme leads to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of operating points.

Finally, a comparative study of the proposed adapted fusion schemes, both user-dependent and quality-based, is given for the case of multimodal authentication based on signature and fingerprint on the real bimodal database MCYT. The proposed approaches are demonstrated to outperform traditional non-adapted fusion schemes. The experimental results favor the adapted fusion schemes based on discriminative formulations with respect to the Bayesian approaches in the case of small training set sizes. The opposite occurs for large training set sizes.

A MIS PADRES Y CUATRO HERMANAS

In a few hundred years, when the history of our time is written from a long-term perspective, it is likely that the most important event those historians will see is not technology, not the Internet, not e-commerce. It is an unprecedented change in the human condition. For the first time—literally—substantial and rapidly growing numbers of people have choices. For the first time, they will have to manage themselves. And society is totally unprepared for it.

–Peter F. Drucker, *Leader to Leader* 16, 2000.

Acknowledgements

This Thesis summarizes the work carried out during my Ph.D. studies with the Biometrics Research Lab.–ATVS since 2002. This research group was established in 1994 at the Dept. of Ingeniería Audiovisual y Comunicaciones of the Universidad Politécnica de Madrid (UPM) and since 2004 is affiliated to the Dept. of Ingeniería Informática of the Universidad Autónoma de Madrid (UAM). The work presented in this Thesis has been conducted at both institutions. The financial support for the first year of the Ph.D. studies came from research grants with Ministerio del Interior (Dirección General de la Guardia Civil, DGGC), and Telefónica Investigación y Desarrollo (TID). Subsequent years have been financially supported by a Ph.D. scholarship from Comunidad de Madrid and Fondo Social Europeo and various National and European projects referenced in the related publications.

Foremost, I would like to thank my advisor Prof. Javier Ortega-García for his guidance and support over the past four years. I really appreciate the confidence he has always shown in me. During these years I have benefited from his courage, self-mastery, and intelligent effort. In the framework of the ATVS research group I have benefited also from a close contact with Prof. Joaquín González-Rodríguez, specially from his vision, discipline, and hard work. On the other hand, I am particulary indebted with Prof. Luis Salgado, he opened his door for me to his group (Grupo de Tratamiento de Imágenes, GTI, UPM) for a challenging project with Prof. Enrique Navarro (Instituto Nacional de Educación Física, INEF, UPM) when I was pursuing my Master degree, and never closed it. The guidance, support, and lessons learned during my collaboration with him and his group clearly shaped my thinking and working attitudes. 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. Ramón García, Luis A. Hernández, and Narciso García.

Having developed the Thesis in two different institutions, I feel lucky to have met many people. From my early days at UPM I was really moved by the support given by Dr. Danilo Simón, and I really enjoyed discussions with Enrique Rendón, Luis I. Ortiz, Antonio Pedrero, José L. Sánchez, and Juan J. Gómez. At UAM I have benefited from a close contact with Profs. Javier Garrido and Juan A. Sigüenza as well as discussions with Ricardo Ribalda, Doroteo T. Toledano, Marino Tapiador, Jesús Bescós, and José M. Martínez.

During my Ph.D. studies I have had the great opportunity to visit a number of research laboratories worldwide. My first three-month stay was at Halmstad University, Sweden, with Prof. Josef Bigün. His pioneer contributions to biometric person authentication were the seed and motivation for this Thesis, and his extraordinary care as host for my visit helped me to open my mind and research perspectives in the current globalized research world. My second stay was with Prof. Davide Maltoni at University of Bologna, Italy. In this stay I really benefited from the vision and experience of Prof. Maltoni, as well as his ability to manage a fantastic group of people, including: Raffaele Cappelli, Loris Nanni, Beatrice Pasolini, Annalisa Franco, Athos Antonelli, Matteo Ferrara, and Denis Baldisserra. The third stay was with Prof. Anil K. Jain at Michigan State University, USA. His example of how to successfully manage multiple tasks and to complete what was started helped me to define and outline the present Thesis. I was really motivated by the close contact with such a distinguished researcher, and the extraordinary group of people under his master direction. In particular, I have to thank his students Stephen Krawczyk and Yi Chen, with whom I worked closely, and the others that helped me in other aspects of my stay, including Umut Uludag, Hong Chen, Karthik Nandakumar, Miguel Figueroa and family, the people from Puerto Rico, and the other research visitors from Spain: Carmen García-Mateo, Francesc F. Ferri and family. My last three-month research stay was conducted with Prof. Josef Kittler at the University of Surrey, UK, where the Dissertation was completed. I really appreciate his support in spite of his multiple commitments as director of the Center of Vision, Speech, and Signal Processing (CVSSP).

I would like also to acknowledge a number of researchers which have helped to shape the Thesis with his interesting discussions. These researchers include: Arun Ross, Norman Poh, Fabio Roli, Bao Ly-Van, Sonia García-Salicetti, Bernadette Dorizzi, Gerard Chollet, Marcos Faúndez-Zanuy, Hartwig Fronthaler, Klaus Kollreider, Jonas Richiardi, and Krzysztof Kryszczuk.

I have been enriched by the opportunity to work with a number of M.Sc. candidates in their degree projects. I have learned much from them. These extraordinary engineers are: Nuria Alonso, Gema Moreno, Alberto Posse, Jaime López, Luis M. Muñoz, Francisco del Valle, and Manuel Freire.

I would like also to thank all the work mates at ATVS with whom I have shared so many projects, deadlines, conversations and laughs. I have to thank you all: Jorge Martín, Marta Sánchez, Daniel G. Romero, Marta G. Gomar, Daniel Ramos, Alberto Montero, Fernando Alonso, Carlos Bousoño, Ignacio López, Javier González, Javier Simón, Javier Franco, Diego Rodríguez, Javier Galbally and Marcos Martínez.

Me siento así mismo realmente afortunado de haber compartido mi tiempo libre con la horda del Sanagus (los 12 de siempre) y con la peña de Tomelloso. Gracias por energizarme semana tras semana.

Por último, pero no menos importante, me gustaría dar las gracias a las personas más importantes en mi vida, mis padres y mis cuatro hermanas. Nada sería sin vuestro apoyo y ejemplo.

Julián Fiérrez Aguilar
Madrid, May 2006

Contents

List of Figures

List of Tables

Chapter 1

Introduction

THIS PHD THESIS IS FOCUSED on automatic person authentication using multiple biometric traits. In particular, this PhD Thesis explores what ancillary information is worth to be considered in person authentication by machines, devises methods to incorporate it in standard multimodal architectures, and provides results of this enhanced decision-making process.

Automatic access of persons to services is becoming increasingly important in the information era. This has resulted in the establishment of a new technological area known as biometric recognition, or simply *biometrics* [Jain *et al.*, 2004b]. 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 fingerprint, face, iris, voice, hand, or written signature. These personal traits are also commonly denoted as *biometrics*.

Although person authentication by machine has been a subject of study for more than thirty years [Atal, 1976; Kanade, 1973; Nagel and Rosenfeld, 1977], and biometric evidences have been used in forensic science for over a century [Maltoni *et al.*, 2003], it has not been until the last decade that biometric recognition has been established as an specific research area. This is evidenced by recent reference books [Jain *et al.*, 1999a; Ratha and Bolle, 2004; Wayman *et al.*, 2005; Zhang, 2002], conferences [Jain and Ratha, 2004; Kittler and Nixon, 2003; Maltoni and Jain, 2004; Zhang and Jain, 2004], common benchmark tools and evaluations [Grother *et al.*, 2003; Maio *et al.*, 2004; Phillips *et al.*, 2000b; Przybocki and Martin, 2004; Wilson *et al.*, 2004; Yeung *et al.*, 2004], cooperative international projects [BioSec, 2004; Biosecure, 2004; COST-275, 2005], international consortia [BC, 2005; EBF, 2005], standardization efforts [BioAPI, 2002; SC37, 2005], and increasing attention both from government [DoD, 2005] and industry [International Biometric Group, 2006].

In this introductory chapter we present the basics of biometric systems including common performance measures, and we outline some of the common biometric traits used in practice. 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 assumed in this chapter, the reader will benefit from introductory readings in biometrics such as Jain *et al.* [2004b].

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 fingerprints, iris, or face. 'What you produce' refers to behavioral patterns that characterize your identity such as the voice or the written signature.

The general architecture of a biometric system can be divided into two categories [Jain *et al.*, 2004b]: 1) *verification* (also referred to as *authentication* in this Thesis), and 2) *identification*. In authentication applications, the *clients* (also *users* or *targets*) are known to the system (through an *enrollment* or *training* process). In such applications, a user provides a biometric sample B (e.g., a written signature) and her claimed identity k and a one-to-one matching is performed with the stored template of the claimed user. The result of the comparison is a *similarity score* s that can be further normalized to x before comparing it to a *decision threshold*. If the score is higher than the decision threshold, then the claim is accepted, otherwise the claim is rejected. On the other hand, identification applications recognize an individual by searching over the registered clients. Identification conducts one-to-many comparisons to establish the identity of the individual.

This Thesis is focused on biometric authentication. The two modes of operation in an authentication system, i.e., enrollment and verification, are sketched in Fig. 1.1.

The objective in biometric authentication is to classify the input biometric signals into two classes, either client or impostor. Depending on the biometric verification system at hand, impostors may know information about the client that worsens verification performance when it is exploited (e.g., signature shape in signature verification). As a result, two kinds of impostors are usually considered, namely: 1) *casual impostors* (producing *random forgeries* in case of signature recognition), when no information about the target user is known, and 2) *real impostors* (producing *skilled forgeries* in case of signature recognition), when some information regarding the biometric trait being forged is used.

A biometric authentication system can commit two types of errors: 1) *False Rejection* (FR), occurring when a client is rejected by the system, and 2) *False Acceptance* (FA), taking place when an impostor is accepted as being a true user. For a given biometric system with fixed client and impostor distributions, the error rates (FRR and FAR) depend on the decision threshold. A common decision-independent performance measure is the Equal Error Rate (EER), which

[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*.

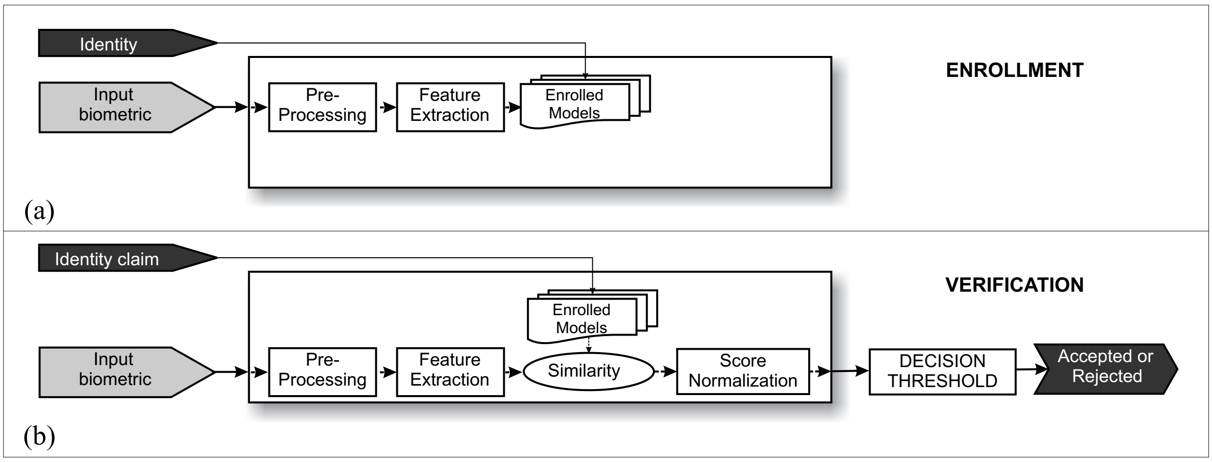


Figure 1.1: *Diagrams of the two modes of operation in a verification system: (a) enrollment, (b) verification.*

is the error rate obtained when the decision threshold is selected in order to have FRR=FAR. More details about performance evaluation in biometric systems will be given in Chapter 4.

1.2. Biometric Modalities

A number of different biometrics have been proposed and are used in various applications. Physiological biometrics include images of the ear, face, hand geometry, iris, retina, palmprint or fingerprint. Behavioral biometrics include voice, written signature, gait or keystroking. The DNA is usually not considered a biometric modality because the person identification systems based on it still require manual operation and cannot be used in (pseudo) real-time. Some of these biometrics have a long history and can be considered mature technologies, while others are still young research arenas. Contrary to the common belief, most of them, even the established traits like fingerprint [Maltoni *et al.*, 2003], are still challenging research topics [Jain *et al.*, 2004a]. Example images from various of these 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 that every person should have the biometric;
- *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 period of time;
- *collectability*, which refers to the ability to measure the biometric quantitatively.

Other criteria required for practical applications include:

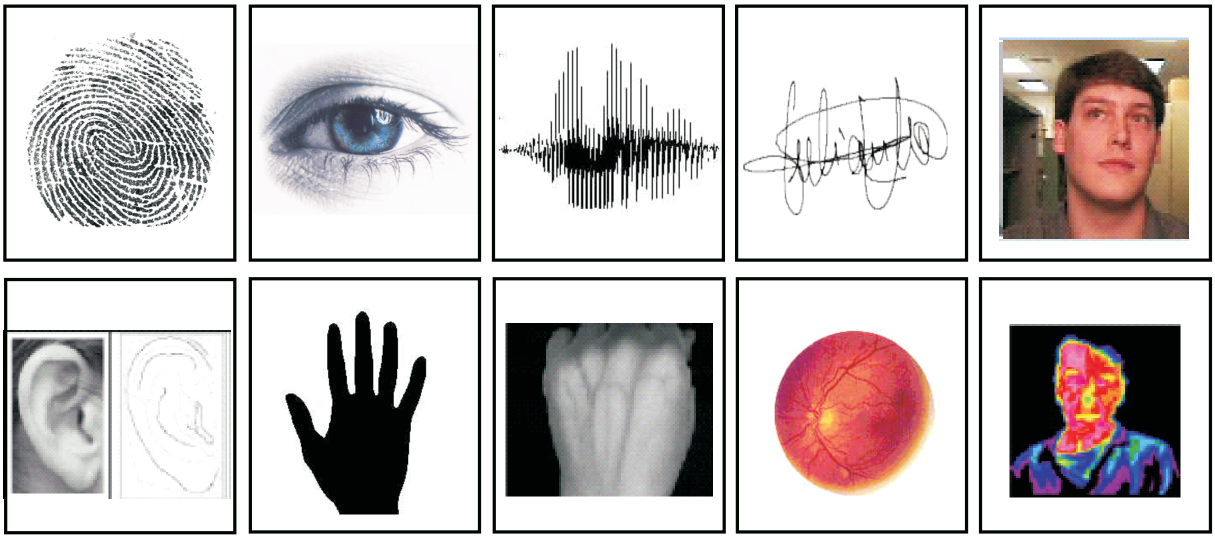


Figure 1.2: *Examples of common biometrics.*

- *performance*, which refers to the efficiency, accuracy, speed, robustness and resource requirements of particular implementations based on the biometric;

- *acceptability*, which refers to the extent to which people are willing to use the biometric and in which terms;

- *circumvention*, which reflects the difficulty to fool a system based on the biometric by fraudulent methods.

Analyzing the state-of-the-art of the different available biometrics, it can be observed that none single biometric outstands according to all criteria. Some biometrics have high distinctiveness with low collectability properties (e.g., iris with acquisition devices which are expensive and complex to use), while others may have excellent collectability but not so good distinctiveness (e.g., face). A comparison among common biometrics using the above criteria is given in Table 1.1. In this table we emphasize the last three rows, which refer to the three biometrics that will be studied in this Thesis, namely: speaker, signature, and fingerprint. Note that when considering the three of them (or only the last two, signature and fingerprint) almost all properties are well fulfilled. This last combination of signature and fingerprint can be found in important applications like electronic identification cards, e.g., the Spanish DNIe [2006].

In Fig. 1.3 we show the current market share by biometric technology according to the International Biometric Group [2006]. It can be observed that the fingerprint modality commands about half of the market, and signature and voice are the only representatives from behavioral traits. Note also the presence in the market of multiple-biometrics solutions, which were not present in previous reports from the same group.

Table 1.1: *Comparison of biometrics. High, Medium, and Low are denoted by H, M, and L, respectively. Adapted from Jain et al. [2004b].*

Biometric	**Universality**	**Distinctiveness**	**Permanence**	**Collectability**	**Performance**	**Acceptability**	**Circumvention**
Gait	M	L	L	H	L	H	M
Face	H	L	M	H	L	H	H
Hand Geometry	M	M	M	H	M	M	M
Iris	H	H	H	M	H	L	L
Speaker	*M*	*L*	*L*	*M*	*L*	*H*	*H*
Signature	*L*	*L*	*L*	*H*	*L*	*H*	*H*
Fingerprint	*M*	*H*	*H*	*M*	*H*	*M*	*M*

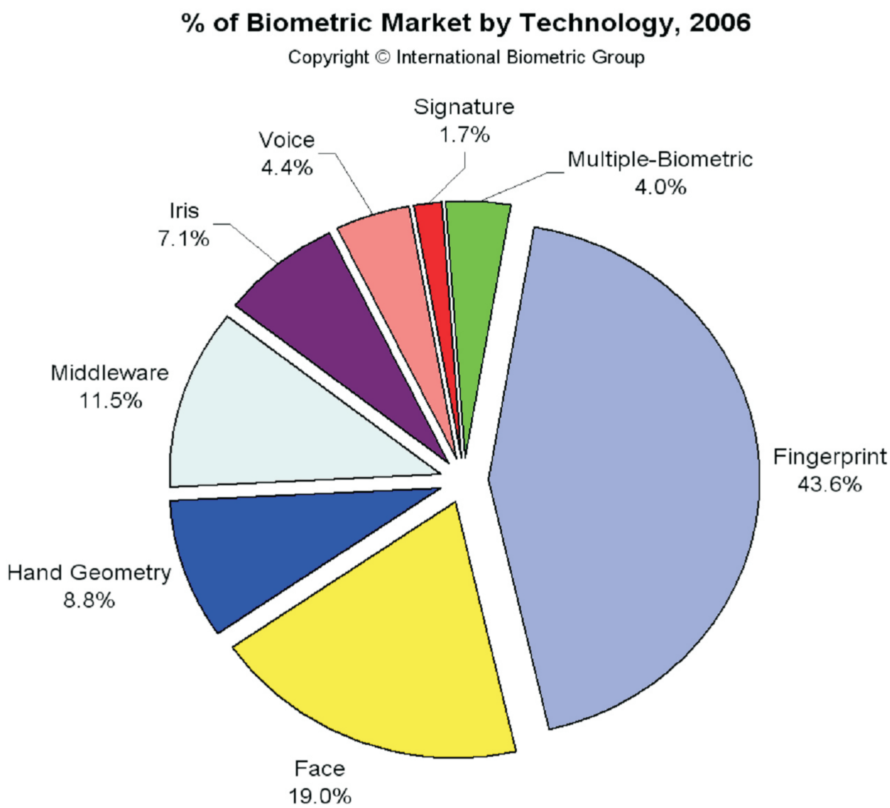


Figure 1.3: *Comparative market share by biometric technology (from International Biometric Group's Biometrics Market Report 2006-2010).*

1.3. Multimodal Biometrics and Multibiometrics

Authentication systems built upon only one biometric modality may not fulfill the requirements of demanding applications in terms of the properties described before, namely: universality, distinctiveness, permanence, collectability, performance, acceptability, and circumvention. This has motivated the current interest in *multimodal biometrics*, in which several biometric traits are simultaneously used [Jain *et al.*, 2004b]. There are a number of benefits in doing so, just to name a few: false acceptance and false rejection error rates decrease, the authentication system becomes more robust against individual sensor or subsystem failures, and the number of cases where the system is not able to make a decision is reduced significantly (e.g. bad quality fingerprints due to manual work). The technological environment is also appropriate because of the widespread deployment of multimodal devices (PDAs, 3G mobile phones, Tablet PCs, laptops, etc.).

First efforts in combining multiple biometrics for person authentication can be traced back to mid nineties [Ben-Yacoub *et al.*, 1999; Bigun *et al.*, 1997a; Brunelli and Falavigna, 1995; Chatzis *et al.*, 1999; Hong and Jain, 1998; Kittler *et al.*, 1998; Verlinde *et al.*, 2000]. In these works the common practice was to combine the matching scores obtained from the unimodal systems by using simple rules (e.g., sum or product), statistical methods, or machine learning procedures. A remarkable characteristic of the score level fusion approach is the possibility of designing structured multimodal systems by using existing unimodal recognition strategies. This multiple classifier approach has been applied not only to biometrics, but also to other pattern recognition problems, and is the source of much recent research [Jain *et al.*, 2000a; Oza *et al.*, 2005].

With respect to biometric authentication, two early theoretical frameworks for combining different machine experts are described by Bigun *et al.* [1997a] and Kittler *et al.* [1998], the former from a risk analysis perspective [Bigun, 1995], and the later from a statistical pattern recognition point of view [Duda *et al.*, 2001]. Both of them concluded (under some mild conditions that may not hold in practice) that the weighted average is a good way of conciliating the different opinions provided by the unimodal systems in the form of similarity scores. Soon after, multimodal fusion was studied as a two-class classification problem by using a number of machine learning paradigms [Ben-Yacoub *et al.*, 1999; Gutschoven and Verlinde, 2000; Verlinde *et al.*, 2000], for example: Neural Networks, Decision Trees and Support Vector Machines (SVM). After a series of experiments, Support Vector Machines outperformed the other approaches in most cases. Based on these results and the common trend in using the score fusion architecture, we set both the weighted average and the SVM-based score fusion approach as our main references for comparing the new techniques proposed in this PhD Thesis.

In the all works referenced above the term *multimodal biometrics* referred to the combination of different biometric traits, therefore *mode* refers to biometric modality. Interestingly, combining different biometric modalities is not the only way to enhance a biometric system, as there are a number of other information sources that can be combined for that purpose. Following recent practices in the process of standardization [SC37, 2005], a biometric system combining any type of biometric information will be referred to as a *multibiometric system* [Jain and Ross, 2004], and these aggregated biometric information sources will be referred to as *multibiometrics* [Ross *et al.*, 2006].

These multiple biometric information sources are originated from the fusion level and the fusion scenario considered.

1.3.1. Fusion Levels

A biometric system is usually divided into four modules (see Fig. 1.1): 1) the sensor acquires the biometric data, 2) the feature extraction module process the biometric data in order to obtain a compact yet discriminative representation of the input biometric data, 3) the matching module compares input feature vectors to stored templates resulting in matching scores, and 4) the decision module releases an identification or verification decision based on the matching scores. Information fusion can be carried out at the output of any of these four modules, resulting in the following fusion levels:

- *Sensor level fusion* refers to the combination of raw data from the biometric sensors. One example is the combination of several face images to obtain a 3D face input biometric.
- *Feature level fusion* refers to the combination of different feature vectors, obtained either with different sensors or by applying different feature extraction algorithms to the same raw data.
- *Score level fusion* refers to the combination of matching scores provided by the different systems.
- *Decision level fusion* refers to the combination of decisions already taken by the individual systems.

A graphical representation of these four fusion levels is given in Fig. 1.4.

The more common fusion level in multibiometrics is the score level, as evidenced by the publications referenced in the previous section. It can be argued that systems that integrate information at an earlier stage, such as sensor or feature level, may be more effective. In practice, score level fusion involves some advantages over these early fusion approaches, for example: 1) the information to be combined in sensor and feature level fusion can be heterogeneous and its (high dimensional) feature space structure can be unknown, 2) the information within the biometric system (either the sensed signals or the extracted features) is hidden in most commercial systems. On the other hand, the information in decision-level fusion is so limited (i.e., binary decisions) that little improvement can be expected with the fusion. As a result, the dominant option in most research works is score level fusion.

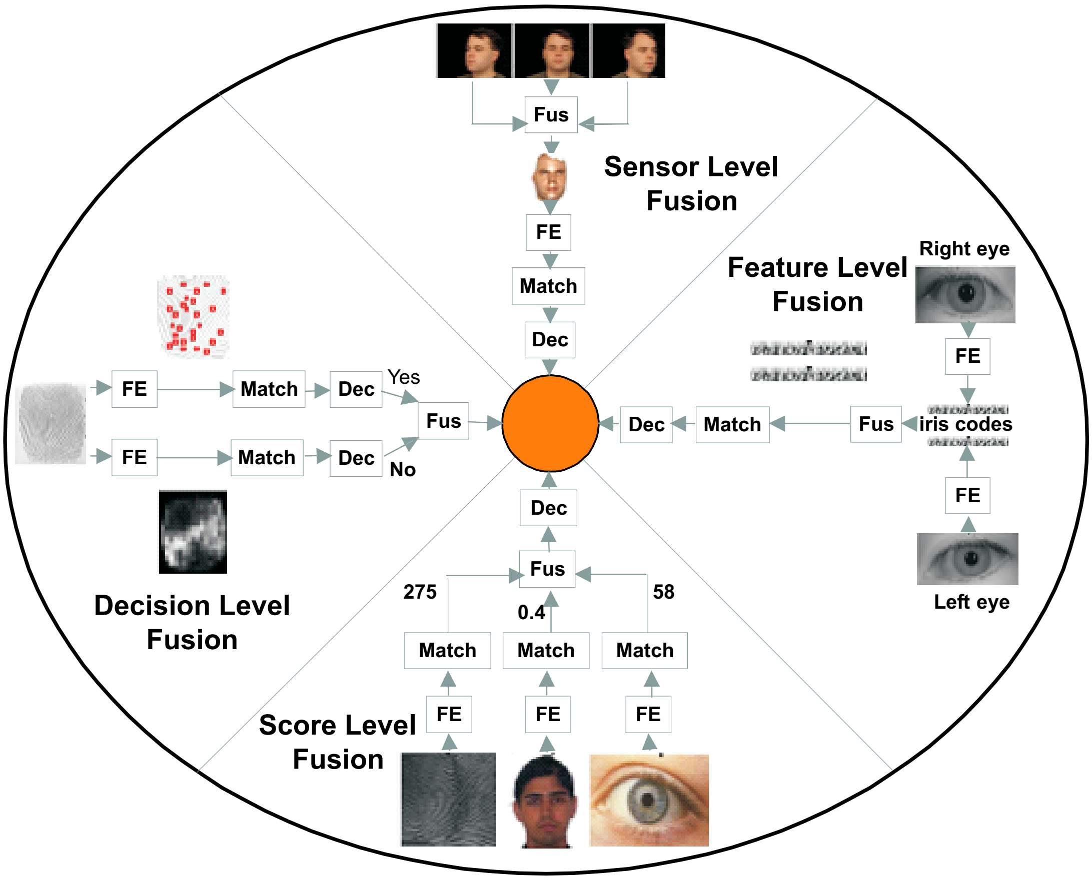


Figure 1.4: Fusion levels in multibiometrics. Adapted from Ross et al. [2006].

1.3.2. Fusion Scenarios

A multibiometric system can be based on one or a combination of the following fusion scenarios:

- *Multiple sensors.* A single biometric modality is acquired by using a number of sensors. One example is multiple face cameras for creating a 3D input face or for combining the output scores of the different baseline face images.
- *Multiple algorithms.* A single biometric input is processed with different feature extraction algorithms in order to create templates with different information content. One example is processing fingerprint images according to minutiae- and texture-based representations.
- *Multiple instances.* A single biometric modality but multiple parts of the human body are used. One example is the use of multiple fingers in fingerprint verification.
- *Repeated instances.* The same biometric modality and instance is acquired with the same sensor multiple times. One example is the sequential use of multiple impressions of the same finger in fingerprint verification. This case is sometimes not considered a multibiometric scenario [SC37, 2005].

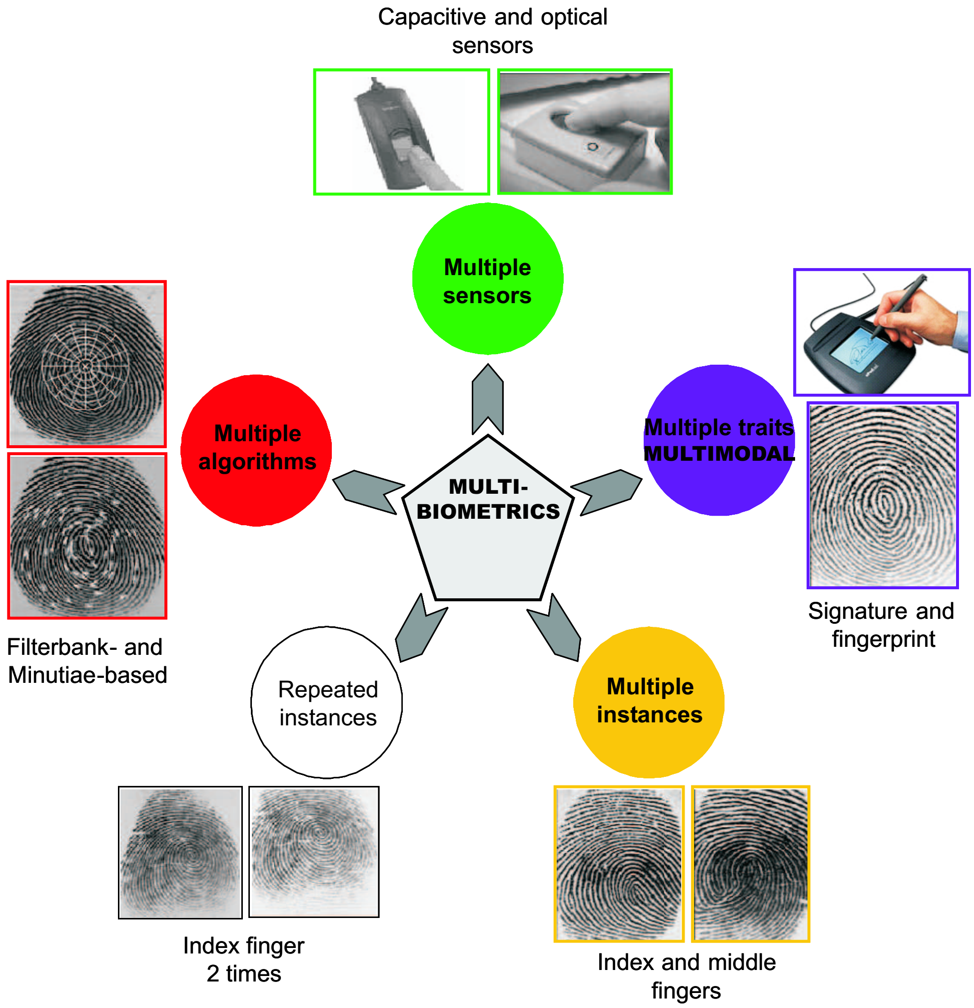


Figure 1.5: *Fusion scenarios in multibiometrics. Adapted from Ross et al. [2006].*

- *Multiple modalities.* Multiple biometric modalities are combined. This is also known as *multimodal biometrics.*

In Fig. 1.5 we illustrate the different multibiometric scenarios with some examples related to this PhD Thesis.

1.4. Motivation of the Thesis

Provided that multiple biometric modalities can overcome some of the practical limitations of current unimodal biometric technologies, this Thesis is focused on multimodal biometrics, with application also to other multibiometric scenarios. The research carried out in this area has been mainly motivated by three observations from the state-of-the-art.

The first observation comes from the seminal contribution by Doddington *et al.* [1998], where the behavior of different individual speakers was studied in the Speaker Recognition Evaluation

1998 organized by the National Institute of Standards and Technology (NIST SRE 1998). This work observed a number of different speaker behaviors in terms of verification performance; for example, some particular speakers were easily accepted by the system, whereas some others were exceptionally unsuccessful at being accepted. This fact has been dealt traditionally in some biometric systems, specially those based on behavioral biometrics, by using user-specific decision thresholds [Furui, 1981; Plamondon and Lorette, 1989]. More recently, the common approach is to apply score normalization techniques trying to map the score distributions of different users to a common domain [Bimbot *et al.*, 2004].

The second observation is strongly related to the first one. The user-dependencies found at the score level in individual systems are related to new research efforts focused on user-dependent score fusion schemes [Jain and Ross, 2002; Toh *et al.*, 2004a]. The basic aim of these approaches is to cope with the fact that some subjects are not well suited for recognition based on some traits even though these traits can be highly discriminant among other subjects. This asseveration has been corroborated experimentally in a number of works. As an example, about 4% of the population have poor quality fingerprints that cannot be easily imaged by some of the existing sensors [Jain and Ross, 2004], due to manual work or other reasons.

The third observation that has motivated this Thesis has been the effect of the input biometric quality on the verification performance of biometric systems [Junqua and Noord, 2001; Simon-Zorita *et al.*, 2003]. In particular, it is known for most unimodal systems that the worse biometric signal quality the worse the performance is. This is for example evidenced by the results of the last International Fingerprint Verification Competition [Cappelli *et al.*, 2006], where fingerprint images with lower image quality than those of previous campaigns were used. The error rates of best systems were found to be more than an order of magnitude worse than those reported in earlier competitions using more controlled data. A related observation is that in a multimodal scenario not all traits will be affected in the same way by the input biometric quality [Jain and Ross, 2004].

These three observations will be further developed in Chapter 2 in relation to the related works existing in the literature.

1.5. The Thesis

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

> *The adaptation of the fusion functions at the score level in multimodal biometric authentication can report significant verification performance improvements. Examples of input information for this adaptation include a reduced number of scores from individual users and signal quality measures of the input biometrics. This statement also applies to other problems in multibiometrics such as multi-algorithm fusion.*

The term *adapted* in this Thesis refers to fusion approaches that are trained using background information, for example a pool of users, and then adjusted considering input information such as user-dependent scores or test-dependent quality measures. In this regard, the user-dependent score fusion methods found in the literature [Jain and Ross, 2002; Toh *et al.*, 2004a], are not adapted to the users, but trained on them. No previous works have been found in the literature on adapted user-dependent fusion. On the other hand, the idea of adapted fusion from quality information was already embedded in some previous works [Chatzis *et al.*, 1999; Toh *et al.*, 2004b], but not in an explicit way as developed in this PhD Thesis.

1.6. Outline of the Dissertation

The main objectives of the PhD Thesis are as follows: 1) reviewing and studying the problem of adapting the score normalization and score fusion stages of a multimodal authentication system, in order to consider the statistics of the user at hand as well as the quality of the input biometric signals; 2) devising practical adapted schemes considering this user-dependent and quality-based information; and 3) applying the proposed techniques to common scenarios, databases and benchmarks widely available for the biometrics research community, with emphasis on signature and fingerprint verification.

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

- Chapter 1 introduces the topic of biometric systems and gives the motivation, outline and contributions of this PhD Thesis.

- Chapter 2 summarizes related works and details the motivations for this Thesis based on these previous works.

- Chapter 3 introduces the set of score level fusion schemes proposed in this Thesis. These methods are divided into two categories, namely: user-dependent fusion and quality-based fusion. User-dependent fusion methods are further classified into three groups: 1) user-dependent score normalization plus simple fusion, 2) user-dependent score fusion, and 3) user-dependent decision. For most of the proposed approaches, two implementations are given, one based on statistical assumptions and the other one based on discriminative criteria using Support Vector Machines.

- Chapter 4 considers the issue of performance evaluation in multimodal biometric systems, and introduces the biometric databases used in this Dissertation.

- Chapter 5 studies the application of user-dependent score normalization and decision to multi-algorithm written signature verification. The two systems used in this chapter are contributions of this PhD Thesis, therefore they will be presented in some detail.

- Chapter 6 studies the application of user-dependent score fusion to multi-algorithm speaker verification. In this case we use multiple speaker verification systems from a third party.
- Chapter 7 studies the application of quality-based score fusion to multi-algorithm fingerprint verification. In this case one of the two systems used is a contribution of this PhD Thesis.
- Chapter 8 conducts a comparative study of the proposed techniques, both user-dependent and quality-based fusion, to the problem of combining signature and fingerprint traits in a multimodal authentication system.
- Chapter 9 concludes the Dissertation summarizing the main results obtained and outlining future research lines.

The dependence among the chapters is illustrated in Fig. 1.6. For example, before reading Chapter 8, one should read first Chapters 5 and 7, for which one should first read Chapters 3 and 4. For this one should start with the introduction in Chapter 1 with the recommendation of reading Chapter 2 as well. The experimental chapters, which are shaded in Fig. 1.6, contain pointers to the particular methods used from previous chapters. Therefore, assuming a background in multibiometrics [Ross *et al.*, 2006], the experimental chapters can be read independently.

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 the topic [Duda *et al.*, 2001; Theodoridis and Koutroumbas, 2003]. This is specially useful for dealing with Chapters 2 and 3. Chapters 5 and 6 are based on particular methods from speech processing and speech recognition [Deller *et al.*, 1999; Quatieri, 2001]. Chapter 7 assumes a knowledge of the fundamentals of image processing [Gonzalez and Woods, 2002], and computer vision [Bigun, 2006].

1.7. Research Contributions

The research contributions of this PhD Thesis are as follows (some publications can appear in different items of the list):

Literature reviews. 1) Score fusion strategies for multimodal biometrics [Fierrez-Aguilar *et al.*, 2003a,b] (best poster paper). 2) User-dependent score normalization [Fierrez-Aguilar *et al.*, 2004c, 2005h]. 3) User-dependent score fusion [Fierrez-Aguilar *et al.*, 2005b].

Theoretical frameworks. Theoretical framework and related taxonomy for score normalization methods [Fierrez-Aguilar *et al.*, 2004c, 2005h].

Novel methods. 1) Novel methods in user-dependent score normalization [Fierrez-Aguilar *et al.*, 2005h]. 2) Novel methods in user-dependent score fusion based on Bayesian adaptation [Fierrez-Aguilar *et al.*, 2005a,c] and Support Vector Machines [Fierrez-Aguilar *et al.*,

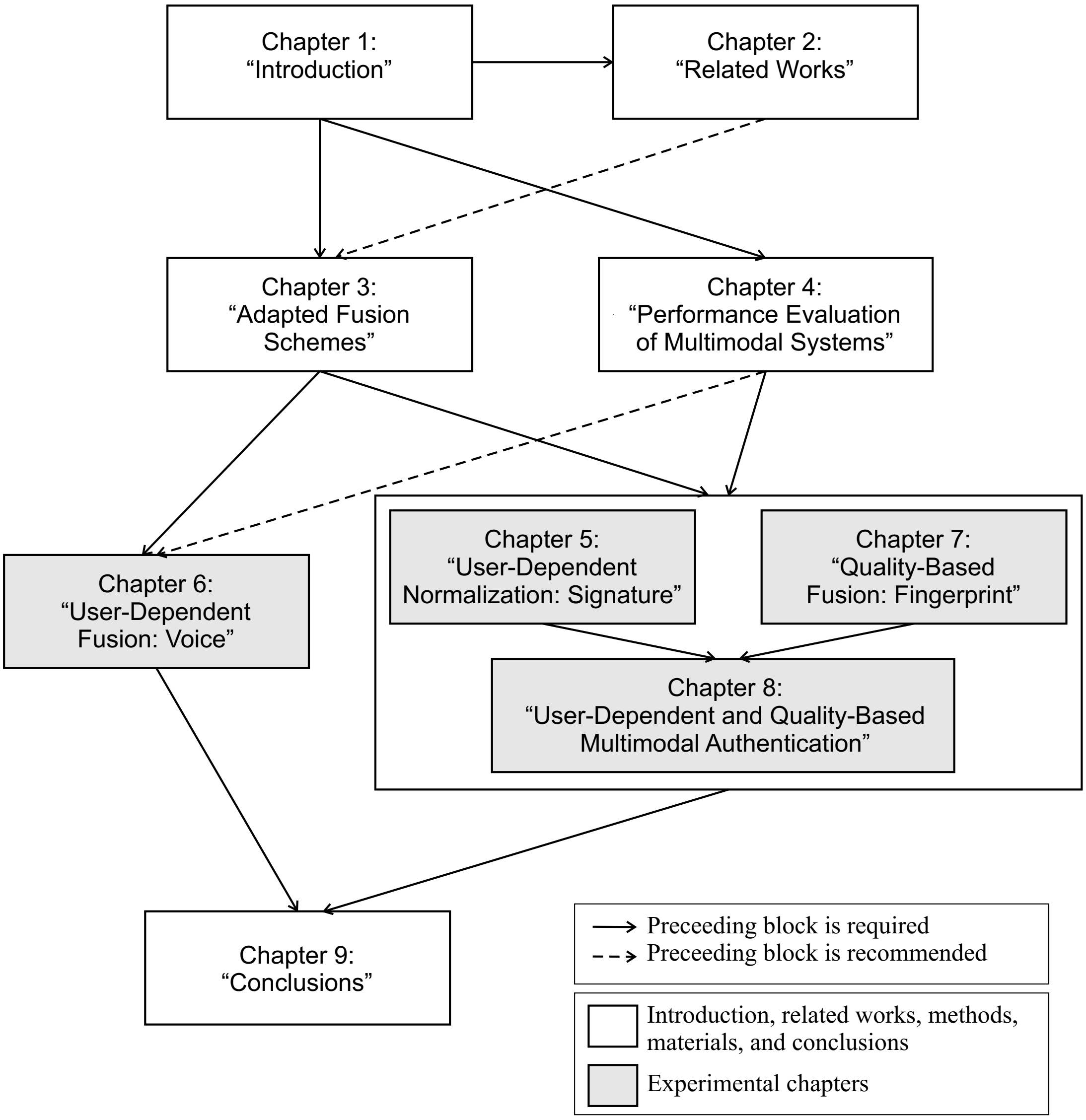


Figure 1.6: *Dependence among chapters.*

2004b, 2005b]. 3) Novel methods in quality-based score fusion based on weighted average [Fierrez-Aguilar *et al.*, 2006] (best student paper), Bayesian theory [Bigun *et al.*, 2003, 2005] (two keynote speeches related to these works at MMUA [2003] and ICIAP [2003], respectively), and Support Vector Machines [Fierrez-Aguilar *et al.*, 2004d, 2005i].

New biometric systems. 1) Two new on-line signature verification systems have been developed [Fierrez-Aguilar *et al.*, 2005f; Ortega-Garcia *et al.*, 2003a] based on previous work at the Biometrics Research Lab.–ATVS [Ortega-Garcia *et al.*, 2002]. One of the two systems was presented at the First International Signature Verification Competition, obtaining highly remarkable results [Yeung *et al.*, 2004]: 1st for random forgeries, and 2nd for skilled forgeries. 2) A new ridge-based system for fingerprint verification [Fierrez-Aguilar *et al.*, 2005e].

New biometric data. A large multimodal biometric database including fingerprint and signature modalities from 330 subjects was collected in the framework of this PhD Thesis [Ortega-Garcia *et al.*, 2003b], which is now publicly available for research purposes. It is now used in more than 30 research groups worldwide.

New experimental studies. 1) Score normalization in signature verification [Fierrez-Aguilar *et al.*, 2004c, 2005h]. 2) Multi-algorithm signature verification [Fierrez-Aguilar *et al.*, 2005f]. 3) Multi-algorithm speaker verification [Fierrez-Aguilar *et al.*, 2005a]. 4) Study of the effects of image quality (automatic assessment) on minutiae- and ridge-based fingerprint verification systems [Fierrez-Aguilar *et al.*, 2005e]. 5) Multi-algorithm fingerprint verification [Fierrez-Aguilar *et al.*, 2006] (invited speech at BQW [2006]). 6) Multimodal fusion of signature and fingerprint modalities [Fierrez-Aguilar *et al.*, 2004b, 2005b,c, 2004d, 2005i].

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

Literature reviews. Review of schemes for fingerprint image quality computation [Alonso-Fernandez *et al.*, 2005b].

Theoretical frameworks. A theoretical framework for the application of biometric evidences in forensic reporting [Gonzalez-Rodriguez *et al.*, 2005].

New methods. A test- and user-dependent fast score normalization method [Ramos-Castro *et al.*, 2006a].

New biometric systems. An off-line signature verification system (i.e., based on the images of written signatures) [Fierrez-Aguilar *et al.*, 2004a].

New biometric data. 1) A new on-line signature database of 53 subjects acquired with Tablet PC [Alonso-Fernandez *et al.*, 2005a]. 2) A new multimodal database including face, iris, fingerprint and voice modalities from 250 subjects acquired in the framework of the European FP6 Integrated Project BioSec [Fierrez-Aguilar, 2005] (invited speech at ICB [2006]). Other current efforts of the Biometrics Research Lab.–ATVS in biometric database acquisition will be detailed in Chapter 4.

New experimental studies. 1) Multi-algorithm off-line signature verification [Fierrez-Aguilar *et al.*, 2004a]. 2) Resilience of on-line signature verification to packet loss in IP networks [Richiardi *et al.*, 2004]. 3) Multi-algorithm on-line signature verification combining local and regional approaches [Fierrez-Aguilar *et al.*, 2005d]. 4) Multi-algorithm on-line signature verification in the framework of the Biosecure Network of Excellence [Garcia-Salicetti *et al.*, 2006]. 5) User-dependent score normalization in speaker verification [Garcia-Romero *et al.*, 2003b]. 6) Multi-algorithm speaker verification using Spanish conversational speech [Garcia-Romero *et al.*, 2003a]. 7) Quality-based multi-algorithm speaker verification using NIST benchmark [Garcia-Romero *et al.*, 2004, 2006]. 8) Test- and user-dependent score normalization in speaker verification [Ramos-Castro *et al.*, 2006a]. 9) Study of the effects of image quality (manual assessment) and position variability on minutiae-based fingerprint verification [Simon-Zorita *et al.*, 2003]. 10) Multi-algorithm fingerprint verification with all the systems from FVC 2004 [Fierrez-Aguilar *et al.*, 2005g]. 11) Multi-algorithm fingerprint verification in the framework of the Biosecure Network of Excellence [Alonso-Fernandez *et al.*, 2006a]. 12) Attacks to fingerprint verification systems [Galbally-Herrero *et al.*, 2006]. 13) Face verification using global representation [Cruz-Llanas *et al.*, 2003].

New biometric applications. 1) Application of biometric evidences to forensic reporting [Gonzalez-Rodriguez *et al.*, 2003, 2002, 2005; Ramos-Castro *et al.*, 2005]. 2) Application of signature verification to Tablet PC [Alonso-Fernandez *et al.*, 2005a,c, 2006b]. 3) Application of on-line written signature to cryptographic key generation [Freire-Santos *et al.*, 2006].

Chapter 2

Related Works

THIS CHAPTER summarizes previous works related to this PhD Thesis.

We start by outlining the related topic of multiple classifier combination. This is a vast and growing research area which find application to the problem studied in this Thesis. Conversely, the approaches developed in this Thesis can be applied to other classifier combination problems different to multimodal biometric authentication.

After that we will concentrate on score fusion schemes and their application to multibiometrics and multimodal biometric authentication.

Related works in multimodal biometric score fusion are divided into: 1) the traditional non-adapted score fusion, and 2) the new attempts in the literature that have motivated what we have called "adapted fusion", namely: user-dependent and quality-based fusion. Also in this case the aim is not to generate a comprehensive review of the topic but to summarize the main works closely related to this Thesis.

This chapter assumes a basic understanding of the fundamentals of pattern recognition and classification [Duda *et al.*, 2001; Jain *et al.*, 2000a; Theodoridis and Koutroumbas, 2003].

This chapter is based on the publications: Fierrez-Aguilar *et al.* [2003a,b].

2.1. Multiple Classifier Combination

The basic aim of pattern recognition is to devise automatic procedures that maximize certain criteria for the recognition problem at hand, usually related to the recognition performance. This is normally achieved by comparing different existing pattern recognition algorithms on the specific problem studied, and selecting the best of them. Worth noting, by observing the errors misclassified by the different approaches, one can observe that some recognition errors committed by the best approach can be well resolved by the inferior methods. These observations motivated the relatively recent interest in combining classifiers [Kittler *et al.*, 1998].

The topic of combining various classifiers has attracted much attention in the last years. The research progress in this topic is well summarized by the successful series of Workshops on

Multiple Classifier Systems, conducted yearly from 2000 [Kittler and Roli, 2000] until 2005 [Oza *et al.*, 2005].

This multiple classifier approach can be found with different names in the literature [Kuncheva, 2004; Kuncheva *et al.*, 2001]: classifier combination, classifier fusion, mixture of experts, committees of neural networks, consensus aggregation, expert conciliation, voting pool of classifiers, dynamic classifier selection, composite classifier design, classifier ensembles, divide-and-conquer classifiers, etc. The differences between these approaches stem mainly from: assumption about classifier dependencies, type of classifier outputs, aggregation procedure, and architecture.

Classifier Dependencies. In general, we may have different classifier outputs because of [Jain *et al.*, 2000a]: different feature sets, different training sets, different classification methods, different parameters in the classification method, or different training sessions. All these reasons result in a set of classifiers whose outputs may be combined with the hope of improving the overall classification accuracy. Classifier combination is specially useful if the individual classifiers are largely independent, which not always occurs. If this has not been guaranteed by the use of different training sets, resampling techniques like rotation or bootstrap may be used to artificially create such differences. Examples of classifier combination based on resampling strategies are the well known stacking [Wolpert, 1992], bagging [Breiman, 1996], and boosting [Shapire, 1990].

In the case of multimodal biometric authentication, the independence between classifiers (one for each modality) is normally assumed.

Type of Classifier Outputs. The outputs of the different classifiers can be classified into three levels [Xu *et al.*, 1992]: 1) abstract, 2) rank, and 3) measurement (or confidence). At abstract level, each classifier only outputs a class label. At rank level, each classifier outputs a ranked list of classes, with the class ranked first being the first choice. At measurement level, each classifier outputs a numerical value indicating the belief or probability that the pattern belongs to a given class.

Aggregation Procedures. Aggregation procedures can be first classified according to trainability and adaptivity. Some combiners do not require training while others are trainable. The trained combiners may lead to better performance at the cost of additional training data and additional training. Some combiners are adaptive in the sense of weighting the contribution of each expert depending on the input pattern. Conversely, nonadaptive combiners consider all input patterns in the same way. Adaptive schemes can exploit the detailed error characteristics of the individual classifiers under different input patterns. Examples of adaptive combination strategies include adaptive weighting [Tresp and Taniguchi, 1995], mixture of local experts (MLE) [Jacobs *et al.*, 1991], and hierarchical MLE [Jordan and Jacobs, 1994].

Architecture. The schemes for multiple classifier combination can also be grouped according to their architecture into three main categories [Jain *et al.*, 2000a]: 1) hierarchical (or tree-like),

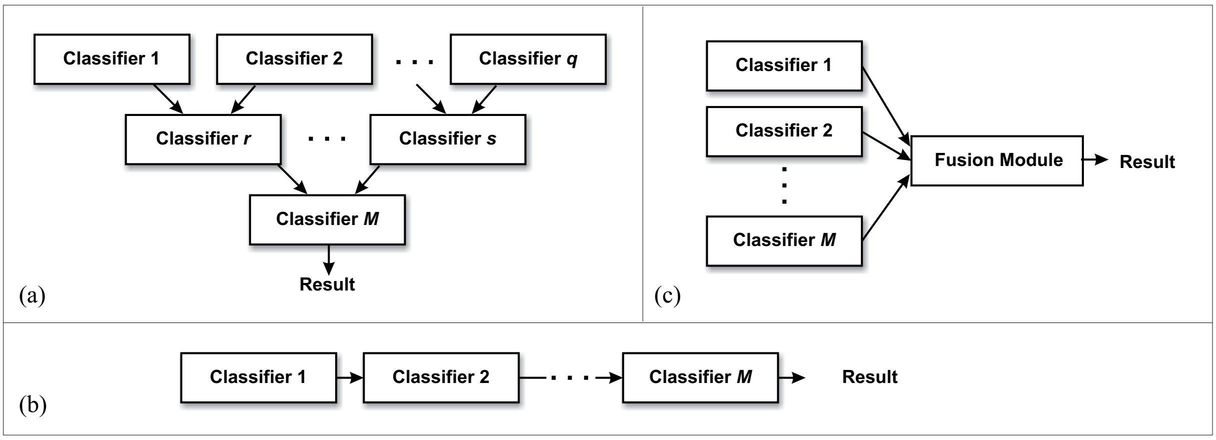


Figure 2.1: *Architectures for multiple classifier combination: (a) hierarchical, (b) serial, (c) parallel.*

2) cascading (or serial), and 3) parallel. A graphic representation of the three categories is given in Fig. 2.1.

In hierarchical classifier combination schemes, the different classifiers are combined into a tree-like structure. This is the more flexible architecture and enables to exploit the different discriminative power that can be embedded in different groups of features.

In the cascade architecture the classifiers are invoked in sequence. Some of them may only be used if certain conditions occur in the outputs of the classifiers invoked first. This architecture enables to improve the efficiency when cheap but inaccurate classifiers are followed by expensive but accurate classifiers.

In the parallel architecture all classifiers are invoked independently and their outputs are combined. Most methods in the literature belong to this category, which can be further divided into two classes: 1) selection, and 2) fusion. In classifier selection, the different individual systems are considered "experts" in local regions of the feature space. The combination gives then more importance to the classifier closest to the input pattern in terms of area of expertise [Alpaydin and Jordan, 1996; Jacobs *et al.*, 1991]. On the other hand, classifier fusion assumes that all the classifiers are trained and their expertise combined over the whole feature space [Xu *et al.*, 1992].

Some well-known combination strategies in multiple classifier systems are compared in Table 2.1 based on the previous properties.

The adapted score fusion approaches for multimodal biometrics developed in this PhD Thesis can be interpreted as trained adaptive parallel classifier fusion methods at the measurement level.

2.1.1. Approaches to Parallel Classifier Fusion

Multiple classifier outputs are usually made comparable by mapping them to the $[0, 1]$ interval. This score normalization step will be detailed in the case of multimodal authentication in Sect. 2.2.2.3. For some classifiers, these normalized output scores can be considered *a posteriori* probabilities for the classes. Assuming further restrictions, e.g., that the individual classifiers

Table 2.1: *Strategies in multiple classifier systems. Adapted from Maltoni et al. [2003].*

Method	Architecture	Level	Train.	Adapt.	Comments
Class set reduction	Serial/Parallel	Rank/Conf.	Yes	No	Efficient
Voting, AND/OR	Parallel	Abstract	No	No	Assumes independency
Associative switch	Parallel	Abstract	Yes	Yes	Explores local expertise
Borda count	Parallel	Rank	Yes	No	Converts ranks to confidences
Logistic regression	Parallel	Rank/Conf.	Yes	No	Converts ranks to confidences
Dempster-Shafer	Parallel	Rank/Conf.	Yes	No	Fuses non-probabilistic scores
Prod, min, max	Parallel	Confidence	No	No	Assumes independency
Sum, median	Parallel	Confidence	No	No	Assumes independency; robust
Gen. Ensemble	Parallel	Confidence	Yes	No	Considers error correlations
Stacking	Parallel	Confidence	Yes	No	Exploits scarcity in data
Fuzzy Integrals	Parallel	Confidence	Yes	No	Fuses non-probabilistic scores
Bagging	Parallel	Confidence	Yes	No	Needs many classifiers
Random subspace	Parallel	Confidence	Yes	No	Needs many classifiers
Adaptive weighting	Parallel	Confidence	Yes	Yes	Explores local expertise
MLE	Parallel	Confidence	Yes	Yes	Explores local expertise
Boosting	Parallel/Hier.	Abstract	Yes	No	Needs many classifiers
Neural tree	Hierarchical	Confidence	Yes	No	Handles many classes
Hierarchical MLE	Hierarchical	Confidence	Yes	Yes	Explores local expertise

use mutually independent subsets of features (which is realistic in the case of multimodal biometrics), fusion can be reduced to simple operators such as product or average. Kittler *et al.* [1998] followed this approach in a probabilistic Bayesian framework and provided an example of multimodal biometric authentication fusing speech, frontal and profile images modalities. Considering M classifiers, C classes, and a given pattern Z that generates the feature vector B_j for classifier j, the classifiers are considered to give the *a posteriori* probability for each class ω_c, $c = 1, \dots, C$: $P(\omega_c|B_j)$. Several ways to implement the fusion of the classifiers are then obtained based on the Bayes theorem and certain hypothesis:

Product Rule. Assign $Z \rightarrow \omega_c$ if

$$P^{(1-M)}(\omega_c) \prod_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \left[P^{(1-M)}(\omega_r) \prod_{j=1}^{M} P(\omega_r|B_j) \right]. \tag{2.1}$$

Sum Rule. Assign $Z \rightarrow \omega_c$ if

$$(1-M)P(\omega_c) + \sum_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \left[(1-M)P(\omega_r) + \sum_{j=1}^{M} P(\omega_r|B_j) \right]. \tag{2.2}$$

Max Rule. Assign $Z \to \omega_c$ if

$$\max_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \max_{j=1}^{M} P(\omega_r|B_j). \tag{2.3}$$

Min Rule. Assign $Z \to \omega_c$ if

$$\min_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \min_{j=1}^{M} P(\omega_r|B_j). \tag{2.4}$$

Median Rule. Assign $Z \to \omega_c$ if

$$\operatorname*{med}_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \operatorname*{med}_{j=1}^{M} P(\omega_r|B_j). \tag{2.5}$$

Majority Vote Rule. In this case the combination is not at score level but at decision level. The *a posteriori* probabilities are thresholded to produce

$$\Delta_{rj} = \begin{cases} 1 & \text{if} \quad P(\omega_r|B_j) = \max_{c=1}^{C} P(\omega_c|B_j) \\ 0 & \text{otherwise} \end{cases}. \tag{2.6}$$

The majority vote rule then assigns $Z \to \omega_c$ if

$$\sum_{j=1}^{M} \Delta_{cj} = \max_{r=1}^{C} \sum_{j=1}^{M} \Delta_{rj}. \tag{2.7}$$

Product rule is obtained from the assumption of statistical independence of the different representations B_j with $j = 1, \ldots, M$. Sum rule is obtained further assuming that the *a posteriori* probabilities computed by the classifiers do not deviate much from the *a priori* probabilities, which is the case in a noisy scenario. The remaining rules are obtained by approximating or bounding the *a posteriori* probabilities. The sum rule outperformed the remainder in the experimental comparison. This was explained by a theoretical analysis of its robustness to the estimation errors of $P(c|B_j)$ made by the individual classifiers. Subsequent works have also reported comparative studies between these simple fusion rules [Alkoot and Kittler, 2000; Kittler and Alkoot, 2003; Kuncheva, 2002].

Another paradigm for parallel classifier fusion is based on considering the combination stage as a second-level pattern recognition problem [Duda *et al.*, 2001]. In this case the outputs from the different classifiers are considered as a new feature vector which is the input to a second-level classifier. The methods specially developed for multiple classifier combination (some of them summarized in Table 2.1), can therefore be extended with any of the large number of classifiers available from the literature.

2.1.2. Theoretical Underpinnings in Multiple Classifier Combination

A large number of experimental studies have demonstrated the benefits of classifier combination [Jain *et al.*, 2000a]. However, very few works have provided some insight into the theoretical explanations.

One preliminary yet rigorous theory for classifier combination was developed by Kleinberg [1990]. Another theoretical analysis of classifier combination was presented by Krogh and Vedelsby [1995], which is based on the well-known bias/variance dilemma [Geman *et al.*, 1992]. Theoretical developments in multiple classifiers systems under severe restrictions usually assume linearly combined classifiers [Fumera and Roli, 2005; Tumer and Ghosh, 1996]. Another more general theoretical framework was presented by Kittler *et al.* [1998], who concluded that the weighted average combination is the most robust technique among the non-trained fusion rules evaluated. This result is also corroborated by the theoretical explanation by Shapire *et al.* [1998] for the effectiveness of the weighted average.

In the particular case of score fusion for biometric authentication, one of the very few works providing some theoretical insight was described by Poh and Bengio [2005b]. This study assumed Gaussian distributions of client and impostor scores and used a theoretical model called Variance Reduction-Equal Error rate (VR-EER). A number of findings linking the correlation and variance of base experts to the performance improvement of score fusion were then obtained.

Although the existence of these theoretical underpinnings, and the success of practical algorithms for classifier fusion, the problem of classifier combination is very complex and most aspects of a general theory still beg explanation [Kittler *et al.*, 1998]. Some of these not well known aspects include: relation between dimensionality expansion (multiple experts) and dimensionality reduction (expert combination), effect of individual expert error distribution on the choice of a combination strategy, etc. Furthermore, a number of practical multiple classifier approaches are either sequential or based on special rules for handling exceptions and rejections, which makes difficult the theoretical advance in this field.

2.2. Non-Adapted Fusion in Multimodal Biometrics

Multimodal biometric authentication can be seen as a two-class (either client or impostor) multiple classifier combination problem. As such, most of the categories presented in Sect. 2.1 for general multiple classifier systems also apply here with some specificities. In particular, a biometric system is usually divided into four modules: 1) the sensor acquires the biometric data, 2) the feature extraction module process the biometric data in order to obtain a compact yet discriminative representation of the input biometric data, 3) the matching module compares input feature vectors to stored templates resulting in matching scores, and 4) the decision module releases an identification or verification decision based on the matching scores. Considering this architecture of biometric systems based on four modules, we adhere to the taxonomy described by Jain *et al.* [2005] to outline the state-of-the-art in multimodal biometric fusion. This taxonomy is sketched in Fig. 2.2.

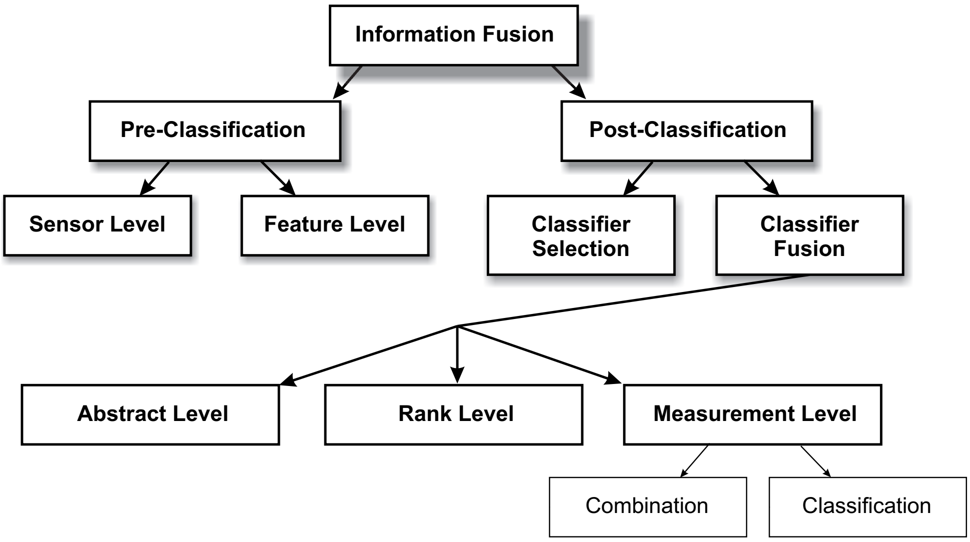


Figure 2.2: *Approaches to information fusion in multimodal biometric authentication. Adapted from Jain et al. [2005].*

2.2.1. Pre-Classification Fusion

Before classification/matching, integration of information can be done either at sensor level or feature level.

In *sensor level fusion*, raw data from the sensors are combined. One example is the combination of several cameras in face verification [Chang *et al.*, 2005].

Feature level fusion refers to the combination of different feature vectors, obtained either with different sensors or by applying different feature extraction algorithms to the same data. Two simple feature fusion schemes are: 1) weighting, when the feature vectors are homogeneous, and 2) concatenation, when the feature vectors are non-homogeneous. Experiments for homogeneous and non-homogeneous feature level fusion with face and hand modalities were reported by Ross and Govindarajan [2005].

2.2.2. Post-Classification Fusion

Approaches for combining information after the matching can be divided into classifier selection and classifier fusion. In the first category, the result is based only on the classifier most likely to give the correct decision for the input pattern. Classifier fusion can be further divided depending on the information to be combined: decisions, ranks, or matching scores.

Abstract or *decision level fusion* refers to the combination of decisions already taken by the individual biometric systems. Examples include: majority voting, weighted voting based on Dempster-Shafer theory [Xu *et al.*, 1992], AND rule, OR rule, etc.

Rank level fusion take place when the individual systems provide a set of possible matches ranked according to confidence. This approach only makes sense in biometric identification, where a number of comparisons between each input pattern and the stored templates in the database are carried out. One example of rank level fusion is Borda count, which is based on the sum of ranks provided by the individual classifiers [Ho *et al.*, 1994].

Score level fusion, also denoted as *measurement* or *confidence level fusion*, refers to the combination of matching scores provided by the different classifiers. In the context of biometric authentication, score level fusion can be classified into two categories: combination and classification. In the combination approach the input matching scores are normalized into the same range and then combined to obtain a scalar fused score. In the classification approach the matching scores are considered as input features for a second-level pattern classification problem between two classes, either client or impostor.

2.2.2.1. Combination Approach

Combination approaches include: product, sum, max, min, median, and majority vote rules as described in Eqs. (2.1) to (2.7).

In the case of multimodal biometric authentication there are only two classes ($C = 2$): ω_0 = impostor and ω_1 = client. Let us also assume that the output similarity matching scores s_j from each system $j = 1, \ldots, M$ are normalized into x_j in order to have $x_j \approx P(\omega_1|B_j)$. The *a posteriori* probabilities for the impostor class are then $P(\omega_0|B_j) = 1 - x_j$. Under these common assumptions in multimodal biometric authentication, the classification rules in Eqs. (2.1) to (2.7) are simplified significantly. As an example, the sum rule in Eq. (2.2) is based on the evaluation of

$$(1-M)P(\omega_1) + \sum_{j=1}^{M} x_j > (1-M)P(\omega_0) + \sum_{j=1}^{M} (1-x_j), \tag{2.8}$$

which is equivalent to evaluating

$$y = \sum_{j=1}^{M} x_j > \frac{(1-M)P(\omega_0) - (1-M)P(\omega_1) + M}{2} = \text{Decision Threshold}. \tag{2.9}$$

This last result indicates that the general sum rule for combining classifiers reduces to simple matching score sum plus a decision based on a threshold. This decision threshold depends on the number of systems M and the *a priori* probabilities of client and impostor classes. The remaining rules can be similarly demonstrated to reduce to simple product, max, min, and median of matching scores plus a decision threshold. Variants including weighting parameters for each system can be also found in the literature [Jain and Ross, 2002]

$$\sum_{j=1}^{M} w_j x_j > \text{Decision Threshold}. \tag{2.10}$$

The parameters w_j can be computed heuristically, by exhaustive search in order to minimize certain error criterion on a training set, or by using a trained approach based on linear classifiers.

The previous rules assume that the output matching scores from the individual systems s_j have been mapped to *a posteriori* probabilities x_j, which by no means is a straightforward task and in most cases is not realistic. This issue is considered in more detail in Sect. 2.2.2.3.

Another theoretical framework which does not rely on the assumption of posterior probabilities released by the individual systems was developed by Bigun *et al.* [1997a]. This work used Bayesian statistics to estimate the accuracy of individual classifiers during the fusion process. As this work is closely related to one of the methods proposed in this PhD Thesis, it will be described in more detail in Chapter 3. In brief, this Expert Conciliation approach results in a combination function based on weighted average of similarity scores x_j

$$y = \begin{cases} \sum_{j=1}^{M} w_j^{\mathcal{C}} x_j + w_0^{\mathcal{C}} & \text{if} \quad \left| 1 - \sum_{j=1}^{M} w_j^{\mathcal{C}} x_j + w_0^{\mathcal{C}} \right| < \left| \sum_{j=1}^{M} w_j^{\mathcal{I}} x_j + w_0^{\mathcal{I}} \right| \\ \sum_{j=1}^{M} w_j^{\mathcal{I}} x_j + w_0^{\mathcal{I}} & \text{otherwise} \end{cases} , \tag{2.11}$$

where the superindexes $\mathcal{C}$ and $\mathcal{I}$ denote parameters computed over a training set of client and impostor scores, respectively. Because this method is not built on the assumption of scores matching *a posteriori* probabilities, this combination approach does not rely so heavily on score normalization as the simple rules mentioned before.

Note that the combination approaches mentioned in this section are either fixed or trained. Simple rules such as product, sum, or max are fixed, although they rely on score normalization which may be subject to training. On the other hand, the Expert Conciliation scheme in Eq. (2.11) is a trained fusion approach.

As in every pattern recognition problem, the success of fixed rules depends heavily on the prior assumptions. On the other hand, the success of trained approaches relies heavily on the amount and representativeness of the training data. This tradeoff can be used to explain the contradictory results obtained in a number of works when comparing fixed to trained approaches, [Roli *et al.*, 2002a; Ross and Jain, 2003]. In general, the success of a trained fusion scheme will depend on the conditions of the problem at hand including the prior information and the amount of training data [Duin, 2002].

2.2.2.2. Classification Approach

In this category of methods, the normalized matching scores x_j, $j = 1, \ldots, M$ are joined together in a feature vector $[x_1, \ldots, x_M]^T$, which is the input to a two-class pattern classification problem, either client or impostor. Although some classification methods may work better when the input features are in the same range, the classification approach to fusion does not necessarily rely on score normalization, so we can assume either $x_j = s_j$ or a basic fixed score normalization just to make homogeneous the score ranges between different systems.

One early study using the classification approach in multimodal biometrics was reported by Brunelli and Falavigna [1995]. This pioneer work combined face (3 classifiers) and voice (2 classifiers) by using various forms of rank and measurement level fusion, including a Neural Network.

Chatzis *et al.* [1999] combined in different ways five different unimodal experts, four for face and one for speech authentication. Experiments were performed by considering repeatedly each person as an impostor and the remaining persons as clients for every shot, with four shots per person. Fusion methods used were the following: OR and AND logical operators on thresholded scores, k-means algorithm, fuzzy k-means algorithm, fuzzy vector quantization algorithm, fuzzy k-means for fuzzy data, fuzzy vector quantization for fuzzy data, and median radial basis function network. For algorithms which operate on fuzzy data, data was fuzzified by quality measures of experts' opinions. This is one of the first published works that used quality measures in the framework of multimodal biometric fusion.

Verlinde *et al.* [2000] followed the classification approach to fusion and compared a number of pattern classification techniques combining face profile, frontal face, and voice. The results sorted by relative decreasing performance were the following: Logistic Regression, Maximum a Posteriori, k-Nearest Neighbors, Multilayer Perceptrons, Binary Decision Trees, Maximum Likelihood, Quadratic Classifiers and Linear Classifiers. In a subsequent contribution [Gutschoven and Verlinde, 2000], the paradigm of Support Vector Machines (SVM) was compared with all the above-mentioned techniques on the same experimental scenario, outperforming all of them. This is corroborated by other comparative studies [Ben-Yacoub *et al.*, 1999], which favored the SVM approach over Neural Networks and Decision Trees. The comparisons were only based on recognition error rates. Therefore the comparative results should be taken with care, as other important factors may be considered in practical implementations, namely: ease of training, ease of implementation, scalability, etc.

Bengio *et al.* [2002] performed fusion of two experts, face verification based on Neural Networks and voice verification based on Gaussian Mixture Models by using three different fusion algorithms: Multi-Layer Perceptrons (MLP), Support Vector Machines (SVM) and Bayes Classifiers using Gaussian Mixture Models (GMM) as density estimators. They compared the performance of each of these methods with and without estimation of confidence of unimodal scores. Intuitively, knowledge of confidence measures on these scores should help in the weighting process, i.e., if one multimodal system produces scores not very precisely, its score should be given less weight. Thus, they proposed and compared three methods to estimate a measure of confidence over a score. The first method is based on Gaussian hypothesis of the score distribution. The second method estimates the confidence by using a resampling technique based on groups of training scores. The third method is based on the adequacy of the trained models to explain the input biometric data. The conclusion of this study is that some confidence measures were able to enhance the fusion performance, but not systematically. In this study the confidence measures were obtained directly either from the available training scores or from parameters of the trained models, and not from the quality of the input biometric signals.

Roli *et al.* [2002b] estimated the performance of classifier ensembles consisting of two to eight different experts. Experts' opinions were combined by using five fixed and two trained fusion

Table 2.2: *Summary of works on multimodal biometrics. M denotes the total number of classifiers combined. Architecture is either Serial or Parallel. Level is either Rank or Confidence. Performance gain over the best single classifier is given for IDentification or VERification either as FR@FA pair, EER or Total Error TE=FR+FA (in %). Adapted from Maltoni et al. [2003].*

Work	Modalities	M	Arch.	Level	Gain
Brunelli and Falavigna [1995]	Speaker, face	5	P	C	ID:17→2 (TE)
Duc *et al.* [1997]	Speaker, face	2	P	C	VER:6.7→0.5 (TE)
Kittler *et al.* [1998]	Speaker, face	3	P	C	VER:1.4→0.7 (EER)
Hong and Jain [1998]	Face, fingerprint	2	S	R/C	ID:6.9→4.5 (FR@0.1%FA)
Jain *et al.* [1999b]	Speaker, face, finger	3	P	C	VER:15→3 (FR@0.1%FA)
Ben-Yacoub *et al.* [1999]	Speaker, face	3	P	C	VER:4→0.5 (EER)
Choudhury *et al.* [1999]	Speaker, face	3	P	C	ID:16.5→6.5 (TE)
Chatzis *et al.* [1999]	Speaker, face	4	P	C	ID:6.7→1.07 (TE)
Verlinde *et al.* [2000]	Speaker, face	3	P	C	VER:3.7→0.1 (TE)
Ross and Jain [2003]	Face, finger, hand	3	P	C	VER:16→2 (FR@0.1%FA)
Kumar and Zhang [2003]	Face, palmprint	2	P	C	VER:3.6→0.8 (EER)
Wang *et al.* [2004]	Speaker, finger	2	P	C	VER:2→0.7 (EER)
Poh and Bengio [2006]	Speaker, face	8	P	C	VER:2.2→0.7 (TE)

rules. Fixed rules included: sum, majority vote, and order statistics operators such as min, med and max. Trained rules included: weighted average, and Behavior Knowledge Space method. They concluded that it is better to combine the most complementary experts rather than the best performing ones. They also concluded that, in real applications, the poor quality and/or the limited size of the training set "can quickly cancel the theoretical advantages of trained rules". Among fixed rules, the vote majority rule exhibited good performance.

Ross and Jain [2003] compared the performance of weighted sum, Decision Tree and Linear Discriminant Classifier for the fusion of face, fingerprint and hand geometry modalities. By using simple fixed score normalization, sum rule outperformed both Decision Tree and Linear Discriminant Classifiers.

In Table 2.2 we summarize some of the approaches for multimodal fusion found in the literature.

2.2.2.3. Score Normalization

In general, the similarity matching scores s_j can be modelled as [Jain *et al.*, 2005]

$$s_j = f[P(\omega_1|B_j)] + \eta(B_j), \tag{2.12}$$

where f is a monotonic function and η is the error made in the estimation of the *a posteriori* probability by the individual system j. This error can be due to noise in the input biometric signals or errors in the feature extraction or matching.

A number of works have focused on mapping output similarity scores s_j to *a posteriori* probabilities $P(\omega_1|B_j)$ by using different assumptions. Most of them assume $\eta(B_j) = 0$ in Eq. (2.12) and particular distributions for the similarity scores. Snelick *et al.* [2005] assumed the conditional densities $P(s_j|\omega_0)$ and $P(s_j|\omega_1)$ to be Gaussian. A more general assumption was developed by Prabhakar and Jain [2002] by using non-parametric density estimation based on Parzen Windows.

Either because of the unrealistic assumptions, or because of problems with density estimation on a finite training set, the prevalent method in the combination approach is not to map scores to probabilities but just to transform them into a common domain by using an operational technique for *score normalization* [Jain *et al.*, 2005]. These techniques can be either *fixed* or *adaptive*. The topic of adaptive score normalization will be further detailed in Chapter 3. Here we summarize the most common techniques for fixed score normalization:

Min-max. The matching scores s are normalized according to

$$x = \frac{s - \min}{\max - \min}, \tag{2.13}$$

where the maximum and minimum are computed from a given set of training scores. This normalization method is specially prone to errors due to outliers.

Z-score. The matching scores s are normalized with

$$x = \frac{s - \mu}{\sigma}, \tag{2.14}$$

where μ and σ are respectively the arithmetic mean and standard deviation of a given set of training scores.

Exponential functions. This include various forms of exponentials [Fierrez-Aguilar *et al.*, 2005b]

$$x = c_1 \exp(c_2 s) + c_3, \tag{2.15}$$

sigmoids [Cappelli *et al.*, 2002a; Snelick *et al.*, 2005]

$$x = \frac{c_1}{1 + \exp(c_2 s + c_3)} + c_4, \tag{2.16}$$

or hyperbolic functions [Jain *et al.*, 2005]

$$x = c_1 \tanh(c_2 s + c_3) + c_4, \tag{2.17}$$

where c_1 to c_4 are parameters. As demonstrated by Jain *et al.* [2005], exponential-based score normalization is more robust and efficient than min-max and z-score, where robust refers to insensitivity to the presence of outliers, and efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known.

After mapping the matching scores s_j to a common domain x_j, simple combination rules as in Eq. (2.9) are then usually applied.

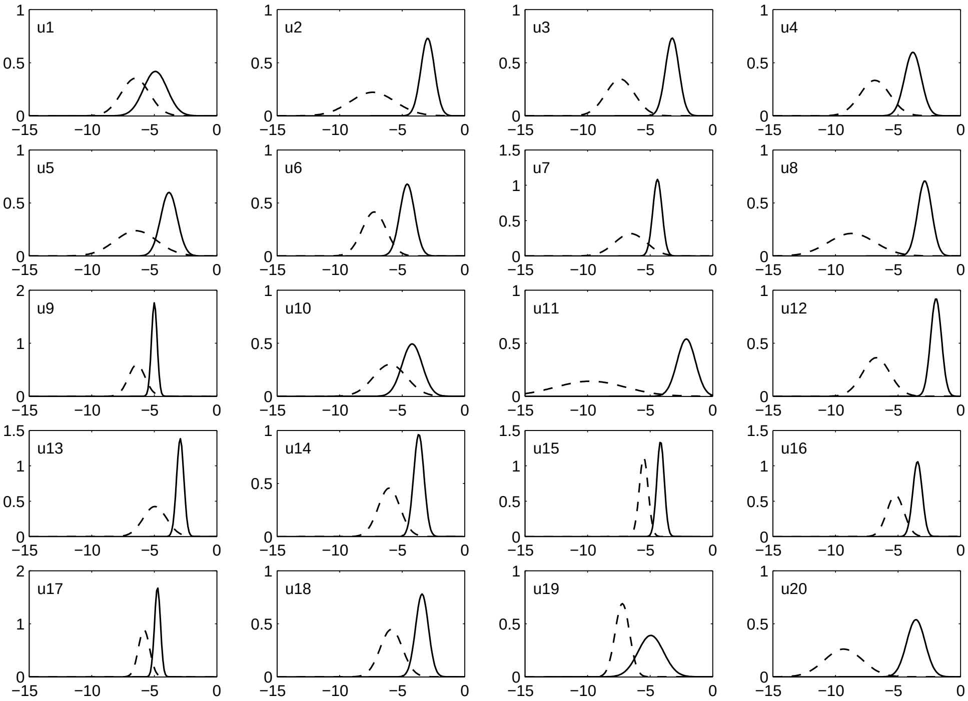
2.2.2.4. The NIST SRE and SVC Experiences

Doddington *et al.* [1998] studied the behavior of different individual speakers in the Speaker Recognition Evaluation organized by the National Institute of Standards and Technology (NIST SRE 1998). It was observed a number of different speaker behaviors in terms of verification performance, further classifying groups of particular users in one of these categories: 1) *sheeps* were speakers whose voice patterns were easily accepted by the system, 2) *goats* were speakers who were exceptionally unsuccessful at being accepted, 3) *lambs* were speakers who were exceptionally vulnerable to impersonation, and 4) *wolves* were speakers who were exceptionally successful at impersonation. This categorization is also known as the Doddington's zoo [Bolle *et al.*, 2004a].

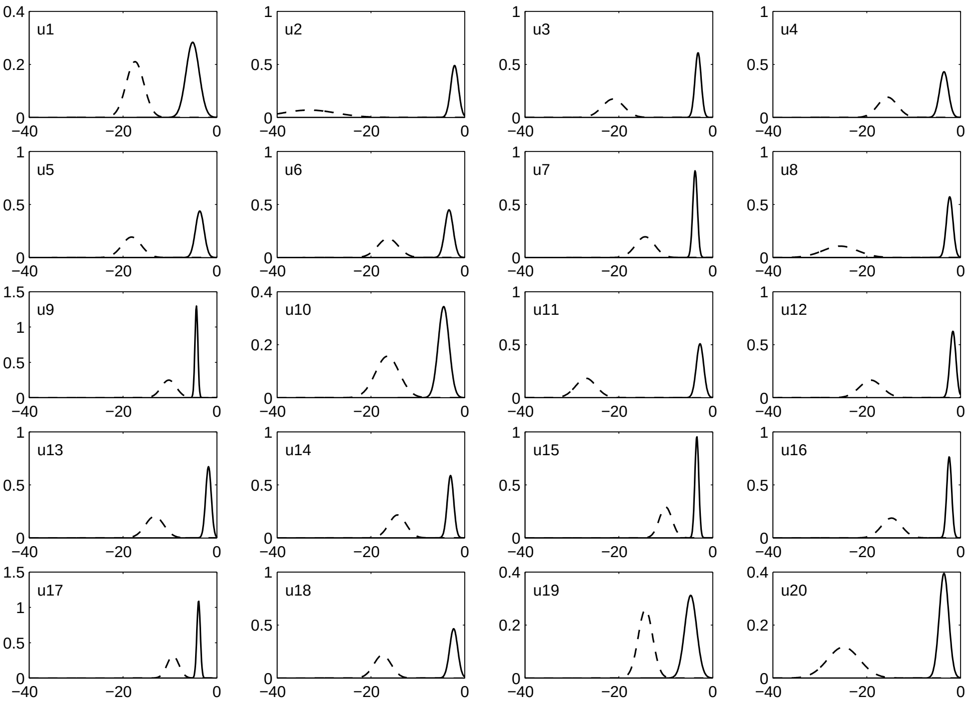
Although clearly detailed for the first time in this popular work, this user dependency has been traditionally assumed in a number of biometric verification systems, specially those based on behavioral traits such as written signature [Paulik *et al.*, 1994; Plamondon and Lorette, 1989] and voice [Furui, 1981; Matsui *et al.*, 1996]. In these preliminary works, the verification performance across users was balanced by introducing user-dependent decision thresholds. More recently, the common approach is to apply test- or user-dependent score normalization techniques, which try to map the score distributions of different users to a common range [Auckenthaler *et al.*, 2000; Bimbot *et al.*, 2004].

This dependency of the score distributions on the particular subject has been observed not only in published works but also in our practice at the Biometrics Research Lab.–ATVS. On the one hand, ATVS has been participating to NIST SRE benchmarks since 2002, observing the same effects evidenced by Doddington *et al.* [1998]. These effects are now compensated with advanced forms of score normalization as KL-Tnorm [Ramos-Castro *et al.*, 2006a].

On the other hand, ATVS participated in the First International Signature Competition in 2004 [Yeung *et al.*, 2004], with the local system presented in Chapter 5, observing also these user-dependencies. We include here a preliminary experiment in order to visualize this effect. User-dependent client and impostor score distributions are plotted in Fig. 2.3 when testing either with skilled (a) or random forgeries (b). The different curves represent Gaussian fits of the matching scores obtained following the SVC experimental protocol on the SVC development corpus as detailed by Yeung *et al.* [2004], see Sect. 4.4. In this preliminary experiment we can observe large differences both in the individual verification performance (e.g., u1 and u8 when testing with skilled forgeries), and in the client-impostor scoring regions (e.g., u1 and u9 when testing with random forgeries). The main objective of user-dependent score normalization techniques [Auckenthaler *et al.*, 2000; Bimbot *et al.*, 2004] is to prevent such misalignments.



(a) Skilled forgeries.



(b) Random forgeries.

Figure 2.3: *Gaussian fit of client (solid) and impostor (dashed) score distributions for users u1 to u20 of SVC 2004 development corpus.*

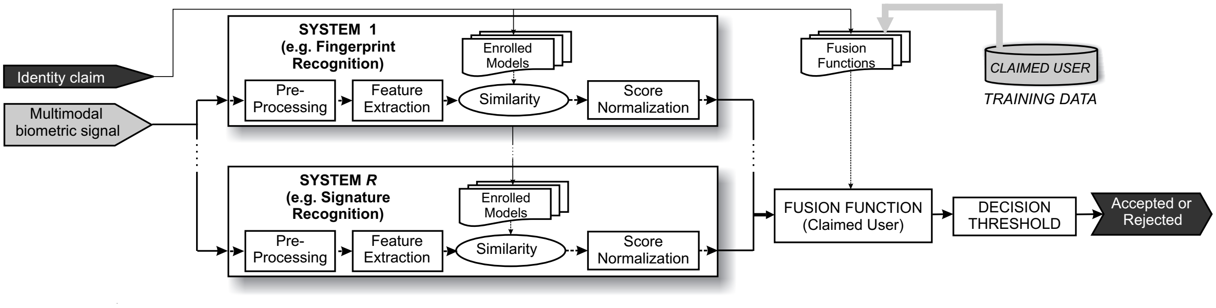


Figure 2.4: *System model of multimodal biometric authentication with user-dependent score-level fusion.*

2.2.2.5. Contribution: User-Dependent Score Normalization

Summarizing the related works presented thus far, the main reasons that have motivated us to concentrate our research efforts towards user-dependent score normalization are:

- Previous works in speaker and signature verification dealing using user-dependent thresholds, and recent works using score normalization.
- The previous evidence of strong user-dependencies found in NIST SRE 1998 which resulted in the Doddington's zoo.
- Our practice at Biometrics Research Lab.–ATVS in international benchmarks for speaker and signature verification, where we have also observed such strong user-dependencies.

The result is a strong motivation to explore in more detail the benefits of adapting the score normalization blocks shown in Fig. 1.1 to every user enrolled in the system, with emphasis on signature verification.

2.3. Adapted Fusion in Multimodal Biometrics

This Thesis is focused on score fusion approaches for multimodal biometric authentication, adapted both to user-specificities and to the input biometric quality. In the following sections we summarize the previous works that have motivated these approaches. Note that adapted approaches for fusion are (by definition) trained approaches.

2.3.1. User-Dependent Fusion

The system models of user-dependent score fusion and user-dependent score decision are shown in Figs. 2.4 and 2.5, respectively.

The idea of exploiting user-specific parameters at the score level in multimodal biometrics was introduced, to the best of our knowledge, by Jain and Ross [2002]. In this work, user-independent weighted linear combination of similarity scores was demonstrated to be improved by using either user-dependent weights or user-dependent decision thresholds, both of them computed by

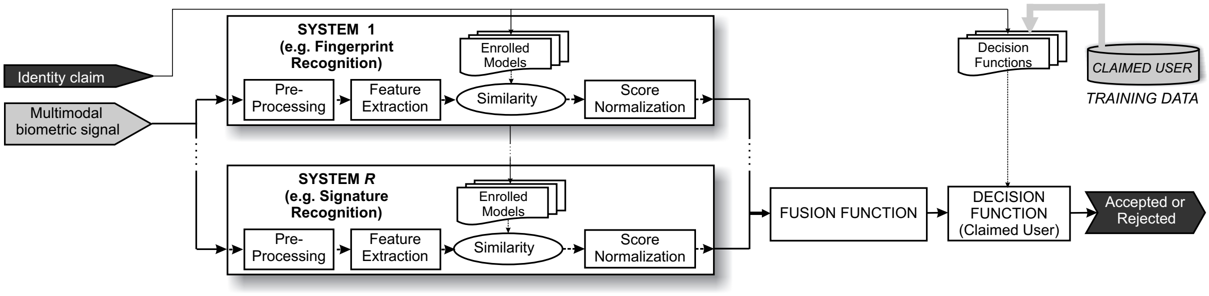


Figure 2.5: *System model of multimodal biometric authentication with user-dependent decision functions.*

exhaustive search on the testing data. The idea of user-dependent fusion parameters was also explored by Wang *et al.* [2004] using non-biased error estimation procedures. Other attempts to personalized multimodal biometrics include the use of the claimed identity index as a feature for a global trained fusion scheme based on Neural Networks [Kumar and Zhang, 2003], computing user-dependent weights using lambness metrics [Snelick *et al.*, 2005], and using personalized Fisher ratios [Poh and Bengio, 2005d].

Toh *et al.* [2004a] proposed a taxonomy of score-level fusion approaches for multi-biometrics. Existing multimodal fusion approaches are classified as global or local depending firstly on the fusion function (i.e., user-independent or user-dependent fusion strategies) and secondly depending on the decision making process (i.e., user-independent or user-dependent decision thresholds): global-learning-global-decision (GG), local-learning-global-decision LG, and similarly GL and LL. Some example works of each group are:

GG: Ben-Yacoub *et al.* [1999]; Bigun *et al.* [1997a]; Brunelli and Falavigna [1995]; Chatzis *et al.* [1999]; Hong and Jain [1998]; Kittler *et al.* [1998]; Verlinde *et al.* [2000].

LG: Fierrez-Aguilar *et al.* [2004b, 2003a]; Jain and Ross [2002]; Kumar and Zhang [2003]; Snelick *et al.* [2005]; Toh *et al.* [2004a]; Wang *et al.* [2004].

GL: Fierrez-Aguilar *et al.* [2005b]; Jain and Ross [2002]; Toh *et al.* [2004a]

LL: Fierrez-Aguilar *et al.* [2005b]; Toh *et al.* [2004a]

2.3.1.1. Contribution: Adapted User-Dependent Fusion

In the present work we adhere to the taxonomy proposed by Toh *et al.* [2004a] and extend it by incorporating adapted-learning and adapted-decisions. Adapted methods in the context of user-dependent fusion will refer to methods using both global and local information for learning the fusion rule and training the decision scheme, respectively.

The idea of adapted learning is based on the fact that the amount of available training data in localized learning is usually not sufficient and representative enough to guarantee good parameter estimation and generalization capabilities. To cope with this lack of robustness derived from partial knowledge of the problem, the use of robust adaptive learning strategies based on background information has been proposed in related research areas [Lee and Huo, 2000]. As an example of this approach we exploit the fact that general information of the problem (i.e., user-independent data) can constitute a rich source of information for user-specific recognition problems. In general, the relative balance between the background information (pool of users) and the local data (specific user) is performed as a tradeoff between both kinds of information.

2.3.2. Quality-Based Fusion

There is recent interest in studying the effects of signal quality on the performance of biometric systems [Junqua and Noord, 2001; Simon-Zorita *et al.*, 2003; Wilson *et al.*, 2004]. As a result, it is known that the performance of an unimodal system can drop significantly under noisy conditions. Multimodal systems have been demonstrated to overcome this challenge to some extent by combining the evidences provided by a number of different traits. This idea can be extended by explicitly considering quality measures of the input biometric signals and weighting the various pieces of evidence based on this quality information. Following this idea, novel quality-based multimodal authentication schemes are proposed in this Thesis, and their benefits are demonstrated on a publicly available real multimodal biometric database.

Bigun *et al.* [1997a] presented the problem of multimodal biometric authentication by using Bayesian statistics. The result was an Expert Conciliation scheme including weighting factors not only for the accuracy of the experts but also for the confidence of the experts on the particular input samples. Experiments were provided by combining face and voice modalities. The idea of relating the confidence value to quality measures of the input biometric signals was nevertheless not developed, which is one of the contributions of this Thesis.

The concept of confidence measure of matching scores was also studied by Bengio *et al.* [2002]. In this work they demonstrated that the confidence of matching scores can help in the fusion process. In particular, they tested confidence measures based on: 1) Gaussian assumptions on the score distributions, 2) the adequacy of the trained biometric models to explain the input data, and 3) resampling techniques on the set of test scores. This research line was further developed by Poh and Bengio [2005c] who devised confidence measures based on the margin between impostor and client score distributions.

Chatzis *et al.* [1999] evaluated a number of fusion schemes based on clustering strategies. In this case quality measures obtained directly from the input biometric signals were used to fuzzify the scores provided by the different systems. They demonstrated that fuzzy versions of k-means and Vector Quantization including the quality measures outperformed slightly, and not in all cases, the standard non-fuzzy clustering methods. This work is, to the best of our knowledge, the first one reporting results of quality-based fusion. One limitation in the experimental setup of this work was the reduced number of individuals used, only 37.

Another more recent effort in quality-based fusion without the success of previous methods was reported by Toh *et al.* [2004b], who developed a score fusion scheme based on polynomial functions. Quality measures were introduced in the optimization problem for training the polynomials as weights in the regularization term. Unexpectedly, no performance improvements were obtained by including the quality measures. One inconvenience of this work was the use of a chimeric multimodal database combining the data from 3 different face, voice and fingerprint databases.

2.3.2.1. The FVC Experience

In this subsection we outline an important observation from the series of Fingerprint Verification Competitions (FVC) that has influenced us to focus our research efforts towards quality estimation in biometrics (specially in fingerprint images) and its application to quality-based fusion. This observation outlines the importance of the proposed methods in practice.

Recent efforts have been conducted in order to establish common evaluation scenarios enabling a fair comparison between competing systems [Wayman *et al.*, 2005]. In the case of fingerprint recognition, a series of International Fingerprint Verification Competitions (FVC) have received great attention both from the academy and the industry [Cappelli *et al.*, 2006]. These competitions have provided common data and procedures widely available now for further research [Maltoni *et al.*, 2003]. Other recent comparative benchmark studies include the Fingerprint Vendor Technology Evaluations organized by NIST [Wilson *et al.*, 2004]. Details of the last FVC competition can be found elsewhere [Cappelli *et al.*, 2006; FVC, 2004].

The series of FVC competitions have been organized biannually since 2000 by the Biometrics Systems Laboratory of Bologna University, the Pattern Recognition and Image Processing Laboratory of Michigan State University, and the Biometric Test Center of San Jose State University.

Data for the competitions consist of 4 different databases, the first three acquired with different sensors and the last one created with a synthetic generator [Cappelli *et al.*, 2002b]. In the first competition FVC2000, 11 algorithms were evaluated. Data were acquired without any special restrictions at acquisition time. The best system obtained an average EER of 1.73% over the 4 databases. The average EER of the first five systems was 4.52%.

Two years after in FVC2002 the number of participants increased significantly to 31. Similar databases were used, also acquired without any special restrictions and using the different sensors in a natural way. Average error rates over the 4 databases decreased significantly both for the first system (0.19% EER) and for the average of the first 5 systems (0.52% EER). In some sense, these results corroborated the maturity of fingerprint verification systems.

The more interesting result related to our work on quality measures was obtained in FVC2004. In this case the image quality of the 4 fingerprint databases was artificially corrupted by using an acquisition procedure with exaggerated plastic distortions, artificial dryness and moistness. See Figs. 2.6 and 2.7 for selected images of good and low image quality, respectively. Although the number of participants increased to 41, and a performance drop was foreseen, the results were surprisingly much worse than those in FVC2000. The best system achieved an average EER over the four databases of 2.07%, and the average for the first five systems was 2.36%

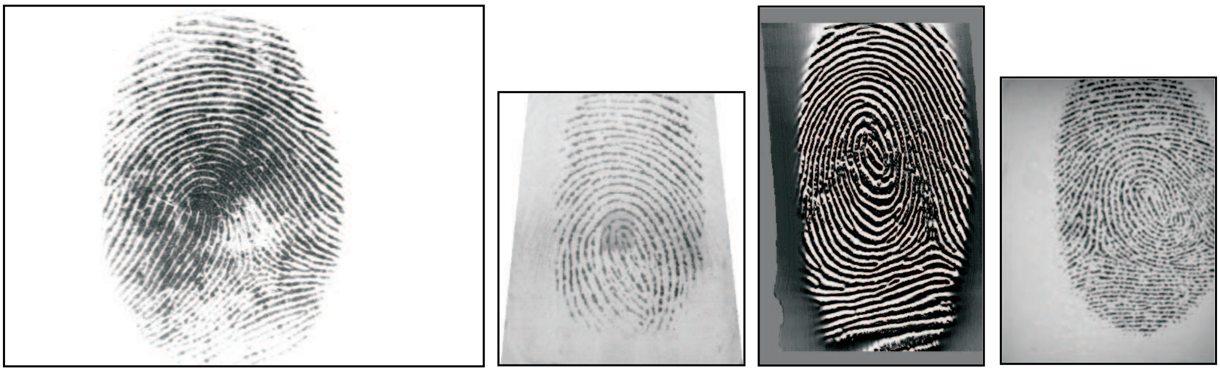


Figure 2.6: *Fingerprint examples of good quality from the four databases used in FVC2004 (left to right): DB1 (CrossMatch V300), DB2 (Digital Persona UareU 4000), DB3 (Atmel FingerChip), and DB4 (SFinGe v3.0).*

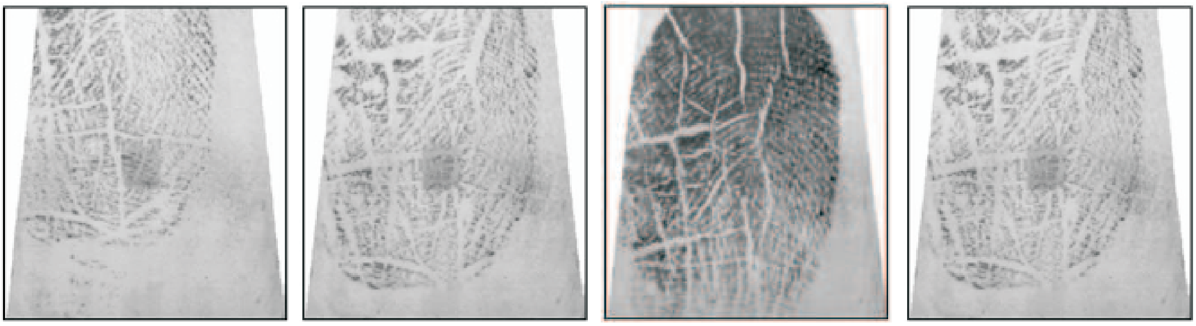


Figure 2.7: *Fingerprint impressions from a low quality finger in FVC2004 (DB2 acquired with Digital Persona UareU 4000).*

EER. A subsequent work demonstrated that by combining the most complementary systems (not necessarily the best ones) the results could be improved to error rates similar to those reported in FVC2000 [Fierrez-Aguilar *et al.*, 2005g]. Worth noting, the following competition has been just announced [FVC, 2006], and will include a study of image quality effects. In this case the Biometrics Research Lab.–ATVS collaborates in the organization of the competition.

In summary, it can be observed that even the best fingerprint verification systems worldwide, both from industry and academy, struggle in the presence of noisy images. This result has been recently recognized by the biometric community and is now considered a big research challenge [BQW, 2006]. As demonstrated in some recent works, this can be overcome by multi-algorithm fusion [Fierrez-Aguilar *et al.*, 2005g]. The purpose of this Thesis is going one step further by explicitly considering the input quality in multibiometric approaches.

2.3.2.2. Contribution: Quality-Based Fusion

Summing up the related works presented thus far, the main reasons that have motivated us to concentrate our research efforts towards quality-based fusion are:

- Self observation about the benefits that intuitively can be obtained by considering the input quality when combining different information in multibiometric systems.

- Previous efforts describing general frameworks for multimodal fusion including confidence measures, but not always related to the input biometric quality.
- Preliminary works including the input quality in biometric fusion with not statistically significant and sometimes contradictory results.
- The importance of biometric quality in practical systems as demonstrated by the FVC experience.

As a result, this PhD Thesis is aimed at developing adapted schemes for quality-based fusion, both based on generative assumptions following the work by Bigun *et al.* [1997a] and discriminative criteria using Support Vector Machines.

2.4. Chapter Summary and Conclusions

In this chapter we have summarized the main works related to this PhD Thesis. We have started by describing the general problem of multiple classifier combination, categorizing the existing approaches with pointers to the theoretical underpinnings behind classifier fusion. Then we have focused on multimodal biometric fusion at the score level, dividing the existing approaches into non-adapted and adapted fusion. Within adapted fusion methods, we have outlined the previous works that have motivated us to focus this Thesis on user-dependent and quality-based fusion. In the next chapter we develop the proposed schemes for both adapted strategies.

No new material has been presented in this chapter. Although the exposition of some parts of chapter is not new (in particular Sect. 2.2 is largely based on Jain *et al.* [2005]), most of the chapter structure has followed a personal perspective.

Chapter 3

Adapted Fusion Schemes

THIS CHAPTER describes the score fusion schemes proposed in this PhD Thesis. These schemes are divided into three classes: 1) user-dependent, 2) quality-based, and 3) user-dependent and quality-based. Although the last class includes the first two classes as particular cases, the three classes are introduced sequentially in order to facilitate the description.

For each class of methods, we first sketch the system model and then we derive particular implementations by using standard pattern recognition methods [Duda *et al.*, 2001], either based on generative assumptions following Bayesian theory, or discriminative criteria using Support Vector Machines. These two classes of implementations aim at minimizing the Bayesian error and the Structural Risk of the verification task, respectively.

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 k. The similarity scores s are normalized to x. Let the normalized similarity scores provided by the different unimodal systems be combined into a multimodal score $\mathbf{x} = [x_1, \ldots, x_M]^T$. The design of a fusion scheme consists in the definition of a function $f : \mathbb{R}^M \rightarrow \mathbb{R}$, so as to maximize the separability of client $\{f(\mathbf{x})|\text{client attempt}\}$ and impostor $\{f(\mathbf{x})|\text{impostor attempt}\}$ fused score distributions. This function may be trained by using labelled training scores $(\mathbf{x}_i, z_i)$, where $z_i = \{0 = \text{impostor attempt}, 1 = \text{client attempt}\}$. The rest of the chapter deals with different schemes for constructing this function adapted both to the individual user and/or to the quality of the input biometric signals according to different separability criteria. In Fig. 3.1 we depict the general system model including all the notations defined above.

This chapter assumes a basic understanding of the fundamentals of pattern recognition and classification [Duda *et al.*, 2001; Jain *et al.*, 2000a; Theodoridis and Koutroumbas, 2003].

This chapter is based on the publications: Bigun *et al.* [2003]; Fierrez-Aguilar *et al.* [2004b, 2005b,c, 2004d, 2005i].

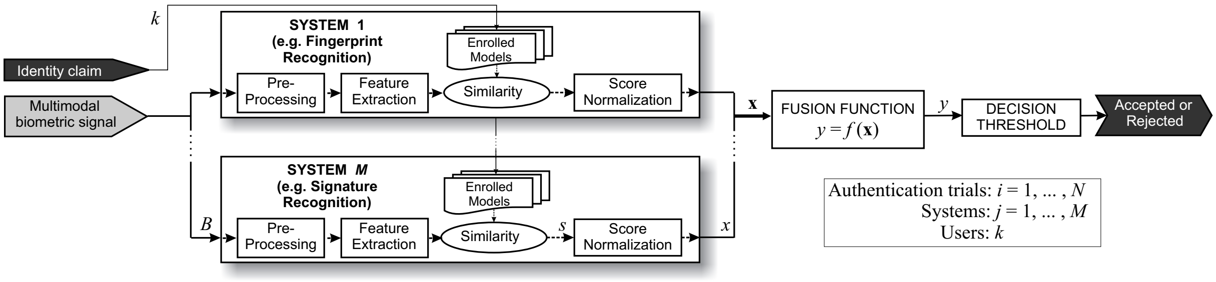


Figure 3.1: *General system model of multimodal biometric authentication using score level fusion including name conventions.*

3.1. User-Dependent Fusion

For user-dependent fusion we assume the use of two sets of training scores each one including both genuine and impostor scores. The first set consists of training scores corresponding to the user being tested. The second set consists of scores corresponding to a pool of background users different to the user being tested. By considering these two sets simultaneously, we experimentally demonstrate that the general information provided by the pool of users can be exploited in user-dependent fusion schemes. For demonstrating this, three algorithms are developed and compared for each user-dependent strategy, namely:

Global. Only the scores from the pool of users are used for training (both genuine and from impostors). This is equivalent to the traditional user-independent fusion.

Local. Only the scores from the user at hand are used for training (both genuine and from impostors). This is equivalent to the traditional user-dependent fusion.

Adapted. The scores from both the pool of users and the user at hand are used for training (both genuine and from impostors). This is an original contribution of this Thesis.

Note that here we use the term *adapted* in the sense of adapted from the general knowledge provided by the background users to the specificities of the individual user at hand. In this sense, the traditional user-dependent fusion methods, which are *local*, are not adapted to individual users but just trained on them.

User-dependent multimodal authentication can be achieved mainly by making user-dependent each one or a combination of the following three modules depicted in Fig. 3.1: 1) score normalization, 2) score fusion, and 3) decision. We develop each one of these three cases in the following sections.

3.1.1. User-Dependent Score Normalization

We first formulate the problem of biometric authentication as an hypotheses test. Examining the factors arising from this hypotheses test, we obtain the general system model for score

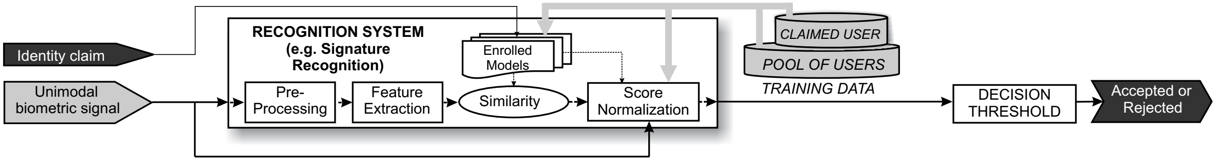


Figure 3.2: *System model of biometric authentication with user-dependent score normalization.*

normalization depicted in Fig. 3.2. This theoretical framework also serves to classify existing algorithms for score normalization in a taxonomy. This taxonomy is completed by introducing some modified versions of existing algorithms better suited to the severe user-dependencies found in biometric modalities such as signature verification.

For the description of user-dependent score normalization in the following sections, we refer to an individual system as depicted in Fig. 3.2.

3.1.1.1. Score Normalization Framework

Given a biometric test sample B, the problem of biometric authentication can be stated as a basic hypotheses test between two hypotheses:

- $H1$: B is from hypothesized client k.
- $H0$: B is *not* from hypothesized client k.

The optimum test to decide between these two hypotheses is a likelihood ratio test [Duda *et al.*, 2001]

$$\frac{p(B|H1)}{p(B|H0)} \begin{cases} > \theta \text{ Accept } H1 \\ < \theta \text{ Accept } H0 \end{cases}, \tag{3.1}$$

where $p(B|H1)$ and $p(B|H0)$ are respectively the probability density functions for the hypotheses $H1$ and $H0$ evaluated for the observed biometric sample B. The decision threshold for accepting or rejecting $H1$ is θ. An equivalent log-likelihood ratio test is obtained transforming Eq. (3.1) into the log domain

$$\log p(B|H1) - \log p(B|H0) \begin{cases} > \log \theta \text{ Accept } H1 \\ < \log \theta \text{ Accept } H0 \end{cases}. \tag{3.2}$$

A common practice in biometric verification, e.g., GMM in case of speaker recognition [Reynolds *et al.*, 2000] and HMM in case of signature recognition [Ortega-Garcia *et al.*, 2003a], consists in characterizing each target client k by a statistical model λ^k (i.e., the enrolled model in Fig. 3.2). In this case, the similarity s is computed as

$$s = \log p(B|\lambda^k), \tag{3.3}$$

which is an estimation of $\log p(B|H1)$. As a result, the optimal score normalization strategy for an authentication system based on statistical modelling is given by

$$x = s - \log p(B|H0). \tag{3.4}$$

The normalizing term $\log p(B|H0)$ is affected, in general, by:

- Input information: the input biometric sample B.
- Information from the client: which is typically extracted from scores $\left\{s_1^k, \ldots, s_{N_k}^k\right\}$ of the hypothesized client k claiming its model λ^k.
- Information from impostors: which is typically extracted from models $\left\{\lambda_1^{\overline{k}}, \ldots, \lambda_{N_I}^{\overline{k}}\right\}$ and scores $\left\{s_1^{\overline{k}}, \ldots, s_{N_{\overline{k}}}^{\overline{k}}\right\}$ from N_I possible impostors (either real or casual) of the hypothesized client k.

Estimation of $\log p(B|H0)$ based on the different information involved (which is also sketched in Fig. 3.2) is nevertheless not a straightforward task. Thus, operational procedures of *score normalization* (also known as *likelihood normalization*) are usually employed in the literature. Much effort has been done in order to derive such procedures, mainly in the speaker recognition community [Auckenthaler *et al.*, 2000; Bimbot *et al.*, 2004]. These operational procedures aim at designing a function

$$x = f(s, B, \left\{s_1^k, \ldots, s_{N_k}^k\right\}, \left\{\lambda_1^{\overline{k}}, \ldots, \lambda_{N_I}^{\overline{k}}\right\}, \left\{s_1^{\overline{k}}, \ldots, s_{N_{\overline{k}}}^{\overline{k}}\right\}) \tag{3.5}$$

so as to minimize the error rate of the verification task. The use of linear functions of various statistics of the information involved in Eq. (3.5) is the prevailing strategy for deriving score normalization algorithms. This is the case of [Bimbot *et al.*, 2004]: 1) z-norm, which considers only scores from impostors, 2) t-norm, based on the input biometric sample and models from impostors, and 3) UBM-norm, which considers the input biometric signal and a universal background model characterizing the average user. Other score normalization examples can be found in face [Sanderson and Paliwal, 2002] and signature recognition [Ortega-Garcia *et al.*, 2003a].

In order to simplify the discussion yet providing a powerful framework for score alignment across users, neither input test information nor models from impostors are considered in this work, i.e.

$$x = f(s, \left\{s_1^k, \ldots, s_{N_k}^k\right\}, \left\{s_1^{\overline{k}}, \ldots, s_{N_{\overline{k}}}^{\overline{k}}\right\}). \tag{3.6}$$

This family of score normalization methods will be referred to as *user-dependent score normalization techniques*. Other normalization methods using the input biometric signal and models from impostors will be referred to as *test-dependent normalization techniques* [Auckenthaler *et al.*, 2000].

Note that here we are assuming only the use of training scores corresponding to the user being tested (genuine and/or impostor scores) but not scores corresponding to other users. In this sense, and according to the definition of global, local, and adapted methods presented in Sect. 3.1, the following score normalization algorithms are *local* but not *adapted*.

3.1.1.2. User-Dependent Score Normalization Algorithms

Following Auckenthaler *et al.* [2000], user-dependent score normalization algorithms have been classified as follows.

Impostor-Centric Score Normalization Algorithms. In impostor-centric methods (IC) no information about client score intra-variability is used. Therefore

$$x_{IC} = f(s, \mathcal{I} = \left\{s_1^{\overline{k}}, \ldots, s_{N_{\overline{k}}}^{\overline{k}}\right\}). \tag{3.7}$$

The following IC methods are considered here:

- IC-1: $x_{IC\text{-}1} = s - \mu_{\mathcal{I}}$,
- IC-2: $x_{IC\text{-}2} = s - (\mu_{\mathcal{I}} + \sigma_{\mathcal{I}})$,
- IC-3: $x_{IC\text{-}3} = (s - \mu_{\mathcal{I}})/\sigma_{\mathcal{I}}$,

where $\mu_{\mathcal{I}}$ and $\sigma_{\mathcal{I}}$ are respectively the mean and standard deviation of the impostor scores $\mathcal{I}$. IC-1 is proposed here as a robust technique for small sample size normalization problems [Raudys and Jain, 1991], IC-2 is equivalent to the *a priori* decision threshold setting described by Furui [1981] and IC-3 is the well known z-norm technique [Auckenthaler *et al.*, 2000].

Note that the impostor scores $\mathcal{I}$ can be, in general, either from casual impostors (in this case leading to a casual-Impostor-Centric method, cIC) or from real impostors (similarly, leading to rIC).

Target-Centric Score Normalization Algorithms. In target-centric methods (TC) no information about impostor score variability is used. Therefore

$$x_{TC} = f(s, \mathcal{C} = \left\{s_1^{k}, \ldots, s_{N_k}^{k}\right\}). \tag{3.8}$$

Similarly to the impostor-centric case, the following methods are obtained

- TC-1: $x_{TC\text{-}1} = s - \mu_{\mathcal{C}}$,
- TC-2: $x_{TC\text{-}2} = s - (\mu_{\mathcal{C}} - \sigma_{\mathcal{C}})$,
- TC-3: $x_{TC\text{-}3} = (s - \mu_{\mathcal{C}})/\sigma_{\mathcal{C}}$,

where $\mu_{\mathcal{C}}$ and $\sigma_{\mathcal{C}}$ are respectively the mean and standard deviation of the client scores $\mathcal{C}$. TC-1 is based on the running average normalization strategy proposed by Naik and Doddington [1986],

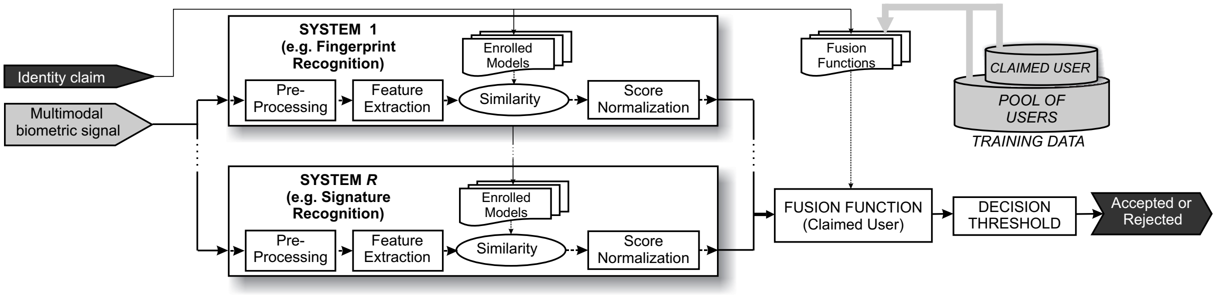


Figure 3.3: *System model of multimodal biometric authentication with adapted user-dependent score fusion.*

TC-2 is a form of the *a priori* decision thresholding technique proposed by Saeta and Hernando [2003] and TC-3 is the normalization scheme proposed by Garcia-Romero *et al.* [2003b].

Client scores $\mathcal{C}$ should be obtained from the available training set. Here we propose to generate $\mathcal{C}$ by using one of the sampling methods described in Sect. 4.1.2.

Target-Impostor Score Normalization Algorithms. In target-impostor methods (TI) information from both client score intra-variability and impostor score variability is used. Therefore

$$x_{TI} = f(s, \mathcal{C} = \left\{s_1^k, \ldots, s_{N_k}^k\right\}, \mathcal{I} = \left\{s_1^{\overline{k}}, \ldots, s_{N_{\overline{k}}}^{\overline{k}}\right\}). \tag{3.9}$$

Based on the decision thresholding techniques described by Matsui *et al.* [1996] and Burton [1987], we obtain the following target-impostor normalization methods

- TI-1: $x_{TI\text{-}1} = s - s_{EER}(\mathcal{C}, \mathcal{I})$,
- TI-2: $x_{TI\text{-}2} = s - (\mu_{\mathcal{I}}\sigma_{\mathcal{C}} + \mu_{\mathcal{C}}\sigma_{\mathcal{I}})/(\sigma_{\mathcal{I}} + \sigma_{\mathcal{C}})$,

where $s_{EER}(\mathcal{C}, \mathcal{I})$ is the target-dependent decision threshold at the empirical Equal Error Rate obtained from the histograms of $\mathcal{C}$ and $\mathcal{I}$.

3.1.2. User-Dependent Score Fusion

The system model of user-dependent score fusion is shown in Fig. 3.3. The aim of the user-dependent score fusion scheme is to obtain the best score function for each particular user.

User-dependent score fusion is confronted with a great challenge: the scarcity of user-dependent training scores. For overcoming this challenge, and as a contribution with respect to the state of the art, we assume the simultaneous use of user-specific and background information for training the user-specific fusion functions. Two algorithms implementing the proposed adapted user-dependent fusion schemes are given in the following sections.

3.1.2.1. Bayesian User-Dependent Score Fusion Algorithm

Impostor and client score distributions are modelled as multivariate Gaussians $p(\mathbf{x}|\omega_0) = N(\mathbf{x}|\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2)$ and $p(\mathbf{x}|\omega_1) = N(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$, respectively[1]. The fused score y_T of a multimodal test score $\mathbf{x}_T$ is defined then as follows

$$y_T = f(\mathbf{x}_T) = \log p(\mathbf{x}_T|\omega_1) - \log p(\mathbf{x}_T|\omega_0), \tag{3.10}$$

which is known to be a Quadratic Discriminant (QD) function consistent with Bayes estimate in case of equal impostor and client prior probabilities [Duda *et al.*, 2001]. The score distributions are estimated using the available training data as follows:

Global. The training set $X_\mathrm{G} = (\mathbf{x}_i, z_i)_{i=1}^{N_\mathrm{G}}$ includes multimodal scores from a number of different clients, and $(\{\boldsymbol{\mu}_{\mathrm{G},0}, \boldsymbol{\sigma}^2_{\mathrm{G},0}\}, \{\boldsymbol{\mu}_{\mathrm{G},1}, \boldsymbol{\sigma}^2_{\mathrm{G},1}\})$ are estimated by using the standard Maximum Likelihood criterion [Reynolds *et al.*, 2000]. The resulting fusion rule $f_\mathrm{G}(\mathbf{x})$ is applied globally at the operational stage regardless of the claimed identity.

Local. A different fusion rule $f_{k,\mathrm{L}}(\mathbf{x})$ is obtained for each client k enrolled in the system by using Maximum Likelihood density estimates $(\{\boldsymbol{\mu}_{k,\mathrm{L},0}, \boldsymbol{\sigma}^2_{k,\mathrm{L},0}\}, \{\boldsymbol{\mu}_{k,\mathrm{L},1}, \boldsymbol{\sigma}^2_{k,\mathrm{L},1}\})$ computed from a set of development scores X_k of the specific client k.

Adapted. The adapted fusion rule $f_{k,\mathrm{A}}(\mathbf{x})$ of client k trades off the general knowledge provided by the user-independent development data X_G, and the user specificities provided by the user-dependent training set X_k, through Maximum a Posteriori density estimation [Reynolds *et al.*, 2000]. This is done by adapting the sufficient statistics as follows

$$\begin{aligned} \boldsymbol{\mu}_{k,\mathrm{A},l} &= \alpha_l \boldsymbol{\mu}_{k,\mathrm{L},l} + (1-\alpha_l)\boldsymbol{\mu}_{\mathrm{G},l}, \\ \boldsymbol{\sigma}^2_{k,\mathrm{A},l} &= \alpha_l(\boldsymbol{\sigma}^2_{k,\mathrm{L},l} + \boldsymbol{\mu}^2_{k,\mathrm{L},l}) + (1-\alpha_l)(\boldsymbol{\sigma}^2_{\mathrm{G},l} + \boldsymbol{\mu}^2_{\mathrm{G},l}) - \boldsymbol{\mu}^2_{j,\mathrm{A},l}. \end{aligned} \tag{3.11}$$

For each class $l = \{0 = \text{impostor}, 1 = \text{client}\}$, a data-dependent adaptation coefficient

$$\alpha_l = N_l/(N_l + r) \tag{3.12}$$

is used, where N_l is the number of local training scores in class l, and r is a fixed relevance factor.

3.1.2.2. Discriminative User-Dependent Score Fusion Algorithm

Let the training set be $X = (\mathbf{x}_i, z_i)_{i=1}^{N}$ where N is the number of multimodal scores in the training set, and $z_i \in \{-1, 1\} = \{\text{Impostor}, \text{Client}\}$. The principle of SVM relies on a linear separation in a high dimension feature space $\mathbb{H}$ where the data have previously been mapped via $\Phi : \mathbb{R}^M \to \mathbb{H}; X \to \Phi(X)$, so as to take into account the eventual non-linearities of the problem

[1] We use diagonal covariance matrixes, so $\boldsymbol{\sigma}^2$ is shorthand for $\mathrm{diag}(\Sigma)$. Similarly, $\boldsymbol{\mu}^2$ is shorthand for $\mathrm{diag}(\boldsymbol{\mu}\boldsymbol{\mu}')$.

[Vapnik, 2000]. In order to achieve a good level of generalization capability, the margin between the separator hyperplane

$$\{\mathbf{h} \in \mathbb{H} | \langle \mathbf{w}, \mathbf{h} \rangle_{\mathbb{H}} + w_0 = 0\} \tag{3.13}$$

and the mapped data $\Phi(X)$ is maximized, where $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ denotes inner product in space $\mathbb{H}$, and $(\mathbf{w} \in \mathbb{H}, w_0 \in \mathbb{R})$ are the parameters of the hyperplane. The optimal hyperplane can be obtained as the solution of the following quadratic programming problem [Vapnik, 2000]

$$\min_{\mathbf{w}, w_0, \xi_1, \ldots, \xi_N} \left(\frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^{N} C_i \xi_i \right) \tag{3.14}$$

subject to

$$\begin{aligned} z_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathbb{H}} + w_0) &\geq 1 - \xi_i, \qquad i &= 1, \ldots, N, \\ \xi_i &\geq 0, \qquad i &= 1, \ldots, N, \end{aligned} \tag{3.15}$$

where slack variables ξ_i are introduced to take into account the eventual non-separability of $\Phi(X)$ and parameter $C_i = C$ is a positive constant that controls the relative influence of the two competing terms.

The optimization problem in Eqs. (3.14) and (3.15) is solved with the Wolfe dual representation by using the kernel trick [Theodoridis and Koutroumbas, 2003]

$$\max_{\alpha_1, \ldots, \alpha_N} \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j z_i z_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \tag{3.16}$$

subject to

$$\begin{aligned} & 0 \leq \alpha_i \leq C_i, \quad i = 1, \ldots, N \\ & \sum_{i=1}^{N} \alpha_i z_i = 0 \end{aligned} \tag{3.17}$$

where the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathbb{H}}$ is introduced to avoid direct manipulation of the elements of $\mathbb{H}$. In particular, radial basis functions

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2 \right), \tag{3.18}$$

and linear kernels

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j, \tag{3.19}$$

are used in this Thesis, resulting in complex and linear separating surfaces between client and impostor distributions, respectively.

The fused score y_T of a multimodal test pattern $\mathbf{x}_T$ is defined as follows

$$y_T = f(\mathbf{x}_T) = \langle \mathbf{w}^*, \Phi(\mathbf{x}_T) \rangle_{\mathbb{H}} + w_0^*, \tag{3.20}$$

which is a signed distance measure from $\mathbf{x}_T$ to the separating surface given by the solution of the SVM problem. Applying the Karush-Kuhn-Tucker (KKT) conditions to the problem in Eqs. (3.14) and (3.15), y_T can be shown to be equivalent to the following sparse expression

$$y_T = f(\mathbf{x}_T) = \sum_{i \in \mathrm{SV}} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}_T) + w_0^*, \tag{3.21}$$

where $(\mathbf{w}^*, w_0^*)$ is the optimal hyperplane, $(\alpha_1^*, \dots, \alpha_N^*)$ is the solution to the problem in Eqs. (3.16) and (3.17), and $\mathrm{SV} = \{i | \alpha_i^* > 0\}$ indexes the set of support vectors. The bias parameter w_0^* is obtained from the solution to the problem in Eqs. (3.16) and (3.17) by using the KKT conditions [Theodoridis and Koutroumbas, 2003].

As a result, the training procedure in Eqs. (3.16) and (3.17) and the testing strategy in Eq. (3.21) are obtained for the problem of multimodal fusion.

Global. The training set $X_{\mathrm{G}} = (\mathbf{x}_i, z_i)_{i=1}^{N_{\mathrm{G}}}$ includes multimodal scores from a number of different clients and the resulting fusion rule $f_{\mathrm{G}}(\mathbf{x})$ is applied globally at the operational stage regardless of the claimed identity.

Local. A different fusion rule $f_{k,\mathrm{L}}(\mathbf{x})$ is obtained for each client enrolled in the system k by using development scores X_k of the specific client k. At the operational stage, the fusion rule $f_{k,\mathrm{L}}(\mathbf{x})$ of the claimed identity k is applied.

Adapted. An adapted user-dependent fusion scheme trading off the general knowledge provided by a user-independent training set X_{G}, and the user specificities provided by a user-dependent training set X_k, is proposed. To obtain the adapted fusion rule, $f_{k,\mathrm{A}}(\mathbf{x})$, for user k, we propose to train both the global fusion rule, $f_{\mathrm{G}}(\mathbf{x})$, and the local fusion rule, $f_{k,\mathrm{L}}(\mathbf{x})$, as described above, and finally combine them as follows

$$f_{k,\mathrm{A}}(\mathbf{x}) = \alpha f_{k,\mathrm{L}}(\mathbf{x}) + (1 - \alpha) f_{\mathrm{G}}(\mathbf{x}), \tag{3.22}$$

where α is a trade-off parameter. This can be seen as a user-dependent fusion scheme adapted from user-independent information. The idea can also be extended easily to trained fusion schemes based on other classifiers. Worth noting, sequential algorithms to solve the SVM optimization problem in Eqs. (3.14) and (3.15) have been already proposed [Navia-Vazquez *et al.*, 2001], and can be used to extend the proposed idea, first constructing the user-independent solution and then refining it by incorporating the local data.

3.1.3. User-Dependent Decision

The system model of user-dependent decision is shown in Fig. 3.4. Once a fused similarity score has been obtained by using either a global, local or an adapted fusion method, the score is compared to a decision threshold in order to accept or reject the identity claim. This decision making process, also subject to training, can also be made globally, locally, or can be adapted

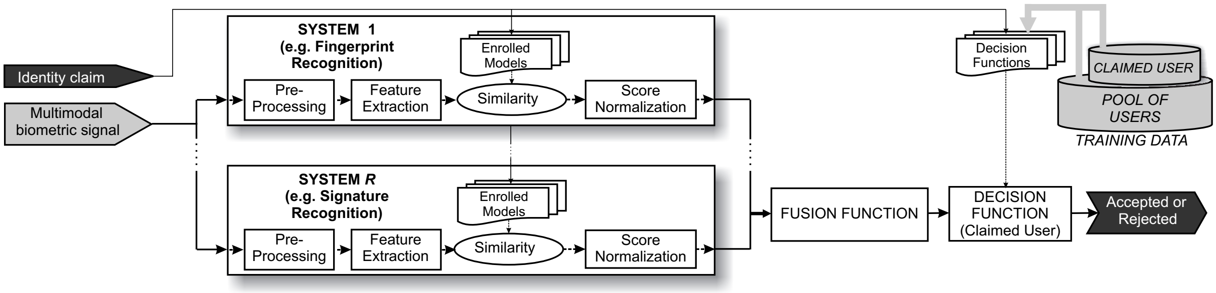


Figure 3.4: System model of multimodal biometric authentication with adapted user-dependent decision.

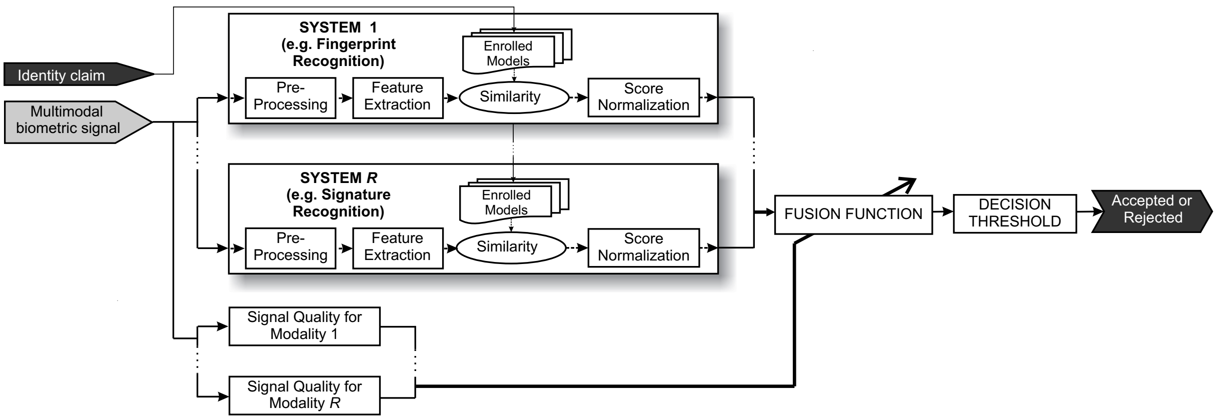


Figure 3.5: System model of multimodal biometric authentication with quality-based score fusion.

from global to local information. For this purpose, the methods presented in Sects. 3.1.2.1 and 3.1.2.2 can be directly applied exchanging the input multimodal scores $\mathbf{x}$ for fused scores y.

3.2. Quality-Based Fusion

Quality measures of the input biometric signals can be used for adapting the different modules of a multimodal authentication system. Although both the score normalization and decision modules are subject to this adaptation based on quality, only quality-based score fusion is considered in this Thesis. In Sect. 9.2 we provide some pointers of ongoing efforts and future works using quality measures for adapting other modules.

The system model of quality-based score fusion proposed in this work is shown in Fig. 3.5.

3.2.1. Quality-Based Combination Approach

One straightforward way to incorporate the input biometric quality to the score fusion approach is by including weights in simple combination approaches (see Sect. 2.2.2.1). In the case of the weighted average presented in Eq. (2.10), this can be achieved by using $w_j = q_j$ in order to obtain the following quality-based score fusion function

$$y = \sum_{j=1}^{M} q_j x_j, \tag{3.23}$$

where q_j is a quality measure of the score x_j. This score quality should be ideally related to the confidence of the system j in providing a reliable matching score for the particular biometric signal being tested. In the present work we use

$$q = \sqrt{Q \cdot Q_{\text{claim}}}, \tag{3.24}$$

where Q and Q_{claim} are the input biometric quality and the average quality of the biometric signals used for enrollment, respectively. The two quality measures Q and Q_{claim} are supposed to be in the range $[0, 1]$ where 0 corresponds to the poorest quality, and 1 corresponds to the highest quality.

Other definitions of score quality include: $q = (Q + Q_{\text{claim}})/2$, $q = \min\{Q, Q_{\text{claim}}\}$, etc.

3.2.2. Bayesian Quality-Based Score Fusion

The name conventions summarized in Fig. 3.1 are extended here:

x_{ij} Similarity score i delivered by system j

v_{ij} Variance of x_{ij} as estimated by system j

z_i The true label corresponding to score i

ζ_{ij} The error score $\zeta_{ij} = z_i - x_{ij}$

With respect to the previous cases developed in this chapter, note that here we introduce the variance v_{ij} of the input scores x_{ij}. The true labels z_i can take only two numerical values corresponding to "Impostor" and "Client". If x_{ij} is between 0 and 1 then these values are chosen to be 0 and 1, respectively. The fusion function is trained on shots $i \in 1 \dots N$ (i.e. x_{ij} and z_i are known for $i \in 1 \dots N$) and we consider the trial $N+1$ as a test shot on the working multimodal system (i.e. $x_{(N+1)j}$ is known, but z_{N+1} is not known).

3.2.2.1. Statistical Model

The model for combining the different systems (here also called machine experts) is based on Bayesian statistics and the assumption of normal distributed expert errors, i.e. ζ_{ij} is considered to be a sample of the random variable $\Theta_{ij} \sim N(b_j, \sigma_{ij}^2)$. It has been shown experimentally [Bigun *et al.*, 1997a] that this assumption does not strictly hold for common audio- and video-based biometric machine experts, but it is shown that it holds reasonably well when client and impostor distributions are considered separately. Taking this result into account, two different fusion functions are constructed, one of them based on genuine scores

$$\mathcal{C} = \{x_{ij}, v_{ij} | 1 \leq i \leq N \text{ and } z_i = 1, 1 \leq j \leq M\}, \tag{3.25}$$

while the other is based on impostor scores

$$\mathcal{I} = \{x_{ij}, v_{ij} | 1 \leq i \leq N \text{ and } z_i = 0, 1 \leq j \leq M\}. \tag{3.26}$$

The two fusion functions will be referred to as *client function* and *impostor function* respectively.

The client function estimates the expected true label of an input claim based on its expertise on recognizing client data. More formally, it computes $M''_{\mathcal{C}} = E\left[Z_{N+1} | \mathcal{C}, x_{N+1,j}\right]$. Similarly, the impostor function computes $M''_{\mathcal{I}} = E\left[Z_{N+1} | \mathcal{I}, x_{N+1,j}\right]$. The conciliated overall score M'' takes into account the different expertise of the two fusion functions and chooses the one which came closest to its goal, i.e. 0 for the impostor function and 1 for the client function:

$$M'' = \begin{cases} M''_{\mathcal{C}} & \text{if } \left|1 - M''_{\mathcal{C}}\right| - \left|0 - M''_{\mathcal{I}}\right| < 0 \\ M''_{\mathcal{I}} & \text{otherwise} \end{cases}. \tag{3.27}$$

Based on the normality assumption of the errors, the fusion training and testing algorithm described by Bigun *et al.* [1997a] is obtained, see Bigun [1995] for further background and details. Here we summarize the resulting algorithm in the two cases where it can be applied.

3.2.2.2. Bayesian Simplified Score Fusion Algorithm

When only the similarity scores x_{ij} are available, the following simplified fusion function is obtained by using $v_{ij} = 1$:

Training. Estimate the bias parameters of each system. The bias parameters for the client function are

$$M_{\mathcal{C}j} = \frac{1}{n_{\mathcal{C}}} \sum_i \zeta_{ij} \quad \text{and} \quad V_{\mathcal{C}j} = \frac{\alpha_{\mathcal{C}j}}{n_{\mathcal{C}}}, \tag{3.28}$$

where i indexes the training set $\mathcal{C}$, $n_{\mathcal{C}}$ is the number of training samples in $\mathcal{C}$ and

$$\alpha_{\mathcal{C}j} = \frac{1}{n_{\mathcal{C}} - 3} \left(\sum_i \zeta_{ij}^2 - \frac{1}{n_{\mathcal{C}}} \left(\sum_i \zeta_{ij} \right)^2 \right). \tag{3.29}$$

Similarly $M_{\mathcal{I}j}$ and $V_{\mathcal{I}j}$ are obtained for the impostor function.

Authentication. At this step, both fusion functions are operational, so that the time instant is $N+1$ and the fusion functions have access to the similarity scores $x_{N+1,j}$ but not to the true label z_{N+1}. First the client and impostor functions are calibrated according to their past performance, yielding (for the client function)

$$M'_{\mathcal{C}j} = x_{n+1,j} + M_{\mathcal{C}j} \;\text{ and }\; V'_{\mathcal{C}j} = (n_{\mathcal{C}} + 1) V_{\mathcal{C}j}, \tag{3.30}$$

and then the different calibrated systems are combined according to

$$M''_{\mathcal{C}} = \frac{\sum_{j=1}^{M} \frac{M'_{\mathcal{C}j}}{V'_{\mathcal{C}j}}}{\sum_{j=1}^{M} \frac{1}{V'_{\mathcal{C}j}}}. \tag{3.31}$$

Similarly, $M'_{\mathcal{I}}$, $V'_{\mathcal{I}}$ and $M''_{\mathcal{I}}$ are obtained. The final fused output is obtained according to Eq. (3.27).

The algorithm described above has been successfully applied by Bigun *et al.* [1997c] in a multimodal authentication system combining face and speech data. Verification performance improvements of almost an order magnitude were reported as compared to the best modality.

3.2.2.3. Bayesian Quality-Based Score Fusion Algorithm

When not only the scores but also the score variances are available, the following algorithm is obtained:

Training. Estimate the bias parameters. For the client function

$$M_{\mathcal{C}j} = \frac{\sum_i \frac{\zeta_{ij}}{\sigma_{ij}^2}}{\sum_i \frac{1}{\sigma_{ij}^2}} \qquad \text{and} \qquad V_{\mathcal{C}i} = \frac{1}{\sum_i \frac{1}{\sigma_{ij}^2}}, \tag{3.32}$$

where the training set $\mathcal{C}$ is used. The variances σ_{ij}^2 are estimated through $\bar{\sigma}_{ij}^2 = v_{ij} \cdot \alpha_{\mathcal{C}j}$, where

$$\alpha_{\mathcal{C}j} = \frac{1}{n_{\mathcal{C}} - 3} \left(\sum_i \frac{\zeta_{ij}^2}{v_{ij}} - \left(\sum_i \frac{\zeta_{ij}}{v_{ij}} \right)^2 \left(\sum_i \frac{1}{v_{ij}} \right)^{-1} \right). \tag{3.33}$$

Similarly $M_{\mathcal{I}j}$ and $V_{\mathcal{I}j}$ are obtained for the impostor function.

Authentication. First we calibrate the systems according to their past performance, for the client function

$$M'_{\mathcal{C}j} = x_{N+1,j} + M_{\mathcal{C}j} \;\text{ and }\; V'_{\mathcal{C}j} = v_{N+1,j}\alpha_{\mathcal{C}j} + V_{\mathcal{C}j}, \tag{3.34}$$

and then the different calibrated systems are combined according to Eq. (3.31). Similarly, $M'_{\mathcal{I}}$, $V'_{\mathcal{I}}$ and $M''_{\mathcal{I}}$ are obtained. The final fused score is obtained according to Eq. (3.27). This combined output can be expressed in the form of Eq. (2.11).

The algorithm described above has been successfully applied by Bigun [1995] in a risk assessment study related to aircraft accidents.

The variance v_{ij} of the score x_{ij} concerns a particular authentication assessment. It is not a general reliability measure for the system itself, but a certainty measure based on the performance of the system and the data being assessed. Typically the variance of the score is chosen as the width of the range in which one can place the score when considering human opinions. Because such intervals can be conveniently provided by a human expert, the algorithm presented here constitutes a systematic way of combining human and machine expertise in an authentication application. An example of such an application is forensic reporting using biometric evidences, where machine expert approaches are increasingly being used [Gonzalez-Rodriguez *et al.*, 2005] and human opinions must be taken into consideration.

In this Thesis, we propose to calculate v_{ij} for a machine expert by using a quality measure of the input biometric signal (see Fig. 3.5). This implies taking into account Eq. (3.34) right, that the trained fusion function adapts the weights of the experts using the input signal quality. First we define the quality q_{ij} of the score x_{ij} according to

$$q_{ij} = \sqrt{Q_{ij} \cdot Q_{\text{claim},j}}, \tag{3.35}$$

where Q_{ij} and $Q_{\text{claim},j}$ are the quality label of the biometric trait j in trial i and the average quality of the biometric signals used by the system j for modelling the claimed identity respectively. The two quality labels Q_{ij} and $Q_{\text{claim},j}$ are supposed to be in the range $[0, Q_{\max}]$ with $Q_{\max} > 1$, where 0 corresponds to the poorest quality, 1 corresponds to normal quality and $Q_{\max}$ corresponds to the highest quality. Finally, the variance parameter is calculated according to

$$v_{ij} = \frac{1}{q_{ij}^2}. \tag{3.36}$$

3.2.3. Discriminative Quality-Based Score Fusion Algorithm

Let $\mathbf{q} = [q_1, \ldots, q_M]^T$ denote the quality vector of the multimodal similarity score $\mathbf{x} = [x_1, \ldots, x_M]^T$, where q_j is a scalar quality measure corresponding to the similarity score x_j with $j = 1, \ldots, M$ being M the number of modalities. As in the case of the Bayesian quality-based fusion algorithm, the quality values q_j are computed as follows:

$$q_j = \sqrt{Q_j \cdot Q_{\text{claim},j}}, \tag{3.37}$$

where Q_j and $Q_{\text{claim},j}$ are the quality measure of the sensed signal for biometric trait j, and the average signal quality of the biometric signals used by unimodal system j for modelling the claimed identity, respectively. The two quality labels Q_j and $Q_{\text{claim},j}$ are supposed to be in the range $[0, Q_{\max}]$ with $Q_{\max} > 1$, where 0 corresponds to the poorest quality, 1 corresponds to standard quality, and $Q_{\max}$ corresponds to the highest quality.

The proposed score-level fusion scheme based on SVM classifiers and quality measures is as follows:

Training. An initial fusion function ($f_{\text{SVM}} : \mathbb{R}^M \to \mathbb{R}, f_{\text{SVM}}(\mathbf{x}_T) = \langle \mathbf{w}, \Phi(\mathbf{x}_T) \rangle + w_0$) is trained by solving the problem

$$\min_{\mathbf{w}, w_0, \xi_1, \ldots, \xi_N} \left(\frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^{N} C_i \xi_i \right) \tag{3.38}$$

subject to

$$y_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i)\rangle_{\mathbb{H}} + w_0) \geq 1 - \xi_i, \quad i = 1, \ldots, N, \tag{3.39}$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N, \tag{3.40}$$

as described in Sect. 3.1.2.2, but using as cost weights

$$C_i = C \left(\frac{\prod_{j=1}^{M} q_{i,j}}{Q_{\max}^{M}} \right)^{\alpha_1}, \tag{3.41}$$

where $q_{i,j}$, $j = 1, \ldots, M$ are the components of the quality vector $\mathbf{q}_i$ associated with training sample $(\mathbf{x}_i, z_i)$, $z_i \in \{-1, 1\} = \{\text{Impostor}, \text{Client}\}$, and C is a positive constant. As a result, the higher the overall quality of a multimodal training score the higher its contribution to the computation of the initial fusion function. Additionally, M SVMs of dimension $M-1$ (SVM_1 to SVM_M) are trained leaving out traits 1 to M respectively. Similarly to Eq. (3.41)

$$C_i = C \left(\frac{\prod_{r \neq j} q_{i,r}}{Q_{\max}^{(M-1)}} \right)^{\alpha_1}, \tag{3.42}$$

for SVM_j with $j = 1, \ldots, M$.

Authentication. Let the sensed multimodal biometric sample generate a quality vector $\mathbf{q}_T = [q_{T,1}, \ldots, q_{T,M}]^T$. Re-index the individual traits in order to have $q_{T,1} \leq q_{T,2} \leq \ldots \leq q_{T,M}$. A multimodal similarity score $\mathbf{x}_T = [x_{T,1}, \ldots, x_{T,M}]'$ is then generated. The combined quality-based similarity score is computed as follows:

$$f_{\text{SVM}_Q}(\mathbf{x}_T) = \beta_1 \sum_{j=1}^{M-1} \frac{\beta_j}{\sum_{r=1}^{M-1} \beta_r} f_{\text{SVM}_j}(\mathbf{x}_T^{(j)}) + (1 - \beta_1) f_{\text{SVM}}(\mathbf{x}_T), \tag{3.43}$$

where $\mathbf{x}_T^{(j)} = [x_{T,1}, \ldots, x_{T,j-1}, x_{T,j+1}, \ldots, x_{T,M}]^T$ and

$$\beta_j = \left(\frac{q_{T,M} - q_{T,j}}{Q_{\max}} \right)^{\alpha_2}, \quad j = 1, \ldots, M-1. \tag{3.44}$$

As a result, the adapted fusion function in Eq. (3.43) is a quality-based trade-off between not using and using low quality traits.

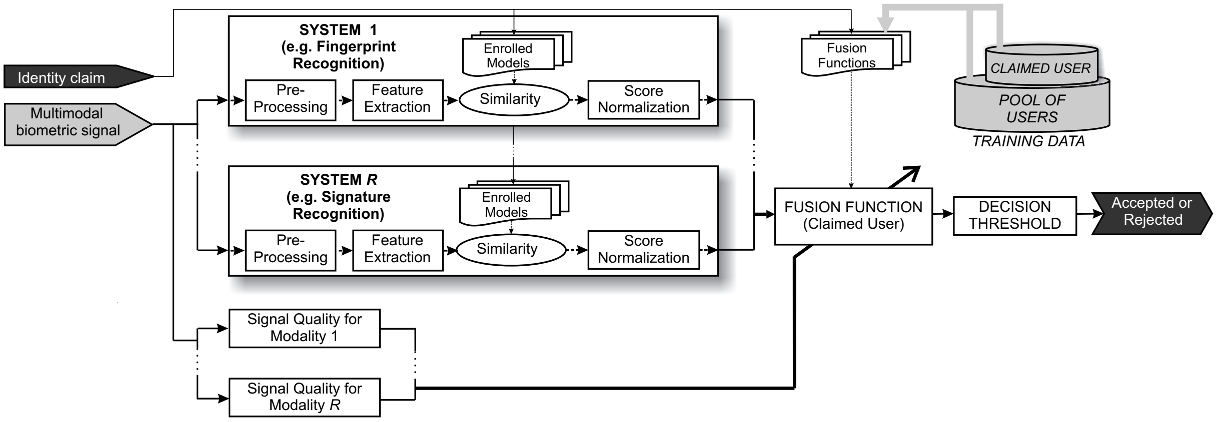


Figure 3.6: System model of multimodal biometric authentication with user-dependent and quality-based score fusion.

3.3. User-Dependent and Quality-Based Fusion

Finally, we also propose a system model for multimodal biometric authentication adapted both to the user specificities and to the input biometric quality, which is shown in Fig. 3.6.

Practical implementations of this scheme can be obtained by combining some of the procedures described in this chapter. One possibility is to use Bayesian user-dependent score fusion plus discriminative quality-based adaptation.

3.4. Chapter Summary and Conclusions

In this chapter we have described the score level fusion methods proposed in this Thesis. These methods are divided into three classes: 1) user-dependent, 2) quality-based, and 3) user-dependent and quality-based. User-dependent fusion methods have been further classified into: *i*) user-dependent score normalization plus simple fusion, *ii*) user-dependent score fusion, and *iii*) user-dependent decision. For each class of methods, we have firstly sketched the system model and then we have derived particular implementations based on generative assumptions or discriminative criteria.

All the methods presented in this chapter are original contributions, except the method proposed in Sect. 3.2.2 which is largely based on Bigun *et al.* [1997a].

Chapter 4

Performance Evaluation of Multimodal Biometric Systems

THIS CHAPTER summarizes the common practices in performance evaluation of biometric systems and describes the biometric databases used in this PhD Thesis.

The chapter is organized as follows. We first summarize the guidelines for performance evaluation followed in this Dissertation. We then provide an overview of the main existing multimodal biometric databases, together with some information on current efforts in the acquisition of new biometric corpora. Finally we describe the bimodal database used in the experiments reported in this Thesis.

This chapter is based on the publication: Ortega-Garcia *et al.* [2003b].

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.*, 2004b; Phillips *et al.*, 2000a]. 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.*, 2000b]; 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]; and the Signature Verification Competition (SVC), organized in 2004 [Yeung *et al.*, 2004]. 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]. In this environment, and as a result of the experience gained in these comparative evaluations, the UK Biometrics Working Group has recently 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.*, 2000a]: 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; Phillips *et al.*, 2000b; Przybocki and Martin, 2004; Yeung *et al.*, 2004].

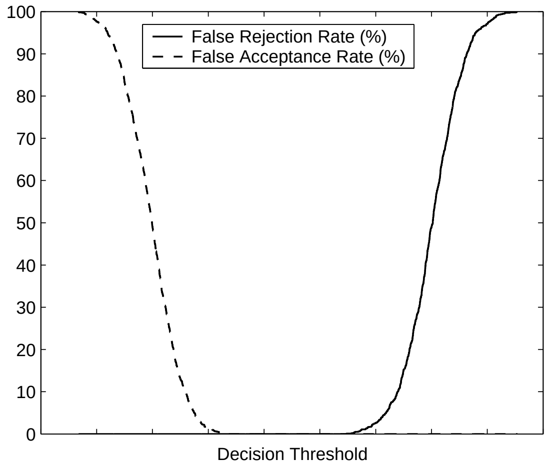
The goal of scenario evaluations is to measure the overall system performance for a prototype scenario that models an application domain. Scenario evaluations are conducted under conditions that model real-world applications [Bone and Blackburn, 2002; Mansfield *et al.*, 2001]. Because each system has 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 experiments as technology evaluations of different unimodal, multi-algorithm, and multimodal strategies for biometric authentication.

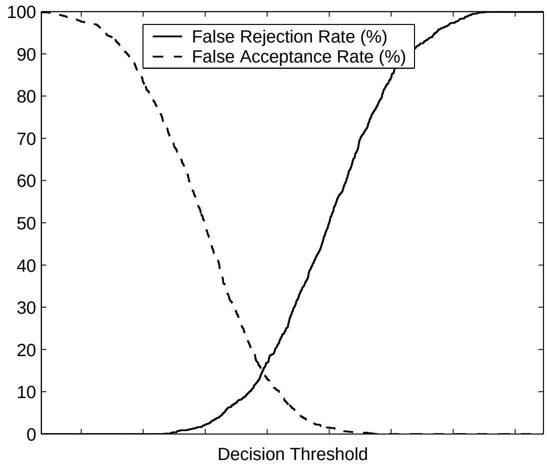
4.1.1. Performance Measures of Authentication Systems

Biometric technologies can be ranked according to several criteria, including [Jain *et al.*, 2004b]: universality, distinctiveness, permanence, collectability, performance, acceptability and circumvention, as it was mentioned in Sect. 1.2. In the experiments of this Thesis we concentrate on performance indicators to compare different methods, and more specifically on authentication error rates.

We do not consider other performance indicators strongly related to particular implementations and hardware/software architectures. These indicators include the computational efficiency, and the computational resources used in terms of storage and memory allocation [Cappelli *et al.*, 2006]. In this Thesis, basic implementations of the strategies studied have been tested on



(a)



(b)

Figure 4.1: *FA and FR curves for ideal (a) and real (b) authentication systems.*

a Pentium IV PC running Microsoft Windows XP.

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, is used in this Thesis [Martin *et al.*, 1997]. In this case, the use of a non-linear scale 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.

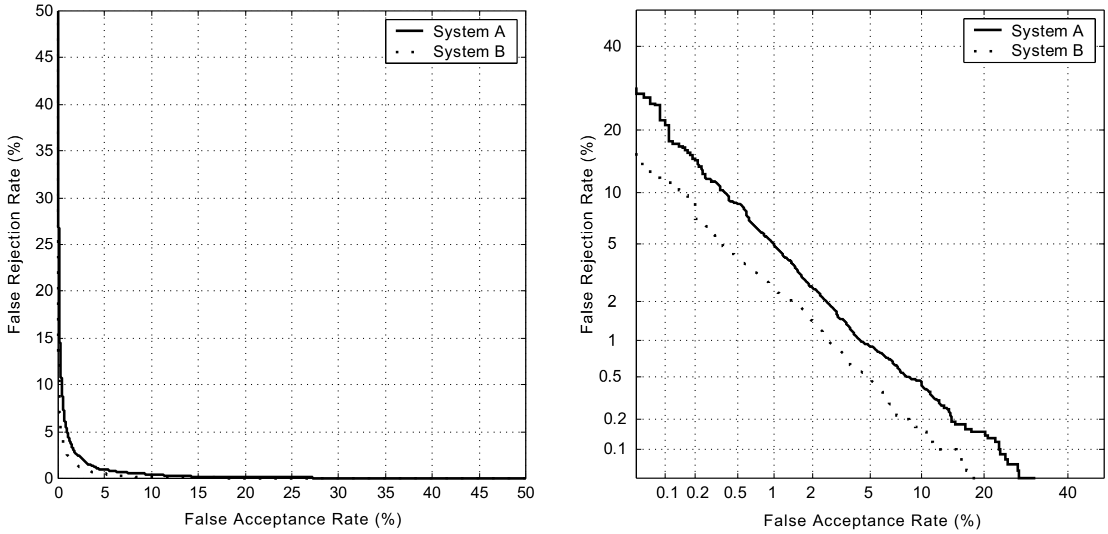


Figure 4.2: *Example of verification performance with ROC (left) and DET curves (right).*

4.1.2. Error Estimation Methods

In order to estimate FRR and FAR, a set of genuine and impostor matching scores have to be generated using the available biometric data. Several methods have been described in the literature in order to exploit the information embedded in the training samples during a test [Jain *et al.*, 2000a; Theodoridis and Koutroumbas, 2003]. Regarding multimodal fusion, some of the methods used include resubstitution [Jain and Ross, 2002], holdout [Kumar and Zhang, 2003; Toh *et al.*, 2004a; Wang *et al.*, 2004], and variants of jackknife sampling using the leave-one-out principle [Bigun *et al.*, 1997a]. In this PhD Thesis, and depending on the experiment at hand, we use one of the following methods or variants [Duda *et al.*, 2001]:

- Resubstitution: all the available data is used for training as well as testing.
- Rotation: this is a version of *cross-validation*. Regarding the available data for each target, the reference model is designed by choosing k consecutive samples as the design set, and the remaining samples constitute the test set; this is repeated for all distinct choices of k consecutive observations. When k is chosen to be equal to the number of samples minus one, the *leave-one-out* procedure is obtained [Jain *et al.*, 2000a]. Note that the rotation approach can also be applied for selecting users from the available ones.
- Bootstrap: a fixed number of samples from the available training data is chosen randomly with replacement (i.e., the same sample can be chosen multiple times). The remaining samples constitute the test set. The procedure is repeated a fixed number of times. Note that bootstrap can be applied either for selecting users from the available database or for selecting training samples within a specific user.

The former strategy leads to optimistically biased estimates whereas the later two give unbiased estimates with larger computational requirements.

When dealing with user-dependent learning, we are confronted with severe data scarcity. Toh *et al.* [2004a] overcome this by augmenting the training set with noisy samples. In this Thesis, we use bootstrap sampling [Bolle *et al.*, 2004b; Duda *et al.*, 2001].

4.1.3. Statistical Significance of Performance Results

Guyon *et al.* [1998] derived the minimum size of the test data set, N, that guarantees statistical significance in a pattern recognition task. The goal was to estimate N so that it is guaranteed, with a risk α of being wrong, that the error rate P does not exceed the estimated from the test set, $\hat{P}$, by an amount larger than $\varepsilon(N, \alpha)$, that is

$$\Pr\left\{P > \hat{P} + \varepsilon(N, \alpha)\right\} < \alpha. \tag{4.1}$$

Letting $\varepsilon(N, \alpha) = \beta P$, and supposing recognition errors as Bernoulli trials [Papoulis, 1991], we can derive the following relation

$$N \approx \frac{-\ln \alpha}{\beta^2 P}. \tag{4.2}$$

For typical values of α and β (0.05 and 0.2, respectively), the following simplified criterion is obtained

$$N \approx 100/P. \tag{4.3}$$

If the samples in the test data set are not independent (due to correlation factors including the recording conditions, some types of sensors, certain groups of users, etc.) then N must be further increased. The reader is referred to Guyon *et al.* [1998] for a detailed analysis of this intricate case, where some guidelines for computing the correlation factors are given. Another approach dealing with correlated errors is described by Guillick and Cox [1989].

4.2. Multimodal Biometric Databases

One key element for performance evaluation of biometric systems is the availability of biometric databases. The availability of multimodal biometric features corresponding to a large population of individuals, together with the desirable presence of biometric variability of each trait (i.e., multi-session, multiple acquisition sensors, different signal quality, etc.), makes database collection a time-consuming and complicated process, in which a high degree of co-operation of the donators is needed. Additionally, the legal issues regarding data protection are controversial [Wayman *et al.*, 2005]. For these reasons, nowadays, the number of existing public multimodal biometric databases is quite limited.

Due to the difficulties in multimodal database collection, some authors have assumed independence between different biometric traits and have performed their experiments on multimodal databases combining biometric signals from different users, thus creating chimeric subjects [Poh and Bengio, 2005a]. The recent trend, on the other hand, and as recommended by best practices, is to conduct the performance evaluations on real multimodal biometric data. This is the approach followed in this Thesis.

The multimodal databases 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 national projects like the French BIOMET [Garcia-Salicetti *et al.*, 2003] or the Spanish MCYT [Ortega-Garcia *et al.*, 2003b]. Other ongoing efforts in multimodal database collection include the BioSec multimodal database [BioSec, 2004], and the database activities of the Biosecure Network of Excellence [Biosecure, 2004].

Multimodal Biometric Databases can be broadly classified into two groups: 1) databases of multimodal biometric signals, and 2) databases of multimodal scores. In the first class the collected data are biometric signals, such as fingerprint images or voice utterances. These signals may be used with a variety of different experimental protocols, both for individual system development and for multimodal experiments at any fusion level (i.e., sensor, feature, or score level). The second class of multimodal databases are intended exclusively for multimodal research based on score fusion. These corpora consist of matching scores from the individual traits considered.

In the following sections we provide an overview of existing multimodal databases of both classes, and provide some information of current efforts in the acquisition of new corpora.

4.2.1. Existing Multimodal Databases

BT-DAVID. The BT-DAVID database contains full-motion video, showing a full-face and a profile view of talking subjects, together with the associated synchronous sound [Chibelushi *et al.*, 1999]. BT-DAVID includes audio-visual material from more than 100 subjects including 30 clients recorded on 5 sessions spaced over several months. The utterances include the English digit set, English alphabet E-set, vowel-consonant-vowel syllables, and phrases for the control of a video-conferencing session. The scenes include variable scene background complexity and illumination. Portions of the database include lip highlighting.

XM2VTS. The XM2VTS database (extended M2VTS) was acquired in the context of the M2VTS project (Multi Modal Verification for Teleservices and Security applications), a part of the EU ACTS programme, which deals with access control by the use of multimodal identification based on face and voice [Messer *et al.*, 1999]. The database contains microphone speech and face image from 295 people. Every subject recorded 4 sessions over a period of 4 months. At each session two head rotation shots and six speech shots (subjects reading three sentences twice) were recorded. The XM2VTS evaluation protocol specifies training, evaluation, and test sets, so that detailed comparisons between algorithms are possible. A variety of subsets of the database are available for purchase from the University of Surrey. To date, the XM2VTS database has been distributed to more than 100 institutions.

BANCA. The BANCA database is a large, realistic and challenging multimodal database intended for training and testing multimodal verification systems [Bailly-Bailliere *et al.*, 2003]. The BANCA database was captured in four European languages and two modalities (face and voice). For recording, both high and low quality microphones and cameras were used. The subjects were recorded in three different scenarios, controlled, degraded, and adverse over 12 different sessions spanning three months in time. In total 208 people were captured, half men and half women. For each recording the subject was instructed to speak a random 12-digit number along with a name, address, and date of birth (client or impostor data). Recordings took an average of 20 seconds. An associated BANCA evaluation protocol is also available.

BIOMET. Five different modalities are present in the BIOMET database [Garcia-Salicetti *et al.*, 2003]: audio, face image, hand image, fingerprint and signature. For the face images, besides a conventional digital camera, a camera designed to suppress the influence of the ambient light is also used. Three different sessions have been realized, with three and five months spacing between them. The number of persons participating to the collection of the database was 130 for the first campaign, 106 for the second, and 91 for the last one. The proportion of females and males was balanced for all the campaigns. 10% of the persons were students (with an average age of 20). The age of the others varies from 35 up to 60 years.

NIST-BSSR1. NIST Biometric Scores Set (BSSR1) is a set of raw output similarity scores from two 2002 face recognition systems and one 2004 fingerprint system, operating on frontal faces, and left and right index live-scan fingerprints, respectively [NIST, 2004]. The release includes true multimodal score data, i.e. similarity scores from comparisons of faces and fingerprints of the same people. This database is available upon request. The data are suited to the study of score-level fusion in multimodal, multi-algorithm, multi-instance and repeated-instance multibiometrics. The database contains three partitions: set 1 is comprised of face and fingerprint scores from the same set of 517 individuals. For each individual, the set contains one score from the comparison of two right index fingerprints, one score from the comparison of two left index fingerprints, and two scores (from two separate matchers) from the comparison of two frontal faces. Set 2 is comprised of fingerprint scores from one system run on images of 6000 individuals. For each individual, the set contains one score from the comparison of two left index fingerprints, and another from two right index fingerprints. Set 3 contains scores from two face systems run on images from 3000 individuals. For each individual, the set contains one score from the comparison of face A with a later face, B, and a score from face A and another later face, C.

IDIAP. This score database [Poh and Bengio, 2006] is built on XM2VTS [Messer *et al.*, 1999], respecting the standard Protocols I and II (LP1 and LP2). LP1 has 8 baseline systems and LP2 has 5 baseline systems. The score database has two fusion protocols: 1) fusion of two experts with specific combinations in order to permit experiments on multimodal fusion, intramodal fusion with different feature sets, and intramodal fusion with the same feature; and 2) fusion with all the possible combinations across protocols.

4.2.2. Multimodal Databases Under Development

MyIdea. The MyIdea database is being acquired in the framework of a collaboration between the University of Fribourg in Switzerland, the Engineering School of Fribourg in Switzerland and the Groupe des Ecoles des Télécommunications in Paris. MyIdea database includes face, audio, fingerprints, signature, handwriting and hand geometry [Dumas *et al.*, 2005]. Further to the independent acquisition of each modality, two synchronized recordings are performed: face-voice and writing-voice. The general specifications of MyIdea are: target of 104 subjects, different quality of sensors, various realistic acquisition scenarios, and organization of the recordings to allow for open-set experimental scenarios.

BioSec. BioSec is an Integrated Project of the 6th European Framework Programme [BioSec, 2004]. The project is aimed at integrating biometrics and security to leverage trust and confidence in a wide spectrum of everyday applications. Over 20 partners from nine European countries participate in the project, including big companies, biometric HW/SW producers, and prestigious universities. Universidad Politécnica de Madrid (where ATVS was formerly until 2004) is in charge of the database activities carried out within BioSec, one of which is the design and acquisition of a new multimodal database. BioSec database includes face images, short speech utterances (low and high quality microphones), fingerprint images (3 different sensors) and iris images from 250 subjects acquired in 4 acquisition sessions. An initial subcorpus of 200 subjects acquired in 2 acquisition sessions is already available [Fierrez-Aguilar, 2005]. Part of this Thesis has been originated from work in this project.

Biosecur-ID. Biosecur-ID is a coordinated project funded by the Spanish Ministerio de Ciencia y Tecnología [Biosecur ID, 2003]. Five academic partners from Spain participate in the project under the coordination of ATVS–Universidad Politécnica de Madrid. One of the objectives of the project is to build a new multimodal database. The new database will consist of the following biometric traits: face, fingerprint, voice, iris, written signature, handwriting, keystroking, palmprint, and hand geometry. The database will be available by late 2006. Part of this Thesis has been originated from work in this project.

BioSecure. Biosecure is a Network of Excellence of the 6th European Framework Programme [Biosecure, 2004]. The project is aimed at coordinating the different research efforts focused on biometrics across Europe. Over 30 research institutions from over 15 countries participate in the network. Universidad Politécnica de Madrid is in charge of the database activities carried out within the network, one of which is the design and acquisition of a new multimodal database to be conducted during 2006.

4.3. MCYT Bimodal Biometric Database

The Biometric Research Lab.–ATVS at the Universidad Politécnica de Madrid promoted the plan and led the development of the Spanish Ministerio de Ciencia y Tecnología (MCYT) project TIC00-1669-C04, in which the design and acquisition of a biometric bimodal database was accomplished [Ortega-Garcia *et al.*, 2003b]. Part of the work described in this Thesis has been originated within this project.

Although there are some other commercial and forensic partners within the project, the participation in the acquisition campaign has been conducted by a consortium of four academic institutions, namely: Universidad Politécnica de Madrid (UPM), Universidad de Valladolid (UVA), Universidad del País Vasco (EHU), and Escuela Universitaria Politécnica de Mataro, Barcelona (EUPMT).

The database acquired in the MCYT project, referred to as MCYTDB from now on, includes fingerprint and signature modalities of each individual enrolled in the database, and a significant number of samples of each modality under different levels of control. The scope of utility of the database involves mainly the performance assessment in the design of automatic recognition systems in civil, commercial and forensic applications, allowing the development and evaluation of biometric recognition algorithms based on single biometric features, and multibiometrics scenarios including multi-sensor, multi-algorithm, multi-instance, repeated instance, and multimodal.

The significance and utility of the database have been optimized by maximizing the number of contributors and biometric samples per contributor, outperforming the figures of existing databases. The number of contributors at each acquisition site are 35, 75, 75 and 145, acquired respectively at EUPMT, UVA, EHU and UPM. The total number of subjects in the database is 330.

4.3.1. Description of MCYTDB fingerprint corpus

The fingerprint databases most currently used for research purposes are the fingerprint corpora by NIST [NIST, 2005], available upon request, and the databases captured in the series of Fingerprint Verification Competitions [Cappelli *et al.*, 2006], available in DVD [Maltoni *et al.*, 2003]. The former corpora include scanned images from plain and rolled fingerprints previously acquired with ink, latent fingerprints, as well as digital videos of live-scan fingerprint data. Already extracted minutiae and manual classification is also available in some cases. The FVC databases include on-line fingerprint images captured with a number of sensors of different technologies.

For the acquisition of the MCYTDB fingerprint corpus two types of acquisition devices were used: 1) a CMOS-based capacitive capture device, model 100SC from Precise Biometrics, with resolution of 500 dpi, and 2) an optical device, model UareU from Digital Persona, also with resolution of 500 dpi. In both cases a ten-print acquisition per individual was carried out. Each input generates a bitmap file representing the image of the fingerprint, in an 8-bit grayscale. The file sizes and the image resolutions are: 1) 89 kB and 300×300 pixels, in the case of the capacitive device, and 2) 102 kB and 256×400 pixels (width $\times$ height), in the case of the optical device.

The whole process of fingerprint acquisition is accomplished under the supervision of an operator following a fixed protocol. A software viewer on the PC screen is available to the operator in order to help in controlling the finger position on the sensor. Based on the core and/or delta position, the operator decides when the acquisition and storage of the image are valid.

With the aim of evaluating automatic recognition systems under different acquisition conditions, the fingerprint corpus includes 12 different impressions of each fingerprint, under different levels of control. The acquisition control is accomplished in three levels, namely:

- Three samples with **low level of control**: the contributor puts the finger on the screen sensor without any position restrictions, without watching the viewer. The operator must regard that at least one core and/or one delta of the fingerprint fall into the restricted area delimited by the rectangle of the interface viewer.

- Three more samples with **medium level of control**: in this stage, the individual must observe the computer screen while the finger is located on the sensor. The image must be centered into the new rectangle of smaller size which appears on the interface viewer.

- Six more samples with **high level of control**: the acquisition is accomplished as in the above stage, but the rectangle has now a smaller size. In this case, the position restrictions are more severe, and one core and/or delta of the fingerprint must always fall under this rectangle.

As a result, each individual provides a total number of 240 fingerprint images to the database (2 sensors $\times$ 12 impressions $\times$ 10 fingers).

Fig. 4.3 shows three impressions belonging to the same finger, acquired both with the optical scanner (top) and the capacitive scanner (bottom) under the three levels of control described above (from left to right). More examples of images included in the MCYTDB fingerprint corpus are depicted in Fig. 4.4.

Additionally, a subjective quality assessment has been accomplished for a total number of 9000 images (all optical samples from a subset of 75 subjects from UPM) [Simon-Zorita *et al.*, 2003]. Basically, each different fingerprint image is assigned an integer subjective quality measure from 0 (lowest quality) to 9 (highest quality) based on image factors like: captured area of the fingerprint, pressure, humidity, dirtiness, etc. By considering this manual quality measures, it

Figure 4.3: Three impressions belonging to a given finger, acquired both with the optical scanner (top) and the capacitive scanner (bottom) under the three levels of control considered in the MCYTDB fingerprint corpus (from left to right).

can be stated that about 5% of the acquired images are of very bad quality; 20% are of low quality; 55% are of medium quality; and 20% are of high quality. The significant percentage of low quality images is due to different factors which appear in the acquisition process: lacks of impression in the image due to the adverse skin conditions (scares, scratches, marks, humidity, dirtiness, etc), the particular configuration of the ridges in some fingers, the excess of pressure applied on the screen sensor, the background noise introduced by the acquisition device, and even the non-cooperative attitude of a few number individuals, producing in these cases a not well defined structure of the ridges. Fig. 4.5 shows four example images and their labeled quality.

4.3.2. Description of MCYTDB signature corpus

A number of different on-line signature databases are currently available [Dolfing, 1998; Garcia-Salicetti *et al.*, 2003; Munich and Perona, 2003], including a subset of the database used in the recent International Signature Verification Competition [Yeung *et al.*, 2004]. In spite of these databases, no clear agreement on the use of a common benchmark has been reached in the on-line signature community. In this respect, the Biosecure NoE [Garcia-Salicetti *et al.*, 2006]

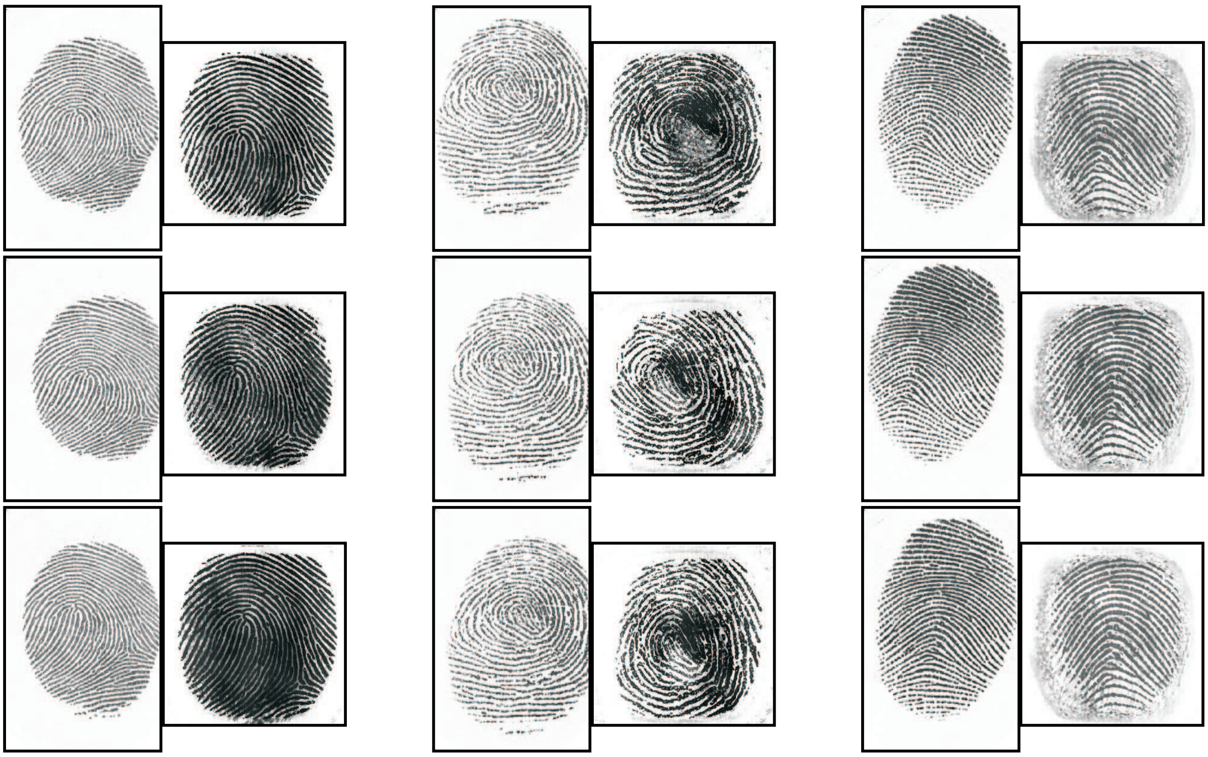


Figure 4.4: *Fingerprint examples from MCYTDB fingerprint corpus. A different fingerprint is depicted in each column. Optical and capacitive sensors correspond to the left and right images of each subplot, respectively. Different impressions of each fingerprint are given in different rows.*

Figure 4.5: *Fingerprint images from the MCYTDB fingerprint corpus. Quality label from left to right: 0 (minimum), 3, 6, and 9 (maximum).*

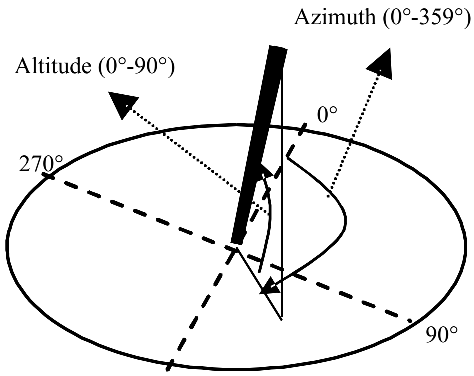


Figure 4.6: *Azimuth and inclination angles of the pen with respect to the plane of the pen tablet Intuos from Wacom.*

is pushing forward to establish common evaluation procedures using the MCYTDB signature corpus described below.

Signature information in MCYTDB was acquired by using an inking pen and paper templates over a WACOM pen tablet, model INTUOS A6 USB. This procedure enables to capture both the signature images and other on-line signature data (pen trajectories, pressure signals and pen inclination signals versus time). Each signature was written within a 3.75×1.75 frame (width $\times$ height in cm.).

The tablet resolution is 2540 lines per inch (100 lines/mm), and the precision is ± 0.25 mm. The maximum detection height is 10 mm (so also pen-up movements are considered), and the capture area is 127×97 mm. This tablet provides the following discrete-time dynamic sequences (dynamic range of each sequence is specified): 1) position in x-axis: $[0-12700]$, corresponding to $0-127$ mm, 2) position in y-axis: $[0-9700]$, corresponding to $0-97$ mm, 3) pressure applied by the pen: $[0-1024]$, 4) azimuth angle of the pen with respect to the tablet (see Fig. 4.6): $[0-3600]$, corresponding to $0^o - 360^o$, and 5) altitude angle of the pen with respect to the tablet (see Fig. 4.6): $[300-900]$, corresponding to $30^o - 90^o$. The sampling frequency of the acquired signals is set to 100 Hz, enough to avoid aliasing taking into account the biomechanics of the hand [Baron and Plamondon, 1989].

Each target user produces 25 genuine signatures and is imitated 25 times. These forgeries are produced by the 5 subsequent target users by observing the static images of the signature to imitate, trying to copy it (at least 10 times), and then producing the forgeries in a natural way (i.e., without artifacts such as breaks or slowdowns). In this way, shape-based skilled forgeries with natural dynamics are obtained. Following this procedure, user n produces a set of 5 samples of her or his genuine signature, and then 5 skilled forgeries of client $n-1$. Then, again a new set of 5 samples of her genuine signature; and then 5 skilled forgeries of user $n-2$; this procedure is repeated 5 times. Summarizing, user n produces 25 samples of her own signature (in sets of 5 samples) and 25 skilled forgeries (5 samples of each user, $n-1$ to $n-5$).

The detection and segmentation of the input signature is automatically accomplished by the acquisition software. The signature start up is determined by the first sample in which the pen contacts the tablet, i.e., the first sample with non-zero pressure value. The signature ending is determined by setting a 3-second timer to the first zero pressure sample found (i.e., a pen up). If no samples with non-zero pressure value are detected in this interval, the capture process is stopped, and the complete signature is stored. Otherwise, the timer is reset until the next pen up is found.

One example signature from MCYTDB signature corpus with its corresponding on-line signals is depicted in Fig. 4.7. More example signatures from different users are depicted in Fig. 4.8.

The MCYTDB signature corpus was released in 2003 by the Biometrics Research Lab.–ATVS [BRL, 2006] and it is used in more than 30 research groups worldwide [Hongo *et al.*, 2005; Igarza *et al.*, 2005; Muramatsu *et al.*, 2006; Nanni and Lumini, 2006; Richiardi and Drygajlo, 2003].

4.3.2.1. Off-line signature subcorpus

Paper templates of 75 signers (and their associated skilled forgeries) have been selected and digitized with an scanner at 600 dpi [Fierrez-Aguilar *et al.*, 2004a]. Resulting subcorpus comprises 2250 signature images, with 15 genuine signatures and 15 forgeries per user (contributed by 3 different user-specific forgers). Some examples of genuine signatures (left and central columns) and forgeries (right column) are given in Fig. 4.9 for the four types of signatures encountered in the MCYTDB signature corpus, namely: simple flourish, complex flourish, name with simple flourish and name with complex flourish.

This off-line signature subcorpus is also available through the Biometrics Research Lab.–ATVS [BRL, 2006].

4.4. SVC2004 Signature Database

The common practice in signature verification research is to evaluate the proposed methods on small data sets acquired at the different research laboratories [Jain *et al.*, 2002]. In this environment, the First International Signature Verification Competition (SVC 2004) was organized providing a common reference for system comparison on the same data and evaluation protocol [Yeung *et al.*, 2004].

Because one of the contributions of this Thesis is the development of novel approaches for on-line signature verification, and SVC 2004 is the only public benchmark with comparative results from a number of state-of-the-art systems already available, we will also use SVC data to assess our systems.

Development corpus of the extended task (including coordinate and timing information, pen orientation and pressure) is used in this Dissertation. This corpus consists of 40 sets of signatures. Each set contains 20 genuine signatures from one contributor (acquired in two separate sessions) and 20 skilled forgeries from five other contributors. The signatures are mostly in either English or Chinese. Some examples are depicted in Fig. 4.10 for two different targets of the data set. Plots of the coordinate trajectories, pressure signal and pen orientation functions are also given.

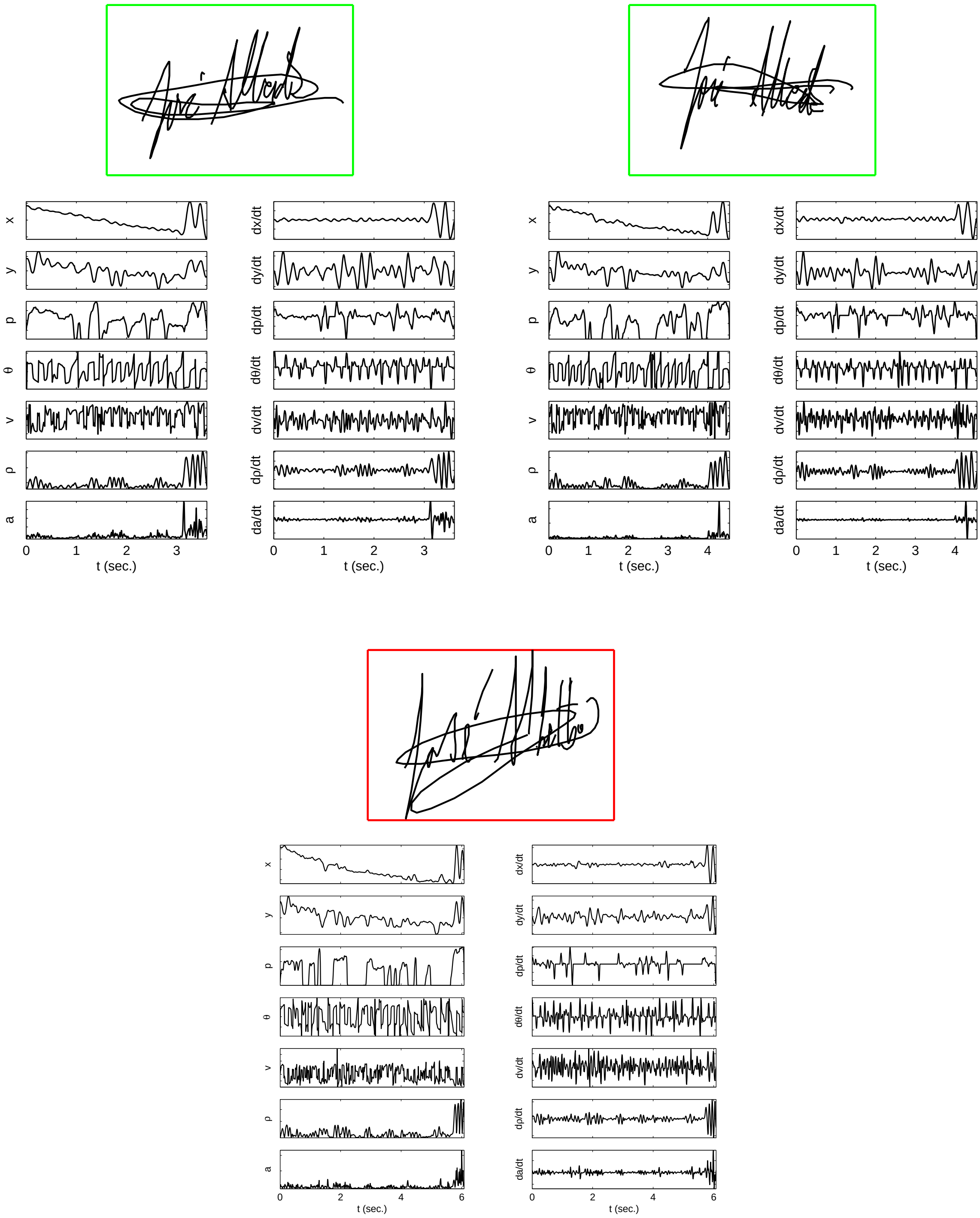


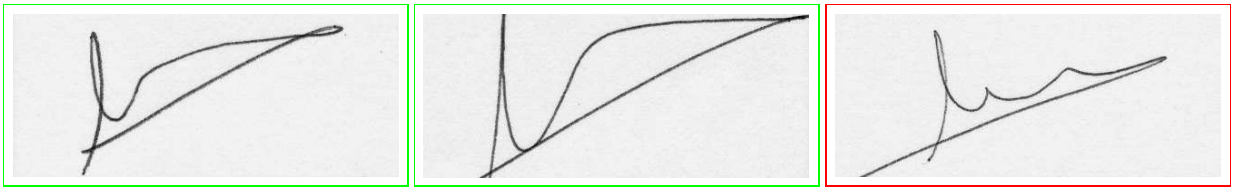
Figure 4.7: *Two genuine signatures (top) and one skilled forgery (bottom) of a given user. The function-based representation of the local system presented in Chapter 5 is depicted below each signature.*

Figure 4.8: *Signature examples from MCYTDB signature corpus. Each row corresponds to a different user. The two left signatures are genuine and the right one is a skilled forgery.*

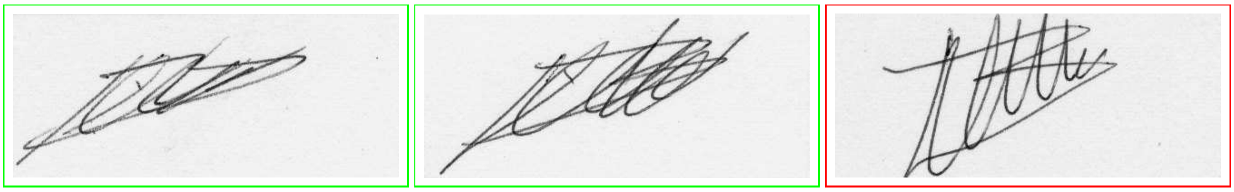
4.5. Chapter Summary and Conclusions

In this chapter we have outlined some guidelines for performance evaluation in biometric authentication following best practices. We have also provided an overview of the main existing multimodal biometric databases, together with some information on current efforts in the acquisition of new biometric corpora. Finally we have described the databases used in this Thesis, namely: MCYTDB consisting of fingerprint images and written signatures (both on-line time sequences and off-line images), and the signature database from the SVC 2004 competition.

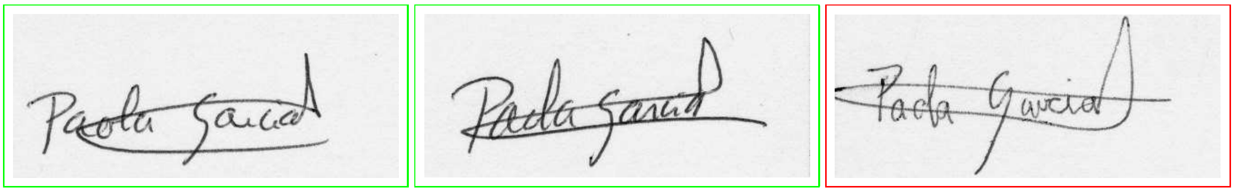
This chapter includes contributions in the survey of the existing multimodal databases, and in the description of the new bimodal corpus MCYT.



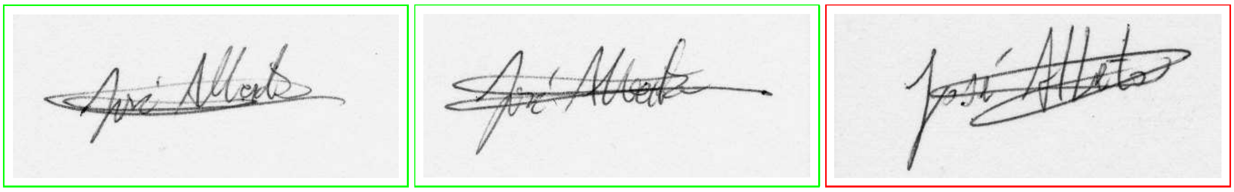
(a) Signature consisting of simple flourish.



(b) Signature consisting of complex flourish.

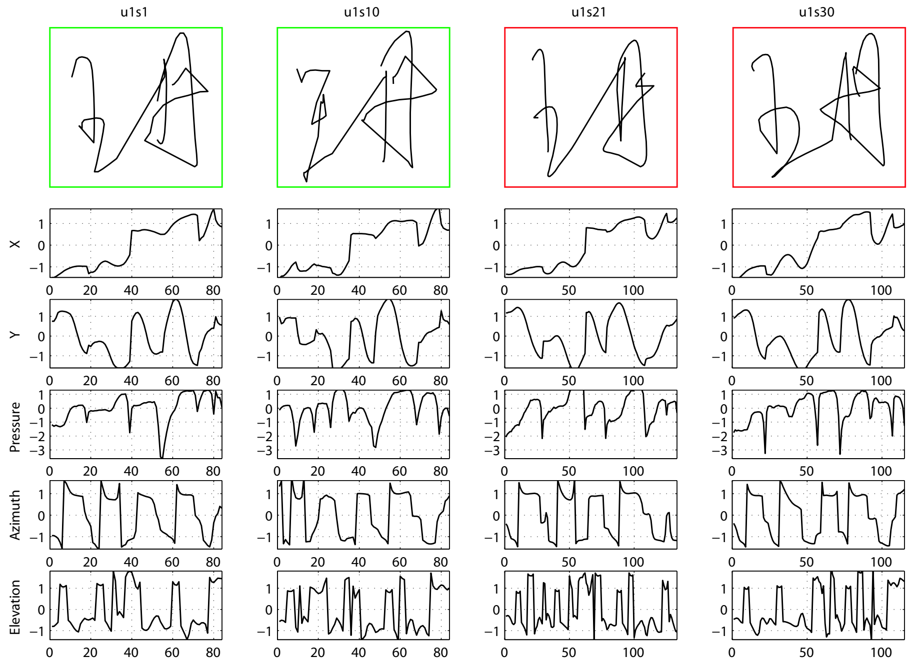


(c) Signature consisting of written name and simple flourish.

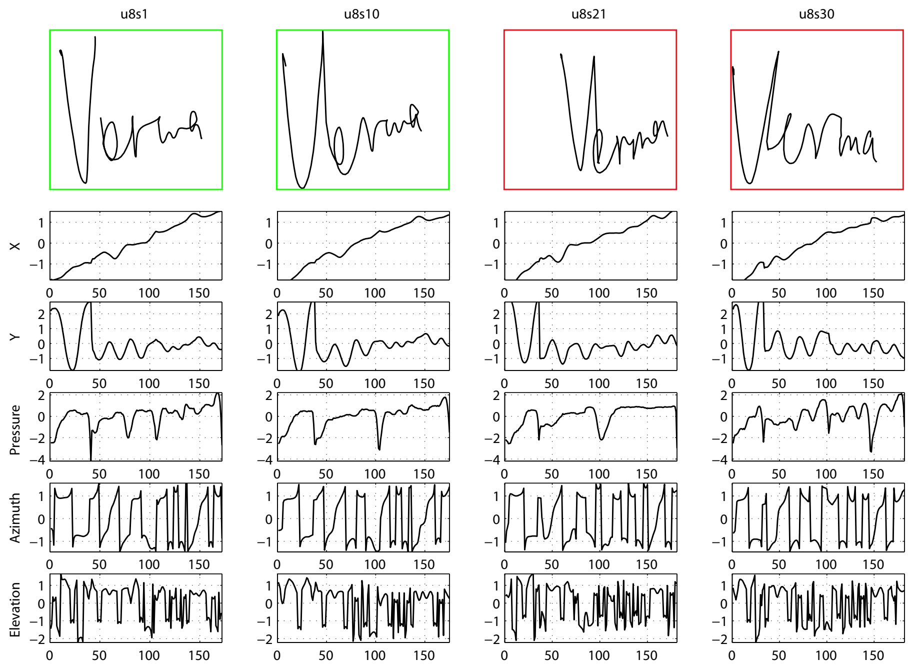


(d) Signature consisting of written name and complex flourish.

Figure 4.9: *Examples from the MCYTDB off-line signature corpus. Genuine signatures (left and central columns) and skilled forgeries (right column) are depicted for the four types of signatures encountered in MCYTDB.*



(a) User 1 (u1).



(b) User 8 (u8).

Figure 4.10: *Signature examples from SVC 2004 corpus. For each one of targets u1 (a) and u8 (b), two genuine signatures (left columns) and two skilled forgeries (right columns) are given.*

Chapter 5

Multi-Algorithm Signature Verification

THIS CHAPTER studies the application of user-dependent score normalization and user-dependent decision to multi-algorithm signature verification.

As indicated in Sect. 3.1, user-dependent multimodal authentication can be achieved by making user dependent each one of the following three modules in the general system model depicted in Fig. 3.1: 1) score normalization, 2) score fusion, and 3) decision. The adaptation to user specificities of the score fusion functions will be applied in Chapter 6 to multi-algorithm speaker verification.

The multi-algorithm on-line signature verification approach used in this chapter is based on two recognition levels exploiting local and global information, respectively. On-line here refers to the acquisition of the time functions of the written signature process, e.g., pen trajectory versus time. Local information is extracted as time functions of various dynamic properties and recognized by using Hidden Markov Models. Global information is extracted with a feature-based representation and recognized by using Parzen windows density estimation. The expert based on local information has been developed in the framework of this Thesis by extending the previous work in function-based signature verification conducted at the Biometrics Research Lab.–ATVS [Ortega-Garcia *et al.*, 2002]. The expert based on global information has been developed in the framework of this Thesis jointly with Lopez-Peñalba [2006]. Therefore this chapter includes novel contributions both in the application of user-dependent score normalization to multi-algorithm signature verification and in the two individual signature systems described.

The chapter is structured as follows. We first summarize the different approaches found in the literature for signature verification. We then describe the system based on local information, analyzing some of its key components on a development set of the MCYT signature corpus. User-dependent score normalization is then studied using the local system on the signature database from the SVC 2004 competition. The system based on global information is then introduced. Experimental results are finally given on the whole MCYT signature database for

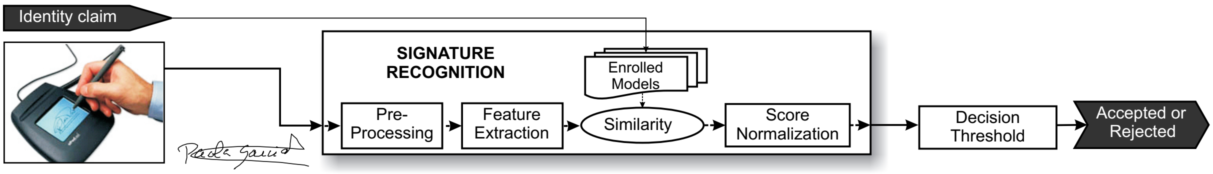


Figure 5.1: *Architecture of the proposed on-line signature verification system based on local information.*

both systems independently, as well as their combination using various forms of user dependent decision thresholds.

This chapter is based on the publications: Fierrez-Aguilar *et al.* [2005f, 2004c, 2005h]; Ortega-Garcia *et al.* [2003a].

5.1. Multilevel Signature Verification

Within biometrics, automatic signature verification has been an intense research area because of the social and legal acceptance and widespread use of the written signature as a personal authentication method [Plamondon and Lorette, 1989; Plamondon and Srihari, 2000]. This chapter deals with on-line signature verification. On-line refers here to the use of some information from the dynamic signing process (e.g., position trajectory, or pressure versus time), which is normally obtained by using a pen tablet or the touch screen of devices like PDAs or Tablet PCs.

Different approaches are considered in the literature in order to extract relevant information from on-line signature data [Plamondon and Lorette, 1989]; they can broadly be divided into: 1) feature-based approaches, in which a holistic vector representation consisting of global features is derived from the acquired signature trajectories [Kashi *et al.*, 1997; Lee *et al.*, 1996; Nelson and Kishon, 1991; Nelson *et al.*, 1994], and 2) function-based approaches, in which time sequences describing local properties of the signature are used for recognition (e.g., position trajectory, velocity, acceleration, force, or pressure) [Jain *et al.*, 2002; Kashi *et al.*, 1997; Nalwa, 1997; Ortega-Garcia *et al.*, 2003a]. Some works combine the two strategies (feature- and function-based) [Kashi *et al.*, 1997; Zhang *et al.*, 2002]. Large-scale experiments in on-line signature verification are described by Fairhurst [1997] and Plamondon *et al.* [1999].

5.2. System Based on Local Information

This system is based on Hidden Markov Models (HMMs), which have shown good recognition capabilities in other behavioral-based biometric traits like speech [Rabiner, 1989]. Some previous works using HMMs for signature recognition are described by Kashi *et al.* [1997]; Yang *et al.* [1995]; and Dolfing *et al.* [1998].

In this system we try to exploit the dynamic signature information as complete time sequences of various physical related parameters [Ortega-Garcia *et al.*, 2003a]. In particular, the contributions of the system are mainly on the following modules (see Fig. 5.1): 1) feature extraction, in this case we propose a new set of time functions including time derivatives; and 2) direct modeling of the time sequences using continuous HMMs, not stroke-based information as the previous works [Dolfing *et al.*, 1998; Kashi *et al.*, 1997; Yang *et al.*, 1995].

5.2.1. Feature Extraction

Basic functions. The signature representation considered in this work is based on the following five time sequences: horizontal x_n and vertical y_n position trajectories, azimuth γ_n and altitude ϕ_n of the pen with respect to the tablet, and pressure signal p_n, where $n = 1, \ldots, N$ is the discrete time index given by the acquisition device and N is the time duration of the signature in sampling units. Although pen inclination trajectories have shown some discriminant capabilities in other works [Hangai *et al.*, 2000; Pacut and Czajka, 2001; Sakamoto *et al.*, 2001], the use of these two functions worsens the verification performance in our system, as demonstrated in Sect. 5.2.3.1. As a result, the basic function set consists only of x_n, y_n and p_n.

Geometric normalization. A signature acquisition process on a restricted size frame is assumed. Therefore users are supposed to be consistent in size and writing dynamics. Nevertheless a geometric normalization consisting of position normalization followed by rotation alignment is applied.

Position normalization consists in aligning the center of mass $\frac{1}{N}\sum_{n=1}^{N}[x_n, y_n]^T$ of the different signatures, where $[\cdot]^T$ denotes transpose. Rotation normalization consists in aligning the average path tangent angle $\bar{\theta} = \frac{1}{N}\sum_{n=1}^{N}\arctan(\dot{y}_n/\dot{x}_n)$ of the different signatures, where the dot notation denotes first order time derivative.

Extended functions. After geometric normalization, some other sequences are derived from the basic function set. Previous results with other dynamic sequences (e.g., path tangent angle, path velocity magnitude, and log curvature radius) have shown good levels of performance [Ortega-Garcia *et al.*, 2003a]. In the present system, four dynamic sequences are used as extended functions, namely [Nelson and Kishon, 1991]:

1. Path-tangent angle: $\theta_n = \arctan(\dot{y}_n/\dot{x}_n)$.
2. Path velocity magnitude: $v_n = \sqrt{\dot{x}_n^2 + \dot{y}_n^2}$.
3. Log curvature radius: $\rho_n = \log(1/\kappa_n) = \log\left(v_n/\dot{\theta}_n\right)$, where κ_n is the curvature of the position trajectory and $\log(\cdot)$ is applied in order to reduce the dynamic range of function values.
4. Total acceleration magnitude: $a_n = \sqrt{t_n^2 + c_n^2}$, where $t_n = \dot{v}_n$ and $c_n = v_n \cdot \dot{\theta}_n$ are respectively the tangential and centripetal acceleration components of the pen motion.

In all cases, (discrete) time derivatives have been computed by using a second-order regression [Young *et al.*, 2002]. As a result, the complete instantaneous function-based feature set, including three basic and four extended time sequences is as follows:

$$\mathbf{u}_n = [x_n, y_n, p_n, \theta_n, v_n, \rho_n, a_n]^T, \qquad n = 1, \ldots, N, \tag{5.1}$$

where N is the time duration of the considered signature in sampling units.

Time derivatives. First order time derivatives of complete instantaneous function-based feature sets are highly effective as discriminant parameters regarding verification with other behavioral traits [Soong and Rosenberg, 1988]. Therefore, we decided to include time derivatives in our function set. Because of the discrete nature of the above-mentioned functions, first order time derivatives are calculated by using a second order regression [Young *et al.*, 2002], expressed through operator Δ:

$$\dot{f}_n \approx \Delta f_n = \frac{\sum_{\tau=1}^{2} \tau (f_{n+\tau} - f_{n-\tau})}{2 \sum_{\tau=1}^{2} \tau^2}. \tag{5.2}$$

In this way, each parameterized signature is formally described as a matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$, where $\mathbf{v}_n = \left[\mathbf{u}_n^T, (\Delta \mathbf{u}_n)^T\right]^T$, $n = 1, \ldots, N$.

Signal normalization. A final signal normalization is applied in order to obtain zero mean and unit standard deviation function values:

$$\mathbf{o}_n = \mathbf{\Sigma}^{-1/2}(\mathbf{v}_n - \boldsymbol{\mu}), \quad n = 1, \ldots, N, \tag{5.3}$$

where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are respectively the sample mean and sample diagonal covariance matrix of vectors $\mathbf{v}_n$, $n = 1, \ldots, N$.

As a result, each signature is represented by a matrix $\mathbf{O} = [\mathbf{o}_1, \ldots, \mathbf{o}_N]$ comprising 14 statistically-normalized discrete time sequences.

5.2.2. Signature Modeling

Basically, the HMM represents a doubly stochastic process governed by an underlying Markov chain with finite number of states and a set of random functions each of which is associated with the output observation of one state [Theodoridis and Koutroumbas, 2003]. At discrete instants of time n, the process is in one of the states and generates an observation symbol according to the random function corresponding to that current state. The model is hidden in the sense that the underlying state which generates each symbol cannot be deduced from simple symbol observation.

Formally, a HMM is described as follows:

1. H, which is the number of hidden states $\{S_1, S_2, \ldots, S_H\}$. The state at discrete time n will be denoted as q_n.

2. The state transition matrix $\mathbf{A} = \{a_{ij}\}$ where

$$a_{ij} = P(q_{n+1} = S_j | q_n = S_i), \qquad 1 \leq i, j \leq H. \tag{5.4}$$

3. The observation symbol probability density function in state j, $b_j(\mathbf{o})$, $1 \leq j \leq H$.

4. The initial state distribution $\boldsymbol{\pi} = \{\pi_i\}$ where

$$\pi_i = P(q_1 = S_i), \qquad 1 \leq i \leq H. \tag{5.5}$$

In this contribution $b_j(\mathbf{o})$ is modeled as a mixture of M multi-variate Gaussian densities:

$$b_j(\mathbf{o}) = \sum_{m=1}^{M} c_{jm} p(\mathbf{o}|\boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}), \qquad 1 \leq j \leq H, \tag{5.6}$$

where $p(\mathbf{o}|\boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm})$ is a multi-variate Gaussian distribution with mean $\boldsymbol{\mu}_{jm}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{jm}$. Thus, the observation symbol density functions can be parameterized as $B = \left\{c_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right\}$, $1 \leq j \leq H, 1 \leq m \leq M$.

A particular signature model is characterized by the set $\lambda = \{\boldsymbol{\pi}, \mathbf{A}, B\}$, which is trained by using K training signatures of a specific client user. The similarity score of an input signature $\mathbf{O} = [\mathbf{o}_1, \ldots, \mathbf{o}_N]$ claiming the identity λ is calculated as $\frac{1}{N} \log \left(P\left(\mathbf{O}\left|\lambda\right.\right)\right)$ by using the Viterbi algorithm [Rabiner, 1989].

The client model λ is trained with K training signatures $\left\{\mathbf{O}^{(1)}, \ldots, \mathbf{O}^{(K)}\right\}$, where $\mathbf{O}^{(k)} = [\mathbf{o}_1^{(k)}, \ldots, \mathbf{o}_{N_k}^{(k)}]$ with $k = 1, \ldots, K$, by means of the following iterative strategy:

- Initialize λ. Each one of the training signatures $\mathbf{O}^{(k)}$, $1 \leq k \leq K$, is divided into H segments $\mathbf{S}_1^{(k)}, \ldots, \mathbf{S}_H^{(k)}$ where

$$\begin{aligned} \mathbf{S}_i^{(k)} &= [\mathbf{o}_{(i-1)\lceil N_k/H \rceil+1}^{(k)}, \mathbf{o}_{(i-1)\lceil N_k/H \rceil+2}^{(k)}, \ldots, \mathbf{o}_{(i)\lceil N_k/H \rceil}^{(k)}], \quad 1 \leq i \leq H-1, \\ \mathbf{S}_H^{(k)} &= [\mathbf{o}_{(H-1)\lceil N_k/H \rceil+1}^{(k)}, \mathbf{o}_{(H-1)\lceil N_k/H \rceil+2}^{(k)}, \ldots, \mathbf{o}_{N_k}^{(k)}], \end{aligned} \tag{5.7}$$

  and $\lceil \cdot \rceil$ denotes equal or higher nearest integer. Observations $\mathbf{o}$ from the segments $\mathbf{S}_j^{(1)}, \mathbf{S}_j^{(2)}, \ldots, \mathbf{S}_j^{(K)}$ are clustered into M groups by using the k-means algorithm [Theodoridis and Koutroumbas, 2003] and the samples from cluster m are used to estimate the initial parameters $B = \left\{c_{jm}, \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right\}$, $1 \leq j \leq H, 1 \leq m \leq M$. Initial $\mathbf{A}$ takes into account a left-to-right topology without skip transitions (see Fig. 5.2). Thus, $a_{ij} = 0$ for $i > j$ or $j > i + 1$, $a_{ii} = (O_i - 1)/O_i$ and $a_{i,i+1} = 1/O_i$, where O_i is the number of observations in the segments $\mathbf{S}_i^{(1)}, \mathbf{S}_i^{(2)}, \ldots, \mathbf{S}_i^{(K)}$. The initial state distribution $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_H\}$ is set up as $\{1, 0, \ldots, 0\}$.

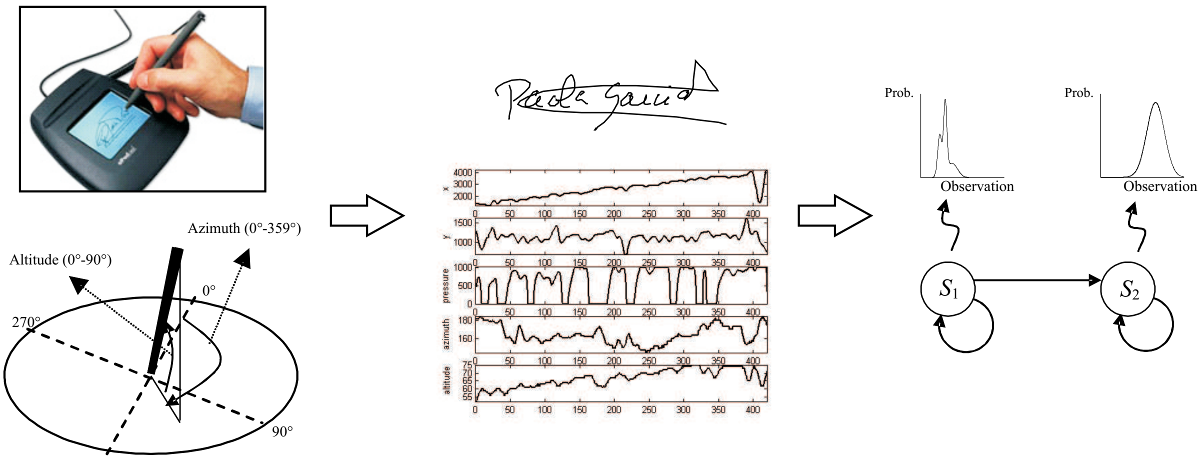


Figure 5.2: *Processing steps of the proposed on-line signature verification system based on local information.*

- Re-estimate a new model $\bar{\lambda}$ from λ by using the Baum-Welch re-estimation equations [Rabiner, 1989], which guarantee that:

$$\prod_{k=1}^{K} P(\mathbf{O}^{(k)}|\bar{\lambda}) \geq \prod_{k=1}^{K} P(\mathbf{O}^{(k)}|\lambda). \tag{5.8}$$

- Replace λ by $\bar{\lambda}$ and go to previous step (iterate) until:

$$\prod_{k=1}^{K} P(\mathbf{O}^{(k)}|\bar{\lambda}) - \prod_{k=1}^{K} P(\mathbf{O}^{(k)}|\lambda) \leq \Theta, \tag{5.9}$$

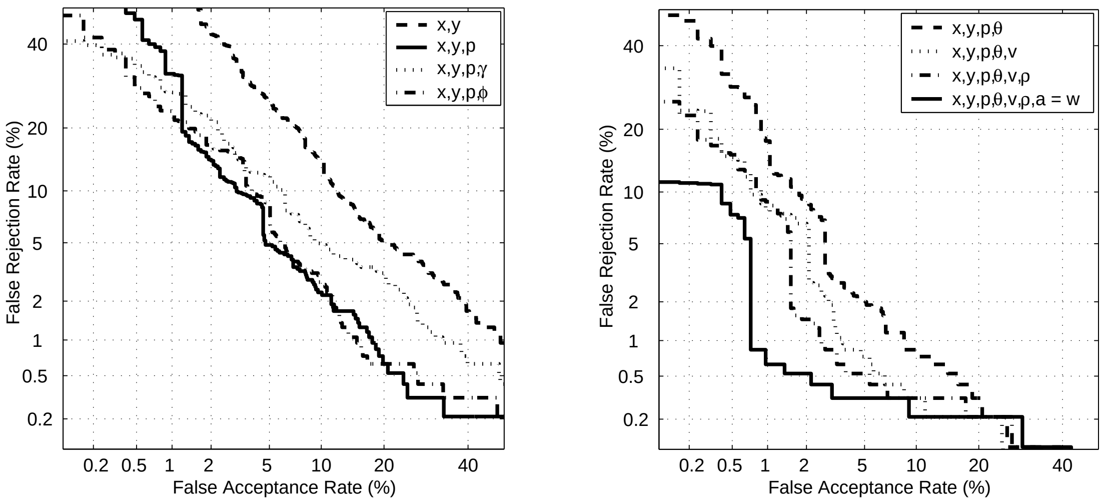
  where the threshold Θ is chosen heuristically and the maximum number of iterations is limited to ten.

In the following experiments, the training algorithm typically converges after five iterations.

5.2.3. Experiments

Experiments have been carried out according to the following procedure. We first adjust the system by using as development set the first 50 subjects from MCYT signature corpus, with all the available signatures (i.e., 25 genuine and 25 forgeries per subject). Results are given both as EER and DET plots applying *a posteriori* score alignment across users based on the user-dependent empirical EER. From these development experiments, and considering the average EER as the cost function to minimize, we obtain the best working point for our system. Simultaneously, we study a number of important factors for signature verification including feature extraction, modeling, and training strategy.

Once the system is adjusted, experiments related to user-dependent score normalization are reported on data from SVC 2004 [Yeung *et al.*, 2004].



(a) (b)

Figure 5.3: *Verification performance results for skilled forgeries with various functions: position trajectories x and y, pressure p, azimuth γ, altitude ϕ, path tangent angle θ, path velocity magnitude v, log curvature radius ρ, and total acceleration magnitude a.*

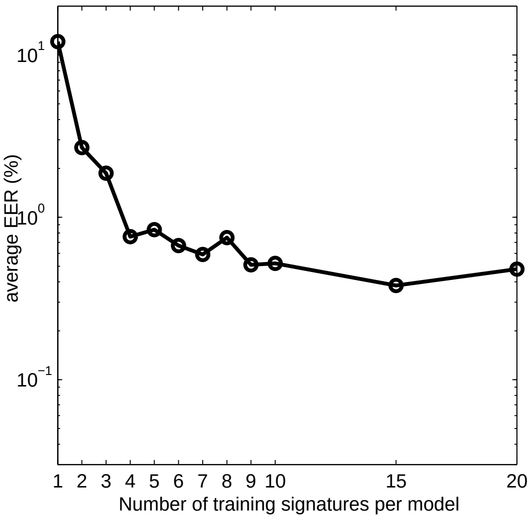
5.2.3.1. System Development

Feature extraction. The initial configuration for the experiments is [Ortega-Garcia *et al.*, 2003a]: 4 HMM states, 8 mixtures per state and 5 training signatures from different acquisition sets. Results for different function parameterizations are shown in Fig. 5.3 for skilled forgeries.

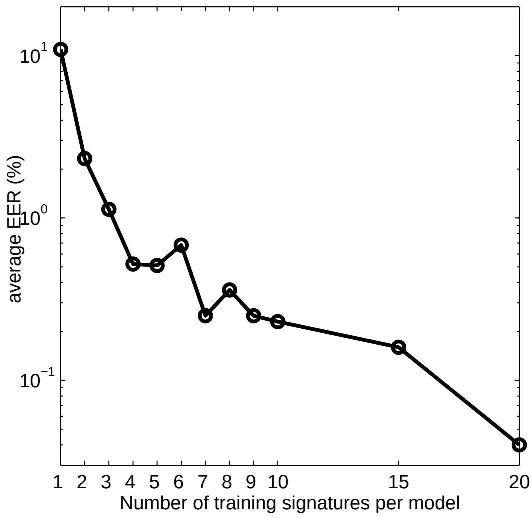
In Fig. 5.3(a), some basic function sets are compared. Although azimuth and altitude signals have shown some discriminative capabilities in other works, the inclusion of these two functions worsens the verification performance in our system. In particular, EER decreases from 10.37% to 4.54% when pressure signal is included to the basic position trajectory information but increases to 4.84% and 6.28% when altitude and azimuth signals are respectively considered. In Fig. 5.3(b) we show the progressive improvement in verification performance when extended functions are included one by one to the basic set $\{x, y, p\}$. In particular, when path tangent angle θ, path velocity magnitude v, log curvature radius ρ, and total acceleration mangnitude a are progressively included, we obtain 2.57%, 1.99%, 1.44%, and 0.68% EER. The set composed by these 7 functions $\{x, y, p, \theta, v, \rho, a\}$ will be referred to as w.

Training Strategy. The initial configuration for the experiments is: $w + \Delta w$ functions, 4 HMM states and 8 mixtures per state. Results for different training strategies are shown in Fig. 5.4. In particular, results for an increasing number of training signatures are shown in Fig. 5.4(a) for *ordered-set* training and in Fig. 5.4(b) for *multi-set* training.

In the *ordered-set* training strategy, see Fig. 5.4(a), training signatures are selected from the minimum number of acquisition sets (i.e., 1 set for 1 to 5 training signatures, 2 sets for 6 to 10 signatures, and so on). As a result, average EER does not improve significantly for more than



(a)



(b)

Figure 5.4: *Training strategy experiments. Verification performance results are given for skilled forgeries with increasing number of training signatures: (a) low variability between training signatures, (b) high variability between training signatures.*

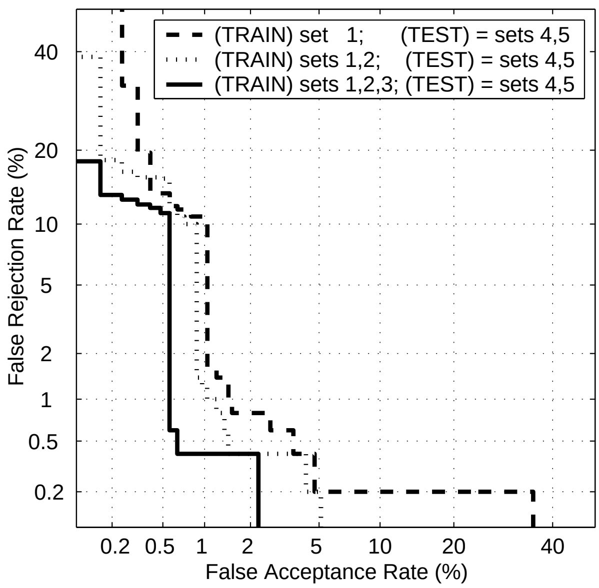


Figure 5.5: *Training strategy experiments. Verification performance results for skilled forgeries for a fixed number of training signatures with increasing variability between training signatures.*

Table 5.1: Average EER (in %) for different HMM configurations (skilled forgeries). H = number of states; M = number of Gaussian mixtures per state.

	$M=1$	$M=2$	$M=4$	$M=8$	$M=16$	$M=32$	$M=64$
$H=1$					1.74	1.05	0.70
$H=2$				1.51	0.74	**0.30**	0.44
$H=4$			1.64	0.87	0.52	0.48	
$H=8$		1.81	0.79	0.76	0.35		
$H=16$		1.20	0.96	0.74			
$H=32$	1.28	0.97					

5 training samples.

Taking into account the fact that the variability between signatures of different acquisition sets is high (see Sect. 4.3.2), in the *multi-set* training strategy, see Fig. 5.4(b), we have selected training signatures from the maximum number of acquisition sets (i.e., 1 set for 1 training signature, 2 sets for 2 training signatures, 3 sets for 3 training signatures, 4 sets for 4 training signatures and 5 sets for 5 to 20 training signatures). As a result, EER improves significantly with the first 5 training signatures (0.85% EER) and keeps improving for higher number of training signatures (0.05% EER for 20 training samples).

Finally, we test the verification performance for a fixed number of training signatures when the variability between training signatures is increased. Results are shown in Fig. 5.5 testing with two acquisition sets, and selecting the 5 training signatures from the other one (0.84% EER), two (0.82% EER) and three (0.48% EER) sets, respectively. As a result, verification performance improves when training data comes from different acquisition sets. This result shows the importance of multi-session training in signature verification.

Signature modeling. The initial configuration for the experiment is: $w + \Delta w$ functions and 5 training signatures from the first acquisition set. Results as EER for different HMM parameters are shown in Table 5.1. The degenerated case of single state HMM is equivalent to modeling based on Gaussian Mixture Models [Richiardi and Drygajlo, 2003].

From Table 5.1 we conclude that $H = 2$, $M = 32$ is the optimal configuration of our system. This result is in accordance with the recent trend of reducing the number of states in HMM-based on-line signature verification systems [Richiardi and Drygajlo, 2003], and can be well explained in our case because the signatures in MCYT corpus usually consist of written name and flourish [Fierrez-Aguilar *et al.*, 2004a]. These two parts usually have quite different dynamic statistics that the 2 state modeling approach may be exploiting.

Finally, we represent in Fig. 5.6 DET plots for $H = 2$ with different values of M, where the verification performance improvement for increasing values of M can be seen, until the optimum ($M = 32$) is reached.

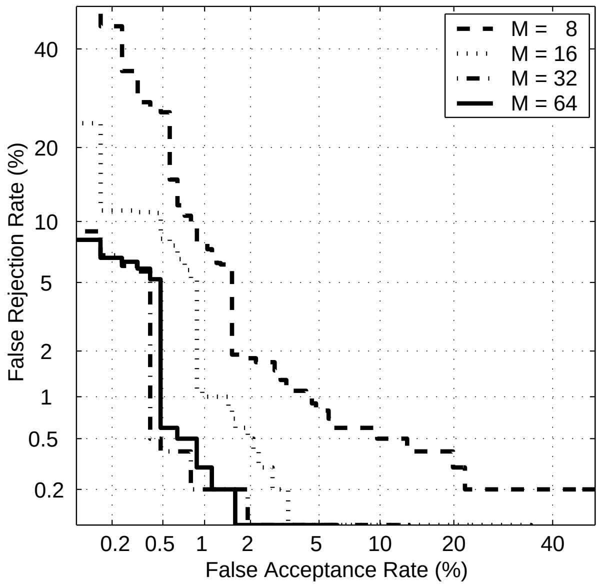


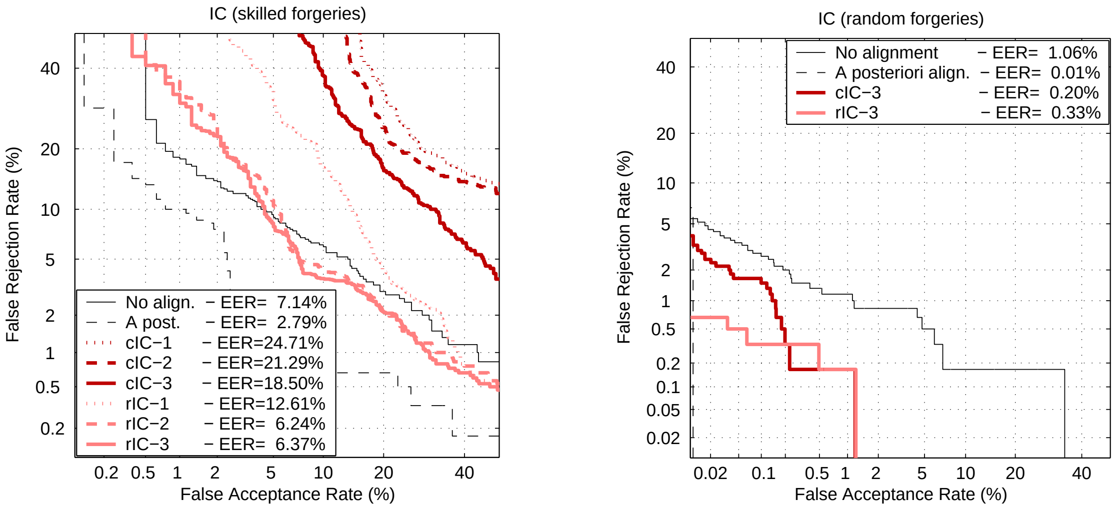
Figure 5.6: *Signal modeling experiments. Verification performance results are given for an increasing number of Gaussian mixtures per state M, being the number of states fixed $H = 2$ (skilled forgeries).*

5.2.3.2. User-Dependent Score Normalization

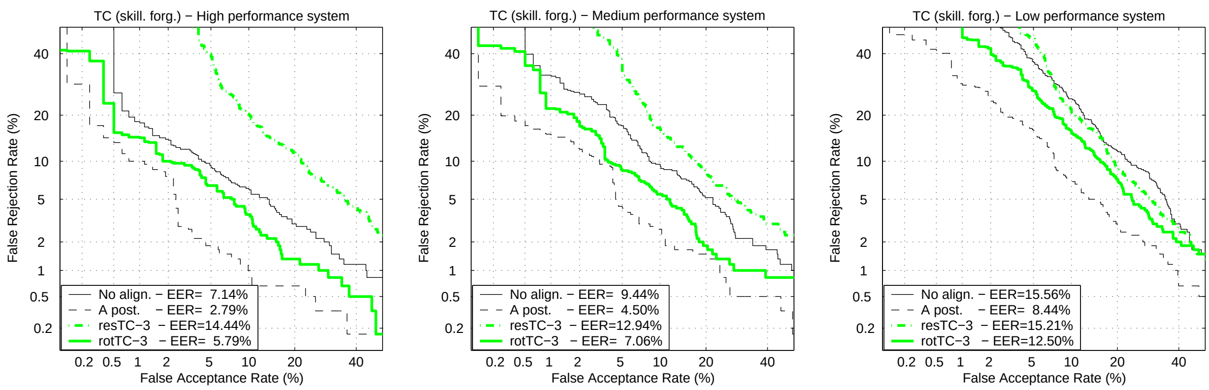
The experimental procedure for evaluating the user-dependent score techniques developed in Sect. 3.1.1.2 is as follows. Signature data from the two acquisition sessions in SVC 2004 are used both for training and testing (see Sect. 4.4 for a description of the SVC 2004 database). For training, we use 5 random genuine signatures from both sessions. For testing, we use the remaining 15 genuine signatures. For a specific target user, casual impostor tests are computed by using signatures from all the remaining targets. Real impostor tests are computed by using the 20 skilled forgeries of each target. Score normalization results are provided using statistics either from casual or real impostors for the computation of the normalization functions.

A priori score normalization methods are compared in the experiments. This means that only the information from the training set is used both for the enrollment of the targets and for the estimation of the parameters of the normalization functions (by using the resampling techniques described in Sect. 4.1.2). In order to have an indication of the level of performance with an ideal score alignment between targets, we also give results using *a posteriori* user-dependent score alignment across users. Only in this case, test information is used both for error estimation and for the computation of the score normalization functions.

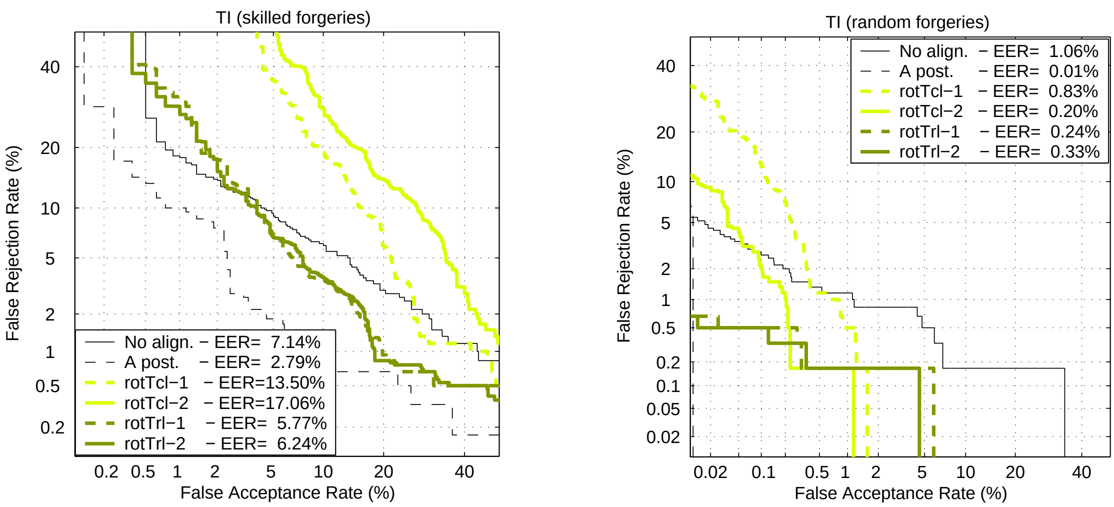
Results. In Fig. 5.7 (a) the different impostor-centric methods (see Sect. 3.1.1.2) are compared either for skilled (left) or random forgeries (right). Raw verification performance with no normalization (7.14% and 1.06% EER for skilled and random forgeries, respectively) is significantly improved by the *a posteriori* normalization scheme (2.79% and 0.01%, respectively). Regarding the skilled forgeries test, method IC-3 outperforms IC-1 and IC-2. Raw performance is



(a) Impostor-Centric: Different impostor-variability estimation methods.



(b) Target-Centric: Different client-variability estimation methods.



(c) Target-Impostor: casual/real information for impostor-variability estimation.

Figure 5.7: *Comparison of user dependent score normalization techniques.*

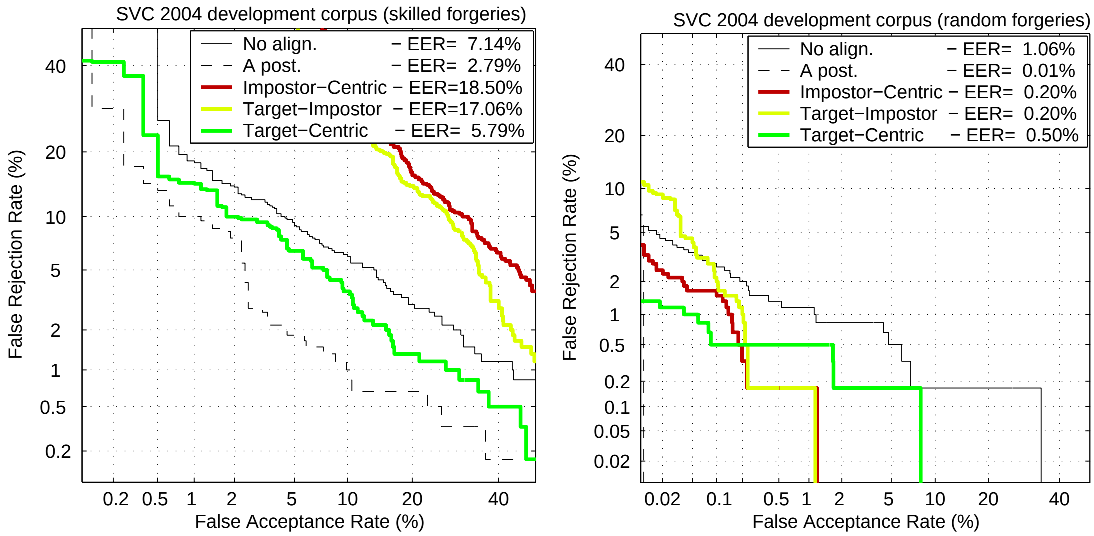


Figure 5.8: *Verification performance for various user dependent normalization methods on SVC 2004 development corpus.*

only improved in this case by considering statistics from real impostors. Regarding the random forgeries test, significant improvements are obtained considering statistics either from casual or from real impostors.

Results of different resampling techniques for the estimation of target variability are summarized in Fig. 5.7 (b) for three different verification systems of decreasing verification performance (from left to right). As it can be observed, the rotation scheme always leads to verification improvements whereas the resubstitution strategy only leads to improvements in the low performance system. This result penalizes the biased estimation provided by the resubstitution scheme in favor of the unbiased rotation procedure, at the cost of higher computational requirements.

Verification performance for the target-impostor methods is shown in Fig. 5.7 (c). As in the impostor-centric experiment, only target-impostor normalization schemes based on real impostor statistics improve the verification performance in case of tests with skilled forgeries. With regard to the test considering random forgeries, verification performance improvements are obtained considering either casual impostor or real impostor statistics.

Summary. Results using a selection of practical *a priori* normalization methods following SVC 2004 guidelines (i.e., using real impostor statistics for the computation of the normalization functions is not permitted) are finally summarized in Fig. 5.8. In this case, only the target-centric method is capable of performance improvements testing both with skilled and random forgeries.

The local system described here participated in the First International Signature Verification Competition in 2004 (see Sect. 4.4). In particular, three variants were submitted without and with user-dependent score normalization (systems 19a and 19b-c, respectively [Yeung *et al.*, 2004]). The best performing score normalization was, as in our previous experiments, the rotTC-3 *a priori* target-centric technique. This is similar to the popular z-norm but using only client scores obtained from rotation training.

In Task 2 the system was ranked first and second for random and skilled forgeries, respectively. Moreover, it was demonstrated that incorporating *a priori* user-dependent score normalization (from system 19a to system 19b), an average of 15% relative performance improvement was obtained.

The local system presented here has been integrated in a Tablet PC environment with a client-server architecture [Alonso-Fernandez *et al.*, 2005c, 2006b], with application to secure web access and file encryption [del Valle-Hernández, 2006]. The Tablet PC system is configured according to the results obtained in the development experiments reported here (e.g., 14 time functions, reduced number of HMM states, 5 training signatures from different acquisition sessions, etc.), and incorporates the user-dependent score normalization strategy that performed best in SVC 2004 (*a priori* target-centric similar to z-norm).

5.3. System Based on Global Information

The subsystem exploiting global information is based on the previous works by Nelson and Kishon [1991]; Nelson *et al.* [1994]; and Lee *et al.* [1996]. In particular, our contributions are the following: 1) the set of features described in these precedent works (approximately 70, considering the three papers) is extended, leading to a 100-dimensional feature vector representation; 2) feature selection experiments on the complete set are carried out, obtaining experimental evidence on the individual relative discriminative capabilities of the proposed and the existing features, and 3) a non-parametric statistical recognition strategy based on Parzen windows is used, obtaining remarkable performance in the common case of small training set size.

5.3.1. Feature Extraction

The complete set of global features is given in Table 5.2. All notations are either defined somewhere in the table or can be found in the three works above mentioned [Lee *et al.*, 1996; Nelson and Kishon, 1991; Nelson *et al.*, 1994]. For example: j is defined in feature ranked 4 and explained in Nelson and Kishon [1991], the notation Δ is defined in the denominator of feature 15, histograms in $34, 51, 61, 70, 93$ are explained in Nelson *et al.* [1994], etc.

The feature extraction process assumes an acquisition process capturing position trajectories and pressure signals both at pen-down and pen-up intervals. Otherwise, the feature set should be reduced discarding features based on trajectory signals during pen-ups (e.g., features 32 and 41). Even though the given set has been demonstrated to be robust to the common distortions encountered in the handwritten scenario, note that not all the parameters are fully rotation/scale invariant, so either a controlled signature acquisition is assumed (as in MCYT database, where users were asked to sign within grid guidelines) or translation/rotation registration should be performed before computing them. Although pen inclination has shown discriminative power in some works [Sakamoto *et al.*, 2001], no features based on pen inclination are introduced in the

Table 5.2: *Set of global features sorted by individual discriminative power (T denotes time interval, t denotes time instant, N denotes number of events, θ denotes angle,* **bold** *denotes novel feature, italic denotes adapted from the literature,* roman *denotes used as in the literature).*

Ranking	Feature Description	Ranking	Feature Description
1	signature total duration T_s	*2*	N(pen-ups)
3	N(sign changes of dx/dt and dy/dt)	**4**	average jerk $\bar{\jmath}$
5	standard deviation of a_y	**6**	standard deviation of v_y
7	(standard deviation of y)/Δ_y	**8**	N(local maxima in x)
9	standard deviation of a_x	**10**	standard deviation of v_x
11	j_{rms}	**12**	N(local maxima in y)
13	t(2nd pen-down)/T_s	14	(average velocity $\bar{v}$)/$v_{x,\max}$
15	$\frac{A_{\min}=(y_{\max}-y_{\min})(x_{\max}-x_{\min})}{(\Delta_x=\sum_{i=1}^{\text{pen-downs}}(x_{\max\mid i}-x_{\min\mid i}))\Delta_y}$	*16*	$(x_{\text{last pen-up}}-x_{\max})/\Delta_x$
17	$(x_{\text{1st pen-down}}-x_{\min})/\Delta_x$	*18*	$(y_{\text{last pen-up}}-y_{\min})/\Delta_y$
19	$(y_{\text{1st pen-down}}-y_{\min})/\Delta_y$	**20**	$(T_w\bar{v})/(y_{\max}-y_{\min})$
21	$(T_w\bar{v})/(x_{\max}-x_{\min})$	*22*	(pen-down duration T_w)/T_s
23	$\bar{v}/v_{y,\max}$	*24*	$(y_{\text{last pen-up}}-y_{\max})/\Delta_y$
25	$\frac{T((dy/dt)/(dx/dt)>0)}{T((dy/dt)/(dx/dt)<0)}$	26	$\bar{v}/v_{\max}$
27	$(y_{\text{1st pen-down}}-y_{\max})/\Delta_y$	*28*	$(x_{\text{last pen-up}}-x_{\min})/\Delta_x$
29	(velocity rms v)/$v_{\max}$	*30*	$\frac{(x_{\max}-x_{\min})\Delta_y}{(y_{\max}-y_{\min})\Delta_x}$
31	(velocity correlation $v_{x,y}$)/$v_{\max}^2$	*32*	$T(v_y>0\mid\text{pen-up})/T_w$
33	$N(v_x=0)$	34	direction histogram s_1
35	$(y_{\text{2nd local max}}-y_{\text{1st pen-down}})/\Delta_y$	36	$(x_{\max}-x_{\min})/x_{\text{acquisition range}}$
37	$(x_{\text{1st pen-down}}-x_{\max})/\Delta_x$	*38*	$T(\text{curvature}>\text{Threshold}_{\text{curv}})/T_w$
39	(integrated abs. centr. acc. a_{Ic})/$a_{\max}$	40	$T(v_x>0)/T_w$
41	$T(v_x<0\mid\text{pen-up})/T_w$	*42*	$T(v_x>0\mid\text{pen-up})/T_w$
43	$(x_{\text{3rd local max}}-x_{\text{1st pen-down}})/\Delta_x$	*44*	$N(v_y=0)$
45	(acceleration rms a)/$a_{\max}$	*46*	(standard deviation of x)/Δ_x
47	$\frac{T((dx/dt)(dy/dt)>0)}{T((dx/dt)(dy/dt)<0)}$	48	(tangential acceleration rms a_t)/$a_{\max}$
49	$(x_{\text{2nd local max}}-x_{\text{1st pen-down}})/\Delta_x$	*50*	$T(v_y<0\mid\text{pen-up})/T_w$
51	direction histogram s_2	**52**	t(3rd pen-down)/T_s
53	(max distance between points)/$A_{\min}$	**54**	$(y_{\text{3rd local max}}-y_{\text{1st pen-down}})/\Delta_y$
55	$(\bar{x}-x_{\min})/\bar{x}$	56	direction histogram s_5
57	direction histogram s_3	58	$T(v_x<0)/T_w$
59	$T(v_y>0)/T_w$	60	$T(v_y<0)/T_w$
61	direction histogram s_8	62	(1st $t(v_{x,\min})$)/T_w
63	direction histogram s_6	**64**	T(1st pen-up)/T_w
65	spatial histogram t_4	66	direction histogram s_4
67	$(y_{\max}-y_{\min})/y_{\text{acquisition range}}$	68	(1st $t(v_{x,\max})$)/T_w
69	(centripetal acceleration rms a_c)/$a_{\max}$	**70**	spatial histogram t_1
71	θ(1st to 2nd pen-down)	*72*	θ(1st pen-down to 2nd pen-up)
73	direction histogram s_7	**74**	$t(j_{x,\max})/T_w$
75	spatial histogram t_2	**76**	$j_{x,\max}$
77	θ(1st pen-down to last pen-up)	**78**	θ(1st pen-down to 1st pen-up)
79	(1st $t(x_{\max})$)/T_w	**80**	$\bar{\jmath}_x$
81	T(2nd pen-up)/T_w	82	(1st $t(v_{\max})$)/T_w
83	$j_{y,\max}$	**84**	θ(2nd pen-down to 2nd pen-up)
85	$j_{\max}$	**86**	spatial histogram t_3
87	(1st $t(v_{y,\min})$)/T_w	**88**	(2nd $t(x_{\max})$)/T_w
89	(3rd $t(x_{\max})$)/T_w	90	(1st $t(v_{y,\max})$)/T_w
91	$t(j_{\max})/T_w$	**92**	$t(j_{y,\max})/T_w$
93	direction change histogram c_2	**94**	(3rd $t(y_{\max})$)/T_w
95	direction change histogram c_4	**96**	$\bar{\jmath}_y$
97	direction change histogram c_3	*98*	θ(initial direction)
99	θ(before last pen-up)	**100**	(2nd $t(y_{\max})$)/T_w

proposed set, as pen inclination turned out to be highly unstable in the previous experiments with the local system. The features in Table 5.2 are sorted by individual discriminative power as described in the next section.

5.3.2. Feature Selection

Due to the high number of proposed features (100), and the large number of signatures considered (16500 in total in MCYT), features have been ranked according to scalar inter-user class separability [Theodoridis and Koutroumbas, 2003]. Feature selection is then based on selecting an increasing number of ranked features.

For each feature F_k, $k = 1, \ldots, 100$, we compute the scalar Mahalanobis distance d^{M}_{i,F_k} between the mean of the F_k-parameterized training signatures of client i, $i = 1, \ldots, 330$, and the F_k-parameterized set of all training signatures from MCYT. Features are then ranked according to the following inter-user class separability measure $S(F_k)$

$$S(F_k) = \sum_{i=1}^{330} \sum_{j=1}^{330} |d^{\mathrm{M}}_{i,F_k} - d^{\mathrm{M}}_{j,F_k}|. \tag{5.10}$$

5.3.3. Signature Modeling

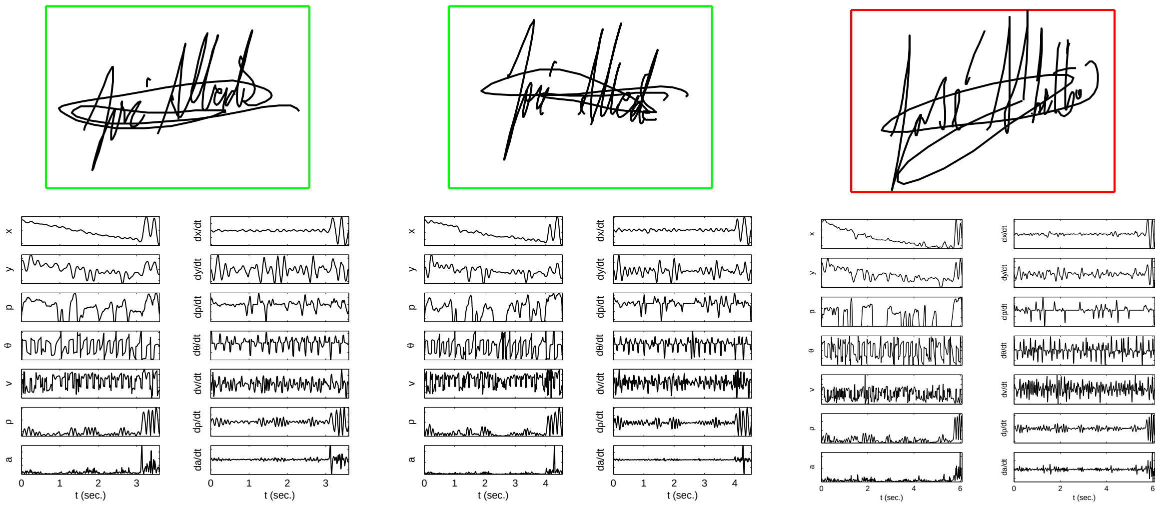
Given the feature vectors of the training set of signatures of a client k, a non-parametric estimation λ_k^{PWC} of their multivariate probability density function is obtained by using Parzen Gaussian windows [Theodoridis and Koutroumbas, 2003]. On the other hand, given the feature vector $\mathbf{o}$ of an input signature and a claimed identity k modeled as λ_k^{PWC}, the following similarity matching score is used

$$s_{\mathrm{PWC}} = p\left(\mathbf{o}|\lambda_k^{\mathrm{PWC}}\right). \tag{5.11}$$

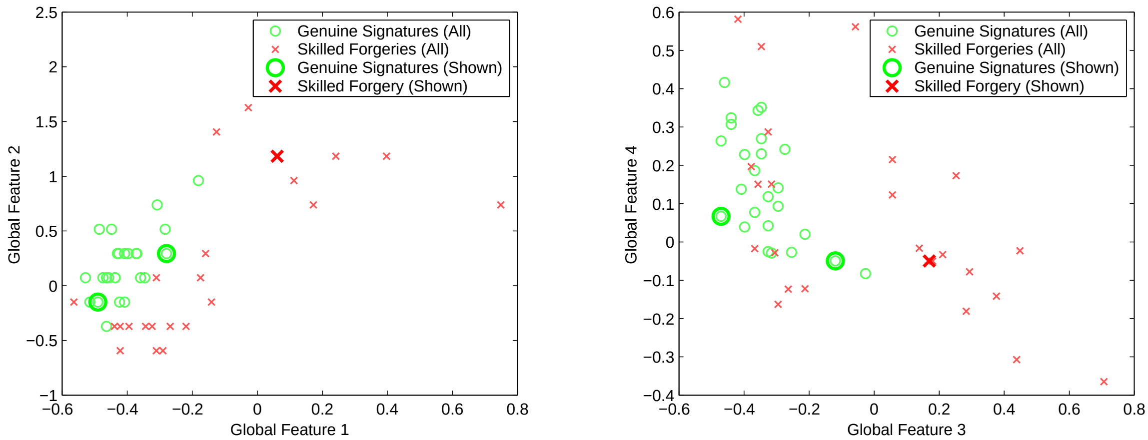
5.4. Experiments Combining Local and Global Systems

All the signatures of the MCYT database are used for these experiments. In total 16500 signatures from 330 signers with 25 genuine signatures and 25 skilled forgeries per signer. Two examples of genuine signatures (left and central columns) and one forgery (right column) are given in Fig. 5.9, which also depicts the related local functions and the four best global parameters for all the signatures corresponding to this subject.

Signature corpus is divided into training and test sets. For the test with skilled forgeries the training set comprises either 5 or 20 genuine signatures and the test set consists of the remaining signatures corresponding to each subject. For random forgeries we use one signature of every other user as impostor data.



(a) Two genuine signatures (left and central columns) and one skilled forgery (right column) for a client using name and complex flourish [Fierrez-Aguilar *et al.*, 2004a]. The function-based description used for local recognition is depicted below each signature.



(b) Best individually performing global features: 1st versus 2nd (left), and 3rd versus 4th (right); for all the signatures of the user above. Features from the genuine signatures and forgery depicted above are highlighted.

Figure 5.9: *Signature examples from MCYT corpus together with the extracted features.*

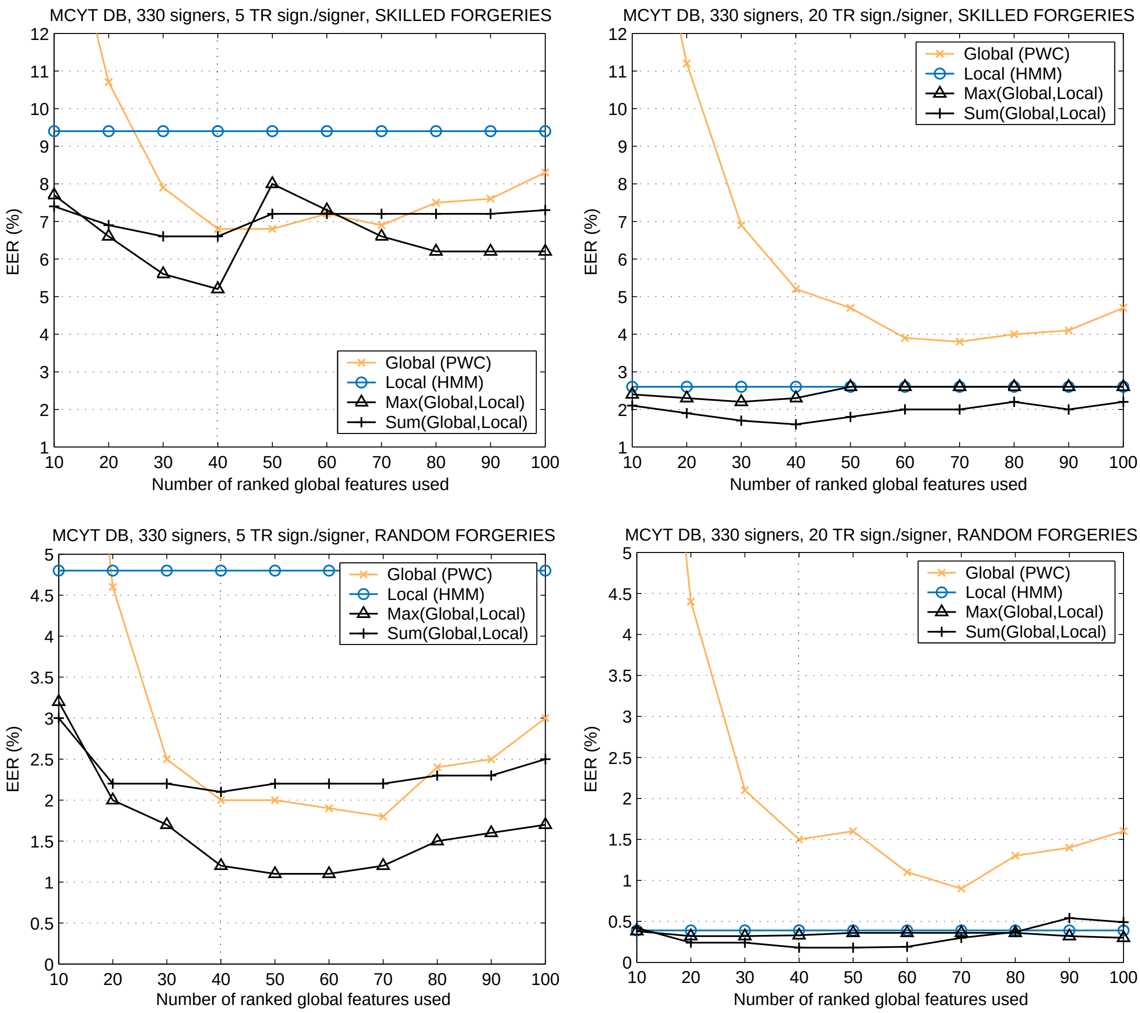


Figure 5.10: *Verification performance with user-independent decision thresholds for an increasing number of ranked global features.*

5.4.1. Results

In Fig. 5.10, verification performance results in four common conditions (few/many training signatures and skilled/random forgeries) are given for: 1) the local expert, 2) the global expert with an increasing number of ranked global features, and 3) their combination through *max* and *sum* rules (see Sect. 2.2.2.1). The similarity scores of each system have been mapped to probabilities by using fixed score normalization based on exponential functions, see Eq. (2.15).

The system based on global analysis outperforms the local approach when training with 5 signatures, and the opposite occurs when training with 20 signatures. The two systems are also shown to provide complementary information for the verification task, which is well exploited in the cases of small and large training set sizes using the *max* and *sum* rules, respectively. We have found a good working point of the combined system in the four conditions depicted in Fig. 5.10 by using the first 40 ranked features for the global approach. This is highlighted with

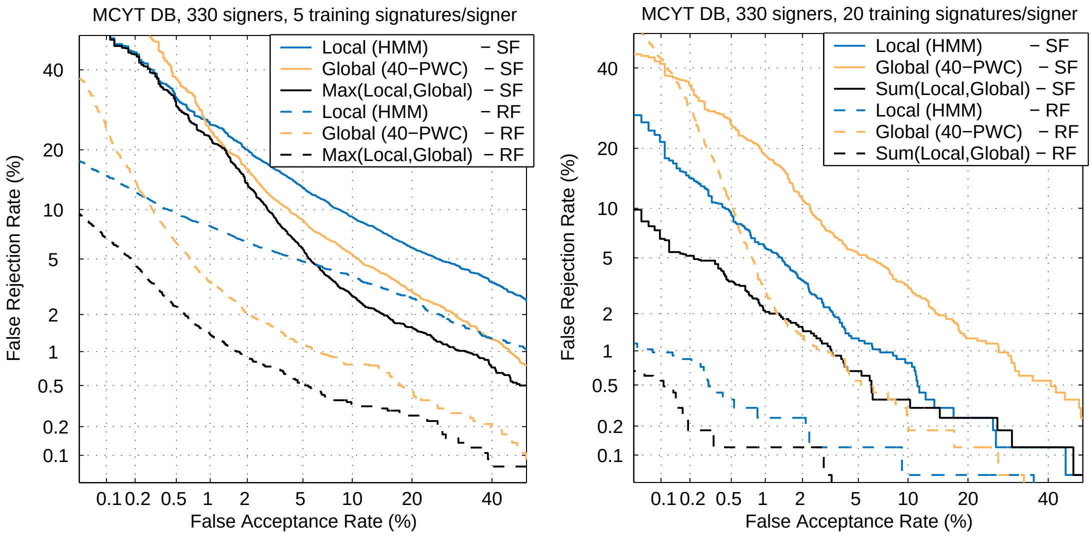


Figure 5.11: *Verification performance of the two individual signature systems as well as their combination using simple fusion rules. Error rates are given both for skilled (SF, solid) and random forgeries (RF, dashed).*

a vertical dashed line. Detection trade-off curves for this working point are given in Fig. 5.11.

Verification performances of individual and combined systems for *a posteriori* user-dependent decision thresholds are given in Tables 5.3 and 5.4. User-dependent decision thresholding leads to error rates significantly lower than user-independent decisions, which corroborates the previous results of score normalization in the local system. When using user-dependent thresholds and for the four conditions considered, the local approach is found to outperform the global one and the *sum* rule performs better than the *max* rule. Also remarkably, the global approach is found to be robust to the score misalignment produced by the strong user-dependencies found in signature recognition, as the difference in performance between user-dependent and user-independent thresholds is not as high as the one found for the local approach.

Table 5.3: *Verification performance with 5 training signatures for* a posteriori *user-independent and user-dependent decision thresholding. Average EERs in %.*

	skilled forgeries		random forgeries	
	user-indep.	user-dep.	user-indep.	user-dep.
Local (HMM)	9.39	2.51	4.86	0.59
Global (40 Feat. + PWC)	6.89	5.61	2.02	1.27
Combined (MAX)	**5.29**	2.39	**1.23**	0.41
Combined (SUM)	6.67	**2.12**	2.14	**0.24**

Table 5.4: Verification performance with 20 training signatures for a posteriori *user-independent and user-dependent decision thresholding. Average EERs in %.*

	skilled forgeries		random forgeries	
	user-indep.	user-dep.	user-indep.	user-dep.
Local (HMM)	2.60	**0.51**	0.39	0.0041
Global (40 Feat. + PWC)	5.21	2.38	1.58	0.3180
Combined (MAX)	2.30	0.53	0.33	0.0064
Combined (SUM)	**1.70**	0.55	**0.18**	**0.0005**

5.5. Chapter Summary and Conclusions

In this chapter we have introduced an enhanced version of a signature verification system based on local information and Hidden Markov Models following the previous work reported by Ortega-Garcia *et al.* [2002], and a novel system based on global features jointly developed with Lopez-Peñalba [2006].

Regarding the local system, we have explored various aspects of feature extraction, modeling, and training strategy. In case of feature extraction, we have shown that the inclusion of azimuth and altitude signals worsens verification performance. With respect to modeling based on HMM we have shown that less states than usually found in the literature results in improved performance, thus resulting in a system configured with 2 HMM states and 32 mixtures per state, which outperforms also the single state GMM recognition approach. Experiments on training strategy have shown that incorporating variability in the training signatures improves the verification performance remarkably. We have also observed that 5 training signatures are enough for obtaining robust models.

The local system has also been used to study the different user-dependent score normalization techniques developed in this Thesis, resulting in various experimental findings. Most remarkably, target-centric techniques based on a variation of the cross-validation procedure provided the best performance improvement testing both with random and skilled forgeries. This is corroborated by the good results of the local system in SVC 2004 [Yeung *et al.*, 2004].

The global system presents a novel feature set which is exploited by non-parametric statistical modeling based on Parzen windows. We have shown comparative results for the discriminative capabilities of various combinations of features using a ranking based on individual discriminative power.

Finally, we have combined the local and global systems using simple score level fusion based on *max* and *sum* rules plus user-specific decision thresholds. The machine expert based on global information is shown to outperform the system based on local analysis in the case of small training set size and user-independent thresholds. The global expert is also found to be quite robust to the severe user-dependencies encountered in signature recognition. The two proposed systems are also shown to give complementary recognition information which is well exploited with simple fusion rules. Relative improvements in the verification performance as high as 44% (for skilled forgeries) and 75% (for random forgeries) have been obtained when including the global information to the local expert.

This chapter includes novel contributions in the enhancement of the local system, the application of user-dependent score normalization to the local system, the development of the global system, and the combination of local and global systems.

Chapter 6

Multi-Algorithm Speaker Verification

THIS CHAPTER studies the application of adapted user-dependent score fusion to multilevel speaker recognition (see Fig. 6.1 for the system model). In particular, we study the application of Bayesian adaptation to derive the personalized fusion functions from prior user-independent data, as described in Sect. 3.1.2.1. Experimental results are reported using the MIT Lincoln Laboratory's multilevel speaker verification system on benchmark data from the Speaker Recognition Evaluation 2004 organized by the National Institute of Standards and Technology (NIST SRE 2004). It is experimentally shown that the proposed adapted fusion method outperforms both user-independent and non-adapted user-dependent traditional fusion approaches.

The systems used in this chapter were developed by MIT-LL [Reynolds *et al.*, 2005], and have traditionally been ranked among the best systems in the NIST SRE campaigns [Przybocki and Martin, 2004]. Although we have at the Biometrics Research Lab.–ATVS our own speaker recognition systems [Garcia-Romero *et al.*, 2006; Ramos-Castro *et al.*, 2006a], we have decided to test the proposed techniques on third party systems. This approach shows the direct applicability of the techniques proposed in this Thesis to systems different to those developed or adjusted within the framework of this PhD work.

Therefore, the systems presented in this chapter are not a contribution of this PhD Thesis. The contributions of this chapter are only related to the adapted fusion scheme applied. Application of the fusion techniques proposed in this Thesis to proprietary systems from Biometrics Research Lab.–ATVS are given in Chapters 5, 7 and 8.

This chapter is structured as follows. Related work on multilevel speaker verification is first summarized. The different modules of the MIT-LL multilevel speaker recognition system are then sketched. The corpus used from the NIST SRE 2004 evaluation is then described. Experiments are then reported. The chapter ends with a summary and some conclusions.

This chapter is based on the publication: Fierrez-Aguilar *et al.* [2005a].

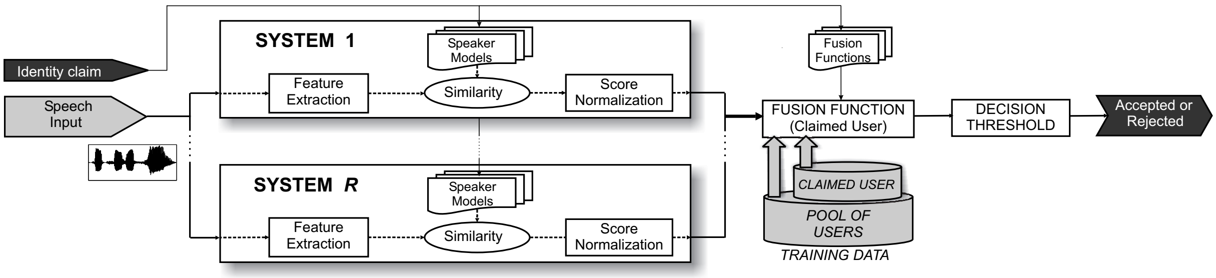


Figure 6.1: *System model of adapted user-dependent multilevel speaker verification.*

6.1. Multilevel Speaker Verification

The state of the art in speaker recognition has been widely dominated during the past decade by the UBM-MAP adapted GMM approach working at the short-time spectral level [Reynolds *et al.*, 2000]. Recently, new approaches based on Support Vector Machines are achieving similar performance [Campbell *et al.*, 2006], working also at the spectral level. These new techniques provide complementary information for the verification task, which has been exploited by the use of score level fusion techniques [Campbell *et al.*, 2006].

On the other hand, higher levels of information conveyed in the speech signal have shown promising discriminative capabilities among speakers, and are a major goal of present speaker recognition research efforts. Some examples in this regard are the SuperSID project [Reynolds *et al.*, 2003], and the NIST Speaker Recognition Evaluations (SRE) [Przybocki and Martin, 2004]. Since the inclusion of the extended data task in the NIST SRE 2002, major advances have been done in finding, characterizing and modeling new high-level sources of speaker information. However, once the similarity scores from each individual system have been computed, little emphasis has been placed in developing new fusion approaches that take into account the well-known speaker specificities present in some groups of subjects [Doddington *et al.*, 1998].

Motivated by these speaker-dependent specificities [Doddington *et al.*, 1998], the present chapter is focused on the application of user-dependent fusion techniques to multilevel speaker verification.

6.2. Baseline Systems

In the present chapter, the scores submitted by the MIT-LL for the NIST SRE 2004 extended data task have been used [Reynolds *et al.*, 2005]. These scores were computed by using seven systems with speaker information from spectral level, pitch, prosodic behavior, and phoneme and word usage. These different types of information were modeled and classified using Gaussian Mixture Models, Support Vector Machines and n-gram language models. In the following list we provide a brief description of the main features of each individual system:

MFCC_GMM. The system is based on short-term acoustic features (Mel Frequency Cepstral

Coefficients [Reynolds, 1994]), and a likelihood ratio detector with target and alternative probability distributions modeled by GMMs [Reynolds *et al.*, 2000]. A Universal Background Model GMM is used as the alternative hypothesis model, and target models are derived using Bayesian adaptation. The techniques of feature mapping [Reynolds, 2003] and t-norm [Auckenthaler *et al.*, 2000] are also used.

MFCC_SVM. The spectral SVM system uses a novel sequence kernel [Campbell *et al.*, 2006]. The sequence kernel compares entire utterances using a generalized linear discriminant. It uses the same front-end processing as the MFCC_GMM system.

PHONE_SVM. The SVM phone system uses a kernel for comparing conversation sides based upon methods from information retrieval. Sequences of phones are converted to a vector of probabilities of occurrence of terms, and co-occurrence of terms (bag of unigram, and bag of bigrams, respectively). A weighting based upon a linearization of likelihoods is then used to compare vectors for SVM training.

PHONE_NGM. A phone n-gram system was developed using the output of the MIT-LL phone recognizer developed with HTK [Young *et al.*, 2002]. This system used the n-gram approach proposed by Doddington [2001].

PROSODY_SLOPE. To capture prosodic differences in the realization of intonation, rhythm, and stress, the $F0$ and energy contours are converted into a sequence of tokens reflecting the joint state of the contours (rising or falling). A n-gram system is used to model and classify distinctive token patterns from token sequences [Adami *et al.*, 2003].

PROSODY_GMM. The aim of this system is to capture the characteristics of the $F0$ and short-term energy features distribution. This system is based on a likelihood ratio detector that uses adapted GMMs for estimating the likelihoods [Adami, 2004].

WORD_NGM. A word n-gram (idiolect) system was developed using the speech-to-text output from the BBN Byblos real-time system. This system used the idiolect word n-gram approach proposed in [Doddington, 2001].

6.3. Database and Experimental Protocol

The experiments presented in this chapter were conducted on the 8sides-1side set of the NIST SRE 2004 corpus [Przybocki and Martin, 2004]. This database comprises conversational telephone speech in five different languages (English, Spanish, Russian, Arabic and Mandarin) over three different channels (Cellular, Cordless and Landline), and four types of transducers (Speaker-phone, Head-mounted, Ear-bud, and Hand-held). Speaker models were trained with 8 single channel conversation sides of approximately five minutes total duration each. Test segments consist of one side of another conversation. All trials were performed between pairs of speakers of the same gender.

In order to provide a development set (DEV) for the experiments, data from Switchboard II phases $1-5$ were used to mimic the conditions in the 8sides-1side set of the NIST SRE 2004 corpus.

The following subsets of the 8sides-1side set were defined for the experiments:

ALL5. All speaker models with at least 5 genuine and 10 impostor attempts. In this way, ALL5 consists of 830 genuine and 4614 impostor attempts of 118 different speaker models.

COMMON5. All speaker models with at least 75% of English enrollment, and at least 5 client and 10 impostor attempts. In this way, COMMON5 consists of 136 genuine and 378 impostor attempts of 19 different speaker models.

Three different types of experiments have been conducted:

User-Independent Fusion. Training on DEV data.

User-Dependent Fusion. For each user and each multilevel test score, 4 different genuine and 9 different impostor multilevel scores of the user at hand are randomly selected (different to the tested one). Local training is performed on the randomly selected multilevel scores. For each multilevel test score, 5 runs of the random sampling are performed.

Adapted User-Dependent Fusion. For each user and each multilevel test score, 4 different genuine and 9 different impostor multilevel scores of the user at hand are randomly selected (different to the tested one). Global training is performed on DEV data whereas local training is carried out on the randomly selected multilevel scores. For each multilevel test score, 5 runs of the random sampling are performed.

6.4. Results

Verification performance of the seven individual systems, along with various user-independent combinations using score level fusion based on Quadratic Discriminant (see Sect. 3.1.2.1), are given in Tables 6.1 and 6.2. Spectral level systems perform remarkably better than the other systems, and their combination with the high-level system WORD_NGM leads to enhanced performance. Worth noting, not all combinations provide improved performance over the best system, and the relative improvement between the best fused system and the best individual system is not very high (10% and 4% on ALL5 and COMMON5 respectively). Finally, performance on COMMON5 is remarkably better than performance on ALL5, specially for the spectral and phonetic systems (60% and 39% relative improvements in the best system of each level respectively).

Verification performance using non-adapted user-dependent fusion is given in Tables 6.3 and 6.4 for the ALL5 and COMMON5 datasets, respectively. The same behavior found in user-independent fusion is also observed here, obtaining similar performance figures. In particular,

Table 6.1: *Verification performance on* ***ALL5*** *data set with* ***user-independent fusion*** *based on Quadratic Discriminant. EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	8.67	**7.39**	*12*	9.28	8.79
	MFCC_SVM	**7.70**		*13*	7.83	6.98
2	PHONE_SVM	16.90	18.21	*14*	7.46	**6.91**
	PHONE_NGM	22.16		*123*	9.05	8.07
3	PROSODY_SLOPE	20.86	16.76	*124*	8.98	8.25
	PROSODY_GMM	22.51		*134*	7.59	6.98
4	WORD_NGM	22.70		*1234*	9.19	7.96

Table 6.2: *Verification performance on* ***COMMON5*** *data set with* ***user-independent fusion*** *based on Quadratic Discriminant. EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	5.98	**3.56**	*12*	3.69	3.06
	MFCC_SVM	**3.06**		*13*	4.32	3.56
2	PHONE_SVM	10.31	10.94	*14*	3.56	**2.93**
	PHONE_NGM	18.32		*123*	3.56	3.56
3	PROSODY_SLOPE	22.14	14.63	*124*	4.32	**2.93**
	PROSODY_GMM	19.08		*134*	3.06	**2.93**
4	WORD_NGM	20.61		*1234*	3.56	3.19

relative improvements between the best fused system and the best individual system are 9% and 12% for ALL5 and COMMON5 datasets, respectively.

Verification performance using the proposed adapted user-dependent fusion approach (with a relevance factor $r = 1$) is given in Tables 6.5 and 6.6 for the ALL5 and COMMON5 datasets, respectively. In this case, all combinations are better than the best individual system, which is outperformed significantly by the best combination (i.e., spectral and lexical systems). In particular, relative improvement between the best fused system and the best individual system are 31% and 61% for ALL5 and COMMON5, respectively. Also worth noting, the unilevel combination of the two spectral level systems gives an interesting combination pair (31% and 34% relative improvement over the best system for ALL5 and COMMON5, respectively). The effect of varying the relevance factor of the adapted fusion scheme on the verification performance is shown in Fig. 6.2. A good working point is found at relevance factor $r = 1$.

Genuine and impostor scatter plots are depicted in Fig. 6.3 for a random data set of the

Table 6.3: *Verification performance on* ***ALL5*** *data set with* ***user-dependent fusion*** *based on Quadratic Discriminant. EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	8.67	**6.84**	*12*	7.86	7.22
	MFCC_SVM	**7.70**		*13*	8.27	8.15
2	PHONE_SVM	16.90	15.74	*14*	8.04	**6.98**
	PHONE_NGM	22.16		*123*	8.08	7.99
3	PROSODY_SLOPE	20.86	18.46	*124*	8.46	7.37
	PROSODY_GMM	22.51		*134*	8.57	8.04
4	WORD_NGM	22.70		*1234*	8.44	8.11

Table 6.4: *Verification performance on* ***COMMON5*** *data set with* ***user-dependent fusion*** *based on Quadratic Discriminant. EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	5.98	**2.95**	*12*	4.40	2.98
	MFCC_SVM	**3.06**		*13*	5.98	4.99
2	PHONE_SVM	10.31	11.60	*14*	5.42	**2.70**
	PHONE_NGM	18.32		*123*	5.60	4.43
3	PROSODY_SLOPE	22.14	18.99	*124*	5.04	2.77
	PROSODY_GMM	19.08		*134*	5.85	3.66
4	WORD_NGM	20.61		*1234*	5.60	3.66

error estimation process. Global, local and adapted fusion function boundaries (i.e., $f(\mathbf{x}) = 0$) are also depicted. Finally, verification performance results comparing individual systems to the studied fusion strategies are summarized in Fig. 6.4.

6.5. Discussion

It can be argued against user-dependent fusion that training data scarcity is a major drawback. In this chapter, it has been demonstrated that the performance of a state-of-the-art multilevel speaker verification system by a third party is significantly improved in an standard evaluation scenario by considering user-dependent information at the fusion level. This has been achieved by using a novel user-dependent fusion technique based on Bayesian adaptation of the fusion functions and only a few training score samples from each user.

Nevertheless, although we have used an un-biased cross-validation experimental procedure,

Table 6.5: *Verification performance on **ALL5** data set with **adapted user-dependent fusion** based on Quadratic Discriminant ($r = 1$). EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	8.67	**5.35**	*12*	6.25	5.66
	MFCC_SVM	**7.70**		*13*	5.85	5.40
2	PHONE_SVM	16.90	13.61	*14*	6.14	**5.36**
	PHONE_NGM	22.16		*123*	5.92	5.39
3	PROSODY_SLOPE	20.86	15.08	*124*	6.72	5.61
	PROSODY_GMM	22.51		*134*	5.95	**5.32**
4	WORD_NGM	22.70		*1234*	6.16	5.37

Table 6.6: *Verification performance on **COMMON5** data set with **adapted user-dependent fusion** based on Quadratic Discriminant ($r = 1$). EERs in %.*

information level	system label	individual performance	unilevel fusion	multilevel fusion		
				levels	best/level	all/level
1	MFCC_GMM	5.98	**2.03**	*12*	2.80	2.06
	MFCC_SVM	**3.06**		*13*	2.37	2.27
2	PHONE_SVM	10.31	8.70	*14*	2.49	**1.20**
	PHONE_NGM	18.32		*123*	2.77	2.11
3	PROSODY_SLOPE	22.14	15.65	*124*	2.92	1.68
	PROSODY_GMM	19.08		*134*	1.91	1.66
4	WORD_NGM	20.61		*1234*	2.42	1.32

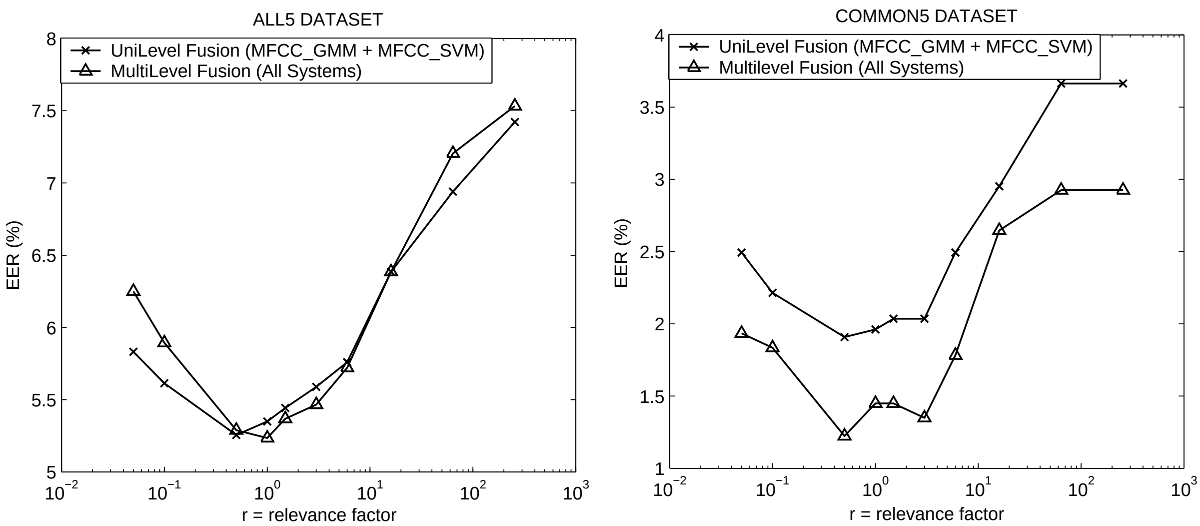


Figure 6.2: *Verification performance of the adapted fusion scheme on ALL5 (left) and COMMON5 (right) data sets for varying relevance factor.*

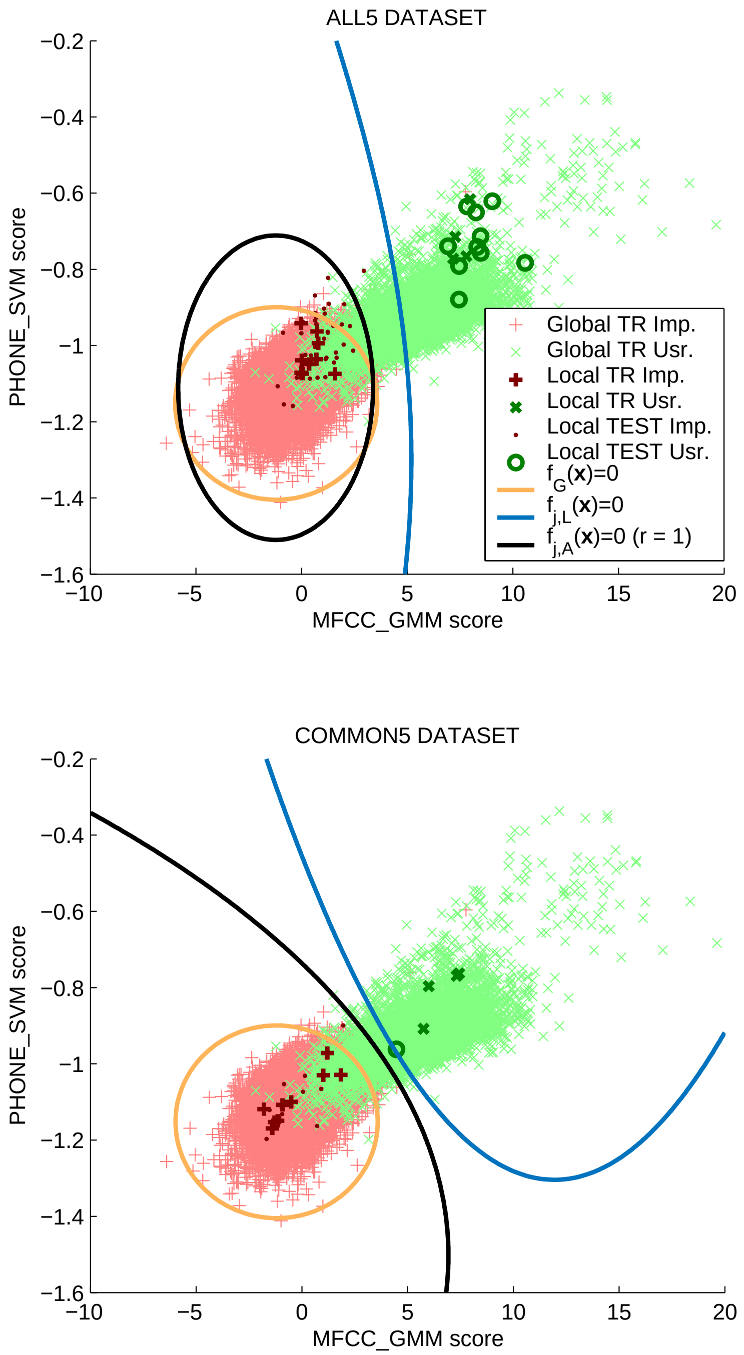


Figure 6.3: *Training/testing 2D scatter plot and decision boundaries of global, local, and adapted approaches for multilevel fusion (one iteration of the error estimation process).*

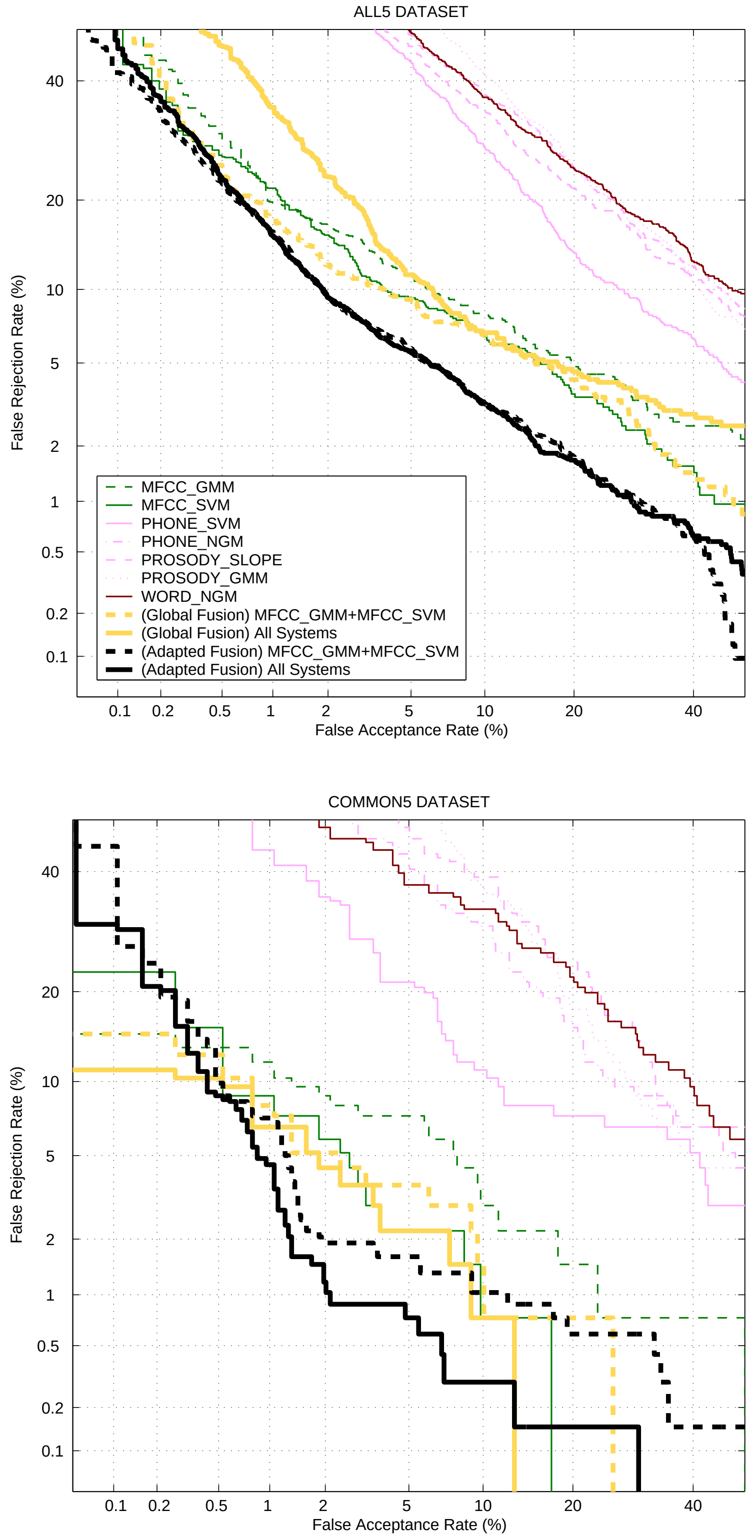


Figure 6.4: *Verification performance of the individual systems and the adapted fusion scheme on ALL5 (left) and COMMON5 (right) data sets.*

it must be emphasized that we have used post-evaluation results for adapting to the user specificities. The study of the case of using only the available training data is to be addressed in future work. In this regard, it is our belief that for the case of large training set size (such as the 8sides-1side or above scenarios defined by NIST), the use of resampling techniques (e.g., resubstitution, leave-one-out, bootstrap, etc. [Jain *et al.*, 2000a]) may result in a significant improvement. As a preliminary justification for this aim, we point out the related work presented in Sect. 5.2.3.2, where resampling techniques were successfully applied to the problem of *a priori* user-dependent score normalization in signature verification.

6.6. Chapter Summary and Conclusions

In this chapter we have evaluated the adapted Bayesian user-dependent fusion scheme presented in Chapter 3. This study has been conducted by using the multilevel speaker verification system from MIT Lincoln Laboratory on data from the NIST SRE 2004 benchmark.

We have compared user-independent, user-dependent, and adapted user-dependent versions of score level fusion based on Quadratic Discriminants (see Sect. 3.1.2.1). It has been shown that the proposed adapted approach outperforms both user-independent and user-dependent traditional fusion schemes. The new approach balances the information provided by a pool of subjects and the user-specific information by using a simple relevance factor.

This chapter includes novel contributions in the application of adapted score fusion but not in the individual systems used.

Chapter 7

Multi-Algorithm Fingerprint Verification

Within biometrics, automatic fingerprint recognition is receiving great attention because of the commonly accepted distinctiveness of the fingerprint pattern, the wide-spread deployment of electronic acquisition devices, and the wide variety of practical applications ranging from access control to forensic identification [Maltoni *et al.*, 2003].

The effect of image quality on the performance of fingerprint verification is studied in this chapter. In particular, we investigate the performance of two fingerprint matchers based on minutiae and ridge information as well as their score-level combination under varying fingerprint image quality. The ridge-based system is found to be more robust to image quality degradation than the minutiae-based system. We exploit this fact by introducing an adapted score fusion scheme based on automatic quality estimation in the frequency domain. The proposed scheme leads to enhanced performance over a wide range of fingerprint image quality and decision thresholds.

This chapter is structured as follows. We first summarize related works on the characterization of fingerprint image quality, and we describe the fingerprint image quality measure used in this chapter. We then outline the individual fingerprint matching systems used. After that we detail the quality-based fusion scheme applied. The experimental setup and results are then described. The chapter ends with a summary and some conclusions.

The adapted quality-based fusion scheme used in this chapter is the simplest of the ones presented in Chapter 3. In Chapter 8 we will provide a general comparison among adapted fusion schemes, both user-dependent and quality-based.

The quality measure used in this chapter is the one proposed by Chen *et al.* [2005], therefore it is not a contribution of this Thesis. In the same way, the minutiae-based system is the one developed and described by Simón-Zorita [2004]. The contributions of this chapter are related to the ridge-based fingerprint matcher, which has been developed jointly with Muñoz-Serrano [2005], the study of the quality effects on the verification performance, and the quality-based fusion approach.

This chapter is based on the publications: Fierrez-Aguilar *et al.* [2006, 2005e].

7.1. Assessment of Fingerprint Image Quality

Our first objective in this chapter is to investigate the effects of varying image quality on the performance of automatic fingerprint recognition systems [Simon-Zorita *et al.*, 2003]. This is motivated by the results of the last Fingerprint Verification Competition [Cappelli *et al.*, 2006]. In this competition fingerprint images with lower image quality than those in earlier benchmarks were used. As a result, the error rates of the best systems were found to be an order of magnitude worse than those reported previously. Similar effects have also been noticed in other recent comparative benchmark studies [Wilson *et al.*, 2004].

Local image quality estimates have been traditionally used in the segmentation and enhancement steps of fingerprint recognition [Hong *et al.*, 1998]. On the other hand, global quality measures have been traditionally used as indicators to identify invalid images. The invalid images are then considered as failure to enroll or failure to acquire events that are handled either manually or automatically [Maltoni *et al.*, 2003].

More recently, there is increasing interest in assessing the fingerprint image quality for a wider variety of applications. Some examples include: study of the effects of image quality on verification performance [Simon-Zorita *et al.*, 2003], comparison of different sensors based on the quality of the images generated [Yau *et al.*, 2004], comparison of commercial systems with respect to robustness to noisy images [Wilson *et al.*, 2004], impact of image quality on forensic reports [Gonzalez-Rodriguez *et al.*, 2005], and many other applications in law enforcement and standardization [BQW, 2006].

A number of fingerprint quality measures have been proposed in the literature. Most of them are based on operational procedures for computing coherence measures of local orientation [Bigun, 2006; Bigun *et al.*, 1991]. Some examples include: local Gabor-based filtering [Hong *et al.*, 1998; Shen *et al.*, 2001], local and global spatial features [Lim *et al.*, 2002], directional measures [Ratha and Bolle, 2004], classification-based approaches [Tabassi *et al.*, August 2004], and local measures based on intensity gradient [Chen *et al.*, 2005]. In the present work we use the global quality index computed in the frequency domain proposed by Chen *et al.* [2005], which is summarized below.

7.1.1. Fingerprint Image Quality Index

Good quality fingerprint images bear a strong ring pattern in the power spectrum, indicating a dominant frequency band associated with the period of the ridges. Conversely, in poor quality images the ridges become unclear and non-uniformly spaced, resulting in a more diffused power spectrum. We thus assess the global quality of a fingerprint image by evaluating the energy distribution in the power spectrum.

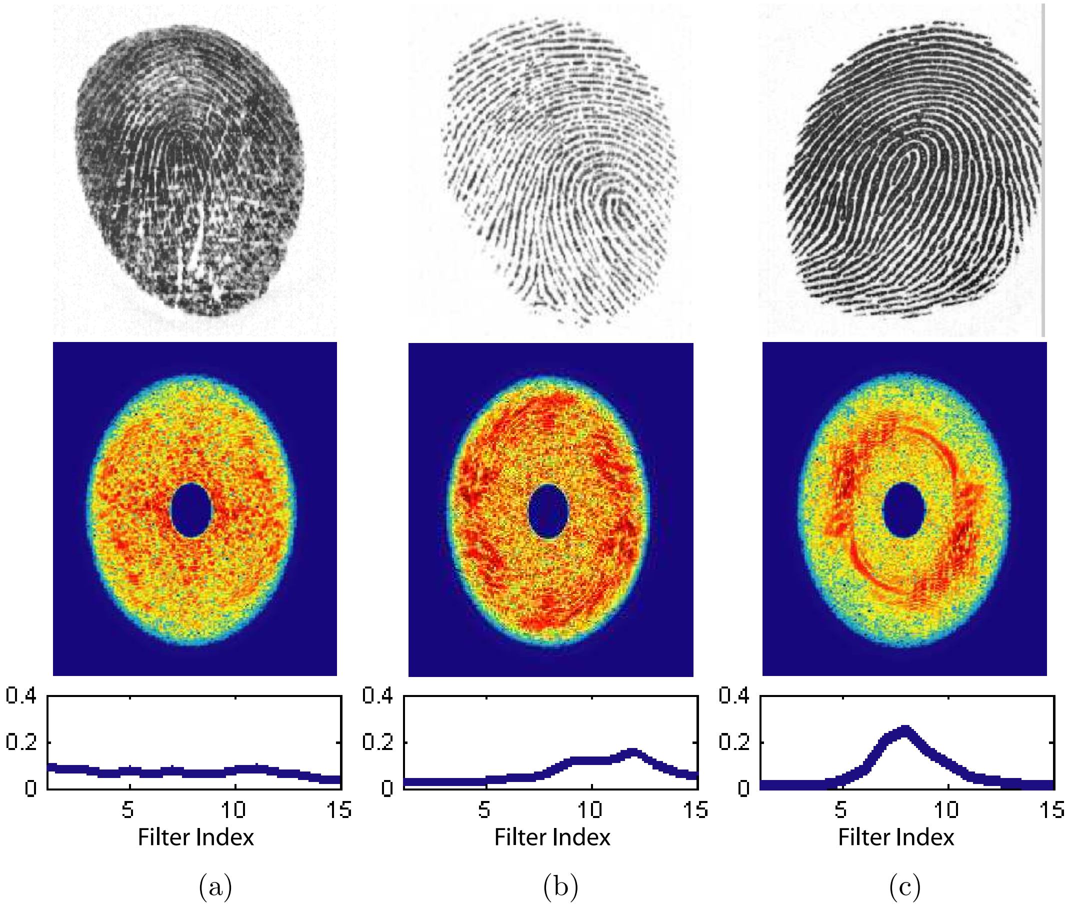


Figure 7.1: *Three sample fingerprint images from MCYT signature database with increasing image quality from left to right (top row), their corresponding power spectrum (middle row), and their energy distribution across concentric rings in the frequency domain. It can be observed that the better the fingerprint quality, the more peaked is its energy distribution, indicating a more distinct dominant frequency band. The resulting quality measure for each fingerprint image from left to right is 0.05, 0.36, and 0.92, respectively.*

A region of interest (ROI) in the power spectrum is defined to be a ring-shaped band with radius ranging from the minimum to the maximum observed frequency of ridges [Chen *et al.*, 2005]. Fig. 7.1 shows three fingerprint images from MCYTDB (see Sect. 4.3.1) with increasing quality from left to right. Their corresponding power spectrums are shown in the second row. Note that the fingerprint image with good quality presents a strong ring pattern in the power spectrum (Fig. 7.1(c)), while a poor quality fingerprint presents a more diffused power spectrum (Fig. 7.1(a)). Multiple bandpass filters are designed to extract the energy in a number of ring-shaped concentric sectors in the power spectrum. The global quality index is defined in terms of the energy concentration across these sectors within the ROI.

In particular, bandpass filters are constructed by taking differences of two consecutive Butterworth functions [Chen *et al.*, 2005]. In the third row of Fig. 7.1, we plot the distribution of the normalized energy across the bandpass filters. The energy distribution is more peaked as

Figure 7.2: *Processing steps of the minutiae-based matcher.*

the image quality improves from (a) to (c). The resulting quality measure Q is based on the entropy of this distribution, which is normalized linearly to the range $[0, 1]$.

7.2. Fingerprint Matcher Based on Minutiae

The minutiae-based matcher follows the approach presented by Jain *et al.* [1997] with the enhancements detailed by Simón-Zorita [2004]. The output of the minutiae-based matcher is a similarity measure based on dynamic programming. Below we provide a summary of the processing steps carried out by this matcher.

Image enhancement. The fingerprint ridge structure is reconstructed by using: 1) grayscale level normalization, 2) orientation field calculation, 3) interest region extraction, 4) spatial-variant filtering according to the estimated orientation field, 5) binarization, and 6) ridge profiling.

Feature extraction. The minutiae pattern is obtained from the binarized profiled image as follows: 1) thinning, 2) removal of structure imperfections from the thinned image using mathematical morphology, and 3) minutiae extraction. For each detected minutia, the following parameters are stored: *i*) the x and y coordinates of the minutia, *ii*) the orientation angle of the ridge containing the minutia, and *iii*) the x and y coordinates of 10 samples of the ridge segment containing the minutia. An example fingerprint image is shown in Fig. 7.2 together with the feature extraction steps.

Pattern comparison. Given a test and a reference minutiae pattern, a matching score s is computed. First, both patterns are aligned based on the minutia whose associated sampled ridge is most likely to be the same. The matching score is computed then by using a variant of the edit distance on polar coordinates and based on a size-adaptive tolerance box. When

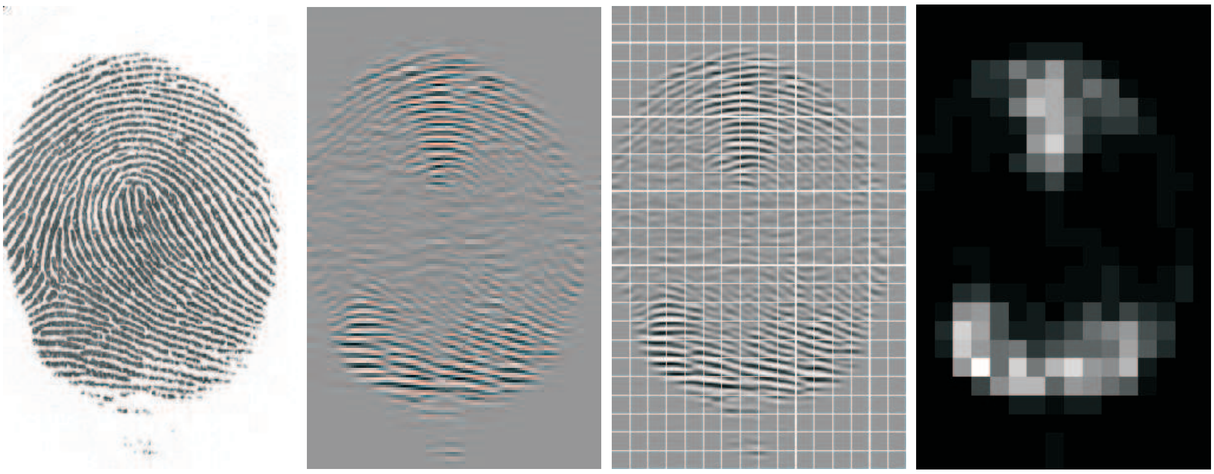


Figure 7.3: *Processing steps of the texture-based matcher.*

more than one reference minutiae pattern per client model are considered, the maximum matching score obtained by comparing the test and each reference pattern is used.

Score normalization. In order to generate a normalized similarity score x between 0 and 1, the matching score s (which is greater than or equal to zero) is further normalized according to

$$x = \tanh\left(c_{\text{minutiae}} \cdot s\right), \tag{7.1}$$

where the parameter c_{minutiae} was chosen heuristically on fingerprint data not used for the experiments reported here.

7.3. Fingerprint Matcher Based on Texture

The ridge-based matcher (also referred to as texture-based) is based on the correlation of Gabor-filter energy responses in a squared grid as proposed by Ross *et al.* [2002] with some modifications as detailed by Muñoz-Serrano [2005]. The result is a dissimilarity measure based on Euclidean distance. Below we provide a summary of the processing steps carried out by this matcher.

Image enhancement. No image enhancement is performed since Gabor filters are robust enough to the typical noise present in the fingerprint images.

Feature extraction. The ridge-based matcher uses a set of 8 Gabor filters to capture the ridge strength in different orientations. The 8 filtered images corresponding to the different orientations are tessellated into square cells. The variance within each cell is computed to form a feature vector with 8 bands. This multi-band feature vector is called Fingercode

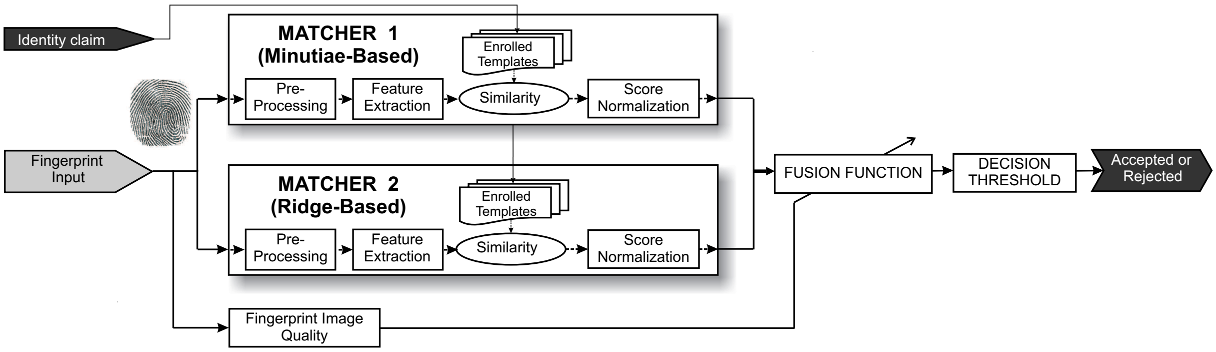


Figure 7.4: *Quality-based multi-algorithm approach for fingerprint verification.*

because of the similarity to previous research works [Jain *et al.*, 2000b]. A sample fingerprint image, the resulting filtered image with an horizontal Gabor filter, the tessellated image, and its corresponding Fingercode band are shown in Fig. 7.3.

Pattern comparison. The matching score is computed as the Euclidean distance between the input and the claimed Fingercodes after an alignment step. To determine the optimal alignment between two Fingercodes, the 2D correlation of the two Fingercodes is computed in the Fourier domain as described by Ross *et al.* [2002]. Although this procedure does not account for rotational offset between the two fingerprints, we have observed that typical rotations between different impressions of the same fingerprint from the MCYT database are typically compensated by using the tessellation.

Score normalization. Distance scores are normalized into similarity matching scores by using the following normalization function

$$x = \exp\left(-c_{\text{texture}} \cdot s\right), \tag{7.2}$$

where normalization parameter c_{texture} is a positive real number chosen heuristically in order to have the normalized scores of the system spread out over the $[0, 1]$ range.

7.4. Quality-Based Score Fusion

The proposed quality-based multi-algorithm approach for fingerprint verification follows the system model depicted in Fig. 7.4.

Let the normalized similarity scores provided by the two matchers be x_M (minutiae-based) and x_T (texture-based). The fused result using the sum rule is $y = (x_M + x_T)/2$ (see Sect. 2.2.2.1).

We assume that verification performance of one of the algorithms drops significantly as compared to the other one under image quality degradation. This behavior is observed in our minutia-based matcher with respect to our ridge-based matcher, as will be demonstrated in the experiments. We evaluate here a simple adapted fusion scheme based on weighted average (see

Sect. 3.2.1). In this case the weights are adjusted depending on the input image quality as follows

$$y = \frac{q}{2}x_M + (1 - \frac{q}{2})x_T, \tag{7.3}$$

with $q = \sqrt{Q \cdot Q_{\text{claim}}}$, where Q and Q_{claim} are the input biometric quality and the average quality of the biometric signals used for enrollment, respectively. As the input image quality Q worsens (i.e., $q \to 0$), more importance is given to the matching score of the more robust system (i.e., $y \to x_T$). Conversely, as the input image quality improves (i.e., $q \to 1$), the fusion function converges to the sum rule.

7.5. Experiments

7.5.1. Database and Experimental Protocol

We use a subcorpus of the MCYT Bimodal Biometric Database for our study (see Sect. 4.3.1). Data consist of 7500 fingerprint images from all the 10 fingers of 75 subjects acquired with the optical sensor. We consider the different fingers as different users enrolled in the system, resulting in 750 users with 10 impressions per user. Some example images are shown in Fig. 7.1.

We use one impression per finger as template (with low control during the acquisition, see Sect. 4.3.1). Genuine matchings are obtained comparing the template to the other 9 impressions available. Impostor matchings are obtained by comparing the template to one impression of all the other fingers. The total number of genuine and impostor matchings are therefore 750×9 and 750×749, respectively.

We further classify all the fingers in the database into five equal-sized quality groups, from I (low quality), to V (high quality), based on the quality measure Q described in Sect. 7.1, resulting in 150 fingers per group. Each quality group contains 150×9 genuine and 150×749 impostor matching scores.

Distribution of fingerprint quality indices and matching scores for the two systems considered are given in Fig. 7.5.

7.5.2. Results

Verification performance results are given in Fig. 7.6 for the individual matchers (minutiae- and texture-based), their combination through the sum fusion rule, and the proposed quality-based weighted sum for different quality groups. We observe that the texture-based matcher is quite robust to image quality degradation. Conversely, the minutiae-based matcher degrades rapidly with low quality images. As a result, the fixed fusion strategy based on the sum rule only leads to improved performance over the best individual system in medium to good quality images.

The proposed adapted fusion approach results in improved performance for all the image quality groups, outperforming the standard sum rule approach, specially in low image quality

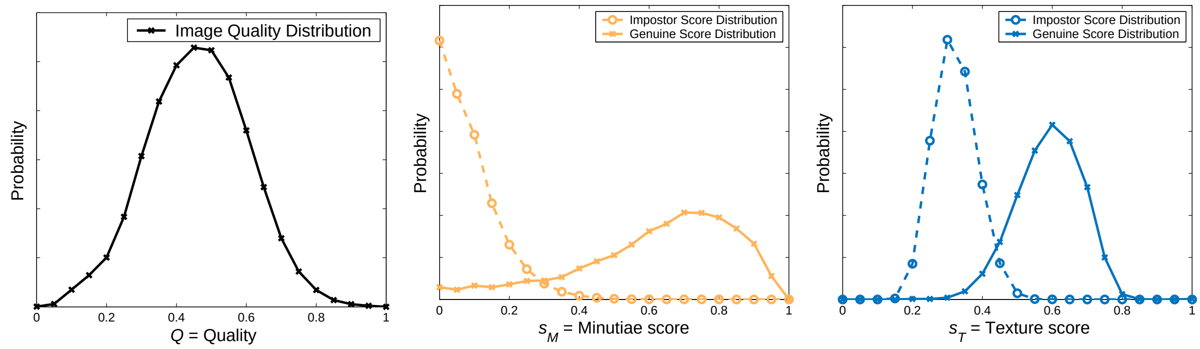


Figure 7.5: *Image quality distribution in the database (left) and matching score distributions for the minutiae (center) and texture matchers (right).*

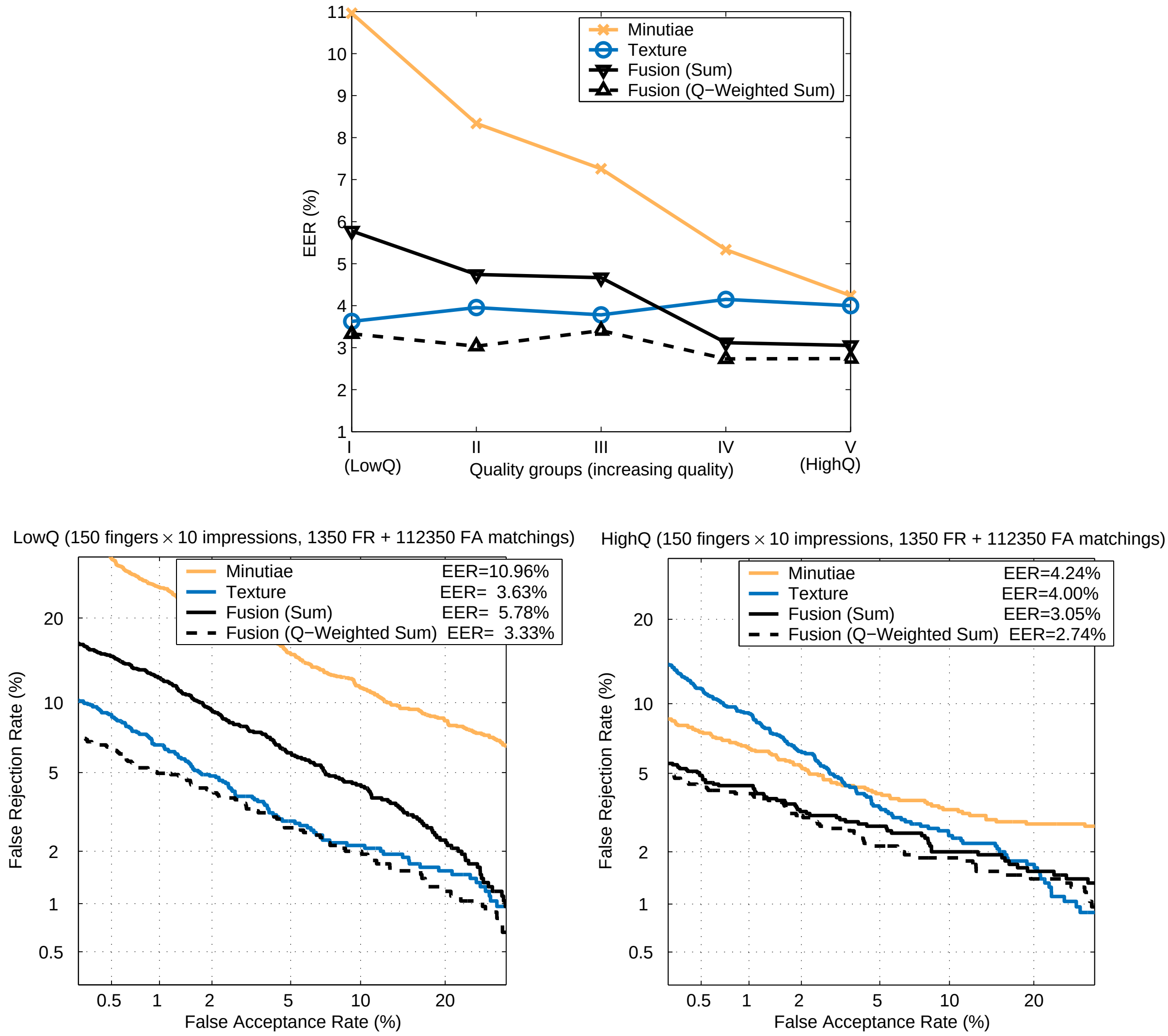


Figure 7.6: *Verification performance of the individual matchers (minutiae- and texture-based), their combination through the sum fusion fusion rule, and the proposed quality-based weighted sum for increasing image quality.*

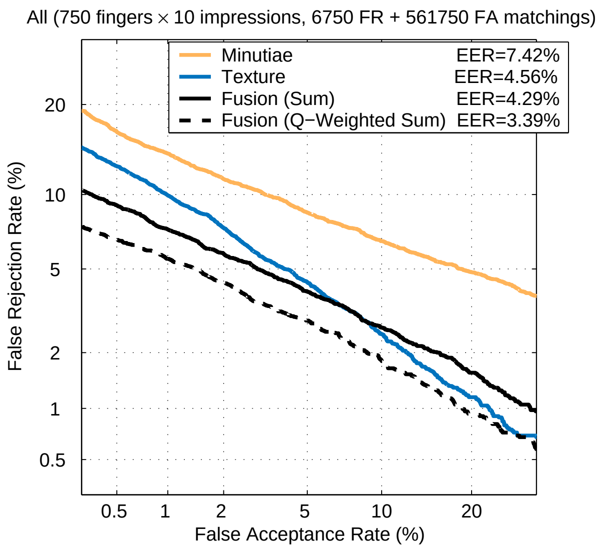


Figure 7.7: *Verification performance for the whole database.*

conditions where the performance of the individual matchers becomes more different.

Finally, in Fig. 7.7 we plot the verification performance for the whole database. Relative verification performance improvement of about 20% is obtained by the proposed adapted fusion approach for a wide range of verification operating points as compared to the standard sum rule.

7.6. Chapter Summary and Conclusions

The effects of image quality on the performance of two common approaches for fingerprint verification have been studied. It has been found that the approach based on ridge information outperforms the minutiae-based approach in low image quality conditions. Comparable performance is obtained on good quality images.

It must be emphasized that this evidence is based on particular implementations of well known algorithms. Other implementations may lead to improved performance of any approach over the other for a given image quality. On the other hand, the robustness of the ridge-based approach as compared to the minutiae-based system has been observed in other studies. One example of this behavior is found in FVC 2004 [Cappelli *et al.*, 2006], where low quality images were used and leading systems used some kind of ridge information [Fierrez-Aguilar *et al.*, 2005g].

This difference in robustness against varying image quality has been exploited by an adapted score-level fusion approach using quality measures estimated in the frequency domain. The proposed scheme leads to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of fingerprint image quality and decision thresholds.

This chapter presents novel contributions in the system based on texture, the study of the image quality effects on minutiae- and texture-based systems, and the quality-based fusion scheme applied to multi-algorithm fingerprint verification.

Chapter 8

User-Dependent and Quality-Based Multimodal Authentication

This chapter studies the application of the proposed adapted score fusion schemes to multimodal biometric authentication based on fingerprint and written signature. In particular, we combine the on-line signature verification system based on local information described in Sect. 5.2 with the minutiae-based fingerprint verification system described in Sect. 7.2.

The chapter starts with a summary of the score fusion methods studied, both user-dependent and quality-based. This is followed by a description of the experimental setup, which is based on a worst-case scenario on the MCYT database. We then present the results. The chapter ends with some conclusions.

This chapter is based on the publications: Bigun *et al.* [2003]; Fierrez-Aguilar *et al.* [2004b, 2005b,c, 2004d, 2005i].

8.1. Methods

The methods studied in this chapter are divided into user-dependent fusion (see Sect. 3.1), and quality-based fusion (see Sect. 3.2). These two classes of fusion schemes are studied independently. For each class, we study the two schemes developed in Chapter 3, the first one based on Support Vector Machines (SVM) and the other one based on Bayesian adaptation.

When training Support Vector Machines, the problem in Eqs. (3.16) and (3.17) is solved by using the decomposition algorithm proposed by Osuna *et al.* [1997], and the interior point optimization solver developed by Vandervei [1999].

User-dependent fusion. In order to study the benefits of adapting the score fusion functions to individual users, we compare the following versions of score level fusion: 1) user-independent (global, see Fig. 3.1), 2) user-dependent (local, see Fig. 2.4), and 3) adapted user-dependent (adapted, see Fig. 3.3). The first scheme studied is based on SVM (see Sect. 3.1.2.2). The second scheme studied is based on Bayesian adaptation (see Sect. 3.1.2.1). As the Bayesian adaptation results in non-linear separating surfaces (i.e., quadratic functions, see Eq. (3.10)), we use non-linear radial basis function kernels for the SVM as expressed in Eq. (3.18) with $\sigma = 0.05$. The information for the adaptation consists of a reduced number of user-specific matching scores, between 1 and 6 scores per class (client and impostor).

Quality-based fusion. In order to study the benefits of adapting the score fusion functions to the input biometric quality of individual verification attempts, we compare the following versions of score level fusion: 1) user-independent without considering the input quality (see Fig. 3.1), and 2) user-independent considering the input quality (see Fig. 3.5). The first scheme studied is based on SVM (see Sect. 3.2.3). The second scheme is based on Expert Conciliation using Bayesian statistics (see Sect. 3.2.2.3). As the Expert Conciliation scheme results in piecewise linear separating surfaces (i.e., the intersection of two linear separating surfaces, see Eq. (2.11)), we use linear kernels for the SVM as expressed in Eq. (3.19). The information for the adaptation consists of the manual quality label of the fingerprint images introduced in Sect. 4.3.1 normalized to the range $[0, 2]$. All the written signatures are assumed to be of uniform quality, i.e., input quality equal to 1.

8.2. Experimental Protocol

In the experiments reported in this chapter we combine the on-line signature verification system based on local information described in Sect. 5.2 with the minutiae-based fingerprint verification system described in Sect. 7.2. The similarity scores of the fingerprint system have been mapped to probabilities by using fixed score normalization based on hyperbolic functions, see Eq. (2.17). The similarity scores from the signature system have been normalized using an exponential function, see Eq. (2.15). The coefficients for the normalization functions are calculated using biometric data from the subjects in MCYT not used in the present experiments.

8.2.1. Database Description

We use 10 impressions of one finger and 17 signatures of each one of the first 75 subjects from the MCYT bimodal database (see Sect. 4.3).

In order to highlight the benefits of the proposed approaches in an scenario showing both user-dependencies and non-uniform image quality, lowest quality finger was used for 10% of the users and highest quality finger was used for the remaining users. In this way we simulate the scenario observed in other large-scale experiments [Wilson *et al.*, 2004]. Other scenarios (e.g., index fingers or 5% lowest quality) can be found in related publications [Fierrez-Aguilar *et al.*, 2005i]. The quality labeling was done manually by a human expert [Simon-Zorita *et al.*, 2003].

For each user, 3 fingerprints are used for fingerprint enrollment and the other 7 are used for testing. A worst-case scenario has been considered by using as impostor data, for each user, the best 10 impostor fingerprints from a pool of 750 different fingers. For each user, 10 genuine signatures are used for enrollment, the other 7 genuine signatures are used for testing, and 10 skilled forgeries from 5 different impostors are used as impostor testing data.

As a result, data for evaluating the proposed fusion strategies consist of 75×7 genuine and 75×10 impostor bimodal attempts in a worst-case scenario.

8.2.2. Experimental Procedure for User-Dependent Fusion

In the case of user-dependent fusion using SVM, the parameters used are $C = 100$ for client scores, and $C = 50$ for impostor scores. For a detailed study of the effect of the parameters we refer the reader to Fierrez-Aguilar *et al.* [2004b].

The experimental procedure for comparing global, local, and adapted versions of the two fusion schemes (i.e., Bayesian and SVM) is as follows:

Global fusion/decision. Bootstrap data sets have been created by randomly selecting M users from the training set with replacement. This selection process has been repeated independently 200 times to yield 200 different bootstrap data sets. Each data set is used then to generate either a user-independent fusion rule or a user-independent decision function. In the latter case, a non-trained sum rule fusion function is assumed and the selected training data are used for training the decision function on combined scores. Testing is finally performed on the remaining users not included in each bootstrap data set.

Local fusion/decision. For each user, 50 bootstrap data sets have been created, selecting randomly N samples without replacement and forcing half of them in each class client/impostor. For each user and bootstrap data set, a different fusion rule (or a decision function on summed scores) is constructed. Testing is performed on the remaining samples not included in the bootstrap data set.

Adapted fusion/decision. Bootstrap sampling of users is performed as in the global case yielding 200 global bootstrap data sets (GBD). Multimodal scores of the remaining users not included in each GBD are then sampled as in the local case. This yields 50 local bootstrap data sets (LBD) per GBD and per client not included in the GBD. Training of the fusion function (or the decision function on summed scores) is performed using the LBD and associated GBD from which the user was left out. Testing is performed on the remaining samples not included in each LBD.

8.2.3. Experimental Procedure for Quality-Based Fusion

All fingerprint images were supervised and labelled according to the image quality by a human expert [Simon-Zorita *et al.*, 2003]. Each fingerprint image was assigned a subjective quality measure from 0 (lowest quality) to 9 (highest quality) based on image factors like: incomplete fingerprint, smudge ridges or non uniform contrast, background noise, weak appearance of the ridge structure, significant breaks in the ridge structure, pores inside the ridges, etc. Fig. 4.5

shows four fingerprints and their corresponding quality labels. These quality measures are linearly mapped to the range $[0, 2]$. In case of written signature, uniform quality $Q = 1$ is used for all signatures.

In the case of quality-based fusion using SVM, the parameters used are $\alpha_1 = 0.5$, $\alpha_2 = 1$ and $C = 100$. For a detailed study of the effect of the parameters we refer the reader to Fierrez-Aguilar *et al.* [2004d].

The experimental procedure is the same described in the previous section for global learning based on bootstrap sampling with $M = 50$. As demonstrated by Bigun *et al.* [2003], this number of users for training the global fusion function is sufficient to get stable results. In order to demonstrate the benefits of including the quality measures into the fusion process, we carry out this user-independent fusion evaluation both with and without considering the input quality measures.

We also compare multimodal results to repeated-instance experiments using individual traits. This is motivated by the recent debate which criticizes the common practice in multimodal biometrics experiments [Bowyer, 2003]. This common practice is to show multimodal authentication performance results in comparison with the performance results using the individual traits. This is criticized as the amount of input information in both experiments is not balanced. The suggested method in order to demonstrate the benefits of incorporating multiple traits is to compare multimodal to repeated-instance experiments on the individual traits. The repeated-instance experimental protocol is as follows:

Repeated-instance fusion. For each user and considering one of the two biometric traits, 7 random matches between genuine scores and 10 random matches between impostor scores are computed (not permitting a match between a score and itself) so as to obtain 7 genuine and 10 impostor score pairs each one corresponding to two independent verification attempts. The two attempts are combined by using the same procedure used for multimodal fusion.

8.3. Results

8.3.1. Results for User-Dependent Fusion

Comparative results of global, local, and adapted fusion/decision functions are given in Fig. 8.1.

In Fig. 8.1 (c) we plot the verification performance of the bimodal authentication system using the proposed trained SVM-based global fusion approach for an increasing number of clients in the fusion function training set. Individual performances of the signature and fingerprint subsystems, and the non-trained sum rule fusion approach are also shown for reference. In this case, baseline equal error rate of the simple fusion approach based on sum rule, 2.28% EER, is improved to 1.39% by using the global SVM-based trained fusion scheme ($M = 74$ users for training the fusion function). We also observe a rapid performance improvement for the first 10 users and stable results for more than 20 users in the training set. Similar effects are observed for the Bayesian adaptation approach depicted in Fig. 8.1 (f). The main difference is the performance drop in the Bayesian scheme with small training set size.

In Fig. 8.1 (a) we compare local approaches for training either the fusion function or the decision function using SVM. It is shown that using training data for learning local fusion functions (1.12% EER for $N = 12$ training samples per user) is significantly better than using a simple common fusion rule and exploiting existing development data for training localized decisions (2.07% EER). The local fusion approach (1.12% EER) also outperforms the global fusion strategy in Fig. 8.1 (c) (1.39% EER) when enough training samples for building the user-specific fusion functions are available (approximately more than 6 in this experiment, i.e., 3 scores per class client/impostor). Local learning for the scheme based on Bayesian adaptation is shown in Fig. 8.1 (d). The performance in this case is similar to the SVM approach when using $N = 12$ for training the user-specific fusion functions, but the performance deteriorates more rapidly with less training samples. This method cannot be applied for less than 3 similarity scores per class client/impostor (i.e., $N = 6$), which are needed for estimating the sufficient statistics of each class. This comparison favors the SVM approach for small training set size, at the cost of higher computational complexity.

In Fig. 8.1 (b) we show the verification results of the proposed adapted approaches using the SVM scheme. In this case, $M = 74$ clients (global) and $N = 12$ samples per client (local) are used for training and α is varied, hence trading off the influence of the global and local information for training the fusion/decision functions. As a result, a minimum of 1.69% EER is found for $\alpha = 0.75$ in the case of sum rule fusion and adapted decisions, outperforming the local decision scheme in Fig. 8.1 (a) (2.07%). Adapted fusion outperforms all other strategies lowering the error rate down to 0.79% also for $\alpha = 0.75$. The same behavior is observed for the Bayesian adaptation scheme in Fig. 8.1 (e), with a maximum of performance of 0.53% EER for relevance $r = 2.5$ (55% relative performance improvement with respect to the local approach), which outperforms the SVM approach in the same conditions. In Fig. 8.1 (e) we also plot a trade-off curve varying the relevance factor when considering only 3 user-dependent training scores for each class user/impostor. In this scenario with severe local training data scarcity the benefits of adapting the fusion functions from general knowledge are even bigger, from 3.43% in the local approach to 0.70% EER with the adapted scheme (80% relative performance improvement).

In Fig. 8.2 we finally show client and impostor scatter plots for a random data set of the bootstrap error estimation process ($M = 74$ users for computing the global parameters and $N = 6$ local training scores). Global, local and adapted fusion function boundaries (i.e., $f(\mathbf{x}) = 0$) are also depicted for the fusion scheme based on Bayesian adaptation.

8.3.2. Results for Quality-Based Fusion

Comparative performance results are given in Figs. 8.3 (a) and (b) for SVM and Bayesian quality-based fusion, respectively.

In Fig. 8.3 (a) we depict the verification performance for the two individual traits, repeated-

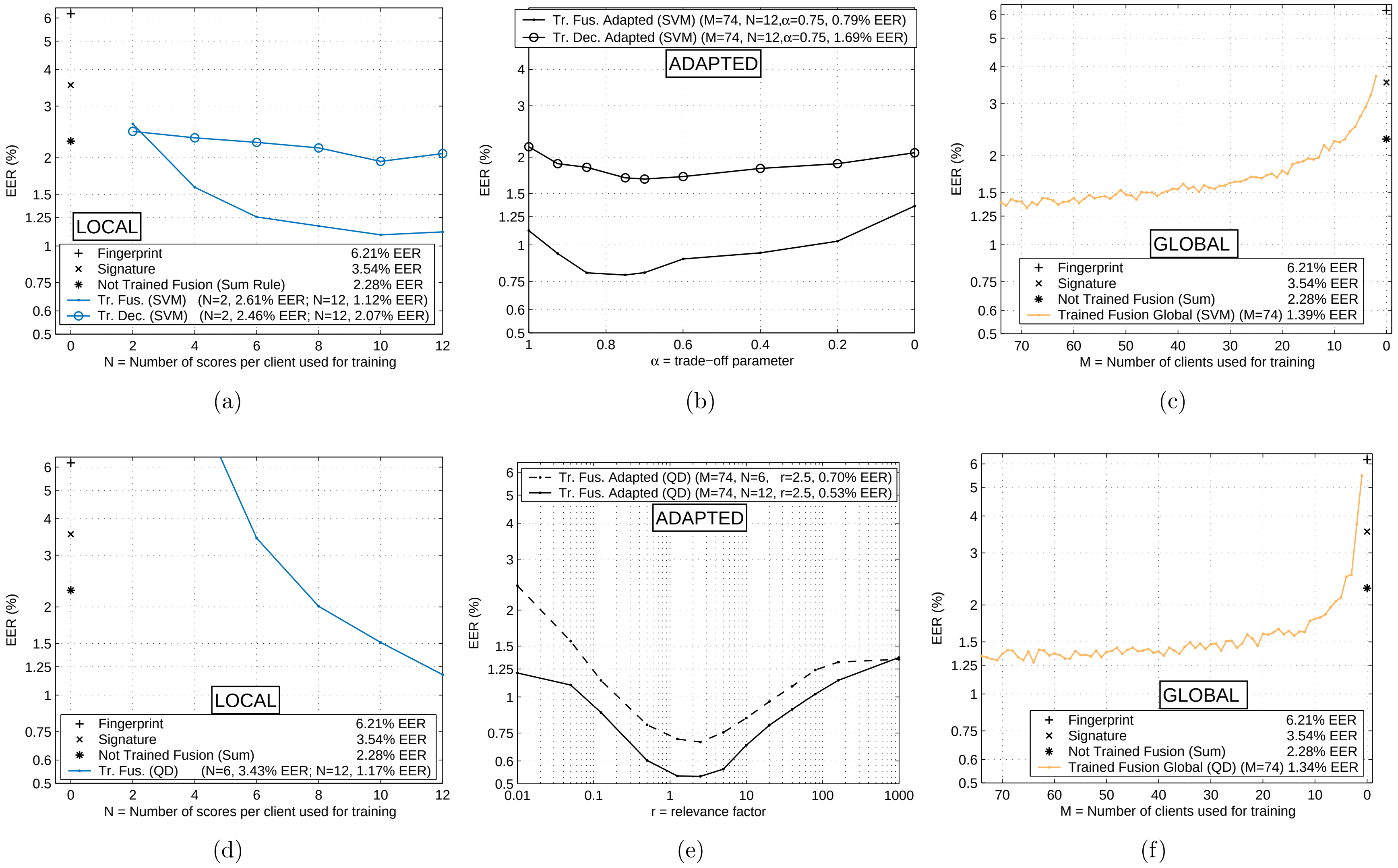


Figure 8.1: *Equal error rates of global (c,f), local (a,d), and adapted (b,e) user-dependent approaches for multimodal fusion based on SVM (a,b,c) and Bayesian adaptation (d,e,f).*

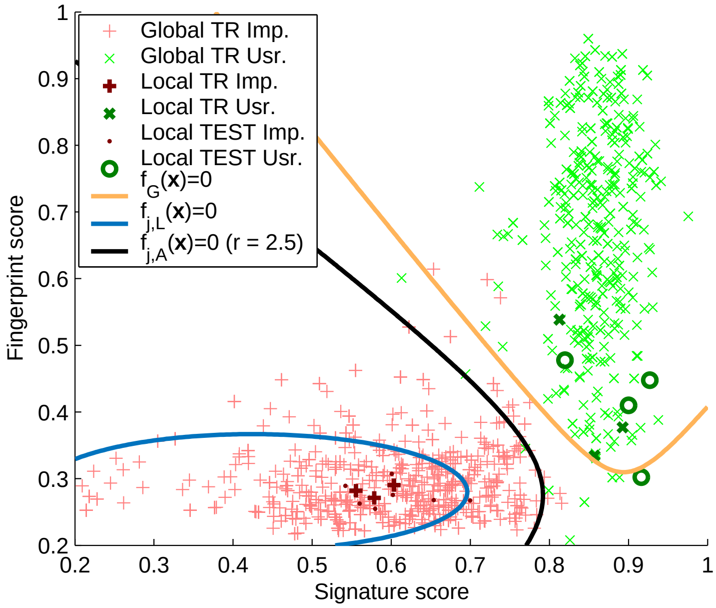
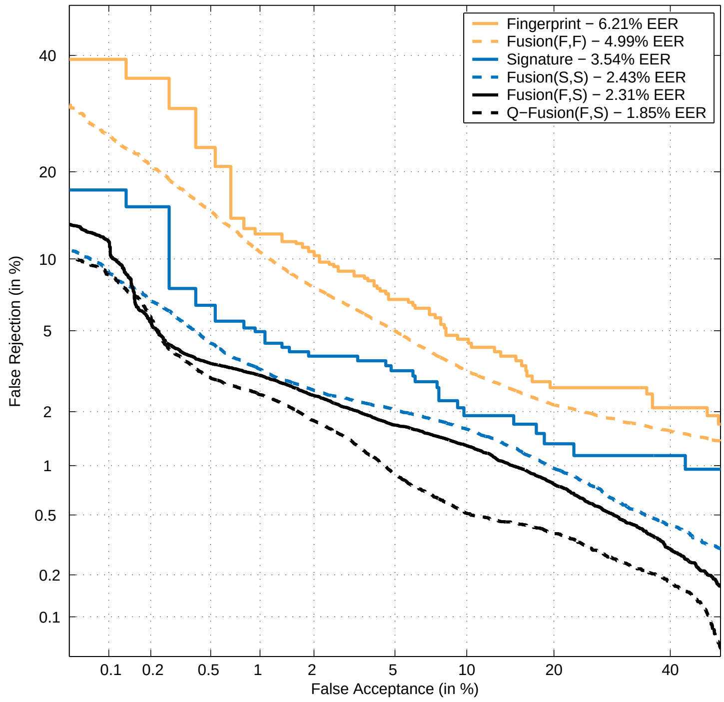


Figure 8.2: *Training/testing scatter plot and decision boundaries of global, local, and adapted approaches for multimodal fusion based on Bayesian adaptation (one iteration of the bootstrap-based error estimation process).*

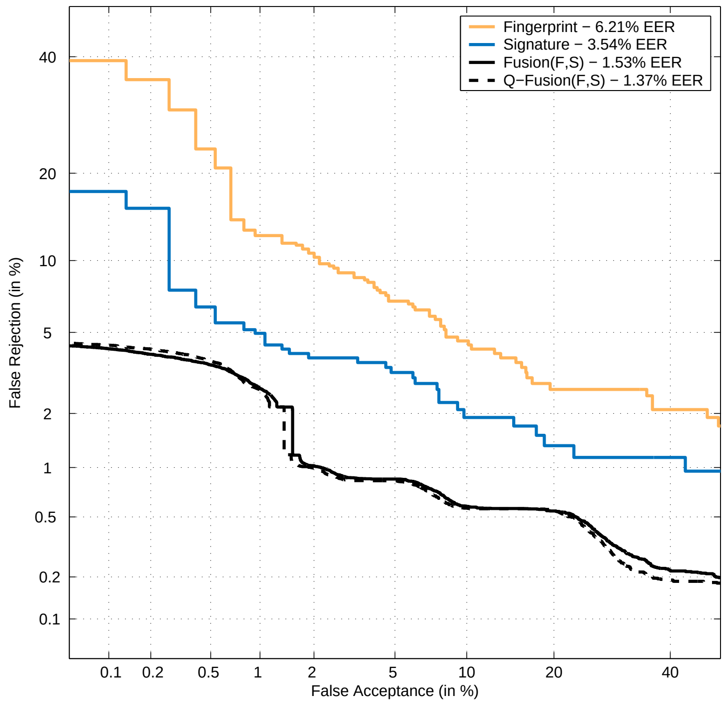
instance of each individual trait, and user-independent fusion of the two traits with and without considering the input fingerprint image quality. We first observe that the performance of individual systems is not as high as in other published works, 6.21% and 3.54% EER for the fingerprint and signature systems, respectively. This is due to the worst-case scenario considered in this chapter. These individual performance measures are improved by using two repeated-instances achieving 4.99% and 2.43% EER, respectively. When combining the two systems without including quality measures we achieve a similar performance measure to the one obtained using repeated-instances with the best individual system. We finally plot the verification performance curve when including the quality measures. The quality-based fusion scheme outperforms the raw fusion strategy without considering signal quality by a relative performance improvement of about 20%. As to the multi-instance experiments with respect to the individual systems, the relative performance improvement in the case of combining multiple signatures (30%) is greater than the relative performance improvement in the case of combining multiple fingerprints (20%).

The results with the Expert Conciliation scheme are shown in Fig. 8.3 (b). In this case we depict the verification performance for the two individual traits, and user-independent fusion of them with and without considering the input fingerprint image quality. When combining the two systems without including quality measures we achieve a relative performance improvement with respect to the best individual trait (57%) much higher than with the SVM approach (35%). Conversely, the performance improvement now when including quality measures is smaller (10% instead of 20% around the EER point with a larger difference at the low False Rejection region).

Finally, we also include some examples that may provide an intuitive idea about how the

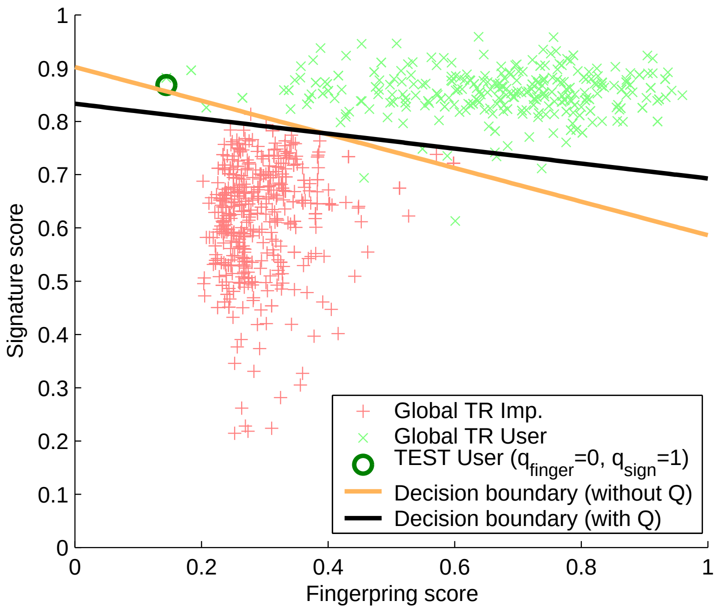


(a) Quality-based multimodal fusion based on SVM.

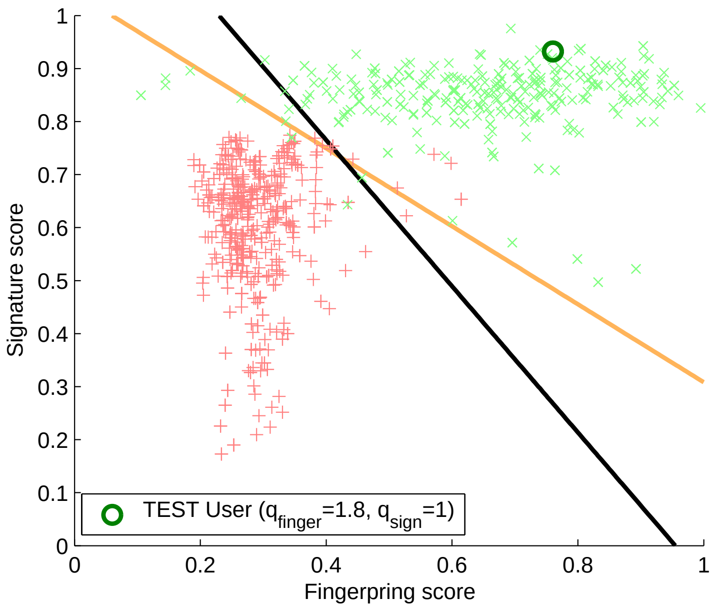


(b) Quality-based multimodal fusion based on Bayesian statistics.

Figure 8.3: *Verification performance results for quality-based multimodal fusion.*



(a) Test sample with low fingerprint image quality.



(b) Test sample with high fingerprint image quality.

Figure 8.4: *Training/testing scatter plot and decision boundaries for SVM-based fusion schemes with and without quality measures.*

fusion scheme is adapted depending on the image quality of the input fingerprints. In particular, two different data sets (a) and (b) of the bootstrap error estimation process are depicted in Fig. 8.4, together with the computed decision boundaries for the quality-based SVM approach.

8.4. Chapter Summary and Conclusions

A set of comparative experiments have been conducted using: 1) a bimodal biometric verification system based on fingerprint and on-line signature traits, 2) real bimodal biometric data from the MCYT database, and 3) a novel experimental protocol based on a worst-case scenario and bootstrap error estimates.

We have first studied the adaptation of the score fusion functions to individual users. For the scenario described in this work, and when enough training data is available for the trained approaches, the following set of experimental findings have been obtained: 1) trained fusion/decision outperforms the non-trained sum rule, 2) for the same amount of training data, local learning of the fusion functions outperforms localized trained decisions on summed scores, 3) local learning outperforms global learning, 4) adapted learning by using both global information from a pool of users and user-specific training data outperforms all other approaches. Most remarkably, we have reported some indications of the critical "enough training data" issue when comparing the trained to the not trained, and the global to the local approaches. Two schemes have been compared based on SVM and Bayesian adaptation, respectively. The SVM approach has been demonstrated to be quite robust to small training set sizes, working similarly to the sum rule in the worst case of 1 training score per class client/impostor, and improving it significantly for more than 3 training scores per class. The Bayesian adaptation approach has been demonstrated to be less robust to small training set size but to achieve best figures with large training set sizes.

We have also explored the adaptation of the score fusion functions to the input biometric quality using schemes based on SVM and Bayesian statistics. The SVM approach was improved significantly by including the quality measures. Conversely, the Bayesian approach did not improve significantly with the quality but achieved anyway similar performance improvements over the best individual system.

This chapter presents novel contributions in the application of adapted user-dependent score fusion and quality-based score fusion to multimodal biometrics.

Chapter 9

Conclusions and Future Work

THIS THESIS has considered the problem of adapting the score fusion functions in multimodal biometric authentication. After a summary of the state-of-the-art in fusion strategies for multimodal biometrics, a number of adapted fusion schemes have been proposed, based either on statistical assumptions or discriminative criteria using Support Vector Machines. These approaches adapt either to individual users through a reduced number of user-specific matching scores or to the input biometric quality. The proposed adapted fusion schemes have been applied to competitive multi-algorithm systems for three different biometrics, namely: signature, voice, and fingerprint; using standard biometric data and benchmarks. Finally, a comparative study of the proposed schemes has been given for the case of multimodal authentication based on signature and fingerprint on the real bimodal database MCYT.

9.1. Conclusions

Chapter 1 introduced the basics of biometric systems, biometric modalities, the motivation of the Thesis, and the research contributions originated from this Thesis. Chapter 2 further detailed the motivations with respect to the related works from the literature. The set of novel adapted score fusion methods proposed in this Thesis were presented in Chapter 3. The first part of the Dissertation concluded with an introduction to performance evaluation of biometric systems in Chapter 4, which also described the state-of-the-art in multimodal biometric databases and the biometric databases used in the Thesis.

The experimental part of the Dissertation started in Chapter 5 studying user-dependent score normalization and decision in multi-algorithm on-line signature verification. This chapter introduced two new systems based on local and global information, respectively. In the local system we have observed that the inclusion of azimuth and altitude signals worsens the verification performance, we have shown that the less modeling states the better the performance, and we have obtained experimental evidence on the importance of multiple training signatures from different acquisition sessions. This system was also used to demonstrate the benefits of incorporating user-dependent score normalization in an standard benchmark test. The second system presented a novel set of global features. In this case, we have shown comparative results for the discriminative capabilities of various combinations of features using a ranking based on individual discriminative capability. Finally, we have combined the local and global systems using simple score level fusion based on *max* and *sum* rules, demonstrating both the complementarity of the two approaches and the benefits of incorporating user-dependent decision thresholds.

Chapter 6 studied the application of adapted user-dependent fusion to multi-algorithm speaker verification using third party systems. We have compared user-independent, user-dependent, and adapted user-dependent versions of score level fusion. It has been shown that the proposed approach based on Bayesian adaptation outperforms both user-independent and user-dependent traditional fusion schemes.

Chapter 7 studied the effects of image quality on the performance of two common approaches for fingerprint verification. It was observed that the approach based on ridge information outperforms the minutiae-based approach in low image quality conditions. This was exploited by a simple adapted score-level fusion approach using quality measures estimated in the frequency domain. The proposed scheme led to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of fingerprint image quality and decision thresholds.

Chapter 8 finally compared the proposed adapted score fusion techniques for multimodal authentication. A set of comparative experiments have been conducted using: 1) a bimodal biometric verification system based on fingerprint and on-line signature traits, 2) real bimodal biometric data from the MCYT database, and 3) a novel experimental protocol based on a worst-case scenario and bootstrap error estimates. We first studied the adaptation of the score fusion functions to individual users, demonstrating the benefits of the proposed adapted approach. The SVM scheme resulted quite robust to small training set sizes and the Bayesian approach provided the best results for large training set sizes. We also demonstrated the benefits of quality-based fusion, either based on SVM or on Bayesian statistics.

Summarizing, the main results and contributions obtained from this Thesis are:

- The novel strategies for adapted score fusion: user-dependent and quality-based.
- The schemes implementing these strategies based on Bayesian statistics and Support Vector Machines.
- The individual systems developed: on-line signature verification using local and global information, and texture-based fingerprint verification.
- The multimodal biometric data acquired, which is now available for research purposes.
- The experimental evidence of the application of the proposed strategies to a number of problems: multi-algorithm signature, multi-algorithm speaker, multi-algorithm fingerprint, and multimodal biometric authentication using signature and fingerprint.

9.2. Future Work

A number of research lines arise from the work carried out in this Thesis. We consider of special interest the following ones:

- Completing the review and the theoretical framework proposed for score normalization by considering test-dependent score normalization techniques [Bimbot *et al.*, 2004] and their application to signature verification.

- Studying and implementing the idea of adapted learning for score normalization.

- The user-dependent score fusion approaches proposed in this work used a reduced number of matching scores for deriving the user-dependent fusion functions. Future work may involve the development of user-dependent fusion approaches directly based on the input biometrics and not on matching scores (e.g., based on the quality of the biometrics used for enrollment [Alonso-Fernandez *et al.*, 2006c]).

- Obtaining automatic quality measures for other biometric modalities. Current work is being done in this regard at the Biometrics Research Lab.–ATVS for voice [Garcia-Romero *et al.*, 2006] and fingerprint images [Alonso-Fernandez *et al.*, 2005b]. Quality measures for other biometrics such as written signature are still open. The topic of biometric quality is attracting much attention in the biometrics community nowadays [BQW, 2006].

- The Bayesian quality-based scheme evaluated in Chapter 8 did not improve the performance significantly when including the quality measures. This is contradictory with other works using the same strategy but combining other types of information [Bigun *et al.*, 2003]. Further work should be done in order to identify the key components for the success of this method, e.g., normalization of the quality measures.

- Integrated theoretical framework for user-dependent and quality-based fusion. Some current efforts in this regard include the work by Poh and Bengio [2005c].

- Studying the application of quality-based score fusion to already adapted used-dependent fusion schemes. The objective in this case will be to obtain a multimodal system capable of adapting to problematic users and noisy acquisitions.

- The adapted SVM-based techniques proposed and implemented in this work are based on trade-off parameters and discriminative considerations. Adaptive kernel methods have been proposed in the literature [Navia-Vazquez *et al.*, 2001], and find direct application to the adapted techniques discussed in this work.

- The experimental evaluations in this Thesis have been based on verification error rates. This evaluation procedure is focused on authentication applications, but it is not well suited to other scenarios where biometric evidences are not used to make a final decision, such as forensic reporting using biometrics [Gonzalez-Rodriguez *et al.*, 2005]. Recent approaches for application-independent evaluation of speaker recognition technologies can be found in the literature [Brummer and Preez, 2006; Ramos-Castro *et al.*, 2006b], and may be applied to other biometric modalities.

Apéndice A

Resumen Extendido de la Tesis

Esquemas Adaptados de Fusión para Autenticación Biométrica Multimodal

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.*, 2004b]. 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 bastante más reciente.

Hoy en día el reconocimiento biométrico se puede considerar como un campo de investigación asentado, con libros de referencia [Jain *et al.*, 1999a; Ratha and Bolle, 2004; Wayman *et al.*, 2005; Zhang, 2002], conferencias específicas en el tema [Jain and Ratha, 2004; Kittler and Nixon, 2003; Maltoni and Jain, 2004; Zhang and Jain, 2004], evaluaciones y pruebas comparativas [Grother *et al.*, 2003; Maio *et al.*, 2004; Phillips *et al.*, 2000b; Przybocki and Martin, 2004; Wilson *et al.*, 2004; Yeung *et al.*, 2004], proyectos internacionales [BioSec, 2004; Biosecure, 2004; COST-275, 2005], consorcios [BC, 2005; EBF, 2005], esfuerzos de estandarización [BioAPI, 2002; SC37, 2005], y un creciente interés tanto por parte de gobiernos [DoD, 2005] como del sector comercial [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; Kanade, 1973; Nagel and Rosenfeld, 1977], el reconocimiento biométrico sigue siendo un área muy activa de investigación, con numerosos problemas prácticos aún por solucionar [Jain *et al.*, 2004a]. 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.

Esta Tesis se centra en la combinación de varios rasgos biométricos para superar algunas de las limitaciones de rasgos individuales, en lo que se conoce como un *sistema biométrico multimodal* [Jain and Ross, 2004].

A.1. Introducción

El paradigma de la autenticación biométrica. El reconocimiento de personas se ha realizado históricamente asociando identidad y "algo que la persona posee" (por ejemplo, una llave o una tarjeta), o bien "algo que la persona sabe" (por ejemplo, una palabra-clave o un PIN). El reconocimiento biométrico añade a este paradigma una nueva dimensión, asociando persona e identidad personal mediante "algo que la persona es (o produce)". "Algo que la persona es" nos indica una característica fisiológica asociada de forma inherente a la persona, mientras que "algo que la persona produce" nos indica una aptitud o acto previamente entrenado que la persona realiza como patrón de conducta.

Sistemas biométricos. El reconocimiento biométrico es un término genérico para denominar a los dos modos de funcionamiento de los sistemas biométricos. De forma más precisa, se denomina *identificación* biométrica a la tarea que pretende asociar una muestra biométrica a uno de los N patrones o modelos disponibles del conjunto conocido de individuos registrados. Por este motivo, a esta tarea también se la conoce como comparación uno-contra-muchos o uno-contra-N. La salida de los sistemas que funcionan bajo este modo suele ser una lista ordenada de candidatos, estando ligado el criterio de ordenación al grado de similitud entre muestra de prueba y patrón registrado. Por el contrario, la *verificación* (o *autenticación*) biométrica es la tarea que pretende decidir si una determinada muestra de entrada coincide o no con un usuario específico (denominado usuario "solicitado", o "pretendido"). Esta tarea es conocida como problema uno-contra-uno, y la salida será una decisión binaria (aceptado/rechazado) basada en la comparación del grado de similitud (en forma de puntuación o *score* entre la muestra de entrada y el modelo de usuario pretendido) respecto a un determinado umbral de decisión. En esta Tesis nos centramos en el modo de verificación, cuyas dos etapas, registro (*enrollment*) y verificación (*verification*), se muestran esquemáticamente en la Figura 1.1.

El objetivo en la verificación biométrica es decidir entre dos clases, cliente o impostor. Dependiendo del rasgo biométrico que se trate, los impostores pueden conocer y utilizar información del rasgo imitado para facilitar el acceso, por ejemplo, la forma de la firma en el caso de verificación de firma escrita. Por ello se suelen considerar dos tipos de impostores: 1) *impostores casuales* (que producen *falsificaciones aleatorias*), cuando no se conoce información del rasgo imitado, y 2) *impostores reales* (que producen *falsificaciones entrenadas*), cuando se conoce y utiliza información del rasgo imitado.

Tipos de errores en verificación. El modo de verificación puede ser considerado como una tarea de detección, comportando un compromiso entre dos tipos de errores: 1) Falso Rechazo

(FR), que se produce cuando un usuario auténtico (lo que se conoce también por usuario genuino o cliente) es rechazado por el sistema, y 2) Falsa Aceptación (FA), que sucede cuando un impostor es aceptado por el sistema como si fuera un usuario auténtico. Estos dos tipos de errores tienen relación inversa entre sí, pudiéndose obtener diversos puntos de funcionamiento del sistema en función del umbral de decisión elegido. El punto de trabajo en cada caso dependerá de cada aplicación en particular. Por esta razón la caracterización de los sistemas biométricos se realiza mediante las curvas completas que relacionan ambos tipos de error (ver Figura 4.1). Por esta razón también, en el caso de caracterizar el rendimiento de un sistema de verificación con tasas numéricas, se suele optar bien por un par (FA,FR) o por el punto en donde coinciden ambas tasas, esto es, el punto de igual error (*Equal Error Rate* –EER).

Representación del funcionamiento en verificación. Tradicionalmente se han venido usando para representar el rendimiento de los sistemas biométricos en modo de verificación las curvas ROC (*Receiver-* o *Relative- Operating Characteristic*), en las que se representa la probabilidad de FA frente a la probabilidad de FR para los diferentes puntos de trabajo (esto es, umbrales de decisión) del sistema. En las curvas ROC, la zona de interés se concentra en la esquina inferior izquierda de la gráfica, que se corresponde con la zona en la que los dos tipos de error se minimizan conjuntamente. El problema de este tipo de representación ocurre cuando los sistemas producen bajas tasas de error, puesto que, en estos casos, las curvas que describen los sistemas tienden a concentrase, impidiéndose de esta forma una visualización comparativa clara de sistemas competitivos. Con el objeto de solventar este problema, más recientemente, se han propuesto las denominadas curvas DET (*Detection Error Tradeoff*) [Martin *et al.*, 1997], que representan también los dos tipos de error pero aplicando una transformación de ejes. Dicha escala produce un efecto de separación de las gráficas de sistema que en las ROC se concentraban en la esquina inferior izquierda, y además consigue que dichas curvas tiendan a ser líneas rectas para distribuciones de puntuaciones Gaussianas, haciendo así que las comparaciones entre sistemas competitivos sean directas y sencillas. En la Figura 4.2 se muestra una comparación entre curvas ROC y DET de dos sistemas hipotéticos de verificación A y B.

Modalidades biométricas. Hay una serie de modalidades fisiológicas que pueden ser consideradas como tecnológicamente "maduras", a saber, la huella dactilar, el iris, la cara, la geometría de los dedos y/o la mano, o la huella palmar. En relación con las modalidades conductuales, rasgos como la voz, la escritura y la firma manuscrita, o el modo de andar (marcha), son modalidades objeto de grandes esfuerzos de investigación. La Figura 1.2 muestra algunos ejemplos de rasgos biométricos. En teoría, cualquier característica humana puede ser considerada como un rasgo biométrico siempre que satisfaga las siguientes propiedades:

- *universal*, que indica que toda persona debe poseer dicho rasgo;
- *distintivo*, que se refiere a que dicho rasgo debe ser lo suficientemente diferente para diferentes personas;

- *permanente*, que indica que dicho rasgo debe poseer una representación que se mantenga a lo largo del tiempo;
- *mensurable*, que se refiere a la habilidad de medir dicho rasgo cuantitativamente.

Otras propiedades deseables de cara al uso de rasgos biométricos en sistemas de autenticación incluyen:

- *rendimiento*, que se refiere a la eficiencia, precisión, velocidad, robustez, y uso de recursos de las implementaciones prácticas basadas en dicho rasgo;
- *aceptabilidad*, que indica el grado en el que la gente está dispuesta a usar dicho rasgo y en qué términos;
- *seguridad*, que se refiere a la dificultad de burlar un sistema basado en dicho rasgo con métodos fraudulentos.

Si se analiza el estado del arte de los sistemas basados en diferentes rasgos biométricos, podremos observar que no existe ningún rasgo individual que maximice todas las propiedades indicadas. Algunos rasgos biométricos son altamente distintivos pero son difícilmente mensurables (p.ej., el iris, con dispositivos caros y difíciles de utilizar), mientras que otros se adquieren fácilmente pero no son tan distintivos (p.ej., la cara). En la Tabla 1.1 se incluye una comparación de los rasgos biométricos comunes de acuerdo a dichas características. En dicha tabla se resaltan las tres últimas filas, que se refieren a los tres rasgos biométricos que se estudian en esta Tesis: locutor (*speaker*), firma escrita (*signature*), y huella dactilar (*fingerprint*). Nótese que cuando se consideran los tres rasgos simultáneamente (o los dos últimos, firma y huella), prácticamente todas las propiedades se satisfacen ampliamente (*High* –H). Esta última combinación de firma escrita y huella dactilar se encuentra en algunas aplicaciones significativas como tarjetas electrónicas de identificación, por ejemplo en el DNIe [2006] Español.

Sistemas biométricos multimodales. En dichos sistemas se utilizan varios rasgos biométricos simultáneamente con objeto de compensar las limitaciones de rasgos individuales. Como resultado, las tasas de error en verificación suelen disminuir, el sistema resultante es más robusto frente a fallos de los sistemas individuales, y el número de casos donde el sistema no es capaz de dar una respuesta se reduce (p.ej., debido a la mala calidad de una muestra biométrica de uno de los rasgos individuales).

La mayoría de estrategias para la combinación de rasgos biométricos existentes en la literatura se basan en la fusión de las puntuaciones o medidas de similitud proporcionadas por los sistemas individuales [Ben-Yacoub *et al.*, 1999; Bigun *et al.*, 1997a; Brunelli and Falavigna, 1995; Chatzis *et al.*, 1999; Hong and Jain, 1998; Kittler *et al.*, 1998; Verlinde *et al.*, 2000]. Estos esquemas se basan normalmente en reglas sencillas de combinación, como la suma o el producto [Kittler *et al.*, 1998], o en clasificadores entrenados [Verlinde *et al.*, 2000], como Redes Neuronales (*Neural Networks* –NN) o Máquinas de Vectores Soporte (*Support Vector Machines* –SVM).

Trabajos recientes en multimodalidad biométrica se incluyen en la Tabla 2.2. Este enfoque de combinación de puntuaciones de sistemas individuales también se aplica en otros problemas de clasificación y es la fuente de mucha investigación en reconocimiento de patrones [Jain *et al.*, 2000a; Oza *et al.*, 2005]. En la Tabla 2.1 se incluye una lista de métodos generales de combinación de clasificadores que pueden encontrar aplicación en multimodalidad biométrica.

En el caso particular de autenticación biométrica, dos marcos teóricos para la combinación de sistemas individuales fueron descritos por Kittler *et al.* [1998] y Bigun *et al.* [1997a]. Ambos trabajos concluyen que la suma ponderada de las puntuaciones de los sistemas individuales es un buen método de fusión. En el primer trabajo [Kittler *et al.*, 1998], el método de combinación preferido es de la forma presentada en la Eq. 2.2, que para el caso de autenticación biométrica se reduce a la Eq. 2.9, esto es, fusión y basada en suma de puntuaciones previamente normalizadas para representar probabilidades *a posteriori* $x_j \approx P(\omega_1|B_j)$, a la que se aplica un umbral de decisión. En dichas ecuaciones x_j representa la similitud entre el patrón de entrada B_j y el modelo de la identidad solicitada en el sistema j, y ω_1 representa la clase usuario frente a ω_0 que representa la clase impostor. En el segundo trabajo [Bigun *et al.*, 1997a], la fusión resultante es del tipo indicado en la Eq. 2.11, donde se incluyen pesos w_j que son calculados en un proceso de entrenamiento de la regla de fusión.

Poco después de dichos esquemas de fusión de puntuaciones basados en marcos teóricos, la fusión biométrica multimodal se estudió como un problema de clasificación con dos clases (cliente e impostor) utilizando varios paradigmas de aprendizaje automático [Ben-Yacoub *et al.*, 1999; Gutschoven and Verlinde, 2000; Verlinde *et al.*, 2000], por ejemplo: Redes Neuronales, Árboles de Decisión, y Máquinas de Vectores Soporte (SVM). Después de varios experimentos comparativos, los métodos basados en SVM superaron a la mayoría de los otros esquemas.

Los esquemas de fusión biométrica multimodal desarrollados en la presente Tesis se basan bien en los marcos teóricos mencionados [Bigun *et al.*, 1997b; Kittler *et al.*, 1998], o en el uso de clasificadores SVM [Ben-Yacoub *et al.*, 1999; Gutschoven and Verlinde, 2000].

Sistemas multibiométricos. En todos los trabajos mencionados anteriormente el término *multimodal* se refería a la combinación de diferentes rasgos biométricos, por lo que *modo* se refiere a rasgo biométrico. Por otra parte, en los sistemas biométricos se pueden combinar no solo diferentes rasgos biométricos sino otro tipo de información con objeto de mejorar las propiedades que se deseen. De acuerdo a las recomendaciones en proceso de estandarización [SC37, 2005], este tipo de sistemas que hacen uso de múltiples fuentes de información biométrica y no necesariamente diferentes rasgos se denominan sistemas multibiométricos [Ross *et al.*, 2006]. Dichas fuentes de información biométrica provienen del nivel y del escenario de fusión considerado.

Niveles de Fusión. Los sistemas de autenticación biométrica se dividen normalmente en cuatro módulos (ver Fig. 1.1): 1) el sensor que adquiere las señales biométricas, 2) el módulo de extracción de características que procesa los datos biométricos con objeto de obtener una representación compacta a la vez que discriminante (vector de características), 3) el módulo de comparación que compara el vector de características de entrada con la muestra previamente registrada del usuario reclamado, y 4) el módulo de decisión. La fusión se puede llevar a cabo a la salida de cualquiera de los cuatro módulos:

- *Fusión a nivel de sensor.*
- *Fusión a nivel de características.*
- *Fusión a nivel de puntuaciones.*
- *Fusión a nivel de decisión.*

En la Fig. 1.4 se muestra una representación gráfica de los cuatro niveles de fusión mencionados. Según se indicó previamente, esta Tesis se centra en la fusión a nivel de puntuaciones.

Escenarios de Fusión. Un sistema multibiométrico se puede basar en uno o en varios de los siguientes escenarios de fusión:

- *Varios sensores.* Un solo rasgo biométrico se adquiere haciendo uso de varios sensores. Un ejemplo es el uso de varias cámaras para crear un modelo de cara 3D o para combinar las puntuaciones de las imágenes de cara individuales.
- *Varios algoritmos.* Una única entrada biométrica se procesa con diferentes módulos de extracción de características o módulos de comparación. Un ejemplo es el procesado de huellas dactilares haciendo uso tanto de patrones de minucias como de patrones de textura.
- *Varias unidades.* Se trata de un solo rasgo biométrico pero varias partes del cuerpo humano. Un ejemplo es el uso de varios dedos en verificación de huella dactilar.
- *Unidades repetidas.* El mismo rasgo y la misma parte del cuerpo se adquiere varias veces.
- *Varios rasgos.* Se combinan varios rasgos biométricos, es lo que se conoce normalmente por *sistemas biométricos multimodales* o *biometría multimodal*.

En la Fig. 1.5 se ilustran los diferentes escenarios multibiométricos con ejemplos relacionados con esta Tesis.

Motivación para la Tesis. Según hemos mencionado, la mayoría de estrategias para la fusión multimodal existentes en la literatura se basan en la fusión de las puntuaciones proporcionadas por los sistemas individuales. En la mayor parte de los casos se asume que dichos esquemas de combinación se mantienen fijos durante la etapa de verificación [Duin, 2002]. Esta Tesis se centra en el estudio de esquemas adaptados de fusión durante la etapa de verificación. Dicho estudio se basa principalmente en tres observaciones de la literatura y de nuestro trabajo práctico en el Lab. de Investigación Biométrica–ATVS.

La primera observación tiene su origen en Doddington *et al.* [1998], donde se analizó el comportamiento de diferentes locutores en la tarea de verificación propuesta por el Instituto Nacional de Estándares y Tecnología de los EEUU (*National Institute of Standards and Technology* –NIST) para la Evaluación de Reconocimiento de Locutor en 1998 (*Speaker Recognition Evaluation* –SRE). Este trabajo observó que ciertos locutores eran fácilmente aceptados por el sistema, mientras que otros eran rechazados sistemáticamente. Este hecho se ha venido compensando tradicionalmente con el uso de umbrales de decisión dependientes de usuario, sobre todo para el caso de rasgos biométricos conductuales como la voz o la firma escrita [Furui, 1981; Plamondon and Lorette, 1989]. Más recientemente se suelen aplicar técnicas de normalización de puntuaciones que intentan alinear las distribuciones de puntuaciones de diferentes usuarios a un mismo rango [Bimbot *et al.*, 2004]. Este mismo comportamiento se ha notado en la práctica del Lab. de Investigación Biométrica–ATVS tanto en ediciones sucesivas de NIST SRE en las que ha participado [Garcia-Romero *et al.*, 2006; Ramos-Castro *et al.*, 2006a], como en la Primera Competición Internacional de Verificación de Firma [Yeung *et al.*, 2004] en la que también participó (ver Fig. 2.3).

La segunda observación está muy relacionada con la primera. En el caso de fusión biométrica multimodal se ha propuesto recientemente el uso de esquemas de fusión dependientes de usuario [Jain and Ross, 2002; Toh *et al.*, 2004a]. El objetivo de estos esquemas es compensar que algunos rasgos pueden no ser adecuados para determinados usuarios pese a tratarse de rasgos altamente discriminantes para el resto de la población, por lo que se pueden esperar mejoras en verificación al restar importancia en la fusión de puntuaciones a dichos rasgos para dichos sujetos.

La tercera observación que ha motivado esta Tesis es el efecto de la calidad de las señales biométricas de entrada en el rendimiento de los sistemas biométricos de verificación [Junqua and Noord, 2001; Simon-Zorita *et al.*, 2003]. En concreto, se sabe que cuanto peor es la calidad de las señales de entrada, peor es el rendimiento en verificación. Esto por ejemplo queda patente en los resultados de la última Competición Internacional de Verificación de Huella (*Fingerprint Verification Competition* –FVC) [Cappelli *et al.*, 2006], en donde se utilizaron imágenes de huella de baja calidad intencionadamente. Como resultado las tasas de error de los mejores sistemas resultaron ser un orden de magnitud peor que en ediciones anteriores con imágenes de calidad más controlada. Este efecto de la calidad de imagen en verificación de huella está suscitando un creciente interés [BQW, 2006], y se estudiará en más detalle en la siguiente competición FVC [2006], en la que el Lab. de Investigación Biométrica–ATVS colabora como organizador. Este efecto de degradación de un sistema individual con la calidad de la imagen se puede compensar en el caso de un sistema multimodal teniendo en cuenta que no todos los rasgos se ven afectados por la calidad de igual manera [Jain and Ross, 2004].

La Tesis. En esta Tesis se plantean nuevas arquitecturas de fusión para autenticación biométrica adaptadas tanto a los diferentes usuarios registrados en el sistema, como a la calidad de las señales biométricas de entrada, resultando en *esquemas adaptados de fusión para autenticación biométrica multimodal*. Dichos esquemas encuentran aplicación asimismo en otros escenarios que combinan diferentes fuentes de información biométrica y no necesariamente diferentes rasgos, como son la fusión multi-sensor, multi-algoritmo, multi-unidad, y de unidades repetidas [Ross *et al.*, 2006].

El término *adaptado* en esta Tesis se refiere a esquemas entrenados utilizando información general, por ejemplo un conjunto de usuarios de referencia, y ajustados al considerar información específica bien del usuario que se trate o bien de la calidad de la muestra biométrica de entrada. En este sentido, los enfoques dependientes de usuario presentados en esta Tesis son originales. Los enfoques similares en la literatura o bien están entrenados únicamente con la información genérica [Bigun *et al.*, 1997b; Kittler *et al.*, 1998], o bien están entrenados con la información particular del usuario [Jain and Ross, 2002], pero no utilizan ambas simultáneamente. Respecto a la idea de la adaptación teniendo en cuenta la calidad biométrica, existen algunos trabajos previos pero que no han desarrollado la problemática de forma explícita (por ejemplo Chatzis *et al.* [1999] usó medidas de calidad en un esquema de fusión basado en lógica borrosa pero no para la adaptación de las funciones de fusión).

La Disertación. En primer lugar se introducen los sistemas biométricos, la motivación de la Tesis, una expresión breve de la Tesis, la organización de la Disertación, y las contribuciones de investigación relacionadas con la Tesis.

Después se resume el estado del arte en fusión para autenticación biométrica multimodal. Acto seguido se introducen los esquemas adaptados propuestos, tanto dependientes de usuario como basados en la calidad de las señales biométricas de entrada. Los esquemas dependientes de usuario se subdividen a su vez en: 1) normalización de puntuaciones dependiente de usuario y fusión sencilla, 2) fusión de puntuaciones dependiente de usuario, y 3) decisión dependiente de usuario. En la mayoría de los casos se presentan implementaciones basadas tanto en modelos estadísticos como en Máquinas de Vectores Soporte (SVM).

A continuación se resumen las prácticas comunes para la evaluación de rendimiento en sistemas de autenticación multimodales, y se presentan las bases de datos biométricos utilizadas en la parte experimental de la Disertación.

La parte experimental de la Disertación comienza con la aplicación de las técnicas de fusión adaptada a tres problemas de autenticación biométrica unimodal multi-algoritmo basados en firma escrita, voz, y huella dactilar, respectivamente.

En el caso de verificación multi-algoritmo de firma escrita se introducen dos nuevos sistemas basados en información local y global, respectivamente. El sistema local se utiliza para estudiar varios aspectos de importancia práctica como son la extracción de características, el modelado, y la normalización de puntuaciones dependiente de usuario. Finalmente se combinan los sistemas local y global usando reglas sencillas de fusión con lo que se demuestra tanto la complementariedad de ambos sistemas como los beneficios del uso de umbrales de decisión dependientes de usuario.

A continuación se estudia la aplicación de la fusión adaptada a usuario en el caso de verificación multi-algoritmo de locutor usando múltiples sistemas desarrollados por otros investigadores. Se comparan versiones de fusión independiente de usuario, dependiente de usuario, y adaptada al usuario, demostrándose la superioridad del esquema adaptado propuesto en esta Tesis.

En el caso de verificación multi-algoritmo de huella dactilar se estudia el efecto de la calidad de las imágenes de entrada en el rendimiento de dos enfoques comunes de reconocimiento, uno basado en patrones de minucias y otro basado en información de textura, habiendo sido este último desarrollado en el marco de esta Tesis. Se observa que este nuevo sistema es bastante robusto frente a imágenes a baja calidad, lo que se aprovecha mediante un esquema de fusión de ambos sistemas adaptado a la calidad de las imágenes de entrada.

Finalmente se realiza un estudio comparativo de los esquemas adaptados presentados en esta Tesis aplicados al problema de autenticación multimodal basada en firma escrita y huella dactilar. See estudian los esquemas de fusión adaptada tanto a usuario como a la calidad de las señales de entrada en sus dos versiones propuestas, basadas en modelos estadísticos o en SVM. Se demuestra la superioridad de los esquemas adaptados a usuario frente a los métodos tradicionales independientes de usuario, o dependientes de usuario sin uso de información general. En el estudio comparativo se demuestra la superioridad del esquema basado en SVM cuando el número de muestras de entrenamiento es reducido, y la del enfoque estadístico cuando el número de muestras es elevado. En el caso de fusión dependiente de calidad se demuestra la mejora proporcionada al tener en cuenta la calidad en ambos enfoques, especialmente en el esquema basado en SVM.

La dependencia entre capítulos se ilustra en la Fig. 1.6. Nótese que los capítulos experimentales, que están sombreados en la Fig. 1.6, contienen referencias a los métodos utilizados de capítulos anteriores. De esta manera, y asumiendo conocimientos generales en sistemas biométricos [Jain *et al.*, 2004b] y fusión multimodal [Ross *et al.*, 2006], los capítulos experimentales se pueden leer independientemente.

Contribuciones de la Tesis. Las contribuciones de la Tesis se pueden clasificar como sigue a continuación (nótese que algunas publicaciones se repiten en puntos diferentes de la lista):

- *Revisiones del estado del arte.* 1) Esquemas de fusión a nivel de puntuaciones para biometría multimodal [Fierrez-Aguilar *et al.*, 2003a,b] (premio al mejor póster). 2) Normalización de puntuaciones dependiente de usuario [Fierrez-Aguilar *et al.*, 2004c, 2005h]. 3) Fusión de puntuaciones dependiente de usuario [Fierrez-Aguilar *et al.*, 2005b].

- *Marcos teóricos.* Marco teórico y taxonomía relacionada para métodos de normalización de puntuaciones [Fierrez-Aguilar *et al.*, 2004c, 2005h].

- *Métodos originales.* 1) Nuevos métodos para normalización de puntuaciones dependientes de usuario [Fierrez-Aguilar *et al.*, 2005h]. 2) Nuevos métodos para fusión de puntuaciones dependientes de usuario basados en adaptación Bayesiana [Fierrez-Aguilar *et al.*, 2005a,c] y en Máquinas de Vectores Soporte [Fierrez-Aguilar *et al.*, 2004b, 2005b]. 3) Nuevos métodos para fusión de puntuaciones dependiente de calidad basada en media ponderada [Fierrez-Aguilar *et al.*, 2006] (premio a la mejor contribución de estudiante), teoría Bayesiana [Bigun *et al.*, 2003, 2005] (discursos clave relacionados en MMUA [2003] e ICIAP [2003]), y SVM [Fierrez-Aguilar *et al.*, 2004d, 2005i].

- *Nuevos sistemas biométricos.* 1) Dos nuevos sistemas de verificación de firma escrita dinámica [Fierrez-Aguilar *et al.*, 2005f; Ortega-Garcia *et al.*, 2003a] basados en información local y global, respectivamente. El sistema local surge como ampliación del trabajo previo realizado en el Lab. de Investigación Biométrica–ATVS [Ortega-Garcia *et al.*, 2002]. Dicho sistema fue presentado a la Primera Competición Internacional de Verificación de Firma, obteniendo muy buenos resultados [Yeung *et al.*, 2004]: 1º para falsificaciones aleatorias, y 2º para falsificaciones entrenadas. El sistema basado en información global fue desarrollado conjuntamente con Lopez-Peñalba [2006]. 2) Un nuevo sistema de verificación de huella basado en información de textura [Fierrez-Aguilar *et al.*, 2005e], desarrollado conjuntamente con Muñoz-Serrano [2005].

- *Nuevos datos biométricos.* Una nueva base de datos biométricos ha sido adquirida en el marco de trabajo de la Tesis incluyendo huella dactilar (12 impresiones de cada uno de los 10 dedos) y firma escrita (25 firmas reales y 25 falsificaciones por usuario) de 330 sujetos [Ortega-Garcia *et al.*, 2003b]. Dicha base de datos, denominada MCYT, se encuentra disponible públicamente en la actualidad, y está siendo usada por más de 30 grupos de investigación en todo el mundo.

- *Nuevos estudios experimentales.* 1) Normalización de puntuaciones en verificación de firma escrita [Fierrez-Aguilar *et al.*, 2004c, 2005h]. 2) Verificación de firma multi-algoritmo [Fierrez-Aguilar *et al.*, 2005f]. 3) Verificación de locutor multi-algoritmo [Fierrez-Aguilar *et al.*, 2005a]. 4) Estudio de los efectos de la calidad de imagen (estimación automática) en sistemas de verificación de huella basados en minucias y textura [Fierrez-Aguilar *et al.*, 2005e]. 5) Verificación de huella multi-algoritmo [Fierrez-Aguilar *et al.*, 2006] (discurso invitado en BQW [2006]). 6) Fusión multimodal de huella y firma escrita [Fierrez-Aguilar *et al.*, 2004b, 2005b,c, 2004d, 2005i].

Otras contribuciones relacionadas con la Tesis no incluidas en el presente volumen incluyen:

- *Revisiones del estado del arte.* Métodos de cálculo de la calidad de imagen en huella dactilar [Alonso-Fernandez *et al.*, 2005b].

- *Marcos teóricos.* Marco teórico para el uso de evidencias biométricas en informes forenses [Gonzalez-Rodriguez *et al.*, 2005].

- *Métodos originales.* Normalización de puntuaciones rápida dependiente tanto de la entrada como del usuario [Ramos-Castro *et al.*, 2006a].

- *Nuevos sistemas biométricos.* Sistema de verificación de firma escrita estática, esto es, basado en imágenes [Fierrez-Aguilar *et al.*, 2004a].

- *Nuevos datos biométricos.* 1) Una nueva base de datos de firma dinámica de 53 sujetos capturada con Tablet PC [Alonso-Fernandez *et al.*, 2005a]. 2) Una nueva base de datos incluyendo cara, iris, huella y voz de 250 sujetos en 4 sesiones capturada en el marco del proyecto integrado del 6º Programa Marco BioSec [Fierrez-Aguilar, 2005] (discurso invitado en ICB [2006]). Otros esfuerzos actuales en este ámbito que se pueden considerar contribución relacionada con esta Tesis incluyen la captura de nuevas bases de datos tanto en el proyecto coordinado del Plan Nacional de I+D+i Biosecur ID [2003], como en la red de excelencia del 6º Programa Marco Biosecure [2004], ambas actividades de adquisición lideradas por el Lab. de Investigación Biométrica–ATVS.

- *Nuevos estudios experimentales.* 1) Verificación de firma estática multi-algoritmo [Fierrez-Aguilar *et al.*, 2004a]. 2) Robustez de la verificación de firma dinámica en redes IP [Richiardi *et al.*, 2004]. 3) Verificación de firma dinámica multi-algoritmo combinando enfoques local y regional [Fierrez-Aguilar *et al.*, 2005d]. 4) Verificación de firma dinámica multi-algoritmo en el marco de la red de excelencia Biosecure [Garcia-Salicetti *et al.*, 2006]. 5) Normalización de puntuaciones dependiente de usuario en verificación de locutor [Garcia-Romero *et al.*, 2003b]. 6) Verificación de locutor multi-algoritmo usando voz conversacional en Español [Garcia-Romero *et al.*, 2003a]. 7) Verificación de locutor multi-algoritmo basada en calidad en el banco de pruebas de NIST [Garcia-Romero *et al.*, 2004, 2006]. 8) Normalización de puntuaciones en verificación de locutor dependiente de la entrada y del usuario [Ramos-Castro *et al.*, 2006a]. 9) Estudio de los efectos de la calidad de imagen (estimación manual) y de la variabilidad de la posición en verificación de huella basada en minucias [Simon-Zorita *et al.*, 2003]. 10) Verificación de huella multi-algoritmo usando todos los sistemas participantes en FVC 2004 [Fierrez-Aguilar *et al.*, 2005g]. 11) Verificación de huella multi-algoritmo en el marco de la red de excelencia Biosecure [Alonso-Fernandez *et al.*, 2006a]. 12) Ataques a sistemas de verificación de huella dactilar [Galbally-Herrero *et al.*, 2006]. 13) Verificación de imagen facial usando representación global [Cruz-Llanas *et al.*, 2003].

- *Nuevas aplicaciones biométricas.* 1) Uso de evidencias biométricas en informes forenses [Gonzalez-Rodriguez *et al.*, 2003, 2002, 2005; Ramos-Castro *et al.*, 2005]. 2) Uso de verificación de firma dinámica en Tablet PC [Alonso-Fernandez *et al.*, 2005a,c, 2006b]. 3) Uso de firmas dinámicas para generación de claves criptográficas [Freire-Santos *et al.*, 2006].

A.2. Esquemas Adaptados de Fusión

Los esquemas adaptados de fusión de puntuaciones que se proponen en esta Tesis se dividen en tres clases: 1) dependientes de usuario, 2) dependientes de la calidad, y 3) dependientes de usuario y de calidad.

Para cada clase de métodos, en primer lugar se incluye un diagrama de bloques del sistema y después se desarrollan algoritmos que implementan dichos métodos haciendo uso de técnicas conocidas de reconocimiento de patrones [Duda *et al.*, 2001]. En concreto se desarrollan implementaciones basadas en teoría Bayesiana de la decisión y en Máquinas de Vectores Soporte (SVM).

Para el desarrollo de los diferentes métodos se usa la siguiente nomenclatura. Dado un sistema multimodal de verificación biométrica compuesto por M sistemas individuales $j = 1, \ldots, M$, cada sistema calcula una puntuación de similitud s entre la señal de entrada biométrica B y el patrón registrado del usuario reclamado k. Las puntuaciones s se normalizan a un rango común x. Las puntuaciones normalizadas se agrupan en un vector de puntuaciones $\mathbf{x} = [x_1, \ldots, x_M]^T$. El diseño de un esquema de fusión consiste en la definición de una función $f : \mathbb{R}^M \rightarrow \mathbb{R}$ que maximice la separabilidad de las distribuciones fusionadas de puntuaciones de cliente $\{f(\mathbf{x})|\text{acceso cliente}\}$ e impostor $\{f(\mathbf{x})|\text{acceso impostor}\}$. Dicha función puede ser construida haciendo uso de puntuaciones de entrenamiento etiquetadas $(\mathbf{x}_i, z_i)$, donde $z_i = \{0 = \text{acceso impostor}, 1 = \text{acceso cliente}\}$. En la Fig. 3.1 se representa el diagrama de bloques general de un sistema de autenticación biométrica multimodal con fusión a nivel de puntuaciones junto con las notaciones mencionadas.

Fusión dependiente de usuario. Para los esquemas de fusión dependiente de usuario se usan dos conjuntos de entrenamiento formados por vectores de puntuaciones $\mathbf{x}$, incluyendo ambos tanto puntuaciones de usuario como de impostor. El primer conjunto está formado por puntuaciones correspondientes al usuario que se esté evaluando. El segundo conjunto está formado por vectores de puntuaciones correspondientes a un conjunto de referencia de usuarios. Haciendo uso simultáneo de ambos conjuntos se demuestra que la información genérica proporcionada por el conjunto de referencia puede ser de ayuda en esquemas dependientes de usuario. Para demostrar esto se desarrollan tres algoritmos para cada uno de los métodos de fusión presentados:

- *Global*. Para el entrenamiento de la función de fusión solo se utilizan las puntuaciones de los usuarios de referencia (tanto de clientes como de impostores). Esto es equivalente a los métodos tradicionales de fusión independiente de usuario.

- *Local*. Para el entrenamiento de la función de fusión solo se utilizan las puntuaciones del usuario que esté siendo evaluado (tanto de él mismo como de impostores suyos). Esto es equivalente a los métodos existes de fusión dependiente de usuario.

- *Adaptado*. Para el entrenamiento de la función de fusión se usan tanto las puntuaciones de referencia como las del usuario en cuestión. Este procedimiento es una aportación original de esta Tesis.

La autenticación multimodal a nivel de puntuaciones se puede hacer depender del usuario haciendo dependiente del usuario uno o varios de los siguientes módulos representados en la

Fig. 3.1: 1) normalización de puntuaciones, 2) fusión de puntuaciones, y 3) decisión.

El primer caso se desarrolla en detalle partiendo de un marco teórico basado en test de hipótesis. El diagrama de bloques en este caso para el sistema individual al que se le aplique la normalización de puntuaciones dependiente de usuario se ilustra en la Fig. 3.2. El resultado es una taxonomía que se divide en: 1) métodos centrados en el impostor, 2) métodos centrados en el usuario, y 3) métodos impostor-usuario. La taxonomía recoge algunos de los trabajos previos en normalización de puntuaciones y se completa con nuevos métodos.

El diagrama de bloques para el segundo caso se ilustra en la Fig. 3.3. Este caso de fusión de puntuaciones dependiente de usuario se desarrolla tanto siguiendo consideraciones probabilísticas como haciendo uso de SVM.

El diagrama de bloques para el tercer caso se ilustra en la Fig. 3.4. En este caso de decisión dependiente de usuario se pueden aplicar de manera directa los esquemas desarrollados para el segundo caso.

Fusión dependiente de calidad. El diagrama de bloques para este caso se ilustra en la Fig. 3.5. En este caso se desarrollan tres esquemas de fusión dependiente de la calidad basados en suma ponderada, enfoque probabilístico, y SVM, respectivamente.

Fusión dependiente de usuario y calidad. Por último se presenta el diagrama de bloques genérico para fusión de puntuaciones dependiente tanto de usuario como de calidad en la Fig. 3.6. En este caso no se desarrollan esquemas de fusión explícitamente al poder tratarse como combinación de los esquemas individuales presentados anteriormente.

A.3. Evaluación del Rendimiento en Sistemas Biométricos Multimodales

El creciente desarrollo de los sistemas biométricos ha hecho necesario la definición de bancos de prueba para la comparación objetiva de diferentes soluciones biométricas [Jain *et al.*, 2004b; Phillips *et al.*, 2000a]. La mayor parte de estos bancos de pruebas se crean a raíz de competiciones internacionales de autenticación personal basada en diferentes rasgos biométricos. En dichas competiciones se proporcionan tanto datos biométricos como protocolos experimentales detallados que, en general, luego se hacen públicos tras la competición. Algunos ejemplos de competiciones biométricas incluyen: Evaluación de Tecnología de Reconocimiento Facial del NIST (*Facial Recognition Technology Evaluation* –FERET), celebrada por primera vez en 1994 [Phillips *et al.*, 2000b]; Evaluación de Reconocimiento de Locutor del NIST, celebrada anualmente desde 1996 [Przybocki and Martin, 2004]; Competición Internacional de Verificación de Huella (*Fingerprint Verification Competition* –FVC), celebrada bianualmente desde 2000 [Cappelli *et al.*, 2006]; y la Competición Internacional de Verificación de Firma (*Signature Verification Competition* –SVC), organizada en 2004 [Yeung *et al.*, 2004]. También existen evaluaciones comparativas de soluciones comerciales organizadas por instituciones como NIST [Grother *et al.*, 2003; Wilson *et al.*, 2004], CESG [Mansfield *et al.*, 2001], o consultoras como International Biometric Group [2006]. En este entorno, y como resultado de la experiencia ganada en las diferentes competiciones y evaluaciones, existen recomendaciones para la evaluación de sistemas biométricos [Mansfield and Wayman, 2002], que se tienen en cuenta en esta Tesis.

La evaluación del rendimiento de un sistema biométrico se puede realizar a tres niveles diferentes [Phillips *et al.*, 2000a]: tecnológico, de escenario, y operacional.

El objetivo en una evaluación tecnológica es comparar varios algoritmos para identificar el más adecuado. La evaluación de los algoritmos se lleva a cabo haciendo uso de bases de datos adquiridas previamente siguiendo un protocolo de pruebas fijo. De esta manera las condiciones de la comparación se pueden repetir en un futuro. Aspectos importantes a tener en cuenta en relación a los datos utilizados son: 1) número de usuarios, 2) número de sesiones de adquisición, y 3) número de muestras biométricas por sesión. Prácticamente todos los bancos de prueba definidos en evaluaciones y competiciones biométricas son de este tipo [Maio *et al.*, 2004; Phillips *et al.*, 2000b; Przybocki and Martin, 2004; Yeung *et al.*, 2004].

El objetivo en evaluaciones de escenario es medir el rendimiento de un sistema en un escenario que modele un campo de aplicación. Debido a que cada escenario tendrá sus propios datos y sensores ligeramente diferentes, los resultados de las evaluaciones de escenario no son directamente comparables [Bone and Blackburn, 2002; Mansfield *et al.*, 2001]. Las evaluaciones operacionales son similares a las de escenario, excepto que en vez de modelar una clase de aplicaciones, se trata de evaluar un sistema específico en una aplicación específica [Bone and Crumbacker, 2001].

En esta Tesis los experimentos consisten en evaluaciones tecnológicas de diferentes esquemas unimodales, multi-algoritmo, y multimodales para autenticación biométrica.

Métodos de estimación de error. Con objeto de obtener tasas de Falso Rechazo y Falsa Aceptación en la tarea de verificación, se utiliza un conjunto de puntuaciones generadas con datos conocidos de usuarios e impostores pertenecientes a la base de datos biométricos que se trate. Existen diferentes métodos para hacer uso de dicha información de entrenamiento para la estimación de los errores [Jain *et al.*, 2000a; Theodoridis and Koutroumbas, 2003]. En esta Tesis, y dependiendo del experimento que se trate, se hace uso de alguna de las siguientes técnicas de estimación de errores en clasificación:

- *Resustitución* (*Resubstitution*): los datos de entrenamiento también se utilizan para evaluación.

- *Rotación*: se trata de una versión del método comúnmente conocido como *validación cruzada* (*cross-validation*). El modelo de cada usuario se entrena con k muestras consecutivas del conjunto de entrenamiento, y el resto se utilizan para evaluación; esto se repite para todas los distintos conjuntos de k muestras consecutivas. Cuando k es igual al número total de muestras de entrenamiento menos uno, se obtiene el método *leave-one-out*. Nótese que el mecanismo de rotación también se puede utilizar para seleccionar diferentes usuarios entre los disponibles.

- *Bootstrap*: se eligen un número de muestras aleatoriamente del conjunto de entrenamiento con reemplazo (esto es, la misma muestra se puede elegir varias veces). El resto de muestras constituye el conjunto de evaluación. El proceso se repite un número fijo de veces. Dicho proceso se puede aplicar tanto para la selección de usuarios como para la selección de muestras de un usuario determinado.

La primera estrategia resulta en estimaciones poco realistas mientras que las dos últimas necesitan mayor tiempo de proceso.

Bases de datos biométricos multimodales. Debido a la dificultad de la adquisición de bases de datos biométricos multimodales, y a los problemas de protección de datos personales relacionados [Wayman *et al.*, 2005], algunos autores han asumido la independencia entre rasgos biométricos y han usado diferentes bases de datos unimodales creando de esto modo individuos quiméricos [Poh and Bengio, 2005a]. Según se indica en las mejores prácticas de evaluación de sistemas biométricos [Mansfield and Wayman, 2002], es recomendable la evaluación sobre información multimodal real. Este es el enfoque seguido en esta Tesis.

Las bases de datos biométricos reales existentes en la actualidad son normalmente el resultado de proyectos de investigación coordinados. Ejemplos de dichos esfuerzos de investigación incluyen proyectos Europeos como M2VTS [Messer *et al.*, 1999] o BANCA [Bailly-Bailliere *et al.*, 2003]; y proyectos nacionales como el Francés BIOMET [Garcia-Salicetti *et al.*, 2003] o el Español MCYT [Ortega-Garcia *et al.*, 2003b]. Otros esfuerzos coordinados para la adquisición de bases de datos biométricos incluyen el proyecto FP6 BioSec [2004], y las actividades de adquisición de datos en la red de excelencia FP6 Biosecure [2004].

Las bases de datos biométricos multimodales se pueden clasificar en dos grupos: 1) bases de datos de señales biométricas, y 2) bases de datos de puntuaciones [Poh and Bengio, 2006]. En la primera clase las bases de datos incluyen muestras biométricas, tales como imágenes de huellas dactilares o señales de voz. Dichas señales se pueden utilizar con diferentes procedimientos experimentales tanto para el desarrollo de sistemas biométricos basados en rasgos individuales como para la fusión a cualquier nivel (sensor, característica, puntuación o decisión). La segunda clase de bases de datos multimodales están orientadas a la investigación en fusión de puntuaciones o decisiones.

Base de datos biométricos bimodal MCYT. Dicha base de datos surge del proyecto del Plan Nacional de I+D+i del Ministerio de Ciencia y Tecnología (MCYT) número TIC00-1669-C04. Como uno de los resultados de dicho proyecto se adquirió una base de datos bimodal incluyendo imágenes de huella dactilar y firma escrita de 330 individuos [Ortega-Garcia *et al.*, 2003b]. Parte del trabajo de esta Tesis se desarrolló en el marco de dicho proyecto.

La adquisición fue llevada a cabo por un consorcio de cuatro universidades: Universidad Politécnica de Madrid (UPM, coordinador), Universidad de Valladolid (UVA), Universidad del País Vasco (EHU), y Escuela Universitaria Politécnica de Mataró (EUPMT). El número de individuos capturados en cada universidad fue 145, 75, 75, y 35, respectivamente.

- *Corpus de huella.* Se usaron dos sensores electrónicos: 1) El sensor capacitivo modelo 100SC de Precise Biometrics, y 2) el sensor óptico modelo UareU de Digital Persona; ambos con resolución de 500 puntos por pulgada. Se capturaron todos los dedos de los individuos, con 12 muestras por dedo. Como resultado cada individuo proporciona al corpus de huella un total de 240 imágenes (2 sensores $\times$ 12 impresiones $\times$ 10 dedos).

  La Fig. 4.3 muestra tres impresiones de un mismo dedo adquiridas con el sensor óptico (arriba) y con el sensor capacitivo (abajo) para los tres niveles de control considerados (de izquierda a derecha). Dichos niveles se controlan durante la adquisición por un supervisor humano de manera que el núcleo de la huella esté en el interior del recuadro indicado (3 muestras para control bajo –izquierda, 3 muestras para control medio –centro, y 6 muestras más para control alto –derecha). Más ejemplos de imágenes de huella en MCYT se pueden encontrar en la Fig. 4.4.

  Para un subconjunto de 9000 imágenes (todas las imágenes de 75 individuos con el sensor óptico) se dispone asimismo de medidas de calidad subjetivas marcadas manualmente por un experto [Simon-Zorita *et al.*, 2003]. Básicamente, a cada imagen se le asignó un entero entre 0 (calidad más baja) y 9 (calidad más alta) de acuerdo a factores como: área de la huella, presión, humedad, suciedad, cortes, etc. Considerando dichas medidas de calidad se obtiene que en torno al 5% de las imágenes son de muy mala calidad, 20% son de baja calidad, 55% son de calidad media, y el 20% son de calidad muy alta. La Fig. 4.5 muestra cuatro ejemplos de imágenes con su calidad subjetiva.

- *Corpus de firma.* La información dinámica de las firmas escritas fue capturada con una tableta digitalizadora Wacom Intuos A6 haciendo uso de un bolígrafo especial con tinta sobre papel común. Este procedimiento permitió capturar por un lado la información dinámica, en forma de trayectorias, presión y ángulos de inclinación del bolígrafo respecto al tiempo (ver Fig. 4.6); y por el otro lado la información estática impresa en las hojas, que posteriormente fue digitalizada a 600 puntos por pulgada para un conjunto total de 2250 firmas de 75 individuos. Cada firma fue escrita en una rejilla de tamaño 3.75 cm $\times$ 1.75 cm (ancho $\times$ alto).

  La resolución de la tableta es de 2540 líneas por pulgada permitiendo la detección de la información dinámica hasta una altura de 10 mm del puntero del bolígrafo sobre la tableta, por lo que el movimiento durante el levantamiento del mismo también se registró. La frecuencia de muestreo es de 100 Hz.

  Cada usuario contribuye con 25 firmas auténticas y con 25 falsificaciones a otros usuarios en grupos alternados de 5 firmas. Para ello el usuario n realizó 5 veces su propia firma, imitó 5 veces al $n-1$, volvió a realizar su firma auténtica 5 veces, luego imitó 5 veces al $n-2$, y así sucesivamente hasta completar las 25 firmas auténticas y las 25 falsificaciones. Las falsificaciones se basan en la forma de las firmas a imitar, para ello los imitadores dispusieron de las hojas impresas de otros usuarios, se les permitió practicar durante unos minutos, y se les indicó que debían falsificar sin cortes ni interrupciones en la firma de forma que la dinámica fuese natural.

Varios ejemplos de firmas del corpus MCYT junto con sus funciones temporales asociadas se pueden encontrar en la Fig. 4.7. Ejemplos adicionales de firmas se incluyen en la Fig. 4.8.

El corpus de firma MCYT ha sido distribuido y está siendo utilizado en más de 30 grupos de investigación en todo el mundo [Hongo *et al.*, 2005; Igarza *et al.*, 2005; Muramatsu *et al.*, 2006; Nanni and Lumini, 2006; Richiardi and Drygajlo, 2003].

Base de datos de firmas escritas SVC2004. Una de las contribuciones importantes de esta Tesis es el desarrollo de nuevos sistemas de verificación de firma dinámica. Teniendo esto en cuenta, y que la competición SVC 2004 [Yeung *et al.*, 2004] es el único banco de pruebas público y reconocido aparte del corpus MCYT usado en esta Tesis, se utilizará asimismo el corpus SVC para evaluar los sistemas presentados.

En concreto se usará el corpus de desarrollo proporcionado por los organizadores de SVC 2004 para la tarea extendida (que incluye información de trayectoria, presión, y ángulos frente al tiempo). La información fue capturada directamente sobre la tableta Wacom Intous con un bolígrafo especial sin tinta. Este corpus consta de 40 individuos, con 20 firmas auténticas por individuo capturadas en dos sesiones y 20 falsificaciones entrenadas (estando a disposición de los impostores la información dinámica de las firmas a imitar). Las firmas son en Inglés o Chino y fueron inventadas para la competición, a diferencia de MCYT, en donde se trata de las firmas usadas normalmente en la vida diaria. Algunos ejemplos del corpus SVC se muestran en la Fig. 4.10.

A.4. Verificación Multi-Algoritmo de Firma

Este primer capítulo experimental se basa en las publicaciones: Fierrez-Aguilar *et al.* [2005f, 2004c, 2005h]; Ortega-Garcia *et al.* [2003a].

El objetivo es estudiar la dependencia de usuario tanto en la normalización de puntuaciones de un sistema individual como en la etapa de decisión de un esquema multi-algoritmo. El caso de funciones de fusión dependientes de usuario se estudia en el segundo estudio experimental centrado en verificación multi-algoritmo de locutor.

En este primer estudio se introducen dos nuevos sistemas de verificación de firma escrita dinámica. El primer sistema se basa en información local y Modelos Ocultos de Markov (*Hidden Markov Models* –HMM) [Rabiner, 1989; Yang *et al.*, 1995], tratándose de una ampliación del trabajo previo en el Lab. de Investigación Biométrica–ATVS [Ortega-Garcia *et al.*, 2002]. El segundo sistema es una aportación original basada en parámetros globales [Lee *et al.*, 1996] y clasificación con ventanas de Parzen (*Parzen Windows Classification* –PWC) [Duda *et al.*, 2001]. Dicho sistema se desarrolló conjuntamente con Lopez-Peñalba [2006].

Sistema basado en información local. En el caso del sistema basado en información local se exploran varios aspectos de la extracción de características, estrategia de entrenamiento y modelado. En primer lugar se demuestra experimentalmente que la inclusión de las señales de inclinación del bolígrafo empeoran el rendimiento del sistema, y se proporcionan tasas de rendimiento al considerar diferentes funciones temporales (ver Fig. 5.3). A continuación se demuestra que el uso de varias firmas de entrenamiento con variabilidad natural multisesión mejora significativamente los resultados respecto al uso de múltiples firmas de entrenamiento adquiridas en la misma sesión (ver Fig. 5.4), observándose que 5 firmas de entrenamiento son suficientes para obtener buenos resultados. Por último se demuestra que la mejor configuración para el HMM es un número reducido de estados (2 en nuestro caso, ver Tabla 5.1) y un número elevado de mezclas Gaussianas por estado (32 en nuestro caso, ver Fig. 5.6).

Dicho sistema también se utiliza para estudiar las técnicas de normalización de puntuaciones dependientes de usuario propuestas en la Tesis. Los mejores resultados se obtienen para una técnica centrada en el usuario y basada en un mecanismo de rotación en el entrenamiento (ver Fig. 5.8). Este hecho se corrobora con los resultados obtenidos por dicho sistema en la Primera Competición Internacional de Verificación de Firma [Yeung *et al.*, 2004].

Sistema basado en información global. Este sistema presenta un nuevo conjunto de parámetros globales que representan varias características temporales, dinámicas y geométricas de las firmas. Dicho conjunto se basa fundamentalmente en los trabajos previos de Nelson and Kishon [1991]; Nelson *et al.* [1994]; y Lee *et al.* [1996], que en conjunto suman aproximadamente 70 parámetros. Además de adaptar algunos de ellos, dicho conjunto se amplía hasta llegar a los 100 parámetros. El reconocimiento se basa en modelado estadístico no paramétrico usando ventanas de Parzen Gaussianas. En la parte experimental (ver Fig. 5.10) se incluyen resultados comparativos para un número creciente de parámetros usando un orden de prelación basado en la capacidad discriminante individual.

Fusión de información local y global. Por último se combinan los sistemas local y global a nivel de puntuaciones usando reglas sencillas de fusión y umbrales de decisión dependientes de usuario (ver Tablas 5.3 y 5.4). El sistema global se comporta mejor que el local para pocas firmas de entrenamiento, lo que se puede justificar con la complejidad del modelado HMM, que necesita de un número elevado de firmas de entrenamiento para empezar a ser competitivo. El sistema global también resulta ser robusto al desalineamiento de las distribuciones de puntuaciones de diferentes usuarios. Se demuestra asimismo que los dos sistemas proporcionan información complementaria que se puede aprovechar con reglas simples de fusión. La combinación de ambos sistemas resulta en mejoras relativas respecto al mejor sistema individual del 44% y del 75% para falsificaciones entrenadas y aleatorias, respectivamente.

A.5. Verificación Multi-Algoritmo de Locutor

Este segundo capítulo experimental se basa en la publicación: Fierrez-Aguilar *et al.* [2005a].

El objetivo es estudiar la adaptación de las funciones de fusión a los diferentes usuarios en el caso de verificación multi-algoritmo de locutor. En concreto se aplica el esquema Bayesiano de adaptación propuesto en la Sección 3.1.2.1.

Sistemas utilizados. Los sistemas individuales utilizados en este estudio fueron desarrollados en el Laboratorio Lincoln del MIT [Reynolds *et al.*, 2005]. Dicho grupo de trabajo ha sido tradicionalmente uno de los que mejores resultados han obtenido en las campañas de evaluación promovidas por el NIST [Przybocki and Martin, 2004]. Pese a que el Lab. de Investigación Biométrica–ATVS también ha participado en dichas evaluaciones con buenos resultados desde el 2002 [Garcia-Romero *et al.*, 2006; Ramos-Castro *et al.*, 2006a], en este segundo capítulo experimental se ha optado por usar sistemas de referencia de terceros. Este enfoque demuestra la aplicabilidad directa de las técnicas desarrolladas en esta Tesis a otros sistemas diferentes de los desarrollados en el marco de trabajo de la Tesis.

Resultados. En total se usan 7 sistemas basados en diferentes niveles de información del locutor en la señal de voz, a saber: acústico, fonético, prosódico y léxico. Los datos para los experimentos están extraídos de la Evaluación de Reconocimiento de Locutor organizada por el NIST en 2004 (NIST SRE 2004) [Reynolds *et al.*, 2005]. En este banco de pruebas público y conocido, y haciendo uso de los sistemas de terceros mencionados, se demuestra experimentalmente que la fusión adaptada a usuario propuesta en esta Tesis (ver Tabla 6.6), supera tanto a los métodos tradicionales de fusión independiente (ver Tabla 6.2) como dependiente de usuario (ver Tabla 6.4).

A.6. Verificación Multi-Algoritmo de Huella

Este tercer capítulo experimental se basa en las publicaciones: Fierrez-Aguilar *et al.* [2006, 2005e].

El objetivo es estudiar la fusión adaptada a calidad en verificación multi-algoritmo de huella. Para ello se utiliza el esquema más sencillo de fusión dependiente de calidad basado en suma ponderada. Una comparación más general de los métodos de fusión adaptada tanto a usuario como a calidad se incluye en el último estudio experimental de fusión multimodal de huella y firma.

Sistemas utilizados. El estudio comienza con un análisis del efecto de la calidad de imagen en el rendimiento de dos sistemas de verificación de huella, el primero basado en patrones de minucias y el segundo basado en información de textura. Para ello en primer lugar se resumen trabajos relacionados con la estimación de la calidad de imágenes de huella. A continuación se describe el método de estimación automática de la calidad usado en este estudio, basado en un análisis de energía en anillos concéntricos del espectro (ver Fig. 7.1), según el método propuesto por Chen *et al.* [2005]. A continuación se describen los sistemas de verificación de huella usados. El primero de ellos está directamente extraído de Simón-Zorita [2004] (ver Fig. 7.2). El segundo de ellos se desarrolló en el marco de esta Tesis conjuntamente con Muñoz-Serrano [2005] (ver Fig. 7.3).

Resultados. De los resultados experimentales se extrae que el sistema basado en textura es más robusto que el de minucias para calidad de imagen baja (ver Fig. 7.6). Este hecho se explota con la fusión dependiente de calidad (ver Fig. 7.4), dando más importancia al sistema basado en textura cuando se estima una imagen de huella de entrada de baja calidad. Dicho esquema dependiente de calidad proporciona resultados en verificación aproximadamente un 20% mejores que la fusión no adaptada de ambos sistemas (ver Fig. 7.7).

A.7. Verificación Multimodal de Firma y Huella

Este último capítulo experimental se basa en las publicaciones: Bigun *et al.* [2003]; Fierrez-Aguilar *et al.* [2004b, 2005b,c, 2004d, 2005i].

El objetivo es comparar los diferentes esquemas propuestos de fusión adaptada a usuario y dependiente de calidad en el problema de autenticación multimodal basada en huella dactilar y firma escrita. Para ello se hace uso de: 1) el sistema de verificación de huella basado en minucias y el sistema de verificación de firma basado en información local, 2) el subconjunto de la base de datos bimodal MCYT para el que se dispone de las medidas subjetivas de calidad en huella (75 individuos, todos los dedos), y 3) un enfoque experimental novedoso basado en el análisis del caso peor y estimación de error basada en bootstrap.

Fusión multimodal dependiente de usuario. En primer lugar se estudia la fusión adaptada a usuario (ver Fig. 8.1). Para el escenario considerado, y cuando se dispone de suficientes datos para el entrenamiento, se obtienen los siguientes resultados: 1) las reglas entrenadas de fusión y decisión superan el enfoque no entrenado de la regla de la suma, 2) para una cantidad fija de datos de entrenamiento dependiente de usuario, es mejor utilizar los mismos para entrenar reglas de fusión locales que reglas de decisión locales, 3) el aprendizaje local es mejor que el aprendizaje global, 4) el aprendizaje adaptado supera tanto al enfoque local como al global. Adicionalmente, se proporcionan indicaciones de la cantidad de datos de entrenamiento necesarios para que se cumplan las condiciones de la comparación. En concreto, se han comparado los enfoques basados en SVM y de adaptación Bayesiana. El enfoque SVM proporciona mejores resultados para conjuntos de entrenamiento reducidos. Por el contrario, el enfoque Bayesiano proporciona mejores resultados con conjuntos de entrenamiento de mayor tamaño.

Fusión multimodal dependiente de calidad. Por último se estudia la fusión dependiente de calidad al combinar huella y firma escrita. En este caso se usan las medidas de calidad subjetiva de huella disponibles en MCYT y calidad uniforme para todas las firmas. En este caso se estudian también los esquemas de fusión Bayesiano y basado en SVM (ver Fig. 8.3). El enfoque de fusión basado en SVM mejora significativamente al considerar las medidas de calidad. Por el contrario, el enfoque Bayesiano no produce una mejora significativa al incluir las mismas. A su favor, el enfoque Bayesiano consigue mejoras en el rendimiento respecto al mejor sistema individual similares al SVM al considerar la calidad.

A.8. Líneas de Trabajo Futuro

Se proponen las siguientes líneas de trabajo futuro relacionadas con el trabajo desarrollado en esta Tesis:

- Completar la revisión del estado del arte y el marco teórico para normalización de puntuaciones considerando también los métodos de normalización de puntuaciones dependientes de *test* [Bimbot *et al.*, 2004], como *t-norm* [Auckenthaler *et al.*, 2000].

- Aplicar dichos métodos de normalización de puntuaciones dependientes de test al problema de verificación de firma.

- Estudiar e implementar la idea de aprendizaje adaptado al usuario para el problema de normalización de puntuaciones.

- Los esquemas dependientes de usuario desarrollados en esta Tesis se basan en un conjunto reducido de puntuaciones del usuario. Una línea de trabajo futuro es el desarrollo de técnicas de este tipo pero directamente basadas en las señales biométricas y no en puntuaciones de similitud, de esta manera evitando la comparación de patrones. Un ejemplo de esta idea es hacer uso de medidas de calidad de las señales [Alonso-Fernandez *et al.*, 2006c].

- Desarrollar medidas de calidad automáticas para otros rasgos biométricos. En el Lab. de Investigación Biométrica–ATVS ya se está trabajando en este sentido para señales de voz [Garcia-Romero *et al.*, 2006] e imágenes de huella [Alonso-Fernandez *et al.*, 2005b]. Otros esfuerzos recientes en este sentido se recogen en BQW [2006].

- El esquema de fusión Bayesiano dependiente de calidad no mejoró significativamente los resultados al incluir las medidas de calidad. Esto es contradictorio con otros resultados publicados con otras bases de datos [Bigun *et al.*, 2003]. Una línea de trabajo futuro es el estudio de los componentes individuales de dicho esquema, por ejemplo el proceso de normalización de las medidas de calidad, para justificar dichas discrepancias.

- Marco teórico integrado de fusión dependiente de usuario y dependiente de la calidad [Poh and Bengio, 2005c].

- Estudiar la adaptación a calidad de esquemas previamente adaptados a usuario. De esta manera se pueden obtener esquemas capaces de compensar tanto usuarios problemáticos en algunos rasgos como muestras de entrada ruidosas.
- Los esquemas basados en SVM desarrollados en esta Tesis se basan en parámetros de ponderación lineal entre funciones componentes. En la literatura se pueden encontrar esquemas de aprendizaje adaptativo para SVM [Navia-Vazquez *et al.*, 2001], que pueden ser de aplicación directa para el problema de fusión biométrica adaptada desarrollado en esta Tesis.
- Las evaluaciones experimentales presentadas en esta Tesis se basan en tasas de error en verificación. Este procedimiento de evaluación no es apropiado para ciertas aplicaciones biométricas en donde no se genera una decisión final de aceptación o rechazo, como es el caso de los informes forenses basados en evidencias biométricas [Gonzalez-Rodriguez *et al.*, 2005]. En la literatura se pueden encontrar métodos de evaluación de tecnología de reconocimiento de locutor independientes de la aplicación [Brummer and Preez, 2006; Ramos-Castro *et al.*, 2006b], que pueden ser de interés para la evaluación de otros rasgos biométricos.

References

NOTE: The numbers at the end of each reference indicate the pages where each reference is cited.

A. G. Adami. *Modeling Prosodic Differences for Speaker and Language Recognition.* PhD thesis, Oregon Graduate Institute, 2004. 93

A. G. Adami, R. Mihaescu, D. A. Reynolds, and J. Godfrey. Modeling prosodic dynamics for speaker recognition. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP*, pages 788–791, 2003. 93

F. Alkoot and J. Kittler. Improving the performance of the product fusion strategy. In *Proc. of Intl. Conf. on Pattern Recognition, ICPR*, volume 2, pages 164–167, September 2000. 21

F. Alonso-Fernandez, J. Fierrez-Aguilar, F. del Valle, and J. Ortega-Garcia. On-line signature verification using Tablet PC. In *Proc. IEEE Intl. Symposium on Image and Signal Processing and Analysis, ISPA*, pages 245–250, Zagreb, Croatia, September 2005a. 14, 15, 135

F. Alonso-Fernandez, J. Fierrez-Aguilar, H. Fronthaler, K. Kollreider, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Combining multiple matchers for fingerprint verification: A case study in Biosecure Network of Excellence. *Annals of Telecommunications, Special Issue on Multimodal Biometrics*, 61, 2006a. (to appear). 15, 135

F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. A review of schemes for fingerprint image quality computation. In *Proc. of 3rd Workshop on Biometrics on the Internet, COST-275*, pages 3–6, Hatfield, UK, October 2005b. Official Publisher of the European Communities. 14, 123, 134, 145

F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. Sensor interoperability and fusion in signature verification: A case study using Tablet PC. In S. Li *et al.*, editors, *Proc. of International Workshop on Biometric Recognition Systems, IWBRS*, pages 180–187. Springer LNCS-3781, 2005c. 15, 83, 135

F. Alonso-Fernandez, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. A web-based secure access system using signature verification over Tablet PC. *IEEE Aerospace and Electronic Systems Magazine*, 21, 2006b. (to appear). 15, 83, 135

F. Alonso-Fernandez, R. N. J. Veldhuis, A. M. Bazen, J. Fierrez-Aguilar, and J. Ortega-Garcia. On the relation between biometric quality and user-dependent score distributions in fingerprint verification. In *Proc. of Intl. Workshop on Multimodal User Authentication, MMUA*, 2006c. 123, 145

E. Alpaydin and M. Jordan. Local linear perceptrons for classification. *IEEE Trans. Neural Networks*, 7(3): 788–792, 1996. 19

B. S. Atal. Automatic recognition of speakers from their voices. *Proceedings of the IEEE*, 64:460–475, 1976. 1, 53, 125

R. Auckenthaler, M. Carey, and H. Lloyd-Tomas. Score normalization for text-independent speaker verification systems. *Digital Signal Processing*, 10:42–54, 2000. 29, 40, 41, 93, 145

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. 58, 59, 139

R. Baron and R. Plamondon. Acceleration measurement with an instrumented pen for signature verification and handwriting analysis. *IEEE Trans. on Instrum. Measurement*, 38(6):1132–1138, 1989. 65

BC, 2005. Biometrics Consortium. (http://www.biometrics.org/). 1, 125

S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Fusion of face and speech data for person identity verification. *IEEE Trans. on Neural Networks*, 10(5):1065–1074, 1999. 6, 26, 27, 32, 128, 129

S. Bengio, C. Marcel, S. Marcel, and J. Mariethoz. Confidence measures for multimodal identity verification. *Information Fusion*, 3(4):267–276, 2002. 26, 33

E. S. Bigun. Risk analysis of catastrophes using experts' judgments: An empirical study on risk analysis of major civil aircraft accidents in Europe. *European J. Operational Research*, 87:599–612, 1995. 6, 48, 49

E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by Bayesian statistics. In J. Bigun, G. Chollet, and G. Borgefors, editors, *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 291–300. Springer LNCS-1206, 1997a. 6, 25, 32, 33, 36, 47, 48, 52, 56, 128, 129

J. Bigun. *Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision*. Springer, 2006. 12, 102

J. Bigun, G. Chollet, and G. Borgefors, editors. *Proceedings of the First International Conference on Audio- and Video-Based Person Authentication, AVBPA*, volume 1206 of *Lecture Notes in Computer Science*. Springer, 1997b. 129, 132

J. Bigun, B. Duc, S. Fischer, A. Makarov, and F. Smeraldi. Multi modal person authentication. In H. Wechsler *et al.*, editors, *NATO-ASI Advanced Study on Face Recogniton*, volume F-163, pages 26–50. Springer, 1997c. 49

J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in mobile communications. In *Proc. of Intl. Conf. on Image Analysis and Processing, ICIAP*, pages 2–13. IEEE CS Press, 2003. 14, 37, 111, 114, 123, 134, 144, 145

J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. *Advanced Studies in Biometrics*, volume 3161 of *Lecture Notes in Computer Science*, chapter Combining Biometric Evidence for Person Authentication, pages 1–18. Springer, 2005. 14, 134

J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(8):775–790, 1991. 102

F. Bimbot, J. F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds. A tutorial on text-independent speaker verification. *Journal on Applied Signal Processing*, 2004:4:430–451, 2004. 10, 29, 40, 123, 131, 145

BioAPI, 2002. ANSI INCITS 358-2002 - Information Technology - BioAPI Specification (Version 1.1). 1, 125

BioSec, 2004. Biometrics and Security, FP6 IP IST-2002-001766. (http://www.biosec.org/). 1, 58, 60, 125, 139

Biosecur ID, 2003. Seguridad Multimodal basada en Autenticación Biométrica mediante Fusión de Expertos Unimodales, MCYT TIC2003-08382-C05. 60, 135

Biosecure, 2004. Biometrics for Secure Authentication, FP6 NoE IST-2002-507634. (http://www.biosecure.info/). 1, 58, 60, 125, 135, 139

R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. *Guide to Biometrics*. Springer Verlag, 2004a. 29

R. M. Bolle, N. K. Ratha, and S. Pankanti. Error analysis of pattern recognition systems–the subsets bootstrap. *Computer Vision and Image Understanding*, 93:1–33, 2004b. 57

M. Bone and D. Blackburn. Face recognition at a chokepoint. Technical report, DoD Counterdrug Technology Development Program Office, November 2002. 54, 138

M. Bone and C. Crumbacker. Facial recognition: Assessing its viability in the corrections environment. *Corrections Today Magazine*, pages 62–64, July 2001. 54, 138

K. W. Bowyer. When is multi-modal better than uni-modal in biometrics? In *Workshop on Multimodal User Authentication, MMUA*, December 2003. 114

BQW. NIST Biometric Quality Workshop. Gaithersburg, MD, USA, March 2006. (http://www.itl.nist.gov/iad/894.03/quality/workshop/). 14, 35, 102, 123, 131, 134, 145

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. 18

BRL. Biometrics Research Lab.-ATVS, 2006. (http://atvs.ii.uam.es/). 66

N. Brummer and J. Preez. Application-independent evaluation of speaker detection. *Computer Speech and Language*, 20:230–275, 2006. 124, 146

R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 17(10):955–966, 1995. 6, 25, 27, 32, 128

D. Burton. Text-dependent speaker verification using vector quantization source coding. *IEEE Trans. Acoust. Speech, Signal Process.*, 35(2):133–143, 1987. 42

W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo. Support Vector Machines for speaker and language recognition. *Computer Speech and Language*, 20(2-3):210–229, 2006. 92, 93

R. Cappelli, D. Maio, and D. Maltoni. A multi-classifier approach to fingerprint classification. *Pattern Analysis and Applications*, 5(2):136–144, 2002a. 28

R. Cappelli, D. Maio, and D. Maltoni. Syntetic fingerprint-database generation. In *Proc. Intl. Conf. on Pattern Recognition, ICPR*, pages 744–747. IEEE Press, 2002b. 34

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. 10, 34, 54, 61, 102, 109, 131, 137

K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2D+3D face biometrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27:619–624, 2005. 23

V. Chatzis, A. G. Bors, and I. Pitas. Multimodal decision-level fusion for person authentication. *IEEE Trans. on System, Man, and Cybernetics, part A*, 29(6):674–680, 1999. 6, 11, 26, 27, 32, 33, 128, 132

Y. Chen, S. Dass, and A. Jain. Fingerprint quality indices for predicting authentication performance. In T. Kanade, N. Ratha, and A. Jain, editors, *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 160–170. Springer LNCS-3546, 2005. 101, 102, 103, 144

C. Chibelushi, S. Gandon, J. Mason, F. Deravi, and D. Johnston. Design issues for a digital integrated audio-visual database. In *IEE Colloquium on Integrated Audio-Visual Processing for Recognition, Synthesis and Communication*, pages 7/1–7/7, November 1999. 58

T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland. Multimodal person recognition using unconstrained audio and video. In *Proc. of Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 176–181, 1999. 27

COST-275, 2005. Biometrics-Based Recognition of People Over the Internet. (http://www.fub.it/cost275/). 1, 125

S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. A comparative evaluation of global representation based schemes for face verification. In *Proc. of the IEEE Intl. Conf. on Image Processing, ICIP*, volume 3, pages 905–908, 2003. 15, 135

F. del Valle-Hernández. Sistema remoto de verificación de firma manuscrita para Tablet PC. Master's thesis, Escuela Politécnica Superior, Universidad Carlos III de Madrid, 2006. 83

J. R. Deller, J. H. L. Hansen, and J. G. Proakis. *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press, 1999. 12

DNIe. El nuevo DNI electrónico llega a los ciudadanos. EL PAIS.es, March 16th 2006. (http://www.elpais.es/). 4, 128

DoD, 2005. Biometrics Management Office, Department of Defense, USA. (http://www.biometrics.dod.mil/). 1, 125

G. Doddington. Speaker recognition based on idiolectal differences between speakers. In *Proc. of ISCA European Conf. on Speech Communication and Technology, EUROSPEECH*, pages 2521–2524, 2001. 93

G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheeps, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 Speaker Recognition Evaluation. In *Proc. of Intl. Conf. on Speech and Language Processing, ICSLP*, 1998. 9, 29, 92, 131

J. Dolfing. *Handwriting Recognition and Verification, a Hidden Markov Approach*. PhD thesis, Technical University of Eindhoven, 1998. 63

J. G. A. Dolfing, E. H. L. Aarts, and J. J. G. M. van Oosterhout. On-line signature verification with Hidden Markov Models. In *Proc. of the Intl. Conf. on Pattern Recognition, ICPR*, pages 1309–1312. IEEE CS Press, 1998. 72, 73

B. Duc, E. S. Bigun, J. Bigun, G. Maitre, and S. Fischer. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, 18:835–843, 1997. 27

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001. 6, 12, 17, 21, 37, 39, 43, 56, 57, 136, 142

R. P. W. Duin. The combining classifier: to train or not to train? In *Proc. of the IAPR Intl. Conf. on Pattern Recognition, ICPR*, pages 765–770. IEEE CS Press, 2002. 25, 130

B. Dumas, J. Hennebert, A. Humm, R. Ingold, D. Petrovska, C. Pugin, and D. Rotz. MyIdea - Sensors specifications and acquisition protocol. Computer Science Department Research Report DIUF-RR 2005.01, University de Fribourg in Switzerland, January 2005. 60

EBF, 2005. European Biometrics Forum. (http://www.eubiometricforum.com/). 1, 125

M. C. Fairhurst. Signature verification revisited: Promoting practical exploitation of biometric technology. *IEE Electronics and Communication Engineering Journal*, 9(6):273–280, 1997. 72

J. Fierrez-Aguilar. Biometric databases: Modalities, privacy, and size. In *BioSec 3rd Workshop*, Helsinki, Finland, June 2005. (http://www.biosec.org). 15, 60, 135

J. Fierrez-Aguilar, N. Alonso-Hermira, G. Moreno-Marquez, and J. Ortega-Garcia. An off-line signature verification system based on fusion of local and global information. In D. Maltoni and A. K. Jain, editors, *Proc. of Intl. Workshop on Biometric Authentication, BIOAW*, pages 295–306. Springer LNCS-3087, 2004a. 14, 15, 66, 79, 86, 135

J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. Jain. Incorporating image quality in multi-algorithm fingerprint verification. In D. Zhang and A. K. Jain, editors, *Proc. of IAPR Intl. Conf. on Biometrics, ICB*, pages 213–220. Springer LNCS-3832, 2006. 14, 102, 134, 143

J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodrigez. Speaker verification using adapted user-dependent multilevel fusion. In N. C. Oza, R. Polikar, J. Kittler, and F. Roli, editors, *Proc. of Intl. Workshop on Multiple Classifier Systems, MCS*, pages 356–365. Springer LNCS-3541, 2005a. 12, 14, 91, 133, 134, 143

J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Exploiting general knowledge in user-dependent fusion strategies for multimodal biometric verification. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 5, pages 617–620, 2004b. 12, 14, 32, 37, 111, 113, 133, 134, 144

J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Adapted user-dependent multimodal biometric authentication exploiting general information. *Pattern Recognition Letters*, 26(16):2628–2639, 2005b. 12, 14, 28, 32, 37, 111, 133, 134, 144

J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Bayesian adaptation for user-dependent multimodal biometric authentication. *Pattern Recognition*, 38(8):1317–1319, 2005c. 12, 14, 37, 111, 133, 134, 144

J. Fierrez-Aguilar, S. Krawczyk, J. Ortega-Garcia, and A. K. Jain. Fusion of local and regional approaches for on-line signature verification. In *Proc. of Intl. Workshop on Biometric Recognition Systems, IWBRS*, pages 188–196. Springer LNCS-3781, 2005d. 15, 135

J. Fierrez-Aguilar, L. M. Muñoz-Serrano, F. Alonso-Fernandez, and J. Ortega-Garcia. On the effects of image quality on minutiae- and ridge-based automatic fingerprint recognition. In *Proc. of IEEE Intl. Carnahan Conf. on Security Technology, CARNAHAN*, pages 79–82, 2005e. 14, 102, 134, 143

J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, and D. Maltoni. An on-line signature verification system based on fusion of local and global information. In *Proc. of IAPR Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 523–532. Springer LNCS-3546, 2005f. 14, 72, 134, 141

J. Fierrez-Aguilar, L. Nanni, J. Ortega-Garcia, R. Cappelli, and D. Maltoni. Combining multiple matchers for fingerprint verification: A case study in FVC2004. In F. Roli and S. Vitulano, editors, *Proc. of Intl. Conf. on Image Analysis and Processing, ICIAP*, pages 1035–1042. Springer LNCS-3617, 2005g. 15, 35, 109, 135

J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez-Rodriguez. A comparative evaluation of fusion strategies for multimodal biometric verification. In J. Kittler and M. S. Nixon, editors, *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 830–837. Springer LNCS-2688, 2003a. 12, 17, 32, 133

J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Fusion strategies in multimodal biometric verification. In *Proc. of IEEE Intl. Conf. on Multimedia and Expo, ICME*, volume 3, pages 5–8, 2003b. 12, 17, 133

J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. In D. Zhang and A. K. Jain, editors, *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 498–504. Springer LNCS-3072, 2004c. 12, 14, 72, 133, 134, 141

J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE Trans. on Systems, Man and Cybernetics, part C*, 35(3): 418–425, 2005h. 12, 14, 72, 133, 134, 141

J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Kernel-based multimodal biometric verification using quality signals. In A. K. Jain and N. K. Ratha, editors, *Proc. of Intl. Conf. on Biometric Technologies for Human Identification, BTHI*, pages 544–554. Proc. SPIE-5404, 2004d. 14, 37, 111, 114, 134, 144

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, 2005i. 14, 37, 111, 112, 134, 144

M. Freire-Santos, J. Fierrez-Aguilar, and J. Ortega-Garcia. Cryptographic key generation using handwritten signature. In *Proc. of Intl. Conf. on Biometric Technologies for Human Identification, BTHI*, volume 6202, pages 225–231. Proc. of SPIE, 2006. 15, 135

G. Fumera and F. Roli. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):942–956, 2005. 22

S. Furui. Cepstral analysis technique for automatic speaker verification. *IEEE Trans. Acoust. Speech, Signal Processing*, 29(2):254–272, 1981. 10, 29, 41, 131

FVC. Fingerprint Verification Competition, 2004. (http://bias.csr.unibo.it/fvc2004). 34

FVC. Fingerprint Verification Competition, 2006. (http://bias.csr.unibo.it/fvc2006). 35, 131

J. Galbally-Herrero, J. Fierrez-Aguilar, J. D. Rodriguez-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprint attacks. In *Proc. IEEE Intl. Carnahan Conf. on Security Technology, CARNAHAN*, 2006. 15, 135

D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. Support Vector Machine fusion of idiolectal and acoustic speaker information in Spanish conversational speech. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 2, pages 229–232, 2003a. 15, 135

D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. On the use of quality measures for text-independent speaker recognition. In J. Ortega-Garcia *et al.*, editors, *ISCA Workshop on Speaker and Language Recognition, ODYSSEY*, pages 105–110, 2004. 15, 135

D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. Using quality measures for multilevel speaker recognition. *Computer Speech and Language*, 20(2-3):192–209, 2006. 15, 91, 123, 131, 135, 143, 145

D. Garcia-Romero, J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, and J. Ortega-Garcia. U-norm likelihood normalization in PIN-based speaker verification systems. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 208–213. Springer LNCS-2688, 2003b. 15, 42, 135

S. Garcia-Salicetti *et al.* BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 845–853. Springer LNCS-2688, 2003. 58, 59, 63, 139

S. Garcia-Salicetti, J. Fierrez-Aguilar, F. Alonso-Fernandez, C. Vielhauer, R. Guest, L. Allano, T. D. Trung, T. Scheidat, B. L. Van, J. Dittmann, B. Dorizzi, J. Ortega-Garcia, J. Gonzalez-Rodriguez, M. B. Castiglione, and M. Fairhurst. Biosecure reference systems for on-line signature verification: A study of complementarity. *Annals of Telecommunications, Special Issue on Multimodal Biometrics*, 61, 2006. (to appear). 15, 63, 135

S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58, 1992. 22

R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2002. 12

J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, and J. Ortega-Garcia. Forensic identification reporting using automatic speaker recognition systems. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 2, pages 93–96, 2003. 15, 135

J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. J. Lucena-Molina. Biometric identification in forensic cases according to the Bayesian approach. In *Proc. of Intl. Workshop on Biometric Authentication, BIOAW*, pages 177–185. Springer LNCS-2359, 2002. 15, 135

J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, and J. Ortega-Garcia. Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *Forensic Science International*, 155(2-3): 126–140, 2005. 14, 15, 50, 102, 123, 134, 135, 146

P. J. Grother, R. J. Micheals, and P. J. Phillips. Face recognition vendor test 2002 performance metrics. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 937–945. Springer LNCS-2688, 2003. 1, 54, 125, 137

L. Guillick and S. Cox. Some statistical issues in the comparison of speech recognition algorithms. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 1, pages 532–535, 1989. 57

B. Gutschoven and P. Verlinde. Multi-modal identity verification using Support Vector Machines (SVM). In *Proc. of Intl. Conf. on Information Fusion, FUSION*, pages 3–8. IEEE Press, 2000. 6, 26, 129

I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik. What size test set gives good error rate estimates? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):52–64, 1998. 57

S. Hangai, S. Yamanaka, and T. Hanamoto. On-line signature verification based on altitude and direction of pen movement. In *Proc. of the IEEE Intl. Conf. on Multimedia and Expo, ICME*, volume 1, pages 489–492, 2000. 73

T. Ho, J. Hull, and S. Srihari. Decision combination in multiple classifier systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(1):66–75, 1994. 24

L. Hong and A. K. Jain. Integrating faces and fingerprints for personal identification. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 20(12):1295–1307, 1998. 6, 27, 32, 128

L. Hong, Y. Wan, and A. K. Jain. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(8):777–789, 1998. 102

Y. Hongo, D. Muramatsu, and T. Matsumoto. Adaboost-based on-line signature verifier. In A. K. Jain and N. K. Ratha, editors, *Proc. of Intl. Conf. on Biometric Technologies for Human Identification, BTHI*, volume 5779, pages 373–380. Proc. of SPIE, 2005. 66, 141

ICB, 2006. IAPR Intl. Conf. on Biometrics, Hong Kong, Jan. 2006. (http://www4.comp.polyu.edu.hk/ icba/). 15, 135

ICIAP, 2003. Intl. Conf. on Image Analysis and Processing, Mantova, Italy, Sept. 2003. (http://iciap2003.unipv.it/). 14, 134

J. Igarza, I. Hernaez, I. Goirizelaia, K. Espinosa, and J. Escolar. Off-line signature recognition based on dynamic methods. In A. K. Jain and N. K. Ratha, editors, *Proc. of SPIE, Biometric Technologies for Human Identification II*, volume 5779, pages 336–343, 2005. 66, 141

International Biometric Group. Biometrics market and industry report 2006-2010, 2006. (http://www.biometricgroup.com/). 1, 4, 54, 125, 138

R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton. Adaptive mixtures of local experts. *Neural Comput.*, 3:79–87, 1991. 18, 19

A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics - Personal Identification in a Networked Society*. Kluwer, 1999a. 1, 125

A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 22(1):4–37, 2000a. 6, 17, 18, 22, 37, 56, 100, 129, 138

A. K. Jain, F. Griess, and S. Connell. On-line signature verification. *Pattern Recognition*, 35(12):2963–2972, 2002. 66, 72

A. K. Jain, L. Hong, and Y. Kulkarni. A multimodal biometric system using fingerprint, face and speech. In *Proc. of Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA-99*, pages 182–187, 1999b. 27

A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. An identity authentication system using fingerprints. *Proceedings of the IEEE*, 85(9):1365–1388, 1997. 104

A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38:2270–2285, 2005. 22, 27, 28, 36

A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman. Biometrics: A grand challenge. In *Proc. of the IAPR Intl. Conf. on Pattern Recognition, ICPR*, volume 2, pages 935–942. IEEE CS Press, 2004a. 3, 125

A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Trans. Image Processing*, 9(5):846–859, 2000b. 106

A. K. Jain and N. K. Ratha, editors. *Biometric Technology for Human Identification, First International Workshop*, volume 5404 of *Proc. of SPIE*. 2004. 1, 125

A. K. Jain and A. Ross. Learning user-specific parameters in a multibiometric system. In *Proc. of IEEE Intl. Conf. on Image Processing, ICIP*, volume 1, pages 57–60, 2002. 10, 11, 24, 31, 32, 56, 131, 132

A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM*, 47(1):34–40, 2004. 7, 10, 126, 131

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, 2004b. 1, 2, 6, 53, 54, 125, 133, 137

M. I. Jordan and R. A. Jacobs. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6: 181–214, 1994. 18

J. C. Junqua and G. V. Noord, editors. *Robustness in Language and Speech Technology*. Kluwer Academic Publishers, 2001. 10, 33, 131

T. Kanade. *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973. 1, 53, 125

R. S. Kashi, J. Hu, W. L. Nelson, and W. Turin. On-line handwritten signature verification using Hidden Markov Model features. In *Proc. of the 4th Intl. Conf. on Document Analysis and Recognition, ICDAR*, volume 1, pages 253–257. IEEE CS Press, 1997. 72, 73

J. Kittler and F. M. Alkoot. Sum versus vote fusion in multiple classifier systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(1):110–115, 2003. 21

J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 20(3):226–239, 1998. 6, 17, 20, 22, 27, 32, 128, 129, 132

J. Kittler and M. S. Nixon, editors. *Audio- and Video-Based Biometric Person Authentication, Fourth International Conference*, volume 2688 of *Lecture Notes on Computer Science*. Springer, 2003. 1, 125

J. Kittler and F. Roli, editors. *Multiple Classifier Systems, First International Workshop*, volume 1857 of *Lecture Notes on Computer Science*. Springer, 2000. 18

R. Kleinberg. Stochastic discrimination. *Annals of Math. and Artificial Intelligence*, 1:207–239, 1990. 22

A. Krogh and J. Vedelsby. Neural network ensembles, cross validation, and active learning. In *Advances in Neural Information Processing Systems, NIPS*, volume 7. MIT Press, 1995. 22

A. Kumar and D. Zhang. Integrating palmprint with face for user authentication. In *Proc. of Workshop on Multimodal User Authentication, MMUA*, pages 107–112, 2003. 27, 32, 56

L. I. Kuncheva. A theoretical study on six classifier fusion strategies. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(4):281–286, 2002. 21

L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004. 18

L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin. Decision templates for multiple classifier fusion: An experimental comparison. *Pattern Recognition Letters*, 34:299–314, 2001. 18

C. H. Lee and Q. Huo. On adaptive decision rules and decision parameter adaptation for automatic speech recognition. *Proceedings of the IEEE*, 88(8):1241–1269, 2000. 33

L. L. Lee, T. Berger, and E. Aviczer. Reliable on-line human signature verification systems. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 18(6):643–647, 1996. 72, 83, 141, 142

E. Lim, X. Jiang, and W. Yau. Fingerprint quality and validity analysis. In *Proc. of the Intl. Conf. on Image Processing, ICIP*, volume 1, pages 469–472, 2002. 102

J. Lopez-Peñalba. Sistema automático de reconocimiento de firma escrita on-line basado en extracción y selección de características discriminatorias. Master's thesis, ETSI Telecomunicación, Universidad Politécnica de Madrid, 2006. 71, 89, 134, 142

D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. FVC2000: Fingerprint Verification Competition. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 24(3):402–412, 2002. 55

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. 1, 54, 125, 138

D. Maltoni and A. K. Jain, editors. *Biometric Authentication, Second International Workshop*, volume 3087 of *Lecture Notes on Computer Science*. Springer, 2004. 1, 125

D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2003. 1, 3, 34, 61, 101, 102

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). 54, 138

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). 54, 138, 139

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. 55, 127

T. Matsui, T. Nishitani, and S. Furui. Robust methods of updating model and a priori threshold in speaker verification. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, pages 97–100, 1996. 29, 42

K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, 1999. 58, 60, 139

MMUA, 2003. Workshop on Multimodal User Authentication, Santa Barbara CA, USA, Dec. 2003. (http://mmua.cs.ucsb.edu/). 14, 134

M. E. Munich and P. Perona. Visual identification by signature tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(2):200–217, 2003. 63

L. M. Muñoz-Serrano. Sistema automático de reconocimiento de huella dactilar basado en información de textura. Master's thesis, ETSI Telecomunicación, Universidad Politécnica de Madrid, 2005. 101, 105, 134, 144

D. Muramatsu, M. Kondo, M. Sasaki, S. Tachibana, and T. Matsumoto. A Markov chain Monte Carlo algorithm for Bayesian dynamic signature verification. *IEEE Trans. on Information Forensics and Security*, 1(1):22–34, 2006. 66, 141

R. Nagel and A. Rosenfeld. Computer detection of freehand forgeries. *IEEE Trans. on Computers*, 26(9):895–905, 1977. 1, 53, 125

J. Naik and G. Doddington. High performance speaker verification using principal spectral components. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, pages 881–884, 1986. 41

V. S. Nalwa. Automatic on-line signature verification. *Proceedings of the IEEE*, 85(2):215–239, 1997. 72

L. Nanni and A. Lumini. Advanced methods for two-class problem formulation for on-line signature verification. *Neurocomputing*, 69:854–857, 2006. 66, 141

A. Navia-Vazquez, F. Perez-Cruz, A. Artes-Rodriguez, and A. R. Figueiras-Vidal. Weighted least squares training of support vector classifiers leading to compact and adaptive schemes. *IEEE Trans. on Neural Networks*, 12 (5):1047–1059, 2001. 45, 123, 146

W. Nelson and E. Kishon. Use of dynamic features for signature verification. In *Proc. of the IEEE Intl. Conf. on Systems, Man, and Cybernetics*, volume 1, pages 201–205, 1991. 72, 73, 83, 142

W. Nelson, W. Turin, and T. Hastie. Statistical methods for on-line signature verification. *Intl. Journal of Pattern Recognition and Artificial Intell.*, 8(3):749–770, 1994. 72, 83, 142

NIST. Image group biometric scores, September 2004. (http://www.nist.gov/biometricscores/). 59

NIST. Image group fingerprint research, August 2005. (http://www.itl.nist.gov/iad/894.03/fing/fing.html). 61

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. 55

J. Ortega-Garcia, J. Fierrez-Aguilar, J. Martin-Rello, and J. Gonzalez-Rodriguez. Complete signal modeling and score normalization for function-based dynamic signature verification. In *Proc. of IAPR Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 658–667. Springer LNCS-2688, 2003a. 14, 39, 40, 72, 73, 77, 134, 141

J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, C. Escudero, and Q.-I. Moro. MCYT baseline corpus: a bimodal biometric database. *IEE Proc. Vision, Image and Signal Processing*, 150(6):391–401, December 2003b. 14, 53, 58, 61, 134, 139

J. Ortega-Garcia, J. Gonzalez-Rodriguez, D. Simon-Zorita, and S. Cruz-Llanas. *Biometric Solutions For Authentication In An E-World*, chapter From Biometrics Technology to Applications Regarding Face, Voice, Signature and Fingerprint Recognition Systems, pages 289–337. Kluwer Academic Publishers, 2002. 14, 71, 89, 134, 141

E. Osuna, R. Freund, and F. Girosi. An improved training algorithm for Support Vector Machines. In *Proc. of IEEE Workshop on Neural Networks for Signal Processing*, pages 276–285, 1997. 111

N. C. Oza, R. Polikar, J. Kittler, and F. Roli, editors. *Multiple Classifier Systems, Sixth International Workshop*, volume 3541 of *Lecture Notes on Computer Science*. Springer, 2005. 6, 18, 129

A. Pacut and A. Czajka. Recognition of human signatures. In *Proc. of the IEEE Joint Intl. Conf. on Neural Networks, IJCNN*, volume 2, pages 1560–1564, 2001. 73

B. Paltridge. Thesis and dissertation writing: An examination of published advice and actual practice. *English for Scientific Purposes*, 21:125–143, 2002. 11

A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991. 57

M. Paulik, N. Mohankrishnan, and M. Nikiforuk. A time varying vector autoregressive model for signature verification. In *Proc. of the 37th Midwest Symposium on Circuits and Systems*, volume 2, pages 1395–1398. IEEE Press, 1994. 29

P. Phillips, A. Martin, C. Wilson, and M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, 33(2):56–63, 2000a. 53, 54, 137, 138

P. J. Phillips. Face and iris evaluations at NIST. In *CardTech/SecurTech*, May 2006. 53

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, 2000b. 1, 53, 54, 125, 137, 138

R. Plamondon, W. Guerfali, and M. Lalonde. Automatic signature verification: A report on a large-scale public experiment. In *Proc. of the 9th Biennial Conference of the International Graphonomics Society, IGS*, pages 9–13, 1999. 72

R. Plamondon and G. Lorette. Automatic signature verification and writer identification: The state of the art. *Pattern Recognition*, 22(2):107–131, 1989. 10, 29, 72, 131

R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. Pattern Anal. and Machine Intell.*, 22(1):63–84, 2000. 72

N. Poh and S. Bengio. Can chimeric persons be used in multimodal biometric authentication experiments? In *2nd Intl. Machine Learning and Multimodal Interaction Workshop, MLMI*, 2005a. 58, 139

N. Poh and S. Bengio. How do correlation and variance of base classifiers affect fusion in biometric authentication tasks? *IEEE Trans. on Signal Processing*, 53(11):4384–4396, 2005b. 22

N. Poh and S. Bengio. Improving fusion with margin-derived confidence in biometric authentication tasks. In *Proc. of Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*, volume Springer LNCS-3546, pages 474–483, 2005c. 33, 123, 145

N. Poh and S. Bengio. An investigation of f-ratio client-dependent normalisation on biometric authentication tasks. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 1, pages 721–724, 2005d. 32

N. Poh and S. Bengio. Database, protocol and tools for evaluating score-level fusion algorithms in biometric authentication. *Pattern Recognition*, 39(2):223–233, 2006. 27, 60, 139

S. Prabhakar and A. Jain. Decision-level fusion in fingerprint verification. *Pattern Recognition*, 35(4):861–874, 2002. 28

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, 53, 54, 91, 92, 93, 125, 137, 138, 143

T. F. Quatieri. *Discrete-Time Speech Signal Processing: Principles and Practice*. Prentice Hall, 2001. 12

L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. 72, 75, 76, 141

D. Ramos-Castro, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. Speaker verification using speaker- and test-dependent fast score normalization. *Pattern Recognition Letters*, 27, 2006a. (to appear). 14, 15, 29, 91, 131, 134, 135, 143

D. Ramos-Castro, J. Gonzalez-Rodriguez, C. Champod, J. Fierrez-Aguilar, and J. Ortega-Garcia. Between-sources modelling for likelihood ratio computation in forensic biometric recognition. In T. Kanade, A. K. Jain, and N. K. Ratha, editors, *Proc. of IAPR Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA*, pages 1080–1089. Springer LNCS-3546, 2005. 15, 135

D. Ramos-Castro, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. Likelihood ratio calibration in a transparent and testable forensic speaker recognition framework. In *ISCA Workshop on Speaker and Language Recognition, ODYSSEY*. IEEE Press, 2006b. (to appear). 124, 146

N. Ratha and R. Bolle, editors. *Automatic Fingerprint Recognition Systems*. Springer, 2004. 1, 102, 125

S. J. Raudys and A. K. Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 13(3):252–264, 1991. 41

D. Reynolds *et al.* The superSID project: Exploiting high-level information for high-accuracy speaker recognition. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, volume 4, pages 784–787, 2003. 92

D. A. Reynolds. Experimental evaluation of features for robust speaker identification. *IEEE Trans. Speech Audio Process.*, 2:639–643, 1994. 93

D. A. Reynolds. Channel robust speaker verification via feature mapping. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP*, pages 53–56, 2003. 93

D. A. Reynolds, W. Campbell, T. T. Gleason, C. Quillen, D. Sturim, P. Torres-Carrasquillo, and A. Adami. The 2004 MIT Lincoln Laboratory speaker recognition system. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP*, volume 1, pages 177–180, 2005. 91, 92, 143

D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted Gaussian Mixture Models. *Digital Signal Processing*, 10:19–41, 2000. 39, 43, 92, 93

J. Richiardi and A. Drygajlo. Gaussian Mixture Models for on-line signature verification. In *Proc. of ACM SIGMM Workshop on Biometric Methods and Applications, WBMA*, pages 115–122, 2003. 66, 79, 141

J. Richiardi, J. Fierrez-Aguilar, J. Ortega-Garcia, and A. Drygajlo. On-line signature verification resilence to packet loss in IP networks. In *Proc. of 2nd Workshop on Biometrics on the Internet, COST-275*, pages 11–16, Vigo, Spain, March 2004. 15, 135

F. Roli, G. Fumera, and J. Kittler. Fixed and trained combiners for fusion of imbalanced pattern classifiers. In *Proc. of the Intl. Conf. on Information Fusion, FUSION*, pages 278–284, 2002a. 25

F. Roli, J. Kittler, G. Fumera, and D. Muntoni. An experimental comparison of classifier fusion rules for multimodal personal identity verification systems. In *Proc. of Third Intl. Workshop on Multiple Classifier Systems, MCS*, pages 252–261, 2002b. 26

A. Ross and R. Govindarajan. Feature level fusion using hand and face biometrics. In *Proc. of Intl. Conf. on Biometric Technologies for Human Identification, BTHI*, volume 5779, pages 196–204. Proc. of SPIE, March 2005. 23

A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003. 25, 27

A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics.* Springer, 2006. 7, 12, 129, 132, 133

A. Ross, J. Reisman, and A. K. Jain. Fingerprint matching using feature space correlation. In M. Tistarelli, J. Bigun, and A. K. Jain, editors, *Proc. of Intl. Workshop on Biometric Authentication, BIOAW*, pages 48–57. Springer LNCS-2359, 2002. 105, 106

J. Saeta and J. Hernando. Automatic estimation of a priori speaker dependent thresholds in speaker verification. In *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pages 70–77. Springer LNCS-2688, 2003. 42

D. Sakamoto, H. Morita, T. Ohishi, Y. Komiya, and T. Matsumoto. On-line signature verification algorithm incorporating pen position, pen pressure and pen inclination trajectories. In *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP*, volume 2, pages 993–996, 2001. 73, 83

C. Sanderson and K. K. Paliwal. Likelihood normalization for face authentication in variable recording conditions. In *Proc. of the IEEE Intl. Conf. on Image Processing, ICIP*, volume 1, pages 301–304, 2002. 40

SC37, 2005. ISO/IEC JTC 1/SC 37 . (http://www.jtc1.org/sc37/). 1, 7, 8, 125, 129

R. Shapire. The strength of weak learnability. *Machine Learning*, 5:197–227, 1990. 18

R. Shapire, Y. Freund, P. Bartlett, and W. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5):1651–1686, 1998. 22

L. Shen, A. Kot, and W. Koo. Quality measures for fingerprint images. In J. Bigun and F. Smeraldi, editors, *Proc. of IAPR Intl. Conf. on Image Analysis and Processing, AVBPA*, pages 266–271. Springer LNCS-2091, 2001. 102

D. Simón-Zorita. *Reconocimiento automático mediante patrones biométricos de huella dactilar*. PhD thesis, ETSI Telecomunicación, Universidad Politécnica de Madrid, 2004. 101, 104, 144

D. Simon-Zorita, J. Ortega-Garcia, J. Fierrez-Aguilar, and J. Gonzalez-Rodriguez. Image quality and position variability assessment in minutiae-based fingerprint verification. *IEE Proceedings Vision, Image and Signal Processing*, 150(6):402–408, 2003. 10, 15, 33, 62, 102, 112, 113, 131, 135, 140

R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. K. Jain. Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):450–455, 2005. 28, 32

F. K. Soong and A. E. Rosenberg. On the use of instantaneous and transitional spectral information in speaker recognition. *IEEE Trans. on Acoust., Speech and Signal Proc.*, 36(6):871–879, 1988. 74

E. Tabassi, C. Wilson, and C. Watson. Fingerprint image quality, August 2004. NIST Research Report NISTIR 7151. 102

S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2003. 12, 17, 37, 44, 45, 56, 74, 75, 85, 138

K. A. Toh, X. Jiang, and W. Y. Yau. Exploiting local and global decisions for multimodal biometrics verification. *IEEE Trans. on Signal Processing*, 52:3059–3072, 2004a. 10, 11, 32, 56, 57, 131

K. A. Toh, W. Y. Yau, E. Lim, and L. C. a C. H. Ng. Fusion of auxiliary information for multi-modal biometrics authentication. In D. Zhang and A. K. Jain, editors, *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 678–685. Springer LNCS-3072, 2004b. 11, 33

V. Tresp and M. Taniguchi. Combining estimators using non-constant weighting functions. In *Advances in Neural Information Processing Systems, NIPS*, volume 7. MIT Press, 1995. 18

K. Tumer and J. Ghosh. Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recognition*, 29:341–348, 1996. 22

R. J. Vandervei. LOQO: An interior point code for quadratic programming. *Optimization Methods and Software*, 12:451–484, 1999. 111

V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2000. 44

P. Verlinde, G. Chollet, and M. Acheroy. Multi-modal identity verification using expert fusion. *Information Fusion*, 1(1):17–33, 2000. 6, 26, 27, 32, 128, 129

Y. Wang, Y. Wang, and T. Tan. Combining fingerprint and voice biometrics for identity verification: An experimental comparison. In D. Zhang and A. K. Jain, editors, *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 663–670. Springer LNCS-3072, 2004. 27, 32, 56

J. Wayman, A. Jain, D. Maltoni, and D. Maio, editors. *Biometric Systems: Technology, Design and Performance Evaluation*. Springer, 2005. 1, 34, 57, 125, 139

C. Wilson *et al.* FpVTE2003: Fingerprint Vendor Technology Evaluation 2003, June 2004. NIST Research Report NISTIR 7123 (http://fpvte.nist.gov/). 1, 33, 34, 54, 102, 112, 125, 137

D. Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992. 18

L. Xu, A. Kryzak, and C. Suen. Methods of combining multiple classifiers and their application to handwritten recognition. *IEEE Trans. on Systems, Man and Cybernetics*, 22(3):418–435, 1992. 18, 19, 23

L. Yang, B. K. Widjaja, and R. Prasad. Application of Hidden Markov Models for signature verification. *Pattern Recognition*, 28(2):161–170, 1995. 72, 73, 141

W. Y. Yau, T. P. Chen, and P. Morguet. Benchmarking of fingerprint sensors. In D. Maltoni and A. K. Jain, editors, *Proc. of Intl. Workshop on Biometric Authentication, BIOAW*, pages 89–99. Springer LNCS-3087, 2004. 102

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, 14, 29, 54, 63, 66, 76, 82, 89, 125, 131, 134, 137, 138, 141, 142

S. Young *et al.* *The HTK Book*. Cambridge University Engineering Department, version 3.2.1 edition, 2002. (available at http://htk.eng.cam.ac.uk/). 74, 93

D. Zhang, editor. *Biometric Solutions for Authentication in an E-World*. Kluwer, 2002. 1, 125

D. Zhang and A. K. Jain, editors. *Biometric Authentication, First International Conference*, volume 3072 of *Lecture Notes on Computer Science*. Springer, 2004. 1, 125

K. Zhang, E. Nyssen, and H. Sahli. A multi-stage on-line signature verification system. *Pattern Analysis and Applications*, 5:288–295, 2002. 72