UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA Y DE LAS COMUNICACIONES

AUTOMATIC SIGNATURE AND GRAPHICAL PASSWORD VERIFICATION: DISCRIMINANT FEATURES AND NEW APPLICATION SCENARIOS

–TESIS DOCTORAL–

Verificación de Firma y Gráficos Manuscritos: Características Discriminantes y Nuevos Escenarios de Aplicación Biométrica

Author: Marcos Martínez Díaz
(Ingeniero de Telecomunicación,
Universidad Autónoma de Madrid)

A thesis submitted for the degree of

Doctor of Philosophy

Madrid, February 2015

Department:	Tecnología Electrónica y de las Comunicaciones Escuela Politécnica Superior Universidad Autónoma de Madrid (UAM), SPAIN
PhD Thesis:	Automatic Signature and Graphical Password Verification: Discriminant Features and New Application Scenarios
Author:	**Marcos Martínez Díaz** Ingeniero de Telecomunicación Universidad Autónoma de Madrid, Spain
Advisor:	**Julián Fiérrez Aguilar** Doctor Ingeniero de Telecomunicación (Universidad Politécnica de Madrid) Universidad Autónoma de Madrid, Spain
Year:	2015
Committee:	President: **Javier Ortega García** Universidad Autónoma de Madrid, Spain
	Secretary:
	Vocal 1:
	Vocal 2:
	Vocal 3:

The research described in this Thesis was carried out within the Biometric Recognition Group – ATVS at the Dept. of Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid.

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 ($^{\text{TM}}$ Adobe Systems Incorporated). The final postscript output was converted to Portable Document Format (PDF) and printed.

Acknowledgements

Me siento muy afortunado por poder estar cerrando esta etapa. Es la meta que me propuse durante mis estudios en la Universidad, y llegué a tener muchas dudas de poder alcanzarla. He podido realizar esta Tesis en el Biometric Recognition Group – ATVS gracias a la ayuda y colaboración de muchas personas. En primer lugar quería agradecer a mi Director, Julián Fierrez su apoyo a lo largo de los últimos años. Sin su empuje y orientación desde que entré en el ATVS allá por 2005 nada de esto habría sucedido. Julián ha sido un ejemplo de cómo enfrentarme de forma inteligente a los problemas y un auténtico guía y fuente de inspiración en esta Tesis. Estoy muy agradecido a Javier Ortega, quien me invitó a unirme al ATVS y desde entonces me ha apoyado y animado a seguir avanzando. Además de los buenos consejos que me ha dado, siempre ha tenido buenas palabras hacia mí, y se lo agradezco especialmente. En el ATVS he tenido la suerte de poder trabajar mano a mano con Javier Galbally, un fenómeno en lo personal y en la investigación. He contado también con la asesoría de Joaquín González y Daniel Ramos quienes con sus inteligentes consejos han contribuido también a esta Tesis. Quiero también agradecer a Pedro Tomé su ayuda durante este periodo, ya sea montando servidores para lanzar experimentos o apoyándome en la última etapa. Muchas gracias a todos vosotros, y al resto del ATVS que habéis estado también ahí.

Antes del doctorado, tuvieron lugar mis estudios en la Escuela Politécnica Superior. Me siento muy afortunado por haber podido formar parte de la primera promoción de Ingeniería de Telecomunicación y, aunque ya lo hice en mi Proyecto de Fin de Carrera, no puedo dejar de agradecer a José María Martínez y a Jesús Bescós su cercanía y orientación a lo largo de todos mis estudios. Le agradezco a ellos, junto a Narciso García, Manuel Alfonseca y Doroteo Torre la pasión que generaron en mí por la ingeniería, las telecomunicaciones y el procesado de señal. Agradezco también muy especialmente a Maria José García-Gutiérrez su disponibilidad, eficacia y la sonrisa con la que me ha tratado siempre desde que comencé en la Escuela.

Fuera de la Universidad he contado con el apoyo y cariño de mi familia a lo largo de estos años. Mi padre me transmitió el interés por la ciencia, creo que no me equivoco al decir que empecé a hacer la Tesis siguiendo su ejemplo. Junto con mi madre me han enseñado el gusto por hacer las cosas bien. En fin, qué decir de mis padres. Cariño, apoyo, paciencia, entrega... este no es el lugar para agradecer todo lo que me han dado hasta hoy, ocuparía más que la propia Tesis. Mis suegros Almudena y Quico han sido otro apoyo importante. Muchas horas dedicadas a esta Tesis han sido en su casa y sus ánimos me han ayudado a continuar cuando faltaban energías.

Y por último agradezco a mi mujer, Isa, que haya estado a mi lado mientras he hecho esta Tesis. Animándome cuando ha hecho falta y ayudándome a levantar el pie del acelerador y a disfrutar de lo importante en la vida cuando me enfrascaba demasiado. Tampoco es este el lugar para contar todo lo que le agradezco a Isa. Carmencita ha nacido hace poco y no ha podido ayudarme mucho con la Tesis. Sólo puedo dar gracias a Dios por haber puesto una cosa tan bonita en mi vida.

Marcos Martínez Díaz
Madrid, febrero de 2015

Abstract[1]

THE PROLIFERATION OF HANDHELD DEVICES such as smartphones and tablets brings a new scenario for biometric authentication, and in particular to automatic signature verification. Research on signature verification has been traditionally carried out using signatures acquired on digitizing tablets or Tablet-PCs.

This PhD Thesis addresses the problem of user authentication on handled devices using handwritten signatures and graphical passwords based on free-form doodles, as well as the effects of biometric aging on signatures. The Thesis pretends to analyze: *(i)* which are the effects of mobile conditions on signature and doodle verification, *(ii)* which are the most distinctive features in mobile conditions, extracted from the pen or fingertip trajectory, *(iii)* how do different similarity computation (i.e. matching) algorithms behave with signatures and graphical passwords captured on mobile conditions, and *(iv)* what is the impact of aging on signature features and verification performance.

Two novel datasets have been presented in this Thesis. A database containing free-form graphical passwords drawn with the fingertip on a smartphone is described. It is the first publicly available graphical password database to the extent of our knowledge. A dataset containing signatures from users captured over a period 15 months is also presented, aimed towards the study of biometric aging.

State-of-the-art local and global matching algorithms are used, namely Hidden Markov Models, Gaussian Mixture Models, Dynamic Time Warping and distance-based classifiers. A large proportion of features presented in the research literature is considered in this Thesis.

The experimental contribution of this Thesis is divided in three main topics: signature verification on handheld devices, the effects of aging on signature verification, and free-form graphical password-based authentication. First, regarding signature verification in mobile conditions, we use a database captured both on a handheld device and digitizing tablet in an office-like scenario. We analyze the discriminative power of both global and local features using discriminant analysis and feature selection techniques. The effects of the lack of pen-up trajectories on handheld devices (when the stylus tip is not in contact with the screen) are also studied.

We then analyze the effects of biometric aging on the signature trait. Using three different matching algorithms, Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and distance-based classifiers, the impact in verification performance is studied. We also study the effects of aging on individual users and individual signature features. Template update techniques are analyzed as a way of mitigating the negative impact of aging.

Regarding graphical passwords, the DooDB graphical password database is first presented. A statistical analysis is performed comparing the database samples (free-form doodles and simplified signatures) with handwritten signatures. The sample variability (inter-user, intra-user

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

and inter-session) is also analyzed, as well as the learning curve for each kind of trait. Benchmark results are also reported using state of the art classifiers.

Graphical password verification is afterwards studied using features and matching algorithms from the signature verification state of the art. Feature selection is also performed and the resulting feature sets are analyzed.

The main contributions of this work can be summarized as follows. A thorough analysis of individual feature performance has been carried out, both for global and local features and on signatures acquired using pen tablets and handheld devices. We have found which individual features are the most robust and which have very low discriminative potential (pen inclination and pressure among others). It has been found that feature selection increases verification performance dramatically, from example from ERRs (Equal Error Rates) over 30% using all available local features, in the case of handheld devices and skilled forgeries, to rates below 20% after feature selection. We study the impact of the lack of trajectory information when the pen tip is not in contact with the acquisition device surface (which happens when touchscreens are used for signature acquisitions), and we have found that the lack of pen-up trajectories negatively affects verification performance. As an example, the EER for the local system increases from 9.3% to 12.1% against skilled forgeries when pen-up trajectories are not available.

We study the effects of biometric aging on signature verification and study a number of ways to compensate the observed performance degradation. It is found that aging does not affect equally all the users in the database and that features related to signature dynamics are more degraded than static features. Comparing the performance using test signatures from the first months with the last months, a variable effect of aging on the EER against random forgeries is observed in the three systems that are evaluated, from 0.0% to 0.5% in the DTW system, from 1.0% to 5.0% in the distance-based system using global features, and from 3.2% to 27.8% in the HMM system.

A new graphical password database has been acquired and made publicly available. Verification algorithms for finger-drawn graphical passwords and simplified signatures are compared and feature analysis is performed. We have found that inter-session variability has a highly negative impact on verification performance, but this can be mitigated performing feature selection and applying fusion of different matchers. It has also been found that some feature types are prevalent in the optimal feature vectors and that classifiers have a very different behavior against skilled and random forgeries. An EER of 3.4% and 22.1% against random and skilled forgeries is obtained for free-form doodles, which is a promising performance.

The author was awarded with the European Biometrics Industry Award 2014 for his study "Graphical Password-based User Authentication with Free-form Doodles", which is based on a part of this Dissertation.

The author was awarded with a Honorable Mention at the Best Student Paper Award at the IEEE Conference on Biometrics: Theory, Applications and Systems 2007 (BTAS 2007), for one publication from this Dissertation: M. Martinez-Diaz, J. Fierrez and J. Ortega-Garcia, "Universal Background Models for Dynamic Signature Verification", in Proc. IEEE BTAS 2007, *Washington DC, USA, September 2007.*

The author obtained top results in international signature verification competitions using methods adapted from this Thesis, namely:

- *BioSecure Signature Evaluation Campaign 2009 (BSEC 2009). The algorithms submitted by the author reached the first position in several categories (BSEC, 2009; Houmani* et al.*, 2012).*
- *ICDAR 2009 Signature Verification Competition (SigComp2009). The algorithm submitted by the author reached the second position in the on-line category (Blankers* et al.*, 2009).*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

How does automatic signature verification perform on handheld devices with touchscreens? Is it feasible to reliably authenticate users with signatures traced with the fingertip, or even just with finger-drawn gestures? How does signature verification performance vary over long periods of time? These topics, among others, arise with the proliferation of touchscreen-enabled handheld devices (e.g. smartphones and tablets), as they have dramatically changed user interaction schemes, from keyboards and mouses to natural gestures.

In the current era of electronic services and pervasive access to information, secure access control and user authentication are common tasks which are usually performed with tokens or passwords. In this field, biometrics has become a focus of interest as it relies on anatomical (e.g. fingerprint, iris) or behavioral (e.g. voice, signature) traits to authenticate a user (Jain *et al.*, 2008). These traits cannot be easily stolen or forgotten. It is now common to find fingerprint verification systems in laptops, face recognition systems on smartphones or for border control purposes and iris verification in a number of airports.

Within biometrics, signature verification is a convenient authentication method that has been an active research field in the last three decades (Fierrez and Ortega-Garcia, 2008; Impedovo and Pirlo, 2008; Impedovo *et al.*, 2012; Plamondon and Lorette, 1989). However, reliable automatic signature verification is a challenging task, mainly because of the notable variability among signatures from the same individual and the risk of highly skilled forgers which, due to their unpredictable nature, are not completely possible to model during the design of a verification system. Since signatures are a behavioral biometric trait, they present a considerable variability even between successive realizations, which can be increased over medium or large periods of time (i.e. biometric aging) (Galbally *et al.*, 2013). Thus, a signature verification system designer must face a high *intra-class* variability (among the signatures of a specific user) and a low *inter-class* variability, when forgeries are considered.

Despite these challenges, signature is one of the most socially accepted biometric traits, as it has been used in financial and legal transactions since long time (Impedovo and Pirlo, 2008; Plamondon and Lorette, 1989).

In contrast with the recent growth of mobile smart devices, little research has been carried out in the field of dynamic signature verification on handheld devices. In most works related to automatic signature verification, experiments are reported using samples captured on a pen tablet (Impedovo and Pirlo, 2008). As a matter of fact, most research-oriented signature databases have been acquired with a pen tablet (Martinez-Diaz and Fierrez, 2009), although there is an emerging interest in signature-based authentication on mobile devices (Blanco-Gonzalo *et al.*, 2013a,b; Houmani *et al.*, 2012, 2008; Impedovo *et al.*, 2012; Sae-Bae and Memon, 2014; Vivaracho-Pascual and Pascual-Gaspar, 2012).

Touchscreens present however some potential drawbacks for signature verification compared to pen tablets. In contrast with touchscreens, most pen tablets usually capture more information than the pen trajectory, namely pen orientation (azimuth and altitude) and pen pressure (see Fig. 1.1). Moreover, pen tablets also detect the pen trajectory when the tip is not in contact with the surface, allowing trajectory acquisition during pen-ups. Thus, due to the reduced amount of available information, it seems reasonable to hypothesize that systems that use signatures captured on a touchscreen for verification may have worse performance that systems using signatures captured with a digitizing tablet.

As an evolution of traditional signature verification, touchscreen devices offer also the possibility to trace the signature, or an equivalent sequence of gestures, directly with the fingertip. Moreover, it has also been found that users tend to type much slower on touchscreen keyboards (Findlater *et al.*, 2011). Thus, finger-drawn gestures (e.g. signatures or graphical passwords) are a convenient and intuitive alternative to traditional alphanumerical passwords. This has been subject of research in the field of *graphical passwords* (Biddle *et al.*, 2012). Graphical user validation methods are also being implemented by major players in the industry (e.g. Google pattern-lock in Android$^{\text{TM}}$ devices and Microsoft Windows 8$^{\text{TM}}$ Picture Password). Authentication based on simple gestures or doodles traced with the fingertip on a touchscreen is gathering as well some interest in the research community (Sae-Bae *et al.*, 2014; Zhao *et al.*, 2014). Unfortunately graphical passwords tend to be much simpler than signatures and may be in general not composed of previously learned or heavily practiced movements. This may imply higher intra-class variability (i.e. variations between different authentication attempts) than signatures or may cause users to forget part or the whole graphical password, that they provided during enrolment. On the other hand, while some users may be concerned about their privacy when registering their signature on an automatic authentication system, doodles may have a higher acceptability.

Due to the fact that biometrics, as an automatic means of human recognition, constitutes a relatively novel field of research (Jain *et al.*, 2008), most efforts undertaken by the different parties involved in the development of this technology (researchers, industry, evaluators, etc.) have been mainly focused on the improvement of its performance (i.e., finding novel methods to obtain lower error rates) (Cappelli *et al.*, 2006; Wayman *et al.*, 2005). As a consequence, other important aspects closely related to this type of systems such as the performance degradation effect known as *aging* have been left partially uncovered (Fairhurst, 2013). Although there always

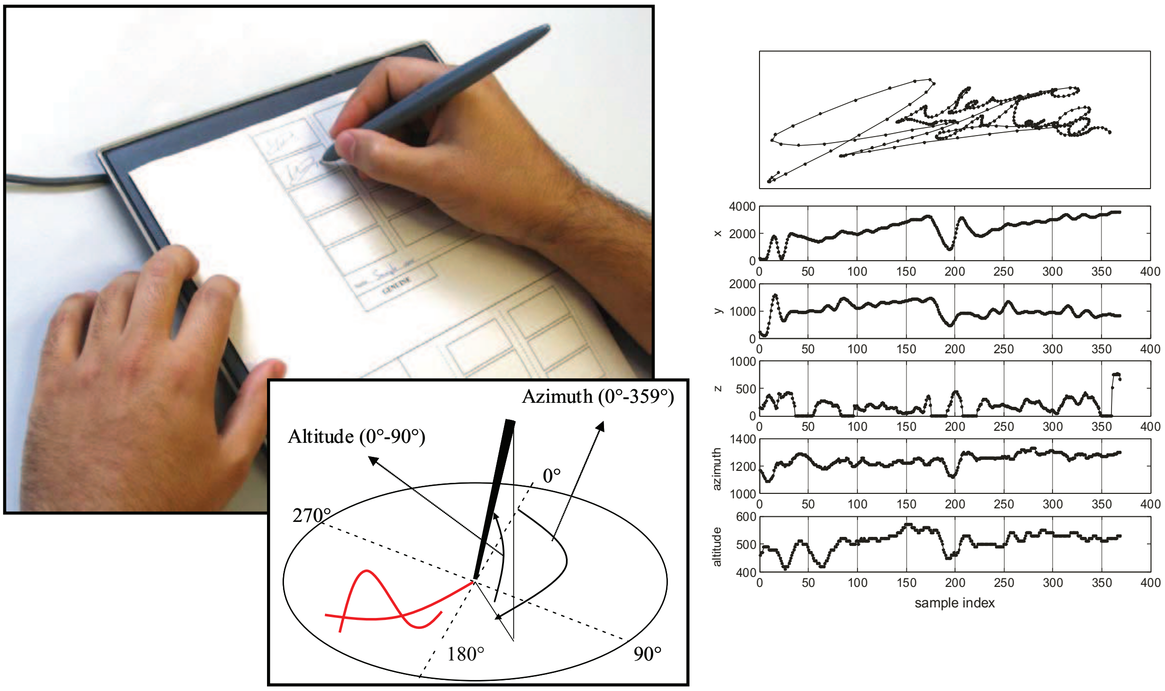


Figure 1.1: *Example of a signature acquisition using a Wacom Intuos 3^{TM} digitizing tablet and a paper template with a delimited signing area for each sample.*

exists a certain variability among biometric samples of one given user (even when they have been acquired successively) (Alonso-Fernandez *et al.*, 2009; Doddington *et al.*, 1998; Houmani *et al.*, 2009), in biometrics the term *aging* is generally used to refer to the gradual decrease in a system performance caused by the changes suffered by the users' trait in the long-term (which cannot be avoided as is inherent to human nature) (Lanitis, 2010). These changes provoked by age imply that, after a sufficiently long period of time, the initial enrolment template of a certain subject substantially differs from his current biometric samples, producing this way lower similarity scores and increasing the error rates of the system. Thus, aging may be considered as a especial type of large intra-class variability caused by the inherent transformations of the human body or behavior over time.

This PhD Thesis addresses the problem of user authentication on handled devices using traditional signatures and graphical passwords based on free-form doodles. The experimental work of the Thesis pretends to analyze: *(i)* which are the effects of mobile conditions on signature and doodle verification, *(ii)* which are the most distinctive features in mobile conditions, extracted from the pen or fingertip trajectory, *(iii)* how do different similarity computation (i.e. matching) algorithms behave with signatures and graphical passwords captured on mobile conditions, and *(iv)* what are the effects of aging on signature verification.

1.1. Biometrics

Biometrics are generally used for *identification* or *verification* purposes (Jain *et al.*, 2004). In the former mode of operation, the biometric trait that individuals present to the system is used to determine which one of the enrolled users in the database they are, leading to a $1 : N$ comparison, where N is the number of users in the database. In the latter, the biometric trait is used to authenticate an individual claiming to be a specific user, which is performed by a $1 : 1$ comparison between the provided biometric trait and the enrolled data of the claimed user. Following a particular identity claim, the user will be *accepted* as *client* or *rejected* as an *impostor* by the system. Throughout this work, we will address the problem of verification, also known as *authentication*.

Verification systems are essentially two-class classifiers, which produce an *accept* or *reject* decision when a biometric trait along with a user identity are presented to the system. Usually, verification is based on a decision threshold. If the similarity (or match score) between the provided trait and the model from the claimed user is higher than a specific threshold, the user is accepted by the system. On the contrary, the user is rejected. In this context, verification systems face two type of errors: False Acceptance (FA) and False Rejection (FR). False Acceptance is produced when a user that falsely claims to be another user is accepted by the system as being the genuine user. False Rejection means that a genuine user is rejected by the system as being an impostor. Given a population of genuine users and impostors and a series of verification trials, the False Acceptance Rate (FAR) and False Rejection Rate (FRR) of the biometric verification system at hand can be computed for any decision threshold.

A common measure to compare the performance of biometric systems is the *Equal Error Rate* (EER). It is computed as the system error rate when the decision threshold is set to satisfy that $FAR = FRR$. Due to the fact that the output of a verification system is in general a binary decision (i.e. accept/reject), the performance of a biometric system is usually represented by a Receiver Operating Characteristic (ROC) or a Detection Error Trade-off (DET) plot (Martin *et al.*, 1997). These plots allow an easy comparison between different systems at any decision threshold.

1.1.1. Biometric Modalities

Several biometric modalities have been proposed in the last decades (Jain *et al.*, 2008). These can be based on physical and behavioral traits depending on their nature. Physical traits are related to anatomical properties of an individual, and include fingerprint, face, iris and hand geometry among others. Behavioral traits refer to how an individual performs an action, and include voice, signature and gait among the most typical. Some examples of popular biometric traits are presented in Fig. 1.2.

Biometric modalities can be further classified by other measures such as the following:

- *Universality*, which states if every person has this biometric.

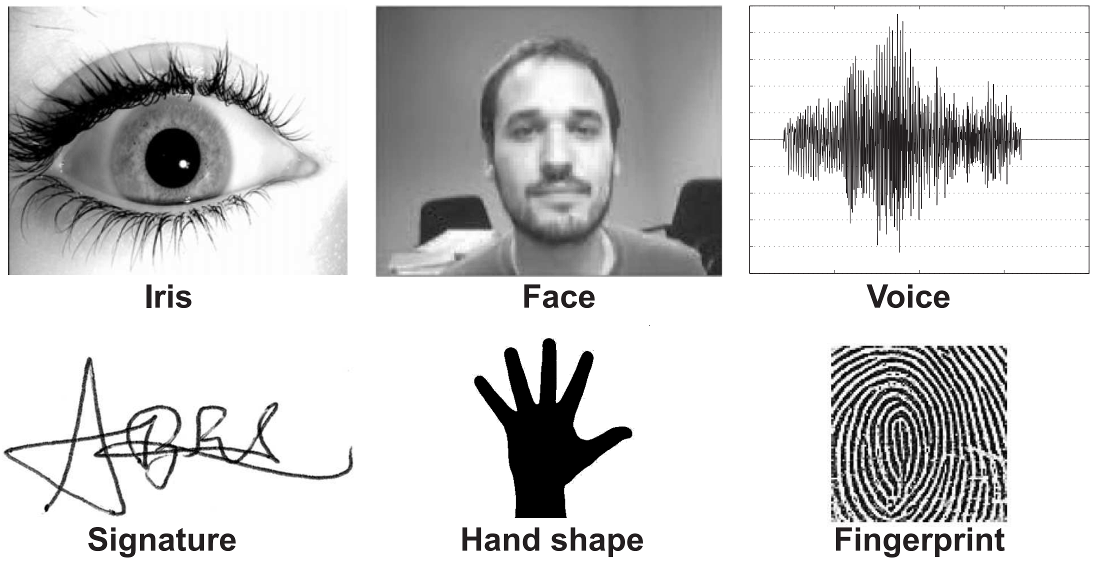


Figure 1.2: *Examples of biometric traits.*

- *Distinctiveness*, related to the discriminative power between different individuals of a biometric modality.
- *Permanence*, which is higher if the traits are invariant along periods of time.
- *Collectability*, which refers to how easy is to acquire the biometric trait.
- *Performance*, related to the speed, or accuracy of systems based on a given biometric.
- *Acceptability*, related to the social perception of the biometric modality.
- *Circumvention*, which refers to the resilience against attacks to security systems based on the biometric.

Other criteria that may be of interest for practical implementation are costs and exception handling, which refers to the case where a manual matching process is required when people cannot interact with the system for any reason. A comparison between some popular biometrics based on the aforementioned measures is presented in Table 1.1. As can be seen, no specific biometric outperforms the rest of them on every category. Consequently, the choice of a modality will depend on the application it is intended to be used for.

1.2. Signature Verification

Signatures have been used since centuries to validate documents and transactions. Therefore, signature is one of the most socially accepted among all biometric traits. In the last few decades, digitizing devices have made possible to perform machine-based signature verification, which has been an intense research field among the biometric and handwriting recognition

Table 1.1: *Qualitative comparison of popular biometric modalities. H, M and L denote High, Medium, and Low respectively. Adapted from (Jain et al., 2004).*

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

research communities. This can be corroborated by the variety of research works conducted during the last decades (Fierrez and Ortega-Garcia, 2008; Impedovo and Pirlo, 2008; Impedovo *et al.*, 2012; Leclerc and Plamondon, 1994; Plamondon and Lorette, 1989; Plamondon and Srihari, 2000) and the amount of competitions held in recent years (Blankers *et al.*, 2009; BMEC, 2007; Houmani *et al.*, 2012, 2011; Liwicki *et al.*, 2011; Malik *et al.*, 2013; Yeung *et al.*, 2004). A number of signature-specific standards have also been published related to biometric data management (ANSI-INCITS 395-2005, 2005; ISO/IEC 19794-11, 2005; ISO/IEC 19794-7, 2005). One of the main challenges in signature verification is related to the signature variability. While signatures from the same user show considerable differences between different captures (high *intra-class* variability), skilled forgers can perform signatures with high resemblance to the user's signature (low *inter-class* variability). Moreover, when a system is designed, only a fraction of information about skilled forgeries can be obtained as forgers with unexpected skills can appear at any time once the system has been deployed.

Two main classes of signature verification systems exist depending on the information extracted from the signature. Off-line systems use only the signature image, while on-line or dynamic systems employ digitized time functions of the signature.

Off-line or **static** signature verification systems use static signature images, which may have been scanned or acquired using a camera, to perform verification. The approaches taken for off-line signature verification have been heterogeneous. Some authors focus on global features using image or shape-oriented pattern recognition techniques (Sabourin, 1997) while others use local features, relying on stroke, texture and structural information (Ammar *et al.*, 1990; Guo *et al.*, 1997; Vargas *et al.*, 2011). Some approaches combine both global and local features (Fierrez-Aguilar *et al.*, 2004; Huang and Yan, 1997).

On-line or **dynamic** systems use captured signature time-functions. These functions are obtained using digitizer tablets or touchscreens (e.g. Tablet-PCs, smartphones, etc.), as illustrated in Fig. 1.1. Traditionally, dynamic systems have presented a better performance than off-line systems as more levels of information than the signature static image are available (Plamondon and Lorette, 1989). This is the approach considered in this Thesis, and will be described in the following chapters.

1.3. Applications of Signature Verification on Handheld Devices

Touchscreen portable devices such as smartphones or tablets provide an appropriate computing platform for signature verification (Martinez-Diaz *et al.*, 2007b, 2009c; Vivaracho-Pascual and Pascual-Gaspar, 2012). In fact, commercial devices already provide handwritten character recognition as a text input alternative (Anquetil and Bouchereau, 2002; Ballagas *et al.*, 2006).

Signature verification can be used for a wide range of applications. Among them, we cite the following:

Payments in commercial environments: the signature is used to validate a payment that is performed via wireless networks. This enables ubiquitous access to commercial transactions. Currently, signatures are not always visually verified at the point of sale, so automatic verification could provide higher security levels.

Legal transactions: legal documents or certificates are signed by the user adding additional security as the signature is verified. This can be a convenient user validation scheme for e-government applications. Using on-line signature verification, the protection against repudiation of signed documents is even increased over traditional signature.

User login: the signature is used to login into a local or remote system as an access control measure (e.g. bank account, personal records, etc.), instead of traditional methods such as PINs or passwords.

Customer validation: a customer is validated by its signature. A client that receives a service or a delivery (e.g. a parcel) signs in a mobile device carried by the deliverer or service provider to certify his conformity.

Paperless office: documents are electronically signed without printing them, providing verification of the signatures and ubiquitous access to them. This allows business process and workflow automation where signatures are needed.

A key advantage of signature verification with respect to other biometric traits in mobile devices is that no additional hardware is needed for acquisition, as it is the case of fingerprint sensors or cameras for fingerprint and face verification systems respectively. Consequently, no extra costs exist and the system complexity does not increase.

1.4. Challenges of Signature Verification on Handheld Devices

Designers of signature verification systems must face many challenges. As has been previously stated, inter- and intra-variability represent two of the main difficulties when trying to reach a good verification performance, specially in the case of skilled forgeries.

Handheld devices such as smartphones or tablets are affected by size and weight constraints due to their portable nature. While processing units, memory chips and battery components are nowadays experimenting higher levels of miniaturization and integration, the input (e.g. keyboard, touchscreen) and output (e.g. display) parts must have reasonable dimensions in order to keep their usability. Poor ergonomics and small input areas on handheld devices are two key factors that increase the variability during the signing process. Moreover, the unfamiliar signing surface may affect the signing process.

The touchscreen digitizing quality should also be taken into account. A typical digitizing pen tablet is based on an electromagnetic principle. The tablet has an embedded wire grid which acts as a transmitter. The pen (which is specifically designed for the tablet) acts as an antenna, which resonates and emits a signal that is captured by the tablet, allowing to detect its position with high accuracy. The tablet detects the pen movement even if it is not in contact with the tablet surface (in a reasonable range of proximity). On the other hand, touchscreens of stylus-oriented handheld devices are based on a resistive principle. Two separated conductive layers are connected when the screen is pressed. The position of the contact point can be accurately detected, but only when the surface is pressed. Capacitive touchscreens are also present in most smartphones and tables. They detect conductive bodies in contact with them and are also unable to detect pressure, although pressure can be estimated by the size of the object in contact. Consequently pressure, pen-azimuth or other signals that have been reported by some authors to increase the verification performance (Muramatsu and Matsumoto, 2007), are not usually captured by touchscreens from handheld devices (although other works suggest that these signals are not among the most discriminative (Houmani *et al.*, 2009)). In addition, the pen trajectory during pen-ups, which is invisible to forgers and provides discriminative information (Sesa-Nogueras *et al.*, 2012), is not available when using touchscreens for acquisition.

Irregular sampling rates and sampling errors, which are common in some mobile devices, may worsen the verification performance and must be addressed during the preprocessing steps.

The interest in security on portable devices has raised in the last decade (Khokhar, 2006). Security is a critical concern while designing a signature verification platform as a breach could give an attacker access to personal data or bank accounts. Gaining access to the matcher could allow an attacker to perform software attacks such as brute force or hill-climbing attacks (Galbally *et al.*, 2007b). The user template must be appropriately secured and encrypted (Argones Rua *et al.*, 2012; Freire-Santos *et al.*, 2006; Maiorana *et al.*, 2008) as well as the communication channels over which signature information may be transmitted.

1.5. Graphical Passwords

Graphical passwords are a topic that has been the subject of active research as a replacement of alphanumerical passwords (Biddle *et al.*, 2012; Suo *et al.*, 2005). The term "graphical password" refers to many different graphical authentication methods, which can be broadly classified in three categories: 1) *recall*, 2) *recognition*, and 3) *cued-recall* passwords. Recall-based systems assume that users remember a graphical password during authentication. Recognition systems present graphical information to the user during authentication, from which the user has to perform a selection matching a set of information previously memorized. Cued-recall systems are a hybrid between the two aforementioned, providing graphical cues that help users recall the previously learned password. An extensive survey of graphical password algorithms has been compiled by Biddle *et al.* (2012).

In the present Dissertation we focus in doodle-based passwords, which are a subset of recall graphical passwords. Individuals are authenticated by using a drawing or sketch, that is captured on a touchscreen during enrollment and is used afterwards for verification. Due to their graphical nature, they are in general easier to remember than classical alphanumerical passwords or PIN codes composed of strings of characters and numbers (Renaud, 2009).

1.6. Motivation of the Thesis

A number of observations from the state of the art have motivated this Thesis.

First, although signature verification has been extensively studied in the literature, little research has been carried out in the field of automatic signature verification on handheld devices. This is seen by the author as one of the currently most natural areas of application of signature-based authentication technology. Unfortunately, the lack of trajectory information during pen-ups, among other limitations, challenges the applicability of traditional approaches usually tested with databases captured with pen-tablets. We understand that the effects of mobility and touchscreens as a capture device should be studied.

Second, the problem of dynamic signature verification has usually been analyzed using predefined sets of features (Fierrez and Ortega-Garcia, 2007; Jain *et al.*, 2002; Kholmatov and Yanikoglu, 2005; Ly-Van *et al.*, 2007), but little attention has been paid to analyzing which specific features are the most discriminative. We have found that feature selection is critical to improve verification performance which has been reflected, for example, in our contribution to the BioSecure Signature Evaluation Campaign (BSEC) 2009 (BSEC, 2009; Houmani *et al.*, 2012). In that competition, the systems presented by the author reached the best performance in a number of categories due to the process of feature selection that was carried out while training our systems.

The third observation is that, in general, signature verification systems are designed and tuned against skilled forgeries (the case where an attacker actively tries to reproduce the forged signature) or against random forgeries (the case where an attacker provides a random signature but claims to be another user), but never both. We have found that, specially in Dynamic Time Warping-based systems, random forgeries and skilled forgeries are completely different problems and systems can be tuned to work specifically against each type of them and afterwards combined. This leads to a better overall performance as also proven in our results in BSEC 2009.

The fourth observation is the lack of research regarding aging and template update in signature verification. It is not easy to find databases where a statistically significant group of people have been captured over a sufficiently long period of time (Rawls and Ricanek, 2009). Furthermore, the acquisition process of such a database should be carried out under almost identical conditions (in terms of acquisition devices, level of control, supervision, etc.) so that the differences in the system performance can be attributed to the elapse of time and not to the variability produced by other external factors. In this context, for the definitive introduction of this biometric technology in the security market, it is relevant to take into account the problem of aging in practical biometric applications, and to implement strategies that compensate the gradual drift of their performance so that their valid life period (in which they are competitive) is increased.

The last observation is that user interaction with handheld devices is becoming increasingly simplified, and the usage of signatures as a daily authentication means may be considered too cumbersome by users. The usage of graphical passwords has been studied in the last decade (Biddle *et al.*, 2012), although the research contributions so far reveal that this field is far from mature (compared to signature verification). As an example, no systematic study, with a reasonably sized and publicly available database for experiments has been carried out, to the extent of our knowledge, except the ones carried out by the author (Martinez-Diaz *et al.*, 2013).

1.7. The Thesis and Main Contributions

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

> *While being convenient and user friendly, signature and graphical password-based authentication on handheld devices is negatively affected by lack of information, sample quality and time variability. This can be partially overcome by the selection of appropriate features and combination of matching algorithms.*

The main contributions of this work are:

- *Signature feature analysis.* A thorough analysis of individual feature performance has been carried out, both for global and local features and on signatures acquired using pen tablets and handheld devices. We have found which individual features are the most robust and which have very low discriminative potential (pen inclination and pressure among others). We study the impact of the lack of trajectory information when the pen tip is not in contact with the acquisition device surface (which happens when touchscreens are used for signature acquisitions), and we have found that the lack of pen-up trajectories negatively affects verification performance.

- *Aging*. We have analyzed the effects of biometric aging in signature verification using a novel dataset (spanning 15 months) and studied a number of ways to compensate the observed performance degradation.

- *Graphical passwords*. We have acquired a new graphical password database and made it publicly available. Algorithms for finger-drawn graphical passwords and simplified signatures have been compared and features analysis has also been performed.

1.8. Outline of the Dissertation

The main objectives of this PhD Thesis are as follows: 1) reviewing and studying the problem of automatic signature verification on handheld devices, focusing on matching algorithms and feature selection; 2) analyzing the effects of aging on signature-based authentication; 3) applying the lessons learned from signature verification to the problem of finger-drawn graphical password authentication on handheld devices.

This Dissertation is structured according to a traditional complex type including background theory, practical methods, and a number of independent experimental studies in which the methods are applied (Paltridge, 2002). The chapter structure is as follows:

- Chapter 1 introduces the topics addressed in this Thesis: signature verification and graphical password-based authentication.

- Chapter 2 summarizes the related works that have motivated this Thesis.

- Chapter 3 describes the verification methods presented in this Thesis, including global and local systems.

- Chapter 4 studies the problem of signature verification on mobile devices compared to pen tablets. The particular effects on signature features are analyzed.

- Chapter 5 studies the effects of aging in handwritten signatures and possible countermeasures.

- Chapter 6 introduces the DooDB Graphical Password Database, which is the first publicly available database of finger-drawn graphical passwords. Quantitative and Qualitative analysis of the database are performed and benchmark results are provided.

- Chapter 7 studies the problem of graphical password-based authentication based on finger-drawn doodles. A number of systems from the signature verification literature are considered and feature selection is performed in order to find the most suitable features.

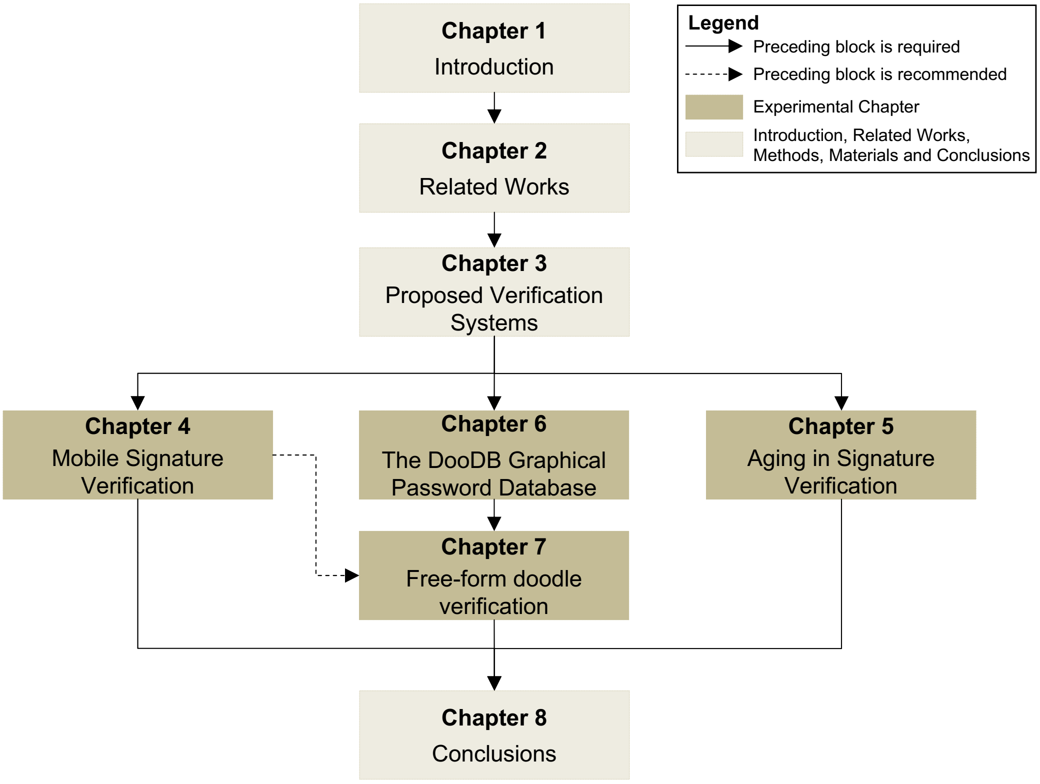

Legend
Preceding block is required
Preceding block is recommended
Experimental Chapter
Introduction, Related Works, Methods, Materials and Conclusions
Chapter 1
Introduction
Chapter 2
Related Works
Chapter 3
Proposed Verification Systems
Chapter 4
Mobile Signature Verification
Chapter 6
The DooDB Graphical Password Database
Chapter 5
Aging in Signature Verification
Chapter 7
Free-form doodle verification
Chapter 8
Conclusions

Figure 1.3: *Dependencies among chapters.*

- Chapter 8 concludes this Dissertation. The main results are discussed and future research areas are proposed.

The dependence between chapters is illustrated in Fig. 1.3. It is recommended to read this Dissertation in consecutive order, although other alternate paths are shown.

If the reader has a background in Biometric Recognition (Jain *et al.*, 2011), the experimental chapters can be read independently.

1.9. Detailed Research Contributions

A list of the research contributions of this PhD Thesis is provided in this section. Some publications appear in several items of the list, they are referenced as citations after the first appearance. Journal articles are highlighted in bold text.

- LITERATURE REVIEWS.

  1. Signature verification.

     - M. Martinez-Diaz and J. Fierrez, ”Signature databases and evaluation”, Stan Z. Li (Eds.), *Encyclopedia of Biometrics*, Springer Verlag, July 2009.
     - M. Martinez-Diaz, J. Fierrez and S. Hangai, ”Signature features”, Stan Z. Li (Eds.), *Encyclopedia of Biometrics*, Springer Verlag, July 2009.
     - M. Martinez-Diaz, J. Fierrez and S. Hangai, ”Signature matching”, Stan Z. Li (Eds.), *Encyclopedia of Biometrics*, Springer Verlag, July 2009.
     - M. Martinez-Diaz, J. Fierrez and J. Ortega-Garcia, ”Automatic signature verification on handheld devices”, S. Kurkovsky (Eds.), Multimodality in *Mobile Computing and Mobile Devices: Methods for Adaptable Usability*, IGI Global, pp. 321-338, May 2009.

- SIGNATURE VERIFICATION.

  1. Experimental studies on the impact of mobility on signature verification.

     - **M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally. “Mobile signature verification: feature robustness and performance comparison”, *IET Biometrics*, Vol. 3, n. 4, pp. 267-277, December 2014.**
     - R. P. Krish, J. Fierrez, J. Galbally, and M. Martinez-Diaz. “Dynamic signature verification on smart phones”, in *Proc. Workshop on User-Centric Technologies and Applications, PAAMS*, pp. 213-222, Salamanca, Spain, May 2013.
     - **N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. Khalil, M. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. R. Alcobé, J. Fabregas, M. Faundez-Zanuy, J. Pascual-Gaspar, V. Cardeñoso-Payo, and C. Vivaracho-Pascual. “BioSecure signature evaluation campaign (BSEC2009): evaluating online signature algorithms depending on the quality of signatures”, *Pattern Recognition*, Vol. 45, n. 3, pp. 993-1003, March 2012.**
     - M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia. “Towards mobile authentication using dynamic signature verification: useful features and performance evaluation”, in *Proc. Intl. Conf. on Pattern Recognition, ICPR*, Tampa, USA, December 2008.
     - M. Martinez-Diaz, J. Fierrez, and J. Ortega-Garcia. “Incorporating signature verification on handheld devices with user-dependent Hidden Markov Models”, in *Proc. International Conference on Frontiers in Handwriting Recognition, ICFHR*, Montreal, Canada, August 2008.
     - M. Martinez-Diaz, J. Fierrez, J. Galbally, F. Alonso-Fernandez and J. Ortega-Garcia, “Signature verification on handheld devices”, in *Proc. MADRINET Workshop*, pp. 87-95, Salamanca, Spain, November 2007.

2. Development of a top-performing algorithm in the BioSecure Signature Evaluation Campaign 2009.

   - Houmani *et al.* (2012)

3. Signature dynamics analysis and signature quality.

   - J. Galbally, R. Plamondon, J. Fierrez and M. Martinez-Diaz. "Quality analysis of dynamic signature based on the Sigma-Lognormal model", in *Proc. IAPR Intl. Conf. on Document Analysis and Recognition, ICDAR*, pp. 633-637, Beijing, China, September 2011.
   - J. Galbally, R. Plamondon, J. Fierrez, C. O'Reilly, M. Martinez-Diaz and J. Ortega-Garcia, "Kinematical analysis of synthetic dynamic signatures using the Sigma-Lognormal model", in *IAPR Proc. Intl. Conf. on Frontiers of Handwriting Recognition, ICFHR*, pp. 113-118, Calcutta, India, November 2010.

4. User-specific model adaptation.

   - Martinez-Diaz *et al.* (2008b)
   - M. Martinez-Diaz, J. Fierrez and J. Ortega-Garcia, "Universal background models for dynamic signature verification", in *Proc. IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS*, pp. 1-6, Washington DC, USA, September 2007.

5. Analysis of the impact of aging on signature verification.

   - **J. Galbally, M. Martinez-Diaz, and J. Fierrez. "Aging in biometrics: an experimental analysis on on-line signature", *PLOS ONE*, Vol. 8, n. 7, pp. e69897, July 2013.**
   - J. Galbally, M. Martinez-Diaz and J. Fierrez, "Ageing in biometrics: a case study in on-line signature", Michael Fairhurst (Ed.), chapter in Age Factors in Biometric Processing, IET, pp. 117-132, 2013.

- GRAPHICAL PASSWORDS

1. Novel datasets.

   - **M. Martinez-Diaz, J. Fierrez, and J. Galbally. "The DooDB graphical password database: data analysis and benchmark results", *IEEE Access*, Vol. 1, pp. 596-605, September 2013.**

2. Novel methods for graphical password authentication based on doodles.

   - M. Martinez-Diaz, C. Martin-Diaz, J. Galbally, and J. Fierrez. "A comparative evaluation of finger-drawn graphical password verification methods", in *IAPR Proc. Intl. Conf. on Frontiers of Handwriting Recognition, ICFHR*, pp. 375-380, Calcutta, India, November 2010.

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

- SYNTHETIC SIGNATURE GENERATION

1. Generation of synthetic signatures and applications.

- J. Galbally, J. Fierrez, M. Martinez-Diaz and J. Ortega-Garcia. "Improving the enrollment in dynamic signature verification with synthetic samples", in *Proc. IAPR Intl. Conf. on Document Analysis and Recognition, ICDAR*, pp. 1295-1299, Barcelona, Spain, July 2009.
- J. Galbally, J. Fierrez, M. Martinez-Diaz and J. Ortega-Garcia. "Evaluation of brute-force attacks to dynamic signature verification using synthetic samples", in *Proc. IAPR Intl. Conf. on Document Analysis and Recognition, ICDAR*, pp. 131-135, Barcelona, Spain, July 2009.
- J. Galbally, J. Fierrez, M. Martinez-Diaz and J. Ortega-Garcia. "Synthetic generation of handwritten signatures based on spectral analysis", in *Defense and Security Symposium, Biometric Technologies for Human Identification, BTHI*, Proc. SPIE, Vol. 7306, pp. 730629, Orlando, USA, April 2009.

▪ BIOMETRIC TEMPLATE PROTECTION

1. Encryption of user signature templates.

- E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia and A. Neri, "Template protection for HMM-based on-line signature authentication", in *Proc. IEEE Computer Society Workshop on Biometrics*, Anchorage, USA, June 2008.
- M. R. Freire, J. Fierrez, M. Martinez-Diaz and J. Ortega-Garcia, "On the applicability of off-line signatures to the fuzzy vault construction", in *Proc. Intl. Conf. on Document Analysis and Recognition, ICDAR*, IEEE Press, Vol. 2, pp. 1173-1177, Curitiba, Brazil, September 2007.

Other doctoral research not included in the Thesis:

▪ ATTACKS TO BIOMETRIC SYSTEMS

1. Direct and Indirect attacks to fingerprint verification systems.

- **M. Martinez-Diaz, J. Fierrez, J. Galbally and J. Ortega-Garcia, "An evaluation of indirect attacks and countermeasures in fingerprint verification systems", Pattern Recognition Letters, Vol. 32, pp. 1643-1651, September 2011.**
- **J. Galbally, J. Fierrez, F. Alonso-Fernandez and M. Martinez-Diaz, "Evaluation of direct attacks to fingerprint verification systems", *Telecommunication Systems, Special Issue on Biometrics*, Vol. 47, n. 3, pp. 243-254, January 2011.**
- M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia and J. A. Siguenza, "Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification", in *Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST*, pp. 151-159, Lexington, USA, October 2006.
- M. Martinez-Diaz, "Vulnerabilidades en sistemas de reconocimiento basados en huella dactilar: Ataques Hill-Climbing" (Vulnerabilities on fingerprint verification systems: Hill-climbing attacks), *MSc Thesis*, Universidad Autonoma de Madrid, September 2006.

Chapter 2

Related Works and State of the Art

In this chapter, a summary of the research in dynamic signature verification, biometric aging, and graphical passwords is provided, presenting related works and available resources.

2.1. Dynamic Signature Verification

2.1.1. Architecture of a Signature Verification System

Dynamic signature verification systems generally share a common architecture. The typical building blocks of an automatic signature verification system are represented in Fig. 2.1. As illustrated, the following steps are performed in most cases (Fierrez and Ortega-Garcia, 2008):

1. **Data Acquisition**: Signature signals are captured from the pen tip using a digitizing tablet or touchscreen. The signature trajectory is sampled and stored as a discrete-time series. While some digitizing tablets provide pressure or pen orientation information, these signals are not commonly available on touchscreens. The sampling rate is usually equal to or above 100 Hz (although lower rates have also been studied (Martinez-Diaz *et al.*, 2007a)). This is a reasonable rate, since it has been observed that the maximum frequencies of the signature time functions are approximately of 20 - 30 Hz (Plamondon and Lorette, 1989). Alternative acquisition techniques have also been studied. Acquisition with a video camera has been proposed by (Munich and Perona, 2003; Muramatsu *et al.*, 2009), as well as using purpose-specific pens which capture the pen dynamics (Martens and Claesen, 1997; Wang *et al.*, 2010). Some examples of different acquisition conditions are illustrated in Figure 2.2.

   After data acquisition, preprocessing steps are commonly performed. These include noise filtering, resampling, or interpolation of missing samples.

2. **Feature Extraction**: Two main approaches have been followed in this step (Martinez-Diaz *et al.*, 2009a): *feature-based* systems extract global features (e.g. signature duration, number of pen-ups, average velocity) from the signature in order to obtain a holistic

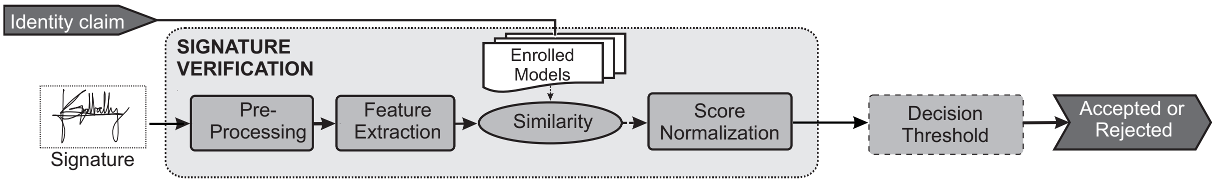


Figure 2.1: *Typical architecture of a signature verification system.*

feature vector (Lee *et al.*, 1996; Sae-Bae and Memon, 2014). On the other hand, *function-based* systems use the signature time functions (e.g. position, pressure) for verification. Traditionally, function-based approaches have yielded better results than feature-based ones (Fierrez-Aguilar *et al.*, 2005a; Kholmatov and Yanikoglu, 2005; Ly-Van *et al.*, 2007).

3. **Enrollment**: In *model-based* systems a statistical client model is computed using a training set of genuine signatures which is used for future comparisons in the matching step (Nanni and Lumini, 2005; Richiardi and Drygajlo, 2003). *Reference-based* systems store the features of each signature provided on the training set as templates. In the matching process the input signature is compared with each reference signature (Lei and Govindaraju, 2005).

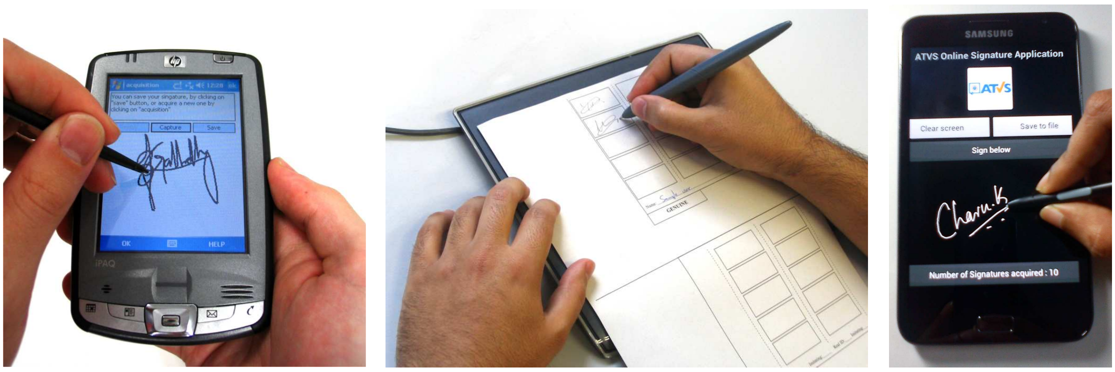
4. **Similarity Computation**: This step involves *pre-alignment* if necessary and a *matching* process, which returns a *matching score* (Martinez-Diaz *et al.*, 2009b). In feature-based systems, statistical techniques like Mahalanobis distance, Parzen Windows or Neural Networks are used for matching (Nelson *et al.*, 1994). Function-based systems use other techniques like Hidden Markov Models - HMM (Dolfing *et al.*, 1998; Fierrez *et al.*, 2007b; Ly-Van *et al.*, 2007), Dynamic Time Warping - DTW (Kholmatov and Yanikoglu, 2005; Martens and Claesen, 1997; Sato and Kogure, 1982), correlation (Parizeau and Plamondon, 1990), and structural descriptors (Bovino *et al.*, 2003; Huang and Yan, 2003; Parizeau and Plamondon, 1990) to compare signature models.

5. **Score Normalization**: The matching score may be normalized to a given range. Score normalization is critical when combining scores from multiple classifiers or in multi-biometric systems (Ross *et al.*, 2006). More sophisticated techniques like target-dependent score normalization can lead to an improved system performance (Fierrez-Aguilar *et al.*, 2005b; Martinez-Diaz *et al.*, 2007c).

An input signature will be considered to belong to the claimed user if its matching score exceeds a given threshold.

2.1.2. Feature-based Systems

Feature-based systems, also known as global systems, have been extensively studied (Fierrez-Aguilar *et al.*, 2005a; Guru and Prakash, 2009; Lee *et al.*, 1996; Lei and Govindaraju, 2005; Richiardi *et al.*, 2005; Sae-Bae and Memon, 2014). In these systems, a holistic vector is formed


SAMSUNG
ATVS Online Signature Application
ATVS
Clear screen
Save to file
Sign below
Number of Signatures acquired : 10

(a) (b) (c)

Figure 2.2: *(a) PDA signature capture process (corresponding to the BIOSECURE DS3 - Mobile Scenario dataset). (b) Pen tablet capture process (corresponding to the BIOSECURE DS2 - Access Control Scenario dataset). (c) Signature capture process on a mobile device.*

by features extracted from the whole signature, such as duration, average speed, number of pen-ups, etc. Despite the large amount of different global feature sets that have been proposed (a maximum of 100 features are considered by Fierrez-Aguilar *et al.* (2005a)), the usually low amount of available training data motivates the usage of feature selection techniques to reduce the feature vector size (due to the curse of dimensionality). Several feature selection techniques have been proposed (see Sect. 2.6), being the Sequential Forward Feature Selection (SFFS) (Pudil *et al.*, 1994) one of the best performing methods reported (Jain and Zongker, 1997). The matching phase is usually performed with statistical classifiers such as Gaussian Mixture Models (Martinez-Diaz *et al.*, 2007c), Parzen Windows (Martinez-Diaz *et al.*, 2007c), majority voting (Lee *et al.*, 1996), or distance measures such as Mahalanobis distance (Galbally *et al.*, 2007b), Manhattan distance (Sae-Bae and Memon, 2014), etc.

2.1.3. Function-based Systems

Function-based systems are also known as local systems. Among these, signature verification systems using Dynamic Time Warping (DTW) (Kholmatov and Yanikoglu, 2005; Martens and Claesen, 1997; Sato and Kogure, 1982), Gaussian Mixture Models (Richiardi and Drygajlo, 2003), and Hidden Markov Models (HMM) (Dolfing *et al.*, 1998; Fierrez *et al.*, 2007b; Ly-Van *et al.*, 2007; Yang *et al.*, 1995) are among the most popular approaches in signature verification. In these systems, the captured time functions (e.g. pen coordinates, pressure, etc.) are used to model each user signature. Additionally, the use of pen orientation features such as azimuth or altitude has been reported to provide good results (Muramatsu and Matsumoto, 2007), although it has been discussed by other authors (Houmani *et al.*, 2009; Lei and Govindaraju, 2005).

Multi-algorithm approaches have been studied for different biometric traits such as fingerprint (Fronthaler *et al.*, 2008) and speech (Rodriguez-Liñares *et al.*, 2003) and can also be applied to signature verification. The combination of the feature- and function-based approaches has been reported to provide better performance than the individual systems (Fierrez-Aguilar *et al.*, 2005a).

2.1.3.1. Gaussian Mixture Models and Hidden Markov Models

Gaussian Mixture Models are popular among the speech recognition literature, and have also been used for signature verification (Richiardi and Drygajlo, 2003). They model a statistical distribution as a linear combination of d-dimensional Gaussian probability density functions (pdf):

$$p\left(\mathbf{x}\left|\lambda_C\right.\right) = \sum_{i=1}^{N} \omega_i p_i\left(\mathbf{x}\right) \tag{2.1}$$

where

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\left|\mathbf{\Sigma}_i\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^T \mathbf{\Sigma}_i^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)\right\}.$$

In order to be a valid pdf, the weights must satisfy $\sum_{i=1}^{N}\omega_i = 1$. The parameters to be estimated are then $\{\omega_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i\}$, $i = 1, ..., N$, where N is the number of Gaussian components, that has to be specified. The covariance matrices are generally chosen to be diagonal, as full matrices do not usually provide an advantage in the model approximation (Reynolds *et al.*, 2000). For a given user C, the model parameters $\{\omega_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i\}$, $i = 1, ..., N$ are estimated from a training set of signatures using the Expectation Maximization (EM) algorithm (Duda *et al.*, 2001).

During the enrollment phase one model is created for each user, which is later used for matching. In addition, a world model $\lambda_{\bar{C}}$ is created, which models the whole set of users. World models, also known as Universal Background Models (Reynolds *et al.*, 2000) are trained using data from a large group of users, as explained in the corresponding experiments.

The match score, given a test vector $\mathbf{x}$ and a target user statistical model λ_C, can be computed as a ratio of the likelihood that the test vector $\mathbf{x}$ is produced by the model λ_C and the likelihood that the test vector has been produced by any other user, which is modeled by the world model $\lambda_{\bar{C}}$.

So, following the previous notation, a match score s is obtained as follows:

$$s = \log p\left(\mathbf{x}\left|\lambda_C\right.\right) - \log p\left(\mathbf{x}\left|\lambda_{\bar{C}}\right.\right). \tag{2.2}$$

Hidden Markov Models (HMM) have also been widely used by the speech recognition community (Rabiner, 1989) as well as in many handwriting recognition applications (Dolfing, 1998). Several approaches using HMMs for dynamic signature verification have been proposed in the last years (Argones Rua and Alba Castro, 2012; Dolfing *et al.*, 1998; Fierrez *et al.*, 2007b; Ly-Van *et al.*, 2007; Muramatsu and Matsumoto, 2003; Yang *et al.*, 1995). An HMM represents a double stochastic process, governed by an underlying Markov chain, with a finite number of states and random function set that generate symbols or observations each of which is associated with one state (Yang *et al.*, 1995). Observations are modeled with GMMs in most speech and

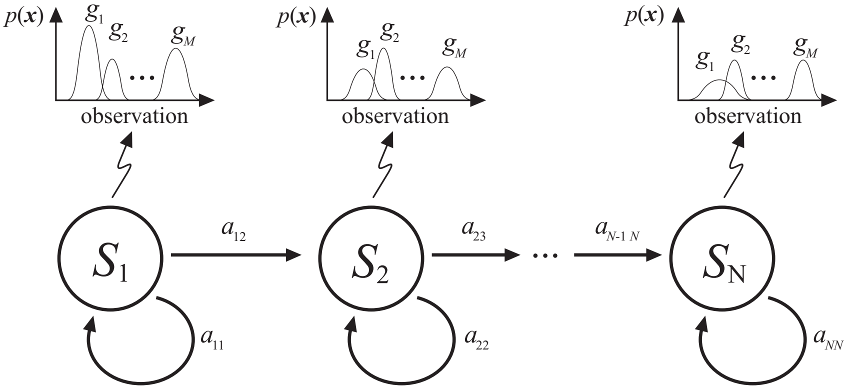


Figure 2.3: *Graphical representation of a left-to-right N-state HMM, with M-component GMMs representing observations and no skips between states.*

handwriting recognition applications. GMMs can, in fact, be considered single-state HMMs.

The basic structure of an HMM using GMMs to model observations is defined by the following elements:

- Number of hidden states N.
- Number of Gaussian Mixtures per state M.
- Probability transition matrix $\mathbf{A} = \{a_{ij}\}$, which contains the probabilities of transitioning from one state to another or staying on the same state.

In Fig. 2.3, an example of a possible HMM configuration is shown. Hidden Markov Models are usually trained in two steps. First, state transition probabilities and observation statistical models are estimated using a Maximum Likelihood algorithm. After this, a re-estimation step is carried out using the Baum-Welch algorithm. A detailed description of the training process is given by Rabiner (1989).

Within HMM-based dynamic signature verification, *regional* and *local* approaches have been proposed. In regional approaches, the extracted time functions are further segmented and converted into a sequence of feature vectors or observations, each one representing regional properties of the signature signal (Dolfing *et al.*, 1998; Kashi *et al.*, 1997; Yang *et al.*, 1995). Some examples of segmentation boundaries are null vertical velocity points (Dolfing *et al.*, 1998) or changes in the quantized trajectory direction (Yang *et al.*, 1995). On the other hand, local approaches directly use the time functions as observation sequences for the signature modeling (Argones Rua and Alba Castro, 2012; Fierrez *et al.*, 2007b; Ly-Van *et al.*, 2007; Richiardi and Drygajlo, 2003).

Finding a reliable and robust model structure for dynamic signature verification is not a trivial task. While too simple HMMs may not allow to model properly the user signatures, too complex models may not be able to model future realizations due to over-fitting. On the other hand, as simple models have less parameters to be estimated, their estimation may be more robust than for complex models. Two main parameters are commonly considered while selecting an optimal model structure: the number of states and the number of Gaussian mixtures per state (Fierrez *et al.*, 2007b). Most of the proposed systems consider a left-to-right configuration without skips between states, also known as Bakis topology, as illustrated in Fig. 2.3.

2.1.3.2. Dynamic Time Warping

Dynamic Time Warping (DTW) is an application of Dynamic Programming to the problem of matching time sequences. Yasuhara and Oka (1977) were the first to report its suitability for dynamic signature verification, by using the algorithm to match time functions extracted from digitized signature signals. Their approach was an adaptation of the original algorithm proposed by Sakoe and Chiba (1978) in the field of speech recognition. The goal of DTW is to find an elastic match among samples of a pair of sequences X and Y that minimize a given distance measure. The algorithm may be defined as follows (Sakoe and Chiba, 1978). Let's define two sequences

$$\begin{aligned}\mathbf{X} &= \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i, ..., \mathbf{x}_I \\ \mathbf{Y} &= \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_j, ..., \mathbf{y}_J\end{aligned} \tag{2.3}$$

and a distance measure as

$$d(i, j) = \|\mathbf{x}_i - \mathbf{y}_j\| \tag{2.4}$$

between sequence samples. A warping path can be defined as

$$\mathbf{C} = \mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k, ..., \mathbf{c}_K \tag{2.5}$$

where each $\mathbf{c}_k$ represents a correspondence (i, j) between samples of $\mathbf{X}$ and $\mathbf{Y}$. The initial condition of the algorithm is set to

$$g_1 = g(1, 1) = d(1, 1) \cdot w(1) \tag{2.6}$$

where g_k represents the accumulated distance after k steps and $w(k)$ is a weighting factor that must be defined. For each iteration, g_k is computed as

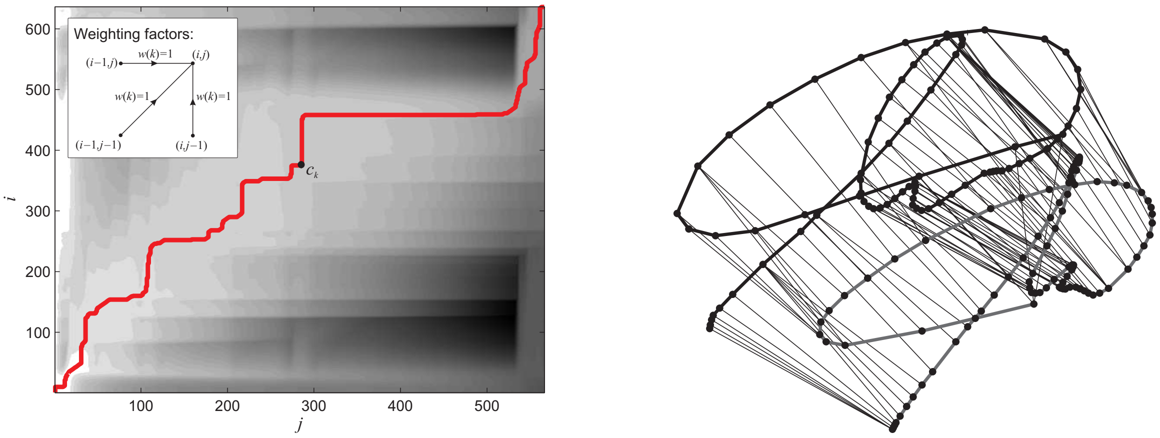
$$g_k = g(i, j) = \min_{c_{k-1}} [g_{k-1} + d(\mathbf{c}_k) \cdot w(k)] \tag{2.7}$$

until the I'th and J'th sample of both sequences respectively is reached. The resulting normalized distance is

$$D(\mathbf{X}, \mathbf{Y}) = \frac{g_K}{\sum_{k=1}^{K} w(k)} \tag{2.8}$$

where $\sum w(k)$ compensates the effect of the length of the sequences.

The weighting factors w_k are defined in order to restrict which correspondences among samples of both sequences are allowed. In Fig. 2.4.a, a possible definition of w_k is depicted, and



(a)

(b)

Figure 2.4: *(a) Optimal warping path between two sequences obtained with DTW and Point-to-point distances are represented with different shades of gray, lighter shades representing shorter distances and darker shades representing longer distances. (b) Example of point-to-point correspondences between two genuine signatures obtained using DTW.*

an example of a warping path between two sequences is given. In this case, only three transitions are allowed in the computation of g_k. Consequently, Eq. (2.7) becomes

$$g_k = g(i,j) = \min \begin{bmatrix} g(i,j-1) + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i-1,j) + d(i,j) \end{bmatrix} \tag{2.9}$$

which is one of the most common implementations found in the literature. In Fig. 2.4.b, an example of point correspondences between two signatures is depicted to visually show the results of the elastic alignment.

The algorithm has been further refined for signature verification by many authors (Faundez-Zanuy, 2007; Kholmatov and Yanikoglu, 2005; Martens and Claesen, 1997; Sato and Kogure, 1982). Moreover, the implementation by Kholmatov and Yanikoglu (2005) won the Signature Verification Competition 2004 (Yeung *et al.*, 2004) and most systems in the BioSecure Signature Evaluation Campaign 2009 (BSEC 2009) used DTW for matching (Houmani *et al.*, 2012). Although the DTW algorithm has been replaced by more powerful ones such as HMMs or SVMs for speech applications, it remains as a highly effective tool for signature verification as it is best suited for small amounts of training data, which is the common case in signature verification (Pascual-Gaspar *et al.*, 2009).

2.1.4. Signature Verification on Handheld Devices

As has been stated before, there is a limited research literature related to dynamic signature verification on handheld devices. Most research-oriented signature databases have been acquired with a pen tablet (Martinez-Diaz and Fierrez, 2009).

Regarding large publicly available datasets, the only existing one is the BioSecure Multimodal

Database (BMDB) (Ortega-Garcia *et al.*, 2010) which contains, among other biometric traits, two signature datasets from the same set of donors. One dataset was captured with a pen tablet (DS2 dataset) and another with a PDA (DS3 dataset), as described in Sect. 2.4. In Fig. 2.2, the capture conditions of both datasets are shown.

In 2007, the BioSecure Multimodal Evaluation was held, where verification algorithms from several European research institutions were compared using the PDA dataset (BMEC, 2007). It was found that error rates were notably higher than in previous competitions, such as SVC 2004 (Yeung *et al.*, 2004), where signatures had been captured on a pen tablet. In 2009, the BioSecure Signature Evaluation Campaign was aimed towards comparing the verification performance between the handheld scenario and the pen tablet scenario (Houmani *et al.*, 2012). Two different tasks were reported. In Task 1, a direct comparison of verification performance using a pen tablet vs. a PDA for signature acquisition was carried out, with signatures from the BMDB database. Task 2 studied the verification performance variation with respect to the information content in signatures (Houmani *et al.*, 2008). Results of Task 1 showed that the participating signature verification algorithms had a significant lower performance against skilled forgeries when signatures were captured on a PDA compared to a pen tablet. On the other hand, verification performance against random forgeries was less negatively affected in the PDA scenario.

A number of works have focused on analyzing the effects on signatures and verification performance when they are captured on handheld devices. It has been found that features extracted from signatures acquired with different devices present statistical distributions that might be significantly different (Elliot, 2004). These statistical differences between features from different devices may affect device inter-operability and may also result in large verification performance differences among sensors. In Alonso-Fernandez *et al.* (2005), the authors compare the error rates of two systems when signatures are captured with two different tablet-PCs. It is shown that the performance depends on the sampling quality of the device used for enrolment. In Simsons *et al.* (2011), the effects of constraining the available signing space are studied, although not specifically for handheld devices. The authors show that the lack of space affects signature complexity, may cause hesitation marks, and reduce fluency, among other factors. Blanco-Gonzalo *et al.* (2013b) have carried out an usability analysis of signature verification on mobile devices. The effects of ergonomics, different kinds of stylus and user position are evaluated. A notable variation in the verification performance is reported depending on the user position among other factors.

Signatures captured with a pen tablet and a handheld device have also been compared from the point of view of their entropy, or information content. In Houmani *et al.* (2008) a client-entropy measure is defined, and it is shown that signatures captured with a PDA have a higher entropy than those captured with a pen tablet. The entropy measure defined in that work increases in general with signature variability and graphical simplicity. Higher verification error rates for signatures with higher entropy are reported.

The performance of a signature verification system using different handheld devices has been analyzed by Blanco-Gonzalo *et al.* (2013a). Signatures produced with a stylus and the fingertip were used in the experiments. It was found that verification using signatures drawn with the fingertip was comparable to signatures captured with a stylus.

2.2. Signature Aging

There are different works where the aging of human biometric traits has been studied from a medical point of view (Coleman and Grover, 2006; Drempt *et al.*, 2011; Morgan *et al.*, 1994; Mueller, 1997), to help in the early diagnosis of diseases (O'Reilly and Plamondon, 2012), or even for its forensic implications (Albert *et al.*, 2007; Walton, 1997). However, not many studies can be found where aging is analyzed from a pure biometrical perspective (two surveys of these works were recently published by Lanitis (2010) and Scheidat *et al.* (2011)). Furthermore, almost all of these aging biometric works are related to the face modality, but, to the best of our knowledge, none of them have been focused on the study of the signature trait.

Among these face-related contributions, there are works dealing with different aspects of aging, for instance, its effect on the performance of face verification systems (Ling *et al.*, 2007; Ramanathan and Chellappa, 2006), methodologies for the synthetic simulation of age (Lanitis, 2008; Lanitis *et al.*, 2002), approaches for the compensation and modeling of the aging effect (Suo *et al.*, 2007), automatic age estimation methods (Geng *et al.*, 2007; Kwon and Lobo, 1999; Lanitis *et al.*, 2004), or descriptions of long-term facial databases (Rawls and Ricanek, 2009). All this interest in the study of the effect of time on face recognition, led in 2004 to the creation of a research group specialized in the analysis of the different factors related to face aging (Face Aging Group, 2004).

Outside the face trait, Modi *et al.* studied the correlation between the quality of fingerprint samples and the age of the users that produced them, and its impact on the final performance of fingerprint recognition systems (Modi and Elliott, 2006; Modi *et al.*, 2007). In the same direction as the fingerprint works by Modi *et al.*, several studies have analyzed the degree of the signing/drawing skill of people belonging to different age groups, their ability to repeat certain valuable recognition features and their vulnerability to eventual imitators (Erbilek and Farihurst, 2012; Faundez-Zanuy *et al.*, 2012; Guest, 2006; Ketcham *et al.*, 2003). Although all these works study an interesting issue related to aging, they are not equivalent to the analysis carried out in the present work, as they do not track individuals over a significant period of their life, but they are focused on establishing a relationship between a certain group of people (e.g., the elderly, youngsters) and a given characteristic (fingerprint quality or signing skill) of their biometric samples (e.g., the elderly-bad quality-poor skill, youngsters-good quality-high skill).

Once the consistent-performance time interval for a given recognition system has been set, an analysis of the best approach to overcome the effect of aging should also be carried out. Among the different palliative methods that have been proposed in the literature, the ones that have received more attention from researchers and industry are automatic template update strategies (Carls, 2009; Rattani *et al.*, 2009). These schemes use some type of target function (e.g., quality measure, similarity score) to automatically select from the most recent biometric samples given by the user to access the system, those which are most suited to be used to recompute (update) the subject's enrolment template. In this field, different fully unsupervised or semiautomatic approaches have been proposed for the fingerprint trait (Lumini and Nanni, 2006; Uludag *et al.*, 2004), for face-based systems (Rattani *et al.*, 2008), signature (Kato *et al.*, 2006), or even in multimodal biometric applications (Roli *et al.*, 2007). Other strategies that try to minimize the effect of aging, are age invariant features (Guest, 2006) and the compensation of age changes (Park *et al.*, 2010).

In addition to the aforementioned works, several authors have also addressed aging-related problems (such as age estimation or age modeling), generally using relatively short-term data, in biometric traits such as the handwriting (Scheidat *et al.*, 2012), the voice (Dobry *et al.*, 2011; Hasan *et al.*, 2012), or even the gait (Veres *et al.*, 2005).

Although it cannot be strictly considered as aging, several works have analyzed the short term variability of signatures using samples captured in the same session (intra-session variability, within minutes), or in different sessions (inter-session variability, within days/weeks) of a regular acquisition campaign (Galbally *et al.*, 2009b; Guest, 2006; Houmani *et al.*, 2009). In these cases, the differences in the systems performance can be attributed more to the inherent variability of the biometric samples (inter and intra-user short term variability) than to a real process of aging, as the time interval between samples is in general too short (Doddington *et al.*, 1998; Poh *et al.*, 2006).

2.3. Other Recent Research Topics Related to Signature Verification

Although not directly related to the dynamic signature verification process, other contributions from the last years are worth noting.

Signature modeling has been studied by O'Reilly and Plamondon (2009). Signatures are modeled as a plan or sequence of strokes executed by the neuromuscular system. Strokes are represented by a Sigma-Lognormal model which models the pen-tip path as a sum of of lognormal primitives. This model has proven to accurately model signatures.

Information content of signatures from an communications theory point of view (i.e. entropy) has also been studied. In Houmani *et al.* (2008) a client-entropy measure is defined, and it is shown that signatures captured with a PDA have a higher entropy than those captured with a pen tablet. The entropy measure defined in that work increases in general with signature variability and graphical simplicity. Signature complexity and feature stability has been also studied in previous works (Brault and Plamondon, 1993; Huang and Yan, 2003).

Generation of synthetic signature samples has also been subject of research. Synthetic signatures can be used for system training or for verification performance. A number of techniques have been proposed (Galbally *et al.*, 2012a,b; Plamondon *et al.*, 2014; Rabasse *et al.*, 2008).

As has been already stated, signature verification systems are exposed to security threats.

While forgeries represent an example of *direct attacks* (i.e. when an attacker has direct access to the acquisition device), *indirect attacks* have also been studied (when an attacker has access to an internal part of the system, such as the matching module). Brute force attacks using synthetic signatures have been studied by Galbally *et al.* (2007b) as well as hill-climbing attacks (Galbally *et al.*, 2009a).

2.4. Dynamic Signature Databases

Publicly available databases make possible for the research community to perform an objective comparison of verification algorithms. Until the last decade, much research had been carried out using private databases, as no large public ones were available. This does not allow reliable performance comparisons of different algorithms, which may have been tuned to a specific database. Moreover, the usage of small datasets reduces the statistical relevance of experiments. Privacy and legal issues have also played a relevant role in the lack of public signature datasets.

The variation of signatures among different cultures must also be taken into account. As an example, in Europe signatures are usually formed by a fast writing followed by a flourish while in North America they usually correspond to the signers name with no flourish. On the other hand, signatures in Asia are commonly formed by Asian characters, which are composed of a larger number of short strokes compared to European or North American signatures.

While some authors have made public the databases used for their research (e.g. Munich and Perona (2003)), most current dynamic signature databases are collected by the joint effort of different research institutions. A summary of the most relevant publicly databases is provided in Table 2.1. In this Section, a brief description of these databases, in chronological order, is provided.

Table 2.1: *Summary of the most popular on-line signature databases. The symbols x, y, p, θ, γ denote pen position horizontal coordinate, vertical coordinate, pen pressure, azimuth and altitude respectively.*

Name	Device	Users	Sessions	Signatures per user		Signals	Interval between sessions
				Genuine	Forgeries		
PHILIPS	Pen tablet	51	3 to 5	30	up to 70	x, y, p, θ, γ	1 week approx.
BIOMET	Pen tablet	84	3	15	up to 12	x, y, p, θ, γ	3 to 5 months
MCYT	Pen tablet	330	1	25	25	x, y, p, θ, γ	-
SVC2004 Task 1	Pen tablet	40	2	20	20	x, y	min. 1 week
SVC2004 Task 2	Pen tablet	40	2	20	20	x, y, p, θ, γ	min. 1 week
SUSIG Blind Subcorpus	Pen tablet	100	1	8 or 10	10	x, y, p	-
SUSIG Visual Subcorpus	Pen tablet	100	2	20	10	x, y, p	1 week approx.
MyIDea	Pen tablet	ca. 100	3	18	18	x, y, p, θ, γ	days to months
BioSecurID	Pen tablet	400	4	16	16	x, y, p, θ, γ	1 month approx.
BioSecure DS2	Pen tablet	ca. 650	2	30	20	x, y, p, θ, γ	1 month approx.
BioSecure DS3	PDA	ca. 650	2	30	20	x, y, p, θ, γ	1 month approx.

PHILIPS Database. Signatures from 51 users were captured using a digitizing tablet at a sampling rate of 200 Hz (Dolfing *et al.*, 1998). The following signals were captured: pen-coordinates, pen-pressure, and the pen-tilt, which is composed by the two angles resulting from

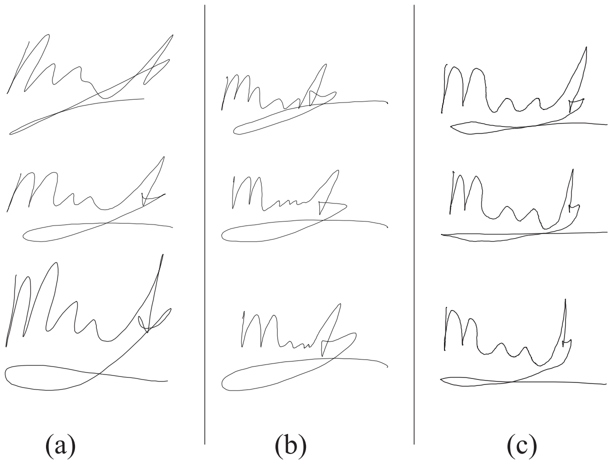


Figure 2.5: *Examples of signatures for a particular subject of the PHILIPS Database. (a) Genuine signatures, (b) over-the-shoulder forgeries, and (c) home improved forgeries. (Adapted from Dolfing* et al. *(1998).)*

the projection of the pen in the (x, z) and (y, z) planes.

Each user contributed 30 genuine signatures, leading to 1530 genuine signatures. Three types of forgeries are present in the database: 1470 over-the-shoulder forgeries, 1530 home-improved and 240 professional forgeries. Over-the-shoulder forgeries were produced by letting the forger observe the signing process. Home-improved forgeries were produced by giving to the forgers the signature static image and letting them to practice at home (see Fig. 2.5). Finally, professional forgeries were performed by forensic document examiners.

MCYT Signature Subcorpus. The MCYT bimodal database is comprised of signatures and fingerprints from 330 individuals (Ortega-Garcia *et al.*, 2003). Signatures were acquired using a WACOM Intuos A6 tablet with a sampling frequency of 100 Hz. The capture area was divided in frames for acquisition of 37.5 mm (width) × 17.5 mm (height). The following time sequences are captured: position coordinates, pressure, azimuth angle and altitude angle. Example signatures and their associated functions are shown in Fig. 2.6.

There are 25 genuine signatures and 25 forgeries per user. Signatures were captured in groups of 5. First, 5 genuine signatures, then 5 skilled forgeries from another user, repeating this until 25 signatures from each type were performed. Each user provided 5 forgeries for the 5 previous users in the database. As the user is forced to concentrate on different tasks between each group of genuine signatures, the variability between groups is expected to be higher than the one within the same group.

BIOMET Signature Subcorpus. This signature subcorpus is part of the BIOMET multimodal database (Garcia-Salicetti *et al.*, 2003). The signatures were captured using a WACOM Intuos2 A6 Pen-tablet and an ink pen with a sampling rate of 100 Hz. The pen coordinates, pen-pressure, azimuth and altitude signals were captured. The database contains data from 84 users, with 15 genuine signatures and 12 forgeries per user. Signatures were captured in two

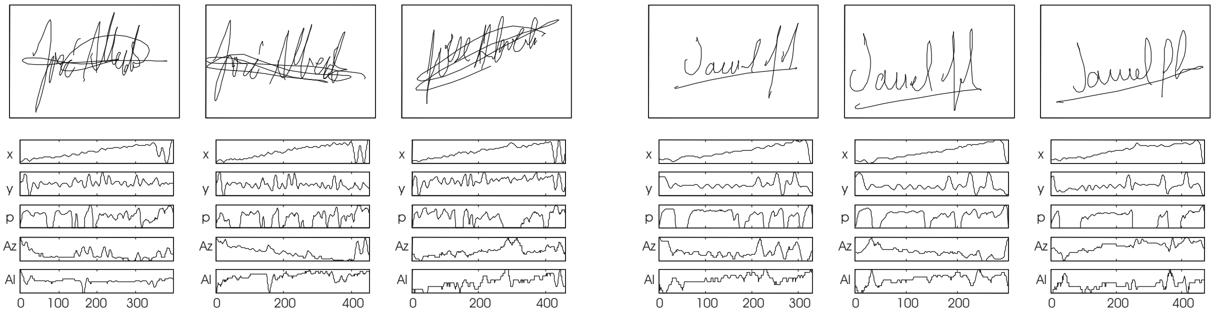


Figure 2.6: *MCYT example signatures and associated functions for two different subjects. One genuine signature (left) and two forgeries (right columns) are presented for each user. (Adapted from* Fierrez et al. *(2007b).)*

sessions separated by 5 months. In the first session, 5 genuine signatures and 6 forgeries were acquired. The remaining 10 genuine signatures and 6 forgeries were captured in the second session. Forgeries were performed by 4 different users (3 forgeries each). This database contains 2201 signatures, since not all users have complete data: 8 genuine signatures and 54 forgeries are missing.

SVC 2004 Database. Two development databases were released prior to the Signature Verification Competition (SVC) 2004 (Yeung *et al.*, 2004). They were captured using a WACOM digitizing tablet and a Grip Pen. Due to privacy issues, users were advised to use invented signatures instead of genuine ones. The two databases differ in the available data, and correspond to the two tasks defined in the competition. One contains only coordinate information while the other provides also pressure and pen orientation signals. Each database contains 40 users, with 20 genuine signatures and 20 forgeries per user acquired in two sessions. Both Occidental and Asian signatures are present in the databases. Examples of signatures from this database are shown in Fig. 2.7.

SUSIG Database. The SUSIG database consists of two sets, one captured using a pen-tablet without visual feedback (Blind subcorpus) and the other using a pen-tablet with an LCD display (Visual subcorpus) (Kholmatov and Yanikoglu, 2008). There are 100 users per database, but these do not coincide, as the Visual subcorpus was captured 4 years after the Blind one. For the Blind subcorpus, a WACOM Graphire2 pen tablet was used. The Visual subcorpus was acquired using an Interlink Electronics ePad-ink tablet, with a pressure-sensitive LCD. In both subcorpus, the pen coordinates and the pen pressure signals were captured, using a sampling frequency of 100 Hz. While performing forgeries, users had prior visual input of the signing process on a separate screen or on the LCD display for the Blind and Visual subcorpus respectively.

For the Blind subcorpus, 8 or 10 genuine signatures were captured in a single session. The users also provided 10 forgeries from another randomly selected user. Two sessions were performed in the Visual subcorpus. During each one, users provided 10 genuine signatures and 5 forgeries.

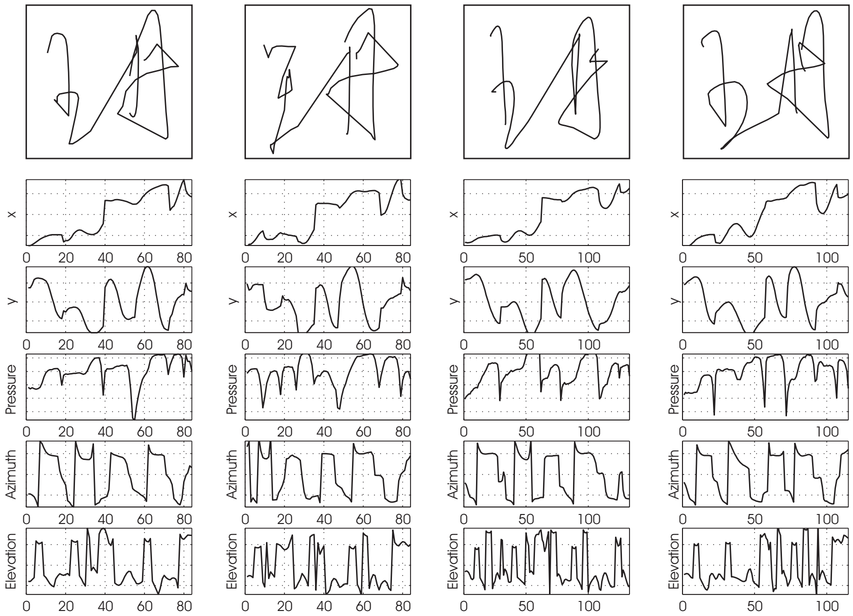


Figure 2.7: SVC 2004 example signatures and associated functions extracted by the pen tablet. For a particular subject, two genuine signatures (left columns) and two forgeries (right columns) are presented. (Adapted from Fierrez et al. *(2007b).)*

MYIDEA CHASM set 1. This signature set is a subset oh the MyIDea multimodal biometric database (Dumas *et al.*, 2005). A WACOM Intuos2 A4 graphic tablet was used at a sampling rate of 100 Hz. Pen position, pressure and azimuth and altitude were captured. This dataset has the particularity that the user must read loud what he is writing, allowing what the authors call CHASM (Combined Handwriting and Speech Modalities). This corpus consists on ca. 70 users. Signatures were captured in 3 sessions. During each session, each user produced 6 genuine signatures and 6 forgeries, with visual access to the images of the target signatures.

BiosecurID Multimodal Database. This database was collected by 6 different Spanish research institutions (Fierrez *et al.*, 2010). It includes the following biometric traits: speech, iris, face, signature, handwriting, fingerprints, hand and keystroke. The data was captured in 4 sessions, distributed in a 4 month time span. The user population was specifically selected in order to contain a uniform distribution of users from different ages and genders. Non-biometric data was also stored, such as age, gender, handedness, vision aids and manual worker (if the user has eroded fingerprints). This allows studying specific demographic groups.

The signature pen-position, pressure, azimuth and altitude signals were acquired using a Wacom Intuos3 A4 digitizer at 100 Hz. During each session, 2 signatures were captured at the beginning and 2 at the end, leading to 16 genuine signatures per user. Each user performed 1 forgery per session of signatures from other 3 users in the database. The skill level of the forgeries is increased by showing to the forger more information of the target signature incrementally. In the first session, forgers have only visual access to one genuine signature, more data (i.e. signature dynamics) is shown in further sessions and forgers are let more time to train. Off-line signature data is also available, since signatures were captured using an inking pen.

BioSecure Signature Subcorpus DS2 - Access Control Scenario. This database was captured under the BioSecure Network of Excellence (Biosecure, 2004; Ortega-Garcia *et al.*, 2010). It is part of the BioSecure multimodal database (Data Set 2) and consists of 667 users. It was acquired at seven different sites in Europe. Acquisition was carried out using a WACOM Intuos3 A6 digitizer at 100 Hz following a procedure similar to the one conducted in MCYT (Ortega-Garcia *et al.*, 2003). The pen coordinates, pressure, azimuth and altitude signals are available.

Signatures were captured in two sessions and in blocks of 5. During each session, users were asked to perform 3 sets of 5 genuine signatures, and 5 forgeries between each set. Each user performed 5 forgeries for the previous 4 users in the database. The users had visual access to the dynamics of the signing process of the signatures they had to forge. Thus, 30 genuine signatures and 20 forgeries are available for each user. An example of the signature capture process of this database including the paper template that was used is depicted in Fig. 2.2.(b).

BioSecure Signature Subcorpus DS3 - Mobile Scenario. The BioSecure Signature Subcorpus DS3 was acquired under the same framework than the Access Control Scenario but on a mobile scenario (BMEC, 2007). It was acquired in 8 different sites in Europe (Alonso-Fernandez *et al.*, 2008; Ortega-Garcia *et al.*, 2010). It is the first multi-session database captured on a PDA. An HP iPAQ hx2790 with a sampling frequency of 100 Hz was used as capture device. Only the pen coordinates and time stamps are available. Users were asked to sign while standing and holding the PDA in one hand. This was done to emulate realistic operating conditions. The acquisition protocol was the same than for the Access Control Scenario Signature Subcorpus, in which signature data was captured using a pen tablet. An average of two months was left between each session. Forgeries for each user are performed by 4 different users (5 forgeries each) in a "worst case" scenario, where each forger has access to the dynamics of the genuine signature in the PDA screen and a tracker tool allowing to see the original strokes. An example of the capture process of this database can be seen in Fig. 2.2.(a). Examples of signatures from the BioSecure Signature subcorpora DS2 and DS3 are presented in Fig. 2.8. Signatures captured with the PDA present missing samples (i.e. sampling errors) due to the PDA touchscreen acquisition process.

2.5. Graphical Password-based Authentication

The term "graphical password" refers to a user authentication method where pictorial information is used for validation, instead of an alphanumerical password. This method poses many challenges, such as memorability (which refers to the easiness to be remembered), usability, and security, since graphical passwords may tend to be visually simple and easily copied by third parties (Biddle *et al.*, 2012).

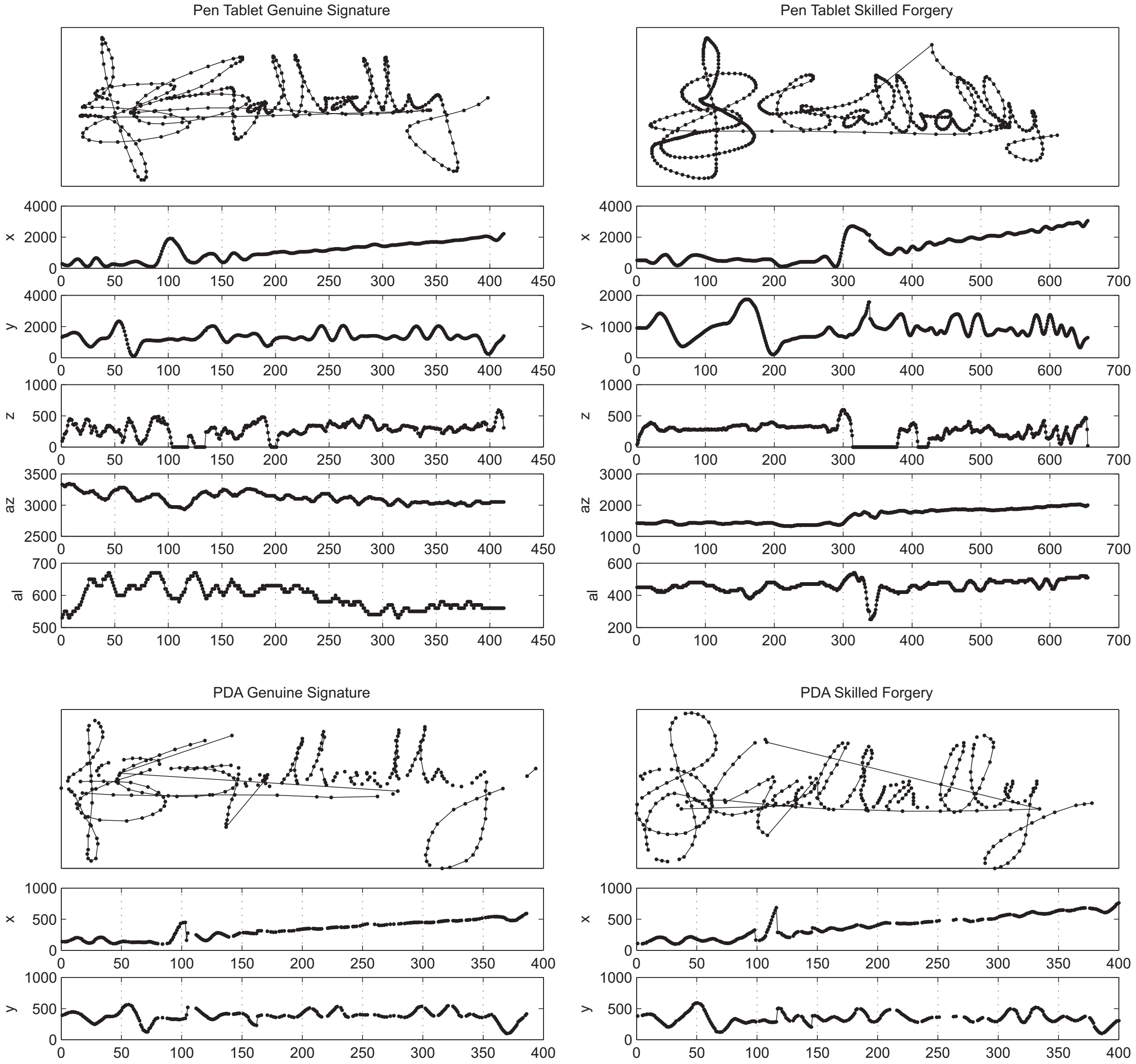


Figure 2.8: Examples of signatures and associated signals from the BioSecure Multimodal Database DS2 and DS3 signature subcorpora captured using a pen tablet (top) and a PDA (bottom), respectively.

Graphical passwords can be broadly classified in three categories: 1) recall, 2) recognition, and 3) cued-recall passwords. In recall-based systems users have to remember a graphical password and provide it during authentication. In recognition systems, graphical information is presented to the user during authentication from which the user has to perform a selection that matches a set of information previously memorized. Cued-recall systems combine the two aforementioned methods, providing graphical cues that help users recall the previously learned password.

In the present work we focus in doodle-based passwords, which fall in the category of recall graphical passwords. Individuals are authenticated by using a drawing or sketch, which is captured on a touchscreen during enrollment and is used afterwards for verification. Due to their graphical nature, they are in general easier to remember than classical alphanumerical passwords or PIN codes composed of strings of characters and numbers (Renaud, 2009).

An extensive survey of graphical password authentication algorithms has been compiled by Biddle *et al.* (2012).

2.5.1. Recall-based Graphical Password Verification

A wide range of approaches for recall-based graphical password authentication have been reported in the literature. Several aspects have been studied such as resilience to forgeries, memorability (i.e. the easiness to remember), user acceptance, error rates, and time to enroll (Biddle *et al.*, 2012). The most relevant methods are surveyed in this section, and summarized in Table 2.2.

Recall-based authentication can be broadly divided in two main categories. *Exact-match* approaches assume that during authentication, a user produces exactly the same drawing provided during enrolment (Jermyn *et al.*, 1999; Tao and Adams, 2008). As a consequence, no biometric information is used. On the other hand, *elastic* approaches allow some variability between enrolment and authentication (Govindarajulu and Madhvanath, 2007; Varenhorst, 2004). Graphical password authentication systems can be also divided into static and dynamic approaches. *Static* or off-line systems use the doodle image for authentication, while *dynamic* or on-line systems use time functions extracted from the doodle trajectory. Dynamic approaches have traditionally reached better verification performances than static systems in the related field of signature verification, since more levels of information are used for authentication.

One of the pioneering contributions in the field is the Draw-A-Secret system (DAS) (Jermyn *et al.*, 1999). The DAS system implements a rectangular 5×5 cell grid where users trace their graphical password. The cell sequence that the users follow is stored as a password. Users are accepted by the system only if they follow the same sequence of cells. The BDAS (Background Draw-a-Secret) was later proposed by Dunphy and Yan (2007), where a background image is shown behind the cell grid. A higher complexity in the password choice and better memorability were reported. The Pass-Go authentication scheme was proposed also as a variation of DAS by Tao and Adams (2008). In that approach, the graphical password is defined by a sequence of grid intersections instead of grid cells, overcoming the limitation of the DAS scheme, where strokes too close to adjacent cell edges could be incorrectly assigned to multiple cells.

The term “passdoodle” was introduced by Goldberg *et al.* (2002). A passdoodle is a graphical password composed of a free-form drawing. In that work, the memorability (i.e. the easiness to remember) of doodles for user authentication is studied, as well as the user preference towards alphanumeric passwords or doodles. However, it is a preliminary study carried out with doodles traced on a sheet of paper. A passdoodle verification system is also proposed by Varenhorst (2004). The stroke spatial distribution and the speed are used for verification.

A doodle authentication system which uses Dynamic Time Warping (DTW) for matching is described by Govindarajulu and Madhvanath (2007). In that work, the trajectory coordinates (x, y), as well as their first and second order derivatives are used as features to characterize each doodle. Recognition performance results are provided using Tamil characters, instead of

Table 2.2: *Summary of related graphical password authentication works, following chronological order. Verification performance is shown if available. Database refers to the number of subjects used in the experiments.*

Method name	**Year**	**Features**	**Matching method**	**Dynamic/Static**	**Verification performance**	**Database**
DAS Jermyn *et al.*	1999	Grid cell sequence	Exact match	Static	N/A	N/A
Passdoodle Goldberg *et al.*	2002	Geometry & color	Visual similarity	Static	N/A	N/A
Passdoodle Varenhorst	2004	Geometry & speed	Multiple measures	Dynamic	98.5% acceptance	10
BDAS Dunphy and Yan	2007	Grid cell sequence	Exact match	Static	N/A	N/A
Pass-Go Tao and Adams	2008	Grid intersection sequence	Exact match	Static	78% acceptance	167
Doodles Govindarajulu and Madhvanath	2008	Geometry, speed, acceleration	Dynamic Time Warping	Dynamic	N/A	N/A
YAGP Gao *et al.*	2008	Stroke orientations	Levenshtein distance	Static	94% acceptance	18
SAS Oka *et al.*	2008	Edge orientation pattern	Correlation	Static	1% EER (random forgeries)	87
PassShapes Weiss and Luca	2008	Stroke orientation	Exact match	Static	94% acceptance	17
Pseudo-signatures Chen *et al.*	2008	Biometric hash	Hash matching	Static	1% EER (skilled forgeries)	37
Graphical Password Zada Khan *et al.*	2011	Predefined symbols	Exact match	Static	N/A	N/A
Multi-touch Sae-Bae *et al.*	2012	Distance between points	Multiple measures	Dynamic	1.58% EER (random forgeries)	34
Password pattern De Luca *et al.*	2012	Coordinates, pressure, speed	Dynamic Time Warping	Dynamic	77% accuracy	31
Lock pattern Angulo and Waestlund	2012	Timing-related features	Random forest	Dynamic	10.39% avg. EER (random forgeries)	32
Touchalytics Frank *et al.*	2013	30 features	k-NN and SVM	Dynamic	3% EER (random forgeries)	41
GEAT Shahzad *et al.*	2013	Velocity, time and acceleration	SVM	Dynamic	0.7% avg. EER (skilled forgeries)	50
SkPWs Riggan *et al.*	2014	Multiple dynamic features	SKS and Fréchet	Dynamic	16.75% EER (skilled forgeries)	35

doodles. Gao *et al.* (2008) presented a static authentication method where free-form sketches are stored as a sequence of cell relative positions. The Levenshtein distance is used to compute distances between sequences. The Scribble-A-Secret (SAS) scheme was later proposed by Oka *et al.* (2008). In that approach, the edge orientation patterns of the doodle static image are used as features, hence no dynamic information is used for authentication. The PassShapes approach considers graphical passwords composed as a sequence of straight strokes following eight possible directions, at 45° angles (Weiss and Luca, 2008). Each stroke is encoded as one of eight different characters, and thus a password is created.

A verification scheme based on predefined visual shapes was described by Chen *et al.* (2009). The system presents a set of cues to the users, which are in general common shapes (e.g. squares, triangles), which the users can follow to define their own free-form password. Cryptographic keys are then generated from the passwords. Similarly, a graphical password verification system based on a set of predefined symbols was proposed by Zada Khan *et al.* (2011). During enrollment, the user first selects a set of predefined symbols (at least 3) and then draws them. The set of symbols constitutes the user password. During authentication, the symbols must be drawn in the same order and are then matched to the predefined templates. If the drawn set is the same as the registered set, the user is validated. No experimental results are provided.

A multi-touch sketch-based authentication approach was described by Sae-Bae *et al.* (2014). In that work, graphical passwords are composed of multi-touch sketches (i.e. drawn with several fingers at the same time). Since the gesture used for authentication is produced with all the fingers, information from the hand geometry is also captured. The GEAT scheme proposed by Shahzad *et al.* (2013) allows user to draw a password composed of a set of ten predefined simple gestures. Many of them are multi-touch gestures. Support Vector Machines (SVM) are used for classification.

Frank *et al.* (2013) presented an authentication scheme based on continuous touchscreen input, instead of specific gestures. SVMs and k-Nearest Neighbor (k-NN) classifiers are used.

A method based on the Simple K-Space (SKS) algorithm and Fréchet distance is proposed by Riggan *et al.* (2014). In this particular approach, dynamic features from the fingertip trajectory as well as the pen tip pressure are used. A usability survey is also carried out on the database users (35 participants) showing, in general, willingness to accept the use of this type of graphical passwords as an authentication means.

Two graphical password approaches have gained popularity in the industry during the last years: the Pattern Lock found in Android OS portable devices and the Picture Password in Windows 8 devices. The Pattern Lock method displays a square grid of 3×3 points on the screen, and users trace a pattern between the points without repeating any of them. This resembles a simplified version of the Pass-Go scheme. Other approaches that also use dynamic information from the Pattern Lock drawing process have been proposed (Angulo and Waestlund, 2012; De Luca *et al.*, 2012). In the Windows 8 Picture Password method a background picture is shown, and users trace on it a password composed of circles, straight lines and points.

The heterogeneity of the existing approaches and lack of public datasets reveals that recall-based graphical password verification is a research field that is still not as established as other closely related fields such as symbol recognition (Llados *et al.*, 2002) or signature verification (Fierrez and Ortega-Garcia, 2008).

In this PhD thesis, doodle-based graphical passwords are considered. The authentication process is performed following the same approach than signature verification, as described in Sect. 2.1.1.

2.5.2. Attacks to Graphical Password-based Systems

Several types of attacks against graphical password authentication systems have been studied in the literature. *Smudge attacks* are those produced when an attacker follows the finger grease path left by the user on the screen (Aviv *et al.*, 2010). *Shoulder-surfing* attacks refer to the case when the attacker has visual access to the password drawing process. Several techniques against shoulder surfing attacks are proposed by Zakaria *et al.* (2011). The authors study how adding fake strokes during the drawing process or removing strokes as they are drawn prevent against forgers.

P. C. van Oorschot and Thorpe (2008) have studied *dictionary attacks* against DAS-like systems. It was shown that users tend to select graphical passwords from a relatively small subspace of cell combinations. Thus an attacker could be successfully accepted after a limited number of random attempts from that particular graphical subspace.

2.6. Feature Selection

Due to the curse of dimensionality (Theodoridis and Koutroumbas, 2006), the performance of a statistical classifier is degraded if the available training data is too small compared to the number of dimensions of the feature vector (Jain and Zongker, 1997). This is usually the case in signature verification, where the average length of a digitized signature is of a few hundreds of samples and the available number of training signatures is relatively small (in practical applications between 3 and 5). The amount of training signatures is mostly conditioned by the willingness of the users to provide many samples during enrollment. Nevertheless, when signatures are captured during only one unique session, their variability is small in general, leading to a poorly trained model.

Feature selection techniques try to reduce the dimensionality of the feature vectors while optimizing the verification accuracy. Their goal is to find the optimal combination of features according to a given optimization criterion. Ideally, given a feature vector of F dimensions, all the possible combinations from 1 to F features should be tested in order to find the optimal combination. Unfortunately, this is not feasible due to the high amount of combinations that have to be tested, which is

$$\sum_{i=1}^{F} \binom{F}{i}.$$

A critical step when performing feature selection is the choice of the optimization criterion. Two main alternatives can be taken: *filter* and *wrapper* methods (Theodoridis and Koutroumbas, 2006). In the former, the optimal feature subset is selected according to intrinsic properties of the training data such as statistical properties. In the latter, the result of the classification problem under consideration is used as the criterion to be optimized. A reasonable choice for a signature verification system is a wrapper method in which the verification performance in terms of the EER is set as the optimization criterion. Wrapper methods require in general more computational resources, as the evaluation of the optimization criterion (e.g. the verification decision) is commonly more complex than the computation of statistical properties of the training data.

Feature selection has been applied to signature verification from several perspectives. Lee *et al.* (1996) propose a method for global features which ranks the discriminative power of each feature for each specific user, based on the distance between the user signatures and the rest of users. They select as an optimal feature vector the one that contains the features that are most commonly ranked among the most discriminative over all the users in the database. Fierrez-Aguilar *et al.* (2005a) perform feature ranking based on their Mahalanobis distance between signatures from different users. The optimal feature vector is then selected by iteratively adding individual features in the order they were ranked and selecting the best performing vector in terms of the system EER. Richiardi *et al.* (2005) propose a distance measure based on the Fisher's Discriminant Ratio and use it to perform Sequential Forward Floating Search Selection (SFFS), which is summarized in this section. Galbally *et al.* (2007a) perform feature selection by using Genetic Algorithms and setting the system EER as the optimization criterion. User-specific feature selection approaches have been also proposed (Kim *et al.*, 1995; Wijesoma *et al.*, 2000).

2.6.1. Feature Selection Algorithms

Several feature selection techniques have been proposed in the literature aimed towards reducing the number of feature combinations that have to be tested. Unfortunately, all of them are only able to find suboptimal solutions. A notable exception is the Branch and Bound algorithm, which is however only applicable when the optimization criterion increases monotonically with the feature subset size. While some of the algorithms are deterministic and always lead to the same suboptimal solution, other algorithms may produce different suboptimal solutions in each execution (Jain and Zongker, 1997). The most popular techniques are summarized next.

2.6.1.1. Scalar Feature Selection

Features are considered individually using this procedure. Each feature is ranked in terms of its class separability using a predefined criterion C, such as the system EER or any distance measure. Then, the N top ranked features in terms of C are selected as the optimal N-dimensional feature vector. This method has the advantage of being computationally simple. Nevertheless, it does not take into account the possible correlations among features. Some techniques to deal with this problem have been proposed in the literature (Theodoridis and Koutroumbas, 2006). This approach is taken by (Fierrez-Aguilar *et al.*, 2005a).

2.6.1.2. Sequential Forward/Backward Selection

In Sequential Forward Selection, given F available features we start by selecting the most discriminative feature x_i. Then, all the possible combinations between this feature and any other feature x_f are computed and the best combination $\{x_i, x_j\}$ is selected. The algorithm continues by iteratively adding features in this manner until the desired feature vector size is reached. The Sequential Backward Selection is similar to this approach but instead of starting with a single feature it starts with all the F features and one feature is removed at a time.

2.6.1.3. Floating Search

Pudil *et al.* (1994) proposed a feature selection algorithm that overcomes some of the limitations of the ones presented above. Namely, when a feature is selected by the previous methods (or discarded, in the backward case), it can no longer be discarded (or selected, in the backward case). This is known as the nesting effect. As with Sequential Selection, both a forward and a backward approach exist. We focus on the forward method, referred to as Sequential Forward Floating Search (SFFS). The algorithm can be summarized as follows (Theodoridis and Koutroumbas, 2006).

Let's consider a set of F features, from which we wish to find the best performing subset of N features, $N \leq F$ in terms of a given criterion C. Let $X_n = \{x_1, x_2, ..., x_n\}$ be the best combination of n features and Y_{F-n} the set of remaining $F - n$ features. In the algorithm, we store the best sets of lower dimensions $X_1, X_2, ..., X_{n-1}$. The following steps are performed until a loop with a stable set X_n is obtained.

1. *Inclusion*
   Choose the element x_{n+1} from Y_{F-n} which, added to X_n produces the best value of the optimization criterion C. Then, $X_{n+1} = \{X_n, x_{n+1}\}$.

2. *Test*

   a) Find the feature x_r that has the least negative (or most positive) effect on the criterion C when it is removed from X_{n+1}.

   b) If $r = n + 1$, change n for $n + 1$ and go to step 1.

   c) If $r \neq n + 1$ and $C(X_{n+1} - \{x_r\}) < C(X_n)$ go to step 1, that is, if removal of any feature does not improve the criterion on the previously selected set X_n, no further backward search is performed.

3. *Exclusion*

a) Remove x_r to get $X'_n = X_{n+1} - \{x_r\}$.

b) Find the feature x_s that has the least negative effect on the criterion C when it is removed from X'_n.

c) If $C(X'_n - \{x_s\}) < C(X_{n-1})$ then $X_n = X'_n$ and go to step 1, that is, if removal of another feature does not improve the criterion on the previously selected set X_n, no further backward search is performed.

d) Remove x_s by putting $X'_{n-1} = X'_n - \{x_s\}$ and $n = n - 1$.

e) Go to step 3.a.

Note that some specific conditions on the first steps have not been considered in order to simplify the algorithm description. The backward algorithm is equivalent to the one explained but removing instead of adding features.

Other algorithms for feature selection include Neural Networks and Genetic Algorithms (Galbally *et al.*, 2007a), although the latter produce variable suboptimal results among different executions. Jain and Zongker (1997) performed an exhaustive comparison of several feature selection algorithms and studied the impact of small training sets on them. The SFFS proved to be highly effective, obtaining results close to the optimal set selected by the Branch and Bound algorithm.

2.7. Chapter Summary and Conclusions

In this chapter, we presented the problem of Dynamic Signature Verification and also described the closely related challenge of automatic verification of recall-based Graphical Passwords. The main verification algorithms from the state of the art were described. We also presented the most popular publicly available databases. Related research areas were also presented as well as feature selection algorithms.

Chapter 3

Proposed Verification Systems

In the present chapter, the automatic signature and doodle verification systems proposed in this Thesis are described. Function-based (referred to as local) and feature-based systems (referred to as global) are considered. The global and a local feature sets that are used are a compendium of existing features from the literature, as will be explained in this chapter.

3.1. Pre-processing

The input coordinate sequence $[\hat{x}_n, \hat{y}_n], n = 1, ..., I$ is sampled from the writing device (or finger-tip) trajectory on a touchscreen, as well as the time interval $\hat{t}_n$ between samples. The trajectory coordinate sequence $[\hat{x}_n, \hat{y}_n]$ is first re-sampled to interpolate missing samples (due to sampling errors or pauses between strokes). Cubic splines are used for interpolation (Catmull and Rom, 1974). The sequences are then normalized to have zero mean, resulting in $[x_n, y_n]$.

In the experimental chapters where signatures are captured on a digitizing pen tablet, the pen azimuth, altitude, and pressure are available, (see Fig. 1.1) and are also processed. This will be explicitly mentioned where applicable. No sampling errors happen in those devices so interpolation is not needed.

3.2. Global Verification System

This feature-based signature verification system extracts a set of 100 global features from each signature $[x_n, y_n]$ normalized coordinate sequence. The feature set was originally described by Fierrez-Aguilar *et al.* (2005a) and is an extension of other sets presented in previous works in the literature (Lee *et al.*, 1996; Nelson and Kishon, 1991; Nelson *et al.*, 1994). A complete description of the feature set is given in Table 3.1. These 100 features can be divided in four categories corresponding to the following magnitudes (the numbering is the same used by Fierrez-Aguilar *et al.* (2005a)):

- **Time** (25 features), related to signature duration, or timing of events such as pen-ups or

Table 3.1: *Set of global features. Table adapted from Fierrez-Aguilar et al. (2005a). T denotes time interval, t denotes time instant, N denotes number of events, and θ denotes angle. Note that some symbols are defined in different features of the table (e.g. Δ in feature 7 is defined in feature 15)*

#	Time related feature	#	Direction related feature
#	Speed and Acceleration related feature	#	Geometry related feature

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

local maxima: 1, 13, 22, 32, 38, 40-42, 50, 52, 58-60, 62, 64, 68, 79, 81-82, 87-90, 94, 100.

- **Speed and Acceleration** (25 features), from the first and second order time derivatives of the position time functions, like average speed or maximum speed: 4-6, 9-11, 14, 23, 26, 29, 31, 33, 39, 44-45, 48, 69, 74, 76, 80, 83, 85, 91-92, 96.
- **Direction** (18 features), extracted from the path trajectory like the starting direction or mean direction between pen-ups: 34, 51, 56-57, 61, 63, 66, 71-73, 77-78, 84, 93, 95, 97-99.
- **Geometry** (32 features), associated to the strokes or signature aspect-ratio: 2, 3, 7-8, 12, 15-21, 24-25, 27-28, 30, 35-37, 43, 46-47, 49, 53-55, 65, 67, 70, 75, 86.

In our implementation, features are normalized into the range $(0, 1)$ using tanh-estimators (Jain *et al.*, 2005). A classifier based on a simplified version of the Mahalanobis distance has been implemented, in order to compare an input signature with a claimed user model. This distance measure has the advantage of being relatively simple to compute and generic enough to provide a reasonable empirical estimate of the statistical class separability achieved by the feature vectors. User models $C = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are created from a training set of genuine signatures, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix obtained from the training signatures. A diagonal covariance matrix is used, and values below a fixed threshold are replaced by the threshold value. This is done to avoid obtaining a singular covariance matrix due to the limited number of training samples in comparison to the problem dimensionality, and to simplify the implementation of this algorithm in handheld devices with limited processing power. The threshold value is 0.00085 and has been heuristically obtained in preliminary experiments. Thus, the matching score s is obtained as the inverse of the "simplified" Mahalanobis distance between the input signature feature vector $\mathbf{x}$ and the claimed user model C:

$$s(\mathbf{x}, C) = \left((\mathbf{x} - \boldsymbol{\mu})^T (\boldsymbol{\Sigma})^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^{-1/2}. \tag{3.1}$$

If the score s computed in 3.1 is above a specific threshold, the signature is considered as genuine. On the contrary it is rejected by the system.

3.3. Local Signature Verification Systems

An HMM system, a GMM system and a DTW-based system have been implemented. In all systems, the $[x_n, y_n]$ normalized signals are used as an input to extract a set of discrete-time functions that model each signature. When available, the $[z_n, \gamma_n, \phi_n]$ pen pressure, pen azimuth and pen altitude are also used. The features considered in this work are an extension from the original set of functions described by Fierrez *et al.* (2007b). We have added features from other contributions (Lei and Govindaraju, 2005; Ly-Van *et al.*, 2007; Richiardi *et al.*, 2005) based on the reported results. In the original set, 7 functions were extracted from the raw signals, from which the first and second order derivatives were computed, leading to a 21-dimensional feature vector. Most second order derivatives have been discarded in this work as they proved to have a very low contribution in the verification performance (as corroborated in Richiardi *et al.* (2005)). In the present Dissertation, an extended set of 15 functions is proposed, plus 12 functions obtained from the first and second order derivatives of some of them. In Table 3.2 we present the resulting set of 27 functions. All features are normalized to have zero mean and variance equal to 1.

Visual examples of the extracted functions can be seen in Fig. 3.1. This set assumes the availability of pressure and pen-inclination information, although this is not the case for signatures acquired on touchscreens. In that case, only 21 features can be extracted from the raw signals.

3.3.1. HMM System

The system implemented in this Thesis is an evolution of the one described by Fierrez *et al.* (2007b). That system participated in the Signature Verification Competition 2004 (Yeung *et al.*, 2004), where it reached the first and second positions against random and skilled forgeries respectively.

In our implementation, based on the HTK Toolkit (Young *et al.*, 2009), an initial step is added to the original HMM training scheme (Fierrez *et al.*, 2007b), leading to the following stages: *i*) the global mean and covariance of the training signatures is assigned to all the mixtures, *ii*) k-means segmentation and Maximum Likelihood training is performed, *iii*) Baum-Welch re-estimation is carried out. The first step allows to have a trainable model for step *iii* (despite being inaccurate) in the case where step *ii* fails due to the large number of parameters to be estimated, or other computational problems.

Similarity scores are computed as the log-likelihood of the signature (using the Viterbi algorithm) divided by the total number of samples of the signature. No score alignment between users is applied (Fierrez-Aguilar *et al.*, 2005b).

In order to keep scores between a reasonable range, normalized scores $\tilde{s}$ between (0,1) are obtained as

$$\tilde{s} = \exp\left(s(\mathbf{x}, C)/30\right), \tag{3.2}$$

where $\mathbf{x}$ and C represent respectively the input signature to verify and the enrolled model of the claimed identity.

The particular implementations used in the experiments (e.g. number of states or number of Gaussian Mixtures) are described in each chapter, where applicable.

3.3.2. GMM System

The GMM system implemented in this Thesis follows the description given in Sect. 2.1.3. The Netlab framework is used for its implementation (Nabney, 2002).

Table 3.2: *Extended set of local features. The upper dot notation (e.g. $\dot{x}_n$) indicates time derivative. Features 3, 10, 15, 16, 17 and 18 are not available on touchscreens.*

#	Feature	Description
1	x-coordinate	x_n
2	y-coordinate	y_n
3	Pen-pressure	z_n
4	Path-tangent angle	$\theta_n = \arctan(\dot{y}_n/\dot{x}_n)$
5	Path velocity magnitude	$\upsilon_n = \sqrt{\dot{y}_n + \dot{x}_n}$
6	Log curvature radius	$\rho_n = \log(1/\kappa_n) = \log(\upsilon_n/\dot{\theta}_n)$, where κ_n is the curvature of the position trajectory
7	Total acceleration magnitude	$a_n = \sqrt{t_n^2 + c_n^2} = \sqrt{\dot{\upsilon}_n^2 + \upsilon_n^2\theta_n^2}$, where t_n and c_n are respectively the tangential and centripetal acceleration components of the pen motion.
8-14	First-order derivative of features 1-7	$\dot{x}_n, \dot{y}_n, \dot{z}_n, \dot{\theta}_n, \dot{\upsilon}_n, \dot{\rho}_n, \dot{a}_n$
15	Pen azimuth	γ_n
16	Pen altitude	ϕ_n
17-18	First-order derivative of features 15-16	$\dot{\gamma}_n, \dot{\phi}_n$
19-20	Second-order derivative of features 1-2	$\ddot{x}_n, \ddot{y}_n$
21	Ratio of the minimum over the maximum speed over a window of 5 samples	$\upsilon_n^r = \min\{\upsilon_{n-4}, ..., \upsilon_n\}/\max\{\upsilon_{n-4}, ..., \upsilon_n\}$
22-23	Angle of consecutive samples and first order difference	$\alpha_n = \arctan(y_n - y_{n-1}/x_n - x_{n-1})$ $\dot{\alpha}_n$
24	Sine	$s_n = \sin(\alpha_n)$
25	Cosine	$c_n = \cos(\alpha_n)$
26	Stroke length to width ratio over a window of 5 samples	$r_n^5 = \frac{\sum_{k=n-4}^{k=n} \sqrt{(x_k-x_{k-1})^2+(y_k-y_{k-1})^2}}{\max\{x_{n-4},...,x_n\}-\min\{x_{n-4},...,x_n\}}$
27	Stroke length to width ratio over a window of 7 samples	$r_n^7 = \frac{\sum_{k=n-6}^{k=n} \sqrt{(x_k-x_{k-1})^2+(y_k-y_{k-1})^2}}{\max\{x_{n-6},...,x_n\}-\min\{x_{n-6},...,x_n\}}$

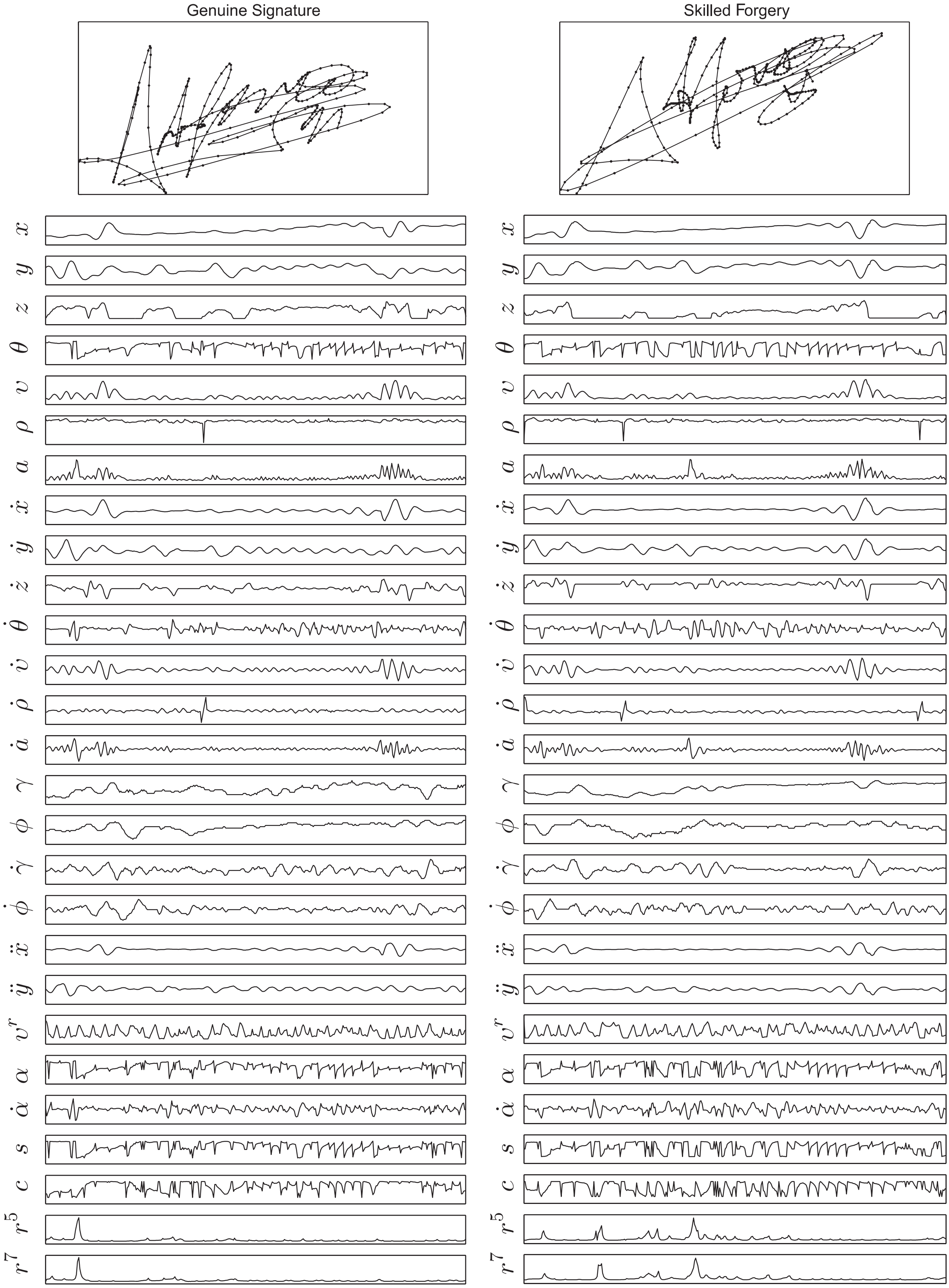


Figure 3.1: *Examples of functions from the 27-feature extended set defined in Table 3.2 for a genuine signature (left) and a skilled forgery (right) of a particular subject from the BIOSECURE DS2 Database.*

Given a number of training samples, the model parameters $\{\omega_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ $i = 1, \ldots, N$ are estimated from a training set of doodles using the Expectation Maximization (EM) algorithm (Duda *et al.*, 2001). The initial data partition (i.e. clustering of data with respect to the Gaussian components) is performed using the k-means algorithm.

In our work, the number of Gaussian components N is chosen to be 32 and diagonal covariance matrices are used, instead of full matrices, due to the limited amount of available data, the better performance reported by Richiardi and Drygajlo (2003) and preliminary experiments which are omitted for the sake of clarity.

During the enrollment phase one model λ_C is created for each user, which is later used for matching. In addition, a world GMM $\lambda_{\bar{C}}$ is created, which models the whole set of users. World models are trained using enrollment samples from a group of users, as explained in the experiments.

Given a test vector $\mathbf{x}$ and a target user statistical model λ_C, the match score is computed as the log-likelihood ratio:

$$s = \log p\left(\mathbf{x} \,|\lambda_C\right) - \log p\left(\mathbf{x} \,|\lambda_{\bar{C}}\right). \tag{3.3}$$

GMM adaptation is a common procedure in speech-related applications (Reynolds *et al.*, 2000). This is not carried out in our implementation since we have observed in preliminary experiments that it does not lead to a better performance.

3.3.3. DTW System

The Dynamic Time Warping System implemented in this work follows the description provided in Sect. 2.1.3.2. Thus, only three transitions are allowed and all weighing factors are equal to 1. Consequently, following Eq. (2.7), g_k is computed as follows:

$$g_k = g(i,j) = \min \begin{bmatrix} g(i,j-1) + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i-1,j) + d(i,j) \end{bmatrix} \tag{3.4}$$

The accumulated distance between the two sequences is computed as

$$D = g(I,J)/K \tag{3.5}$$

where K is the length of the warping path. A normalized match score is obtained as $\tilde{s} = \exp(-D)$.

Given a set of reference samples provided during the enrollment phase and a test signature (or doodle), the scores between all the reference data and the test sample are computed and the average is taken as the match score for that particular test sample.

3.4. Chapter Summary and Conclusions

In this chapter, we described the verification systems implemented in this Thesis and the associated feature sets. Both global and feature sets encompass a notable amount of features already proposed in the literature, plus new ones based on our experience and recent research. The verification algorithms are aligned with the state of the art and have reached top positions in signature verification competitions, namely SVC 2004 for the HMM algorithm (Yeung *et al.*, 2004) and BSEC 2009 for the DTW algorithm implementation (Houmani *et al.*, 2012).

Chapter 4

Mobile Signature Verification

The effects of mobile acquisition conditions in automatic signature verification are studied in this chapter. We focus on the impact of mobility and the usage of touchscreens on the feature discriminative power of different types of features (local and global) compared to the traditional pen tablet scenario. We use for that purpose discriminant analysis of individual features and feature selection algorithms. As described in Chapter 2, signature verification on mobile conditions is affected by a number of factors not present in the pen tablet acquisition scenario. Users are forced to provide their signature on a constrained space, holding a device with their own hands, and using an unfamiliar stylus. This may affect the discriminative power of features. Moreover, since touchscreens do not capture the pen tip trajectory when it is not in contact with the surface, information is lost compared to pen tablets.

We study the performance of state-of-the-art verification systems in both scenarios (pen tablet and handheld device), using the feature sets and the global system and HMM local system described in Chapter 3.

Two different databases are used in the experiments: (i) the BioSecure Multimodal Database (BMDB), as a standard benchmark (Ortega-Garcia *et al.*, 2010); (ii) a signature database captured specifically for this experimental work, using a state-of-the-art device (Samsung Galaxy Note). The BMDB signature database has two subcorpora, one captured on a PDA and other on a digitizing pen tablet. They correspond to the same users in both devices, allowing a fair comparison between them. The reader may refer to Sect. 2.4 for a detailed description of the BMDB database.

This chapter is structured as follows. First, the verification systems used in the experiments are referenced in Sect. 4.1. The experimental protocol is described in Sect. 4.2. Results are reported in Sect. 4.3. Conclusions are finally drawn in Sect. 4.4.

This chapter is based on the publications: (Martinez-Diaz *et al.*, 2008a, 2014).

4.1. Verification systems

Global Verification System. The global system described in Sect. 3.2 is used, as well as the 100-feature set.

Local Verification System. The local HMM system described in Sect. 3.3 is used in the experiments, as well as the 27-feature set. Each user is modeled with a 2-state HMM with 32 Gaussian mixtures per state, following the implementation described by (Fierrez *et al.*, 2007b) that participated in the SVC 2004 competition (Yeung *et al.*, 2004).

4.2. Databases and Experimental Protocol

4.2.1. Databases

Two databases are used in the experiments, the Biosecure Multimodal DataBase (BMDB), acquired using a pen-tablet and a PDA (Ortega-Garcia *et al.*, 2010), and a database captured using a Samsung Galaxy Note device, referred to as SG-NOTE.

A subset of 120 users from the BMDB is used in this work [1]. It contains 20 genuine signatures and 20 skilled forgeries per user and acquisition device (PDA and pen tablet). Genuine signatures were acquired in two different sessions separated by an average period of two months. The first five signatures correspond to the initial session while the remaining 15 belong to the second session. Signatures were captured with a PDA while the user was standing and holding the device with one hand in the handheld scenario, whereas for the pen tablet case they were acquired while the user was sitting, using a pen on a paper placed over the tablet (see Fig.2.2.a and 2.2.b). This emulates real operating conditions.

In both devices, skilled forgeries for each user were performed by 4 different forgers (5 forgeries each) under "worst case" conditions: each forger had visual access to the dynamics of the genuine signature using a tracker tool that allowed replaying the original strokes.

Only the x and y position signals and the sample timestamps are captured by the PDA, while pressure (z) and pen orientation (θ, γ) signals are also acquired by the pen tablet. Pen trajectories during pen-ups (when the pen tip is not in contact with the tablet surface) are recorded by the pen tablet but are not available in the PDA dataset. It is found in the pen tablet dataset that, for each genuine signature, an average of 18% of sampled points correspond to pen-up trajectories (i.e. when the pen tip is not in contact with the tablet surface). A histogram of the proportion of sample points during pen-ups compared to the total signature samples is depicted in Fig. 4.1. In order to evaluate the effect of the lack of pressure and inclination information and pen motion during pen-ups, a third signature dataset is artificially created by removing the samples produced during pen-ups (i.e. having pressure values equal to 0) in the pen tablet dataset. This set will be referred to as "Tablet interpolated pen-ups".

[1] This subset corresponds to the "120 common DS2/DS3" signature dataset available in the BioSecure Foundation web site: http://biosecure.it-sudparis.eu/AB/index.php?option=com_content&view=article&id=72

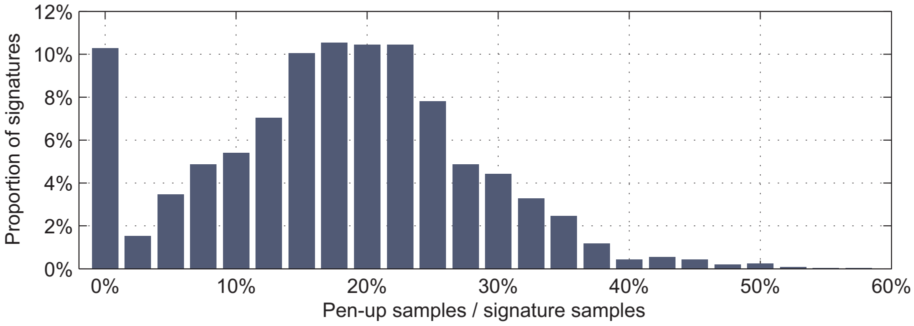


Figure 4.1: *Histogram of signatures classified by the proportion of sampled points during pen-up trajectories vs. total signature sample points, computed on the pen tablet signature dataset.*

Pen-up trajectories are interpolated in the PDA and in the tablet interpolated pen-ups dataset. For the PDA subset, an additional preprocessing step is performed to interpolate erroneous (missing) samples.

From each of the three BMDB subsets (i.e. PDA, Tablet, and Tablet interpolated pen-ups), each one containing 120 users, signatures from the first 50 users are used for development purposes (i.e. individual feature analysis and feature selection), while the remaining 70 are left to validate the performance of the optimal feature vectors selected by the SFFS algorithm. We will refer to the development datasets as BMDB-DEV50 and to the validation datasets as BMDB-VAL70.

This setup follows the protocol of the BioSecure Signature Evaluation Campaign (Houmani *et al.*, 2012), where a subset of 50 users was released for algorithm tuning prior to the competition, which was later carried out using a different test dataset.

The SG-NOTE database[1] is also used for performance validation, in addition to the BMDB-VAL70 subset. This dataset was captured by the authors using a Samsung Galaxy Note mobile phone and contains signatures from 25 users. The SG-NOTE database was captured in two different sessions with an average gap of 5 days between them. In each session, signatures were acquired in two blocks of 5 samples, with a short break between blocks. No skilled forgeries are available in this database. Consequently, the database contains a total amount of 500 signatures (25 users × 2 sessions × 10 signatures per session). See Fig. 2.2.c for an example signature acquisition in SG-NOTE.

The five genuine signatures from the initial session are used for enrolment, both for the global and local systems. *Genuine* user scores are computed using the remaining from the second session (15 signatures in BMDB and 10 signatures in SG-NOTE). *Random forgery* scores (the case where a forger uses his own signature claiming to be a different user) are obtained by comparing the user model to one signature sample of all the remaining users. *Skilled forgery* scores for the BMDB datasets are computed comparing the 20 available skilled forgeries per user with his or her own model (trained with five signatures, as stated before).

[1] This database is available at the ATVS - Biometric Recognition Group web site: http://atvs.ii.uam.es

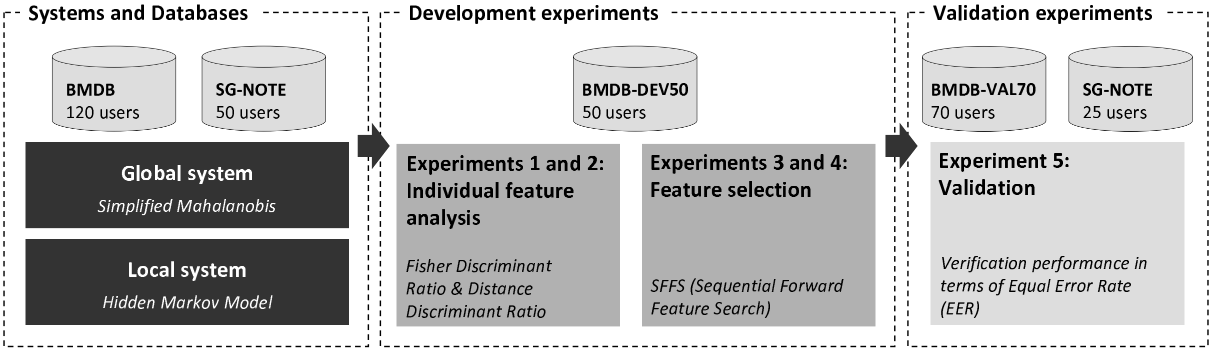

Systems and Databases
BMDB
120 users
SG-NOTE
50 users
Global system
Simplified Mahalanobis
Local system
Hidden Markov Model
Development experiments
BMDB-DEV50
50 users
Experiments 1 and 2: Individual feature analysis
Fisher Discriminant Ratio & Distance Discriminant Ratio
Experiments 3 and 4: Feature selection
SFFS (Sequential Forward Feature Search)
Validation experiments
BMDB-VAL70
70 users
SG-NOTE
25 users
Experiment 5: Validation
Verification performance in terms of Equal Error Rate (EER)

Figure 4.2: *Diagram of the experimental setup followed in this work. The global and local systems are presented in Chapter 3. The experimental protocol and databases are described in Sect. 4.2. Results are reported in Sect. 4.3.*

4.2.2. Development and Validation Experiments

The experiments are structured as follows: first, a global and local individual feature analysis is performed on signatures from the BMDB-DEV50 development dataset (Experiments 1 and 2). Optimal feature combinations are then computed using the SFFS algorithm for feature selection (Experiments 3 and 4). Finally, results are validated using the BMDB-VAL70 and SG-NOTE datasets (Experiment 5).

The experimental approach that has been followed is depicted in Fig. 4.2.

Experiment 1: Global Feature Analysis. The discriminative power of global features can be measured using the Fisher's Discriminant Ratio (FDR) for each individual feature. The FDR provides an intuitive measure of discriminative power, as it increases with the inter-class variability and decreases with the intra-class variability. The FDR D for the i-th feature from user C is computed as follows:

$$D_i(C) = \frac{(\mu_{G_i} - \mu_{F_i})^2}{\sigma^2_{G_i} + \sigma^2_{F_i}} \tag{4.1}$$

where μ and σ are the average and standard deviation respectively of the genuine signature sample set G_i and the forged sample set F_i. We use this measure in this work to compare the discriminative power of each feature defined in Table 3.1 between the mobile and the pen tablet scenario.

Experiment 2: Local Feature Analysis. Contrary to the case of global features, the application of the FDR to compute the discriminative power of individual local features is impractical. This is due to the fact that local features are time functions. As a consequence, the computation of distances between average feature values as defined in the FDR does not represent a realistic measure.

A distance-based discriminative measure using time functions is proposed in (Lei and Govindaraju, 2005) to overcome this limitation. In that work, a consistency value is described, which provides a similar measure to the FDR at least from an intuitive point of view, as it decreases when genuine features are far apart among them and close to forgeries and viceversa. We use the DTW algorithm to compute distances between the time functions, as in (Lei and Govindaraju, 2005). We modify the consistency value definition in order to make its notation similar to the FDR and thus define the Distance Discriminant Ratio (DDR) R for the i-th feature of user C as

$$R_i(C) = \frac{(\mu_{DG_i} - \mu_{DF_i})^2}{\sigma^2_{DG_i} + \sigma^2_{DF_i}}, \tag{4.2}$$

where DG_i is the set of distances among the user genuine signatures and DF_i is the set of distances between the genuine signatures and forgeries. This measure assumes that for each user the mean distance between genuine signatures and forgeries μ_{DF_i} is higher than the mean distance between genuine signatures μ_{DG_i}, which has been tested to be true in the datasets used for experiments. As can be seen, while not being mathematically equivalent to the FDR, the DDR provides a comparable measure in terms of the feature discriminative power. Unlike the FDR, this measure is not scale invariant. However, in our experiments local features are normalized to have zero mean and variance equal to 1.

The median FDR and DDR are computed differently for random and skilled forgeries. In the case of random forgeries, for each user, the FDR and DDR between the user samples and the rest of the genuine signatures in the database are computed, while for skilled forgeries, the FDR and DDR are computed between the genuine signatures and the available skilled forgeries for each user.

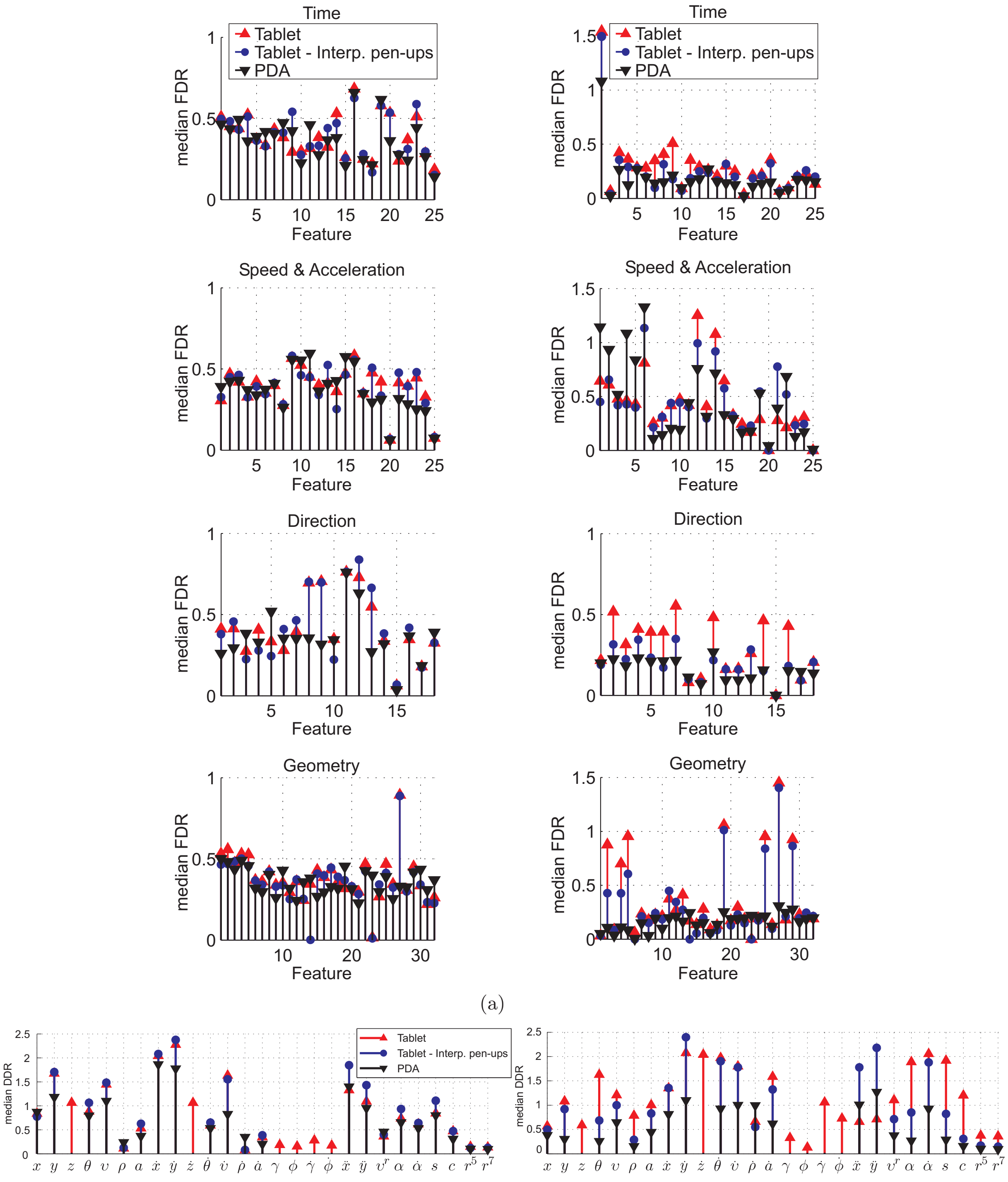
Experiment 3: Feature Selection. In order to select the best performing feature combinations, feature selection on the global 100-feature set and the local 27-feature set is carried out using the Sequential Forward Floating Search (SFFS) algorithm, described in Chapter 2, which is set to minimize the system Equal Error Rate (EER) on the BMDB-DEV50 development dataset.

Experiment 4: Validation. Finally, the verification performance in terms of the Equal Error Rate (EER) using the optimal feature vectors selected by the SFFS algorithm for each scenario are compared on the two available validation sets (BMDB-VAL70 and SG-NOTE).

4.3. Results

4.3.1. Experiments 1 and 2: Individual Feature Analysis

From Fig. 4.3.a, we observe that the median FDR for each feature is similar in the pen tablet and the PDA scenario when random forgeries are considered (left column). Nevertheless, the FDR for pen tablet tends to be always higher or equal than the FDR for PDA. In the case of skilled forgeries, the FDR is higher in most cases for pen tablet than PDA in the case of



(a)

(b)

Figure 4.3: *(a) Fisher's Discriminant Ratio (FDR) of each global feature for random (left) and skilled (right) forgeries. (b) Distance Discriminant Ratio (DDR) of each global feature for random (left) and skilled (right) forgeries.*

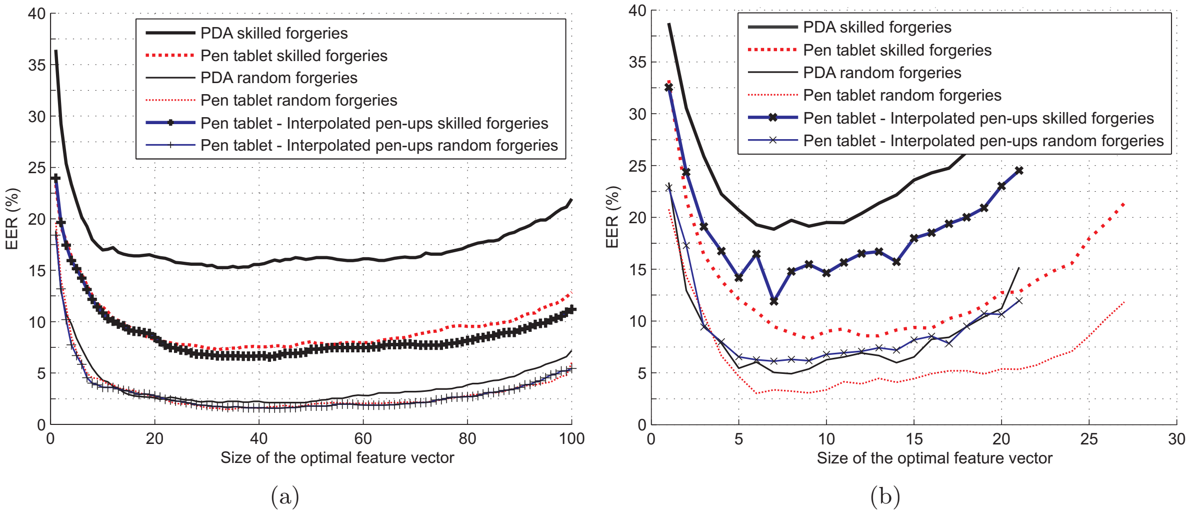


Figure 4.4: *System EER for each possible size of the optimal feature vector as selected by the SFFS algorithm for the global (left) and local (right) system.*

skilled forgeries (right column). This suggests that the verification performance in the PDA scenario against skilled forgeries would be *a priori* lower than for pen tablet independently from the classifier used. Interestingly, the FDR for the interpolated pen-ups tablet subset is in general lower than the original subset, especially for skilled forgeries. This suggests that pen-up trajectories are more resilient to forgeries (i.e. harder to imitate).

The DDR is in general higher for pen tablet than for PDA, independently of the availability of pen-up trajectories (see Fig. 4.3.b). As for global features, when pen-up trajectories are interpolated, the DDR is more negatively affected for skilled forgeries than for random forgeries. In random forgeries, the most relevant difference is observed in the vertical coordinate feature y, which is the one that best characterizes the shape of signatures. The first derivative of y has also a notably higher DDR in the pen tablet scenario. This suggests a higher geometrical variability in the PDA scenario. As can be seen, first and second x, y derivatives are more discriminative when pen-ups are interpolated, which may reflect unstable motion during pen-ups. The path velocity magnitude v and its first derivative are also considerably more discriminative in the pen tablet dataset. This suggests higher variability in the writing speed on the PDA, which can be motivated by the unfamiliar signing surface (touchscreen) and device.

4.3.2. Experiments 3 and 4: Feature Selection

In Fig. 4.4 the evolution of the global and the local system EER using the optimal feature vector, as selected by the SFFS algorithm, is depicted for each possible vector size. It can be observed that while the behavior for the case of random forgeries is similar on both scenarios (mobile and tablet), the optimal verification performance is significantly better for skilled forgeries in the pen tablet scenario.

In the global system, the verification performance for pen tablet does not significantly vary

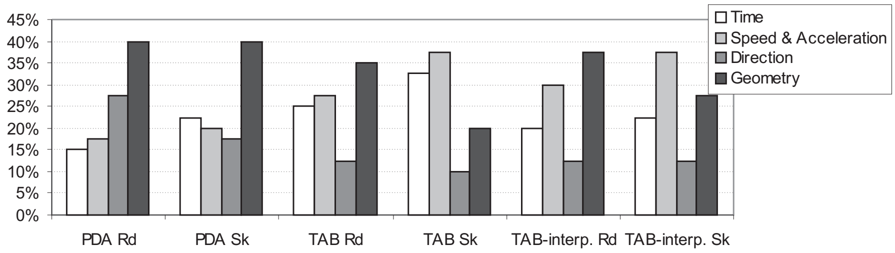


Figure 4.5: *Histogram of global feature types (Time, Speed & Acceleration, Direction, and Geometry) selected by the SFFS algorithm on each optimization scenario using the BMDB-DEV50 subcorpus. Feature vectors of 40 elements are considered. Rd denotes random forgeries, Sk skilled forgeries and "interp." refers to the interpolated pen-ups dataset.*

when pen-up trajectories are interpolated. On the other hand, the EER increases notably in the local system when pen-ups are interpolated. This corroborates the results from the individual feature analysis, that is, trajectories during pen-ups provide considerable discriminative information against skilled forgeries.

Experiment 3: Global Features As can be seen in Fig. 4.4.a, the optimal feature vectors have an approximate size of 40 features. The specific features which conform the optimal 40-feature vectors are shown in Table 4.1. The proportion of each feature type (Time, Speed & Acceleration, Direction and Geometry, as described in Table. 4.1) in each optimization scenario is represented in Fig. 4.5, considering feature vectors of 40 elements. As can be seen, Geometry features have a higher relevance in the PDA dataset. On the contrary, Time and Speed & Acceleration features are more relevant in pen-tablet feature vectors, specially against skilled forgeries. Geometry features are in principle the easiest to forge, so their higher presence in PDA feature vectors may lead to a lower verification performance.

Experiment 4: Local Features The optimal local feature combinations selected by the SFFS algorithm for each optimization scenario are presented in Table 4.2.

Several remarks can be extracted from these results. First, neither pressure nor pen orientation-related features are present in the pen tablet optimal feature vectors, suggesting that the lack of them should not penalize the verification performance (contrary to the results presented by Muramatsu and Matsumoto (2007) and aligned with the findings of Houmani *et al.* (2009)). For the two original datasets (PDA and pen tablet), three features are present in all vectors, namely the x coordinate, the first derivative of the y coordinate and the cosine c of the trajectory angle α.

These results also reveal that less features are needed for HMM-based signature verification compared to the ones commonly considered in other works such as Fierrez *et al.* (2007b); Ly-Van *et al.* (2007); Richiardi *et al.* (2005), at least under these experimental conditions. The absence of pressure in the optimal feature vectors suggests that a pen tablet-based system does not have *a priori* advantage over a handheld device due to the capture of pressure information *per se*.

Table 4.1: *Global feature set described in Chapter 3. The optimal 40-feature subsets, as described in the Experimental Results (Sect. 4.3.2), are shown for each optimization scenario: "Ps" and "Pr" denote PDA skilled and random forgeries, "Ts" and "Tr" pen tablet skilled and random forgeries and "Us" and "Ur" refer to pen tablet with interpolated pen-ups against skilled and random forgeries respectively.*

#	Time related feature	#	Direction related feature
#	Speed and Acceleration related feature	#	Geometry related feature

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

Table 4.2: Local feature sets selected by the SFFS algorithm on the development datasets. "SK." denotes skilled forgeries and "RD." random forgeries.

Optimization scenario		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
		x_n	y_n	z_n	θ_n	υ_n	ρ_n	a_n	$\dot{x}_n$	$\dot{y}_n$	$\dot{z}_n$	$\dot{\theta}_n$	$\dot{\upsilon}_n$	$\dot{\rho}_n$	$\dot{a}_n$	γ_n	ϕ_n	$\dot{\gamma}_n$	$\dot{\phi}_n$	$\ddot{x}_n$	$\ddot{y}_n$	υ_n^r	α_n	$\dot{\alpha}_n$	s_n	c_n	r_n^5	r_n^7
SK.	PDA	✓	✓			✓	✓			✓				✓												✓		
	Pen tablet	✓	✓			✓				✓		✓	✓									✓		✓		✓		
	Pen tablet interp.	✓	✓		✓	✓															✓	✓			✓			
RD.	PDA	✓	✓				✓		✓	✓		✓												✓		✓		
	Pen tablet	✓				✓				✓		✓												✓		✓		
	Pen tablet interp.	✓	✓					✓		✓												✓		✓		✓		

Table 4.3: System performance in terms of EER on the BMDB-VAL70 validation set using global or local features on both scenarios for random (rd) and skilled (sk) forgeries. The combined EER (EER_c) is also presented, as described in Sect. 4.3.3. Vectors of 40 features have been selected in every configuration for the global system.

Optimization scenario		Global			Local		
		$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_c(\%)$	$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_c(\%)$
SK.	PDA	7.2	16.3	9.7	6.0	17.5	9.1
	Pen tablet	5.6	11.3	7.5	4.5	9.3	5.7
	Pen tablet interp.	6.9	10.9	7.9	6.8	12.1	8.1
RD.	PDA	5.4	17.7	9.2	5.8	22.2	9.5
	Pen tablet	6.7	13.0	8.6	3.8	11.1	7.1
	Pen tablet interp.	6.7	10.9	7.7	5.8	15.3	8.9

The main disadvantage of a handheld device would be the lack of trajectories during pen-ups, which penalizes verification performance.

4.3.3. Experiment 5: Validation

The verification performance (in terms of EER) on the BMDB-VAL70 validation set using the optimal feature vectors in each scenario is shown in Table 4.3. As can be seen, global features provide better results in general on mobile conditions, at least compared to an HMM-based system. It can also be observed that when pen-up trajectories are not available, the performance of the local system is significantly degraded against skilled forgeries. This corroborates the reduction of the individual feature discriminative power (FDR and DDR) against skilled forgeries observed in the individual feature analysis (Sect. 4.3.1).

It can also be observed in Table 4.3 that, comparing both optimization scenarios, when the systems are optimized against random forgeries, there is a significant degradation in the performance against skilled forgeries. On the contrary, the EER against random forgeries is nearly not degraded (or even enhanced) when the systems are optimized against skilled forgeries.

A combined EER (EER_c) is also presented in Table 4.3, where all available scores (genuine, random forgeries and skilled forgeries) are used for its computation. This implies that, for each user, 15 genuine user scores, 20 skilled forgery scores and 69 random forgery scores are used for

Table 4.4: *System performance in terms of EER on the SG-NOTE set using global or local features on both scenarios for random (rd) forgeries. Vectors of 40 features have been selected in every configuration for the global system. "SK." denotes skilled forgeries and "RD." random forgeries.*

Optimization scenario	**Global**	**Local**
	$EER_{rd}(\%)$	$EER_{rd}(\%)$
SK.	4.2	6.2
RD.	2.1	6.8

Table 4.5: *System performance in terms of EER in the BSEC 2009 Signature Evaluation Campaign both for random (rd) and skilled (sk) forgeries. Table data has been extracted from Houmani et al. (2012).*

System ID	**DS2 Pen tablet Dataset**		**DS3 PDA Dataset**	
	$EER_{sk}(\%)$	$EER_{rd}(\%)$	$EER_{sk}(\%)$	$EER_{rd}(\%)$
UPM1	4.9	2.3	7.4	1.9
UPM2	4.4	1.9	8.2	2.0
SKU	2.9	1.6	7.9	1.3
ASU	3.8	2.7	31.6	30.6
VDU	**2.2**	1.0	6.6	1.7
SU	3.0	2.2	**5.0**	4.3
UAM-DTWr	4.2	**0.5**	12.2	**0.6**
UAM-DTWs	2.9	1.5	5.8	1.5
UAM-HMM	19.2	24.2	25.8	21.3
UAM-GLO	6.7	3.3	13.2	4.7
UAM-FUS	2.2	0.6	5.5	0.7
Reference	4.5	1.7	11.3	4.8

the (EER_c) computation. It can be observed that in most cases the systems optimized against skilled forgeries present a better overall performance under these experimental conditions.

In Table 4.4, the verification performance in terms of EER against random forgeries is shown for the SG-NOTE validation dataset. As can be seen, the performance is similar than in the BMDB database when the local system is used. On the contrary, the global system verification performance is better than with the BMDB database.

Results of the BSEC 2009 Signature Evaluation Campaign (Houmani *et al.*, 2012) Task 1, are reported in Table 4.5. Performance in terms of EER of the eleven participating systems and a reference system is shown. The BMDB signature corpus was used for the competition, which contains 382 users. As can be seen performance is degraded on mobile conditions. The UAM-GLO system is based on the global system presented in this work, and the UAM-HMM system is based on the local system. Unfortunately, the UAM-HMM system had an implementation error that led to a poor performance in the BSEC 2009 competition. The DTW systems presented by the authors (UAM-DTWs and UAM-DTWr) reached top positions in many categories of BSEC 2009, and were based on the same local features and approach followed in this chapter, using a DTW-based matcher. A detailed report of the results may be found in BSEC (2009).

4.4. Chapter Summary and Conclusions

The effects on the feature discriminative power produced by the usage of handheld devices for signature acquisition have been studied. It has been observed that mobile conditions negatively affect feature discriminative power, specially when local features are considered, at least for the HMM-based system used in the experiments, which is based on the one that reached top positions in the SVC-2004 competition (Kholmatov and Yanikoglu, 2005).

The performance difference against skilled forgeries between the mobile and pen tablet BMDB datasets may also be due to the different forgery acquisition protocols. On the mobile scenario, forgers had access to an on-screen replay of the signature while the replay shown was on a separate screen when using pen-tablet. Nevertheless, it has been clearly seen that verification performance decreases when pen-up samples are not available, except for the case of the global system and skilled forgeries. These results indicate that trajectories during pen-ups contain relevant biometric information, corroborating the findings reported by Sesa-Nogueras *et al.* (2012) in the field of handwriting recognition. The verification performance when using global features presents a more robust behavior than the local approach based on discrete-time functions against the lack of pen-up samples.

It has also been observed that the optimal feature set selected by the SFFS algorithm has a similar performance on the SG-NOTE database in the case of local features, while it presents lower error rates for the global system. The performance of the global system is better while the local system has a similar performance. This corroborates the apparent robustness of global features against degraded signature acquisition conditions, at least in this experimental setup.

At an individual feature level, it has also been observed that on handheld devices the feature discriminative power is more negatively affected for skilled forgeries than for random forgeries. The discriminative power on the mobile scenario is penalized by the lack of pen-up trajectories, the unfamiliar screen surface where users must sign and the poor ergonomics of a handheld device stylus. Features related to pen inclination and pen pressure, not available in this scenario, have not proven to be among the most discriminant in the pen tablet setup, corroborating the observations reported by Houmani *et al.* (2009).

Chapter 5

Aging in Signature Verification

It has not been until recently that different European and national efforts have led to the acquisition of compatible (regarding certain traits) multimodal databases with a relatively large number of common users which have been captured in different sessions over a several year time span. Some examples include the Biosec (Fierrez *et al.*, 2007a), BiosecurID (Fierrez *et al.*, 2010) and BioSecure (Ortega-Garcia *et al.*, 2010) projects. For the current Thesis, the signature modality of the common subset of users in BiosecuriID and BioSecure has been used to generate a new Long-Term dynamic signature dataset. This new dataset has been used to analyze the effect of aging on three state-of-the-art on-line signature verification systems working on totally different features and matchers. In addition to the study of the signature performance stability over time, several template update strategies have also been explored in order to assess their efficiency as a way to maintain the consistency of the system performance in the long-term. Furthermore, several experiments regarding the changes suffered by signatures with time and their most/least robust features have also been carried out.

This way, although some novelty may be found in the algorithms and techniques used in the experiments, the most relevant contributions of the present chapter lie on: *i*) the presentation of the first dataset where the signature of different subjects may be tracked over more than a year; *ii*) the rigorous methodology followed to reach the experimental results, which may be generalized in the future for similar aging studies focused on other biometric traits; *iii*) the experimental findings and practical conclusions extracted from them, which help to shed some light into the difficult problem of handwriting evolution over time.

The chapter is structured as follows. The on-line signature Long-Term DB used in the experiments is presented in Sect. 5.1. The experimental protocol followed is described in Sect. 5.2, while results are given in Sect. 5.3. Limitations of this study and open questions are discussed in Sect. 5.4. Conclusions are finally summarized in Sect. 5.5.

The findings of this chapter have been published by the author with Galbally *et al.* (2013).

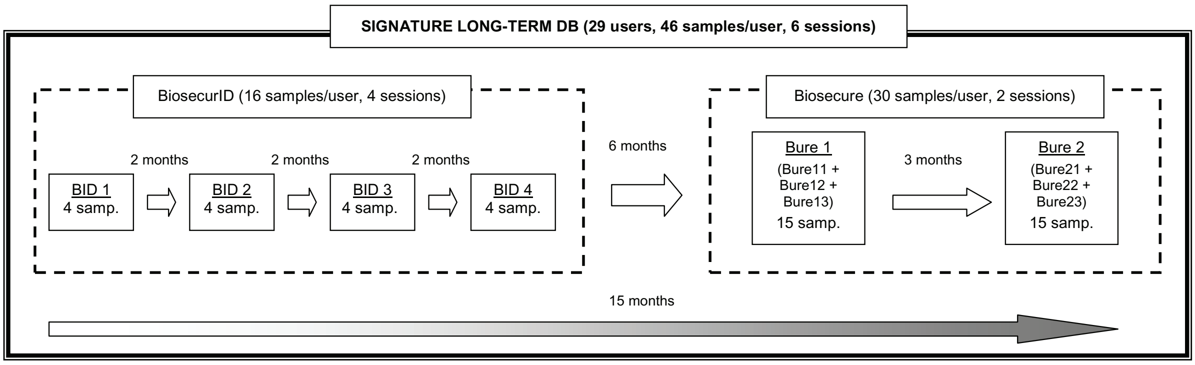

SIGNATURE LONG-TERM DB (29 users, 46 samples/user, 6 sessions)
BiosecurID (16 samples/user, 4 sessions)
2 months
2 months
2 months
BID 1 4 samp.
BID 2 4 samp.
BID 3 4 samp.
BID 4 4 samp.
6 months
Biosecure (30 samples/user, 2 sessions)
Bure 1 (Bure11 + Bure12 + Bure13) 15 samp.
3 months
Bure 2 (Bure21 + Bure22 + Bure23) 15 samp.
15 months

Figure 5.1: *General time diagram of the different acquisition sessions that conform the Signature Long-Term Database.*

5.1. The On-Line Signature Long-Term Database

The dataset used in the experimental section of this work comprises the on-line signature data of the 29 common users to the BiosecurID and the BioSecure databases. These two signature subsets, which were acquired in a 15 month time span, present some unique features that make them especially suited for the aging evaluation performed in the present work. A description of the databases can be found in Sect. 2.4

- **The BiosecurID Signature Subset** (Fierrez *et al.*, 2010). As described in Sect. 2.4, it comprises 16 genuine signatures and 12 skilled forgeries per user, captured in 4 separate acquisition sessions (named here *BID1*, *BID2*, *BID3* and *BID4*). A two month interval was left between capture sessions, and signatures were acquired in a controlled and supervised office-like scenario.

- **The BioSecure Signature Subset** (Ortega-Garcia *et al.*, 2010). This dataset was captured 6 months after the BiosecurID acquisition campaign had finished (the time sequence of the two databases is shown in Fig. 5.1). As described in Sect. 2.4, it comprises 30 genuine signatures per user, and 20 skilled forgeries, distributed in two acquisition sessions separated three months (named here *Bure1* and *Bure2*). The 15 original samples corresponding to each session were captured in three groups of 5 consecutive signatures with an interval of around 15 minutes between groups (named here *Bure11-12-13* and *Bure21-22-23*, respectively). The signature dataset was designed to be fully compatible with BiosecurID.

For the final dataset used in the present work, only the genuine signatures were considered. This way, it comprises 1,334 signatures coming from the 29 common users of the two databases with 46 samples per user (16 from BiosecurID, and the remaining 30 from BioSecure) which are distributed in 6 sessions (BID1-2-3-4 and Bure1-2) according to the general time diagram shown in Fig. 5.1.

It constitutes the first signature dataset where we can track, over a 15 month time span (as there are 6 almost uniformly distributed acquisition sessions in this interval), the signature of a given user, and assess if that period of time is sufficient to detect a decrease in the verification performance of signature-based biometric systems. Furthermore, as all the samples of the same subject have been acquired under almost identical conditions we may discard external factors as the cause of a possible degradation in the recognition rates.

All users in the database are Spanish, white Caucasian with higher level education, between 18 and 51 years of age. In particular, the age distribution of the subjects is: 24 donors between 18 and 25; 3 donors between 25 and 45; and 2 donors above 45 years old. The gender distribution within the database is quite balanced with 11 women and 18 men.

It should also be noted that all the users included in the database may be considered as adults in terms of writing. This means that their signature is a well learned sequence of movements which may be considered as permanent and that has already gone through the transitional learning period which usually happens in the youth. The effect of aging during the time in which the signature has not yet been fully fixed may be different and would be the subject of future work.

Some typical examples of the signatures that can be found in the different sessions comprised in the Signature Long-Term DB are shown in Fig. 5.2. The Signature Long-Term DB is publicly available for research purposes[1].

5.2. Experimental Protocol

The experimental framework has been designed to evaluate the effect of aging on the performance of signature-based systems and to assess the stability of signatures through time. In particular, five different objectives are pursued in the experiments, which may be divided into two main groups:

- **Signature recognition performance**. On the one hand, *i*) to evaluate the loss of performance of different competitive signature recognition systems as a consequence of the changes suffered by the signature trait with time (i.e., aging); *ii*) to determine the dependencies of this performance degradation (e.g., signature-dependent *vs.* user-dependent); and *iii*) to assess the effectiveness of different template update approaches to compensate this effect.

- **Signature evolution**. On the other hand, *iv*) to determine which are the changes over time that motivate the previously evaluated decrease in the signature recognition performance; and *v*) to establish which are the most stable features in the signature trait.

In order to achieve these goals the experimental protocol includes two groups of tests which are described in the next sections.

[1] Available at: http://atvs.ii.uam.es/databases.jsp

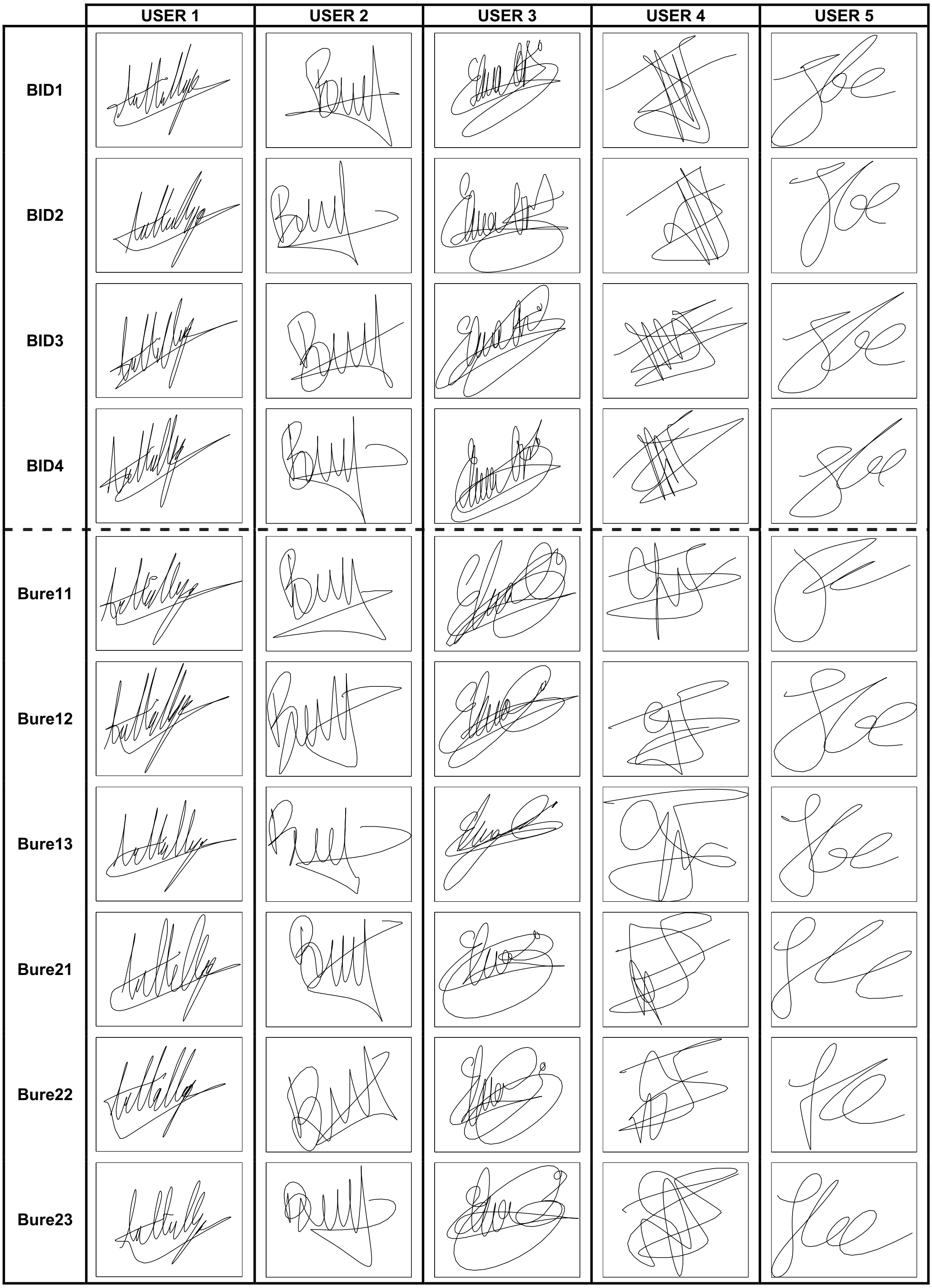


Figure 5.2: *Typical samples that can be found in the Signature Long-Term DB. Each signature corresponds to each of the acquisition sessions of five different users.*

5.2.1. Signature Recognition Performance Experiments

The first objective of this group of experiments is to evaluate the degree of aging that may be observed in the recognition performance of signature-based systems. The results will also shed some light on the user- and signature-dependency of aging, that is, if certain type of signatures are more prone to worsen their performance in the long term, or if this only depends on the signer (second objective).

The third objective of these tests is to analyze different template update approaches that can help to reduce the performance deterioration that signature recognition systems suffer with time.

In order to reach these goals, several sets of genuine matching scores (i.e., those computed between samples of the same user and therefore affected by aging) are computed on the Signature Long-Term DB simulating two different scenarios:

- **Aging experiments: Fixed template and varying test**. In this case the user models enrolled to the system are always computed using the same samples (i.e., those belonging to the first session of the Signature Long-Term DB, BID1), while the test signatures are taken from the following sessions (BID2-3-4 and Bure1-2).

- **Template update experiments: Varying template and fixed test**. In this case the test samples are always taken from session Bure13, while the enrolled models are updated with signatures coming from different previous sessions (BID2-3-4 and Bure11-12).

Not all the systems working on a given trait may be necessarily affected in the same way by aging. In order to account for possible differences, we have carried out this set of experiments on three different competitive on-line signature verification systems using totally diverse feature sets (feature- and function-based) and matchers (Mahalanobis distance, Hidden Markov Models, and Dynamic Time Warping, presented in Chapter 3). A brief description of each of the three systems is given next:

- **System A: function-based + HMM**. An HMM system, as described in Chapter 3, is used. A subset of 12 discrete-time signals are derived from the coordinate (x and y) and pressure (p) functions, while no pen inclination signals are used as its utility for automatic signature recognition is at least unclear (Martinez-Diaz *et al.*, 2014). This subset corresponds to the first 12 features described in Table 3.2. After some preprocessing (position and rotation alignment) and the computation of the 12 functions, similarities are computed using a 12-state left-to-right HMM with 4 Gaussian mixture components per state, which is a common implementation in the literature (Dolfing *et al.*, 1998; Ly-Van *et al.*, 2007).

- **System B: feature-based + Mahalanobis distance**. This system models the signature as a holistic multidimensional vector composed of 100 global features described in Table 3.1.

	Features #
Static	2,7,8,12,15-19,24,27-28,30,34-37,43,46,51,53-57,61,63,65-67, 70-73,75,77-78,84,86,93,95,97-99.
Dynamic	1,3-6,9-11,13-14,20-23,25-26,29,31-33,38-42,44-45,47-48,50,52,58-60, 62,64,68-69,74,76,79-83,85,87-92,94,96,100.

Table 5.1: *Division of the feature set introduced in Table 3.1 according to the type of information they contain.*

The similarity scores are computed using the simplified Mahalanobis distance method described in Sect. 3.2.

- **System C: function-based + DTW**. The Dynamic Time Warping system described in Sect. 3.3 is used. A set of 9 functions is extracted from the signatures, namely features 1, 2, 5, 9, 11, 12, 21 23 and 25 from Table 3.2. These functions correspond to the best performing set of features against skilled forgeries on the training set of the BioSecure Signature Evaluation Campaign 2009. It outperformed other systems based on HMMs and global features (Houmani *et al.*, 2012)

5.2.2. Signature Evolution Experiments

In this case, the aim of the experiments is to give some indication on whether there is a common trend in the evolution through time of signatures coming from different users (objective four), and if there are certain types of features (e.g., static *vs* dynamic) which are more stable (objective five).

To reach these objectives, the Signature Long-Term DB is parameterized using the set of 100 features described in Table 3.1. This 100-feature set may be divided into two classes according to the information contained by each of the parameters, namely: static or dynamic. All the features assigned to each class are specified in Table 5.1 (the numbering criterion is the same used in Table 3.1).

5.3. Results

The results obtained for the two sets of experiments described in Sect. 5.2 are presented in the next sections.

5.3.1. Signature Recognition Performance Results

As already mentioned, aging may be defined as the loss of performance experimented by biometric systems due to the transformations suffered by biometric traits in the long term. With this in mind, the questions raised in this section are: Is aging present in the signature

	Aging Experiments	
	Enrollment	Test
Exp. A	BID1 (4 sig.)	BID2 (4 sig.)
Exp. B	BID1 (4 sig.)	BID3 (4 sig.)
Exp. C	BID1 (4 sig.)	BID4 (4 sig.)
Exp. D	BID1 (4 sig.)	Bure1 (15 sig.)
Exp. E	BID1 (4 sig.)	Bure2 (15 sig.)

Table 5.2: *Enrollment and test signatures used to compute the genuine scores in the aging experiments.*

trait? To what extent? Are some users more prone to be affected by aging than others? How can it be corrected?

In order to give an answer to these questions, several sets of genuine scores (i.e., those affected by aging) are computed in order to evaluate the performance of signature recognition systems.

Before presenting the results, it is very important to notice that, given a fixed set of impostor scores, the best possible performance results are reached when the genuine similarity score distributions have a mean value as high as possible and a variance as low as possible. Therefore, a worsening of the systems performance with time (i.e., aging) may be caused by two factors: *i*) a decrease of the genuine distributions mean value, or *ii*) an increase of the genuine distributions variance.

5.3.1.1. Objective 1: Aging analysis

As mentioned before, these experiments are aimed at estimating the impact of aging on signature recognition systems. For this purpose, the enrolled models of the 29 users present in the Signature Long-Term DB are trained using the 4 signatures corresponding to the first session (BID1). Then, the sets of genuine and impostor scores are computed as follows:

- Genuine scores are generated matching the models against the signatures of the following 5 sessions: BID2-3-4 and Bure1-2. This way, for each user 5 different sets of genuine scores are computed: BID1 vs BID2, BID1 vs BID3, BID1 vs BID4, BID1 vs Bure1, and BID1 vs Bure2 (see Table 5.2).

- On the other hand, the same set of impostor scores is used for all the experiments A-E (i.e., we assume impostor signatures may come from any of the acquisition sessions as they are not affected by aging). To compute the set of impostor scores one signature from each session of the rest of the users is matched against the enrolled model of the subject at hand, leading this way to a total of $29 \times 6 \times 28 = 4,872$ impostor scores.

As the impostor score distribution is fixed for all the scenarios, any changes observed in the performance of signature recognition systems among experiments A-E will be caused by changes in the genuine score distributions.

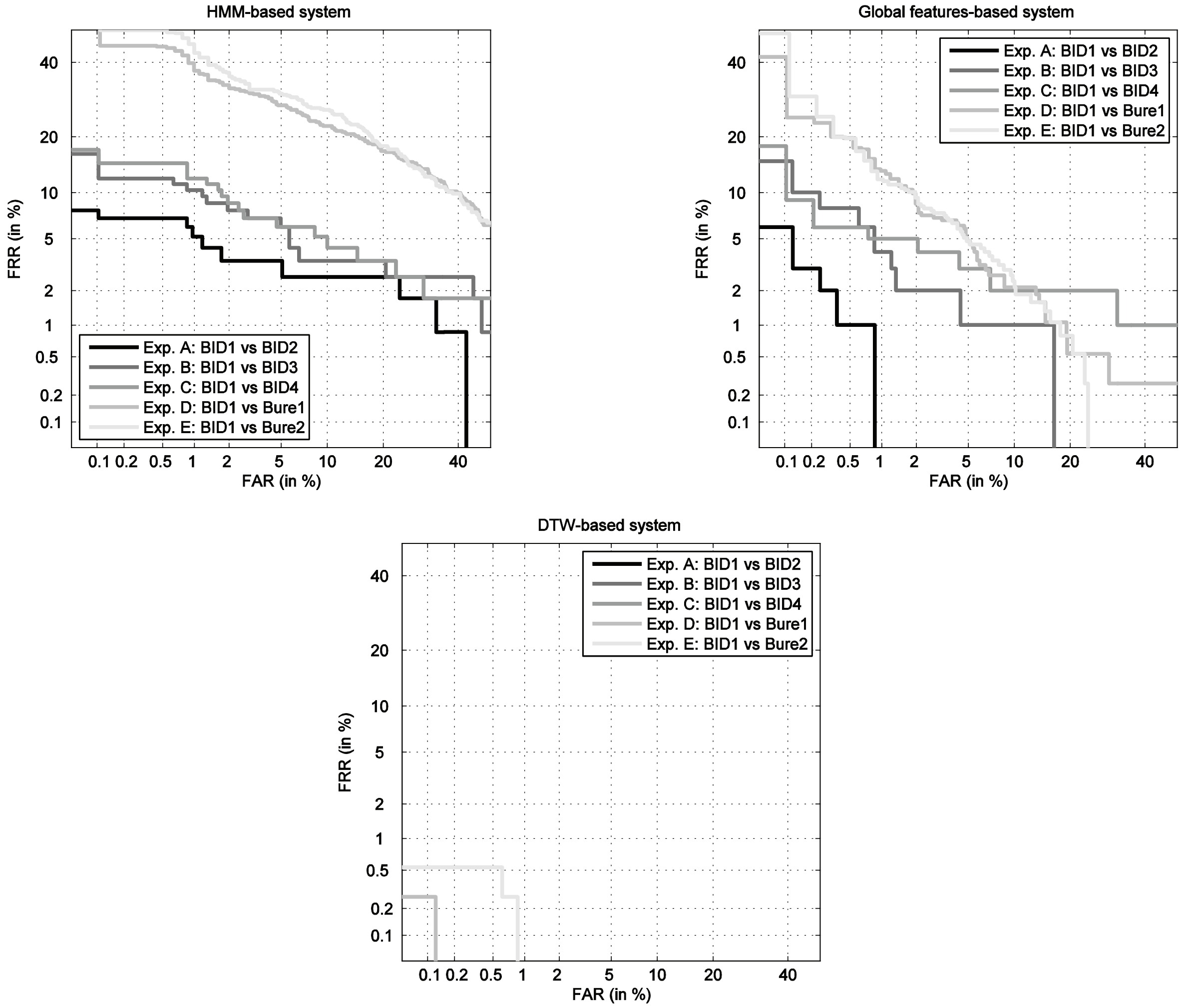


Figure 5.3: *Performance evolution of the three signature recognition systems considered in the experiments. For the DTW-based system only two curves appear as for experiments A-C its EER is close to zero. The EER for the three systems and for the different experiments are reported in Table 5.3.*

The DET (Detection Error Trade-off) curves obtained with the aforementioned genuine and impostor scores for the five scenarios (A-E) and for the three recognition systems are shown in Fig. 5.3. A darker gray level corresponds to a better performance of the evaluated system. It may be observed that, as the test signatures are more distant in time from those samples used for enrollment, the performance of all the three systems drops. For completion, the Equal Error Rate (EER) corresponding to the curves shown in Fig. 5.3 is given in Table 5.3.

In order to further analyze this performance loss, in Fig. 5.4 we show the evolution of the genuine scores when the test signatures move away (in terms of time) from the model. The distributions for each of the five sets of genuine scores are depicted on the right planes (in vertical) with a darker gray representing a better performance. On the left planes we can see the mean (circles) and variance (vertical lines) for each of the five distributions. Several observations can be extracted from the results shown in Figs. 5.3 and 5.4:

	Aging Experiments - **EER** (%)				
	Exp. A	Exp. B	Exp. C	Exp. D	Exp. E
HMM-based	3.2	5.5	5.6	22.7	27.8
GF-based	1.0	2.0	4.2	4.9	5.0
DTW-based	0.0	0.0	0.0	0.1	0.5

Table 5.3: *EER for the aging experiments defined in Table 5.2. The whole DET curves for these experiments are shown in Fig. 5.3.*

- The performance of the three systems consistently decreases as the testing signatures move away from the model (the DET curves in Fig. 5.3 are further away from the origin), which means that the users discriminant power decreases with time or, in other words, that all the three recognition approaches are affected by aging. The previous observation indicates that this effect is not particular of a certain signature recognition technology, but that, as expected, it is inherent to the signature trait itself.

- Not all the systems are affected in the same way by the passing of time, that is, not all the curves in Fig. 5.4 present the same decreasing slope. In particular, the system based on DTW presents a decrease in the average genuine score between the first and the last test set of signatures of 5.6%, compared to a 16.7% of the one based on global features and a 21.8% for the HMM. Thus, we may conclude that the signature recognition technology based on DTW is not only more accurate but also more robust to aging.

- The effect of aging may also be observed in the worsening of the scores variance through time, that is, the scores are not only lower but also more disperse. This way we can see how the variance increases around 45% from experiment A to E for all the three technologies tested.

- Another important observation to be made from the results shown in Fig. 5.4 is that the effect of aging on the signature trait is not negligible. There is a significant drift in the genuine score distributions (from the first to the last signature test set) in a relatively short period of time (15 months).

5.3.1.2. Objective 2: Aging user-dependency analysis

The sets of genuine scores generated in the previous experiments (Sect. 5.3.1.1) are used here to determine if certain users are more prone to suffer from aging. For this purpose we compute an Aging Coefficient (AC) defined as: $AC = \Delta\mu \cdot \Delta\sigma$, where $\Delta\mu$ and $\Delta\sigma$ are respectively the mean and variance relative variation between two sets of scores. This way both aging effects (i.e., decrease of the genuine scores mean value and increase of the variance) are taken into account in one metric, so that the higher the AC of a user, the more affected that subject's signature is by the elapse of time.

The AC is computed for all the users in the database between the genuine scores of experiments A and E, which are the two score distributions more separated in time. In Fig. 5.5 the AC is shown for all the subjects ordered according to their level of aging, from the lowest to the highest, for all the three systems considered in the experiments. Please note that the least affected user, the most affected user, or any of the users in between, do not necessarily have to coincide (i.e., be the same signer) for all three systems. The three AC curves are shown on the same figure for an easier visual comparison across systems.

The five most/least affected subjects by aging (i.e., those with respectively a higher/lower AC) are shown in Table 5.4 for all the three systems tested. For completion, the individual mean and variance variation indexes (i.e., $\Delta\mu$ and $\Delta\sigma$) are also given.

Different observations may be extracted from the results shown in Fig. 5.5 and Table 5.4:

- As expected, not all the systems present the same AC values. The DTW-based system has the lowest values (i.e., most consistent system over time), compared to the one based on global features (GF-based) and the HMM. This is consistent with the results obtained in Sect. 5.3.1.1 and confirms that the AC is a valid metric to evaluate the level of aging.

- In all the three systems there is a very big difference (around 95% on average) between the AC of the least and most affected users. Thus, even for the most robust technologies (DTW), the degree of aging is very dependent on the signer.

- In general the users tend to perform consistently well (3, 19) or badly (1, 17, 11) regardless of the recognition system used. Furthermore, none of the top five users in a system (i.e., those least affected by aging) appear in the list of the worst five users of the other two systems, and vice versa. This means that, as a general rule, a subject that despite of the aging effect presents high recognition rates on a given system, will be very likely to be consistently recognized if the system is changed.

Therefore, we may conclude that, although some technologies are more robust than others to aging, the degree of deterioration of a subject's signature depends mainly on the subject and not on the recognition system being used.

Those subjects with the highest number of appearances in the AC rows of Table 5.4 (shown in bold) are considered to be those with a more/less stable signature. The signatures of these users are depicted in Fig. 5.6 where we can see that the complexity of the signature is not a key factor in the level of aging. That is, complex signatures (i.e., long signatures, with the written name and flourish) may be very affected by aging or, on the contrary, can also be very stable through time. The same happens for short and simple signatures. In other words, these initial results suggest that the degree of aging does not depend on the type of signature, but on the signer. However, these findings regarding aging and signature complexity should be further addressed on a specific database where signatures are classified into different complexity groups.

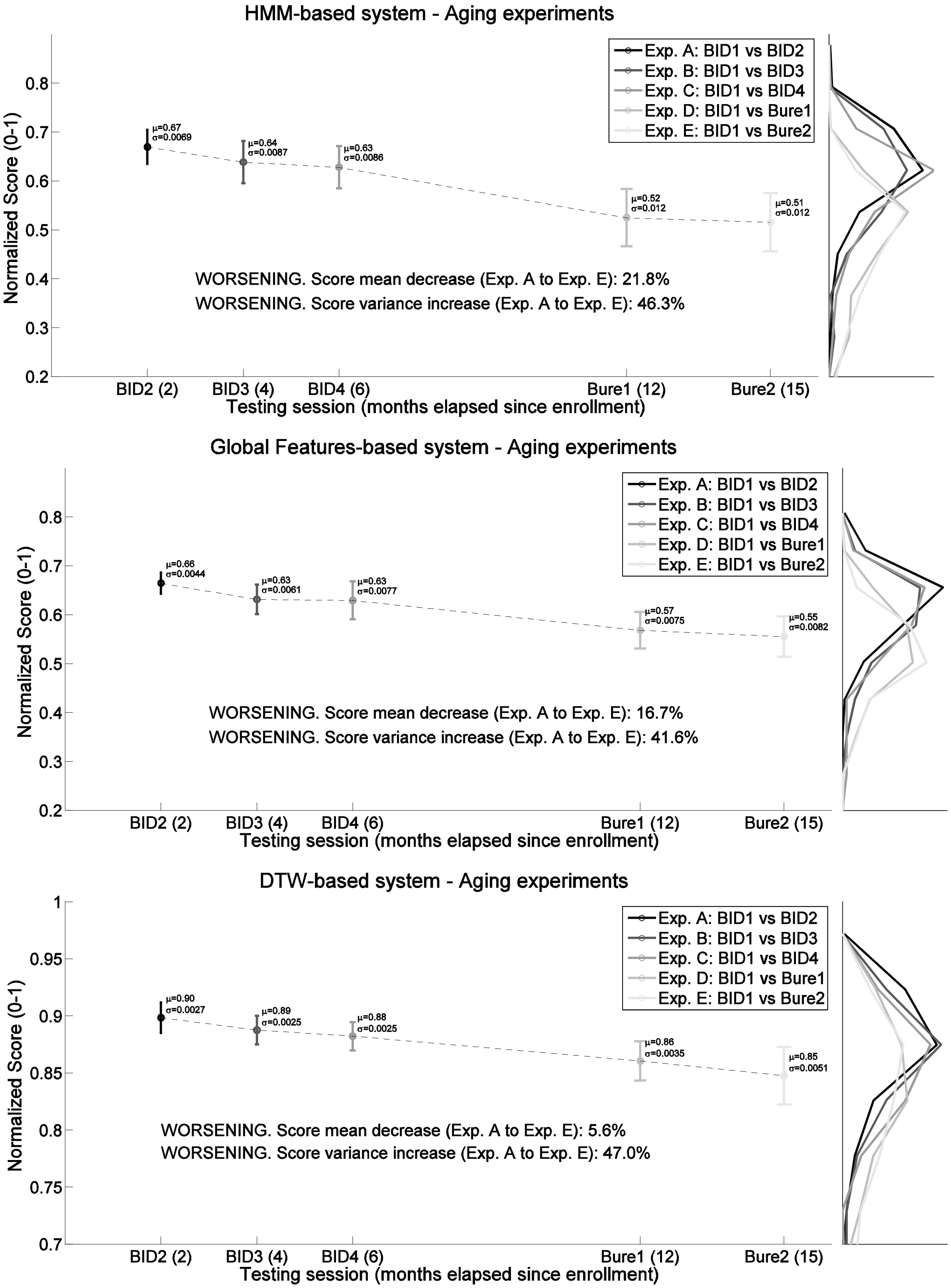


Figure 5.4: *Evolution through time of the mean (circles) and variance (vertical lines) of the genuine score distributions (in vertical on the right) for the three systems considered in experiments A-E. A darker gray level represents a better performance of the given system.*

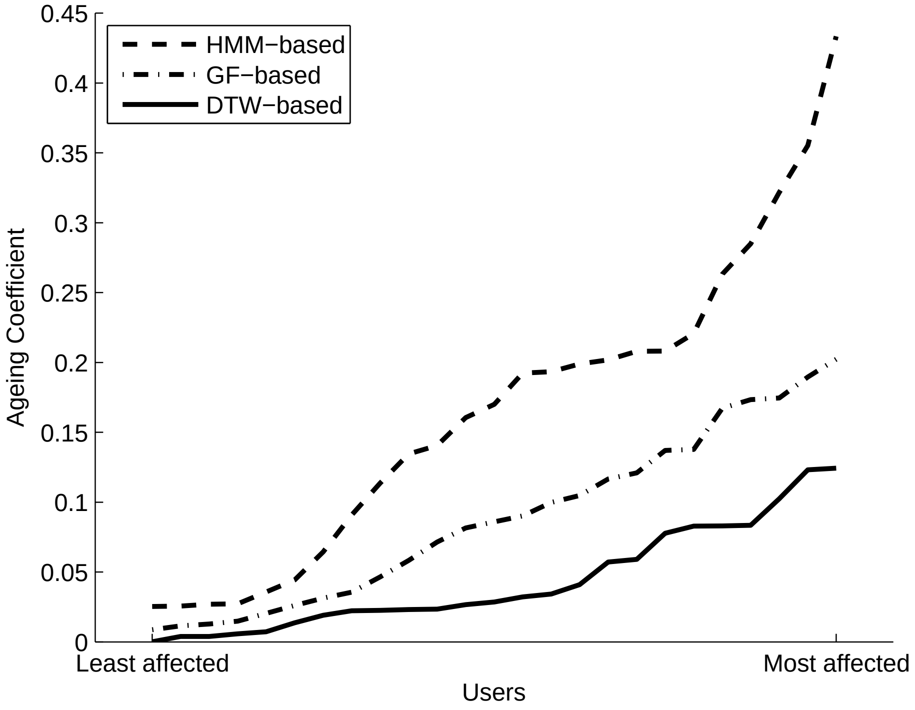


__Figure 5.5:__ Aging Coefficient (AC) from the least affected to the most affected user by aging in the Signature Long-Term DB, for the three systems considered in the experiments. Please note that the least affected user, the most affected user, or any of the users in between, do not necessarily have to coincide (i.e., be the same signer) for all three systems. The three AC curves are shown on the same figure for an easier visual comparison across systems.

		Aging: user dependency	
		Most affected users	Least affected users
HMM	$\Delta\mu$	15, 17, 16, 22, 4	19, 27, 3, 9, 28
	$\Delta\sigma$	17, 4, 5, 26, 12	28, 6, 1, 9, 3
	AC	17, 4, 5, 22, **11**	28, **3**, 6, 27, **19**
GF	$\Delta\mu$	16, 24, 11, 23	19, 21, 3, 2, 27
	$\Delta\sigma$	14, 1, 21, 6, 9	18, 12, 16, 17, 13
	AC	**1**, 24, **7**, **11**, 21	18, 12, 21, **19**, **3**
DTW	$\Delta\mu$	7, 16, 11, 1, 8	19, 13, 3, 14, 26
	$\Delta\sigma$	11, 9, 16, 14, 2	19, 24, 26, 8, 29
	AC	16, **11**, **1**, 18, **7**	**19**, 13, 26, **3**, 24

__Table 5.4:__ Most and least affected users by aging in the Signature Long-Term DB according to the three systems considered in the experiments. Users with the most appearances in the AC rows (in bold) are depicted in Fig. 5.6.

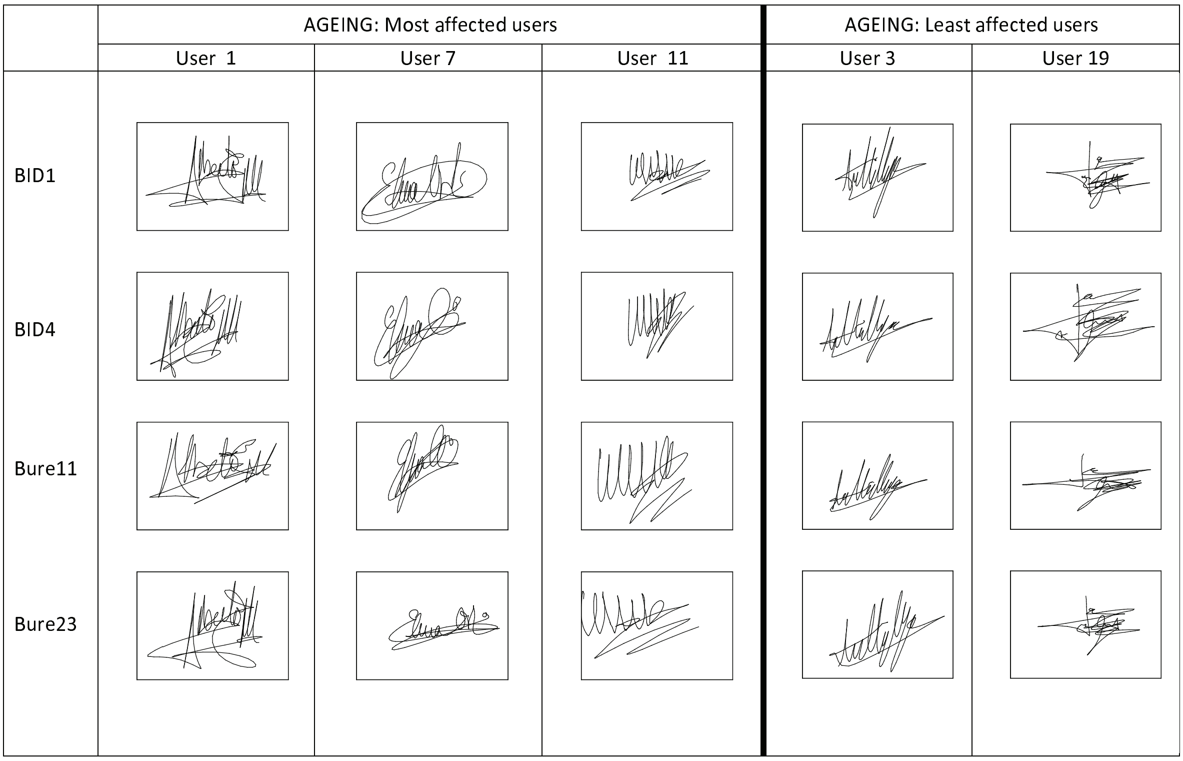


Figure 5.6: Most (left) and least (right) affected users by aging in the Signature Long-Term DB according to Table 5.4.

5.3.1.3. Objective 3: Template update analysis

The results presented in Sects. 5.3.1.1 and 5.3.1.2 confirm the necessity to develop strategies that can help to minimize the effect of aging, especially in those behavioral or learned traits, such as the signature, which are more sensitive to time. Here, we analyze the efficiency of different template update approaches varying the enrollment signatures used to compute the users models and testing always with the same set of samples, as shown in Table 5.5. In particular, the scenarios considered are:

- Baseline result (Exp. F). This represents the scenario with no template update strategies to correct aging. There is a 14 month difference between the enrolled model (BID1) and the test set (Bure13).
- Complete update (Exp. G). The first template update approach considered is to discard the old enrollment samples (BID1) and replace them by new samples acquired very close in time to the test set (Bure11).
- Mixed update (Exp. H). In this case we do not discard the old samples but we update the enrolled model with newly acquired samples (BID1+Bure11). Thus, in this scenario there will be more available data to train the model than in the previous two cases (experiments

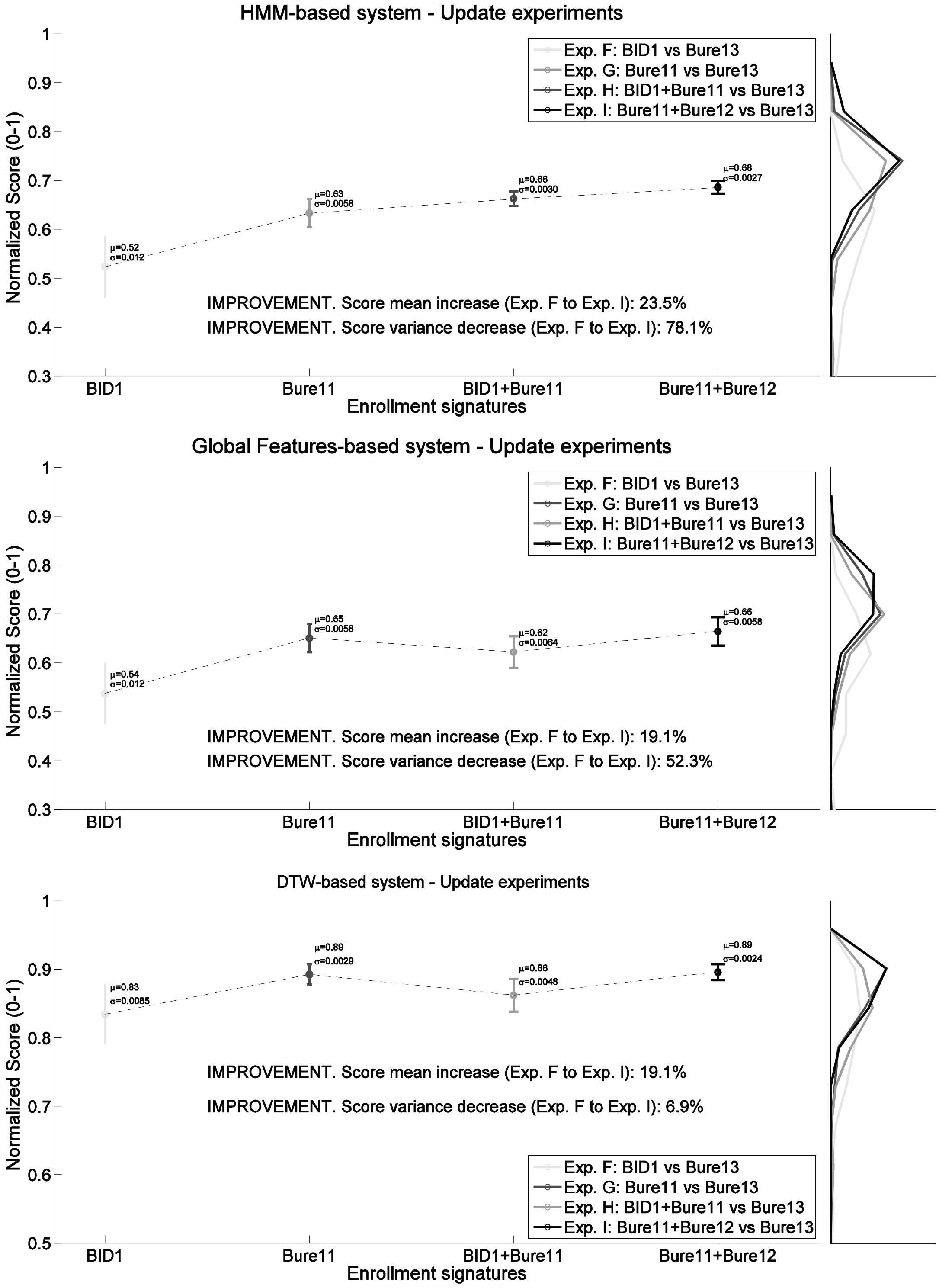


Figure 5.7: *Mean (circles) and variance (vertical lines) of the genuine score distributions (in vertical on the right) for the 4 different template update strategies tested and for the three systems considered in the experiments. A darker gray shade represents a better performance of the given system.*

	Template Update Experiments	
	Enrollment	Test
Exp. F (baseline)	BID1 (4 sig.)	Bure13
Exp. G (complete)	Bure11 (4 sig.)	Bure13
Exp. H (mixed)	BID1 (4 sig.) + Bure11 (4 sig.)	Bure13
Exp. I (complete)	Bure11 (4 sig.) + Bure12 (4 sig.)	Bure13

Table 5.5: *Enrollment and test signatures used to compute the genuine scores in the template update experiments.*

F and G).

- Complete update (Exp. I). Here, we consider the same amount of training data as in experiment H, but all of it comes from recent acquisitions (Bure11+Bure12).

The results of the previously described setups for the three considered systems are shown in Fig. 5.7. As in the case of the aging experiments the score distributions for each of the four considered scenarios is shown on the right planes in vertical with a darker gray shade representing a better performance of the given system. On the left plane we can see the evolution of the mean (circles) and variance (vertical lines) of the score distributions. Although all the template update strategies studied improve the performance with respect to the baseline experiment (in all cases there is an increase of the mean value and a decrease of the variance), two different behaviors may be observed in Fig. 5.7 depending on the signature recognition system considered:

- HMM system. HMM-based systems heavily depend on the amount of training data available (Fierrez *et al.*, 2007b). As a consequence, it is better to perform a mixed update (i.e., do not discard the old samples, exp. H) so that the model is trained with as many signatures as possible (8 signatures, in this particular case), instead of using few recent samples (i.e., exp. G, where only 4 signatures are used for enrollment).

- Global features and DTW systems. On the other hand, the systems based on DTW and global features do not rely as much on the amount of enrollment data, but on the quality of these data (Martinez-Diaz *et al.*, 2009b). Therefore, the performance reached using 4 recently acquired samples (exp. G) is almost the same as the one obtained using 8 of those signatures (exp. I). This means that, as can be seen in Fig. 5.7, in these cases it is preferable to perform a complete update with the most recent samples (i.e., exp. G) than to keep the old ones (i.e., exp. H) even if this means training the enrolled model with a smaller number of signatures.

As could be expected, in all cases the best possible template update strategy is to use for enrollment all the most recent samples available (i.e., exp. I). However, this may represent a somewhat unrealistic scenario, as we are assuming that we have access to as many as 8 signatures captured in a time period very close to the test set. The amount of new collected data will rarely comply with this condition.

5.3.2. Signature Evolution Results

The results presented in Sect. 5.3.1 clearly show that the effect of aging is patent in the signature trait. The purpose of the present set of experiments is to further investigate the causes of the deterioration in the performance of signature recognition systems.

From a human perspective, the changes experienced with age by certain biometric traits are easily distinguished. For instance, we know that the face gradually loses its oval shape and that the wrinkles and sun-stains make its texture less smooth (in fact, these characteristics are successfully used for automatic age estimation purposes). However, what are the changes and transformations, if any, undergone by signatures with age?

In order to shed some light on this difficult question, the aging-related issues raised in this section are: How do signatures typically evolve over time? What type of transformations do they suffer? Are some signature-defining features more stable over time than others?

5.3.2.1. Objective 4: Signature evolution analysis

In order to determine the way in which signatures typically evolve with time, five of the most representative global features given in Table 3.1 have been analyzed for the whole Signature Long-Term DB. Not all the features proposed in Table 3.1 have a direct physical meaning, thus, the selected parameters have been those with an easy interpretation, namely: duration of the signatures (parameter 1 in Table 3.1), number of maxima points in x (parameter 8) and y (parameter 12), number of pen-ups (parameter 2) and the average speed (parameter 26).

These parameters have been averaged for all the users in the database in a sample by sample basis. That is, in the end, for each of the features, a 46-dimensional vector is computed where each element is the result of averaging the value of that parameter for the corresponding sample (from 1 to 46) of all the users in the database. In that way, we can see the evolution of the feature value from the first acquisition (month 0) to the last one (month 15). The results are shown in Fig. 5.8.

We can observe that, regardless of the user, the general trend for the signatures is to become: shorter, with fewer singular points and penups, and faster. That is, the results imply that signatures tend to be simplified with time.

5.3.2.2. Objective 5: Parameter evolution analysis

In this case the goal is to determine which of the global features proposed in Table 3.1 are more stable through time and, on the contrary, which are those that suffer the largest variations in the long term. For this purpose we use a Variation Coefficient (VC) analogue to the Aging Coefficient (AC) computed in Sect. 5.3.1.2. This new Variation Coefficient is defined

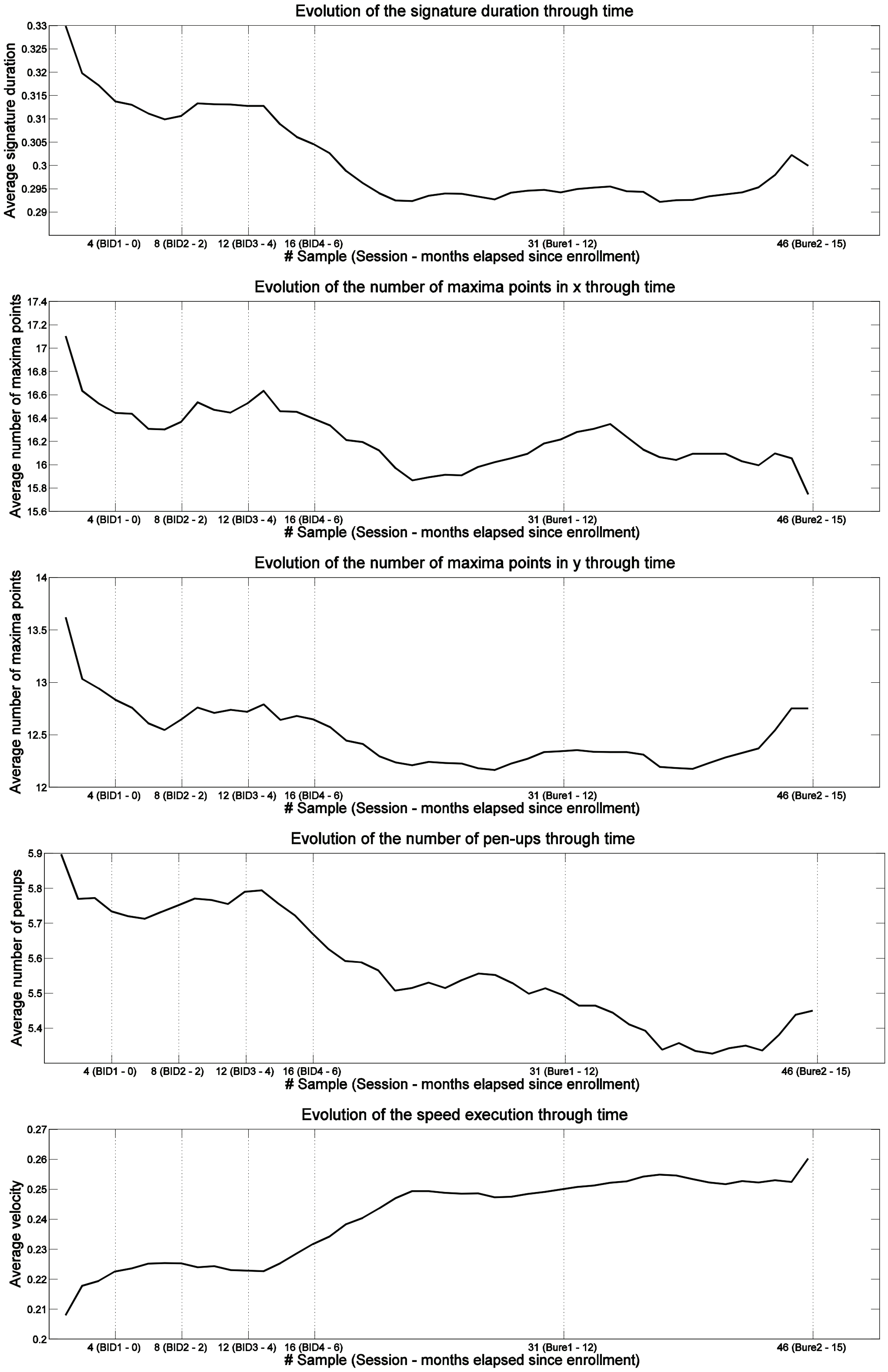


Figure 5.8: *Evolution through time of the duration, maxima points in x, maxima points in y, number of penups and speed of the signatures in the Signature Long-Term Database.*

as: VC = $\Delta\mu_{\text{gf}} \cdot \Delta\sigma_{\text{gf}}$, where $\Delta\mu_{\text{gf}}$ and $\Delta\sigma_{\text{gf}}$ are respectively the mean and variance relative variation of a certain global feature between two acquisition sessions.

Prior to compute the VC, the values of the global features are averaged for all the users in the database on a sample by sample basis. That is, for each sample (1-46) we compute a 100-dimensional vector where each dimension is the mean value of that global feature for all the users in the dataset. Then, in order to evaluate the degree of variation through time of each global feature, the VC is computed between the samples of acquisition sessions BID1 and Bure2, which are the two most distant in time.

In Fig. 5.9 we show the value of the Variation Coefficient from the least variable to the most variable static and dynamic features. On the other hand, in Table 5.6 the 10 most and least variable features are shown following the numbering criterion used in Table 3.1. The 'S' and 'D' stand for Static and Dynamic features respectively, according to the classification given in Table 5.1.

In Table 5.6 we can see that 9 out of the total 10 most unstable features correspond to parameters measuring dynamic information. Furthermore, Fig. 5.9 shows how, in general, dynamic features present a higher variability with time. From these results it may be concluded that the static information of a signature (e.g., geometric, spatial, or angular) is more robust over time than the dynamic data (e.g., velocity or acceleration). In other words, with time, signers tend to be more consistent repeating the shape of their signature rather than the way in which this shape is produced. These results are in line with the findings of previous related studies (Dixon *et al.*, 1993; Houmani *et al.*, 2009; Walton, 1997).

5.4. Limitations of the Study and Open Questions

The main limitations of the present study are derived from the characteristics of the database used in the experiments. It has been mentioned in the chapter that the On-Line Signature Long-Term DB is unique regarding the number of subjects whose signature has been uniformly tracked over more than a year. Nevertheless, although this was the best available possibility, it is still limited both in terms of individuals (29) and time span considered (15 months).

The present work sets a first landmark in the understanding of aging in a behavioral biometric. Its conclusions should be confirmed by further analysis and assessment on databases comprising a big number of uniformly-acquired samples for a larger number of individuals (several hundreds) and over a longer period of time (several years). However, we do believe that the experimental protocol and posterior analysis carried out in the present work is general and may serve as a baseline to be applied in future studies.

Therefore, the results, findings and conclusions presented in the article should be taken as a first approximation to the challenging problem of assessing aging in the signature trait, but not as conclusive and demonstrated facts. Furthermore, the study is also constrained to the type of subjects present in the database: Spaniards white Caucasians, mostly between 20 and 25 years of age, with a higher education degree (or pursuing it). For similar studies concerning other

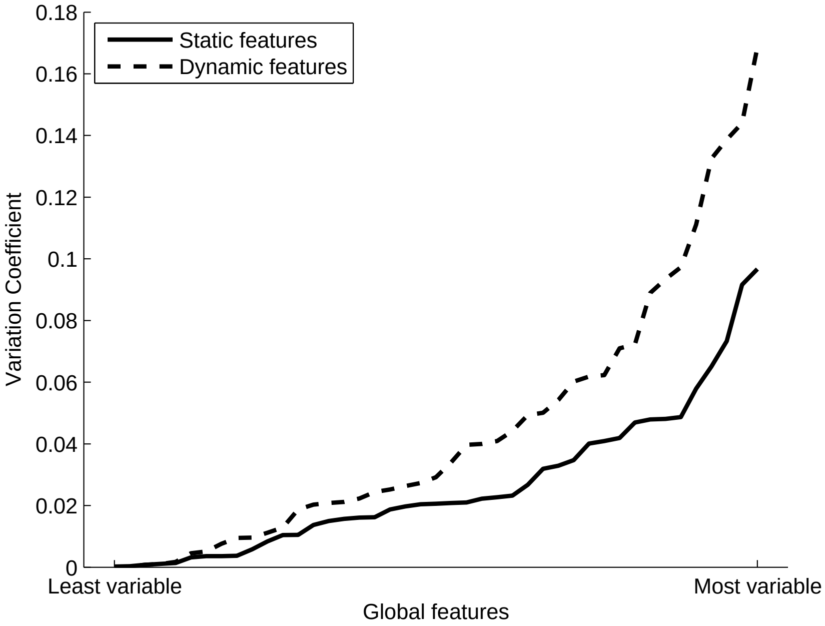


Figure 5.9: *Variation Coefficient (VC) from the least variable to the most variable dynamic and static features (see Table 5.1).*

	Most variable global features
$\Delta\mu_{gf}$	33(D), 36(S), 47(D), 95(S), 66(S), 64(D), 31(D), 10(D), 76(D) ,85(D)
$\Delta\sigma_{gf}$	73(S), 86(S), 76(D), 19(S), 85(D), 13(D), 90(D), 77(S), 65(S), 28(S)
VC	33(D), 47(D), 76(D), 85(D), 10(D), 64(D), 31(D), 36(S), 9(D), 32(D)
	Least variable global features
$\Delta\mu_{gf}$	38(D), 59(D), 3(D), 17(S), 20(D), 7(S), 19(S), 40(D), 46(S), 60(D)
$\Delta\sigma_{gf}$	93(S), 72(S), 58(D), 45(D), 17(S), 97(S), 21(D), 62(D), 67(S), 54(S)
VC	17(S), 58(D), 38(D), 93(S), 59(D), 72(S), 3(D), 45(D), 97(S), 7(S)

Table 5.6: *Most and least variable features over time. The numbering criterion is the same used in Table 3.1. 'S' stands for Static and 'D' for Dynamic according to the classification established in Table 5.1.*

sectors of the population, specific data should be acquired.

Accordingly, the present study should be understood as a valuable but limited start which leaves different open questions to be addressed in similar future works. For instance:

- Is 15 months a sufficiently long period of time to be in the presence of real "aging"? Although all the results given in the present work point in that direction, as mentioned above, this end should still be fully confirmed on a database acquired over a larger time span.
- What is the relationship (if any) between signature complexity and aging? In the current work an initial approach to address this issue has been followed. However, more rigorous studies should be carried out on databases where signatures have been grouped into different complexity levels either by experts, different human observers, or some type of objective measure.
- Can the results presented here (using data acquired in laboratory conditions) be generalized to real world scenarios? For this type of study specific data from a real application should be employed.
- Are the signatures from men/women more prone to aging? A large gender-balanced database may be used to study this issue.
- Is the aging effect more pronounced in individuals with low writing skills? The current study was carried out only taking into account subjects with higher education degrees.

5.5. Chapter Summary and Conclusions

We have conducted the first systematic study on the degradation of on-line signature with time and how this aging effect may be compensated. For this purpose, we have introduced the Signature Long-Term DB which contains the dynamic signature samples of the 29 common users of the BiosecurID and the BioSecure databases. All the subjects were captured under very similar conditions over a 15 month time span. The experiments, carried out using three totally different state-of-the-art systems representing the most usual technologies in on-line signature recognition, have proven that the aging effect is present in this trait even for time lapses of several months. Several conclusions have been extracted throughout the work thanks to the consistent and reproducible experimental protocol followed:

- Aging in the signature trait is a user-dependent effect. This means that:
    - In general, a user affected by aging perform badly regardless of the system being used (this deterioration will be higher in those systems more sensitive to time).
    - Complex and simple signatures can present the same amount of aging. Aging does not seem to depend on the type of signature but on the signer.

- Not all signature recognition technologies are equally affected by aging. The one based on DTW has demonstrated that it is not only the most accurate (Houmani *et al.*, 2012), but also the most robust against time.
- Global features containing dynamic information are in general less stable with time than those which comprise static information.
- With time, signatures evolve towards a higher simplicity. They become: shorter, faster and with fewer singular points and pen-ups.
- Depending on the signature recognition system being used some template update strategies are more efficient than others.

In summary, due to its very high user-dependency, the analysis and subsequent compensation of aging in the signature trait should be done, ideally, on a user by user basis. Given a specific signature recognition technology, different template update approaches should be adopted for different users, depending on the performance degradation that each of the subjects present with time. This is consistent with previous research works which also emphasize the strong user dependencies found in signature recognition (Fierrez-Aguilar *et al.*, 2005b; Jain *et al.*, 2002).

In light of the experimental results obtained in the present work, a possible strategy to detect the appearance of aging in the signature of a given individual would be to follow a constant monitoring over time of the Aging Coefficient. A possible "aging detection" protocol for a signature-based application would be:

1. Set a suitable AC threshold (i.e., δ_{AC}) for the given application depending on the amount of aging allowed.
2. With every new genuine access attempt, estimate the mean and variance of the last known N genuine access attempts and compare them to the mean and variance of the first N attempts (i.e., attempts that were recorded when the individual first started using the application).
3. Given the variation of the mean and variance between both sets of scores (new and old) compute the AC.
4. If δ_{AC} is exceeded, apply a suitable template update strategy depending on the signature recognition technology being used.

In this suggested protocol both δ_{AC} and N will depend on the type of application where it is being implemented (e.g., high security, commercial, high convenience), and on the level of restriction that will be imposed on aging. If only a small amount of aging is allowed a small value of both variables should be selected. On the contrary, if the designer prefers to be quite flexible with aging, larger values would be acceptable.

In summary, the main contribution of this chapter is the theoretical and practical new knowledge built in the fields of signature recognition and biometric aging, which may be directly applied by researchers and companies for the future development of the biometric technology.

Chapter 6

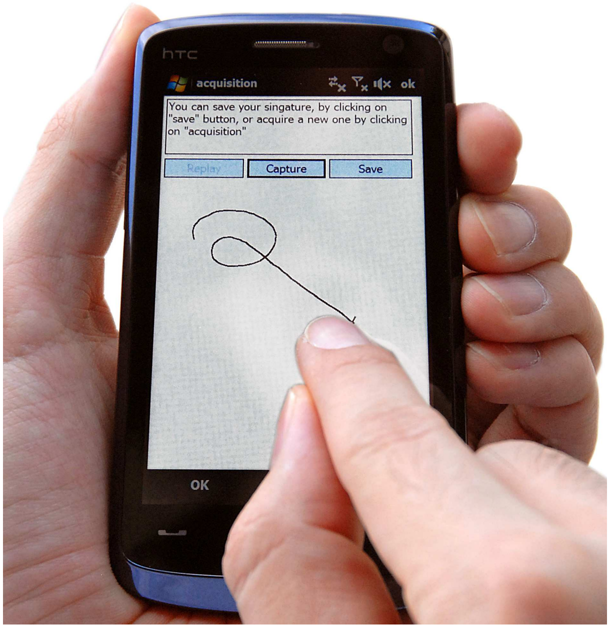
The DooDB Graphical Password Database

It is well known that publicly available databases together with their associated evaluation protocols make possible that researchers develop and objectively compare pattern recognition algorithms on the same benchmark. Experiments carried out using private databases are usually hard to replicate since database-specific effects, which cannot be reproduced by a third party, may take place. Unfortunately, there is no such a public database in the field of doodle-based graphical passwords, to the extent of our knowledge. Research on doodle verification has traditionally relied on private databases (Goldberg *et al.*, 2002; Govindarajulu and Madhvanath, 2007; Jermyn *et al.*, 1999; Oka *et al.*, 2008; Sae-Bae *et al.*, 2014). Moreover, in those works there is no reference to forgeries, since only genuine doodles are considered.

The objective of this chapter is the presentation and analysis of *DooDB*, a doodle and pseudo-signature database containing data from 100 users. Pseudo-signatures are doodles based on a simplified version of the user signature, being thus composed of learned and natural movements. The database has been captured on a handheld device under realistic conditions. It has two main advantages compared to other databases used in the literature: two acquisition sessions were performed, so inter-session variability effects can be analyzed, and skilled forgeries are provided for each user. The DooDB database is publicly available from the ATVS - Biometric Recognition Group website (http://atvs.ii.uam.es).

Another objective of this analysis is to obtain a baseline doodle verification performance that can be used to compare this method with current well known authentication alternatives such as signatures or with future doodle-based recognition algorithms. We also analyze the differences in the verification performance between doodles and pseudo-signatures. Since pseudo-signatures are simplified versions of real signatures, and thus composed of learned movements, it can be hypothesized that they present a lower variability and a better verification performance. The effects of inter-session variability are also studied.

The chapter is structured as follows. In Sect. 6.1 the database is described. Quantitative



You can save your singature, by clicking on "save" button, or acquire a new one by clicking on "acquisition"

Figure 6.1: *Doodle acquisition setup.*

and qualitative properties of the database are analyzed in Sect. 6.2. Preliminary verification experiments using the data from DooDB are reported in Sect. 6.3 and conclusions are finally drawn in Sect. 6.4.

This chapter is based on the publications: (Martinez-Diaz *et al.*, 2013, 2010a).

6.1. The DooDB Database

The DooDB database comprises two subcorpora, each one containing a different modality:

- **Subcorpus 1: Doodles**. Participants were asked to draw with their fingertip a doodle on a handheld device touchscreen that they would use as a graphical password on a regular basis for authentication (e.g. instead of the PIN code). There were no restrictions regarding duration or shape. In most cases, users invented their own doodle at the time of acquisition.

- **Subcorpus 2: Pseudo-signatures**. Participants were also asked to draw with their fingertip a simplified version of their signature, which they would also use as a graphical password on a regular basis. This could be, for example, their initials or part of their signature flourish. The main difference between doodles and this modality is that in this case, the dynamic process to produce the drawing is in general composed of natural and well trained movements.

6.1.1. Acquisition Protocol

Acquisition was performed using an HTC Touch HD smartphone (see Fig. 6.1). The device has a resistive touchscreen of 2×3.5 in (ca. 5×8.5 cm). The x and y coordinates of the fingertip position are sampled at discrete time values t at 100Hz when the user presses the screen. The coordinate values represent milli-inches, so x_t values range between $[0, 2000]$ (width) and y_t values between $[0, 3500]$ (height). The time interval Δ_t between consecutive samples is also stored. However, the device has some sampling errors, such as lost samples or samples that are not captured due to insufficient pressure. The device assigns [0,0] coordinate values to the erroneous samples. To summarize, each drawing is stored as a sequence of discrete values $[x_t, y_t, \Delta_t]$. Some examples of doodles and pseudo-signatures are shown in Fig. 6.2.

The acquisition process was divided in two sessions, separated by an average period of two weeks. This period was chosen in order to allow enough inter-session variability while trying to avoid that users forgot their doodles. Participants were briefed in the first session about the purpose of the acquisition. Each modality (doodles and pseudo-signatures) was explained to them following the same instructions so that each user received the same information. The donors were asked to draw with their fingertip on the handset screen holding it in their own hand, simulating thus real operating conditions. They were allowed to practice their drawings until they felt comfortable with them.

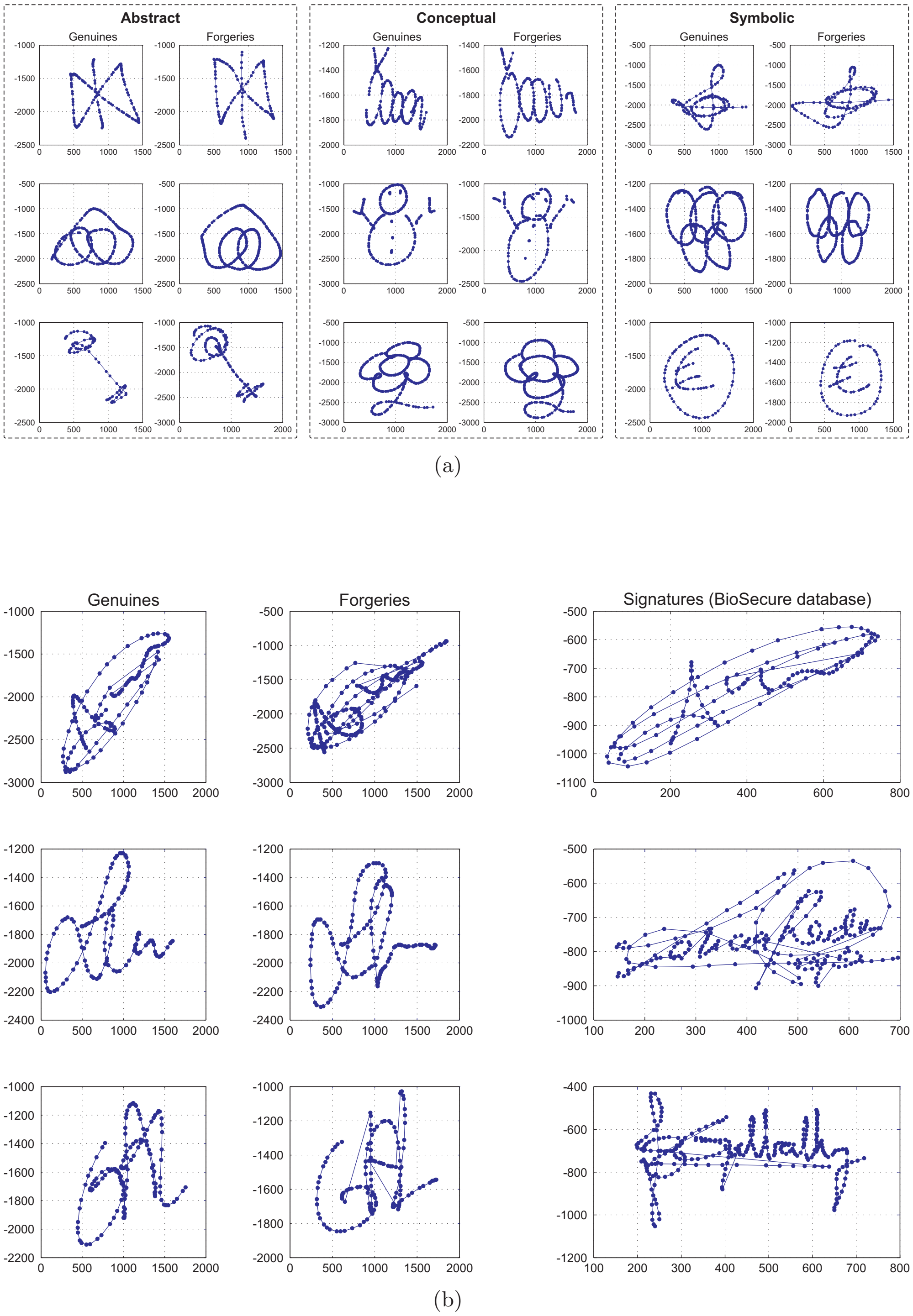
Forgeries have also been captured in this database. To perform forgeries, users had visual access to the doodle or pseudo-signature they had to imitate. The acquisition software replayed the strokes on the screen showing their dynamic properties (e.g. speed). This animation was shown to users up to three times, and then they were allowed to train until they felt confident with their forgery. The usage of the replay software makes possible to produce forgeries with a notable degree of accuracy, as can be observed in Fig. 6.2.

During the two sessions, the same protocol was followed for each user and modality: 5 genuine samples, then 5 forgeries, 5 genuine samples, followed by 5 forgeries and finally 5 genuine samples. This separation in blocks of 5 signatures allows analyzing intra-session variability. Consequently, at the end of the two sessions, each user had produced 30 genuine drawings (15 per session) and 20 forgeries. In the first session, user n produced forgeries for users $n - 1$ and $n - 2$, while in the second, forgeries for users $n - 3$ and $n - 4$ were produced.

6.1.2. Demographics and Memorability

The 100 participants in the database present the following age distribution: 75 are less than 25 years old, 14 are between 25 and 40 years old, and 11 are older. The gender distribution is 44 women and 56 men. It was observed during the capturing process that participants not familiar with touchscreen devices required a significant longer training time than the rest. This case was more common in older participants.

A subset of 13 participants of this database have also participated in the BioSecure Multimodal Database (BMDB) (Ortega-Garcia *et al.*, 2010). In that database, on-line signatures



(a)

(b)

Figure 6.2: *(a) Example of doodles from the database, classified following the criteria explained in Sect. 6.2. The doodle on the right is a forgery of the one on the left. (b) Example of pseudo-signatures from the database. Genuine pseudo-signatures (left), forgeries (middle) and the corresponding handwritten signature (right) from the BioSecure database (Ortega-Garcia* et al.*, 2010).*

were captured using both a pen-tablet and a PDA with a stylus. This overlap makes possible to observe the evolution of signatures from a controlled scenario (signature with ink pen and paper placed on a pen-tablet), towards more degraded conditions (signature on a PDA with a stylus) and, finally, the most challenging case of pseudo-signature (simplified signature traced with the fingertip). Some examples of genuine signatures and their corresponding pseudo-signatures from the same user are shown in Fig. 6.2.b.

One of the critical issues in graphical passwords is memorability. During the second acquisition session, it was observed that approximately 90% of the participants remembered correctly their pseudo-signature. On the other hand, nearly 40% of the participants had difficulties to recall their doodle from the first session. Users could request to see the tracing process of their own drawings from the first session. This was done by using the aforementioned functionality designed to train forgers. The high percentage of users that requested help to recall their doodles is related to the fact that they did not use them between sessions on a regular basis. In a real scenario with more frequent use, memorability may certainly improve.

6.2. Database Analysis

6.2.1. Statistical Properties

Given the different nature of doodles and signatures it is expected that they present differences in their properties such as their length or graphical complexity. A statistical analysis of the properties from the two captured subcorpora has been performed. They have also been compared with the ones from a BioSecure PDA Signature subcorpus of 120 users (also captured by the ATVS - Biometric Recognition Group), allowing thus a comparison between handwritten signatures, finger-traced pseudo-signatures and doodles. The following properties have been analyzed: graphical complexity (as the number of trajectory intersections), average speed and duration.

In Fig. 6.3.a, the distribution of the number of intersections in the drawings is represented. We observe that signatures present a considerably higher number of intersections, as expected. The difference between doodles and pseudo-signatures is small in this case. A low amount of intersections can be associated to low graphical complexity. This lower complexity indicates that doodles and pseudo-signatures may be easier to forge.

The stroke average speed distributions are compared in Fig. 6.3.b[1]. As can be seen, doodles are the “slowest” from the three datasets. The main cause for this may be that doodles are in general newly invented drawings for the participants, while pseudo-signatures are (or at least contain) previously learned movements. It can also be observed that pseudo-signatures are on average also produced faster than signatures. This is a reasonable result, since the motor process is different for the production of doodles and signatures. When producing a signature, the writer moves the stylus with a combination of his fingers and wrist movements (i.e. the

[1] This graph is a corrected version from the one presented in Martinez-Diaz *et al.* (2010a), which had an erroneous scaling for the signature duration histogram.

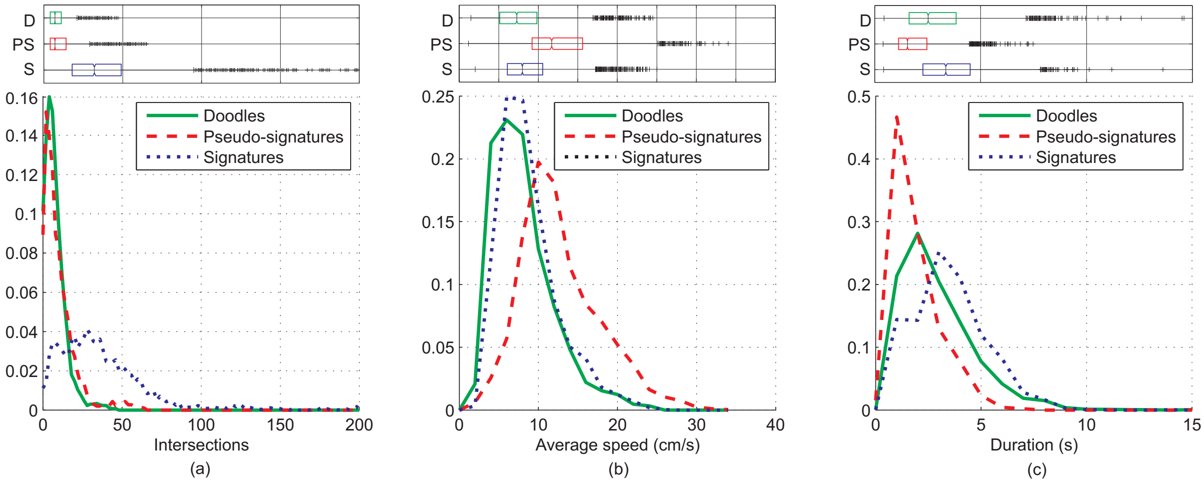


Figure 6.3: *Histograms normalized to [0,1] and box plots of (a) number of intersections, (b) average drawing speed and (c) duration.*

natural writing process), while in the case of finger-drawn sketches, the wrist is the main motor element, as the finger used for drawing is kept almost fixed. This way, signatures are based on more precise movements than doodles, and composed of small graphical elements compared to pseudo-signatures, which are produced by faster movements and larger shapes.

In Fig. 6.3.c, the statistical distribution of the three sets in terms of their total duration is represented. As can be seen, handwritten signatures tend to have a higher duration than the finger-traced drawings. Moreover, signatures present a higher variability in terms of duration. Doodles also tend to require more time than pseudo-signatures, which are in general composed of initials or simplified signature flourish.

6.2.2. Variability Analysis

Three types of variability may increase the error rate of a verification system. Intra-user variability reflects the difference between genuine samples of the same user. Inter-user variability represents the variance between samples of different users. Last, inter-session variability is related to the difference between samples of the same user over time. In general, verification performance will be best if intra-user and inter-session variability are low and inter-user variability is high.

An analysis of the three variability classes in DooDB is carried out in this section. A simple DTW-based verification system trained with the 5 first samples from session 1 is implemented (Martinez-Diaz *et al.*, 2009b), using three pairs of features: the coordinate sequence $[x, y]$, the speed sequence, $[x', y']$ and the acceleration sequence $[x'', y'']$.

Skilled and random forgeries are considered. To compute skilled forgery scores, the 20 available forgeries per user are employed. Random forgeries represent the case where a user claims to be a different one while providing his or her own doodle or pseudo-signature to the system. Random forgery scores are obtained by comparing the user reference set to the first genuine signature sample from each of the remaining users.

The verification performance for the three feature pairs is shown in Table 6.1 using separately genuine samples from session 1 and from session 2 as test samples. In the case of Session 1, the 10 remaining samples are used for verification (since the first 5 are used for training), while for session 2, all 15 samples are used for verification.

The score distributions of genuine samples from session 2, random forgeries and skilled forgeries are represented for each modality and for each feature pair in Fig. 6.4. The Equal Error Rates (EERs) of these systems are also shown in Table 6.1.

Several observations can be made from Fig. 6.4 and Table 6.1:

Intra-user Variability In Fig. 6.4, we observe that the genuine score distribution for doodles presents a long tail towards low scores. This effect reflects the presence of users who vary significantly the aspect or the dynamics (including stroke order) of their doodles. The highest intra-user variability (i.e. the most spread genuine score distribution) is observed for the acceleration features on doodles, which reflects the variation not only in the doodle aspect but also in the dynamics between different sessions. This indicates that in general users concentrate in reproducing the shape of their own doodles, but tend to vary the speed and acceleration of their strokes. The effect is reduced with pseudo-signatures, since generally these are based on better learned movements, and is clearly minimized for signatures, which are the best trained passwords of the three categories.

Inter-user Variability Regarding random forgeries, it can be observed in Fig. 6.4 that random forgery score distributions for doodles are shifted significantly towards lower scores, compared to pseudo-signatures and signatures. This is especially visible for the $[x, y]$ feature pair, revealing a higher inter-user variability, at least in shape, for doodles. This is not reflected in a lower EER in Table 6.1, since the tail towards lower scores for the genuine score distribution overlaps with forgery scores. When skilled forgeries are considered, inter-user variability is inversely related to the easiness of forging samples from another user. As can be seen in Fig. 6.4, there is a high overlap between skilled forgeries scores and genuine user scores for doodles. Skilled forgery scores decrease when dynamic features (speed and acceleration) are selected. However, since genuine user scores also decrease for these features on doodles, the overlap does not decrease significantly nor does the EER (for doodles). A predictable effect is that dynamic features such as speed and acceleration provide a higher separation between genuine and skilled forgery scores for signatures since they are harder to imitate, leading to lower EERs.

Inter-session Variability As expected, the error rates are higher in every case when genuine samples from session 2 are used (see Table 6.1). We observe that the performance degradation between sessions for doodles and pseudo-signatures is significantly higher than for signatures both in relative and absolute terms. It is also worth noting that the verification performance against random forgeries is in some cases better for doodles and pseudo-signatures than for

__Table 6.1:__ Verification performance in terms of EER (%) using samples from different sessions for authentication. EER_{sk} refers to the EER for skilled forgeries and EER_{rd} for random forgeries.

Features	Session	Doodles		Pseudo-signatures		Signatures	
		EER_{rd}	EER_{sk}	EER_{rd}	EER_{sk}	EER_{rd}	EER_{sk}
$[x, y]$	1	2.7	28.0	3.5	28.6	3.2	23.9
$[x', y']$	1	3.4	26.7	1.6	23.9	2.1	18.0
$[x'', y'']$	1	4.5	28.1	2.2	19.8	2.8	13.8
$[x, y]$	2	7.6	36.4	5.0	34.5	4.6	27.0
$[x', y']$	2	6.3	33.9	3.8	29.7	3.2	21.5
$[x'', y'']$	2	7.3	34.1	4.3	25.0	4.0	17.8

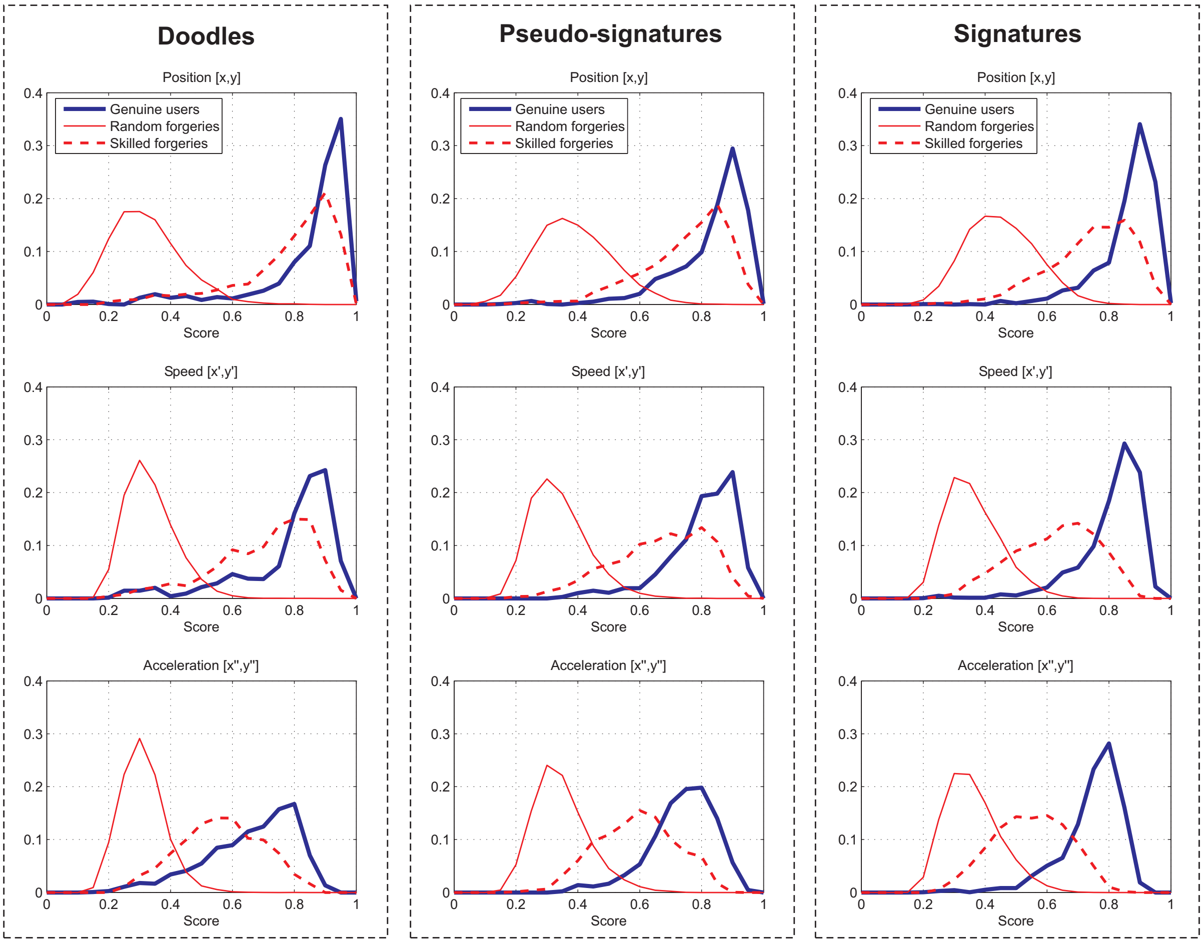


__Figure 6.4:__ Score distributions for Doodles (left), pseudo-signatures (middle) and signatures (right) using different feature pairs.

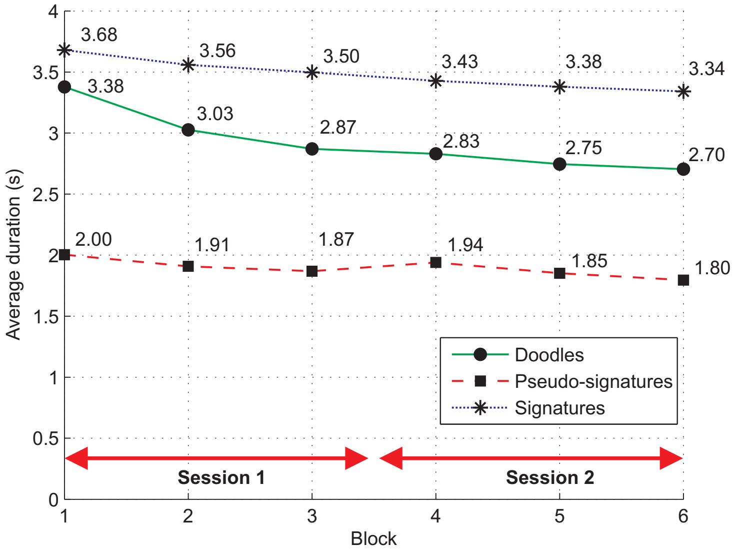


Figure 6.5: *Average genuine sample duration for each capture block during database acquisition (3 blocks of 5 samples per session).*

signatures. This suggests a higher variability in size and shape between users, compared to signatures. However, the higher error rates against skilled forgeries also reflects that pseudo-signatures, and especially doodles are significantly easier to forge.

6.2.3. Learning Curve

The learning curve for the three modalities (doodles, pseudo-signatures and signatures) is studied by analyzing the average genuine sample duration for each capture block during the database acquisition. As described in Sect. 6.1.1, during the database acquisition process, users were asked to draw genuine samples in blocks of 5, separated by the production of forgeries.

It can be hypothesized that if the average duration significantly decreases between different blocks, the users are still not used to the acquisition method or they are still learning how to produce their graphical password. The average duration for each modality among consecutive blocks is represented in Fig. 6.5. The average duration between the first block and the last block for the case of doodles has a 20% difference, while for pseudo-signatures and signatures there is only a 10% difference.

These observations corroborate the fact that doodles were in general specifically created for the experiments while pseudo-signatures are composed of well-learned movements.

6.2.4. Graphical and Qualitative Properties

When the whole doodle dataset is visually inspected, it can be seen that there are three main types of doodles:

- **Abstract** doodles, which cannot be directly interpreted as representing an object or idea.
- **Conceptual** doodles, which represent an object or idea (e.g. a flower).

- **Symbolic** doodles, which are known and recognizable symbols, like currency or musical notation.

Doodles that are abstract for an observer may be conceptual to another that is able to interpret them. However, it seems reasonable to assume that abstract doodles may be more resilient to forgers with visual access to them, since they are harder to remember (Renaud, 2009). The proportion of these three doodle types in the DooDB database is: 43 abstract, 37 conceptual, and 20 symbolic doodles, although this is based on a subjective evaluation. It has also been observed some repetitions among the doodles provided by participants, specially for common drawings. Some examples of repeated doodles are a flower symbol and a smiling face. Examples of each type of doodle are shown in Fig. 6.2.

Regarding pseudo-signatures, a clear classification between different types cannot be established. It is observed that most participants tend to produce a simplified version of the signature, including flourish. However, approximately 20% of the participants have written their initials, their name or a shortened version of their name without flourish.

6.3. Benchmark Results

In order to assess the authentication performance based on doodles and pseudo-signatures, preliminary experiments have been carried out. A simple verification system, based on Dynamic Time Warping (DTW) to compare the captured time sequences has been used, following the algorithm as described in Martinez-Diaz *et al.* (2009b).

Two representative local feature sets from the state of the art are studied in this benchmark. First, the one from the doodle authentication system proposed in Govindarajulu and Madhvanath (2007). In that system, 6 local features are extracted from the doodle trajectory. These are the coordinate sequence $[x, y]$, and its first and second derivatives (speed and acceleration). Thus, each doodle is described by the 6-dimensional sequence $[x, y, x', y', x'', y'']$. Matching is performed using the DTW algorithm. We refer to this feature set as HP-LOCAL.

The other system is based on the one presented by the Biometric Recognition Group - ATVS to the BioSecure Signature Evaluation Campaign BSEC 2009 (Houmani *et al.*, 2012). In particular, the system is the one based on DTW that was tuned to maximize its performance against skilled forgeries, identified as system “DTWs” in Houmani *et al.* (2012). It was one of the best performing systems in most evaluation scenarios against skilled forgeries. This feature set is referred to as ATVS-BSEC. The system extracts the following 7 local features:

- x-coordinate, x
- Second-order derivative of x-coordinate, x''
- First-order derivative of y-coordinate, y'
- Second-order derivative of y-coordinate, y''

- Path velocity, $\upsilon = \sqrt{(y')^2 + (x')^2}$
- First-order derivative of path velocity, υ'
- First-order derivative of the *log* curvature radius, ρ', where $\rho = \log(\upsilon/\theta')$ and $\theta = \arctan(y'/x')$ is the curvature of the position trajectory.

6.3.1. Experimental Protocol

The experimental protocol follows the one described in Sect. 6.2.2, but only genuine signatures from session 2 are used for authentication.

The whole sets of doodles and pseudo-signatures from the DooDB database are used for the experiments. The first 5 genuine samples from the first session of each user are used for enrollment as reference templates. The 15 genuine signatures of the second session are used to compute genuine user scores, simulating thus real operating conditions, in which inter-session variability affects the verification performance.

Random and skilled forgery scores are obtained following the same protocol described in Sect. 6.2.2.

For each comparison against the 5 reference templates, an output score is generated by averaging the inverse of the 5 DTW distances obtained.

6.3.2. Results

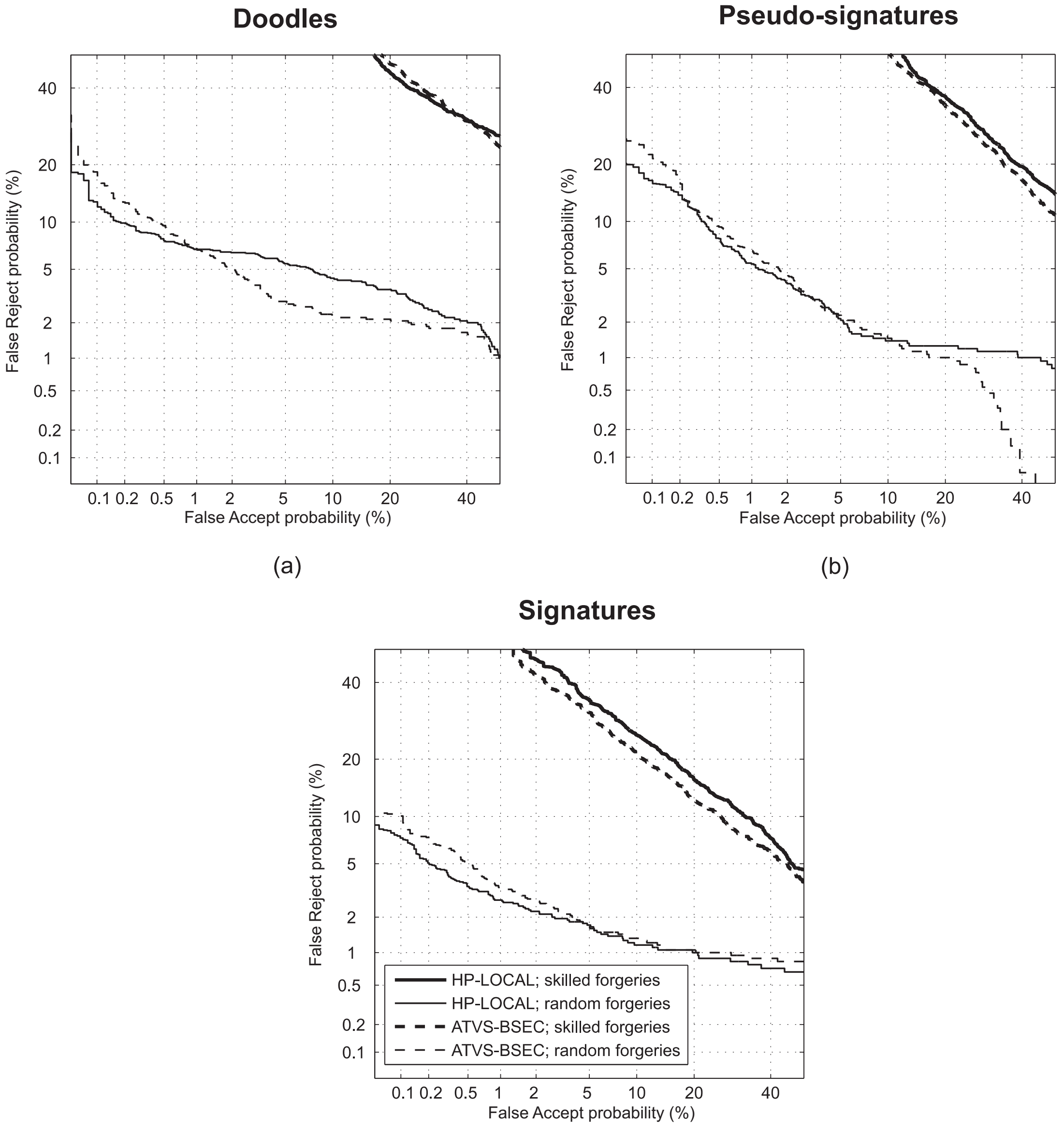
The verification performance in terms of Equal Error Rate (EER) is shown in Table 6.2 and DET (Detection Error Tradeoff) curves for each dataset are represented in Figure 6.6. As can be seen, the performance is higher (i.e. lower error) for pseudo-signatures compared to doodles both for random and skilled forgeries.

Comparing Table 6.2 (which considers state-of-the-art feature sets) to the results shown in Table 6.1 using only samples from session 2 (with simple feature pairs) for verification, we can see that the performance is similar. This is an indication that the selected state-of-the-art feature sets may not be totally adequate for doodles, and better performance may be achieved by considering feature extraction adjusted to the doodle recognition problem. This is subject to future work.

In Table 6.1 we also saw that the performance against skilled forgeries improved for pseudo-signatures when dynamic properties (i.e. speed or acceleration) were used. This effect may be due to the higher consistency in the drawing process of pseudo-signatures, since they are composed in general of natural or learned movements. On the other hand, when doodles are considered, the usage of speed or acceleration properties does not increase the performance in the same proportion. This may be due to an increased variability in the drawing process. In fact, it was observed during the doodle subset acquisition, that some users varied the stroke order of their doodles even in the same session. This was not the case for pseudo-signatures.

Table 6.2: *Verification performance in terms of EER (%) using samples from session 2 for authentication. EER_{sk} refers to the EER for skilled forgeries and EER_{rd} for random forgeries.*

Features	Doodles		Pseudo-signatures		Signatures	
	EER_{rd}	EER_{sk}	EER_{rd}	EER_{sk}	EER_{rd}	EER_{sk}
HP-LOCAL	5.4	33.8	3.1	28.4	2.1	17.8
ATVS-BSEC	3.4	34.4	3.1	26.9	2.5	15.8



(a)

(b)

(c)

Figure 6.6: *DET plots for (a) doodles, (b) pseudo-signatures and (c) signatures.*

6.4. Chapter Summary and Conclusions

The DooDB database has been presented. This database comprises doodles and pseudo-signatures from 100 users and skilled forgeries for all of them. The acquisition protocol has been described and various data analyses have been performed. Benchmark verification experiments have been carried out, revealing that one of the main challenges of doodle and pseudo-signature verification may be the protection against forgeries.

We have also observed that there is a high intra-user variability in the production of doodles, which negatively affects the verification performance. Unlike the case of signature verification, where dynamic features such as acceleration of velocity clearly increase the verification accuracy (Fierrez and Ortega-Garcia, 2008), the variability found in doodles defies the utility of dynamic features for doodle-based authentication. On the other hand, pseudo-signatures are more stable and thus provide promising results. Users may produce doodles more naturally over time, assuming a frequent usage, leading to an improvement in their verification performance which would become closer to pseudo-signatures in the long term.

Based on the results, doodles and pseudo-signatures are seen as a potential lightweight authentication method oriented to mobile devices. One of the main advantages of this kind of graphical password is its convenience and the possibility of performing user authentication without extra hardware unlike, for example, fingerprint authentication. As previously stated, revocability is an advantage of doodles with respect to other biometric traits.

Chapter 7

Free-form doodle verification

GRAPHICAL PASSWORDS have become popular due to the proliferation of touchscreen devices, in particular smartphones and tablets. However, the prevalent approaches are based on simple graphical passwords which can be easily remembered and reproduced by potential attackers. This chapter focuses on free-form graphical passwords captured on touchscreen devices. Authentication is based on features extracted from the dynamics of the doodle drawing process (e.g. speed or acceleration). These features contain behavioral biometric information, which has been successfully used for automatic user verification based on handwritten signatures (Fierrez and Ortega-Garcia, 2008). As a consequence, a potential attacker would have to copy not only *what* the user draws, but also *how* the user draws it. Unfortunately graphical passwords tend to be much simpler than signatures and are not composed, in general, of previously learned or heavily practiced movements. This can lead to a higher intra-user variability (i.e. variations between samples produced by the same person) than in the case of signatures or may cause users to forget part or the whole graphical password that they provided during enrolment. On the other hand, while users may be concerned about their privacy when registering their signature on an authentication system, doodles can be a potential solution to overcome this type of legal and social issues. Doodles have also a high revocability compared to signatures.

In this chapter, we study the advantages, drawbacks and limits of user authentication based on finger-drawn free-form doodles and authentication based on pseudo-signatures, which are simplified versions of the signature drawn with the fingertip (see Fig. 7.1). To the extent of our knowledge, this is the first exhaustive and systematic analysis of user authentication on touchscreens based on free-form sketches, using a publicly available database. The recently acquired DooDB Graphical Password Database, described in Chapter 6 is used for this purpose (Martinez-Diaz *et al.*, 2013). The contributions of this chapter can be summarized as follows:

1. Two verification systems are proposed, one based on Gaussian Mixture Models (GMMs), and another based on the Dynamic Time Warping (DTW), which are state-of-the-art approaches for signature verification. We analyze the performance of these systems against

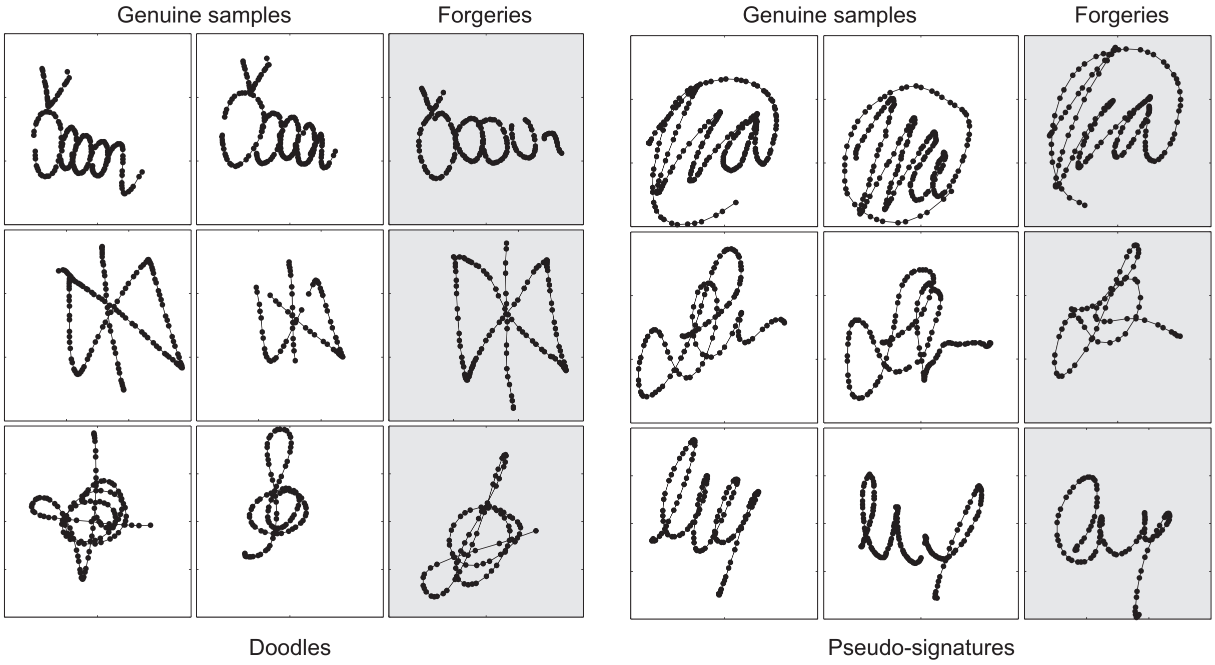


Figure 7.1: *Examples of doodles and pseudo-signatures from the DooDB database (Martinez-Diaz* et al.*, 2013).*

random forgeries (when attackers claim to be another user but use their own password) and intentional forgeries (when attackers have visual access to the password being forged).

2. Feature selection is carried out in order to understand which features provide the highest discriminative power for doodles and pseudo-signatures.

3. The impact of inter-session variability (i.e. the effects of time between enrolment and authentication) is studied.

4. We study the impact of the number of available training samples during enrolment on the verification performance.

5. An improved authentication system based on the best selected features and the fusion of the two aforementioned matchers (GMM and DTW) is presented.

The chapter is structured as follows. In Sect. 7.1 the proposed verification systems are described. Experiments and results are reported in Sect. 7.2, and conclusions are finally drawn in Sect. 7.3.

Preliminary results of the work presented in this chapter were reported by the author in Martinez-Diaz *et al.* (2010b).

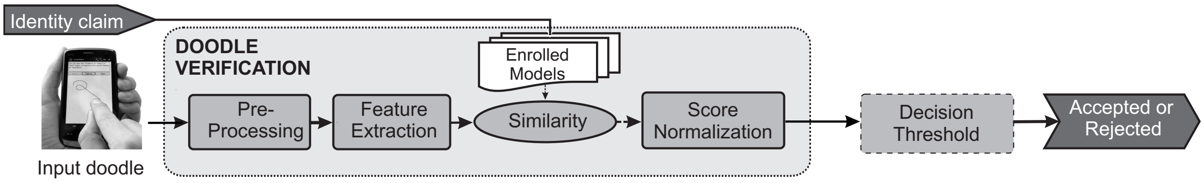


Figure 7.2: *Main components of a Doodle Verification System.*

7.1. Proposed Algorithms

In this section, the proposed Doodle Verification Systems are described. In both systems, the input coordinate sequence $[\hat{x}_n, \hat{y}_n]$ is sampled from the finger-tip trajectory on a touchscreen, as well as the time interval $\hat{t}_n$ between samples. A generic model of a doodle verification system is shown in Fig. 7.2 (following the signature verification architecture described in Chapter 2).

7.1.1. Preprocessing and Feature Extraction

The trajectory coordinate sequence $[\hat{x}_n, \hat{y}_n]$ $i = 1, \ldots, I$ is first resampled to interpolate missing samples (due to sampling errors or pauses between strokes). Cubic splines are used for interpolation. The sequences are then normalized to have zero mean, resulting in $[x_n, y_n]$.

A set of 19 additional features are extracted from the $[x_n, y_n]$ coordinate sequence. A description of the feature set can be found in Table 3.2. All features are normalized to have zero mean and variance equal to 1. Thus, each doodle is described by a total amount of 21 time functions.

7.1.2. Gaussian Mixture Model system

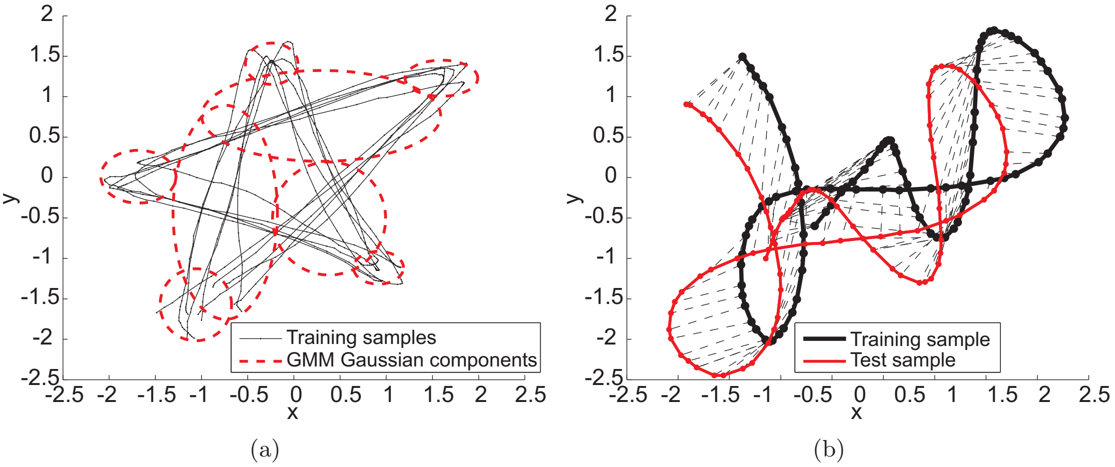
For each user u, the distribution of d features extracted from the fingertip motion is modeled by a d-dimensional Gaussian Mixture Model λ_u, as described in Sect. 2.1.3.

In our work, the number of Gaussian components N is chosen to be 32, and diagonal covariance matrices $\mathbf{\Sigma}_i$ are used, based on the benchmark results reported by Richiardi and Drygajlo (2003), and preliminary experiments which are omitted for the sake of clarity. The model parameters $\{\omega_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i\}$ $i = 1, \ldots, N$ are estimated from a training set of doodles using the Expectation Maximization (EM) algorithm.

During the enrollment phase one model is created for each user, which is later used for matching. In addition, a world GMM is created, which models the whole set of users. World models are used during the matching phase and are trained using doodles from a separate group of users, as explained in the experiments.

A graphical representation of a GMM is depicted in Figure 7.3.a. A GMM with 8 Gaussian components (represented by circles of 2σ width) trained with the pair of features $[x_n, y_n]$ of 5 user samples is shown.

The match score, given a test vector $\mathbf{x}$ and a target user statistical model λ_C, can be computed as a ratio of the log-likelihood that the test vector $\mathbf{x}$ is produced by the model λ_C and



(a)

(b)

Figure 7.3: *(a) Representation of an 8-component GMM trained with the $[x_n, y_n]$ features of 5 user samples. The 2σ contour is depicted for each Gaussian component. (b) Representation of the point to point correspondences between two doodles obtained using the DTW algorithm.*

the log-likelihood that the test vector has been produced by any other user, which is modeled by the world model $\lambda_{\bar{C}}$.

So, following the previous notation, a match score s is obtained as follows:

$$s = \log p\left(\mathbf{x}\,|\lambda_C\right) - \log p\left(\mathbf{x}\,|\lambda_{\bar{C}}\right). \tag{7.1}$$

7.1.3. Dynamic Time Warping system

The DTW system described in Sect. 2.1.3 is used.

An example of a set of corresponding samples between two doodles from the same subject, using the $[x_n, y_n]$ functions is depicted in Fig. 7.3.b.

7.2. Experiments

7.2.1. Database and Experimental Protocol

The doodle and pseudo-signature sets from the DooDB database[1] are used for the experiments (Martinez-Diaz *et al.*, 2013). As described in Chapter 6, the doodle dataset consists of free-form doodles, while the pseudo-signature dataset is composed of simplified finger-drawn signatures. Doodle and pseudo-signature examples are provided in Figure 7.1. A brief overview of the database is given in this chapter for the sake of clarity.

The database was captured in an HTC Touch HD touchscreen mobile phone at a sampling rate of 100Hz. Both datasets were produced by the same set of 100 users in two sessions, separated by an average of 2 weeks. Users were requested to hold the handheld device in their own hands while drawing. Participants were briefed to provide a graphical password that they

[1] Available at: http://atvs.ii.uam.es/databases.jsp

would use as an authentication method and were left to train until they felt comfortable with the capture method. For each password, the $[\hat{x}_n, \hat{y}_n]$ coordinate sequence is captured, and the time interval between each sample. The time interval is in general constant, except in the transitions between consecutive strokes.

During each session, each user provided 15 genuine samples of each type (doodle and pseudo-signature) and 10 forgeries. To increase the quality of forgeries, the system replayed the target sample drawing process.

In the experiments, the first 50 users of the database are selected as the development set for feature selection purposes, while the remaining users are left for validation. In the development experiments, the GMM world models are estimated using the genuine samples from the validation set and vice-versa. Enrollment is done with the first 5 genuine samples from the first session of each user. Unless stated otherwise, genuine scores are obtained with the 15 genuine doodles from the second session, to take into account inter-session variability.

Two types of forgeries are considered. *Skilled forgery* scores are obtained using the 20 available forgeries per user. *Random forgery* scores are computed for each user by comparing the user reference set (DTW system) or model (GMM system) to one sample from each of the other users. Random forgeries represent the situation where a forger claims to be a different user but provides his or her own doodle or pseudo-signature.

Throughout the next section, when results are presented, EER_{sk} refers to the Equal Error Rate (EER) for skilled forgeries and EER_{rd} for random forgeries.

7.2.2. Experiment 1: Feature Selection

First, we analyze which are the most discriminative features for each verification system. Feature selection on the local 21-feature set using the Sequential Forward Floating Search (SFFS) algorithm is carried out (see Sect. 2.6.1). The algorithm, is used here to find a near-optimal feature set that minimizes the system EER on the development datasets.

Feature selection is performed in 2 different scenarios for each dataset (doodles and pseudo-signature):

- **PSEUDO-SK** & **DOODLE-SK**: minimize the system EER against skilled forgeries.
- **PSEUDO-RD** & **DOODLE-RD**: minimize the system EER against random forgeries.

In both scenarios, the 15 doodles and pseudo-signatures from the second session are used for genuine score computation, while the first 5 signatures from Session 1 are used for enrollment. Thus, inter-session variability is taken into account.

The best performing feature sets selected by the SFFS algorithm for each optimization scenario are shown in Table 7.1, where several patterns can be observed. Feature $\ddot{y}_n$ (vertical acceleration) is present in 7 of the 8 sets, and features $\dot{y}_n$ (vertical speed) and $\dot{\rho}_n$ (variation of log curvature radius) are present in 6 of the 8 sets. This indicates that vertical dynamic features may be more stable than horizontal features. However, it can also be observed that feature $\ddot{x}$ is present in the four GMM optimal feature sets. This implies that GMMs may be more robust to users that change the usual left-to-right drawing order of their sketches. This last observation is aligned with the fact that GMMs, contrary to DTW, do not consider the temporal order of time series for matching.

The performance in terms of EER against random (EER_{rd}) and skilled (EER_{sk}) forgeries using the optimal feature sets on the development and validation datasets is shown in Table 7.2. The average of the user-specific EERs (referred to as $aEER$) is also reported. It is computed by averaging the individual user EERs that are obtained with user-specific decision thresholds. This represents the best EER that can be obtained if user scores are optimally normalized. As can be seen, the verification performance on the development and on the validation set is similar in general.

It can be observed that the GMM has a notably better verification performance against skilled forgeries than the DTW system. On the other hand, the DTW system has a significantly higher performance against random forgeries. The error rates against skilled forgeries are higher for doodles, contrary to the case of random forgeries, where doodles have a better performance. This may imply that pseudo-signatures are harder to imitate but are more similar between them than doodles.

It can be seen in Table 7.2 that, for the GMM system, the EER for random and skilled forgeries does not vary significantly independently of whether the system is optimized for either of the two forgery types. This is not the case for the DTW system, where the EERs vary significantly between the two optimization scenarios. This may reveal that for DTW-based doodle authentication, different features are suitable for random and skilled forgeries respectively. That behavior is corroborated by the results of the BSEC 2009 signature verification competition, where DTW systems tuned separately for random or skilled forgeries reached top performances against each kind of forgery (Houmani *et al.*, 2012).

7.2.3. Experiment 2: Inter-session variability

Using the feature sets obtained in Experiment 1, we analyze the impact in the verification performance of providing samples from the first session for authentication (instead of samples from Session 2). Consequently, user models are trained with the first 5 samples from Session 1, and genuine scores are computed using the 10 remaining samples of Session 1. Results are shown in Table 7.3. As can be seen, the EER improves significantly in all scenarios, compared to the previous experiment (where the test samples were taken from the second session). This reflects a high inter-session variability, which may be due to the limited training period that users had while defining their own graphical password.

For the GMM system, the EER improvement is homogeneous in relative terms (around 35% – 45%), except in the case of doodle random forgeries. An improvement of nearly 70% in the EER against random forgeries is observed (from 7,2% to 2,2% in the development subsets). This corroborates that users may be failing to reproduce accurately their own doodle in Session 2.

Regarding the DTW system, the EER improvement against skilled forgeries is around

Table 7.1: *Feature sets selected by the SFFS algorithm on the development datasets.*

System	Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		x_n	y_n	θ_n	υ_n	ρ_n	a_n	$\dot{x}_n$	$\dot{y}_n$	$\dot{\theta}_n$	υ_n	$\dot{\rho}_n$	$\dot{a}_n$	$\ddot{x}_n$	$\ddot{y}_n$	υ_n^r	α_n	$\dot{\alpha}_n$	s_n	c_n	r_n^5	r_n^7
GMM	PSEUDO-SK		✓		✓						✓	✓		✓	✓							
	DOODLE-SK		✓		✓							✓		✓	✓							
	PSEUDO-RD		✓			✓		✓		✓	✓	✓		✓	✓							
	DOODLE-RD	✓	✓			✓						✓		✓	✓							
DTW	PSEUDO-SK										✓			✓	✓							
	DOODLE-SK								✓			✓	✓		✓							
	PSEUDO-RD		✓					✓	✓	✓						✓		✓				
	DOODLE-RD		✓								✓	✓			✓	✓				✓		

Table 7.2: *Verification performance in terms of EER and average individual EER (aEER) using the feature sets selected by the SFFS algorithm (Table 7.1). Results on the development (left) and validation (right) datasets are shown. Enrollment with Session 1 (5 signatures) and testing with Session 2. Data in (%).*

System	Scenario	Development subset				Validation subset			
		EER_{sk}	EER_{rd}	$aEER_{sk}$	$aEER_{rd}$	EER_{sk}	EER_{rd}	$aEER_{sk}$	$aEER_{rd}$
GMM	PSEUDO-SK	**17.2**	12.9	13.5	7.6	**20.9**	12.0	14.9	6.8
	DOODLE-SK	**24.3**	9.2	18.5	4.9	**23.0**	7.9	17.8	4.1
	PSEUDO-RD	18.6	**9.5**	14.8	4.8	23.1	**12.9**	17.2	6.4
	DOODLE-RD	24.6	**7.2**	20.4	2.9	23.7	**6.7**	17.2	3.4
DTW	PSEUDO-SK	**21.6**	5.2	15.4	1.1	**29.0**	2.7	19.5	0.9
	DOODLE-SK	**31.9**	4.1	24.8	0.9	**33.0**	5.2	29.0	1.3
	PSEUDO-RD	29.1	**2.0**	23.2	0.7	33.6	**1.3**	21.0	0.4
	DOODLE-RD	36.7	**1.6**	26.5	0.3	32.7	**1.4**	27.3	0.3

Table 7.3: *Verification performance using samples from Session 1 both for enrollment and testing. The feature sets described in Table 7.1 are considered. Data in (%).*

System	Scenario	Development subset				Validation subset			
		EER_{sk}	EER_{rd}	$aEER_{sk}$	$aEER_{rd}$	EER_{sk}	EER_{rd}	$aEER_{sk}$	$aEER_{rd}$
GMM	PSEUDO-SK	**11.5**	7.3	8.3	3.3	**16.2**	8.8	11.0	4.0
	DOODLE-SK	**15.5**	5.1	10.7	2.1	**14.4**	3.6	10.4	1.5
	PSEUDO-RD	12.4	**5.9**	8.2	3.3	16.4	**7.5**	12.5	3.2
	DOODLE-RD	14.60	**2.2**	11.3	0.8	13.5	**2.6**	9.2	1.0
DTW	PSEUDO-SK	**15.2**	1.4	8.4	0.3	**22.8**	2.2	12.8	1.1
	DOODLE-SK	**25.2**	1.2	15.6	0.1	**26.1**	3.3	17.5	1.1
	PSEUDO-RD	20.2	**0.6**	10.8	0.0	27.0	**0.8**	15.3	0.2
	DOODLE-RD	29.3	**0.4**	16.2	0.2	23.7	**1.4**	15.5	0.3

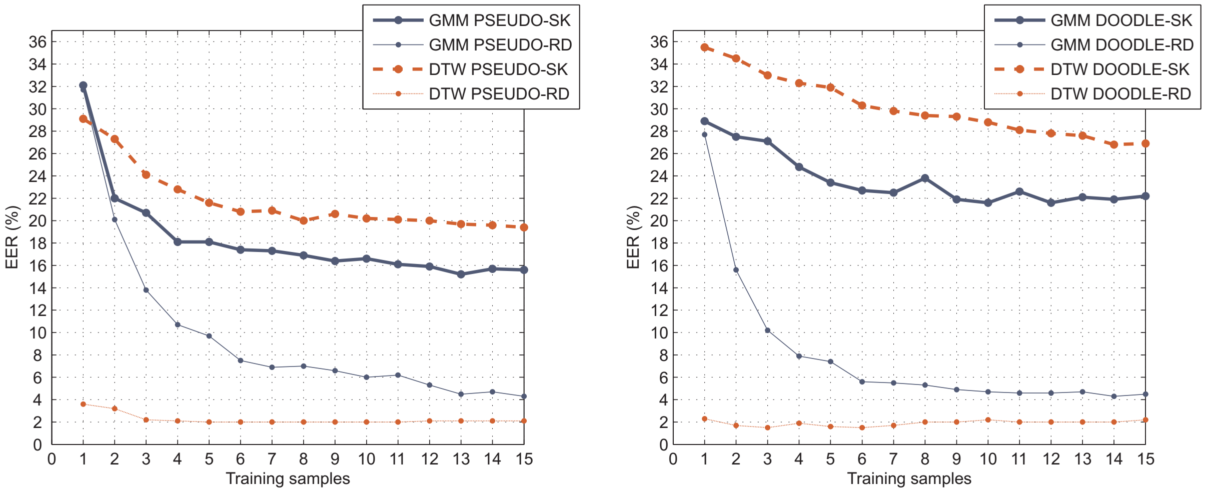


Figure 7.4: *Evolution of the EER in each scenario in terms of the number of training samples.*

20% – 30% in relative terms, while against random forgeries is around 70% in most cases. This reinforces the previous observations about a high inter-session variability. It is worth noting that the DTW system reaches remarkably low EERs, below 1%, and average EERs near 0%.

7.2.4. Experiment 3: Training set size

The effect of the available number of samples during enrollment is also studied. Maintaining the previously computed optimal feature sets, the EER is obtained on each scenario using from 1 to 15 samples from Session 1 for training. Samples from Session 2 are used for authentication.

In Fig. 7.4, the EER evolution with respect to the number of graphical samples used for training is shown. As might be expected, the EER decreases in general when more training samples are available. However, this is not the case for the DTW system against random forgeries on both datasets. The EER does not vary significantly when additional samples are available. In the rest of the cases, the EER starts to stabilize at 6-7 training samples.

7.2.5. Experiment 4: Fusion

Finally, the verification performance combining the best systems of Experiment 1 is studied by applying score fusion. Thus, the GMM system optimized against skilled forgeries and the DTW system optimized against random forgeries are combined and results are computed for both datasets (doodles and pseudo-signatures)

A simple fusion scheme based on score weighted sum is used. This approach has shown a remarkable performance over other techniques (Kittler *et al.*, 1998). The fusion score is computed as $s = (1-k)s_D + ks_G$, where s_D and s_G are the DTW and GMM system scores respectively and k is the fusion weighting factor. The optimal value of k is estimated heuristically on the development dataset and is equal to 0.5 on both datasets.

The verification performance of both resulting fused systems on the validation datasets is

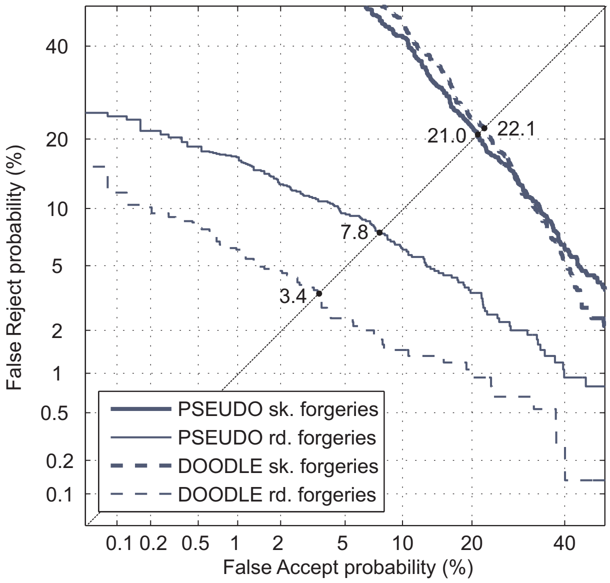


Figure 7.5: *Verification performance applying score fusion.*

shown in Fig. 7.5, represented by the DET plot and the corresponding EERs. As can be observed, there is no significant performance increase on the pseudo-signature dataset. On the other hand, combining both systems on the doodle dataset improves the performance of the GMM system against skilled forgeries and random forgeries, due to the DTW system contribution.

7.3. Chapter Summary and Conclusions

Two different algorithms have been analyzed for the problem of free-form graphical password verification, and the effects of feature selection, inter-session variability, and training set size have been studied. It has been observed that vertical features tend to be more prevalent than horizontal ones in the optimal feature sets, indicating a possible higher discriminative power.

It has also been noticed that, using DTW, different feature sets provide highly different performances against random and skilled forgeries. This is aligned with the results of the DTW systems presented by the author to the BSEC 2009 Signature Evaluation Campaign (Houmani *et al.*, 2012). It was found that, using DTW, systems could be optimized independently (using feature selection) against random forgeries and for skilled forgeries. These systems could be afterwards combined, via score fusion, and result in a very high performance against both types of forgeries.

Session inter-variability has proven to cause a considerable negative impact in verification performance, as already observed by Martinez-Diaz *et al.* (2013), probably due to users that fail to reproduce correctly their own graphical passwords. Although the GMM systems may overcome partially this issue (since they do not take into account the stroke order), verification performance is still considerably degraded. It has also been found that the optimal enrollment set size is around 7 samples, a bit higher than the common trend in the signature verification literature (5 samples) (Fierrez and Ortega-Garcia, 2008).

It has also been observed in the experimental results, that depending on the optimization scenario (skilled or random forgeries) very different optimal feature sets are selected by the SFFS algorithm. In addition, the GMM system has a better performance against skilled forgeries while the DTW system has a better performance against random forgeries. This suggests that random and skilled forgeries may be a different problem from a pattern recognition point of view. This corroborates results already observed in the signature verification field, namely in the BioSecure Signature Evaluation Campaign 2009 (Houmani *et al.*, 2012), where verification systems from many international research groups were compared. It was found that the best performing systems against random and skilled forgeries were tuned for each scenario respectively, and fusion of both systems provided an overall good performance in both scenarios. In our case, score fusion has shown to provide better results than individual systems, especially in the case of doodles.

Chapter 8

Conclusions and Future Work

THIS THESIS has studied the problem of automatic user authentication on handheld devices using signatures and gesture-based graphical passwords. The effects of aging on signature recognition have also been analyzed. A summary of the state of the art on these topics has been first presented and the verification systems used in this Thesis have been also described. The experimental studies have analyzed the effects of mobility on signature features compared to traditional capture conditions, the feasibility and performance of user authentication based on doodles, and the influence of aging on the verification performance.

8.1. Conclusions

Chapter 1 outlines the basic principles and methods of biometric recognition, focusing on signature-based authentication. The motivation of this Thesis and the specific contributions are also described. The state of the art on related topics is reviewed in Chapter 2, covering signature verification, graphical password-based authentication, feature selection and biometric aging. In Chapter 3, the verification systems and feature sets used in the experimental body of this Thesis are described.

The experimental contributions start in Chapter 4. The effects of mobile capture conditions on signature verification performance are first studied. Signature datasets captured on mobile conditions and on a traditional scenario (pen tablet) are used for the experiments. A local HMM system and a global system are used for the experiments. It is found that the lack of trajectory information during pen-ups (which happens on mobile conditions due to the use of touchscreens for signature acquisition) negatively affects verification performance. This is specially remarkable for the local system used in the experiment, compared to the global system. Thus, global features appear to be more robust in mobile conditions, although it is well known that they usually provide a worse verification performance than local features. The methodology followed in these experiments (i.e. comprehensive set of features, feature subset selection, and robust matchers) has led the author to remarkable success in the BioSecure Signature Evaluation Campaign, ranking first in several tasks (BSEC, 2009; Houmani *et al.*, 2012). The author also reached second position in the on-line category of the ICDAR 2009 Signature Verification Competition (SigComp2009) (Blankers *et al.*, 2009), following this methodology.

The effects of aging on signature verification are studied in Chapter 5. This is the first systematic study on the degradation of on-line signature with time and how this aging effect may be compensated. The experiments are carried out using three state-of-the-art systems (HMM, DTW and distance-based). It has been observed that the aging effect is present in this trait even for time lapses of several months. We have found that aging in the signature trait is a user-dependent effect. A user affected by aging performs badly regardless of the system being used. Not all signature recognition technologies have been found to be equally affected by aging. The one based on DTW is the most robust to the passing of time. Regarding different feature types, we observe that global features containing dynamic information are in general less stable with time than those which comprise static information. We have also found that in the datasets used in the experiments, signatures evolve towards a higher simplicity over time. They become shorter, faster and with fewer singular points and pen-ups. Template update strategies have been studied, in order to mitigate the effects of aging. Their efficiency has found to be variable depending on the signature recognition system being used.

Research works such as the one presented here try to shed some light into the difficult problem of biometric aging. Performing systematic studies of biometric systems sensitivity to time is essential before effective strategies that minimize the impact of the detected effects can be developed, so that the user acceptability of this rapidly emerging technology is improved.

This way, we believe that this work can be of great utility not only for researchers, but also for developers and vendors in order to produce more secure and trustful applications based on the signature trait, to better understand its strengths, and to be able to foresee the weaknesses of this biometric modality. Furthermore, this type of study can also help to develop the ongoing biometric standards and to better define the requirements that real applications should comply with (ANSI-INCITS 395-2005, 2005; ISO/IEC 19794-11, 2005; ISO/IEC 19794-7, 2005).

In Chapter 6, the DooDB graphical password database has been presented. This database contains doodles (free-form finger-drawn sketches) and pseudo-signatures, which are finger-drawn and simplified versions of signatures. The acquisition protocol has been described and various data analyse have been performed. Benchmark verification experiments have been conducted, revealing that one of the main challenges of doodle and pseudo-signature verification may be the protection against forgeries. A high intra-user variability in the production of doodles has also been observed. On the other hand, pseudo-signatures are more stable and thus provide promising results. Based on the results, doodles and pseudo-signatures are seen as a potential lightweight authentication method oriented to mobile devices. One of the main advantages of this kind of graphical password is its convenience and the possibility of performing user authentication without extra hardware unlike, for example, fingerprint authentication.

Chapter 7 studies the problem of graphical password-based user authentication using doodles. The DooDB database, presented in Chapter 6, is used for that purpose. Two algorithms from the signature verification state of the art are used: DTW and GMMs. A feature selection process is also carried out in order to study which features are most discriminative for graphical passwords. It is found that vertical movement related features are more prevalent in the optimal features sets. We have also observed that different feature sets work best against random or skilled forgeries respectively. That is, random forgery or skilled forgery detection is a different problem and may be approached using different classifiers (which can be afterwards combined). This was also observed for the case of signature verification in the results of the BioSecure Signature Evaluation Campaign (BSEC) 2009 (Houmani *et al.*, 2012), where the algorithms developed by the author reached top positions against random and skilled forgeries independently. It has also been found that the optimal number of training features is slightly higher than in the case of signature verification (7 vs. 5).

To summarize, the main contributions of this Thesis are:

- The up-to-date survey on mobile signature verification, biometric aging and recall-based graphical passwords.
- The experimental analysis of which signature features are more robust on mobile conditions and the evidence of performance degradation when pen-up trajectories are not captured.
- The novel experimental setup to analyze the effects of biometric aging on signature verification.
- Experimental evidence on the effect of aging in signature verification, and its compensation through template update approaches.
- The acquisition and analysis of the first publicly available finger-drawn graphical password database, including also pseudo-signatures.
- The experimental analysis of graphical password-based user authentication, showing that using feature selection and score fusion, a promising verification performance can be obtained.

8.2. Future Work

Based on the work presented in this Thesis, several research paths arise. The following ones are considered of interest by the author:

- This Thesis has focused on the problem of user authentication on mobile conditions, compared to traditional desktop scenarios. There is however a need to analyze the impact of device inter-operability, that is, the impact of acquiring biometric samples in a device that is different than the one that has been used to train the user model (e.g. two different brands of smartphones, or one smartphone and one pen tablet). Some research contributions already addresses this challenging scenario (Alonso-Fernandez *et al.*, 2005), but a systematic study on inter-operability with a large and publicly available database is yet to be performed. Recent efforts are going in this line (Blanco-Gonzalo *et al.*, 2013a; Vera *et al.*, 2015).

- Regarding aging, this Thesis has shed some light on its effects in signature verification. However, several questions arise. Is 15 months a sufficiently long period of time to be in the presence of real "aging"? Although all the results given in the present work point in that direction, this should still be fully confirmed on a database acquired over a larger time span. Other factors, such as gender, or writing skills may also be of interest when aging is considered. A recent reference in this regard is Fairhurst (2013).

- Sketch-based graphical passwords are still a relatively novel field of research and still represent a challenging authentication scenario. Future research using additional datasets should be carried out, (like the set recently presented by Riggan *et al.* (2014)), taking into account other usage profiles compared to the acquisition scenario of the DooDB database, which was captured considering only a span of two weeks between sessions.

- Inter-session variability has been found to be one of the main factors for performance degradation, when doodles are used for authentication. Template update techniques (Didaci *et al.*, 2014; Uludag *et al.*, 2004) could help to alleviate this problem.

- This Thesis has studied free-form graphical passwords produced on touchscreens. With the proliferation of front cameras on smartphones and body motion capture technologies (such as Microsoft Kinect^{TM}), in-air gestures should be studied as an authentication means. It has also been found that users tend to prefer gestures than touchscreen interaction in some scenarios (Parada-Loira *et al.*, 2014). Some recent works in this field are (Lai *et al.*, 2012; Wu *et al.*, 2014)

- As we have seen, the acquisition scenario (e.g. smartphone, tablet) affects verification performance. There is in general a trade-off between usability and performance. Some recent works have studied the impact of usability and acquisition conditions on signature verification (Blanco-Gonzalo *et al.*, 2014, 2013b; Brockly *et al.*, 2014). Future research may be carried out in this area, in order to identify measurable usability features and analyze their correlation with verification performance.

- This work has relied on well-established matching techniques (HMM, GMM, DTW and distance measures). Other complementary approaches that have gained popularity, such as Support Vector Machines (SVM) (Ferrer *et al.*, 2005), Deep Neural Networks and hybrid HMM/Neural Network approaches (Dahl *et al.*, 2012; Povey *et al.*, 2011) should be explored in order to identify other possible top performing verification algorithms.

Apéndice A

Resumen Extendido de la Tesis

Verificación de Firma y Gráficos Manuscritos: Características Discriminantes y Nuevos Escenarios de Aplicación Biométrica

En la actualidad los sistemas de reconocimiento biométrico son una alternativa a métodos tradicionales de autenticación, como contraseñas, llaves físicas o electrónicas. La biometría permite validar la identidad de un usuario mediante la utilización de un rasgo anatómico (p.ej. huella dactilar) o comportamental (p.ej. firma manuscrita) inherente a una persona (Jain *et al.*, 2008), y es algo habitual dentro del ámbito forense y judicial desde hace más de un siglo.

En comparación con los métodos clásicos comúnmente utilizados, como llaves o claves, los rasgos biométricos no pueden, en general, ser prestados, robados o copiados. El usuario emplea directamente su propia huella dactilar, retina, voz u otro rasgo para ser reconocido. Por otro lado, esta clase de sistemas suele ser fácil de mantener y en general no requiere la intervención de más agentes que el propio usuario para funcionar.

Los rasgos biométricos pueden clasificarse según varias características (Jain *et al.*, 2008). Entre ellas cabe mencionar su unicidad, su distintividad o individualidad, su universalidad, su facilidad de proceso y adquisición o su variabilidad con el tiempo. La firma manuscrita reúne muchas de estas características y es además uno de los medios más utilizados desde la antigüedad para validar la autoría de documentos escritos. Otra ventaja es la facilidad de captura electrónica de la firma, especialmente tras la masiva proliferación de dispositivos con pantallas táctiles. Son ejemplos de ello las tabletas, los teléfonos inteligentes (*smartphones*), los ordenadores portátiles, y terminales de punto de venta con pantalla táctil.

El reconocimiento biométrico es un área de investigación madura, con libros de referencia (Jain *et al.*, 2008, 2011; Ratha and Govindaraju, 2008; Ross *et al.*, 2006), conferencias específicas en el área (ICB, 2015; IJCB, 2014; Vijaya-Kumar *et al.*, 2010), revistas específicas (Fairhust, 2012), proyectos internacionales (BBfor2, 2010; Biosecure, 2004; Tabula Rasa, 2010), consorcios dedicados al reconocimiento biométrico (BC, 2015; BI, 2015; EAB, 2015) y estándares internacionales (ANSI/NIST, 2009; SC37, 2005). La investigación en firma manuscrita es además un área que, si bien está muy activa desde hace varias décadas (Plamondon and Lorette, 1989), continúa suponiendo un problema para el que se siguen buscando soluciones. Prueba de ello es el número de competiciones en firma manuscrita celebradas en los últimos años (Blankers *et al.*, 2009; BMEC, 2007; Houmani *et al.*, 2012, 2011; Liwicki *et al.*, 2011; Malik *et al.*, 2013; Yeung *et al.*, 2004) y el alto volumen de publicaciones científicas en el área (Fierrez and Ortega-Garcia, 2008; Impedovo and Pirlo, 2008; Impedovo *et al.*, 2012; Plamondon and Lorette, 1989).

Los dispositivos portátiles con pantalla táctil (*smartphones*, tabletas, etc.) están motivando un cambio en cuanto a la interacción hombre-máquina. Permiten, por un lado la captura de firmas en movilidad, y por otro una interacción basada en gestos trazados con los dedos sobre la pantalla. La autenticación puede estar, de hecho, no solo basada en firma sino en un conjunto diferente de trazos escogidos por el usuario (denominado *password* gráfico), o en una versión simplificada de la firma. Se abre por tanto un nuevo escenario de aplicación, en donde estos trazos, ya sean la firma completa trazada con el dedo o un conjunto de trazos, pueden ser utilizados como rasgo biométrico.

A.1. Resumen y Conclusiones

Esta Tesis aborda la verificación de firma manuscrita centrándose en tres ejes principales, la autenticación en dispositivos móviles, el efecto del paso del tiempo (conocido como *biometric aging*) y un nuevo escenario de aplicación: la autenticación basada en gestos o firmas simplificadas realizadas con el dedo sobre una pantalla táctil.

A.1.1. Capítulo 1: Introducción

En el Capítulo 1 se presenta en primer lugar una introducción a la biometría y en particular el reconocimiento de firma manuscrita. Se explica adicionalmente la motivación de la Tesis y se describen y detallan las contribuciones de la misma, proporcionando un listado de publicaciones científicas del autor resultantes del trabajo de la Tesis, clasificadas por temática.

A.1.2. Capítulo 2: Trabajos Relacionados y Estado del Arte

El Capítulo 2 contiene una descripción del estado del arte en verificación automática firma manuscrita, con especial atención a la autenticación en condiciones de movilidad, así como del estado del arte en verificación de passwords gráficos y en el estudio de los efectos *aging*. Se proporciona también una descripción de las bases de datos disponibles de firma manuscrita *on-line* en la comunidad científica y una revisión de algoritmos de selección de características.

A.1.3. Capítulo 3: Sistemas de Verificación Propuestos

En el Capítulo 3 se describen los sistemas de verificación utilizados en los experimentos desarrollados a lo largo de la Tesis. Se emplean dos tipos de sistemas: locales y globales. El sistema global está basado en la extracción de un vector de 100 parámetros globales de cada firma (Tabla 3.1), las cuales han sido previamente muestreadas en una pantalla táctil o tableta digitalizadora y normalizadas, y realiza el cálculo de similitud empleando la distancia de Mahalanobis. En cuanto a los sistemas locales, se han implementado tres sistemas diferentes, uno basado en Modelos Ocultos de Markov (Hidden Markov Models, HMM), otro en Modelos de Mezclas Gaussianas (Gaussian Mixture Models, GMM) y otro basado en el método de alineamiento elástico *Dynamic Time Warping* (DTW). En el caso de los sistemas locales, se extraen un total de 27 funciones (Tabla 3.2) de cada firma, que suponen un compendio de las funciones utilizadas en la literatura científica en los últimos años.

A.1.4. Capítulo 4: Verificación de Firma en Movilidad

Las contribuciones experimentales comienzan en el Capítulo 4. En él se aborda el problema de la verificación de firma manuscrita en dispositivos móviles. Se basa en las publicaciones del autor (Martinez-Diaz *et al.*, 2008a, 2014). En este capítulo se utiliza la base de datos de firmas BioSecure Multimodal Database (BMDB) (Ortega-Garcia *et al.*, 2010) la cual contiene un conjunto de firmas capturadas tanto en una tableta digitalizadora como en una PDA, del mismo grupo de usuarios. El principal objetivo del capítulo es evaluar cuáles son los efectos provocados por la captura en condiciones de movilidad en el rendimiento de los sistemas de verificación de firma. Es razonable suponer que la menor ergonomía de un dispositivo móvil comparado con una tableta así como las condiciones de captura, en movimiento frente a en una superficie firme, pueden afectar negativamente al proceso de verificación, incrementando las tasas de error. Se observa además en la base de datos que, dado que las pantallas táctiles no capturan la trayectoria del estilete cuando éste no está en contacto con la pantalla (al contrario que en el caso de las tabletas digitalizadoras), se dejan de muestrear en promedio el 18% aproximadamente de la trayectoria de las firmas. En el capítulo se realiza un análisis del poder discriminante individual de los vectores de características locales y globales definidos en el Capítulo 3. Este análisis se realiza sobre 3 conjuntos de firmas, el capturado en la PDA, el capturado en la tableta digitalizadora, y uno adicional que se crea a partir del capturado en tableta, eliminando los puntos muestreados cuando el bolígrafo no está en contacto con la superficie e interpolándolos.

Para el análisis individual de los parámetros globales, se utiliza el *Fisher's Discriminant Ratio* (FDR), y se define una medida ad-hoc para los parámetros locales, que denominamos *Distance Discrimant Ratio* (DDR), descrita en el Apartado 4.2. Se realiza también un análisis del poder discriminante de combinaciones de parámetros. Para ello se realiza selección de características mediante el método *Sequential Forward Floating Search* (SFFS) (Theodoridis and Koutroumbas, 2006), con el fin de encontrar cuáles son los conjuntos de características que proporcionan una menor tasa de error en cada escenario (móvil y tableta) y tipo de impostor (casual o intencionado). Se realiza selección de características sobre los conjuntos de parámetros globales y locales descritos en el Capítulo 3 y empleando el sistema global basado en distancia de Mahalanobis y el sistema local basado en HMMs.

En los resultados se observa que en general las características tienen un un poder discriminante individual superior en el caso de la tableta con respecto a la PDA (véase Figura 4.3), y que el hecho de interpolar las trayectorias cuando el bolígrafo no está en contacto con la superficie (imitando lo que sucede en la PDA) también reduce el poder discriminante. Esto se observa especialmente en el caso de las imitaciones intencionadas, por lo que cabe suponer que las trayectorias del bolígrafo en el aire son más difíciles de imitar.

En cuanto a la selección de características, se observa que en general el número óptimo de parámetros globales está en torno a 40, con respecto al total de 100 (véanse Figura 4.4 y Tabla 4.1). Se observa además que los parámetros de naturaleza geométrica (relacionados con características geométricas de la firma) son los que prevalecen en los vectores óptimos (véase Figura 4.5).

El tamaño de los vectores óptimos de parámetros locales oscila entre 6 y 9 características (véase Tabla 4.2). Se observan varios comportamientos en el caso de los parámetros locales. En primer lugar, ninguna característica relacionada con presión u orientación del bolígrafo está presente en los conjuntos óptimos de características. Se entiende por tanto que su ausencia en el caso de la PDA, dado que no es posible capturarlas en pantallas táctiles, no debería afectar la tasa de error del sistema (al contrario que lo los resultados presentados por Muramatsu and Matsumoto (2007) y alineado con las observaciones de Houmani *et al.* (2009)). Se observa además que la coordenada x, la derivada de y, el coseno c y el ángulo de la trayectoria α están presentes en la mayoría de los vectores (estos parámetros están descritos en la Tabla 3.2). Se observa también que el número de características es menor al comúnmente empleado en la literatura (Fierrez *et al.*, 2007b; Ly-Van *et al.*, 2007; Richiardi *et al.*, 2005).

Finalmente se realizan experimentos de validación, sobre conjuntos de firmas diferentes a los empleados para la selección de características (véase el protocolo detallado en el Apartado 4.2). El rendimiento de los sistemas en términos de *Equal Error Rate* (ERR) se muestra en la Tabla 4.3. Se observa que el sistema global presenta en general un mejor rendimiento en el caso de la PDA. Cuando se interpolan las trayectorias en el aire, la tasa de error se ve incrementada, especialmente en el caso de imitaciones intencionadas. Se puede observar también que cuando los sistemas se optimizan frente a imitaciones aleatorias, el rendimiento empeora notablemente frente a imitaciones intencionadas. Al contrario, si se optimizan frente a imitaciones intencionadas, el rendimiento frente a imitaciones aleatorias no empeora en gran medida.

La metodología seguida en este capítulo (conjunto inicial muy amplio de características, selección de características y algoritmos robustos) ha permitido al autor obtener resultados muy exitosos en la competición de firma manuscrita BioSecure Signature Evaluation Campaign, alcanzando el primer puesto en varias categorías (BSEC, 2009; Houmani *et al.*, 2012). Obtuvo de forma similar el segundo puesto en la categoría *on-line* de la competición ICDAR 2009 Signature Verification Competition (SigComp2009) (Blankers *et al.*, 2009).

A.1.5. Capítulo 5: *Aging* en Firma Manuscrita

El Capítulo 5 estudia los efectos del paso del tiempo (conocido como "aging") en el rendimiento de los sistemas de verificación de firma manuscrita. Se basa en las publicación del autor con Galbally *et al.* (2013). Supone la primera contribución científica en la que se presenta un conjunto de firmas capturadas durante un periodo superior a un año y se analizan los efectos del *aging* sobre el mismo. Se propone una metodología nueva para analizar los efectos del paso del tiempo, que puede ser extrapolable a otros rasgos biométricos. Esta metodología es utilizada para extraer conclusiones acerca de los efectos del *aging* y cómo pueden ser mitigados.

La base de datos utilizada se denomina Signature Long-Term Database y contiene firmas de 29 usuarios comunes de las bases de datos BiosecurID (Fierrez *et al.*, 2010) y BioSecure (Ortega-Garcia *et al.*, 2010). En total, el periodo de captura se extiende a lo largo de 15 meses. Los detalles del protocolo de adquisición se muestran en la Figura 5.1.

En el apartado experimental, se analizan dos principales aspectos. En primer lugar cuál es el impacto del *aging* en el rendimiento (en cuanto a tasas de error) de la verificación de firma manuscrita, y en segundo lugar cuáles son los cambios específicos que experimentan las firmas con el paso del tiempo, y qué parámetros son más estables. Se utilizan para los experimentos los sistemas locales HMM, DTW y el sistema global basado en distancia de Mahalanobis descritos en el Capítulo 3.

En cuanto a los experimentos relacionados con el rendimiento de los sistemas, se observa que las tasas de error (EER) se incrementan con el paso del tiempo, según las firmas de test han sido capturadas más tarde con respecto a las capturadas en el registro de los usuarios. Este efecto se observa en los tres sistemas de verificación empleados (véanse Figuras 5.3 y 5.3). El sistema DTW es el más robusto ante el paso del tiempo. Se observa además que la varianza de las puntuaciones de los sistemas de verificación aumenta con el paso del tiempo. Se analiza también cómo afecta el *aging* a cada usuario de la base de datos individualmente (véanse Figura 5.5 y Tabla 5.4), apreciándose que el *aging* afecta en medida muy diferente a cada usuario, existiendo usuarios a los que les afecta en escasa medida y otros en gran medida. Finalmente, se analiza el efecto de actualizar las plantillas de usuario con firmas más recientes, con respecto a las de test. Se comprueba que cuanto más recientes son las firmas de entrenamiento, mejor es el rendimiento, entre otras observaciones (véase Figura 5.7).

En relación a los experimentos acerca de los cambios que experimentan las firmas con el tiempo, se analiza en primer lugar la evolución de los parámetros globales con el *aging* (véase Figura 5.8). Se observa que en general, con el paso del tiempo, las firmas son más cortas, rápidas, con menores puntos singulares y número de trazos. En definitiva, las firmas se simplifican con el paso del tiempo. Se mide también cuál es el efecto del *aging* en los parámetros globales, observándose que en general los parámetros relacionados con características dinámicas son los que más varían con el paso del tiempo, frente a los parámetros geométricos que son más robustos.

Este capítulo trata por tanto de mostrar los efectos del *aging* en la firma manuscrita y cómo pueden ser mitigados mediante la actualización de las plantillas de usuario. Se considera que este tipo de análisis puede ser de utilidad para la comunidad científica y la industria, en la medida que ayudan a desarrollar sistemas más robustos ante el paso del tiempo. Este tipo de estudios pueden además ayudar a la definición de requerimientos para aplicaciones biométricas reales en estándares como ANSI-INCITS 395-2005 (2005); ISO/IEC 19794-11 (2005); ISO/IEC 19794-7 (2005).

A.1.6. Capítulo 6: DooDB, Base de Datos de *Passwords* Gráficos

El Capítulo 6 describe la adquisición y características de la base de datos DooDB, que contiene *passwords* gráficos y firmas simplificadas trazados con el dedo sobre la pantalla táctil de un dispositivo móvil. Está basado en las publicaciones (Martinez-Diaz *et al.*, 2013, 2010a). La base de datos DooDB es la primera base de datos disponible para la comunidad científica que contiene *passwords* gráficos capturados de forma sistemática. Contiene muestras de 100 usuarios, para cada una de las dos modalidades, capturadas en dos sesiones separadas por dos semanas en promedio. Cada sesión se divide en 3 bloques de captura de 5 muestras por bloque. En la Figura 6.2 se pueden observar algunas muestras de la base de datos. En primer lugar se realiza un análisis estadístico de las muestras capturadas, comparándolas con una base de datos de firmas capturadas en una PDA (de la base de datos BioSecure). Se observa que en general los *passwords* gráficos tienden a ser más sencillos gráficamente que las firmas y que se trazan más rápido (véase Figura 6.3). Se analiza también la variabilidad de los *passwords* gráficos y firmas simplificadas, observándose una mayor variabilidad de las muestras genuinas en comparación a la firma manuscrita, así como una variación sensiblemente mayor con el paso del tiempo entre sesiones de captura (véanse Figura 6.4 y Tabla 6.1). Para este análisis se emplea un clasificador DTW y se analizan por separado la variabilidad de la secuencia de coordenadas $[x, y]$, su primera derivada (velocidad) y su segunda derivada (aceleración). La mayor variabilidad se observa en la aceleración. Se puede apreciar también que las firmas simplificadas presentan menor variabilidad que los *passwords* gráficos, probablemente porque están compuestas por movimientos más practicados que los *passwords*. Se analiza también la curva de aprendizaje de cada rasgo (*passwords*, firmas simplificadas y firmas tradicionales), comparando la duración en promedio de las muertas a lo largo de cada bloque de captura. Se observa que en general la duración va decreciendo con el tiempo, pero en mucha mayor medida en el caso de los *passwords* que las firmas (véase Figura 6.5).

En el capítulo también se lleva a cabo un análisis cualitativo de los *passwords* gráficos. Se observan tres grandes tendencias: *passwords* abstractos, que no pueden ser interpretados; *passwords* conceptuales, que representan un concepto reconocible y *passwords* simbólicos, que representan un símbolo o letra reconocible. En cuanto a las firmas simplificadas, un 80% de los usuarios realizan una versión simplificada de su firma original, mientras que el resto emplean sus iniciales o una versión acortada de su nombre. Por último, se realizan experimentos de verificación a modo de *benchmark*, utilizando un sistema basado en DTW y los parámetros globales del sistema con mejor rendimiento ante imitaciones intencionadas en la competición de firma manuscrita BSEC 2009 (Houmani *et al.*, 2012). Se observa en general tasas de error mayores al caso de firma manuscrita, especialmente en el caso de las imitaciones intencionadas. A pesar de las mayores tasas de error, que pueden derivar en escenarios de aplicación diferentes a la firma manuscrita, una de las principales ventajas observadas de este rasgo biométrico es la usabilidad y facilidad de captura, que puede realizarse en cualquier *smartphone* o tableta del mercado de forma intuitiva y sin elementos adicionales.

A.1.7. Capítulo 7: Verificación de *Passwords* Gráficos

El Capítulo 7 estudia el rendimiento de diferentes sistemas y conjuntos de características ante el problema de la verificación de *passwords* gráficos y firmas simplificadas trazadas con el dedo. Para ello se emplea la base de datos DooDB descrita en el Capítulo 6. Se proponen dos sistemas de verificación basados en parámetros locales, utilizando GMMs y el algoritmo DTW. Se analiza mediante selección de características cuáles son los parámetros óptimos en diferentes escenarios: *passwords* gráficos o firmas simplificadas e imitaciones aleatorias o intencionadas. Se estudia también el impacto de la variabilidad inter-sesión, el impacto del número de firmas de entrenamiento y se propone finalmente un sistema de verificación basado en fusión de los sistemas GMM y DTW analizados.

En cuanto a la selección de características, se aprecia que las características relacionadas con el movimiento vertical, especialmente la aceleración en la coordenada vertical, están presentes en la mayoría de los vectores óptimos, al contrario que los parámetros relacionados con el movimiento horizontal (véase Tabla 7.1). Se observa que el sistema GMM presenta en general un rendimiento mejor ante imitaciones intencionadas, y al contrario en el caso de aleatorias, donde el sistema DTW alcanza una tasa de error EER mucho menor (véase Tabla 7.2). En la Tabla 7.3 se muestra el rendimiento en términos de EER de los sistemas bajo estudio y los vectores de características previamente seleccionados cuando se utilizan muestras de test de la misma sesión que las de entrenamiento (sesión 1). Se observa un rendimiento mucho mayor de los sistemas, alcanzándose en algunos casos tasas de error cercanas a 0%. Esto da lugar a entender que existe una muy alta variabilidad inter-sesión, probablemente debida a que los *passwords* gráficos en general no están compuestos por movimientos muy practicados y naturales, como es el caso de la firma manuscrita. En cuanto al número óptimo de muestras de entrenamiento, se comprueba que se sitúa en torno a 6-7 muestras, al contrario que el caso general de firma que está en torno a 5 muestras (Fierrez and Ortega-Garcia, 2008) (véase Figura 7.4). Por último, mediante la fusión de ambos sistemas, se alcanza un rendimiento combinado frente a imitaciones aleatorias e intencionadas mejor que los sistemas independientes.

A.1.8. Contribuciones de la Tesis

En resumen, las principales contribuciones de esta Tesis Doctoral son:

- La revisión actualizada del estado del arte en verificación de firma manuscrita en condiciones de movilidad, el *aging* y el reconocimiento basado en *passwords* gráficos.

- El análisis experimental de qué características de la firma son más robustas en condiciones móviles y la evidencia de la degradación del rendimiento si las trayectorias del bolígrafo o estilete no se capturan cuando no está en contacto con la superficie.

- Un nuevo protocolo experimental para analizar los efectos del *aging* sobre la firma como rasgo biométrico.

- La evidencia experimental del efecto del *aging* en la verificación de firmas, y su compensación a través de métodos de actualización de plantillas.

- La adquisición y el análisis de la primera base de datos disponible para la comunidad científica de *passwords* gráficos trazados con el dedo, incluyendo también firmas simplificadas.

- El análisis experimental de la autenticación de usuario basada en *passwords* gráficos (y firmas simplificadas), mostrando que si se lleva a cabo selección de características y fusión de clasificadores, se obtienen tasas de reconocimiento prometedoras.

A.2. Líneas de Trabajo Futuro

En base al trabajo presentado en esta Tesis, surgen varias líneas de trabajo futuro relacionadas. Las siguientes líneas de investigación son consideradas de interés por parte del autor:

- Esta Tesis se ha centrado en el problema de la autenticación de usuarios en condiciones de movilidad, en comparación con escenarios tradicionales basados en una tableta digitalizadora sobre un escritorio. Sin embargo, existe la necesidad de analizar el impacto de la inter-operabilidad entre dispositivos, es decir, el impacto de la adquisición de muestras biométricas en un dispositivo que es diferente de la que se utiliza para autenticar posteriormente al usuario (por ejemplo, dos diferentes marcas de smartphones, o un smartphone y una tableta). Algunas contribuciones ya abordan este escenario (Alonso-Fernandez *et al.*, 2005) pero está todavía por realizar un estudio sistemático sobre la inter-operabilidad con una gran base de datos a disposición del público. Algunos trabajos recientes van en esta línea (Blanco-Gonzalo *et al.*, 2013a; Vera *et al.*, 2015).

- En cuanto al *aging*, esta Tesis ha arrojado algo de luz sobre sus efectos en la verificación de firmas. Sin embargo, surgen varias preguntas. ¿Es 15 meses un período suficientemente largo de tiempo para analizar los efectos del *aging*? Aunque los resultados obtenidos en el presente trabajo apuntan en esa dirección, esto aún debe ser confirmado por completo con una base de datos adquirida a lo largo de un lapso mayor de tiempo. Otros factores, como el sexo, o las habilidades en la escritura pueden ser también de interés cuando se considera el envejecimiento. Una referencia reciente en este sentido es Fairhurst (2013).

- Los *passwords* gráficos son todavía un campo de investigación relativamente novedoso y continúan representando un escenario de autenticación desafiante, en comparación con otros rasgos biométricos. Debe llevarse a cabo más estudios con otras bases de datos (como la base de datos presentada recientemente por Riggan *et al.* (2014)), teniendo en cuenta otros perfiles de uso en comparación con el escenario de adquisición de la base de datos DooDB, que fue capturada considerando sólo un lapso de dos semanas entre las sesiones.

- Se ha observado que la variabilidad inter-sesión es uno de los principales factores para el aumento en las tasas de error en la autenticación en el caso de los *passwords* gráficos. Las técnicas de actualización de plantillas de usuario (Didaci *et al.*, 2014; Uludag *et al.*, 2004) podrían ayudar a mitigar este problema.

- Como hemos visto, el escenario de adquisición (por ejemplo, *smartphone*, tableta) afecta a la tasa de error en la autenticación. Hay en general un compromiso entre la facilidad de uso y el rendimiento. Algunos trabajos recientes han estudiado el impacto de las condiciones de usabilidad y de adquisición en la verficación de firma (Blanco-Gonzalo *et al.*, 2014, 2013b; Brockly *et al.*, 2014). Se puede llevar a cabo más investigación en este área, con el fin de establecer características de usabilidad medibles y analizar su correlación con las tasas de error de los sistemas.

- Esta Tesis se ha basado en técnicas de comparación de similitud (*matching*) bien establecidas (HMM, GMM, DTW y medidas de distancia). Otros enfoques complementarios y populares en los últimos años, tales como *Support Vector Machines* (SVM) (Ferrer *et al.*, 2005), redes neuronales profundas y modelos híbridos HMM/red neuronal (Dahl *et al.*, 2012; Povey *et al.*, 2011) deberían ser explorados con el fin de identificar otros algoritmos de verificación con buen rendimiento.

References

A. M. Albert, K. Ricanek, and E. Patterson. A review of the literature on the aging adult skull and face: implications for forensic science research and applications. *Forensic Science International*, 172:1–9, 2007. 25

F. Alonso-Fernandez, J. Fierrez, A. Gilperez, and J. Ortega-Garcia. Impact of time variability in off-line writer identification and verification. In *Proc. of Intl. Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 16–18, 2009. 3

F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia. Dealing with sensor interoperability in multi-biometrics: the UPM experience at the BioSecure Multimodal Evaluation 2007. In *Proc. of SPIE Defense and Security Symposium, Biometric Technologies for Human Identification (BTHI)*, volume 6944, 2008. 31

F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. Sensor interoperability and fusion in signature verification: a case study using Tablet PC. In *Proc. of Intl. Workshop on Biometric Recognition Systems (IWBRS)*, LNCS-3781, pages 180–187. Springer, 2005. 24, 109, 118

M. Ammar, Y. Yoshida, and T. Fukumura. Structural description and classification of signature images. *Pattern Recognition*, 23(7):697–710, 1990. 6

J. Angulo and E. Waestlund. Exploring touch-screen biometrics for user identification on smart phones. In *Privacy and Identity Management for Life*, volume 375 of *IFIP Advances in Information and Communication Technology*, pages 130–143. 2012. 34, 35

E. Anquetil and H. Bouchereau. Integration of an on-line handwriting recognition system in a smart phone device. In *Proc. of 16th Intl. Conf. on Pattern Recognition (ICPR)*, volume 3, pages 192–195, 2002. 7

ANSI-INCITS 395-2005. ANSI-INCITS 395-2005: Information technology - biometric data interchange formats: Signature/sign data., 2005. 6, 108, 116

ANSI/NIST. NIST ITL American National Standards for Biometrics. http://fingerprint.nist.gov/standard/, 2009. (Accessed February 2015). 111

E. Argones Rua and J. Alba Castro. Online signature verification based on generative models. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics,*, 42(4):1231–1242, 2012. 20, 21

E. Argones Rua, E. Maiorana, J. Alba Castro, and P. Campisi. Biometric template protection using universal background models: An application to online signature. *IEEE Trans. on Information Forensics and Security*, 7(1):269–282, 2012. 8

A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proc. of 4th USENIX Conf. on Offensive Technologies*, pages 1–7, 2010. 36

R. Ballagas, J. Borchers, M. Rohs, and J. Sheridan. The smart phone: a ubiquitous input device. *IEEE Pervasive Computing*, 5(1):70–77, 2006. 7

BBfor2. BBfor2: Bayesian Biometrics for Forensics, FP7-ITN-2008-238803. http://www.bbfor2.net/, 2010. (Accessed February 2015). 111

BC. Biometrics consortium. http://www.biometrics.org/, 2015. (Accessed February 2015). 111

BI. Biometrics institute. http://www.biometricsinstitute.org/, 2015. (Accessed February 2015). 111

R. Biddle, S. Chiasson, and P. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Comput. Surv.*, 44(4):19:1–19:41, 2012. 2, 9, 10, 31, 33

Biosecure. Biometrics for Secure Authentication, FP6 NoE IST-2002-507634. http://biosecure.it-sudparis.eu/AB/, 2004. (Accessed February 2015). 31, 111

R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and E. Bella-Pulgarin. Automatic usability and stress analysis in mobile biometrics. *Image and Vision Computing*, 32(12):1173 – 1180, 2014. 110, 119

R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez. Performance evaluation of handwritten signature recognition in mobile environments. *IET Biometrics*, 3 (3):139–146, 2013a. 2, 25, 110, 118

R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez. Usability analysis of dynamic signature verification in mobile environments. In *2013 Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2013b. 2, 24, 110, 119

V. L. Blankers, C. E. van den Heuvel, K. Y. Franke, and L. G. Vuurpijl. The ICDAR 2009 signature verification competition. In *Proc. of 10th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 1403–1407, 2009. vii, 6, 108, 112, 114

BMEC. BioSecure Multimodal Evaluation Campaign, BMEC 2007 - Mobile Scenario, 2007. URL `http://biometrics.it-sudparis.eu/BMEC2007/files/Results_mobile.pdf`. 6, 24, 31, 112

L. Bovino, S. Impedovo, G. Pirlo, and L. Sarcinella. Multi-expert verification of hand-written signatures. In *Proc. of 7th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 932–936, 2003. 18

J. Brault and R. Plamondon. A complexity measure of handwritten curves: modeling of dynamic signature forgery. *IEEE Transactions on Systems, Mand, and Cybernetics*, 23:400–413, 1993. 26

M. Brockly, S. Elliott, J. Burdine, M. Frost, M. Riedle, and R. Guest. An investigation into biometric signature capture device performance and user acceptance. In *Proc. of Intl. Carnahan Conf. on Security Technology (ICCST)*, pages 1–5, 2014. 110, 119

BSEC. BioSecure Signature Evaluation Campaign, BSEC 2009, 2009. URL `http://biometrics.it-sudparis.eu/BSEC2009/downloads/BSEC2009_results.pdf`. VII, 9, 59, 107, 114

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. 2

J. W. Carls. *A framework for analyszing biometric template aging and renewal prediction*. PhD thesis, US Air Force Institute of Technology, 2009. 25

E. Catmull and R. Rom. A class of local interpolating splines. *Computer aided geometric design, Academic Press*, pages 317–326, 1974. 41

J. Chen, D. Lopresti, and F. Monrose. Toward resisting forgery attacks via pseudo-signatures. In *Proc. of 10th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 51–55, 2009. 34, 35

S. R. Coleman and R. Grover. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Anesthetic Surgery Journal*, 26:4–9, 2006. 25

G. Dahl, D. Yu, L. Deng, and A. Acero. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. on Audio, Speech, and Language Processing*, 20(1):30–42, 2012. 110, 119

A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and I know it's you!: implicit authentication based on touch screen patterns. In *Proc. of ACM annual conference on Human Factors in Computing Systems (CHI)*, pages 987–996, 2012. 34, 35

L. Didaci, G. L. Marcialis, and F. Roli. Analysis of unsupervised template update in biometric recognition systems. *Pattern Recognition Letters*, 37:151 – 160, 2014. 110, 119

R. A. Dixon, D. Kurzman, and I. C. Friesen. Handwriting performance in younger and older adults: age, familiarity, and practice effects. *Psychology and Aging*, 8:360–370, 1993. 78

G. Dobry, R. Hecht, M. Avigal, and Y. Zigel. Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal. *IEEE Trans. on Audio, Speech and Language Processing*, 19:1975–1985, 2011. 26

G. Doddington, W. Liggett, A. Martin, M. Pryzbocki, and D. Reynolds. Sheep, goats, lambs, and wolves: a statistical analysis of speaker performance in the NIST 1998 Speaker Recognition Evaluation. In *Proc. of Intl. Conf. on Spoken Language Processing (ICSLP)*, pages 1–5, 1998. 3, 26

J. G. A. Dolfing. *Handwriting Recognition and Verification, a Hidden Markov Approach*. PhD thesis, Technical University of Eindhoven, 1998. 20

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 14th Intl. Conf. on Pattern Recognition (ICPR)*, pages 1309–1312, 1998. xiii, 18, 19, 20, 21, 27, 28, 65

N. Drempt, A. McCluskey, and N. A. Lannin. A review of factors that influence adult handwriting performance. *Australian Occupational Therapy Journal*, 58:321–328, 2011. 25

R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, 2001. 20, 47

B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacretaz, A. Humm, F. Evequoz, R. Ingold, and D. V. Rotz. MyIDea - multimodal biometrics database, description of acquisition protocols. In *Proc. of 3rd COST 275 Workshop (COST 275)*, pages 59–62, 2005. 30

P. Dunphy and J. Yan. Do background images improve ”draw a secret” graphical passwords? In *Proc. of 14th ACM Conf. on Computer and Communications Security (CCS)*, pages 36–47, 2007. 33, 34

EAB, 2015. European Association for Biometrics. (http://www.eab.org). 111

S. Elliot. Differentiation of signature tratis vis-à-vis mobile- and table-based digitizers. *ETRI Journal*, 26(6):641–646, 2004. 24

M. Erbilek and M. Farihurst. Framework for managing ageing effects in signature biometrics. *IET Biometrics*, 1:136–147, 2012. 25

Face Aging Group. http://www.faceaginggroup.com/, 2004. (Accessed February 2015). 25

M. Fairhurst, editor. *Age Factors in Biometric Processing*. The Institution for Engineering and Technology, IET, 2013. 2, 110, 118

M. Fairhust. Editorial: First issue editorial. *IET Biometrics*, 1(1):1–2, 2012. 111

M. Faundez-Zanuy. On-line signature recognition based on VQ-DTW. *Pattern Recognition*, 40 (3):981 – 992, 2007. 23

M. Faundez-Zanuy, E. Sesa-Nogueras, and J. Roure-Alcobe. On the relevance of aging in handwritten biometric recognition. In *IEEE Intl. Carnahan Conf. on Security Technology*, pages 105–109, 2012. 25

M. A. Ferrer, J. B. Alonso, and C. M. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):993–997, 2005. 110, 119

J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano-Rey, G. G. de Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardenoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Urunuela, F. Martinez-Contreras, and J. J. Gracia-Roche. BiosecurID: A multimodal biometric database. *Pattern Analysis & Applications*, 13(2):235–246, 2010. 30, 61, 62, 115

J. Fierrez and J. Ortega-Garcia. *Advances in biometrics: sensors, systems and algorithms*, chapter Function-based on-line signature verification. Springer, 2007. 9

J. Fierrez and J. Ortega-Garcia. *Handbook of Biometrics*, chapter On-line signature verification, pages 189–209. Eds. A. K. Jain and A. Ross and P. Flynn, Springer, 2008. 1, 6, 17, 36, 95, 97, 106, 112, 117

J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez. BioSec baseline corpus: A multimodal biometric database. *Pattern Recognition*, 40:1389–1392, 2007a. 61

J. Fierrez, D. Ramos-Castro, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. HMM-based on-line signature verification: feature extraction and signature modeling. *Pattern Recognition Letters*, 28(16):2325–2334, 2007b. xiii, xiii, 18, 19, 20, 21, 22, 29, 30, 43, 44, 50, 56, 75, 114

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 *Proc. of Intl. Workshop on Biometric Authentication (BIOAW)*, pages 295–306. Springer LNCS-3087, 2004. 6

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, 2005a. xvii, 18, 19, 37, 38, 41, 42

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, 2005b. 18, 44, 81

L. Findlater, J. O. Wobbrock, and D. Wigdor. Typing on flat glass: Examining ten-finger expert typing patterns on touch surfaces. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pages 2453–2462, 2011. 2

M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Trans. on Information Forensics and Security*, 8(1):136–148, 2013. 34, 35

M. Freire-Santos, J. Fierrez-Aguilar, and J. Ortega-Garcia. Cryptographic key generation using handwritten signature. In *Proc. SPIE*, volume 6202, pages 225–231, 2006. 8

H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Fingerprint image quality estimation and its application to multialgorithm verification. *IEEE Trans. on Information Forensics and Security*, 3(2):331–338, 2008. 19

J. Galbally, J. Fierrez, M. R. Freire, and J. Ortega-Garcia. Feature selection based on genetic algorithms for on-line signature verification. In *Proc. of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 198–203, 2007a. 37, 39

J. Galbally, J. Fierrez, M. Martinez-Diaz, and J. Ortega-Garcia. Evaluation of brute-force attacks to dynamic signature verification using synthetic samples. In *Proc. of 10th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 131–135, 2009a. 27

J. Galbally, J. Fierrez, M. Martinez-Diaz, and J. Ortega-Garcia. Improving the enrollment in dynamic signature verification with synthetic samples. In *Proc. of 10th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 1295–1299, 2009b. 26

J. Galbally, J. Fierrez, and J. Ortega-Garcia. Bayesian hill-climbing attack and its application to signature verification. In *Proc. of IAPR Intl. Conf. on Biometrics (ICB)*. Springer LNCS 4642, 2007b. 8, 19, 27

J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Plamondon. Synthetic on-line signature generation. Part II: Experimental validation. *Pattern Recognition*, 45(7):2622 – 2632, 2012a. 26

J. Galbally, M. Martinez-Diaz, and J. Fierrez. Aging in biometrics: An experimental analysis on on-line signature. *PLoS ONE*, 8(7):e69897, 2013. 1, 61, 115

J. Galbally, R. Plamondon, J. Fierrez, and J. Ortega-Garcia. Synthetic on-line signature generation. Part I: Methodology and algorithms. *Pattern Recognition*, 45(7):2610 – 2621, 2012b. 26

H. Gao, X. Guo, X. Chen, L. Wang, and X. Liu. YAGP: Yet another graphical password strategy. In *Proc. of Annual Computer Security Applications Conference (ACSAC)*, pages 121–129, 2008. 34, 35

S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L.-L. Jardins, J. Lanter, Y. Ni, and D. Petrovska-Delacretaz. 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. 28

X. Geng, Z. H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging pattern. *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, 29:2234–2240, 2007. 25

J. Goldberg, J. Hagman, and V. Sazawal. Doodling our way to better authentication. In *Proc. of CHI '02 extended abstracts on Human factors in computing systems*, pages 868–869, 2002. 33, 34, 83

N. S. Govindarajulu and S. Madhvanath. Password management using doodles. In *Proc. of 9th Intl. Conf. on Multimodal Interfaces (ICMI)*, pages 236–239, 2007. 33, 34, 83, 92

R. Guest. Age dependency in handwritten dynamic signature verification systems. *Pattern Recognition Letters*, 27:1098–1104, 2006. 25, 26

J. K. Guo, D. Doermann, and A. Rosenfeld. Local correspondence for detecting random forgeries. In *Proc. of 4th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 319–323, 1997. 6

D. S. Guru and H. N. Prakash. Online signature verification and recognition: An approach based on symbolic representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31 (6):1059–1073, 2009. 18

M. Hasan, M. McLaren, H. V. Hamme, and D. V. Leeuwen. Age estimation from telephone speech using i-vectors. In *Proc. of InterSpeech*, pages 1–4, 2012. 26

N. Houmani, A. M. S. Garcia-Salicetti, B. Dorizzi, M. I. Khalil, M. N. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. R. Alcobé, J. Fabregas, M. Faundez-Zanuy, J. M. Pascual-Gaspar, V. Cardeñoso-Payo, and C. Vivaracho-Pascual. BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures. *Pattern Recognition*, 45 (3):993–1003, 2012. vii, xviii, 2, 6, 9, 14, 23, 24, 48, 51, 59, 66, 81, 92, 102, 105, 106, 107, 109, 112, 114, 117

N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. A novel personal entropy measure confronted with online signature verification systems' performance. In *Proc. of 2nd IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2008. 2, 24, 26

N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. On assessing the robustness of pen coordinates, pen pressure and pen inclination to short-term and long-term time variability with personal entropy. In *Proc. of 3rd Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2009. 3, 8, 19, 26, 56, 60, 78, 114

N. Houmani, S. Garcia-Salicetti, B. Dorizzi, J. Montalvao, J. Canuto, M. Andrade, Y. Qiao, X. Wang, T. Scheidat, A. Makrushin, D. Muramatsu, J. Putz-Leszczynska, M. Kudelski, M. Faundez-Zanuy, J. Pascual-Gaspar, V. Cardenoso-Payo, C. Vivaracho-Pascual, E. Argones Riia, J. Alba-Castro, A. Kholmatov, and B. Yanikoglu. BioSecure signature evaluation campaign (ESRA'2011): evaluating systems on quality-based categories of skilled forgeries. In *Proc. of Intl. Joint Conf. on Biometrics (IJCB)*, pages 1–10, Oct 2011. 6, 112

K. Huang and H. Yan. Off-line signature verification based on geometric feature extraction and neural network classification. *Pattern Recognition*, 30(1):9–19, 1997. 6

K. Huang and H. Yan. Stability and style-variation modeling for on-line signature verification. *Pattern Recognition*, 36(10):2253–2270, 2003. 18, 26

ICB. Proc. of 8th IAPR International Conference on Biometrics. IEEE Press (to appear), 2015. 111

IJCB. Proc. of International Joint Conference on Biometrics. IEEE Press, 2014. 111

D. Impedovo and G. Pirlo. Automatic signature verification: The state of the art. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(5):609–635, 2008. 1, 2, 6, 112

D. Impedovo, G. Pirlo, and R. Plamondon. Handwritten signature verification: New advancements and open issues. In *Proc. of Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, pages 367–372, 2012. 1, 2, 6, 112

ISO/IEC 19794-11. ISO/IEC 19794-11:2005, information technology - biometric data interchange formats - part 11: Signature/sign processed dynamic data, 2005. 6, 108, 116

ISO/IEC 19794-7. ISO/IEC 19794-7:2005, information technology - biometric data interchange formats - part 7: Signature/sign time series data, 2005. 6, 108, 116

A. K. Jain, F. Griess, and S. Connell. On-line signature verification. *Pattern Recognition*, 35 (12):2963–2972, 2002. 9, 81

A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005. 43

A. K. Jain, A. Ross, and P. Flynn, editors. *Handbook of Biometrics*. Springer, 2008. 1, 2, 4, 111

A. K. Jain, A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer, 2011. 12, 111

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, 2004. xvii, 4, 6

A. K. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997. 19, 36, 37, 39

I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. The design and analysis of graphical passwords. In *Proc. of 8th USENIX Security Symposium*, 1999. 33, 34, 83

R. S. Kashi, J. Hu, W. L. Nelson, and W. Turin. On-line handwritten signature verification using Hidden Markov Model features. In *Proc. of 4th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, volume 1, pages 253–257, 1997. 21

Y. Kato, D. Muramatsu, and T. Matsumoto. A sequential Monte Carlo algorithm for adaptation to intersession variability in on-line signature verification. In *Proc. of 10th Intl. Workshop on Frontiers in Handwriting Recognition (IWFHR)*, 2006. 26

C. J. Ketcham, N. V. Dounskaia, and G. E. Stelmach. Control of multijoint drawing movements: a comparison of young and elderly adults. In *Proc. of Conf. of the International Graphonomics Society (IGS)*, pages 42–44, 2003. 25

R. Khokhar. Smartphones — a call for better safety on the move. *Network Security*, 2006(4): 6–7, 2006. 8

A. Kholmatov and B. Yanikoglu. Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408, 2005. 9, 18, 19, 23, 60

A. Kholmatov and B. Yanikoglu. SUSIG: an on-line signature database, associated protocols and benchmark results. *Pattern Analysis & Applications*, 12(3):227–236, 2008. 29

S. H. Kim, M. S. Park, and J. Kim. Applying personalized weights to a feature set for on-line signature verification. In *Proc. of 3rd Intl. Conf. on Document Analysis and Recognition (ICDAR)*, volume 2, pages 882–885, 1995. 37

J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998. 104

Y. H. Kwon and N. V. Lobo. Age classification from facial images. *Computer Vision and Image Understanding*, 74:1–21, 1999. 25

K. Lai, J. Konrad, and P. Ishwar. Towards gesture-based user authentication. In *Proc. of IEEE 9th Intl. Conf. on Advanced Video and Signal-Based Surveillance (AVSS)*, pages 282–287, 2012. 110

A. Lanitis. Evaluating the performance of face-aging algorithms. In *Proc. of IEEE Intl. Conf. on Automatic Face and Gesture Recognition (ICAFGR)*, pages 1–6, 2008. 25

A. Lanitis. A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, 2:34–52, 2010. 3, 25

A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. on Systems, Man and Cybernetics, Part B: Cybernetics*, 34: 621–628, 2004. 25

A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, 24:442–455, 2002. 25

F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art-1989-1993. *Intl. Journal of Pattern Recognition and Artificial Intelligence*, 8(3):643–660, 1994. 6

L. L. Lee, T. Berger, and E. Aviczer. Reliable on-line human signature verification systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(6):643–647, 1996. 18, 19, 37, 41

H. Lei and V. Govindaraju. A comparative study on the consistency of features in on-line signature verification. *Pattern Recognition Letters*, 26(15):2483–2489, 2005. 18, 19, 43, 52, 53

H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. A study of face recognition as people age. In *Proc. of IEEE Intl. Conf. on Computer Vision (ICCV)*, pages 1–8, 2007. 25

M. Liwicki, M. Malik, C. van den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, and B. Found. Signature verification competition for online and offline skilled forgeries (SigComp2011). In *Proc. of 11th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 1480–1484, 2011. 6, 112

J. Llados, E. Valveny, G. Sanchez, and E. Marti. Symbol recognition: Current advances and perspectives. In *Graphics Recognition Algorithms and Applications, Proc. of 4th Intl. Workshop on Graphics Recognition (GREC)*, Lecture Notes in Computer Science, pages 104–128. Springer LNCS 2390, 2002. 36

A. Lumini and L. Nanni. A clustering method for automatic biometric template selection. *Pattern Recognition*, 39:495–497, 2006. 26

B. Ly-Van, S. Garcia-Salicetti, and B. Dorizzi. On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 37(5):1237 – 1247, 2007. 9, 18, 19, 20, 21, 43, 56, 65, 114

E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia, and A. Neri. Template protection for HMM-based on-line signature authentication. In *Proc. of IEEE Computer Society Workshop on Biometrics (CVPR)*, 2008. 8

M. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, and B. Found. ICDAR 2013 competitions on signature verification and writer identification for on- and offline skilled forgeries (SigWiComp 2013). In *Proc. of 12th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 1477–1483, 2013. 6, 112

R. Martens and L. Claesen. Dynamic programming optimisation for on-line signature verification. In *Proc. of 4th. Intl. Conf. on Document Analysis and Recognition (ICDAR)*, volume 2, pages 653 – 656, 1997. 17, 18, 19, 23

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. 4

M. Martinez-Diaz and J. Fierrez. *Encyclopedia of Biometrics*, chapter Signature databases and evaluation, pages 1178–1184. Springer, 2009. 2, 23

M. Martinez-Diaz, J. Fierrez, M. R. Freire, and J. Ortega-Garcia. On the effects of sampling rate and interpolation in HMM-based dynamic signature verification. In *Proc. of 9th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, volume 2, pages 1113–1117, 2007a. 17

M. Martinez-Diaz, J. Fierrez, and J. Galbally. The DooDB graphical password database: Data analysis and benchmark results. *IEEE Access*, 1:596–605, 2013. xv, 10, 84, 97, 98, 100, 105, 116

M. Martinez-Diaz, J. Fierrez, J. Galbally, F. Alonso-Fernandez, and J. Ortega-Garcia. Signature verification on handheld devices. In *Proc. of MADRINET Workshop*, pages 87–95, 2007b. 7

M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia. Towards mobile authentication using dynamic signature verification: useful features and performance evaluation. In *Proc. of Intl. Conf. on Pattern Recognition (ICPR)*, pages 1–6, 2008a. 49, 113

M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Encyclopedia of Biometrics*, chapter Signature features, pages 1185–1192. Springer, 2009a. 17

M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Encyclopedia of Biometrics*, chapter Signature matching, pages 1192–1196. Springer, 2009b. 18, 75, 88, 92

M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally. Mobile signature verification: Feature robustness and performance comparison. *IET Biometrics*, 3(4):267–277, 2014. 49, 65, 113

M. Martinez-Diaz, J. Fierrez, C. Martin-Diaz, and J. Ortega-Garcia. DooDB: a graphical password database containing doodles and pseudo-signatures. In *Proc. of Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, pages 339–344, 2010a. 84, 87, 116

M. Martinez-Diaz, J. Fierrez, and J. Ortega-Garcia. Universal Background Models for dynamic signature verification. In *Proc. of IEEE Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2007c. 18, 19

M. Martinez-Diaz, J. Fierrez, and J. Ortega-Garcia. Incorporating signature verification on handheld devices with user-dependent Hidden Markov Models. In *Proc. of Intl. Conf. on Frontiers in Hadwriting Recognition (ICFHR)*, 2008b. 14

M. Martinez-Diaz, J. Fierrez, and J. Ortega-Garcia. *Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability*, chapter Automatic Signature Verification on Handheld Devices, pages 321–338. IGI Global, May 2009c. 7

M. Martinez-Diaz, C. Martin-Diaz, J. Galbally, and J. Fierrez. A comparative evaluation of finger-drawn graphical password verification methods. In *Proc. of Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, pages 375–380, 2010b. 98

S. K. Modi and S. J. Elliott. Impact of image quality on performance: age comparison of young and elderly fingerprints. In *Proc. of Intl. Conf. on Recent Advances in Soft Computing (ICRASC)*, pages 10–12, 2006. 25

S. K. Modi, S. J. Elliott, and K. Hakil. Impact of age groups on fingerprint recognition performance. In *Proc. of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 19–23, 2007. 25

M. Morgan, J. L. Bradshaw, J. G. Phillips, J. B. Mattingley, R. Iansek, and J. A. Bradshaw. Effects of hand and age upong abductive and adductive movements – a kinematic analysis. *Brain Cognition*, 25:194–206, 1994. 25

P. B. Mueller. The aging voice. *Seminars in Speech and Language*, 18, 1997. 25

M. E. Munich and P. Perona. Visual identification by signature tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(2):200–217, 2003. 17, 27

D. Muramatsu and T. Matsumoto. An HMM signature verifier incorporating signature trajectories. In *Proc. of 7th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, volume 1, pages 438–442, 2003. 20

D. Muramatsu and T. Matsumoto. Effectiveness of pen pressure, azimuth, and altitude features for online signature verification. In *Proc. of IAPR Intl. Conf. on Biometrics (ICB)*, pages 503–512. Springer LNCS 4642, 2007. 8, 19, 56, 114

D. Muramatsu, K. Yasuda, and T. Matsumoto. Biometric person authentication method using camera-based online signature acquisition. In *Proc. of 10th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 46–50, 2009. 17

I. T. Nabney. *NETLAB: algorithms for pattern recognition*. Advances in Pattern Recognition. Springer, 2002. 44

L. Nanni and A. Lumini. Ensemble of Parzen Window classifiers for on-line signature verification. *Neurocomputing*, 68:217–224, 2005. 18

W. Nelson and E. Kishon. Use of dynamic features for signature verification. In *Proc. of IEEE Intl. Conf. on Systems, Man, and Cybernetics*, volume 1, pages 201–205, 1991. 41

W. Nelson, W. Turin, and T. Hastie. Statistical methods for on-line signature verification. *Intl. Journal of Pattern Recognition and Artificial Intelligence*, 8(3):749–770, 1994. 18, 41

M. Oka, K. Kato, X. Yingqing, L. Liang, and F. Wen. Scribble-a-secret: Similarity-based password authentication using sketches. In *Proc. of Intl. Conf. on Pattern Recognition (ICPR)*, pages 1–4, 2008. 34, 35, 83

C. O'Reilly and R. Plamondon. Development of a sigma–lognormal representation for on-line signatures. *Pattern Recognition*, 42(12):3324 – 3337, 2009. 26

C. O'Reilly and R. Plamondon. Design of a neuromuscular disorders diagnostic system using human movement analysis. In *Proc. of Information Science, Signal Processing and their Applications (ISSPA)*, pages 787–792, 2012. 25

J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, *et al.* The multi-scenario multi-environment biosecure multimodal database (BMDB). *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(6):1097–1111, 2010. xv, 24, 31, 49, 50, 61, 62, 85, 86, 113, 115

J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. MCYT baseline corpus: a bimodal biometric database. *IEE Proc. Vision, Image and Signal Processing*, 150 (6):391–401, 2003. 28, 31

P. C. van Oorschot and J. Thorpe. On predictive models and user-drawn graphical passwords. *ACM Trans. on Information and System Security*, 10(4):1–33, 2008. 36

B. Paltridge. Thesis and dissertation writing: an examination of published advice and actual practice. *English for Specific Purposes*, 21(2):125 – 143, 2002. 11

F. Parada-Loira, E. Gonzalez-Agulla, and J. L. Alba-Castro. Hand gestures to control infotainment equipment in cars. In *Proc. of IEEE intelligent Vehicles Symposium (IV)*, pages 1–6, 2014. 110

M. Parizeau and R. Plamondon. A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 12(7):710–717, 1990. 18

U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32:947–954, 2010. 26

J. M. Pascual-Gaspar, V. Cardenoso-Payo, and C. E. Vivaracho-Pascual. Practical on-line signature verification. In *Proc. of 3rd Intl. Conf. on Biometrics (ICB)*, pages 1180–1189. Springer LNCS-5558, 2009. 23

R. Plamondon and G. Lorette. Automatic signature verification and writer identification: the state of the art. *Pattern Recognition*, 22(2):107–131, 1989. 1, 6, 7, 17, 112

R. Plamondon, C. O'Reilly, J. Galbally, A. Almaksour, and E. Anquetil. Recent developments in the study of rapid human movements with the kinematic theory: Applications to handwriting and signature synthesis. *Pattern Recognition Letters*, 35:225 – 235, 2014. 26

R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: a comprehensive survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22:63–84, 2000. 6

N. Poh, S. Bengio, and A. Ross. Revisiting Doddington's zoo: A systematic method to assess user-dependent variabilities. In *Proc. of Workshop on Multimodal User Authentication (MMUA)*, pages 1–7, 2006. 26

D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. The Kaldi speech recognition toolkit. In *Proc. of IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, 2011. 110, 119

P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994. 19, 38

C. Rabasse, R. Guest, and M. Fairhurst. A new method for the synthesis of signature data with natural variability. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38 (3):691–699, 2008. 26

L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. 20, 21

N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Trans. on Image Processing*, 15:3349–3361, 2006. 25

N. K. Ratha and V. Govindaraju, editors. *Advances in biometrics: Sensors, algorithms and systems*. Springer, 2008. 111

A. Rattani, B. Freni, G. L. Marcialis, and F. Roli. Template update methods in adaptive biometric systems: a critical review. In *Proc. of IAPR/IEEE Intl. Conf. on Biometrics (ICB)*, pages 847–856. Springer LNCS-5558, 2009. 25

A. Rattani, G. L. Marcialis, and F. Roli. Biometric template update using the graph mincut: a case study in face verification. In *Proc. of IEEE Biometric Symposium (BSYM)*, pages 23–28, 2008. 26

A. W. Rawls and K. Ricanek. MORPH: Development and optimization of a longitudinal age progression database. In *Proc. of COST 2101 Workshop on Biometrics and Identity Management (BIOID)*, pages 17–24. Springer LNCS-5707, 2009. 10, 25

K. Renaud. On user involvement in production of images used in visual authentication. *Journal of Visual Languages & Computing*, 20(1):1–15, 2009. 9, 33, 92

D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted Gaussian Mixture Models. *Digital Signal Processing*, 10:19–41, 2000. 20, 47

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. 18, 19, 20, 21, 47, 99

J. Richiardi, H. Ketabdar, and A. Drygajlo. Local and global feature selection for on-line signature verification. In *Proc. of 8th Intl. Conf. on Document Analysis and Recognition (ICDAR)*, pages 625 – 629, 2005. 18, 37, 43, 44, 56, 114

B. S. Riggan, W. E., Snyder, X. Wang, and J. Feng. A human factors study of graphical passwords using biometrics. In *Proc. of 36th German Conf. on Pattern Recognition (GCPR)*, pages 464–475. Springer LNCS 8753, 2014. 34, 35, 110, 119

L. Rodriguez-Liñares, C. Garcia-Mateo, and J. L. Alba-Castro. On combining classifiers for speaker authentication. *Pattern Recognition*, 36(2):347 – 359, 2003. 19

F. Roli, L. Didaci, and G. L. Marcialis. Template co-update in multimodal biometric systems. In *Proc. of IAPR/IEEE Intl. Conf. on Biometrics (ICB)*, pages 1194–1202. Springer LNCS-4642, 2007. 26

A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer, 2006. 18, 111

R. Sabourin. *Off-line signature verification: recent advances and perspectives*, volume 1339, chapter Lecture Notes in Computer Science, LNCS-1339, pages 84–98. Springer, 1997. 6

N. Sae-Bae and N. Memon. Online signature verification on mobile devices. *IEEE Trans. on Information Forensics and Security*, 9(6):933–947, 2014. 2, 18, 19

N. Sae-Bae, N. Memon, K. Isbister, and K. Ahmed. Multitouch gesture-based authentication. *IEEE Trans. on Information Forensics and Security*, 9(4):568–582, 2014. 2, 34, 35, 83

H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 26:43–49, 1978. 22

Y. Sato and K. Kogure. Online signature verification based on shape, motion and writing pressure. In *Proc. of 6th Intl. Conf. on Pattern Recognition (ICPR)*, pages 823–826, 1982. 18, 19, 23

SC37. ISO/IEC JTC 1/SC 37. http://www.iso.org/iso/jtc1_sc37_home/, 2005. (Accessed February 2015). 111

T. Scheidat, J. Heinze, C. Vielhauer, J. Dittmann, and C. Kraetzer. Comparative review of studies on aging effects in context of biometric authentication. In *Proc. of SPIE Intl. Conf. on Multimedia and Mobile Devices*, SPIE Vol. 7881, page 788110, 2011. 25

T. Scheidat, K. Kummel, and C. Vielhauer. Short term template aging effects on biometric authentication performance. In *Proc. of Intl. Conf on Communications and Multimedia Security*, Springer LNCS-7394, pages 107–116, 2012. 26

E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska. An information analysis of in-air and on-surface trajectories in online handwriting. *Cognitive Computation*, 4(2):195–205, 2012. 8, 60

M. Shahzad, A. X. Liu, and A. Samuel. Secure unlocking of mobile touch screen devices by simple gestures: You can see it but you can not do it. In *Proc. of 19th Annual Intl. Conf. on Mobile Computing and Networking*, pages 39–50, 2013. 34, 35

D. Simsons, R. Spencer, and S. Auer. The effects of constraining signatures. *Journal of the American Society of Questioned Document Examiners*, 14(1):39–50, 2011. 24

J. L. Suo, F. Min, S. C. Zhu, S. G. Shan, and X. L. Chen. A multi-resolution dynamic model for face aging. In *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition (ICCVPR)*, pages 1–8, 2007. 25

X. Suo, Y. Zhu, and G. Owen. Graphical passwords: a survey. In *Proc. of 21st Annual Computer Security Applications Conf. (ACSAC)*, pages 463–472, 2005. 9

Tabula Rasa. TABULA RASA: Trusted Biometrics under Spoofing Attacks, FP7-ICT-257289. http://www.tabularasa-euproject.org, 2010. (Accessed February 2015). 111

H. Tao and C. Adams. Pass-go: A proposal to improve the usability of graphical passwords. *International Journal of Network Security*, 7(2):273–292, 2008. 33, 34

S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2006. 36, 37, 38, 113

U. Uludag, A. Ross, and A. Jain. Biometric template selection and update: a case study in fingerprints. *Pattern Recognition*, 37(7):1533–1542, 2004. 26, 110, 119

C. Varenhorst. Passdoodles; a lightweight authentication method. Technical report, Research Science Institute, Massachusetts Institute of Technology, 2004. 33, 34

J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso. Off-line signature verification based on grey level information using texture features. *Pattern Recognition*, 44(2):375 – 385, 2011. 6

R. Vera, R. Tolosana, J. Ortega-Garcia, and J. Fierrez. e-BioSign: Stylus- and finger-input multi-device database for dynamic signature recognition. In *Proc. of Intl. Workshop on Biometrics and Forensics (IWBF)*, 2015. 110, 118

G. Veres, M. Nixon, and J. Carter. Model-based approaches for predicting gait changes over time. In *Proc. of IEEE Intl. Conf. on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 325–330, 2005. 26

B. Vijaya-Kumar, S. Prabhakar, and A. Ross, editors. *Proc. of 7th SPIE Conf. on Biometric Technology for Human Identification (BTHI VII)*, volume 7667, 2010. SPIE. 111

C. Vivaracho-Pascual and J. Pascual-Gaspar. On the use of mobile phones and biometrics for accessing restricted web services. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 42(2):213–222, 2012. 2, 7

J. Walton. Handwriting changes due to aging and Parkinson's syndrome. *Forensic Science International*, 88:197–214, 1997. 25, 78

D. Wang, Y. Zhang, C. Yao, J. Wu, H. Jiao, and M. Liu. Toward force-based signature verification: A pen-type sensor and preliminary validation. *IEEE Trans. on Instrumentation and Measurement*, 59(4):752–762, 2010. 17

J. Wayman, A. Jain, D. Maltoni, and D. Maio. *Biometric Systems. Technology, Design and Performance Evaluation*. Springer, 2005. 2

R. Weiss and A. D. Luca. PassShapes: Utilizing stroke based authentication to increase password memorability. In *Proc. of 5th Nordic Conference on Human-computer Interaction: Building Bridges (NordiCHI)*, pages 383–392, 2008. 34, 35

W. Wijesoma, M. Mingming, and E. Sung. Selecting optimal personalized features for online signature verification using ga. In *Proc. of IEEE Intl. Conf. on Systems, Man, and Cybernetics*, volume 4, pages 2740–2745, 2000. 37

J. Wu, P. Ishwar, and J. Konrad. The value of posture, build and dynamics in gesture-based user authentication. In *Proc. of IEEE Intl. Joint Conf. on Biometrics (IJCB)*, pages 1–8, 2014. 110

L. Yang, B. K. Widjaja, and R. Prasad. Application of Hidden Markov Models for signature verification. *Pattern Recognition*, 28(2):161–170, 1995. 19, 20, 21

M. Yasuhara and M. Oka. Signature verification experiment based on nonlinear time alignment: a feasibility study. *IEEE Trans. on Systems, Man and Cybernetics, part C*, 12(3):212–216, 1977. 22

D. Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First international signature verification competition. In *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, pages 16–22. Springer LNCS-3072, 2004. 6, 23, 24, 29, 44, 48, 50, 112

S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, *et al.* *The HTK Book (for HTK Version 3.4)*. Cambridge University Engineering Department, revised for HTK version 3.4 edition, 2009. 44

W. Zada Khan, M. Y. Aalsalem, and Y. Xiang. A graphical password based system for small mobile devices. *Intl. Journal of Computer Science Issues*, 8(5):145–154, 2011. 34, 35

N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan. Shoulder surfing defence for recall-based graphical passwords. In *Proc. of 7th Symposium on Usable Privacy and Security (SOUPS)*, volume 6, pages 1–12, 2011. 36

X. Zhao, T. Feng, W. Shi, and I. Kakadiaris. Mobile user authentication using statistical touch dynamics images. *IEEE Trans. on Information Forensics and Security*, 9(11):1780 − 1789, 2014. 2