UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA Y DE LAS COMUNICACIONES

Disruptive Approaches to Handwriting and Signature Authentication for Security-Enhanced Schemes

–TESIS DOCTORAL–

APROXIMACIONES DISRUPTIVAS PARA LA MEJORA DE SISTEMAS DE AUTENTICACIÓN BASADOS EN FIRMA Y ESCRITURA MANUSCRITA

Author: Rubén Tolosana Moranchel
(Ingeniero de Telecomunicación).

A Thesis submitted for the degree of:

Doctor of Philosophy

Madrid, March 2019

Colophon

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

Copyright © 2019 by Ruben Tolosana Moranchel. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the author. Universidad Autonoma de Madrid has several rights in order to reproduce and distribute electronically this document.

This Thesis was printed with the financial support from EPS-UAM and the Biometrics and Data Pattern Analytics Laboratory - BiDA Lab.
contact: ruben.tolosana@uam.es

Department: Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid (UAM), SPAIN

PhD Thesis: Disruptive Approaches to Handwriting and Signature Authentication for Security-Enhanced Schemes

Author: **Rubén Tolosana Moranchel**
Ingeniero de Telecomunicación
(Universidad Autónoma de Madrid)

Advisors: **Rubén Vera Rodríguez**
Doctor Ingeniero de Telecomunicación
(Swansea University)
Universidad Autónoma de Madrid, SPAIN

Javier Ortega García
Doctor Ingeniero de Telecomunicación
(Universidad Politécnica de Madrid)
Universidad Autónoma de Madrid, SPAIN

Year: 2019

Committee: President: **Julián Fiérrez Aguilar**
Universidad Autónoma de Madrid, SPAIN

Secretary: **Juan Alberto Sigüenza Pizarro**
Universidad Autónoma de Madrid, SPAIN

Vocal 1: **Richard Guest**
University of Kent, UNITED KINGDOM

Vocal 2: **Marcos Faúndez Zanuy**
Universitat Pompeu Fabra, SPAIN

Vocal 3: **Donato Impedovo**
Università degli Studi di Bari, ITALY

BiDA Lab

The research described in this Thesis was carried out within the Biometrics and Data Pattern Analytics Laboratory - BiDA Lab at the Dept. of Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid (from 2014 to 2019). The project was partially funded by a FPU Fellowship from Spanish MECD.

The author was awarded with a PhD Fellowship "Formación de Profesorado Universitario (FPU)" from Spanish Ministerio de Educación, Cultura y Deporte between 2015 and 2019 which supported the research summarised in this Dissertation.

The author was awarded with a travel grant to the 7th IEEE International Workshop on Information Forensics and Security by the IEEE Signal Processing Society in November 2015.

The author was awarded with a travel and fees grant to the 12th International Summer School for Advanced Studies on Biometrics: Biometrics for Forensics, Security and Mobile Application by the European Commission, IAPR, MORPHO and the IEEE Biometrics Council in June 2015.

The author was awarded with a mobility grant from FPU Spanish MECD, which supported his research stay carried out at University of Kent, Canterbury, United Kingdom from September 2016 to December 2016.

The author was awarded with a mobility grant from BATL research project (IARPA), which supported his research stay carried out at da/sec - Biometrics and Internet-Security Research Group, Darmstadt, Germany from May 2018 to July 2018.

The author was awarded with the European Biometrics Industry Award 2018, from the European Association for Biometrics (EAB), for part of the research work originated from this Dissertation.

Abstract

Handwritten signature is one of the most socially accepted biometric traits as it has been used in financial and legal agreements for over a century. However, is signature biometric technology really adapted to current scenarios? With the massive deployment of mobile general purpose devices such as smartphones and tablets, new very interesting and user-friendly scenarios have appeared beyond the traditional office-like scenario considering high quality devices especifically designed for signature acquisition. In addition, despite the high technological evolution, and concretely, the success of deep learning techniques in combination with Graphics Processing Units (GPUs), the core of most of the state-of-the-art signature verification systems is still almost the same than 20 years ago. Why deep learning techniques do not outperform traditional systems as it happens in other fields?

The last motivation for this Thesis is related to password-based systems. Traditionally, the two most prevalent user authentication approaches have been Personal Identification Numbers (PIN) and One-Time Passwords (OTP). However, and despite the high popularity and deployment of PIN- and OTP-based authentication systems in real scenarios, many studies have highlighted the weaknesses of these approaches as they are very easy to guess or steal (i.e., through shoulder-surfing and smudge attacks). Is it possible to increase the security of these traditional authentication systems at the same time that we provide a good experience to the users?

As a way of finding the answers to these questions, this Thesis is mainly focused on the analysis of the new opportunities that bring up these novel scenarios and technologies and the challenges that must be tackled in order to achieve state-of-the-art results.

This Dissertation comprises five different parts. *Part I* first concentrates on the problem statement and main contributions of the Thesis. The experimental chapters are then divided into three parts, *Part II*, *Part III*, and *Part IV*. Lastly, *Part V* concludes the Thesis.

Part I first introduces the basics of biometrics, focusing on handwritten signature biometrics, which is the main topic of study in this Thesis, and the challenges and opportunities for it along an exhaustive overview of the state-of-the-art. Then, we concentrate on describing the most relevant features of existing on-line signature databases, making special emphasis on all the databases acquired during this Thesis. Finally, *Part I* concludes explaining first the specific details of the traditional on-line signature verification systems considered in the experimental parts of the Thesis, and then our novel end-to-end writer-independent RNN signature verification systems proposed in this Dissertation.

The first experimental part (*Part II* of this Dissertation) starts analysing the system performance of traditional signature verification systems on emerging scenarios such as finger input, device interoperability and mixed writing-input. Due to the high system performance degradation of them, in this Thesis we propose a two-stage approach based on robust preprocessing

and feature selection techniques. We then study the novel scenario where the number of stored samples or templates per user can grow very fast, making it possible to train more robust statistical user models, improving the performance of biometric systems, and in particular, reducing the template aging effect. The research carried out in this part aims to answer the following questions: How is the system performance affected on these novel scenarios? What approach should we consider to overcome these challenges?

In the second experimental part (*Part III* of this Dissertation) we propose new ways to improve traditional signature verification systems. Concretely, we first evaluate the potential of including deep learning technology through a new architecture (Siamese) more adapted to the signature verification task. We then focus on the concept of complexity in signature and enhance the traditional systems through the selection of the most robust features for each signature complexity level.

Finally, *Part IV* of this Dissertation evaluates the potential of incorporating handwriting biometric information to traditional authentication systems based on passwords, asking the user to draw each digit of the password on the touchscreen instead of typing them as usual.

The research carried out in this Dissertation has led to novel contributions which include: *i*) analysis and adaptation of on-line signature verification systems to emerging scenarios such as finger input, device interoperability and mixed writing-input through robust preprocessing and feature selection techniques, *ii*) an exhaustive experimental analysis of template update strategies for three popular on-line signature verification approaches, extracting various practical findings related to the template aging effect in signature biometrics, and configuring time-adaptive improved versions of the considered baseline approaches overcoming to some extent the template aging, *iii*) exploring the potential of deep learning approaches for on-line signature verification. We have proposed a novel end-to-end writer-independent on-line signature verification system based on Recurrent Neural Networks with a Siamese architecture, which has outperformed other state-of-the-art systems, *iv*) improvement of traditional signature verification systems through the incorporation of the signature complexity concept, *v*) enhancement of traditional PIN and OTP authentication systems through the incorporation of handwriting biometric information as a second level of user authentication, *vi*) acquisition of new unprecedented handwriting and signature databases and release of them to the research community, and *vii*) part of the research presented in this Thesis has been deployed successfully in a pilot project in which on-line signature verification will be used massively in the Spanish banking sector.

Don't ever let someone tell you
that you can't do something.
Not even me. You got a dream,
you gotta protect it.
When people can't do
something themselves,
they're gonna tell you
that you can't do it.
You want something,
go get it. Period.

The Pursuit of Happiness

Acknowledgements

Even though this section should be one of the easiest to write, it is not at all, at least for me. Many great people come to my mind. People who have made me like I'm: a funny, outgoing, optimistic, and hard-working person. People who have taught me how to face challenges and invest my time in what really matters. People... who have helped me to achieve this Dissertation and all my goals (never it is too much!). So, let's start expressing my gratitude step by step...

Foremost, I would like to thank Dr. Ruben Vera-Rodriguez, one of my first great teachers, my MSc and PhD Thesis co-advisor, and after all these years working and travelling together, my friend. It's been a great pleasure to learn from you, always helping me to see problems in a different way and saying yes to all my research proposals (even in your holidays). Of course, this Dissertation would not have been possible without the support of my co-advisor Prof. Javier Ortega-Garcia who one day asked me: "Tolosana, have you ever thought about doing the PhD degree?". Many thanks Javier for giving me this fantastic opportunity. Finally, and although his name does not appear in the front page, I would also like to thank Prof. Julian Fierrez, my "hidden" co-advisor, for all your time, advice (both in research and personal life), and... your 2080 RTX Ti GPUs. Thank you very much to each of you for all your time and great moments together.

This entire long journey would have been completely different without the excellent professionals and colleagues from the ATVS (now BiDA and AUDIAS). If you had asked me 5 years ago "what do you want to be in life?" My answer would probably have been "I want to be like them". First, I would like to thank Dr. Ruben Zazo, for being my partner on a daily basis since we started going to the university over 10 years ago (¿Una palmerita?), and Dr. Ester Gonzalez-Sosa, for being my PhD-sister, always supporting me every time I received unfair revisions (I still buy "Platanos de Canarias", I promise you!). My thanks also to Dr. Aythami Morales (for your research support and good talks at any time of the day), Dr. Ram Krish (for introducing me to the real programming world), Dr. Marta Gomez-Barrero (for teaching me since my PhD beginnings how things work in this field and accepting me in your new wonderful place, *Danke sehr!*), Dr. Alicia Lozano (for our special late evenings in the lab), Dr. Javier Franco and Dr. Daniel Ramos (always sharing the same philosophy than I do: life is much easier when you smile. Many thanks for all your jokes and good moments lived in the lab), Alejandro Acien (I wish you had lunch with us a single day), Javier Hernandez-Ortega (always surprising me), Juan Maroñas (calm down and trust you, you will go far my friend!). I would also like to thank Prof. Joaquin Gonzalez-Rodriguez and Dr. Doroteo Torre-Toledano for all their good deeds, interest and constructive reviews at "Cafes ATVS". My thanks also to all BSc and MSc young students of the group: Ivan Bartolome, Ignacio de la Serna, Alvaro Escudero, Adrian Garcia, Maria Pilar Fernandez, Marta Blazquez, Javier Gismero, Ruben Barco, Pablo Lazaro, Diego de Benito,

Beltran Labrador, Sara Gamiz, Natalia Delgado, Sandra Gaytan, and many more I'm sure I'm forgetting (sorry!). Finally, I would also like to thank my other colleagues from ATVS, amazing people that I have had the opportunity to meet them along the years: Dr. Javier Galbally, Dr. Javier Gonzalez, Dr. Pedro Tome, Dr. Ignacio Lopez-Moreno, Dr. Marcos Martinez-Diaz and Dr. Fernando Alonso-Fernandez. You will be always my source of inspiration. To all of you, many thanks for making the ATVS more than a simple research group.

But my knowledge not only came from the ATVS, also for many other sources. I would like to thank first to Prof. Richard Guest from University of Kent (Canterbury) for giving me in 2016 the opportunity to perform my first internship. It was the first time in my life I spent so much time abroad, and I have to say it was wonderful. Many thanks Richard for all your time, and the opportunity to meet your great family. I will never forget. Also, I would like to thank Prof. Christoph Busch for giving me the opportunity to spend three months in his wonderful research group (da/sec - Biometrics and Internet Security Research Group). Many thanks for all your good advice, time, and funny conversations during our casual lunches. I would also like to thank all members of the research group for making possible I felt like home, specially to: Andreas Nautsch, Christian Rathgeb, Daniel Fischer, Jascha Kolberg, Hareesh Mandalapu, Sergey Isadskiy, Ahmed Madhun, Karina, and of course, the unique Ulrich Scherhag. I will never forget my time with all of you and our Friday cake appointment (I'm still waiting for my winner certificate...). I would also like to thank all great researchers I have met since my PhD beginnings, for all your good advice, moments, and conversations in coffe breaks, gala dinners or while drinking a cold beer. Special mention to Dr. Moises Diaz, Prof. Miguel Angel Ferrer, Prof. Giuseppe Pirlo, Prof. Réjean Plamondon, Prof. Anil K. Jain, Prof. Raymond Veldhuis, Prof. Maria De Marsico, Prof. Björn Schuller, and of course, Prof. Arun Ross, who in 2015 surprised me seating next to me knowing my name despite of being you like my hero and me just a beginner researcher. Many thanks to all of you. You are not only great researchers, but also great people.

Time to Spanish!

En primer lugar, me gustaría dar las gracias a todos mis grandes amigos de Madrid, por estar siempre pendientes de mí, apoyándome y cargándome las pilas siempre que me hacía falta para seguir adelante. En especial, me gustaría destacar a 4 grandes personajes de este mundo: Jesús Calvo, Iván Martínez, Adrián Alvaredo y Rubén Martín. Por muchos más viajes y momentos juntos disfrutando de Francisco de Goya. También me gustaría agradecer el apoyo de todos mis compañeros Telecos, en especial: Daniel Izquierdo, Álvaro Foguet, Víctor Sánchez, Fran Fernández, Rubén Jiménez, Javier López, Alexis Moreno, Diego Arribas, Borja Maza y Pablo Sanz. Gracias por todas las aventuras y risas vividas juntos. Por supuesto, como olvidarme de la tierra en la que crecí, Carrascosa del Tajo. Un pueblo en el que a pesar de ser muy pequeño (sólo hay un bar), está repleto de gente encantadora y llena de vida. Gracias a todos vosotros también carrascoseños por el apoyo dado desde mis origenes.

Y finalmente, falta lo mejor, mi gran familia. Gracias a cada uno de vosotros (tanto los que estáis presentes como los que desgraciadamente nos habéis dejado) por haberme dado tanto en

esta vida. En especial me gustaría dar las gracias a mi madre, Bienvenida Moranchel, por tantos y tantos “tapers” de comida que han salvado mi día a día, mi padre, el genio Tomás Tolosana, por apoyarme siempre dando el toque de humor en la familia, mi hermano mellizo, el otro futuro doctor, el verdadero crack de la familia, Álvaro Tolosana, por aguantarme desde que nacimos (que no es poco) y finalmente, mi hermano mayor, Jorge Tolosana, por enseñarme tantas cosas importantes de la vida y estar siempre pendiente de mí. Mi más sincero agradecimiento a todos vosotros. Todo lo que diga aquí es poco comparado con todo lo que os merecéis.

A todos vosotros, gracias.

Rubén Tolosana Moranchel
Madrid, March 2019

Glossary

- **BRNN**: Bidirectional Recurrent Neural Network.
- **CNN**: Convolutional Neural Network.
- **COTS**: Commercial Off-The-Shelf.
- **CVPR**: Conference on Computer Vision and Pattern Recognition.
- **DET**: Detection Error Trade-off Curve.
- **DL**: Deep Learning.
- **DTW**: Dynamic Time Warping.
- **ECG**: Electrocardiograph.
- **EEG**: Electroencephalograph.
- **EER**: Equal Error Rate.
- **FDR**: Fisher Discriminative Ratio.
- **FA**: False Acceptance.
- **FAR**: False Acceptance Rate.
- **FR**: False Rejection.
- **FRR**: False Rejection Rate.
- **GA**: Genetic Algorithm.
- **GAN**: Generative Adversarial Network.
- **GMM**: Gaussian Mixture Models.
- **GPU**: Graphics Processing Unit.
- **GRU**: Gated Recurrent Unit.
- **HMM**: Hidden Markov Models.
- k**NN**: k-Nearest-Neighbours.
- **LSAD**: Least Squares Anomaly Detection.
- **LSTM**: Long Short-Term Memory.
- **MLP**: Multilayer Perceptron.

- **NN**: Neural Network.
- **OCR**: Optical Character Recognition.
- **OTP**: One-Time Password.
- **PA**: Presentation Attack.
- **PAD**: Presentation Attack Detection.
- **PAI**: Presentation Attack Instrument.
- **PCA**: Principal Component Analysis.
- **PIN**: Personal Identification Number.
- **RNN**: Recurrent Neural Network.
- **ROC**: Receiver Operating Characteristic.
- **SFFS**: Sequential Forward Floating Search.
- **SVC**: Signature Verification Competition.
- **SVM**: Support Vector Machines.
- **TDNN**: Time Delay Neural Network.
- **VSA**: Virtual Skeletal Arm.

Contents

List of Figures

List of Tables

Part I

Problem Statement and Contributions

Chapter 1

Introduction

Is Handwritten signature technology adapted to current scenarios? Is it making the most of the available resources? Certainly not. Signatures have been traditionally acquired in pen-based office-like scenarios using devices specifically designed to capture dynamic signatures and handwriting (i.e., so called graphic or writing tablets such as those manufactured by Wacom and others), in which the stylus has always been considered as input, and achieving, in general, very good results. However, the high deployment of mobile devices such as smartphones and tablets has given rise to new very interesting scenarios and opportunities with their corresponding challenges for the system performance that must be tackled.

This rapid and continuous deployment of mobile devices around the world has been motivated not only by the high technological evolution and new features incorporated by the mobile device sector but also to the new internet infrastructures that allow the communications and use of social media in real time, among many other factors [Salehan and Negahban, 2013]. In this way, both public and private sectors are aware of the importance of mobile devices in our lives and they are putting all their efforts in order to deploy their services through user-friendly mobile applications ensuring data protection and high security at the same time.

Traditionally, passwords have been the most prevalent user authentication approach. However, despite the high popularity and deployment of them in practical scenarios, many studies have highlighted the disadvantages of these approaches as they may be easy to be stolen/forged [Bonneau *et al.*, 2012; Galbally *et al.*, 2017]. Biometric recognition schemes are able to cope with these challenges by combining both security and convenience [Meng *et al.*, 2015].

Biometrics is a technological area whose aim is to authenticate subjects through the use of biological (e.g., face and fingerprint) or behavioural (e.g., voice and handwritten signature) traits [Jain *et al.*, 2016]. It is easy to find biometric systems all around nowadays, e.g., in our smartphones for unblocking them, payments, banking, insurance, and access control such as borders, among many others.

Traditionally, biometric recognition systems have been based on features manually designed by researchers for a specific task (a.k.a. handcrafted features). However, this trend has begun to change in the last years. The reasons mainly reside in the massive amount of available data

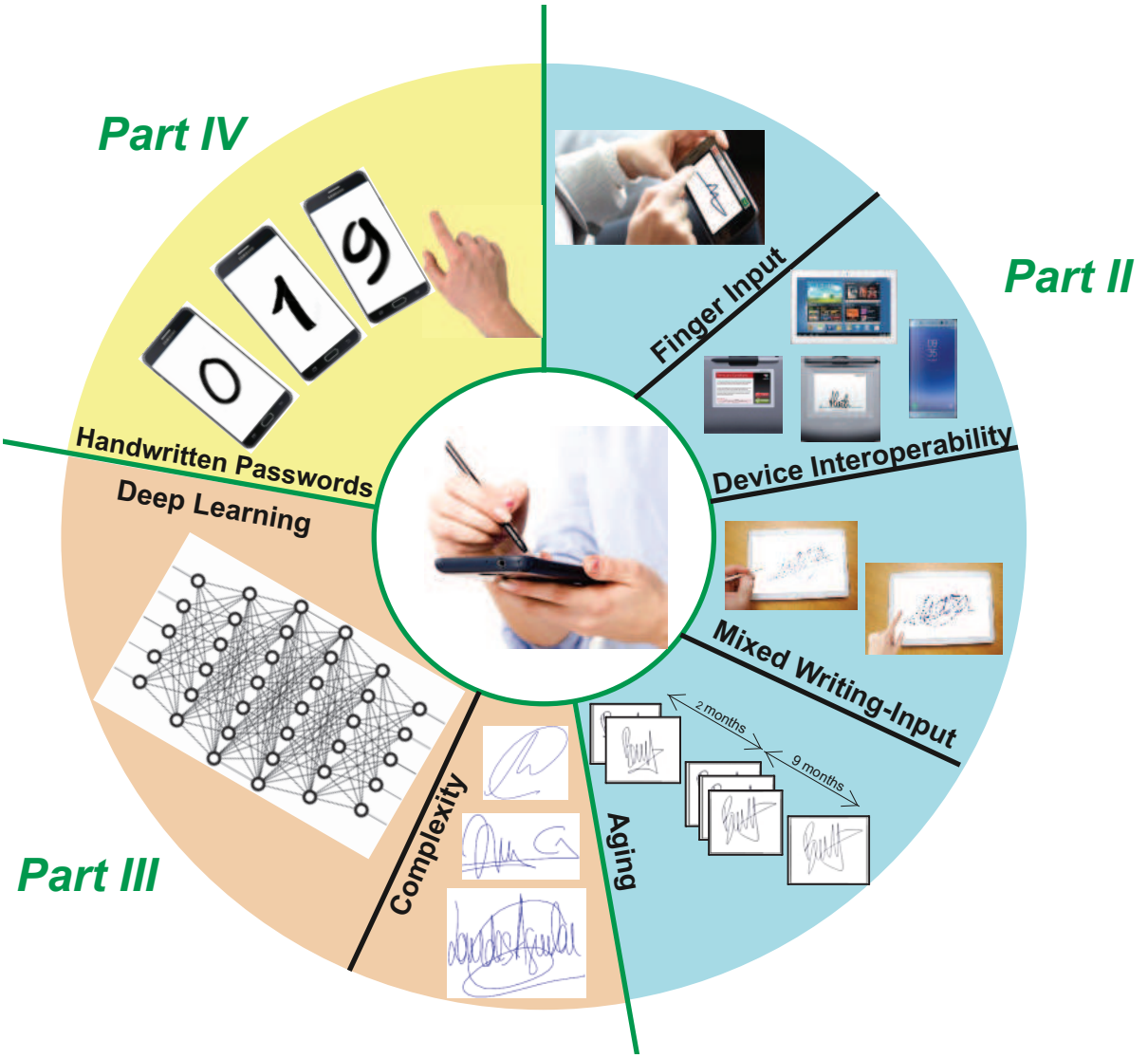


Figure 1.1: *Handwriting and signature scenarios addressed in this Thesis.*

together with the increased computer resources available these days. In this sense, Deep Learning (DL) has become a thriving topic [Goodfellow *et al.*, 2016], allowing computers to learn from experience and understand the world in terms of hierarchy of simpler units. DL has enabled significant advances in complex domains such as natural language processing [I. Sutskever, O. Vinyals and Q.V. Le, 2014], computer vision [B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, 2016] and also for biometrics [Bhanu and Kumar, 2017; Sundararajan and Woodard, 2018], among many others.

This Thesis is mainly focused on the functioning of on-line handwritten signature verification systems on current mobile scenarios, the new opportunities that arise for this biometric trait with the increasing amount of data and computer resources, and the exploration of handwritten passwords for touchscreen biometrics. Fig. 1.1 graphically summarises the handwriting and signature scenarios addressed in this Thesis. The experimental work of the Thesis pretends

to: *i*) analyse and alleviate the impact of emerging scenarios on the system performance by the exploration of robust preprocessing and feature selection techniques, *ii*) enhance the core of traditional signature verification systems through the exploration of DL approaches and the concept of complexity, and *iii*) incorporate biometrics to traditional password-based mobile authentication systems, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual.

This introductory chapter first presents the basics of biometric systems, including properties, and biometric traits. Then, we focus on handwritten signature biometrics, which is the main topic of study in this Thesis, and the challenges and opportunities for it on the emerging scenarios. Later on we summarise the success of DL techniques in many different biometric applications. Then we motivate the incorporation of handwriting biometric information in traditional password-based systems as a second level of user authentication. We finish the chapter by stating the Thesis, giving an outline of the Dissertation, and summarising the research contributions originated from this Thesis.

Although no special background is required for this chapter, the reader will benefit from introductory readings in biometrics [Jain *et al.*, 2008, 2016, 2006, 2004, 2011], handwritten signature verification [Diaz *et al.*, 2018b; Impedovo and Pirlo, 2008; Leclerc and Plamondon, 1994; Plamondon and Lorette, 1989; Plamondon and Srihari, 2000], and DL [Goodfellow *et al.*, 2016; Schmidhuber, 2015].

1.1. Biometrics

Since biometric traits are generally inherent to an individual, there is a strong and reasonably permanent link between a person and his/her biometric traits. Thus, biometric recognition can be used to identify individuals [Jain *et al.*, 2016]. The first scientific study that proved the possibility of using personal anatomical traits for identity verification dates from the 60s. In [Trauring, 1963], the author analysed the minutiae in finger-ridge patterns, concluding with some evidences of the feasibility of those patterns for automatic identity verification. Although the article was written more than 50 years ago, it is incredible to see the capacity of Trauring for coming ahead to the biometric scenario applications.

Despite the appearance of biometric studies on early stages, it wasn't until the last decade when it was established as an specific research area. This is evidenced by recent reference texts [Jain *et al.*, 2016, 2011; Ratha and Govindaraju, 2008; Ross *et al.*, 2006; Tistareli *et al.*, 2009], specific conferences [Boult *et al.*, 2014; Fierrez *et al.*, 2013; Hoque *et al.*, 2017; Singh *et al.*, 2016], peer-review journals such as the new IEEE Transactions on Biometrics, Behavior, and Identity Science, common benchmark tools and evaluations [Beveridge *et al.*, 2013; Kemelmacher-Shlizerman *et al.*, 2016; Neves and Proença, 2016; Phillips *et al.*, 2000; Phillips, 2006; Phillips *et al.*, 2011, 2009a,b; Przybocki and Martin, 2004; Yeung *et al.*, 2004], cooperative international projects [BBfor2, 2010; BioSec, 2004; Biosecure, 2004; COST, 2007; MTIT, 2009; Tabula Rasa, 2010], international consortia dedicated specifically to biometric recognition [EAB,

2015, 2017; BC, 2009; BF, 2009; BI, 2009; EBF, 2009], standardization efforts [ANSI/NIST, 2009; BioAPI, 2009; ISO/IEC JTC 1/SC 27 , 2009; SC37, 2005], and increasing attention both from government [BWG, 2009; DoD, 2005] and industry [IBIA, 2009; International Biometric Group, 2006].

In general, two different operational modes are studied in biometrics: *identification*, and *verification* [Jain *et al.*, 2016]. The former tries to predict the user's identity of the query biometric sample in a database. Therefore, the query biometric sample is compared against all available templates of the database (one-to-many match). The result of the identification operation can be one of the following two decisions: *i*) the identity of one or more users of the database whose templates produce the highest similarity with the query biometric sample, or *ii*) a response from the system indicating that the biometric sample does not match with any of the user templates of the database. In this case, if the system is forced to output an identity, this is referred as *close-set*. Otherwise, it is known as *open-set*. This operational mode has become a hot topic for the police forces in the last years as this is usually a time consuming process that requires a lot of computational resources due to the high number of comparisons that must be performed in order to identify possible terrorists or criminals among all the world population. The latter operational mode is the *verification*. In this case, the query biometric sample is only compared with the template of the claimed user (one-to-one match).

From the first studies performed on fingerprint until now, biometrics has widen to many different traits such as face [Taigman *et al.*, 2014], iris [Bowyerin and Burge, 2016], fingerprint [Cao and Jain, 2018], palmprint [Svoboda *et al.*, 2016], ear [Chen *et al.*, 2015], keystroke [Morales *et al.*, 2016], handwritten signature [Diaz *et al.*, 2018b], touchscreen gestures [Shen *et al.*, 2018], and voice [Ghahabi and Hernando, 2017], among many others. The variety is so large and the system performance so good in most of them that sometimes it is difficult to answer the question: Which biometric trait is the best one? Which one should I choose for my security system?

Despite the impressive results achieved in many different biometrics traits, it is important not to forget that these biometric recognition systems have to withstand different types of possible attacks. Among all possible attack points [ISO/IEC JTC1 SC37 Biometrics, 2016; Ratha *et al.*, 2001], the biometric capture device is probably the most exposed one: no further knowledge about the inner functioning of the system is required to perform an attack. Such attacks are known in the ISO/IEC IS 30107 [ISO/IEC JTC1 SC37 Biometrics, 2016] as *Presentation Attacks* (PA), and refer to the presentation to the capture device of a *Presentation Attack Instrument* (PAI), such as gummy fingerprints [Marasco and Ross, 2015; Sousedik and Busch, 2014] and 3D facial masks [Galbally *et al.*, 2014; Marcel *et al.*, 2014]. Therefore, *Presentation Attack Detection* (PAD) methods must be considered in order to detect such attacks in a first authentication stage [Tolosana *et al.*, 2018a,e].

1.1.1. Modalities and Applications of Biometric Systems

Biometric traits can be classified into physiological or behavioural. Physiological biometrics includes those traits describing *what the person is* such as face, fingerprint, iris, hand

geometry, palmprint, ear, retina, sclera, periocular region, vascular patterns, DNA, electrocardiograph (ECG) and electroencephalograph (EEG). Conversely, behavioural biometrics incorporates information regarding *what the person does* or the way humans behave such as the way we talk (speech), sign/write (signature/handwriting), walk (gait or footsteps), type keyboards (keystroking), use the mouse (mouse dynamics), or interact with mobile devices (touchscreen gestures), among others.

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

- **Universality**, which indicates to what extent a biometric is present in the world population.
- **Uniqueness or Distinctiveness**, which means that the trait has to be unique for every single person, or at least discriminative enough to distinguish between subjects.
- **Permanence**, which entails that the biometric trait should have a compact representation permanent or invariant over a sufficiently large period of time.
- **Collectability**, which refers this biometric trait has to be easily measured quantitatively.

Apart from these aforementioned requirements, the following practical criteria are also desired:

- **Performance**, which involves the efficiency, accuracy, speed, robustness and resource requirements of particular implementations based on the biometric trait.
- **Acceptability**, which refers to whether people are willing to use the biometric trait for authentication purposes and under which conditions.
- **Circumvention**, which reflects the difficulty to fool a system based on a given biometric trait by fraudulent methods.
- **Cost**, which refers to all costs that would be necessary to introduce the system in a real-world scenario.
- **Proportionality**, which refers to the trade-off between the amount of privacy you give to the system and the services you have in return. For instance, it is not logical to compromise your fingerprint information to have access to the gym.

In all biometric traits, it is desirable to have both low intra-user variability (i.e., the biometric information remains stable across measurements of the same subject) and high inter-user variability (i.e., the biometric information differs across measurements of different subjects). In this case, a biometric security approach will provide good and reliable results. However, biometrics traits are usually subject to variations. For physiological biometric traits such as the face, the intra-user variability is highly affected by environment changes (e.g., illumination and

background scene) or other factors such as occlusion, pose or make up [Chan *et al.*, 2014; Chen *et al.*, 2014; Ding and Tao, 2017]. For behavioural biometric traits, the main source of intra-user variability resides on the own user. Behavioural traits such as the voice or the handwritten signature are the result of a complex process that depends not only of the physiological model of the human (e.g., mouth cavities, vocal chords, and wrist and arm muscles) but also to the mood (e.g., happy, sad, and nervous). Other aspects that can reduce the accuracy of both physiological and behavioural biometric systems are related to sensor interoperability (i.e., the case in which different acquisition devices are considered for the enrolment and test stages) due to the different acquisition properties presented on them such as the resolution (dots per inch), sampling frequency, size of the screen, frames per rate, acquisition spectrum, signal to noise ratio, distance from camera, etc [Nogueira *et al.*, 2016; Poh *et al.*, 2007; Ross and Jain, 2004].

The aforementioned biometric traits are also known as *hard biometrics*, as they have the capacity to discriminate identities by themselves. New sources of information such as gender, age or ethnicity, which are known as *soft biometrics* as they are not able to distinguish subjects by themselves, are becoming more and more studied nowadays in order to reduce, for example, the range of search in an identification process [Dantcheva *et al.*, 2016; Gonzalez-Sosa *et al.*, 2018].

Therefore, at this point, the unanswered question "Which biometric trait is the best one?" that arose in Sec. 1.1 seems easier to reply: it depends. It depends on the final application scenario, the user acceptance, the risk of the operation, the usability, and feasibility of the approach, among many other factors [Jain *et al.*, 2011]. No single biometric trait is likely to be optimal and satisfy the requirements of all applications. For this reason it is common to find *multi-biometric* systems based on the fusion of multiple biometric traits. The weight of each of them in the final system performance can be modified for example regarding illumination or noise conditions [Kasprowski and Harezlak, 2018; Ross *et al.*, 2006].

1.2. Handwritten Signature Verification

Handwritten signature is the result of a complex process that depends on the psychophysical state of the signer and the conditions under which the signature apposition process occurs [Impedovo and Pirlo, 2008]. We usually start learning how to write at the age of about three years old. At that age, most children understand that writing is made by combining lines, curves, and repeated patterns. About a year later, children begin to use letters in their own style. They usually start experimenting with the letters of their own names, as these are the most familiar letters to them. Thus, children begin to learn the shape and sequence of the letters in their name although their motor control is not yet accurate [Ferrer *et al.*, 2015]. In some parts of the world it is common for the children to improve their handwriting skills through printed worksheets.

Many efforts have been carried out in the last years in order to model the handwritten signature process and analyse the human movement for handwriting. Some studies describe a movement with analytical expressions [Alimi, 2003; Hogan, 1984; Plamondon, 1995], while oth-

ers proceed through the numerical resolution of a system of differential equations [Harris and Wolpert, 1998; Neilson, 1993]. Among all these types of model representation, the studies carried out in [Plamondon and Parizeau, 1988; Wolpert *et al.*, 1995] prove that kinematics signals (in particular velocity oriented models) should be preferred over kinetics ones. These guidelines were followed by the authors in [Reilly and Plamondon, 2009], considering a physiological model of the human movement production for the generation of signatures. This choice was motivated specially for the advantage of being invariant in regard of cultural or language differences, whereas systems based on visual characteristics often need to be tailored for Chinese, Arabic, European, or American signatures. As a result of that work, the Sigma-Lognormal writing generation model, which is further used nowadays in many different fields such as health [Impedovo *et al.*, 2013], appeared. This model was later used in [Galbally *et al.*, 2012a,b] for the generation of synthetic signatures. The approach was based on the combination of the spectral analysis of real signatures with the Kinematic Theory of rapid human movements in order to generate totally synthetic specimens. It is worth to highlight the research carried out by the authors Diaz M. and Ferrer M.A. for the understanding of the intra-personal variability of the signatures of a signer. The results of their investigation have concluded with great achievements in the generation of synthetic data [Diaz *et al.*, 2017a; Ferrer *et al.*, 2017a, 2015], improving the system performance in scenarios with lack of training signatures [Diaz *et al.*, 2016a], and proposing new discriminative features for user authentication [Diaz *et al.*, 2018a], among many other lines.

Handwritten signature verification is the task of authenticating users through the way they sign. It has been fully studied along the last 50 years proving to be one of the most reliable and convenient biometric traits in many relevant sectors such as security, e-government, healthcare, education, banking or insurance regardless of the age of the user. This fact has been demonstrated in many different surveys published since its origins [Diaz *et al.*, 2018b; Impedovo and Pirlo, 2008; Leclerc and Plamondon, 1994; Plamondon and Lorette, 1989; Plamondon and Srihari, 2000]. Despite its high deployment in actual scenarios, handwritten signature verification is a complex task as signatures from the same user (a.k.a. genuine signatures) can differ significantly (high intra-user variability) whereas forgeries performed by other subjects can be very similar to the genuine signatures (low inter-user variability).

Handwritten signature verification can be divided into two main areas regarding the data acquisition method:

- **Off-line** or **static** signature verification uses only the image of the signature that is acquired from the paper sheet through imaging devices such as scanners and cameras providing gray scale images. Different approaches have been proposed in this area [Hou *et al.*, 2004]. In [Nguyen *et al.*, 2009], the authors proposed the use of global features (e.g., vertical and horizontal projections of the signatures, focusing on key strokes of them) based on the boundary of a signature. Other authors have focused on the global image level and measured the gray level variations in the signature images by using statistical texture feature [Vargas *et al.*, 2011]. Recent studies have also proved the potential of DL technology for the task [Dey *et al.*, 2017].

- **On-line** or **dynamic** signature verification uses special hardware for the acquisition of the signatures such as digitizing tablets manufactured by Wacom [Jain *et al.*, 2002]. Therefore, the dynamic information of the complete writing process is acquired. Besides, information related to the pressure that the user perform on the screen device while signing as well as the orientation of the pen can be available depending on the quality of the acquisition device and the input. Therefore, on-line signature verification systems have traditionally achieved much better results compared to the off-line systems as not only the image of the signature is available, but also the dynamics [Galbally *et al.*, 2015; Plamondon and Lorette, 1989]. This is the approach considered in this Thesis, and it will be further described in the following chapters.

For both off- and on-line approaches, the output of the system is binary: *accept* or *reject*. Usually, this decision depends on a similarity threshold. If the similarity (or match score) resulting of the comparison between the query signature and the model of the claimed user is higher than a specific threshold, the user is accepted in the system. Otherwise, the user is rejected. Therefore, two types of errors can be produced in the system: False Acceptance (FA) and False Rejection (FR). FA is produced when a user that falsely claims to be another user is accepted by the system as being the genuine user. FR 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 verification system can be computed for any similarity threshold.

A common measure to compare the performance of biometric systems is the Equal Error Rate (EER), which represents the error rate when the decision threshold is set to satisfy that FAR = FRR. In addition, for an easy comparison between different systems at any decision threshold, the Receiver Operating Characteristic (ROC) or Detection Error Trade-off (DET) plots are generally used [Martin *et al.*, 1997].

1.3. Challenges and Opportunities for On-Line Handwritten Signature Verification on Emerging Scenarios

Signatures have been traditionally acquired in office-like scenarios using devices specifically designed to capture dynamic signatures and handwriting, in which the stylus has always been considered as the input device achieving, in general, very good results. This fact is mainly produced due to the very good conditions and controlled scenarios considered, e.g., the high-quality digitizer tablets that provide information related to the pressure that the user performs on the screen device, the pen inclination, and the pen trajectory during pen-ups, which is invisible information for the impostors.

The high technological evolution and the significant improvement of sensors quality have given rise to new interesting scenarios and opportunities for the task of on-line handwritten signature verification. The acquisition of signatures have been expanded from the traditional

high-quality Wacom devices to new general purpose devices such as smartphones and tablets, given rise to device interoperability scenarios, i.e., different devices are considered for the acquisition of signatures during the enrolment and test. In addition, the use of the finger as the writing input has become a thriving scenario for many real applications due to the stylus is rarely available in devices such as smartphones. All these factors, together with the fact that handwritten signature is one of the most socially accepted traits, have produced a massive deployment of this technology in many relevant sectors regardless of the age of the user [Guest, 2006]. Also, the high acceptance of the society to use their mobile devices on daily activities [Salehan and Negahban, 2013] introduces a new scenario where the number of signatures for a specific user can increase with time, ending up with even dozens or hundreds of signatures acquired in multiple sessions, unlike the traditional scenario considered in signature verification where just a few genuine signatures from the enrolment session are used for modelling the users.

However, all these novelties and opportunities also bring up some unanswered questions: What quality and type of information is provided by general purpose devices such as tablets or smartphones? What is the system performance when using the finger as input? Is the intra- and inter-user variability highly affected in these new scenarios? How do long-term scenarios affect the final system performance? Should we consider template and system configuration update strategies?

1.4. Deep Learning

DL has become a thriving topic in the last years, allowing computers to learn from experience and understand the world in terms of hierarchy of simpler units [Goodfellow *et al.*, 2016; LeCun *et al.*, 2015; Schmidhuber, 2015]. DL has enabled significant advances in complex domains such as natural language processing [Sutskever *et al.*, 2014], computer vision [Szegedy *et al.*, 2016], healthcare [Miotto *et al.*, 2017] and fashion [Wang *et al.*, 2018], among many others. Biometrics has also made the most of it [Bhanu and Kumar, 2017; Sundararajan and Woodard, 2018], e.g., in speech [Graves and Jaitly, 2014], and facial and fingerprint recognition [Chugh *et al.*, 2018; Parkhi *et al.*, 2015]. The main reasons to understand the high deployment of DL lie on the increasing amount of available data and also the technological evolution produced in the field of GPU, allowing the development and training of deep size neural network models. However, there are still some tasks in which DL has not achieved state-of-the-art results due to the scarcity of available data and therefore, the inability to train from scratch traditional deep learning architectures.

One of the fields in which DL has caused more impact in the last years is in handwriting recognition due to the relationship that exists between current inputs and past and future contexts. These architectures have been deployed with success in both on- and off-line handwriting [Graves *et al.*, 2009; Graves and Schmidhuber, 2009; Zhang *et al.*, 2017].

Despite the good results obtained in the field of on-line handwriting recognition, and the similarity with the task of on-line handwritten signature, few studies have analysed the use of DL

approaches to the task of handwritten signature verification [Otte *et al.*, 2014; Tiflin and Omlin, 2003], concluding that DL systems trained with standard mechanisms are not appropriate as the amount of available data is scarce compared to other tasks such as handwriting recognition.

In this context, the following question comes to my mind: What deep learning architecture shall we propose to facilitate the training process and generalise well to new unseen samples?

1.5. Handwritten Passwords for Touchscreen Biometrics

Traditionally, the two most prevalent user authentication approaches have been Personal Identification Number (PIN) and One-Time Password (OTP). While PIN-based authentication systems require users to memorise their personal passwords, OTP-based systems avoid users to memorise them as the security system is in charge of selecting and providing to the user a different password each time is required, e.g., sending messages to personal mobile devices or special tokens. Despite the high popularity and deployment of PIN- and OTP-based authentication systems in real scenarios, many studies have highlighted the weaknesses of these approaches [Bonneau *et al.*, 2012; Galbally *et al.*, 2017]. First, it is common to use passwords based on sequential digits, personal information such as birth dates, or simply words such as "password" or "qwerty" that are very easy to guess. Second, passwords that are typed on mobile devices such as tablets or smartphones are susceptible to *smudge attacks*, i.e., the deposition of finger grease traces on the touchscreen can be used for the impostors to guess the password [Aviv *et al.*, 2010]. Finally, password-based authentication is also vulnerable to *shoulder surfing*. This type of attack is produced when the impostor can observe directly or use external recording devices to collect the user information. This attack has attracted the attention of many researchers in recent years due to the increased deployment of handheld recording devices and public surveillance infrastructures [Shukla *et al.*, 2014; Yue *et al.*, 2014]. Biometrics can cope with these challenges by combining both security and convenience [Meng *et al.*, 2015].

Two-factor authentication approaches have gained a lot of success in the last years in order to improve the level of security. These approaches are based on the combination of two authentication stages. For example, one possible case could be the following: 1) the security system checks that the claimed user introduces its unique password correctly, and 2) its behavioural biometric information is used for an enhanced final verification [Luca *et al.*, 2012]. This way the robustness of the security system increases as impostors need more than the traditional password to get access to the system. This approach has been studied in previous works. In [Angulo and Wastlund, 2011], the authors proposed a two-factor verification system based on dynamic lock patterns, achieving a final average value of 10.39% EER against impostors. A similar approach based on OTP with dynamical lock patterns was considered in [Lacharme and Rosenberger, 2016] extracting features such as the X and Y position, pressure or finger size with very good results. This approach has also been expanded to periocular biometrics [Jenkins *et al.*, 2017].

In the present Dissertation we evaluate the advantages and potential of incorporating biometrics to password-based mobile authentication systems, asking the users to draw each digit

of the password on the touchscreen instead of typing them as usual. This way, the traditional authentication systems are enhanced by incorporating dynamic handwritten biometric information. One example of use that motivates our proposed approach is on internet payments with credit cards. Banks usually send a numerical password (typically between 6 and 8 digits) to the user's mobile device. This numerical password must be inserted by the user in the security platform in order to complete the payment. Our proposed approach enhances such scenario by including a second authentication factor based on the user biometric information while drawing the digits.

However, the novelty of our approach brings up some unanswered questions: What is the discriminative power of each handwritten digit? How robust is our biometric system regarding the length of the handwritten password or the number of available enrolment samples per user? Should we follow any password generation strategy to improve the system performance?

1.6. Motivation of the Thesis

The research carried out in this Thesis has been mainly motivated by the following five observations:

The first observation comes from the fact that at the beginning of this Thesis, there were no publicly available databases that considered the acquisition of handwritten signatures using commercial off-the-shelf (COTS) devices over the emerging finger, device interoperability and mixed writing-input scenarios. All signature databases such as the MCYT or the BiosecurID considered devices specifically designed to capture dynamic signatures and handwriting using the stylus as input [Fierrez *et al.*, 2010; Ortega-Garcia *et al.*, 2003].

The second observation is strongly related to the first one. Due to the novelty of these scenarios and the lack of publicly available databases, there were almost no studies that performed a complete analysis of the functioning of signature verification systems using COTS devices such as smartphones and tablets general purpose devices. Additionally, scenarios such as finger input, device interoperability, and mixed writing-input (i.e., different writing tools are considered for the acquisition of signatures during the enrolment and test) were scarcely studied [Alonso-Fernandez *et al.*, 2005; Martinez-Diaz *et al.*, 2013; Robertson and Guest, 2015].

The third observation is motivated due to the high acceptance of the society to use their mobile devices on daily activities [Salehan and Negahban, 2013]. This fact opens the way to new scenarios where the number of available signatures for a specific user can increase with time, being possible to configure time-adaptive improved versions of the users' models in order to overcome to some extent the aging effect (i.e., the gradual degradation of the system performance due to the changes suffered by the user's trait along the time).

The fourth observation comes from the fact that despite the astonishing results achieved using DL approaches in many different fields, concretely in handwriting recognition and writer identification [Graves *et al.*, 2009; Graves and Schmidhuber, 2009; Zhang *et al.*, 2017], almost no research has been carried out on handwritten signature verification. The core of most of the

state-of-the-art signature verification systems is still almost the same than 20 years ago (e.g., Dynamic Time Warping (DTW), Hidden Markov Models (HMM), Gaussian Mixture Models (GMM) or Support Vector Machines (SVM) [Diaz *et al.*, 2018b; Impedovo and Pirlo, 2008; Leclerc and Plamondon, 1994; Plamondon and Lorette, 1989; Plamondon and Srihari, 2000]).

The last observation is that, in general, traditional authentication methods based on passwords such as PIN and OTP are still commonly used nowadays, despite they can be easily stolen or forged [Bonneau *et al.*, 2012; Galbally *et al.*, 2017]. Two-factor authentication approaches can enhance the security of these scenarios including biometric information such as handwriting. This way impostors need more than the traditional password to get access to the system.

1.7. The Thesis and Main Contributions

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

The high technological evolution and massive deployment of mobile devices in our society make possible to use signature- and handwriting-based authentication approaches in unprecedented scenarios. However, these new opportunities also bring up challenges that must be carefully tackled. These can be overcome through an exhaustive analysis of the new scenarios considering robust preprocessing and feature selection techniques, template update strategies, and also by the improvement of traditional authentication core matchers through Deep Learning techniques and new concepts such as the signature complexity.

The **main contributions** of this Thesis are:

- *Emerging scenarios:* We have performed a complete analysis of traditional on-line signature verification systems on finger input, device interoperability and mixed writing-input scenarios. In order to do that, we have first acquired a new signature database (e-BioSign, which is already publicly available to the research community). Different preprocessing and feature selection techniques have been proposed, analysing the robustness of both global and local approaches and selecting the optimal feature subsets for each scenario. Finally, we have concentrated on the new scenario where the number of available signatures for a specific user can increase with time, acquiring signatures in multiple sessions. In order to do that, we have first extended the ATVS On-Line Signature Long-Term database including skilled forgeries. This database comprises a total of 6 different acquisition sessions within a 15-month time span. Then, we have carried out an exhaustive experimental analysis of template update strategies for three very popular on-line signature verification systems, extracting various practical findings related to the template aging effect in signature biometrics, and configuring time-adaptive improved versions of the considered baseline approaches overcoming to some extent the template aging.

- *Towards the near future:* We have explored the potential of DL approaches for on-line signature verification. Our proposed approach has outperformed other state-of-the-art systems even with small number of training signatures. Besides, we have evaluated the concept of complexity in signature verification, extracting specific features for each signature complexity level, resulting in better system performances compared to the traditional approaches.

- *Handwritten passwords for touchscreen biometrics:* We have proposed the incorporation of handwriting biometric information to traditional password-based mobile authentication systems, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual. A new database (e-BioDigit) that comprises handwritten numerical digits from 0 to 9 has been acquired. We have performed a complete analysis of the touch biometric system regarding the robustness and discriminative power of each handwritten digit. In addition, we have analysed the potential of our proposed approach when increasing the length of the handwritten password and the number of available enrolment samples per user. Our proposed approach has achieved remarkable results compared to other verification traits such as the handwritten signature and graphical passwords, as well as other recent touchscreen biometrics.

1.8. Outline of the Dissertation

The Thesis is structured according to a traditional complex type with background theory, practical methods, and experimental studies in which the methods are applied [Paltridge, 2002].

The Dissertation is divided into five parts. *Part I* introduces the problem statement and the contributions originated from this Dissertation. Then, there are three experimental parts: *Part II* focuses on the new challenging and current signature verification scenarios, *Part III* describes the experimental work carried out in order to enhance the core of traditional signature verification systems, and finally *Part IV* addresses the experimental work carried out for incorporating handwriting biometric information to traditional password-based authentication systems. The Dissertation concludes with *Part V*. The chapter structure is as follows:

- *Part I: Problem Statement and Contributions*

  - Chapter 1 introduces the topics addressed in this Thesis: biometrics, on-line handwritten signature verification, deep learning, and handwritten passwords for touchscreen biometrics.
  - Chapter 2 summarises related works which are in line with this Thesis.
  - Chapter 3 first gives an overview of the most relevant features of existing on-line signature databases, making special emphasis on the databases used in the experimental work of this Thesis. We then present the new e-BioSign and e-BioDigit databases, as well as the extension of the ATVS On-Line Signature Long-Term database. These

new databases have been acquired during the realization of the Thesis and are publicly available to the research community nowadays.

- Chapter 4 describes all the details of the traditional and novel on-line handwritten signature verification systems considered in this Thesis.

- *Part II: Emerging Scenarios*

  - Chapter 5 first analyses the system performance of traditional signature verification systems on emerging scenarios such as finger input, device interoperability and mixed writing-input scenarios. Both Biosecure and e-BioSign databases are considered in the experimental work. Then, we propose a two-stage approach based on robust preprocessing and feature selection techniques in order to alleviate the degradation of the system performance on these novel scenarios.
  - Chapter 6 explores the scenario where the number of available signatures for a specific user can increase with time, ending up even with dozens or hundreds of signatures acquired in multiple sessions. This chapter focuses on the template aging effects on popular signature verification systems and proposes template and system configuration update strategies in order to reduce the template aging.

- *Part III: Towards the Near Future*

  - Chapter 7 evaluates the potential of our proposed on-line signature verification system based on DL approaches and compares it with state-of-the-art signature verification systems.
  - Chapter 8 explores the concept of complexity in signature biometrics and proposes on-line signature verification systems adapted to the signature complexity level of the user

- *Part IV: Handwritten Passwords for Touchscreen Biometrics*

  - Chapter 9 studies the incorporation of handwriting biometric information to password-based mobile authentication systems, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual.

- *Part V: Conclusions*

  - Chapter 10 concludes the Thesis summarising the main results obtained and outlining future research lines.

The dependence among the chapters is illustrated in Fig. 1.2. For example, before reading any of the experimental Chapters 5, 6, 7, 8, and 9 (yellow boxes in Fig. 1.2), one should read first Chapters 1, 2, 3 and 4. Before Chapter 8 it is recommended to read Chapter 5.

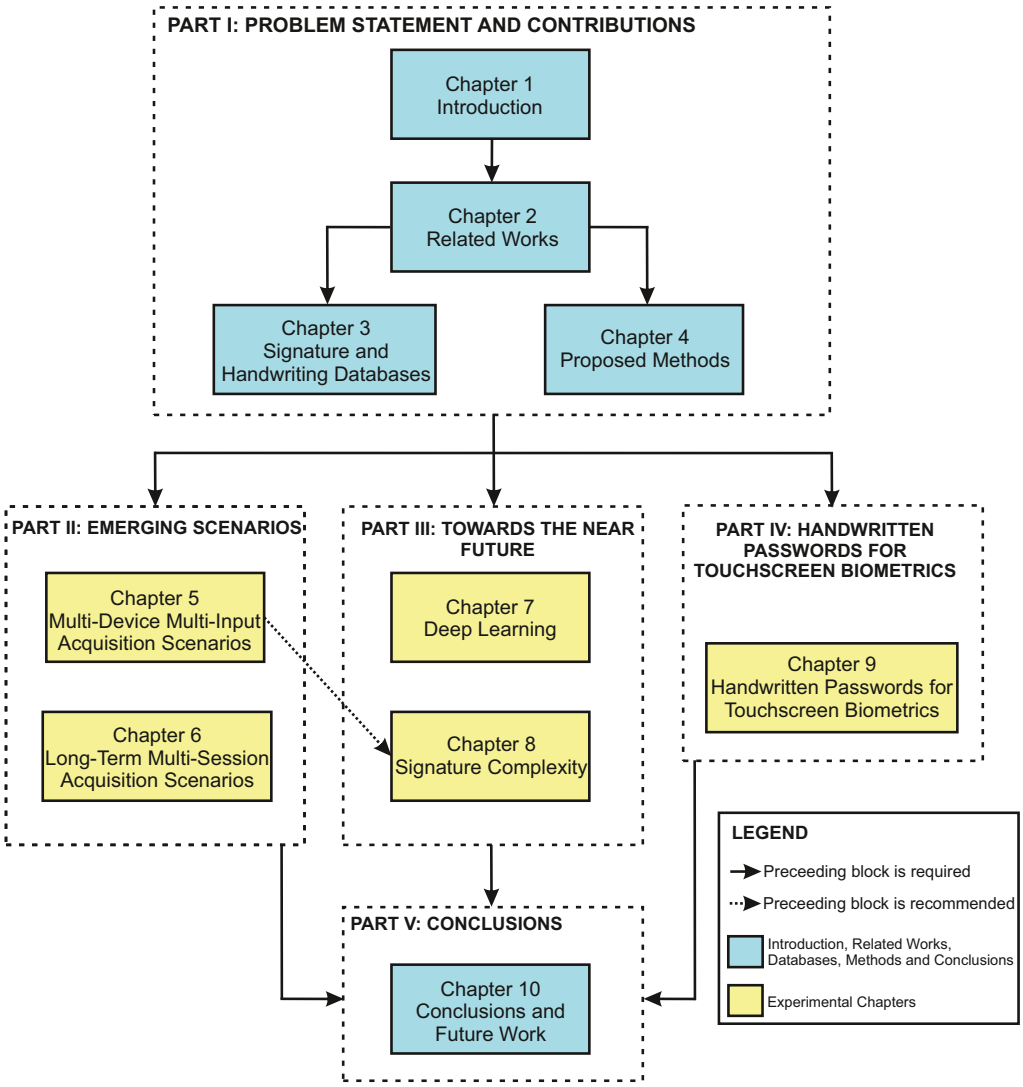


Figure 1.2: *Dependence among chapters.*

The methods considered in this Thesis are based on two different approaches: *i*) traditional approaches from the handwriting and signature recognition literature, and *ii*) novel deep learning approaches. The reader is referred to standard texts for a background on the topic [Duda *et al.*, 2001; Goodfellow *et al.*, 2016; Schmidhuber, 2015; Theodoridis and Koutroumbas, 2008].

1.9. Detailed Research Contributions

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

- LITERATURE REVIEWS.

  1. Presentation attacks in signature biometrics.

- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega Garcia, "Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection", S. Marcel, M.S. Nixon, J. Fierrez and N. Evans (Eds.), *Handbook of Biometric Anti-Spoofing (2nd Edition), Springer*, 2018.

- SIGNATURE VERIFICATION.

1. Novel databases.

- **R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database", *PLOS ONE*, Vol. 5, n. 12, 2017.**
- **R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Reducing the Template Aging Effect in On-Line Signature Biometrics", *IET Biometrics* (under review).**
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Do You Need More Data? The DeepSignDB On-Line Handwritten Signature Biometric Database", *in Proc. 15th Int. Conference on Document Analysis and Recognition, ICDAR*, Sydney, Australia, 2019 (under review).
- R. Vera-Rodriguez, R. Tolosana, J. Ortega-Garcia and J. Fierrez, "e-BioSign: Stylus- and Finger-Input Multi-Device Database for Dynamic Signature Recognition", *in Proc. 3rd International Workshop on Biometrics and Forensics, IWBF*, Gjovik, Norway, 2015.
- R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Update Strategies for HMM-Based Dynamic Signature Biometric Systems", *in Proc. 7th IEEE Int. Workshop on Information Forensics and Security, WIFS*, Rome, Italy, 2015.

2. Multi-device multi-writing-tool acquisition scenarios.

- **R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database", *PLOS ONE*, Vol. 5, n. 12, 2017.**
- **R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Preprocessing and Feature Selection for Improved Sensor Interoperability in Online Biometric Signature Verification", *IEEE Access*, Vol. 3, pp. 478 - 489, 2015.**
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Do You Need More Data? The DeepSignDB On-Line Handwritten Signature Biometric Database", *in Proc. 15th Int. Conference on Document Analysis and Recognition, ICDAR*, Sydney, Australia, 2019 (under review).
- R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Assessment of Using the Finger in On-Line Handwritten Signature Verification Systems", *in Proc. 18th Int. Graphonomics Society Conference, IGS*, Gaeta, Italy, 2017.
- R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Optimal Feature Selection and Inter-Operability Compensation for On-Line Biometric Signature Authentication", *in Proc. IEEE/IAPR Int. Conf. on Biometrics, ICB*, Phuket, Thailand, 2015.
- R. Vera-Rodriguez, R. Tolosana, J. Ortega-Garcia and J. Fierrez, "e-BioSign: Stylus- and Finger-Input Multi-Device Database for Dynamic Signature Recognition", *in Proc. 3rd International Workshop on Biometrics and Forensics, IWBF*, Gjovik, Norway, 2015.

3. Long-term multi-session acquisition scenarios.

- **R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Reducing the Template Aging Effect in On-Line Signature Biometrics", *IET Biometrics* (under review).**
- R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Update Strategies for HMM-Based Dynamic Signature Biometric Systems", *in Proc. 7th IEEE Int. Workshop on Information Forensics and Security, WIFS*, Rome, Italy, 2015.

4. Deep learning on signature verification systems.

- **R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics", *IEEE Access*, pp. 5128-5138, 2018.**
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Do You Need More Data? The DeepSignDB On-Line Handwritten Signature Biometric Database", *in Proc. 15th Int. Conference on Document Analysis and Recognition, ICDAR*, Sydney, Australia, 2019 (under review).
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Biometric Signature Verification Using Recurrent Neural Networks", *in Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, Kyoto, Japan, 2017.

5. Complexity-based signature verification systems.

- R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez and J. Ortega-Garcia, "Complexity-based Biometric Signature Verification", *in Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, Kyoto, Japan, 2017.
- R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Morales, J. Fierrez and J. Ortega-Garcia, "Modeling the Complexity of Biomechanical Tasks using the Lognormality Principle: Applications to Signature Recognition and Touch-Screen Children Detection", *in Proc. IAPR Intl. Conf. on Pattern Recognition and Artificial Intelligence, ICPRAI*, 2018.
- R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Acien, A. Morales, J. Fierrez and J. Ortega-Garcia, "Modeling the Complexity of Signature and Touch-Screen Biometrics using the Lognormality Principle", R. Plamondon *et al.* (Eds.), *The Lognormality Principle and its Applications, World Scientific*, 2019.

6. Exploring new global features.

- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Feature-Based Dynamic Signature Verification under Forensic Scenarios", *in Proc. 3rd International Workshop on Biometrics and Forensics, IWBF*, Gjovik, Norway, 2015.

7. Signature biometric template protection.

- R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia and J. Fierrez, "Increasing the Robustness of Biometric Templates for Dynamic Signature Biometric Systems", *in Proc. 49th Annual Int. Carnahan Conf. on Security Technology, ICCST*, Taipei, Taiwan, 2015.

- HANDWRITTEN PASSWORDS FOR TOUCHSCREEN BIOMETRICS.

1. Novel Databases.

- **R. Tolosana, R. Vera-Rodriguez and J. Fierrez, "BioTouchPass: Handwritten Passwords for Touchscreen Biometrics", *IEEE Transactions on Mobile Computing* (under review).**
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "MobileTouchDB: Mobile Touch Character Database in the Wild and Biometric Benchmark", *in Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, Long Beach, California, USA, 2019 (under review).
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Incorporating Touch Biometrics to Mobile One-Time Passwords: Exploration of Digits", *in Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, Salt Lake City, Utah, USA, June 2018.

2. Novel authentication methods.

- **R. Tolosana, R. Vera-Rodrigue and J. Fierrez, "BioTouchPass: Handwritten Passwords for Touchscreen Biometrics", *IEEE Transactions on Mobile Computing* (under review).**
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "MobileTouchDB: Mobile Touch Character Database in the Wild and Biometric Benchmark", *in Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, Long Beach, California, USA, 2019 (under review).
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Incorporating Touch Biometrics to Mobile One-Time Passwords: Exploration of Digits", *in Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, Salt Lake City, Utah, USA, June 2018.

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

- FORENSIC TOOLS.

  1. Handwritten signature.

     - R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, A. Acien and R. Tolosana, "e-BioSign Tool: Towards Scientific Assessment of Dynamic Signatures under Forensic Conditions", *in Proc. IEEE 7th International Conference on Biometrics: Theory, Applications and Systems, BTAS*, Arlington, Virginia, USA, 2015.

- APPLICATIONS OF THE SIGMA LOGNORMAL WRITING GENERATION MODEL.

  1. Children detection.

     - R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Morales, J. Fierrez and J. Ortega-Garcia, "Modeling the Complexity of Biomechanical Tasks using the Lognormality Principle: Applications to Signature Recognition and Touch-screen Children Detection", *in Proc. IAPR Intl. Conf. on Pattern Recognition and Artificial Intelligence, ICPRAI*, 2018.
     - R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Acien, A. Morales, J. Fierrez and J. Ortega-Garcia, "Modeling the Complexity of Signature and Touch-Screen Biometrics using the Lognormality Principle", R. Plamondon *et al.* (Eds.), *The Lognormality Principle and its Applications*, 2019.

- ESTABLISHING HUMAN BASELINE PERFORMANCE.

  1. Signature recognition.

     - D. Morocho, A. Morales, J. Fierrez and R. Tolosana, "Signature Recognition: Establishing Human Baseline Performance via Crowdsourcing", *in Proc. 4th Int. Workshop on Biometrics and Forensics, IWBF*, pp. 1-6, Limassol, Cyprus, 2016.

Other doctoral research not included in the Thesis:

- COMPETITIONS.

  1. Keystroke biometrics.

     - **A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos and S. Marcel, "Keystroke Biometrics Ongoing Competition", *IEEE Access*, Vol. 4, pp. 7736-7746, 2016.**

- FINGERPRINT BIOMETRICS.

  1. Presentation attack detection (PAD).

     - **R. Tolosana, M. Gomez-Barrero, C. Busch and J. Ortega-Garcia, "Biometric Presentation Attack Detection: Beyond the Visible Spectrum", *IEEE Transactions on Information Forensics and Security* (under review).**
     - R. Tolosana, M. Gomez-Barrero, J. Kolberg, A. Morales, C. Busch and J. Ortega-Garcia, "Towards Fingerprint Presentation Attack Detection Based on Convolutional Neural Networks and Short Wave Infrared Imaging", *In Proc. 17th International Conference of the Biometrics Special Interest Group*, 2018.

- HEALTHCARE.

  1. Eye diseases.

     - A. Morales, F. M. Costela, R. Tolosana and R. L. Woods, "Saccade Landing Point Prediction: A Novel Approach based on Recurrent Neural Networks", *in Proc. 2018 International Conference on Machine Learning Technologies, ICMLT*, Jinan, China, 2018.

Chapter 2

Related Works

This chapter summarises previous studies related to the Thesis. First, Sec. 2.1 describes each module of traditional on-line signature verification systems as well as the two modalities considered, i.e., feature-based systems (a.k.a. global systems) and time functions-based systems (a.k.a. local systems). Then, we present in Sec. 2.2 some of the signature verification emerging scenarios considered in this Dissertation and perform a thorough overview of related works in this line. Other interesting on-line signature research topics are described in Sec. 2.3. The importance and success of DL approaches are described in Sec. 2.4 together with a brief overview of the most famous DL architectures considered nowadays. Lastly, Sec. 2.5 surveys and compares advantages and limitations of recent touchscreen biometrics approaches.

This chapter is based on the following publications: [Tolosana *et al.*, 2018b, 2017a, 2018c, 2015d].

2.1. On-Line Signature Verification

2.1.1. System Architecture

On-line signature verification systems usually contain the same modules that other biometric traits. Fig. 2.1 shows the architecture of a traditional on-line signature verification system [Diaz *et al.*, 2018b; Fierrez and Ortega-Garcia, 2008; Impedovo and Pirlo, 2008; Plamondon and Lorette, 1989; Plamondon and Srihari, 2000]. In general, the following modules are considered:

1. **Data Acquisition:** Many different devices allow the acquisition of handwritten signatures nowadays. From the traditional Wacom devices designed specifically for the acquisition of signatures and handwriting [Ortega-Garcia *et al.*, 2010, 2003], to general purpose devices such as smartphones and tablets that we use on a daily basis [Antal and Bandi, 2017; Blanco-Gonzalo *et al.*, 2014; Tolosana *et al.*, 2017a]. The wide variety of acquisition devices has extended even to the development of new ink pens with the technology necessary for

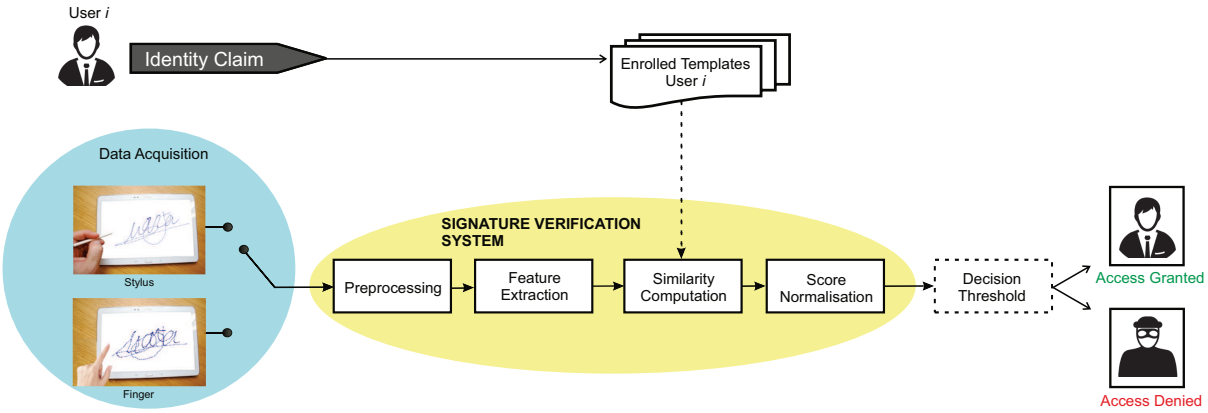


Figure 2.1: *Traditional architecture of a handwritten signature verification system.*

the acquisition of the biometric information during the whole signing process, like those manufactured by Anoto [Malik *et al.*, 2015]. The signature information captured is highly dependent on the device (i.e., specifically designed for acquiring handwriting and signature or general purpose devices such as smartphones and tablets) and also the writing input considered (i.e., stylus or finger). In general, high-quality devices like those manufactured by Wacom or Signotec GmbH provide information related to the signature trajectory (i.e., X and Y spatial coordinates), pressure and pen orientation (i.e., altitude and azimuth). However, this information is reduced to just spatial coordinates and pressure when using general purpose devices. The information available is even more reduced when we consider the case of using the finger as input. In this scenario only information related to the spatial coordinates is available. In addition to all these variabilities, it is common to find different spatial resolutions and also sampling rates among different devices, ranging from 100 Hz to 200 Hz (the maximum frequencies of the signature time functions are approximately of 20 - 30 Hz [Plamondon and Lorette, 1989]). Finally, all these signature signals are stored in a file as discrete-time series.

2. **Preprocessing:** This is an optional module, although it is commonly used nowadays. The main purpose is to enhance the quality of the raw signature signals captured by the sensor in order to extract more robust features over different acquisition scenarios and improve the final performance of the system. Among all the different preprocessing techniques, it is common to correct sampling errors, normalise the sampling frequency among different devices using interpolation techniques, normalise the size and position of the signatures, and remove the first and last samples of the signatures as they correspond to the time between the operator clicks to start/finish the acquisition and the time the user starts/finishes signing [Martinez-Diaz *et al.*, 2007; Tolosana *et al.*, 2015c,d].

3. **Enrolment:** In this step the biometric information of the user is extracted in a preliminary registration stage before using the authentication system for the first time. Two different systems are considered, *model-* and *reference-based* systems. In model-based systems a

statistical user model is computed using a set of genuine signatures. The user model is used for future comparisons against input signatures in the similarity computation module [Galbally *et al.*, 2013; Nanni and Lumini, 2005; Tolosana *et al.*, 2015e]. On the other hand, reference-based systems store the features of each genuine signature as templates. These templates are then used in the similarity computation module to measure the similarity against the input signature features [Galbally *et al.*, 2013; Lei and Govindaraju, 2005; Sae-Bae and Memon, 2014; Tolosana *et al.*, 2015d].

4. **Feature Extraction:** Two main approaches have been traditionally considered in the literature [Diaz *et al.*, 2018b; Martinez-Diaz *et al.*, 2015a]. On the one hand, *feature-based systems* (a.k.a. global systems) consist on the extraction of global features from the signature in order to obtain a holistic feature vector that represents the user signature [Martinez-Diaz *et al.*, 2014; Sae-Bae and Memon, 2014; Tolosana *et al.*, 2015a]. On the other hand, *time functions-based systems* (a.k.a. local systems) consider the time sequences of the signature provided by the sensor (e.g., X and Y coordinates and pressure) and other time sequences extracted from the raw signals [Liu *et al.*, 2014; Martinez-Diaz *et al.*, 2014]. Time sequences related to pen orientation such as the azimuth and altitude have been reported to be useful for some cases [Houmani *et al.*, 2011; Lei and Govindaraju, 2005; Muramatsu and Matsumoto, 2007]. Due to time sequences provide much more discriminative biometric information of the user, local systems usually outperform global systems [Martinez-Diaz *et al.*, 2014; Tolosana *et al.*, 2015d; Van *et al.*, 2007].

5. **Similarity Computation:** This module measures the similarity between the reference signatures acquired in the enrolment stage and the input query signatures, returning a *matching score* value as the output [Diaz *et al.*, 2018b; Martinez-Diaz *et al.*, 2015b]. In global systems, some of the most famous matching techniques are based on Euclidean and Mahalanobis distance, Random Forest, Parzen Windows, Support Vector Machines, and Neural Networks (NNs) [Kareem *et al.*, 2010; Liu *et al.*, 2014; Martinez-Diaz *et al.*, 2014; Parodi and Alewijnse, 2014; Tolosana *et al.*, 2015a]. Local systems try to make the most of the dynamic information of the signature using techniques like DTW, HMMs, GMMs, Time Delay Neural Networks (TDNNs) and Recurrent Neural Networks (RNNs) [Bromley *et al.*, 1993; Fierrez *et al.*, 2007; Jonas and Andrzej, 2003; Martinez-Diaz *et al.*, 2014; Tolosana *et al.*, 2018c].

6. **Score Normalisation:** The matching score may be normalised to a given range using different techniques such as *min-max*, *z-score* or *tanh-estimators*, among many others [Jain *et al.*, 2005]. This module is critical when combining scores from multiple classifiers or in multi-biometric systems [Alonso-Fernandez *et al.*, 2010; Castrillon-Santana *et al.*, 2016; Kittler *et al.*, 1998; Poh *et al.*, 2007]. More sophisticated techniques like target-dependent score normalisation can lead to an improved system performance [Fierrez-Aguilar *et al.*, 2005c].

Finally, an input query signature will be considered to belong to the claimed user if its matching score exceeds a given threshold. This threshold could be modified in real applications depending on the importance of the operation, e.g., do the shopping on Amazon or make a transaction of 10,000 euros.

2.1.2. Global Systems

Global systems (a.k.a. feature-based systems) have been exhaustively analysed in the last centuries as a robust way to verify the identity of the user [Diaz *et al.*, 2018b; Fierrez and Ortega-Garcia, 2008; Plamondon and Srihari, 2000]. This approach consists of extracting discriminative features from the whole signature, creating a final feature vector that represents a specific user. Many different global features have been proposed along the years. Fierrez *et al.* originally proposed a set of 100 features related to time, kinematic, direction and geometry information [Fierrez-Aguilar *et al.*, 2005b]. Other authors have also evaluated the discriminative power of other features for user authentication. In [Sae-Bae and Memon, 2014], the authors proposed a feature vector derived from attributes of several histograms that can be computed in linear time. Those histogram-based features were designed to capture essential attributes of the signature as well as relationships between these attributes. Other approaches have extracted features related to the Kinematic Theory of the rapid human movements and its associated Sigma LogNormal model. In [Gomez-Barrero *et al.*, 2015], the authors proposed a set of 4 features in order to analyse the variations of the neuromuscular responses in the whole signature. A novel set of features was recently proposed in [Diaz *et al.*, 2018a]. In that study, Diaz *et al.* proposed a new feature space based on characterising the movement of the shoulder, the elbow and the wrist joints when signing. As this motion is not directly obtained from a digital tablet, the new features were calculated by means of a virtual skeletal arm (VSA) model, which simulated the architecture of a real arm and forearm. It is also interesting to highlight the approach proposed in [Zeinali *et al.*, 2017]. In that study the authors applied the concept of i-vectors, widely used in speaker and language recognition, to the signature verification task extracting a fixed-length vector for each signature. Other interesting approaches have been also proposed in the following studies [Galbally *et al.*, 2015; Guru and Prakash, 2009; Lee *et al.*, 1996; Parodi and Alewijnse, 2014; Parziale *et al.*, 2013; Richiardi *et al.*, 2005; Sharma and Sundaram, 2016]. Due to the large number of features proposed in the literature, it is common to use different algorithms in order to select the most discriminative features for each scenario. Among all the feature selection techniques proposed, one of the best performing techniques is Sequential Forward Floating Search (SFFS) [Jain and Zongker, 1997; Pudil *et al.*, 1994]. This technique is explained in detail in Sec. 2.1.4. Finally, the similarity computation module is based on techniques such as Euclidean and Mahalanobis distance, Random Forest, Parzen Windows, Support Vector Machines, Neural Networks, etc [Kareem *et al.*, 2010; Liu *et al.*, 2014; Martinez-Diaz *et al.*, 2014; Parodi and Alewijnse, 2014; Tolosana *et al.*, 2015a].

2.1.3. Local Systems

Local systems (a.k.a. time functions-based systems) consider the time sequences of the signatures to discriminate between users. This way more comprehensive user models can be created as every single moment of the signing process is considered. Different local features have been proposed in the literature. In [Fierrez-Aguilar *et al.*, 2005b; Martinez-Diaz *et al.*, 2014] the authors proposed a set of 27 time functions related to spatial coordinates, pressure, pen angular orientations and geometric information of the signatures. Other local features related to the Kinematic Theory of the rapid human movements have been studied in [Diaz *et al.*, 2016a; Fischer and Plamondon, 2017]. In [Fischer and Plamondon, 2017], the authors proposed a set of 18 new dynamic features extracted from the Sigma LogNormal writing generation model, demonstrating that this neuromuscular analysis is complementary to a well-established signature verification system. Among all the possible techniques studied in signature verification, the most famous are DTW, HMM, and GMM [Fierrez *et al.*, 2007; Jonas and Andrzej, 2003; Martinez-Diaz *et al.*, 2014]. These algorithms are explained with more details in Sec. 2.1.3.1 and 2.1.3.2. Multi-algorithm approaches have been also studied in order to enhance the robustness of the on-line signature verification systems [Fierrez-Aguilar *et al.*, 2005a; Pirlo *et al.*, 2014; Sharma and Sundaram, 2017]. Finally, authentication systems based on the combination of global and local systems have been proposed in the literature as well [Fierrez-Aguilar *et al.*, 2005b].

2.1.3.1. Dynamic Time Warping

DTW is one of the most popular algorithms in on-line signature verification. It achieves very robust results and does not require to build any statistical user model as it is an elastic technique algorithm. DTW is an application of Dynamic Programming to the problem of matching time sequences. Yasuhara and Oka were the first to report in [Yasuhara and Oka, 1978] its suitability for on-line 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 [Yasuhara and Oka, 1977] in the field of speech recognition. The goal of DTW is to find an elastic match among samples of a pair of sequences $\mathbf{X}$ and $\mathbf{Y}$ that minimise a given distance measure. The algorithm may be defined as follows [Yasuhara and Oka, 1977]. Let's define two sequences

$$\begin{aligned} \mathbf{X} &= \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i, ..., \mathbf{x}_I \\ \mathbf{Y} &= \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_j, ..., \mathbf{y}_J \end{aligned} \tag{2.1}$$

and a distance measure as

$$d(i, j) = \|\mathbf{x}_i - \mathbf{y}_j\| \tag{2.2}$$

between sequence samples. A warping path can be defined as

$$\mathbf{C} = \mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k, ..., \mathbf{c}_K \tag{2.3}$$

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

$$g_1 = g(1,1) = d(1,1) \cdot w(1) \tag{2.4}$$

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

$$g_k = g(i,j) = \min_{c_{k-1}} \left[g_{k-1} + d(\mathbf{c}_k) \cdot w(k) \right] \tag{2.5}$$

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

$$D(\mathbf{X}, \mathbf{Y}) = \frac{g_K}{\sum_{k=1}^{K} w(k)} \tag{2.6}$$

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.

Despite the algorithm was first used 40 years ago, many authors have recently proposed new advancements and approaches over the traditional DTW algorithm [Faundez-Zanuy, 2007; Fischer and Plamondon, 2017; Kholmatov and Yanikoglu, 2005; Sharma and Sundaram, 2017; Xia *et al.*, 2018]. This fact demonstrates its potential for signature verification, especially in those scenarios with low number of training signatures [Diaz *et al.*, 2018b, 2016b]. In fact, this algorithm has defeated other similarity computation algorithms in international competitions such as BioSecure Signature Evaluation Campaign 2009 (BSEC 2009) and SigWiComp 2013 and 2015 [Houmani *et al.*, 2011; Malik *et al.*, 2015, 2013].

2.1.3.2. Hidden Markov Models and Gaussian Mixture Models

HMM and GMM have been further studied in on-line signature verification, proving to achieve remarkable results in scenarios where the number of available signatures per user is large [Dolfing *et al.*, 1998; Fierrez *et al.*, 2007; Galbally *et al.*, 2013; Jonas and Andrzej, 2003; Martinez-Diaz *et al.*, 2007; Sharma and Sundaram, 2017; Van *et al.*, 2007; Yang *et al.*, 1995]. HMM algorithm [Rabiner, 1989] 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. The basic structure of an HMM-based system comprises the following elements:

- Number of hidden states N.
- Number of Gaussian mixtures per state M.

- Probability transition matrix A which contains the probabilities of jumping from one state to another or staying on the same state.

Finding a reliable and robust model structure for on-line signature verification is not a trivial task. The selection of the optimal parameters N and M can severely affect the system performance of our systems as it has been analysed in previous studies [Fierrez *et al.*, 2007]. In that work, Fierrez *et al.* evaluated the performance of an HMM-based system for different values of N and M. In the present Thesis we propose to go further, considering template and system configuration update strategies when the number of user signatures increases with time. In addition, a GMM-based system, which can be seen as a particular case of an HMM-based system with only one hidden state, is also considered.

2.1.4. Feature Selection Algorithms

One of the the key factors in signature verification systems is the selection of the optimal features (i.e., global and local features). In Sec. 2.1.2 and 2.1.3 we have described the widen variety of features proposed in the literature. However, due to the intra- and inter-user variability, among other factors like the application scenario and writing tool, the best system performance is usually achieved through feature selection techniques.

Feature selection techniques try to reduce the dimensionality of the feature vector while optimising the verification performance. Their goal is to find the optimal combination of features according to a given optimisation criteria. Ideally, given a feature vector of F dimensions, all the possible feature combinations from 1 to F should be tested in order to find the optimal feature subset. Unfortunately, this is not feasible in many cases due to the high amount of combinations to perform, which is

$$\sum_{i=1}^{F} \left(\begin{array}{c} F \\ i \end{array} \right)$$

A critical step when performing feature selection is the choice of the optimisation criterion. Two main alternatives can be taken: *filter* and *wrapper* methods [Theodoridis and Koutroumbas, 2008]. 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 system performance of the task under consideration is used as the criterion to be optimised. 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 optimisation criterion. Wrapper methods require in general more computational resources, as the evaluation of the optimisation criterion (e.g., the verification decision) is commonly more complex than the computation of statistical properties of the training data.

Many different feature selection approaches have been proposed in the last years for on-line signature verification. Lee *et al.* [1996] studied algorithms for selecting and orthogonalising features in accordance with the availability of training data and the system complexity level. The authors evaluated these algorithms using several classifiers proposing for the final system a

selection of only 15 features over the original 42 features. Fierrez-Aguilar *et al.* [2005a] carried out some feature selection experiments based on feature ranking according to scalar inter-user class separability. In order to do that, they computed the scalar Mahalanobis distance between the features. Richiardi *et al.* [2005] proposed a new feature selection technique based on a modification of the traditional Fisher ratio for the cost function of the algorithm. This way the authors were able to adapt to the signature verification task where small number of samples are commonly found. Galbally *et al.* [2007] applied two different Genetic Algorithm (GA) architectures to the feature selection problem. They first considered standard GAs with binary coding to find the optimal subset of features that minimise the verification error rate of the system. Then, they studied the phenomenon of the curse of dimensionality using a GA with integer coding. Very interesting findings were extracted regarding what type of features are more discriminative for each type of impostor. Other feature selection techniques based on GAs have been proposed in [Rúa and Castro, 2012]. Martinez-Diaz *et al.* [2008] proposed a new feature selection technique based on Fisher Discriminative Ratio (FDR) in order to analyse the discriminative power of each feature on mobile scenarios. Finally, Parodi and Gomez [2014] proposed a fixed-length representation of the time functions associated with the signatures based on Legendre polynomials series expansions. They also analysed feature combinations in order to provide some insight on their actual discriminative power for Western and Chinese publicly available signature databases.

One of the best feature selection techniques is the SFFS algorithm (a specific case of the floating search), which is considered in this Thesis in order to select the optimal global and local features for each specific scenario. This algorithm was first proposed by Pudil *et al.* [1994] so as to solve some of the problems presented in traditional feature selection techniques, e.g., when a feature is selected, it can no longer be discarded (the nesting effect). SFFS algorithm offers a suboptimal solution since it does not take into account all the possible feature combinations, although it considers correlations between features. This way we can obtain robust feature vectors in a reasonable amount of time. In all cases we consider the EER as the optimisation criterion. This algorithm has been used in previous studies with very good results [Galbally *et al.*, 2013; Jain and Zongker, 1997; Martinez-Diaz *et al.*, 2008, 2014; Tolosana *et al.*, 2015b].

2.1.5. Sequential Forward Floating Search

This section describes the operation of the SFFS algorithm following [Theodoridis and Koutroumbas, 2008].

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 optimisation 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.

2.2. On-Line Signature Verification on Emerging Scenarios

The high deployment of mobile devices and acceptance of the society towards the use of them on daily operations have given rise to new very interesting scenarios and opportunities for on-line signature verification. However, these new scenarios and opportunities also bring up some challenges for the system performance that must be tackled.

2.2.1. Device Interoperability and Finger Input

On-line handwritten signature can be easily acquired through many different devices nowadays. Yet more, the acquisition of the signatures has been expanded from the traditional stylus on high-quality Wacom devices to COTS general purpose devices using even our own finger as the writing input. These aspects have made possible an unprecedented deployment of the signature technology in many different scenarios and applications. However, it is important not to forget that each device and acquisition tool provides different user information and quality that

must be taken into account in order not to degrade the system performance, as it was described in Sec. 2.1.1.

To the best of our knowledge, one of the first studies that evaluated device interoperability scenarios was [Alonso-Fernandez *et al.*, 2005]. In that work, the authors studied the variability of the signatures acquired on access control scenarios using two different tablet PCs (Hewlett-Packard TC1100 and Toshiba Portege M200) and the stylus as the writing input. The authors considered a signature verification system based on a total of 14 local features and HMM algorithm for the similarity computation. They evaluated the performance of the system considering both monosensor and multisensor enrolment and fusion of sensors. Their results shown that, when using the sensor providing less reliable information, verification performance was not much affected by the Tablet PC used for enrolment. However, this fact did not occur when testing with the more reliable sensor, where verification performance drops significantly if they used the other Tablet PC for enrolment. These results remarked the importance of having enrolment models generated with good quality data.

Since that preliminary study carried out in 2005, few works have focused on this important scenario. In [Blanco-Gonzalo *et al.*, 2014], the authors evaluated device interoperability scenarios using a new database in which signatures were captured using tablet PCs, smartphones and tablets. They considered random (zero-effort) forgeries with a signature verification system based on only 4 local features and DTW algorithm. In order to achieve a higher similarity between signatures acquired through different devices, time- and spatial-based preprocessing normalisation techniques were applied. In [Smejkal *et al.*, 2017], the authors evaluated this new scenario using a total of 8 different devices (specifically designed for the acquisition of signatures) from the company Signotec GmbH, setting up the sampling frequency to 250 Hz in all devices in order to decrease the variability between them. The authors evaluated the stability of the users when signing through each of the devices, concluding that signatures are not very affected on those specific devices. Other studies such as [Sae-Bae and Memon, 2014] indirectly considered this scenario through the acquisition of a new database in which each user had to sign in his/her personal smartphone. Despite the importance of the device interoperability scenario in our society where acquisition devices are replaced constantly, none of these studies have completely analysed and compensated the countermeasures of device interoperability scenarios for on-line signature verification.

Finger input scenarios have also attracted the attention of many researchers in the last years. In [Martinez-Diaz *et al.*, 2013], both pen and finger were considered as input in the experimental work. For the finger case, users were asked to perform a simplified version of their signatures (a.k.a. pseudo-signatures) based on their initials or part of their signature flourish. The results using both inputs were analysed, showing a high degradation of the system performance for the finger scenario with results in the range of 20.0% EER. In [Robertson and Guest, 2015], a statistical analysis was conducted to assess consistency between signatures acquired using pen and finger. The results showed a set of local and global features that maintain stability in both scenarios. In [Sae-Bae and Memon, 2014], the authors acquired a database composed

of 6 different sessions. Users were asked to perform their signatures using the finger as input on their own devices. Regarding the experimental work, they considered a global system whose features were extracted from histograms related to X and Y coordinates, speed, angles, pressure, and their derivatives. That approach was evaluated only for random forgeries achieving results between 3.0% and 8.0% EERs. In [Antal and Bandi, 2017], both pen and finger were considered as input. For the pen case, the MCYT database was used whereas for the finger case a new database named MOBISIG was captured using a Nexus 9 tablet with a total of 83 users and 3 acquisition sessions. The results obtained using both global and local signature verification systems showed the worsening of the system performance when the finger was used as input, especially for skilled forgeries with EERs ca. 20.0%. Similar results have been also obtained in other recent studies on the finger scenario using approaches based on autoencoders or simplified versions of DTW [Nam *et al.*, 2016; Tang *et al.*, 2016].

Finally, it is also interesting to remark that in most studies the authors have focused on the analysis of the system performance for each individual writing tool. However, it is also very interesting to analyse the scenario where enrolment and test signatures are captured using different writing tools as this can be a potential application scenario, for example in banking.

This Thesis performs a complete analysis of the effects of these emerging acquisition scenarios on the system performance, covering all details of our proposed approach for both research and industrial applications. Additionally, we acquired and made publicly available the first handwritten signature database (i.e., e-BioSign) that considers all these acquisition scenarios.

2.2.2. Signature Template Aging

The effect of aging on human biometric traits have been studied for many different applications. Some studies analyses the aging effect from a medical point of view and early diagnoses of diseases [Coleman and Grover, 2006; Drempt *et al.*, 2011; Reilly and Plamondon, 2012], while others analyse the countermeasures for the biometric authentication system performance, especially for both face and fingerprint traits [Deb *et al.*, 2018; Galbally *et al.*, 2018; Ling *et al.*, 2007; Mahajan and Sondur, 2018; Modi and Elliott, 2006; Modi *et al.*, 2007; Ramanathan and Chellappa, 2006].

Aging, in terms of the gradual degradation of a system performance due to the changes suffered by the user's trait along the time, has been also studied for on-line signature verification. Although it cannot be strictly considered as aging, several works have analysed 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) [Galbally *et al.*, 2009; Guest, 2006; Houmani *et al.*, 2009]. Among all these studies, it is important to remark the work carried out in [Sae-Bae and Memon, 2014]. In that study the authors evaluated the impact of signature aging in short term and the effectiveness of using a cross-session training strategy. For that purpose, they acquired a database composed of 6 different sessions. Users were asked to perform their signatures using the finger as the writing tool on their own devices. Regarding the experimental work, they considered a global system whose

features were extracted from histograms related to X and Y coordinates, speed, angles, pressure, and their derivatives. Results obtained in that work showed the degradation of the system performance when training and test samples belonged to different sessions. Additionally, they analysed the system performance when signatures from multiple sessions were considered for training, achieving better results compared to the case of using just one session for training.

The first consistent and reproducible evaluation of the template aging effect for on-line signature verification was carried out in [Galbally *et al.*, 2013]. In that study, Galbally *et al.* generated a new database (ATVS Signature Long-Term) from two previous datasets which were acquired, under very similar conditions, in 6 sessions distributed in a 15-month time span. In the evaluation, the authors considered in the experimental work three different systems, representing the current most popular approaches in signature recognition, proving the degradation suffered by this trait with the passing of time.

In order to reduce the impact of the aging effect on the system performance, different template update strategies have been proposed in [Galbally *et al.*, 2013; Sae-Bae and Memon, 2014]. However, in [Galbally *et al.*, 2013] template update strategies were studied considering only the case of random forgeries. Additionally, signatures from the same session were considered for training and testing, so not meaningful conclusions could be extracted. In [Sae-Bae and Memon, 2014] the database considered was acquired with a very small time gap between the first and last sessions (i.e., only seven days) being difficult to extrapolate these results to real long-term scenarios (e.g., time gap of several months between the training and test signatures). It would be also difficult to know whether the improvement achieved in that work was produced due to the increasing number of signatures used in the different experiments or due to the reduction of the template aging effect. It is also worth mentioning that only the case of random forgeries was considered in that work, as skilled forgeries were not performed during the acquisition of the database. Therefore, we consider necessary to make an exhaustive study of template update strategies on real long-term scenarios in order to reduce the template aging effect for on-line signature verification and get feasible authentication systems regarding computational cost and resources.

In this Thesis we focus on current scenarios where the number of signatures acquired per user can rapidly increase as in real banking or commercial applications nowadays. Therefore, these signatures can be used to update users templates and reduce the aging effect. In addition to perform a complete template update analysis, we also study system configuration update strategies in order to select the optimal system configuration parameters regarding the number of available training signatures per user.

2.3. Signature Complexity

Handwritten signature is a biometric trait highly sensitive to the signature complexity. This aspect has been analysed in previous studies. In [Fairhurst and Kaplani, 1998], a total of 36 subjects were asked to assign a score based on visually appearance complexity to five different

users whose signatures were of varying length, number of strokes, and with differing degrees of embellishment in signing execution. The results demonstrated that while at the extremes of the scale there is a modest spread in the perceived degree of complexity, the intermediate complexity level appears to be much more difficult to assess and categorise quantitatively. A similar study focused on assessing how signature complexity affects when forging signatures was carried out in [Brault and Plamondon, 1993]. In that work an automatic difficulty coefficient was proposed to measure the difficulty that could be experienced by a typical imitator in reproducing signatures both visually and dynamically. Results obtained using their proposed difficulty coefficient were compared to the opinions of the imitators themselves and an expert document examiner. In [Alonso-Fernandez *et al.*, 2007], the authors evaluated the effect of complexity and legibility of signatures for off-line signature verification (i.e., signatures with no available dynamic information) pointing out the differences in performance for several matchers. Signature complexity has also been associated to the concept of entropy, defining entropy as the inherent information content of biometric samples [Daugman, 2003; Lim and Yuen, 2016]. In [Houmani *et al.*, 2008] a "personal entropy" measure based on HMM was proposed in order to analyse the complexity and variability of on-line signatures regarding three different levels of entropy. Results proved that lower entropy is achieved for those signatures with a longer production time and an appearance more related to handwriting. In addition, the same authors have proposed a new metric known as "relative entropy" for classifying users into animal groups (see the biometric menagerie [Yager and Dunstone, 2010a]) where skilled forgeries are also considered [Houmani and Garcia-Salicetti, 2016]. Despite all the studies performed in the on-line signature trait, none of them have exploited, as far as we are aware, the concept of complexity in order to develop more robust and accurate on-line signature verification systems, which is one of the objects to study in this Thesis.

2.4. Deep Learning

2.4.1. Introduction

In general, the main purpose of DL is to change the representation of our original data into a new dimensional space able to achieve a higher separability between different classes. This fact is achieved through the combination of linear and nonlinear transformations. Among all nonlinear transformations, the most common activation functions are sigmoid, softmax, hyperbolic tangent (a.k.a. tangh) and rectified linear unit (a.k.a. ReLU) [Goodfellow *et al.*, 2016]. The basis of DL relies on the feedforward deep network (a.k.a. multilayer perceptron (MLP)). These models are called feedforward as the information flows from the input to the output. There are no feedback connections in which outputs of the model are fed back into itself. These networks consist of three or more layers, i.e., one input and output layer and one or more hidden layers. The input layer is also known as visible layer as it contains the variables of our problem that we are able to observe whereas hidden layers extract features whose values are not given in the

original data. Additionally, each layer can be composed of one of more several units. For the case of the input layer, each input node can refer to a different source of information of our task. In the hidden layers, these nodes are usually known as hidden units or neurons. Each of them applies nonlinear operations (e.g., sigmoid or tanh) in order to change the dimensional space of our original data. Finally, in the output layer, the number of output units change regarding the purpose of the task. For example, for classification is common to use a softmax activation with one output unit per class whereas for the task of verification, the sigmoid activation with just one output unit is considered.

2.4.2. Architectures

2.4.2.1. Multilayer Perceptron

The functioning of MLP is first described for a good understanding of the neural network basis. The information available to face our task is first inserted to the input layer. This information flows from the input units to the neurons of the first hidden layer, receiving each of the neurons a weighted amount of information from the input units that is controlled through the weight matrix W and bias b. After that, each of the neurons of the hidden layer apply a nonlinear operation such as sigmoid or tangh in order to change the dimensional space of the original features. Therefore, the output of one hidden unit can be denoted as $y = f(Wx + b)$, where y is the output of the hidden units, f the nonlinear operation and x the original information of the input layer. Then, the output of this hidden layer serves as the input of the next hidden layer repeating the same procedure commented before (or directly to the output layer in case the network consists of a single hidden layer). Therefore, the weight matrices W are the only parameters that must be learnt for the correct operating of the network. To achieve that, backpropagation algorithm is usually considered for the minimisation of errors through stochastic gradient descendent [Goodfellow *et al.*, 2016; Schmidhuber, 2015]. From the original MLP networks up to now, many different neural network architectures have been proposed.

2.4.2.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been one of the most successful network architectures for input images in the last years. Some of their key design principles were drawn from the findings of the Neurophysiologists Nobel Prizes David Hubel and Torsten Wiesel in the field of human vision [Goodfellow *et al.*, 2016]. CNNs are mainly composed of convolutional and pooling layers. The former extracts patterns from the images through the application of several convolutions in parallel to local regions of the images. These convolutional operations are carried out by means of different kernels (adapted by the learning algorithm) that assign a weight to each pixel of the local region of the image depending on the type of patterns to be extracted. Therefore, each kernel of one convolutional layer is focused on extracting different patterns such as horizontal or vertical edges. The output of these operations produces a set of linear activations (a.k.a. feature map) that serve as input to nonlinear activations such as the

ReLU function. Finally, it is common to use pooling layers to make the representation invariant to small translations of the input. The pooling function replaces the output of the network at a certain region with a statistical summary of the nearby outputs. For instance, the max-pooling function selects the maximum value of the region.

2.4.2.3. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are becoming more and more important nowadays for modelling sequential data with arbitrary length. They are defined as a connectionist model containing a self-connected hidden layer. One benefit of the recurrent connection is that a memory of previous inputs remains in the network internal state, allowing it to make use of past context. However, the range of contextual information that standard RNNs can access is very limited due to the well-known vanishing gradient problem [Graves *et al.*, 2009]. Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997] and Gated Recurrent Unit (GRU) [Cho *et al.*, 2014a,b; Chung *et al.*, 2014] are RNN architectures that arose with the aim of resolving the shortcomings of standard RNNs. Additionally, bidirectional schemes (i.e., BRNNs) have been studied in order to provide access not only to the past context but also to the future [Schuster and Paliwal, 1997]. Due to the nature of the on-line signature verification task (time sequences), this neural network architecture is studied in Chapter 4.2 of the Thesis.

2.4.2.4. Others

In addition to the aforementioned models, it is important to mention some other well-known DL architectures.

An **autoencoder** is the combination of an encoder function, which converts the input data into a different representation, similar to Principal Component Analysis (PCA), and a decoder function, which converts the new representation back into the original format. Therefore, an autoencoder is trained to attempt to copy its input to its output. Some examples of use can be seen in [Hong *et al.*, 2015; Zeng *et al.*, 2018] for the task of human pose recovery and facial expression recognition.

Generative Adversarial Networks (GANs) comprise two different networks, a generative model (G) that captures the data distribution, and a discriminative model (D) that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game [Goodfellow *et al.*, 2016, 2014]. This network has achieved tremendous impact in the last years as it can be seen in the last big conferences such as Conference on Computer Vision and Pattern Recognition (CVPR) [Chen and Hays, 2018; Choi *et al.*, 2017; Wang *et al.*, 2017; Yang *et al.*, 2017].

Most of the aforementioned neural network models belong to the **supervised learning** area. These networks are trained feeding the input layer of the network with the data of our task and indicating to the training algorithm the output of the network expected. Therefore, each input

of the network has its corresponding label. However, **unsupervised learning** has received a lot of interest in the last years due to the majority of the available information nowadays is not labelled. Additionally, unsupervised learning techniques can be very useful to train neural networks in the first stages [Erhan *et al.*, 2010; Radford *et al.*, 2015].

Finally, DL can be used in two different modes. On the one hand, some authors have proposed the use of the networks as an *end-to-end* approach. This term refers to the case where the network is used for both feature extractor and classification. Therefore, the network is fed with the raw data and the network is in charge of both selecting the relevant features for the task and performing the classification. This approach has been studied in many different tasks such as speech and text recognition [Graves and Jaitly, 2014; Wang *et al.*, 2012]. On the other hand, the network can be used as a feature extractor in order to obtain a more rich representation of the task (for example removing the last fully-connected layers of a CNN model). These features can be then used to feed traditional algorithms such as HMM or SVM [Lozano-Diez *et al.*, 2017; Nogueira *et al.*, 2016].

2.4.3. Deep Learning for On-Line Signature Verification

Despite the good results obtained in the field of handwriting recognition, and the similarity with the task of handwritten signature, very few studies have successfully applied RNN DL architectures to on-line handwritten signature verification. In [Tiflin and Omlin, 2003], the authors proposed the use of a system based on LSTM for on-line signature verification. Different configurations based on the use of forget gates and peephole connections were studied considering in the experimental work a small database with only 51 users. The LSTM system proposed in that work seemed to authenticate genuine and impostor cases very well. However, as it was pointed out in [Otte *et al.*, 2014], the method proposed for training the LSTM system was not feasible for real applications for various reasons. First, the authors considered the same users for both development and evaluation of the system. Moreover, the system should be trained every time a new user was enroled in the application. In addition, forgeries were required in that approach for training, which may not be feasible to obtain as well. Besides, the results obtained in [Tiflin and Omlin, 2003] cannot be compared to any state-of-the-art signature verification system as the traditional measures such as the EER or calibrated likelihood ratios were not considered. Instead, they just reported the errors of the LSTM-outputs. In order to find some light on the feasibility of RNNs for signature verification purposes, Otte *et al.* performed in [Otte *et al.*, 2014] an analysis considering three different real scenarios: *i)* training a general network to distinguish forgeries from genuine signatures on a large training set, *ii)* training a different network for each writer that works perfectly on the training set, and *iii)* training the network considering only genuine signatures. However, all experiments failed obtaining a 23.75% EER for the best configuration, far away from the best state-of-the-art results and concluding that LSTM RNN systems trained with standard mechanisms were not appropriate for the task of signature verification as the amount of available data for this task is scarce compared to other tasks such as handwriting recognition.

After the publication of our novel DL study presented in Chapters 4 and 7 of the Thesis, new DL approaches have been applied to on-line handwritten signature verification with success [Ahrabian and Babaali, 2017; Lai *et al.*, 2017]. In [Ahrabian and Babaali, 2017], the authors proposed first the use of autoencoders as feature extractor, and then a Siamese network based on MLP for the final verification process. Experiments were carried out using the SigWiComp2013 and GPDS Synthetic databases with very good results. Lai *et al.* proposed in [Lai *et al.*, 2017] a new RNN system based on GRU as feature extractor. The training objective was focused on the minimisation of intra-class variations and the maximisation of the distances between skilled forgeries and genuine signatures, which was achieved through triplet loss and center loss [Hoffer and Ailon, 2015; Wen *et al.*, 2016]. Experiments were carried out using SVC-2004 database achieving a final value of 2.37% EER for skilled forgeries.

2.5. Handwriting Biometrics and Beyond

Touch biometrics are becoming a very attractive way to verify users on mobile devices [Fierrez *et al.*, 2018b; Tolosana *et al.*, 2017a]. Table 2.1 summarises relevant approaches in this area. For each study, we include information related to the verification method, features, classifiers and datasets considered. We also report in Table 2.1 the verification performance for the two impostor scenarios commonly considered in this area [Tolosana *et al.*, 2018e]: *i) imitation attack*, the case in which impostors have some level of information about the user being attacked; and *ii) random attack*, the case in which no information about the user being attacked is known. Note that most algorithms and experimental conditions vary between the listed works, e.g., the amount and type of training and testing data. Therefore, Table 2.1 should be mainly interpreted in general terms to compare different scenarios of use based on touch biometrics, but not individual algorithms.

Angulo and Wastlund [2011] evaluated the use of lock pattern dynamic systems for user authentication. Users were asked to draw three different lock patterns a certain number of times (50 trials for each pattern), with each pattern consisting of six dots. Authors considered a total of 11 timing-related features extracted from the finger-in-dot time (i.e., the time in milliseconds from the moment the participant finger touches a dot to the moment the finger is dragged outside the dot area), and the finger-in-between-dots time (i.e., representing the speed at which the finger moves from one dot to the next) achieving results above 10.0% EER for imitation attacks. In [Lacharme and Rosenberger, 2016], the authors incorporated biometric dynamic features related to the position of the finger, pressure, finger size and accelerometer sensor to the traditional Android unlock patterns, achieving a final 15.0% EER for imitation attacks using a matching algorithm based on Hamming Distance. Zezschwitz *et al.* [2016] presented a similarity metric for Android unlock patterns to quantify the effective password space of user-defined gestures. The proposed metric was evaluated using 506 user-defined patterns revealing very similar shapes that only differ by simple geometric transformations such as rotation. Consequently, they presented an approach to increase the pattern diversity in order to strengthen user lock patterns.

Table 2.1: *Comparison of different touch biometric approaches for mobile scenarios. Acc = Accuracy.*

Study	Method	Features	Classifiers	Verification Performance		# Participants (Dataset)
				Random Attack	Imitation Attack	
[Angulo and Wastlund, 2011]	Lock Pattern Dynamics	Timing-related Features	Random Forest	-	EER = 10.39%	32
[Lacharme and Rosenberger, 2016]	Lock Pattern Dynamics	Dynamic Features	Hamming Distance	-	EER = 15.0%	34
[Zezschwitz *et al.*, 2016]	Lock Pattern Dynamics	Shape Features	Greedy Clustering	-	-	506
[Buschek *et al.*, 2015b]	Keystroke	Font Adaptation Features	Manual	Acc = 94.8%	-	91
[Buschek *et al.*, 2015a]	Keystroke	Touch-specific Features	GM, kNN, LSAD	EER = 13.74%	-	28
[Li *et al.*, 2013]	Touchscreen Gestures	Static Features	SVM	EER = 3.0%	-	75
[Sae-Bae *et al.*, 2014]	Touchscreen Gestures	Distance between Points	DTW	EER = 1.58%	-	34
[Shen *et al.*, 2016]	Touchscreen Gestures	Static Featurs	SVM, Random Forest, kNN, Neural Networks	EER $\sim$ 3.0%	-	71
[Fierrez *et al.*, 2018b]	Touchscreen Gestures	Static Features	SVM, GMM	EER = 10.7%	-	190
[Sae-Bae and Memon, 2014]	Handwritten Signatures	Histogram Static Features	Manhattan Distance	EER = 5.04%	-	180
[Tolosana *et al.*, 2017a]	Handwritten Signatures	Dynamic Features	DTW	EER = 0.5%	EER = 17.9%	65
[Khan *et al.*, 2011]	Graphical Passwords	Predefined Symbols	Exact Match	-	-	100
[Martinez-Diaz *et al.*, 2016]	Graphical Passwords	Dynamic Features	DTW, GMM	EER = 3.4%	EER = 22.1%	100
[Kutzner *et al.*, 2015]	Handwritten Password	Static and Dynamic Features	Bayes-Nets KStar, kNN	-	FAR = 10.42% FRR = unknown	32
[Nguyen *et al.*, 2017a]	Handwritten Digits	Dynamic Features	DTW	-	EER = 4.84%	20
Proposed Approach	**Handwritten Digits**	**Dynamic Features**	**DTW, RNNs**	**-**	**EER = 3.8%**	**93**

Other studies have focused on the potential of keystroke biometrics for user authentication on mobile scenarios. Buschek *et al.* [2015b] introduced qualitative aspects like personal expressiveness in order to enhance traditional keystroke biometric systems based on quantitative factors such as error rates and speed. They introduced a dynamic font personalisation framework, TapScript, which adapted a finger-drawn font according to user behavior and context, such as finger placement, device orientation, and position of the user while typing (i.e., walking or sitting), resulting in a handwritten-looking font. Following their new approach, users were able to distinguish pairs of typists with 84.5% accuracy and walking/sitting scenarios with 94.8%. The same authors compared in [Buschek *et al.*, 2015a] touch-specific features between three different hand postures (i.e., one-thumb, two-thumb and index finger typing) and evaluation schemes: Gaussian Model without covariance (GM), k-Nearest-Neighbours (kNN) and Least Squares Anomaly Detection (LSAD). Authors concluded that spatial touch features reduce the EER by 26.4 - 36.8% compared to the traditional temporal features.

Biometric verification systems based on touchscreen gestures (i.e., scrolling, zooming and clicking) while using mobile devices in scenarios such as document reading, web surfing or free tasks are gaining a lot of impact nowadays [Fierrez *et al.*, 2018b; Li *et al.*, 2013; Sae-Bae *et al.*, 2014; Shen *et al.*, 2016]. These approaches enable active or continuous authentication schemes, in which the user is transparently authenticated [Patel *et al.*, 2016; Serwadda *et al.*, 2013]. Different features and algorithms have been proposed in this field achieving very good results against random attacks. In [Sae-Bae *et al.*, 2014], the authors proposed a set of 22 multitouch gestures using characteristics of hand and finger movements with an algorithm robust to orientation and translation achieving a final result of 1.58% EER. In [Fierrez *et al.*, 2018b], a set of 100 static features extracted from swipe gestures and systems based on SVM and GMM were considered obtaining performances up to 10.7% EER. Very good results have been also achieved in [Li *et al.*, 2013; Shen *et al.*, 2016] using verification algorithms such as SVM, kNN, Random Forest and Neural Networks.

Handwritten signature is one of the most socially accepted biometrics as it has been used in financial and legal agreements for many years [Diaz *et al.*, 2018b; Fierrez and Ortega-Garcia, 2008; Plamondon and Srihari, 2000; R. Plamondon and G. Pirlo and D. Impedovo, 2014], and it also finds applications in mobile scenarios. However, a considerable degradation of the system performance with results around 20.0% EER is obtained for imitation attacks when testing on mobile scenarios using finger touch as input [Sae-Bae and Memon, 2014; Tolosana *et al.*, 2017a]. The main reason for such degradation of the system performance is studied in the experimental chapters of this Thesis. Graphical passwords were studied in [Khan *et al.*, 2011; Martinez-Diaz *et al.*, 2016]. In [Martinez-Diaz *et al.*, 2016], the authors proposed an approach based on graphical passwords (doodles) achieving final results above 20.0% EER for imitation attacks. The main reason for such degradation of the system performance laid down on the specific task that the user needed to perform to be authenticated, e.g., doodles were difficult to memorise for most of the users as they didn't use them on a daily basis.

Finally, strongly related to the study carried out in this Thesis, in [Kutzner *et al.*, 2015;

Nguyen *et al.*, 2017a] the authors proposed the use of handwritten passwords to be authenticated. In [Kutzner *et al.*, 2015], users had to perform an 8-digit password on the screen of a tablet device. For each handwritten password, a total of 25 global and local features were extracted and tested using many different authentication algorithms. However, the authentication scenario considered in that approach restricts the deployment of the technology in real mobile applications as: *i)* the authors considered a large number of training samples (12), and *ii)* it seems to be only applicable to devices with large screens (such as tablets) as it would be very difficult for the users to perform such a long password (8 digits) on a screen of much smaller size. Nguyen *et al.* [2017a] evaluated the use of handwritten touch biometrics for PIN-based authentication systems. Their proposed authentication approach overcame some of the drawbacks previously cited as they asked users to draw each digit of the PIN one by one. A final 4.84% EER was achieved using a biometric system composed of 5 local features and a matcher algorithm based on DTW.

2.6. Chapter Summary and Conclusions

In this chapter we have summarised the main studies related to this Thesis. We have first described each module of traditional on-line signature verification systems as well as the two modalities considered, i.e., feature-based systems (a.k.a. global systems) and time functions-based systems (a.k.a. local systems). Second, we have concentrated on some of the signature verification emerging scenarios considered in this Dissertation and perform a thorough overview of related works in this line. Then, other recent interesting on-line signature research topics have been described. In the fourth section of this chapter we have highlighted the importance and success of DL approaches. We have also described the basis of deep neural networks, the different topologies commonly used and its feasibility for on-line signature verification through preliminary studies. The last section surveys and compares advantages and limitations of current state-of-the-art touchscreen biometric approaches.

Chapter 3

Signature and Handwriting Databases

This chapter is organised as follows. First, Sec. 3.1 gives an overview of the most relevant features of existing on-line signature databases, making special emphasis on the databases used in the experimental work of this Thesis. Then, Sec. 3.2 presents the new e-BioSign database, as well as the extension of the ATVS On-Line Signature Long-Term database, acquired during the execution of the Thesis. These new on-line signature databases have been already released to the research community. Sec. 3.3 introduces the new e-BioDigit database acquired for the purpose of incorporating handwriting biometric information to traditional passwords. Conclusions are finally drawn in Sec. 3.4.

This chapter is based on the following publications: [Tolosana *et al.*, 2018b, 2017a, 2018d, 2015e; Vera-Rodriguez *et al.*, 2015].

3.1. Existing On-Line Signature Databases

3.1.1. Overview

Many efforts have been carried out in the signature biometrics community in order to capture large and reliable databases. In Table 3.1 we summarise the most relevant features of the main existing on-line signature databases. It is important to highlight the two largest databases acquired (i.e., Biosecure [Ortega-Garcia *et al.*, 2010] and BiosecurID [Fierrez *et al.*, 2010]) with several hundreds of users, which are extensions of the largely used MCYT [Ortega-Garcia *et al.*, 2003]. These three databases were collected by public institutions and have been extensively used by the signature research community for improving the state-of-the-art on many different scenarios [Martinez-Diaz and Fierrez, 2015]. However, the type and quality of the devices used nowadays, apart from the acquisition scenarios considered, have significantly changed compared to the same ones followed in [Fierrez *et al.*, 2010; Ortega-Garcia *et al.*, 2010, 2003]. New databases have recently appeared in the last years considering COTS devices for the acquisition

of the signatures. In [Blanco-Gonzalo *et al.*, 2014], the authors considered a total of 7 devices (tablet PCs, smartphones and tablets) for the acquisition of the signatures. Device interoperability scenarios were evaluated for the case of random forgeries, encouraging further research to improve the system performance. However, as far as we know, that database is not publicly available to the research community. Antal and Bandi released the MOBISIG database [Antal and Bandi, 2017], which comprises pseudo-signatures for a total of 83 users. The database was captured in three sessions resulting in 45 genuine signatures and 20 skilled forgeries per user. Regarding the acquisition device, pseudo-signatures were acquired using the finger on a Nexus 9 Tablet device. A more realistic analysis was carried out in [Sae-Bae and Memon, 2014]. In that work, Sae-Bae and Memon captured a new database for a total of 180 users and 6 different acquisition sessions. Users were asked to perform their signatures using the finger as input on their own devices. Their proposed approach based on features extracted from histograms achieved results between 3.0% and 8.0% EER for random forgeries. However, as far as we know, that database is not publicly available to the research community.

The effect of the signature aging on the system performance was first analysed in [Galbally *et al.*, 2013] through the ATVS On-Line Signature Long-Term database. In that database, signatures were acquired in six different sessions during a 15-month time interval. An exhaustive analysis of the aging, template and system configuration update strategies have been carried out in this Thesis through an extension of the ATVS On-Line Signature Long-Term database. This database is described in Sec. 3.2.2. An assessment of the age dependency for on-line signature verification was performed in [Guest, 2006] considering a database with a total of 274 users. In [Martinez-Diaz *et al.*, 2013], also both stylus and finger were considered as writing tools in the experimental work. For the finger case, users were asked to perform a simplified version of their signature (a.k.a. pseudo-signatures) based on their initials or part of their signature flourish. Results obtained in that preliminary work showed its feasibility and the necessity of further research toward practical application of such mixed writing-input. In this Thesis, we have acquired and made public to the research community a large signature database (e-BioSign) acquired from 5 different COTS devices in total, and considering both pen stylus and also the finger. The complete design and acquisition of the e-BioSign database is described in Sec. 3.2.1.

Table 3.1: *Most relevant features of existing on-line signature databases.*

	Year	Users	Sessions	#genuine samples/user/device	#forgeries/user/device	Device (writing tool)	Best performance (EER(%))
MOBISIG* [Antal and Bandi, 2017]	2017	83	3	15	20	Nexus 9 Tablet (finger)	Finger. Skilled: 14.3 [Antal and Bandi, 2017] Finger. Random: 1.7 [Antal and Bandi, 2017]
e-BioSign* [Tolosana *et al.*, 2017a]	2016	65	2	8	6	Wacom STU-500 (stylus) Wacom STU-530 (stylus) Wacom DTU-1031 (stylus) Samsung Gal. Note (stylus/finger) Samsung ATIV7 (stylus/finger)	Stylus. Skilled: 7.9 [Tolosana *et al.*, 2017a] Stylus. Random: 0.0 [Tolosana *et al.*, 2017a] Finger. Skilled: 17.9 [Tolosana *et al.*, 2017a] Finger. Random: 0.3 [Tolosana *et al.*, 2017a]
GPDS Synthetic* [Ferrer *et al.*, 2017b]	2016	10,000	-	24	30	Simulated stylus	Stylus. Skilled: 0.25 [Ahrabian and Babaali, 2017] Stylus. Random: 1.83 [Ferrer *et al.*, 2017b]
SigWiComp 2015 [Malik *et al.*, 2015]	2015	30	-	15	10	Anoto Pen (stylus)	Stylus. Skilled: 0.29 (cllr min) [Malik *et al.*, 2015]
ATVS-SLT DB* [Galbally *et al.*, 2013; Tolosana *et al.*, 2015e]	2015	29	6	46	10	Wacom Intuos 3 (stylus)	Stylus. Skilled: 1.4 [Tolosana *et al.*, 2015e] Stylus. Random: 0.0 [Tolosana *et al.*, 2015e]
[Sae-Bae and Memon, 2014]	2014	180	6	5	-	User-Own Devices (finger)	Finger. Random: 3.0-8.0% [Sae-Bae and Memon, 2014]
ATVS-DooDB* [Martinez-Diaz *et al.*, 2013]	2013	100	2	30	20	HTC Touch HD (finger)	Finger. Skilled: 21.0 [Martinez-Diaz *et al.*, 2016] Finger. Random: 7.8 [Martinez-Diaz *et al.*, 2016]
[Blanco-Gonzalo *et al.*, 2014]	2013	43	3	60	-	Wacom Intuos 4 (stylus) Wacom STU-500 (stylus) Asus Eee PC Touch (stylus) Samsung Gal. Note (stylus/finger) BlackBerry Playbook (finger) Apple Ipad2 (finger) Samsung Gal. Tab (finger)	Stylus. Random: 0.58 [Blanco-Gonzalo *et al.*, 2014] Finger. Random: 0.19 [Blanco-Gonzalo *et al.*, 2014]
SUSIG* [Kholmatov and Yanikoglu, 2009]	2009	100	2	20 (visual subcorpus) 8 - 10 (blind subcorpus)	10	Wacom Graphire2 (stylus) ePad-ink (stylus)	Stylus. Skilled: 0.77 [Diaz *et al.*, 2016b] Stylus. Random: 1.23 [Diaz *et al.*, 2016b]
Biosecure* [Ortega-Garcia *et al.*, 2010]	2008	667 (DS2) 713 (DS3)	2	30	20	Wacom Intuos3 (stylus) PDA HP iPAQ (stylus)	Stylus. Skilled: 6.2 Tolosana *et al.* [2015d] Stylus. Random: 2.0 [Tolosana *et al.*, 2015d]
BiosecurID* [Fierrez *et al.*, 2010]	2007	400	4	16	12	Wacom Intuos3 (stylus)	Stylus. Skilled: 4.77 [Gomez-Barrero *et al.*, 2015] Stylus. Random: 0.50 [Gomez-Barrero *et al.*, 2015]
MBioID [Dessimoz and *et al.*, 2007]	2007	120 (approx.)	2	20	-	Wacom Intuos2 (stylus)	-
[Guest, 2006]	2006	274	variable	10 - 74	-	Graphic tablet (stylus)	-
MyIDEA* [Dumas and *et al.*, 2005]	2005	104 (approx.)	3	18	18	Wacom Intuos2 (stylus)	Stylus. Skilled: 13.7 [Humm *et al.*, 2009] Stylus. Random: 4.0 [Humm *et al.*, 2009]
SVC2004* [Yeung and *et al.*, 2004]	2004	100	2	20	20	Wacom Intuos (stylus) PDA (stylus)	Stylus. Skilled: 0.83 Diaz *et al.* [2016b] Stylus. Random: 0.12 [Diaz *et al.*, 2016b]
MCYT-100* [Ortega-Garcia *et al.*, 2003]	2003	100	1	25	25	Wacom Intuos (stylus)	Stylus. Skilled: 2.85 [Diaz *et al.*, 2016b] Stylus. Random: 1.04 [Diaz *et al.*, 2016b]
BIOMET* [Garcia-Salicetti and *et al.*, 2003]	2003	130 106 91	1	15	17	Wacom Intuos2 (stylus)	-

* publicly available databases.

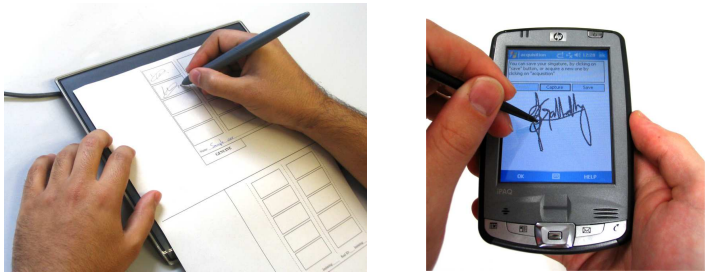
3.1.2. BiosecurID Database

The BiosecurID database [Fierrez *et al.*, 2010] is considered in the experimental work of the Thesis. This database is composed of 16 genuine signatures and 12 skilled forgeries per user, captured in 4 separate acquisition sessions leaving a two-month interval between them. There are a total of 400 users and signatures were acquired considering a controlled and supervised office-like scenario. Users were asked to sign on a piece of paper, inside a grid that marked the valid signing space, using an inking pen. The paper was placed on a Wacom Intuos 3 pen tablet that captured the following time signals: X and Y spatial coordinates (resolution of 0.25 mm), pressure (1024 levels) and timestamp (100 Hz). In addition, pen-up trajectories are available. All the dynamic information is stored in separate text files following the format used in the first Signature Verification Competition (SVC) [Fierrez-Aguilar *et al.*, 2005c; Yeung *et al.*, 2004]. The acquisition process was supervised by a human operator whose task was to ensure that the collection protocol was strictly followed and that the captured samples were of sufficient quality (e.g., no part of the signature outside the designated space), otherwise the subjects were asked to repeat the signature.

3.1.3. Biosecure Database

The Biosecure database [Ortega-Garcia *et al.*, 2010] is also considered in the experimental work of the Thesis. This database comprises two different on-line signature datasets, DS2 and DS3. DS2 dataset was captured under access control scenario where users had to sign while sitting, whereas the DS3 dataset considered a mobile scenario where users had to sign while standing and holding the device in one hand, emulating realistic operating conditions. Furthermore, it is important to highlight that intra-class variability problem is also considered, as Biosecure DS2 and DS3 datasets contain two different acquisition sessions separated by a 3-month time gap between them. DS3 dataset was captured using a PDA HP iPAQ hx2790 with a sampling frequency of 100 Hz, whereas the DS2 dataset was captured with a WACOM Intuos3 A6 pen tablet with a sampling frequency of 100 Hz and signing on a paper sheet placed on top of the device as it can be seen in Fig. 3.1. For both DS2 and DS3 datasets, there is a subset of 120 common users that is considered in the experimental work reported in this Thesis. This subset is considered in Chapter 5 of the thesis in order to study the device interoperability effect. The available information in Biosecure DS2 is the following: X and Y spatial coordinates, pressure, pen angular orientation (azimuth and altitude angles) and timestamp. However, in Biosecure DS3 just X and Y spatial coordinates and timestamp are available.

Regarding the number of signatures, there are a total of 30 genuine signatures (i.e., 15 genuine signatures per session) and 20 skilled forgeries (i.e., 10 skilled forgeries per session) per user and dataset. For the skilled forgeries, users had visual access to the dynamics of the signing process of the signatures they had to forge.



(a) (b)

Figure 3.1: *(a) Pen tablet acquisition scenario in the Biosecure DS2 - Access Control Scenario dataset. (b) PDA signature acquisition scenario in the Biosecure DS3 - Mobile Scenario dataset.*

3.2. Novel Databases

3.2.1. e-BioSign Database

The e-BioSign database, which is publicly available[1], comprises five capturing devices. Three of them are specifically designed for capturing handwritten data (Wacom devices), while the other two are general purpose tablets not designed for that specific task (Samsung tablets). Fig. 3.2 shows an image of the setup used to acquire the database, with all five considered devices.

It is worth noting that all five devices were used with their own pen stylus. Additionally, the two Samsung devices were used with the finger as the writing input, allowing us to analyse the effect of the writing tool on the system performance. The same capturing protocol was used for all five devices: they were placed on a desktop and subjects were told to feel comfortable when writing on them, so a small rotation of the devices was allowed.

The software for capturing handwriting and signatures was developed in the same way for all devices in order to minimise the variability of the user during the acquisition process. A rectangular area with an horizontal line in the bottom part was represented on the device screen, including two buttons "OK" and "Cancel" to press after writing if the sample was good or bad respectively. If the sample was not good, then it was repeated. The nomenclature and a brief description of each device considered in e-BioSign are given next:

1. ***W1: Wacom STU-500***. 5-inch TFT-LCD B/W display, with VGA resolution of 640 × 480 pixels. It has a sampling rate of 200 Hz, and 512 pressure levels. This device gives a very natural feel of writing.

2. ***W2: Wacom STU-530***. Newer version of W1 device. 5-inch TFT-LCD color display, with VGA resolution of 640 × 480 pixels. It has a sampling rate of 200 Hz, and 1024 pressure levels. This device allows safe transactions as it has AES 256 bit / RSA 2048 embedded data encryption.

[1] http://atvs.ii.uam.es/atvs/eBioSign-DS1.html

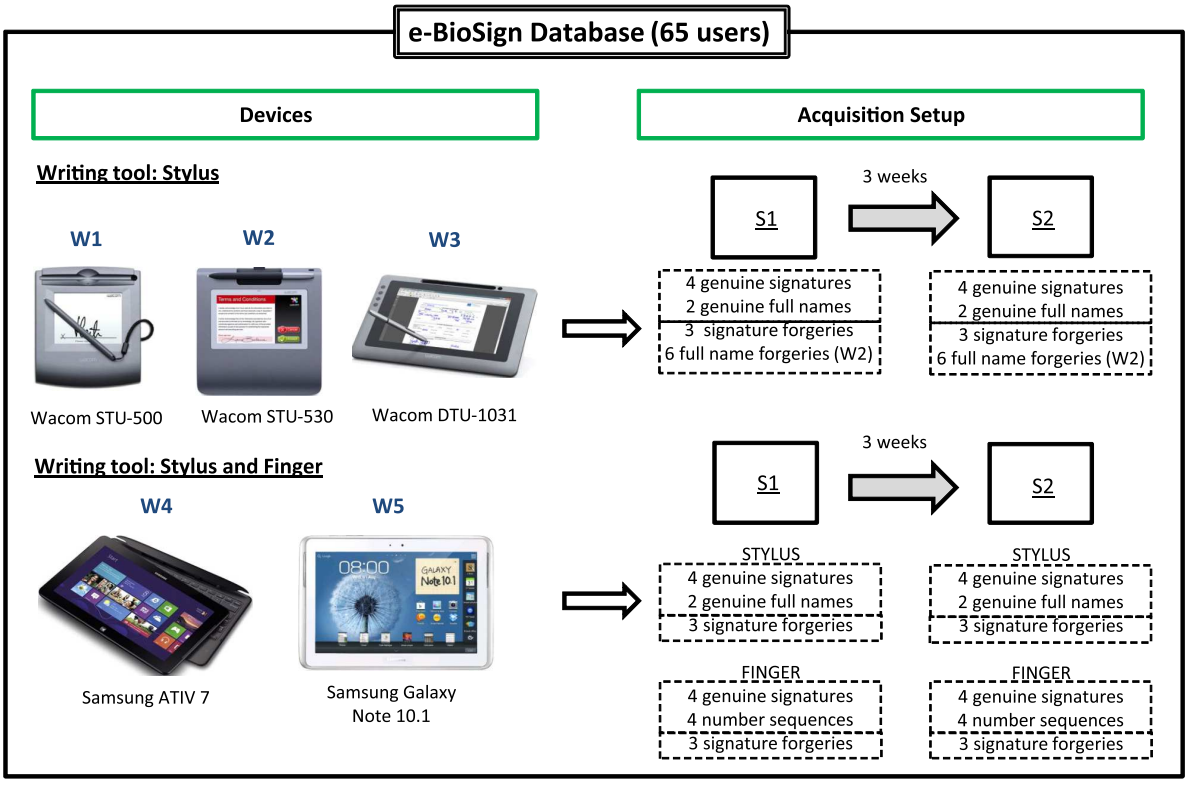


Figure 3.2: *Description of the devices and the acquisition setup considered in the new e-BioSign database. A total of 65 users and 5 different COTS devices are considered (three Wacom and two Samsung general purpose devices). For the two Samsung devices, data is collected using both a pen stylus and also the finger.*

3. ***W3: Wacom DTU-1031***. This device has a larger 10.1-inch color LCD display with a resolution of 1280 × 800 pixels. It has a sampling rate of 200 Hz, and 512 pressure levels. It also provides the same data encryption as W2. It allows to visualize documents on the display before signing them.

4. ***W4: Samsung ATIV 7***. This is a device with Windows 8. It has a 11.6-inch LED display with a resolution of 1920 × 1080 pixels. It has 1024 pressure levels, and contrary to the Wacom devices, the sampling rate is not uniform in this case. This tablet allows to use its own stylus or also the finger, but no pressure information is recorded in this last case.

5. ***W5: Samsung Galaxy Note 10.1***. This is an Android device. It has a 10.1-inch LCD display with a resolution of 1280 × 800 pixels. It has 1024 pressure levels and not uniform sampling rate. This device also allows to use its own stylus or the finger.

Table 3.2 shows the number of samples acquired for each user per session. As previously mentioned, the database was collected in two sessions with a time gap of at least three weeks between them. In each session there were three capturing blocks namely *Genuine 1*, *Genuine*

Table 3.2: *Handwritten samples captured per user and device in each of the two sessions.*

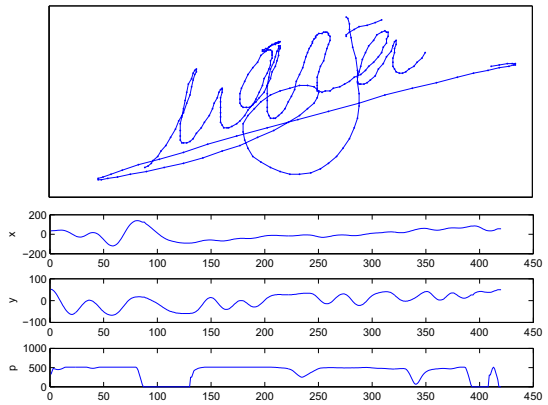
	Block	Stylus	Finger
Signature	Genuine 1	2 (W1 - W5)	2 (W4, W5)
	Genuine 2	2 (W1 - W5)	2 (W4, W5)
	Forgeries	3 (W1 - W5)	3 (W4, W5)
Full name	Genuine 1	1 (W1 - W5)	-
	Forgeries	3 (only W2)	-
Full name capital letters	Genuine 2	1 (W1 - W5)	-
	Forgeries	3 (only W2)	-
Number sequence	Genuine 1	-	2 (W4, W5)
	Genuine 2	-	2 (W4, W5)

2 and *Forgeries*. In *Genuine 1* block, two signatures plus the full name are performed for each device using their own pen stylus. Then, two signatures and a number sequence composed of numbers from 0 to 9 plus a random letter are performed for the two Samsung devices with the finger. Next, *Genuine 2* block is recorded, which comprises the same information as *Genuine 1* block, but in this case the full name is written in capital letters. Finally, the last block *Forgeries* is performed, where each user carries out a forgery of the signatures of the three previous users in the database for each of the 5 devices using the stylus, and also with the finger for the two Samsung devices. Regarding forgeries of the full name, they are only performed for the Wacom STU-530 both for lower and upper case writing. In order to perform high-quality forgeries, users are allowed to visualize a recording of the dynamic realization of the signature to forge.

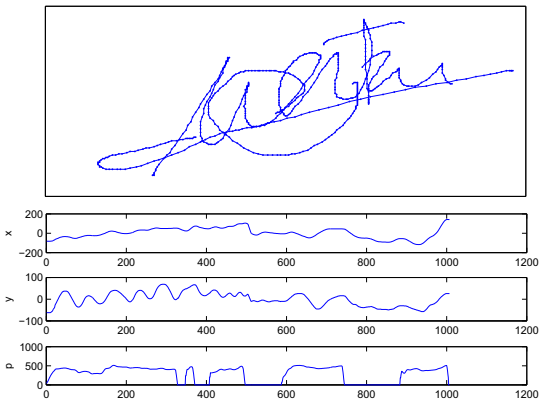
In the second session, the procedure is identical, except one difference in the *Forgeries* block. In this case, impostors forge the same users as in session one, but this time a paper with the image of the signatures and names to forge is placed over the screen devices so the users can overwrite to perform the forgeries. Nevertheless, they are not allowed to see the recordings of the signatures in this case.

In total there are 6,370 signatures, of which 3,640 are genuine samples and 2,730 are skilled forgeries. From the total, 4,550 were performed with the stylus and 1,820 with the finger. There are a total of 2,080 handwritten names, of which 1,300 are genuine samples and 780 are forgeries (only for Wacom STU-530). Also, half of the samples are done with natural writing and the other half in capital letters. Finally, there are 1,040 genuine alphanumeric sequences carried out for the two Samsung devices using the finger.

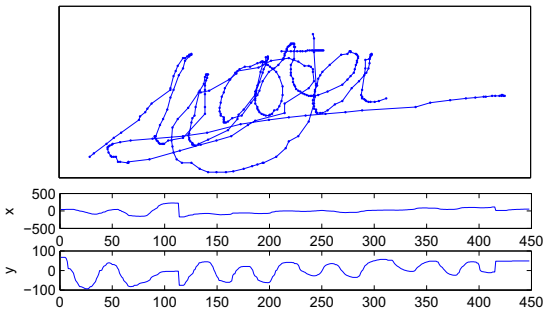
The whole capturing process was supervised by an operator who explained all the steps that donors had to follow. Therefore, this is a multi-session and multi-device database with samples captured using both stylus and finger as the writing input for signature and handwritten data. Fig. 3.3 shows examples of the data collected in e-BioSign for the Samsung Galaxy Note 10.1 (W5), as this device contains all types of information collected, i.e., signatures (genuine and forgeries) using the pen stylus and the finger, full name in lower and upper cases (only genuine as the forgeries were only performed for the Wacom STU-530) and number sequences made with the finger. Fig. 3.3.E and 3.3.F are just examples in order not to reveal the name of any user of the database. The rest of the samples are contained in the database. It is worth noting that data collected using the finger for Samsung ATIV 7 and Galaxy Note 10.1 do not contain



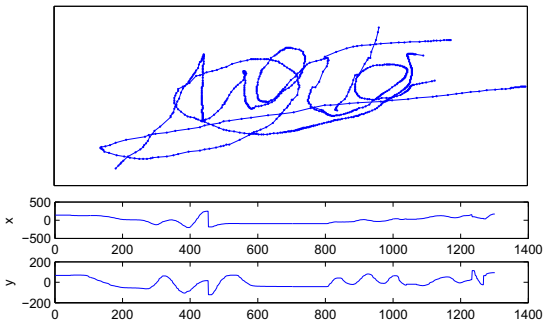
(a) Genuine Signature Pen Stylus



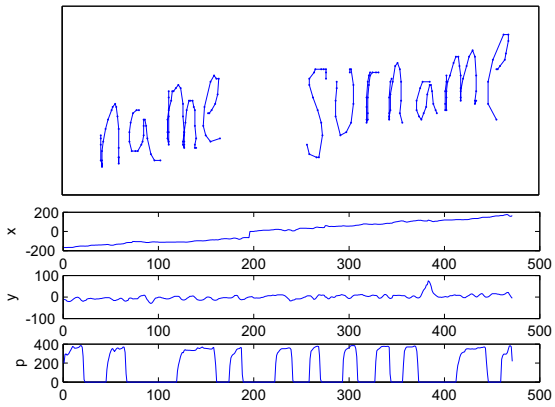
(b) Forgery Signature Pen Stylus



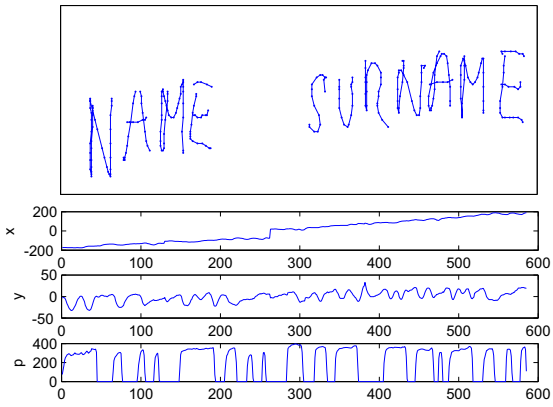
(c) Genuine Signature Finger



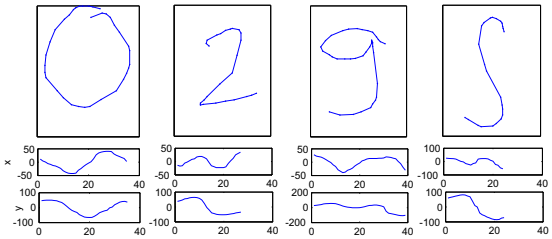
(d) Forgery Signature Finger



(e) Name Lower-case



(f) Name Upper-case



(g) Number Sequence

Figure 3.3: *Example of the data collected in e-BioSign database for the Samsung Galaxy Note 10.l.*

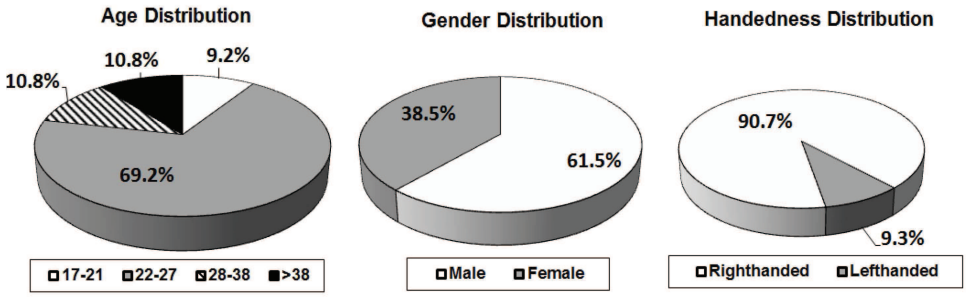


Figure 3.4: *Population statistics of e-BioSign database.*

pressure information as this was not provided by these devices, and there is also no information of the trajectory (X and Y coordinates) during pen-ups. For the case of signatures acquired using the stylus, pressure information and pen-up trajectories are available for all devices.

Fig. 3.4 shows the population statistics of the e-BioSign database. Regarding the age distribution, the majority of the subjects (69.2%) are between 22 and 27 years old, as the database was collected in a university environment. Fig. 3.4 also shows the handedness and the gender distributions. The gender was designed to be as balanced as possible, having 61.5% of males and 38.5% of females. Regarding the handedness distribution, 90.7% of the population is righthanded.

3.2.2. ATVS On-Line Signature Long-Term Extended Database

The ATVS On-Line Signature Long-Term Extended database is an extension of the database published in [Galbally *et al.*, 2013]. Fig. 3.5 shows the number of genuine signatures per user and the general time diagram of the different acquisition sessions of it. This database was used in [Galbally *et al.*, 2013] taking into account only random forgeries. However, skilled forgeries have been included in this extended version of the database, which is already publicly available at `https://github.com/BiDAlab/xLongSignDB`. This database comprises a total of 29 users. The inter-session variability problem is also considered in this database as signatures were acquired in 6 different sessions (S1 to S6 in Fig. 3.5) within a 15-month time span. Sessions from S1 to S4 are composed of 4 genuine signatures per user each and have a two-month interval between them in a first acquisition campaign (i.e., BioscurID Signature Subset [Fierrez *et al.*, 2010]). The acquisition of S5 and S6 sessions was performed in a different campaign (i.e., Biosecure Signature Subset [Ortega-Garcia *et al.*, 2010]) that started 6 months after the first campaign had finished. It comprises 30 genuine signatures per user distributed in two acquisition sessions separated three months. Therefore, the total number of genuine signatures and skilled forgeries per user are 46 and 10, respectively. To perform skilled forgeries, the users had visual access to the dynamics of the signing process of the signatures they had to forge as many times as they wanted. This database allows the research community to perform evaluations over current emerging scenarios.

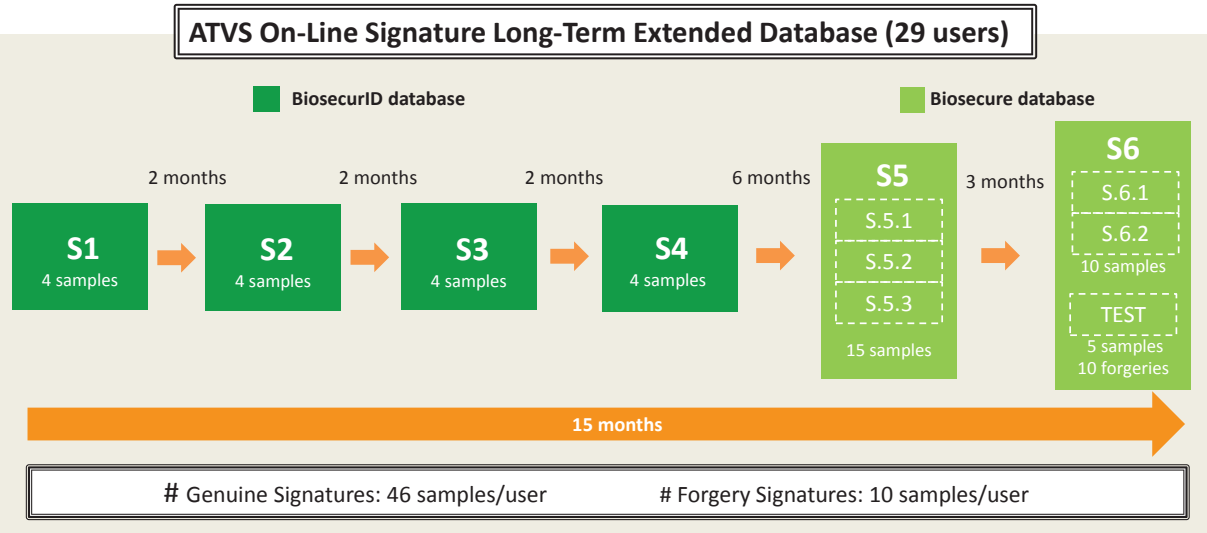


Figure 3.5: *General time diagram of the different acquisition sessions and number of genuine signatures per user that form the ATVS On-Line Signature Long-Term Extended Database.*

3.3. Handwriting Touchscreen Databases

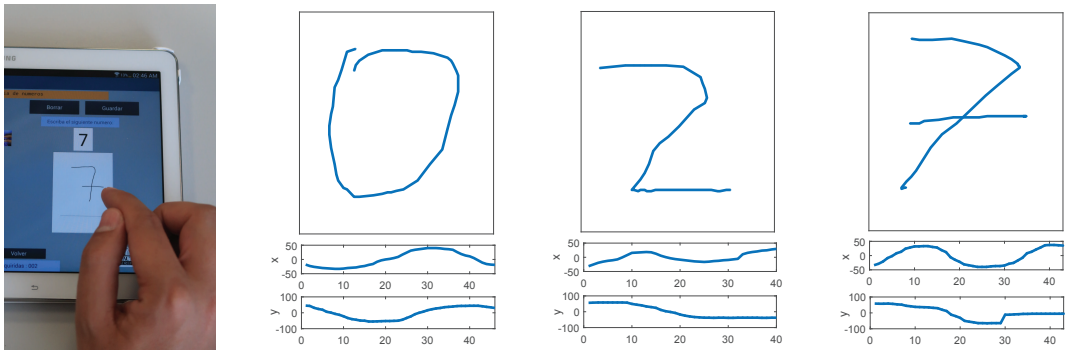
3.3.1. e-BioDigit Database

The e-BioDigit database has been captured during this Thesis in order to evaluate the advantages and potential of incorporating handwriting biometric information to traditional password-based mobile authentication systems. This database comprises on-line handwritten numerical digits from 0 to 9 acquired using a Samsung Galaxy Note 10.1 general purpose tablet. This device has a 10.1-inch LCD display with a resolution of 1280×800 pixels.

Regarding the acquisition protocol, subjects had to perform handwritten numerical digits from 0 to 9, one at a time. The acquisition setup and some examples of the handwritten numerical digits of the e-BioDigit database are depicted in Fig. 3.6. Additionally, samples were collected in two sessions with a time gap of at least three weeks between them in order to consider inter-session variability, very important for behavioural biometric traits [Galbally *et al.*, 2013]. For each session, users had to perform a total of 4 numerical sequences from 0 to 9 using the finger as input. Therefore, there are a total of 8 samples per numerical digit and user. Information related to X and Y spatial coordinates and timestamp is captured and stored in the database.

The software for capturing handwritten numerical digits was developed in order to minimise the variability of the user during the acquisition process. A rectangular area with a writing surface size similar to a 5-inch screen smartphone was considered, see Fig. 3.6(a). A horizontal line was represented on top of the drawing rectangular area, including two buttons "OK" and "Cancel" to press after writing if the sample was good or bad respectively.

The database comprises a total of 93 users. Regarding the age distribution, the majority of the subjects (85.0%) are between 17 and 27 years old, as the database was collected in a



(a) (b) (c) (d)

Figure 3.6: *(a) Acquisition setup. (b-d) examples of different handwritten numerical digits of the e-BioDigit database. X and Y denote horizontal and vertical position versus the time samples.*

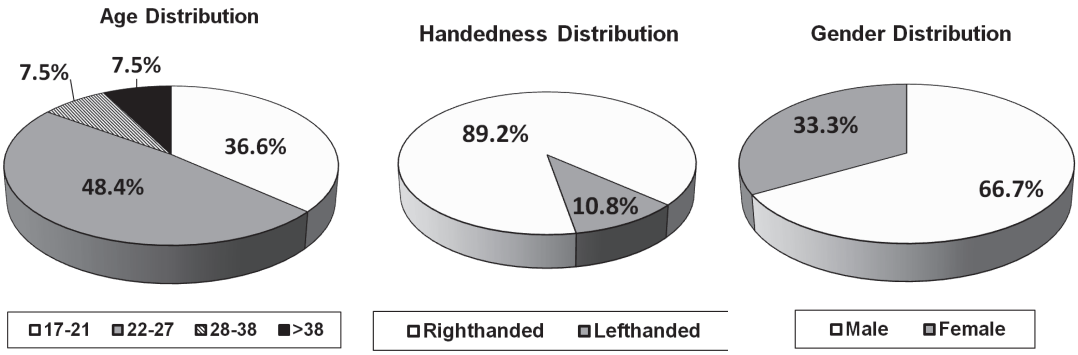


Figure 3.7: *Population statistics for the e-BioDigit database.*

university environment (36.6% between 17 and 21). Regarding the gender, 66.7% of the subjects were males and 33.3% females whereas for the handedness distribution, 89.2% of the population was righthanded. The e-BioDigit statistics are depicted in Fig. 3.7.

3.4. Chapter Summary and Conclusions

This chapter gives an overview of the most well-known on-line signature databases. We have first described the most relevant features of existing on-line signature databases, making special emphasis on the databases used in the experimental work of this Thesis. Then, we have described the new e-BioSign and ATVS On-Line Signature Long-Term Extended databases, which have been captured during the execution of this Thesis and constitute some of the main contributions. Finally, the new e-BioDigit database, which has been also captured during this Thesis for the evaluation and improvement of traditional password-based systems, has been described in detail. We would like to highlight that all these new databases are already publicly available to the research community.

Chapter 4

Proposed Methods

The present Chapter aims to describe all the details of the on-line signature verification systems considered in this Dissertation.

The chapter is organised as follows. Sec. 4.1 is focused on traditional signature verification systems. It describes the specific features and matching algorithm configurations considered in this Thesis. Sec. 4.2 concentrates on novel signature verification systems based on deep learning architectures. It first explains the basics of RNN systems and gives an overview of the main relevant studies. Then, the specific details of our proposed end-to-end writer-independent RNN signature verification systems are described. Conclusions are finally drawn in Sec. 4.3.

This chapter is based on the following publications: [Tolosana *et al.*, 2015a, 2017b, 2018c,f, 2015d,e].

4.1. Traditional Signature Verification Systems

4.1.1. Global System

The global system (a.k.a. feature-based system) considered in this Dissertation is mainly based on previous studies carried out in our research group. Concretely, we extract for each signature a set of 117 global features from the normalised signals X and Y spatial coordinates and pressure $[x_n, y_n, p_n]$. From those global features, the first 100 were already proposed in [Fierrez-Aguilar *et al.*, 2005b] as an extension of other sets presented in [Lee *et al.*, 1996; Nelson and Kishon, 1991; Nelson *et al.*, 1994]. The remaining 17 global features have been proposed in this Thesis and are mainly based on discriminative information related to the pressure. Tables 4.1 and 4.2 describe the set of 100 and 17 global features considered in this Dissertation.

The entire set of 117 global features can be divided into five different categories corresponding to the following magnitudes:

- **Time (25 features):** related to signature duration, or timing of events such as pen-ups or local maxima. Feature numbers are: 1-19, 21-26.

Table 4.1: *Set of 100 global features originally proposed in [Fierrez-Aguilar* et al.*, 2005b]. Table adapted from [Martinez-Diaz, 2015]. T denotes time interval, t denotes time instant, N denotes number of events, and θ denotes angle. All notations are defined or referenced in the table.*

#	Time related feature	#	Direction related feature
#	Kinematic related feature	#	Geometry related feature
#	Pressure 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(1st pen-up)/T_w
13	T(2nd pen-up)/T_w	14	T(2nd pen-down)/T_s
15	T(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 $v_{x,y}$)/$v^2_{\max}$	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(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: *Set of 17 novel global features proposed in this Thesis. z denotes pressure.*

#	Feature Description	#	Feature Description
101	average pressure $\overline{z}$	102	median pressure
103	N(Pen Downs samples)	104	N(Pen Ups samples)
105	median N(Pen Ups samples) individually	106	average N(Pen Ups samples) individually
107	median N(Pen Downs samples) individually	108	average N(Pen Downs samples) individually
109	$\overline{z}$ / p_{max}	110	($\overline{z}$ - z_{min}) / $\overline{z}$
111	median pressure last pen-down	112	average pressure last pen-down
113	median pressure first pen-down	114	average pressure first pen-down
115	(z_{max} - z_{min}) / $\overline{z}$	116	average velocity $\overline{v}$
117	average acceleration $\overline{a}$		

- **Kinematic (27 features):** extracted from the first and second time order derivatives of the position time functions, like average speed or maximum speed. In this category the two 116 and 117 new features have been added to the existing ones. Feature numbers are: 27-51, 116-117.
- **Direction (18 features):** extracted from the path trajectory like the starting direction or mean direction between pen-ups. Feature numbers are: 55-72.
- **Geometry (32 features):** associated to the strokes or signature aspect-ratio. Feature numbers are: 20, 52-54, 73-100.
- **Pressure (15 features):** associated to pressure information like the mean pressure or number of pen-down samples. This is a new category proposed in this Dissertation. Feature numbers are: 101-115.

For the similarity computation, we always consider the Mahalanobis distance as it has provided very good results in previous studies [Fierrez-Aguilar, 2006; Martinez-Diaz, 2015]. The Mahalanobis distance [Theodoridis and Koutroumbas, 2008] is used to compare the similarity between a query signature and a claimed user model. A user model is created from a training set of genuine signatures. This model is defined as $C = (\mu,\Sigma)$, where μ is a feature vector with the mean of feature vectors extracted from each signature of this user and Σ is a diagonal covariance matrix. The matching score is obtained as the inverse of the Mahalanobis distance between the input signature feature vector x and the claimed user model C:

$$s(x, C) = \left((x - \mu)^T (\Sigma)^{-1} (x - \mu)\right)^{-1/2} \tag{4.1}$$

If the score $s(x, C)$ is above a specific threshold, the signature is considered genuine. Otherwise, it is rejected by the system.

4.1.2. Local Systems

The local system (a.k.a. time functions-based system) considered in this Dissertation is mainly based on previous studies carried out in our research group. Concretely, we extract for each signature a set of 23 local features from the normalised signals X and Y spatial coordinates

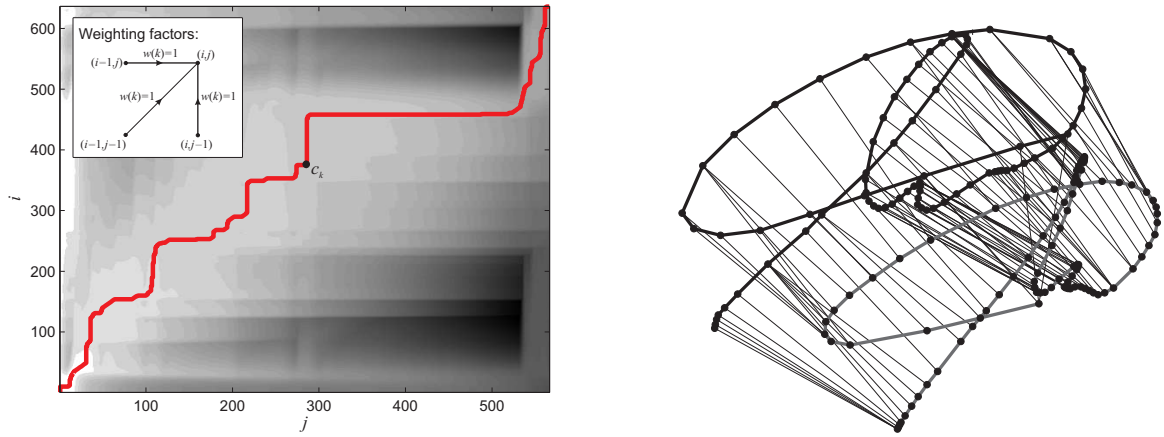
Table 4.3: *Set of 23 local features considered in this Thesis. Local Features 3 and 10 (highlighted in yellow colour) are not available when using the finger as input of the signature verification system.*

#	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	$\dot{x}_n$, $\dot{y}_n$, $\dot{z}_n$, $\dot{\theta}_n$, $\dot{\upsilon}_n$, $\dot{\rho}_n$, $\dot{a}_n$
15-16	Second-order derivative of features 1-2	$\ddot{x}_n, \ddot{y}_n$
17	Ratio of the minimum over the maximum speed over a window of 5 samples	$\upsilon_n^r = \min\{\upsilon_{n-4}, ..., \upsilon_n\}/\max\{\upsilon_{n-4}, ..., \upsilon_n\}$
18-19	Angle of consecutive samples and first order difference	$\alpha_n = \arctan(y_n - y_{n-1}/x_n - x_{n-1})$ $\dot{\alpha}_n$
20	Sine	$s_n = \sin(\alpha_n)$
21	Cosine	$c_n = \cos(\alpha_n)$
22	Stroke length to width ratio over a window of 5 samples	$r_n^5 = \frac{\sum_{k=n-4}^{k=n}\sqrt{(x_k-x_{k-1})^2+(y_k-y_{k-1})^2}}{\max\{x_{n-4},...,x_n\}-\min\{x_{n-4},...,x_n\}}$
23	Stroke length to width ratio over a window of 7 samples	$r_n^7 = \frac{\sum_{k=n-6}^{k=n}\sqrt{(x_k-x_{k-1})^2+(y_k-y_{k-1})^2}}{\max\{x_{n-6},...,x_n\}-\min\{x_{n-6},...,x_n\}}$

and pressure $[x_n, y_n, p_n]$. These local features were proposed in [Fierrez *et al.*, 2007; Lei and Govindaraju, 2005; Martinez-Diaz, 2015; Richiardi *et al.*, 2005]. Other information considered in previous studies such as the altitude and azimuth (local features related to the pen orientation) have been discarded in this Dissertation as they are not acquired using general purpose devices. Table 4.3 describes the set of 23 local features considered in this Dissertation. It is important to remark that local features 3 and 10 (highlighted in Table 4.3 in yellow colour) are not available when using the finger as input of the signature verification system.

For the similarity computation, we have considered three different state-of-the-art signature verification systems. The specific configuration details are described below:

- For the **DTW system**, we consider the same basis described in Sec. 2.1.3.1 with the implementation details proposed in [Fierrez-Aguilar, 2006; Martinez-Diaz, 2015]. For the



(a) (b)

Figure 4.1: *(a) Optimal warping path (red colour) between two sequences obtained with DTW. 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. Images extracted from [Martinez-Diaz, 2015].*

computation of the distance measure between sequence samples (i.e., $d(i,j)$), we use Euclidean distance. For the definition of the weighting factors (i.e., w_k), only three transitions with the same value equal to 1 are allowed for the computation of g_k. Consequently, Eq. (2.5) becomes

$$g_k = g(i,j) = \min \begin{bmatrix} g(i,j-1) + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i-1,j) + d(i,j) \end{bmatrix} \tag{4.2}$$

The accumulated distance between the two sequences is computed as

$$D = g(I,J)/K \tag{4.3}$$

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

Fig. 4.1.(a) represents our definition of w_k together with an example of a warping path between two time sequences. In Fig. 4.1.(b), an example of point correspondences between two signatures is depicted to visually show the results of the elastic alignment.

- For the **HMM and GMM systems**, we consider the same basis described in Sec. 2.1.3.2 with the implementation details proposed in preliminary studies carried out in our research group [Fierrez-Aguilar, 2006; Martinez-Diaz, 2015]. They are based on a left-to-right configuration without skipping state transitions (see Fig. 4.2). Systems are trained in two different steps. First, state transition probabilities and observation statistical models

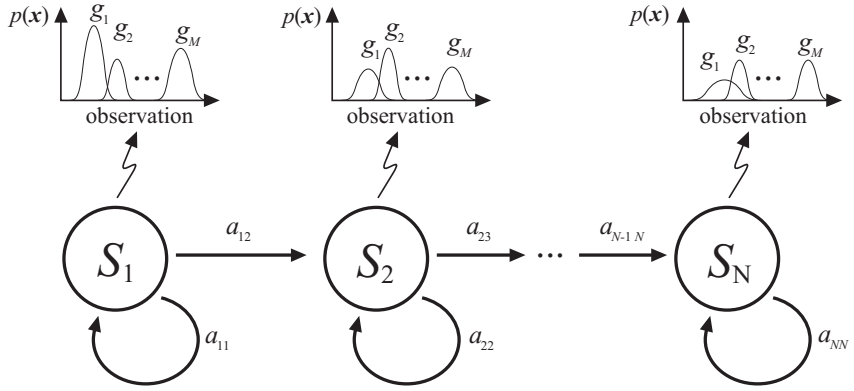


Figure 4.2: *Graphical representation of a left-to-right N-state HMM, with M Gaussian Mixtures per state. Image extracted from [Martinez-Diaz, 2015].*

are estimated using a Maximum Likelihood algorithm. After this, a re-estimation step is carried out using the Baum-Welch algorithm [Rabiner, 1989].

For the similarity computation module, the final score is computed as the log-likelihood of the target signature (using the Viterbi algorithm) divided by the total number of samples of the signature signal. In order to keep scores between a reasonable range, normalised scores s_n between (0,1) are obtained as $s_n = exp(s(x, C)/30)$, where $s(x, C)$ is the score returned by the HMM algorithm and x and C represent respectively the input signature to verify and the enroled model of the claimed identity.

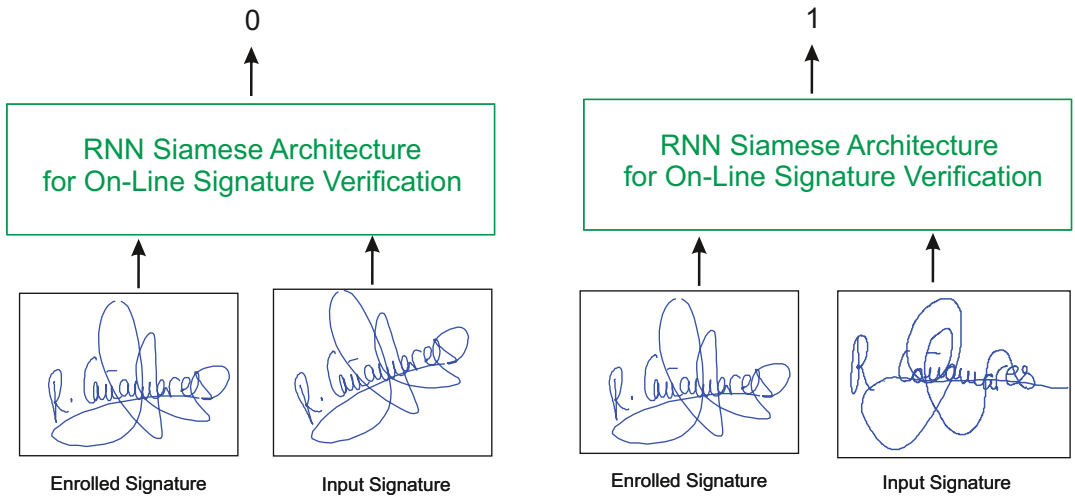
4.2. Deep Learning Signature Verification Systems

This section describes our proposed novel writer-independent on-line signature verification system based on Recurrent Neural Networks with a Siamese architecture whose goal is to learn a dissimilarity metric from pairs of signatures. To the best of our knowledge, this is the first time these recurrent Siamese networks are applied to the field of on-line signature verification, which provides our main motivation. We consider both LSTM and GRU architectures. Additionally, a bidirectional scheme, which is able to access both past and future context, is considered for both LSTM- and GRU-based systems.

4.2.1. Exploring RNN DL Architectures

4.2.1.1. Siamese Architecture

The Siamese architecture has been used for recognition and verification applications where the number of categories is very large and not known during training, and where the number of training samples for a single category is small. In our case, the main goal of this architecture is to learn a dissimilarity metric from data minimising a discriminative cost function that drives the dissimilarity metric to be small for pairs of genuine signatures from the same user, and longer for pairs of signatures coming from different users. Fig. 4.3 shows this idea visually. In previous



(a) Genuine case (b) Impostor case

Figure 4.3: *Examples of our proposed LSTM and GRU RNN systems based on a Siamese architecture for minimising a discriminative cost function.*

studies such as [Chopra *et al.*, 2005], the authors proposed the use of CNNs with a Siamese architecture for face verification. Experiments were performed with several databases obtaining very good results where the number of training samples for a single category was very small. Siamese architectures have been also studied in early works for on-line signature verification [Bromley *et al.*, 1993] although not considering RNNs. In [Bromley *et al.*, 1993], the authors proposed an on-line signature verification system composed of two separated sub-networks based on Time Delay Neural Networks (TDNNs) that are one-dimensional convolutional networks applied to time series. Different architectures regarding the number and size of layers were studied. A total of 8 time functions fixed to the same 200 points length were extracted for X and Y pen coordinates using an old-fashion NCR 5990 Signature Capture Device. The best performance was obtained using two convolutional layers with 12 by 64 units in the first layer and 16 by 19 units in the second one. The threshold was set to detect 80.0% of forgeries and 95.5% of genuine signatures, far away from the results that can be achieved nowadays with state-of-the-art systems [Diaz *et al.*, 2018b, 2016b; Gomez-Barrero *et al.*, 2015; Liu *et al.*, 2014; Martinez-Diaz *et al.*, 2014].

4.2.1.2. Long Short-Term Memory

LSTM RNNs [Hochreiter and Schmidhuber, 1997] have been successfully applied to many different tasks such as language identification considering short utterances [Zazo *et al.*, 2016] or biomedical problems [Petrosian *et al.*, 2000], among many others. However, the analysis and design of LSTM architectures for new tasks are not straightforward [Pascanu *et al.*, 2014].

LSTM RNNs comprise memory blocks usually containing one memory cell each of them with

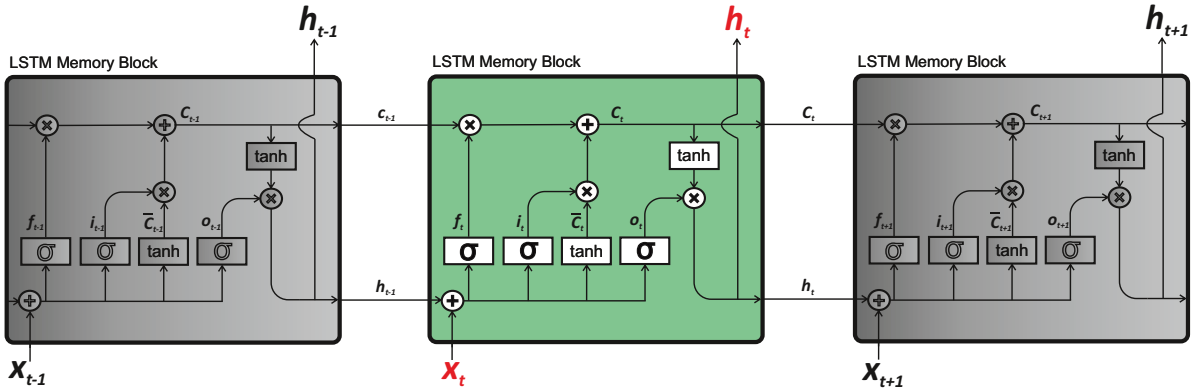


Figure 4.4: *Scheme of a single LSTM memory block at different time steps (i.e., X_{t-1}, X_t and X_{t+1}).*

a forget gate f, an input gate i, and an output gate o. For a time step t:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{4.4}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4.5}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4.6}$$

$$\widetilde{C_t} = \tanh(W_C x_t + U_C h_{t-1} + b_C) \tag{4.7}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{4.8}$$

$$h_t = o_t \odot \tanh(C_t) \tag{4.9}$$

where W_* and U_* are weight matrices and b_* is the bias vector. The symbol $\odot$ represents a pointwise product whereas σ is a sigmoid activation that outputs values between 0 and 1. The LSTM does have the ability to remove old information from $t-1$ time or add new one from t time. The key is the cell state C_t that is carefully regulated by the gates. The f gate decides the amount of previous information (i.e., h_{t-1}) that passes to the new state of the cell C_t. The i gate indicates the amount of new information (i.e., $\widetilde{C_t}$) to update in the cell state C_t. Finally, the output of the memory block h_t is a filtered version of the cell state C_t, being the o gate in charge of it. Fig. 4.4 shows a single LSTM memory block at different time steps (i.e., X_{t-1}, X_t and X_{t+1}) for clarification.

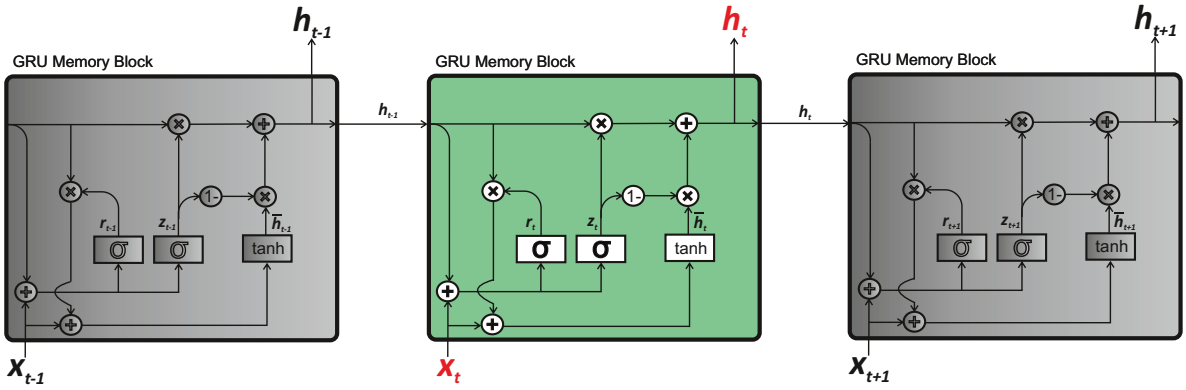


Figure 4.5: *Scheme of a single GRU memory block at different time steps (i.e., X_{t-1}, X_t and X_{t+1}).*

4.2.1.3. Gated Recurrent Unit

GRU [Cho *et al.*, 2014a,b] is a relatively new type of RNN that has been inspired by the LSTM unit but is much simpler to compute and implement. In addition, the results obtained using this novel RNN system seems to be very similar to the LSTM RNN system [Jozefowicz *et al.*, 2015]. The main difference between GRU and LSTM RNNs resides in the number of gates used to control the flow of information. Whereas the LSTM unit contains three different gates (i.e., forget f, input i and output o), the GRU unit only owns two gates (i.e., reset r and update z). For a time step t:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{4.10}$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{4.11}$$

$$\widetilde{h_t} = \tanh(W_h x_t + U_h(h_{t-1} \odot r_t) + b_h) \tag{4.12}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h_t} \tag{4.13}$$

where W_* and U_* are the weight matrices and b_* is the bias vector. The symbol $\odot$ represents a pointwise product whereas σ is a sigmoid activation that outputs values between 0 and 1. The GRU does have the ability to remove old information from $t-1$ time or add new one from t time. The reset gate r_t is in charge of keeping in the current cell state (i.e., $\widetilde{h_t}$) the information of the previous time step (i.e., h_{t-1}) or reset it with the information of only the current input (i.e., x_t). Finally, the update gate z_t filters how much information from the previous time step and current cell state will flow to the current output of the memory block (i.e., h_t). Fig. 4.5 shows a single GRU memory block at different time steps (i.e., X_{t-1}, X_t and X_{t+1}) for clarification.

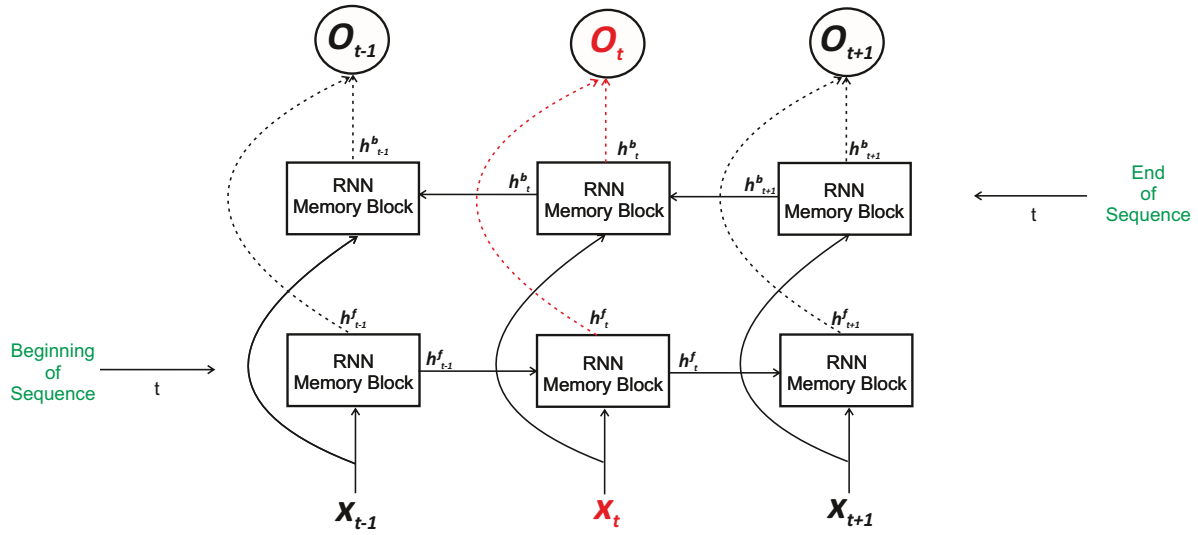


Figure 4.6: *Scheme of a typical Bidirectional RNN system at different time steps (i.e., X_{t-1}, X_t and X_{t+1}). The bottom part of the scheme propagates the information forward in time (towards the right) while the top part of the scheme propagates the information backward in time (towards the left). Thus at each point t, the output units O_t can benefit from a relevant summary of the past in its h_t^f input and from a relevant summary of the future in its h_t^b input. Figure adapted from [Goodfellow* et al.*, 2016].*

4.2.1.4. Bidirectional RNNs

The RNN schemes explained before in Sec. 4.2.1.2 and 4.2.1.3 are the original ones. These schemes have access only to the past and present contexts. However, for some applications such as handwriting or speech recognition the chance of having access to the future context can further improve the system performance [Graves and Jaitly, 2014; Graves *et al.*, 2009]. Schemes that also allow access to the future context are known as Bidirectional RNNs (BRNNs) [Schuster and Paliwal, 1997]. BRNNs combine a RNN that moves forward through time beginning from the start of the sequence with another RNN that moves backward through time beginning from the end of the sequence [Goodfellow *et al.*, 2016]. Fig. 4.6 shows a typical scheme of a BRNN system at different time steps (i.e., X_{t-1}, X_t and X_{t+1}) for clarification. The bottom part of the scheme propagates the information forward in time (towards the right) while the top part of the scheme propagates the information backward in time (towards the left). Thus at each point t, the output units O_t can benefit from a relevant summary of the past in its h_t^f input and from a relevant summary of the future in its h_t^b input [Goodfellow *et al.*, 2016].

4.2.2. Proposed RNN On-Line Signature Verification Systems

Our proposed end-to-end writer-independent on-line signature verification system is depicted in Fig. 4.7. This system has been obtained after carrying out an exhaustive analysis in terms of the number of time functions used to feed the network and the complexity level of the RNN system (i.e., the number of hidden layers and memory blocks per hidden layer). All details are described in the experimental part of Chapter 7. The present section aims to summarise the

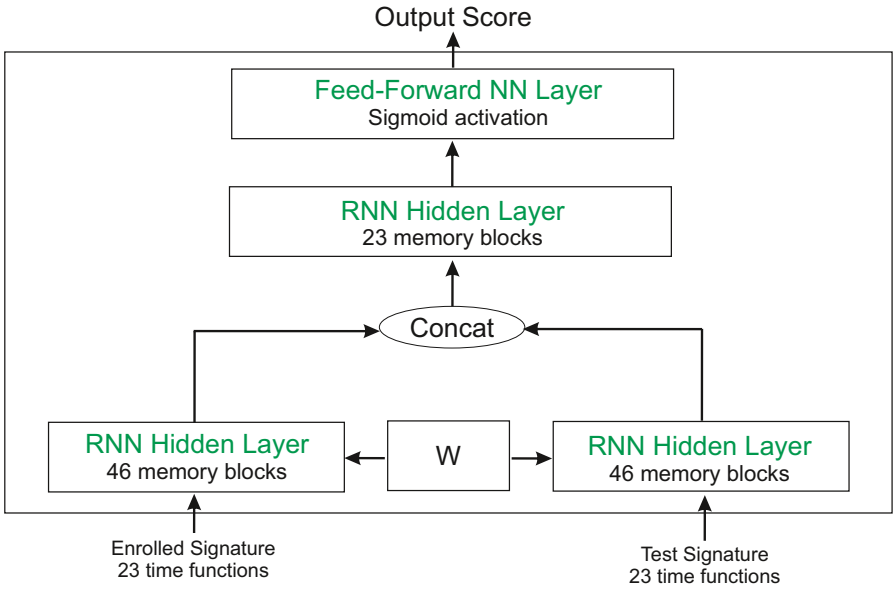


Figure 4.7: *End-to-end writer-independent on-line signature verification system proposed in this Thesis based on the use of LSTM and GRU RNNs with a Siamese architecture.*

final configuration selected.

We propose both LSTM and GRU systems with a Siamese architecture. In addition, a bidirectional scheme is considered for both LSTM- and GRU-based systems in order to be able to access both past and future context. For the input of the RNN system, we extract a set of 23 local features (Table 4.3) per signature from signals related to X and Y spatial coordinates and pressure. The first layer is composed of two LSTM/GRU hidden layers with 46 memory blocks each, sharing the weights between them. The outputs of the first two parallel LSTM/GRU hidden layers are concatenated and serve as input to the second layer, which corresponds to a LSTM/GRU hidden layer with 23 memory blocks. Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score between 0 and 1 for each pair of signatures.

4.3. Chapter Summary and Conclusions

In this chapter we have concentrated on describing the on-line signature verification systems considered in this Dissertation. First, Sec. 4.1 has focused on traditional signature verification systems, explaining the specific features and matching algorithm configurations considered in this Thesis. Then, Sec. 4.2 has described our proposed novel signature verification systems based on deep learning. We have first explained the basics of RNN systems and gave and overview of the main relevant studies. Finally, the specific details of our proposed end-to-end writer-independent RNN signature verification systems have been stated. This architecture has been also adapted to the task of handwritten passwords for touchscreen biometrics described in Chapter 9.

Part II

Emerging Scenarios

Chapter 5

Multi-Device Multi-Input Acquisition Scenarios

In this chapter, we analyse and adapt traditional on-line signature verification systems to emerging scenarios focusing on device interoperability, finger input, and mixed writing-input. Both Biosecure and e-BioSign databases are considered in the experimental work carried out in this chapter in order to evaluate the system performance using traditional and new COTS devices.

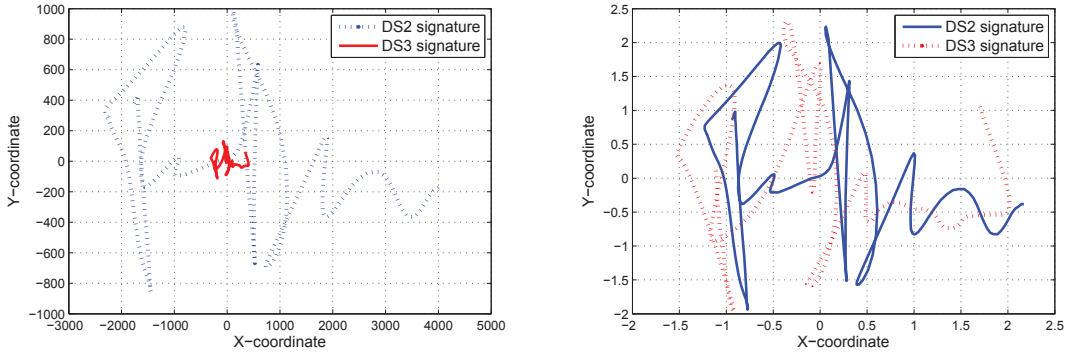
The chapter is organised as follows. Sec. 5.1 explains our two-stage approach proposed in order to alleviate the degradation of the system performance on these novel scenarios. Sec. 5.2 describes the experimental protocol considered for both Biosecure and e-BioSign databases. Then, in Sec. 5.3 we analyse finger input scenarios for on-line signature verification and compare the results obtained with the traditional stylus scenario. In Sec. 5.4 we first evaluate the system performance of traditional signature verification systems on device interoperability scenarios, and then we apply the two-stage approach proposed in the Thesis. Sec. 5.5 finally evaluates the possibility of using different writing tools during the acquisition of enrolment and test signatures (i.e., mixed writing-input scenario). Conclusions are finally drawn in Sec. 5.6.

This chapter is based on the following publications: [Tolosana *et al.*, 2017a, 2018e, 2015c,d, 2017d; Vera-Rodriguez *et al.*, 2015].

5.1. Proposed Approach

This section describes the two main stages proposed in this Thesis in order to increase the robustness of on-line signature verification systems on these novel scenarios. First, a data preprocessing stage is applied in order to achieve a high similarity between signatures coming from different devices and writing tools. Second, a new criterion is proposed in order to select the optimal features for each specific scenario.

In this chapter we use both global and local systems. For the global system, we consider



(a) Original Signatures

(b) Preprocessed Signatures

Figure 5.1: *Signatures from DS2 and DS3 datasets before and after applying the mean and standard deviation normalisation technique.*

the same system described in Sec. 4.1.1 based on 117 initial global features and Mahalanobis distance algorithm for the similarity computation. For the local system, we consider the same system described in Sec. 4.1.2 based on 23 initial local features and DTW algorithm for the similarity computation.

5.1.1. Data Preprocessing Stage

The first stage we propose to compensate multi-device and multi-input acquisition scenarios is related to data preprocessing. The aim of this first stage is to obtain signatures with the same type of information (i.e., X and Y spatial coordinates, pressure, etc.) and time and spatial position standard formats so as to improve the performance of the system in these novel scenarios. Several statistical data normalisation techniques have been studied in order to correct variabilities between devices. A graphical example of this effect can be seen in Fig. 5.1 for the devices considered in Biosecure DS2 and DS3 datasets. The different spatial position of DS2 signatures is due to the acquisition protocol followed in Biosecure where users had to sign in different boxes on a sheet of paper (see Fig. 3.1(a)) whereas the different size among signatures from DS2 and DS3 could be due to the screen resolution of the devices (see Fig. 5.1(a)).

In order to compensate the position and spatial variabilities described before, we apply normalisation techniques based on the mean and standard deviation. Fig. 5.1(b) represents DS2 and DS3 signatures after applying the proposed normalisation. In addition, we also consider interpolation techniques based on splines [Martinez-Diaz *et al.*, 2007] in order to correct sampling errors (missing samples) and get the same sampling frequency in all acquisition devices (fixed to 200Hz). Finally, for the e-BioSign database, we remove the first and last samples of the signatures as they correspond to the time between the operator clicks to start/finish the acquisition and the time the user starts/finishes signing.

5.1.2. Feature Extraction and Selection Stage

The second stage of our proposed approach is focused on the selection of the optimal global and local features for each specific scenario. The SFFS algorithm described in Sec. 2.1.5 is considered here in order to obtain an optimal subset of the initial 117 and 23 global and local features that improves the system performance in terms of EER (%).

In our proposed approach, in order to increase the robustness of the systems on these novel scenarios, the criterion of the SFFS algorithm has been adapted to each specific scenario considering the EER obtained from different signature comparisons at the same time. For example, in order to tackle device interoperability scenarios, the EER of intra- and inter-device signature comparisons are considered at the same time in our proposed SFFS criterion. All details are included inside of each scenario.

5.2. Experimental Protocol

5.2.1. Biosecure Database

The first 50 users of the database are considered for the development and training of the systems whereas the remaining 70 users are used for the final evaluation of the proposed systems.

For each user, the first 5 genuine signatures of the first session are used for training, whereas the 15 genuine signatures of the second session are left for testing in order to take into account the inter-session variability. Therefore, the 10 remaining genuine signatures of the first session are not used in our experiments. Skilled forgery scores are obtained by comparing training signatures against the 20 available skilled forgeries for the same user whereas random (zero-effort) forgery scores are obtained by comparing the training signatures with one genuine signature of the remaining users. For the global system, scores are obtained by comparing signatures against the user model, while for the local system, the average score of the five one-to-one comparisons is performed.

The nomenclature followed in the Thesis for the analysis of both intra- and inter-device scenarios is denoted as:

$$a \text{ — } b \text{ — } c$$

where a indicates skilled or random forgery cases, and b and c represent the device used for training and testing respectively.

5.2.2. e-BioSign Database

The experimental protocol proposed for this database has been designed to cover all emerging scenarios, i.e., device interoperability, finger input, and mixed writing-input. The e-BioSign database is divided into two different datasets. The first 30 users of the database are considered for the development and training of the systems, whereas the remaining 35 users are considered for the final evaluation of the proposed systems.

Table 5.1: *Local features considered in the local baseline system. Local feature # taken from Table 4.3.*

#	**Feature description**
1	x-coordinate: x_n
2	y-coordinate: y_n
8-9	First-order derivate of features 1-2: $\dot{x_n}, \dot{y_n}$
15-16	Second-order derivate of features 1-2: $\ddot{x_n}, \ddot{y_n}$
19	First order difference of angle of consecutive samples: $\dot{\alpha_n}$

Regarding the development and training phase, different experimental protocols have been proposed for each of the three emerging scenarios considered in this Thesis. All details about the training procedures are given inside each scenario.

Regarding the evaluation phase, the same experimental protocol is considered for all experiments. The 4 genuine signatures of the first session are used for training, whereas the remaining 4 genuine signatures of the second session are left for testing. Skilled forgery scores are obtained by comparing the training signatures against the 6 available skilled forgeries per user whereas random (zero-effort) forgery scores are obtained by comparing the training signatures with one genuine signature of each of the remaining users. For the global system, scores are obtained by comparing test signatures against the user model obtained with the 4 training signatures whereas for the local system, the average score of the four one-to-one comparisons is performed.

5.3. Finger Input Scenarios

In this section we assess the feasibility of signature verification systems based on finger writing input. The local system described in Sec. 5.1 and based on DTW is considered in the study. The W4 and W5 devices of the new e-BioSign database have been considered in this experimental work as signatures were acquired using both stylus and finger for the same group of users. This way we can perform a clear evaluation of this novel scenario compared to the traditional one based on the stylus.

Two different approaches are considered in this experiment. First, baseline systems whose local features are fixed from previous works (see Table 5.1). Second, the proposed systems are adjusted using our proposed two-stage approach described in Sec. 5.1 over the development dataset in order to improve the system performance. The SFFS algorithm has been individually applied to each device and writing tool in order to select the optimal feature subsets for each specific case. The first approach is considered as Baseline (B) whereas the second one is considered as Proposed (P). Table 5.2 shows results for both Baseline and Proposed approaches considering the 35 users of the evaluation dataset.

Analysing Table 5.2 for the case of using the stylus as the writing tool, the baseline systems achieve an average EER of 11.5% and 1.5% for skilled and random forgeries respectively whereas for the proposed systems the average EER improves to 9.3% and 0.9% for skilled and random forgeries respectively. These results show the benefits of using our two-stage approach over the development dataset. A thorough analysis of the compensation effects of our proposed two-stage

Table 5.2: *System performance results in terms of EER (%) on the evaluation dataset. B = Baseline and P = Proposed.*

	STYLUS				FINGER			
	W4		W5		W4		W5	
	B	P	B	P	B	P	B	P
Skilled	10.0	7.9	12.9	10.7	24.0	22.1	27.0	26.4
Random	0.8	0.7	2.1	1.0	1.4	0.3	2.3	1.0

approach will be then carried out over the device interoperability scenario. From the results, it is also important to highlight the good system performance obtained for both W4 and W5 devices, showing the possibility of considering general purpose devices in real banking and commercial applications and not only the traditional Wacom devices.

Analysing Table 5.2 for the case of using the finger as the writing input, the baseline systems achieve an average EER of 25.5% and 1.9% for skilled and random forgeries respectively whereas for the proposed systems the average EER improves to 24.3% and 0.7% for skilled and random forgeries respectively, showing again the benefits of using our proposed two-stage approach. An important effect that can be observed from Table 5.2 is how the system performance changes regarding the writing input (i.e., stylus and finger) considered during the acquisition process, especially for skilled forgeries with results almost three times worse than the stylus case.

In order to find out the reasons for such difference in the system performance, an exhaustive analysis of the finger scenario has been carried out. Two main aspects justify the results obtained. First, we have observed that in general users who perform their signatures using closed letters (i.e., a, e, o, l, p, q, etc.) tend to perform much larger writing executions in comparison with other letters due to the lower precision they are able to achieve using the finger. Besides, users whose signatures are composed of a long name and surname (or two surnames) tend to simplify some parts of their signatures on the finger scenario. Regarding the sampling frequency of the acquisition process, it is important to highlight the differences that exist between the stylus and finger scenarios. For the stylus scenario, all samples of the signature are uniformly distributed across the whole signing process. However, for the case of using the finger as input, there are many samples distributed in small parts of the signature instead of the whole signature as it happens in the stylus scenario. This non-desirable effect is due to the lack of precision obtained using the finger and the friction produced between the screen and the finger. Therefore, it might not be related to the specific device considered in the experimental work, but to the use of the finger as the writing input instead. Some differences that exist between both stylus and finger scenarios are depicted in Fig. 5.2. Despite this effect, and although the number of samples are very similar in both scenarios, an additional interpolation step based on splines is required in order to reduce the difference in the sampling effect between the stylus and finger scenarios.

Finally, it is important to remark the very challenging scenario considered in this experiment as forgers had access to the dynamic realization of the signatures to forge. A recommendation for the usage of signature recognition on mobile devices would be for the users to protect themselves

(a) Stylus (b) Finger

Figure 5.2: *Signatures from the e-BioSign database acquired using both stylus and finger.*

from other people that could be watching while signing, as this is more feasible to do in the mobile scenario compared to the office scenario. This way skilled forgers might have access to the global shape of the signature but not to the dynamic information. Therefore, the higher variability together with the challenging forgeries considered in this experiment conclude that these new finger input scenario can be applied to those scenarios where the knowledge of the impostor about the user to forge is scarce, e.g., random forgeries.

5.4. Device Interoperability Scenarios

5.4.1. Biosecure Database

We first assess device interoperability scenarios using the traditional devices considered in Biosecure database. In Sec. 5.4.1.1, we evaluate the data preprocessing first stage of our proposed approach for the standard case of having a recognition system adjusted specifically to each device, without considering device interoperability conditions. Signature verification systems are optimised for the skilled forgery case as it is the most challenging case in on-line signature verification [Diaz *et al.*, 2018b; Martinez-Diaz *et al.*, 2014; Plamondon and Srihari, 2000]. Due to the importance of the data preprocessing stage, this first stage is always considered in the rest of experiments in order to evaluate the second stage based on the selection of the optimal global and local features for this device interoperability scenario. In Sec. 5.4.1.2, the systems obtained in Sec. 5.4.1.1 are considered as baseline systems in order to measure the importance of our proposed second stage in the following experiments. Sec. 5.4.1.3 evaluates the ideal case where for each system (i.e., global and local) and comparison case, a different optimal subset of global and local features are selected (i.e., eight different optimal subsets are selected per system, four for random forgery cases and other four for skilled forgery cases) achieving therefore the best possible performance (although this could be unrealistic). In Sec. 5.4.1.4, we propose to apply our two-stage approach in order to select just one optimal subset of local and global features able to alleviate the degradation of the system performance on device interoperability scenarios. In Sec. 5.4.1.5, we propose a final fusion of both systems via weighted sum of the match scores

Table 5.3: Exp. 1: *System performance results in terms of EER (%) on the development dataset with and without applying the first data preprocessing stage proposed in this Thesis. Top: local system cases. Bottom: global system cases.*

Local System	*Skilled forgeries*	
Training vs Testing	Without stage 1	With stage 1
DS2 - DS2	7.1	8.6
DS3 - DS3	28.6	17.1
DS2 - DS3	45.5	27.3
DS3 - DS2	56.6	17.6
Global System	*Skilled forgeries*	
Training vs Testing	Without stage 1	With stage 1
DS2 - DS2	4.3	4.1
DS3 - DS3	25.3	14.8
DS2 - DS3	30.6	23.5
DS3 - DS2	49.3	46.9

considering the optimal global and local features selected in Sec. 5.4.1.4. In all these experiments only the development dataset of 50 users is considered.

Finally, our proposed optimal system obtained in Sec. 5.4.1.5 is tested using the remaining 70 users of the database not considered during the development and training process.

5.4.1.1. Experiment 1: Data Preprocessing Stage

This experiment aims to evaluate our proposed data preproccesing first stage. Both global and local systems are adjusted to each specific device (i.e., intra-device) without considering device interoperability scenarios yet, which is the common procedure in on-line signature verification. SFFS algorithm has been applied in order to improve the individually EERs for DS2 and DS3 datasets. Therefore, we consider two optimal feature subsets per system, one adjusted for DS2 dataset and another one fixed for DS3 dataset. Table 5.3 shows the performance for both systems. Analysing the device interoperability cases, we can observe a significantly system performance improvement after applying our proposed data preprocessing stage, specially for the local system. In case we do not consider this first stage, the performance of the systems on this device interoperability scenario increases to EERs around 50% in most cases. This first experiment proves the importance of the proposed preprocessing stage in this novel scenario. Analysing intra-device cases, the performance of global and local systems are very similar for the DS2 dataset. A small degradation of the system performance is produced for the local system after applying the data preprocessing stage due to pen-up information was discarded from DS2 as well as pressure information as this information is not recorded by the DS3 device. Therefore, for all experiments carried out using our first data preprocessing stage, we consider a total of 100 and 21 global and local initial features respectively. For the DS3 dataset, a high system performance improvement is achieved after applying the data preprocessing stage due to the sampling errors were corrected using the interpolation based on splines [Martinez-Diaz *et al.*, 2007].

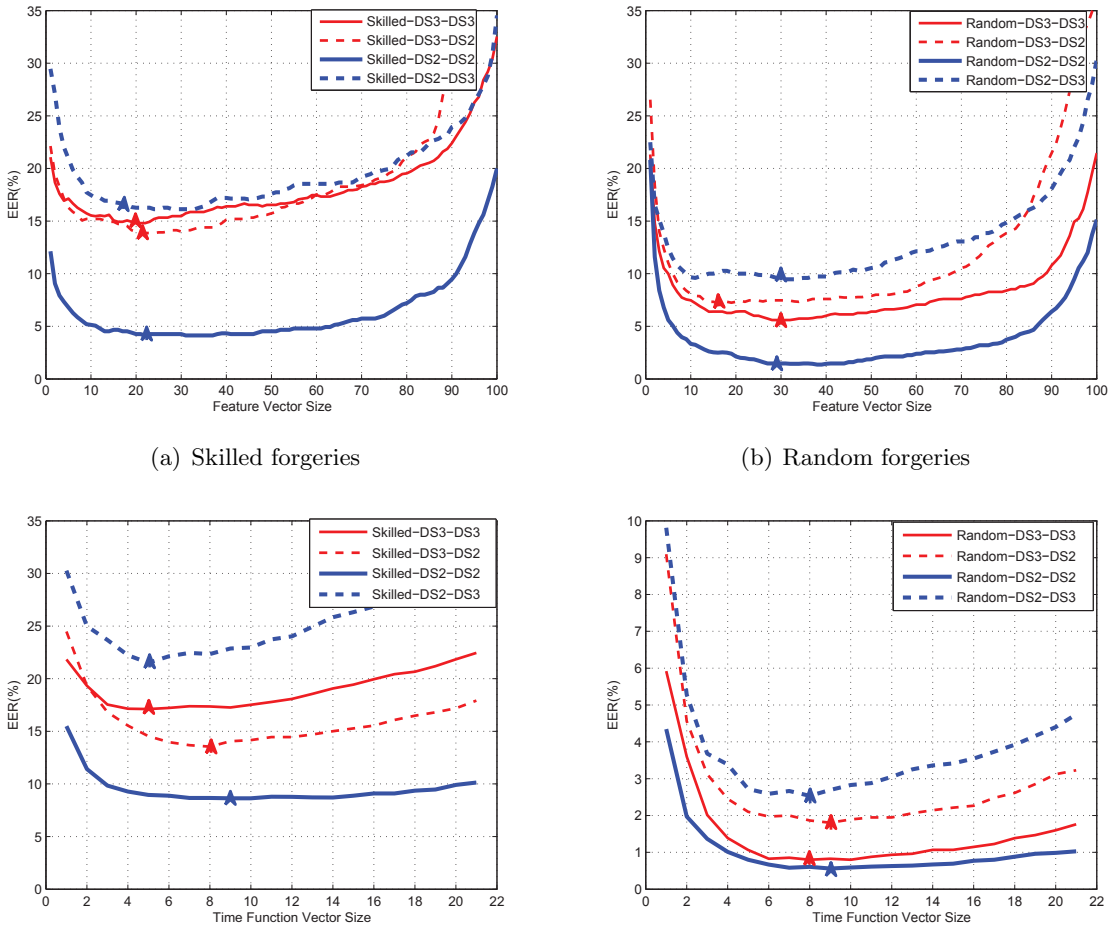
Table 5.4: Exp. 2, 3, and 4: *System performance results in terms of EER (%) on the development dataset. Top: local system cases. Bottom: global system cases.*

Local System	*Skilled forgeries*			*Random forgeries*		
Training vs Testing	Baseline *Exp. 2*	Individually optimised *Exp. 3*	Proposed *Exp. 4*	Baseline *Exp. 2*	Individually optimised *Exp. 3*	Proposed *Exp. 4*
DS2 - DS2	8.6	8.6	9.3	1.2	0.6	0.9
DS3 - DS3	17.1	17.1	18.1	2.1	0.8	1.5
DS2 - DS3	27.3	21.5	22.9	4.7	2.5	4.3
DS3 - DS2	17.6	13.6	15.7	5.1	1.8	2.9
Global System	*Skilled forgeries*			*Random forgeries*		
Training vs Testing	Baseline *Exp. 2*	Individually optimised *Exp. 3*	Proposed *Exp. 4*	Baseline *Exp. 2*	Individually optimised *Exp. 3*	Proposed *Exp. 4*
DS2 - DS2	4.1	4.1	8.3	3.5	1.5	4.0
DS3 - DS3	14.8	14.8	20.5	10.8	5.5	8.6
DS2 - DS3	23.5	16.1	20.0	23.3	9.5	13.7
DS3 - DS2	46.9	13.9	21.9	44.4	7.2	13.2

5.4.1.2. Experiment 2: Baseline System

In this experiment, the systems obtained in Sec. 5.4.1.1, which only considers the first data preprocessing stage, are now used as baseline systems. The optimal features were selected considering intra-device scenarios but not device interoperability scenarios. Therefore, the idea is to use these baseline systems to measure the improvement of our proposed second stage based on the selection of the optimal feature subsets considering both intra- and inter-device scenarios. The results obtained using both global and local baseline systems are included in Table 5.4.

Analysing the intra-device cases in Table 5.4, the performance of the system is much better for DS2 compared to DS3 datasets for both systems and impostor scenarios. This is due to the fact that DS2 device (pen tablet Wacom) is a higher quality device designed for capturing signatures and besides, in DS3 dataset signatures were captured under a mobility scenario where people had to sign standing and holding the PDA in one hand. Analysing the device interoperability cases, a high degradation of the system performance is produced in both systems and impostor scenarios. This degradation is more critical for the DS2 device, e.g., the system performance of the DS2 - DS3 device interoperability case increases up to 6 times for the global system. Therefore, in this experiment we can conclude that training and testing with different devices has a high impact on the performance, being the critical case when the quality of the device used for testing is worse than the quality of the device used for training. The performance of the system on device interoperability scenarios have been evaluated in recent works for random forgery cases [Blanco-Gonzalo *et al.*, 2014], but no solutions have been proposed for compensating the degradation of the performance between different quality devices apart from the data preprocessing stage. For this reason, the aim of the next experiments is to select just one optimal subset of global and local features able to alleviate the degradation of the system performance on device interoperability scenarios.



(a) Skilled forgeries

(b) Random forgeries

(c) Skilled forgeries

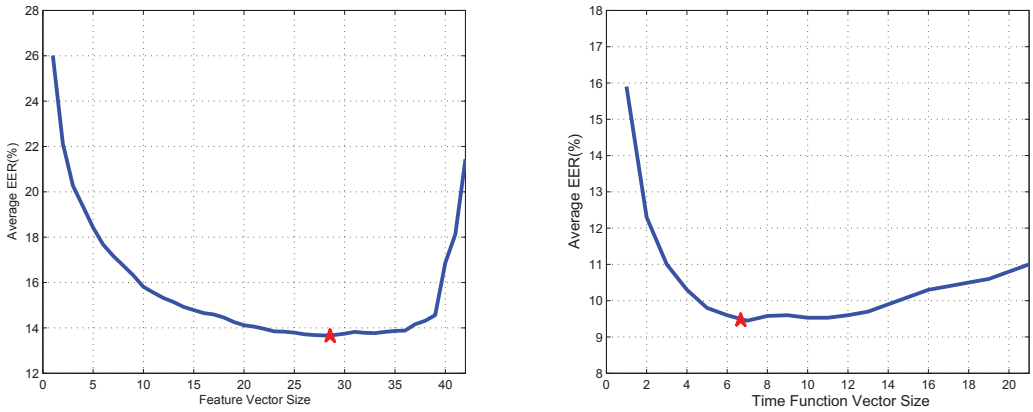
(d) Random forgeries

Figure 5.3: Exp. 3: *System performance results in terms of EER (%) on the development dataset for each possible size of the optimal feature vector selected by the SFFS algorithm. Top: global system cases. Bottom: local system cases.*

5.4.1.3. Experiment 3: Individually Optimised Systems

In this experiment, the goal is to obtain the best ideal possible performance achieved for both systems in an individually optimised case. It is important to highlight that this approach would not be possible to deploy in realistic applications as SFFS algorithm has been individually applied to each system (global and local) and impostor cases (i.e., 4 for random and 4 for skilled forgeries). Fig. 5.3 depicts the system performance results in terms of EER (%) for each possible size of the optimal feature vector selected by the SFFS algorithm. Table 5.4 represents the best EER obtained for each individually optimised case. The optimal global and local feature subsets are different for each case as it can be seen in Fig. 5.3, where the number of features selected for each case is depicted with a marker.

The performance of each individually optimised system is much better compared to the baseline system, especially for device interoperability cases. This is due to the fact that device



(a) Global system (b) Local system

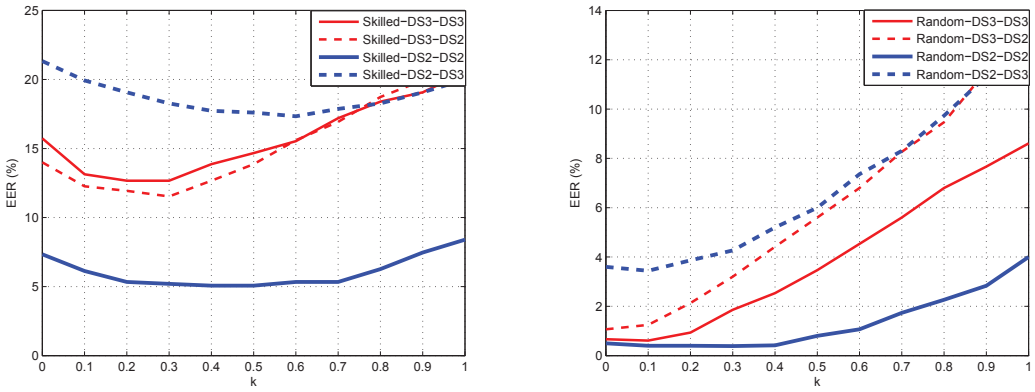
Figure 5.4: Exp. 4: *System performance results in terms of the Average EER (%) on the development dataset for each possible size of the optimal feature vector selected by the SFFS algorithm. The new SFFS criterion is considered in order to optimise the systems against device interoperability scenarios.*

interoperability scenarios have been taken into account by the SFFS algorithm in this individually optimised systems. In addition, it considers 16 different optimal feature vectors (one for each case and system), so this would not be possible to consider in realistic applications. These results allow us to know the performance limits we would be able to achieve in the best cases.

It is very interesting to remark the cases when the systems are trained and tested with DS3 and DS2 devices respectively (DS3 - DS2) compared to the case of training and testing using only the DS3 device (DS3 - DS3) for skilled forgery cases, as the system performance results achieved on device interoperability scenarios are even better than intra-device scenarios for both global and local systems. Finally, it is also important to note that for both systems the case DS2 - DS3 provides the worst system performance, so this specific challenging case would be taken into account by the SFFS algorithm in the next experiment.

5.4.1.4. Experiment 4: Proposed System

The main goal of this experiment is to select just one optimal subset of global and local features able to alleviate the degradation of the system performance on device interoperability scenarios. To achieve this, the two-stage approach proposed in this PhD Thesis has been applied, modifying therefore the criterion of the SFFS algorithm in order to obtain the lowest total EER (average of the EERs obtained for all the comparison cases) and the lowest EER for the DS2 - DS3 skilled forgery cases as this specific case provided the worst results in both global and local systems in Sec. 5.4.1.3. Fig. 5.4 shows the average system performance after applying the SFFS algorithm with the new criterion proposed. The optimal feature subsets are composed of 28 and 7 global and local features respectively. Features related to the geometry, speed and acceleration have been the most important ones for the global system whereas for the local system, local features related to Y-coordinate and velocity are the best performing.



(a) Skilled forgeries case

(b) Random forgeries case

Figure 5.5: Exp. 5: *System performance results in terms of EER (%) on the development dataset for the fusion of the global and local systems at score level for different values of the fusion weighting coefficient k.*

The system performance of both global and local systems is represented in Table 5.4 for both skilled and random forgeries. In general, good results have been achieved for device interoperability scenarios compared to the original baseline systems. This improvement is especially noticeable for the global systems.

Analysing the device interoperability cases for the global system, our proposed system provides an average relative improvement of 40.5% EER for skilled forgeries and 60.3% EER for random forgeries compared to the baseline system. Besides, it is important to note that an absolute improvement of 3.5% EER has been achieved for the most challenging case (skilled - DS2 - DS3) compared to the baseline system.

Analysing the device interoperability cases for the local system, the proposed system provides an average relative improvement of 14.0% EER for skilled forgeries and 26.5% EER for random forgeries compared to the baseline system. In addition, and as it happens for the global system, an absolute improvement of 4.4% EER has been achieved for the most challenging case (skilled - DS2 - DS3) compared to the baseline system.

5.4.1.5. Experiment 5: Fusion of the Proposed Systems

In this experiment we propose a final fusion of the optimal global and local systems obtained in Sec. 5.4.1.4 to further improve the system performance. This fusion is performed via weighted sum of the match scores [Kittler *et al.*, 1998]. Before applying fusion of the systems, all scores are normalised in a range [0,1] using tanh-estimators [Jain *et al.*, 2005]. The final fusion score s_f is obtained as:

$$s_f = k \cdot s_g + (1 - k) \cdot s_l \tag{5.1}$$

where s_f is the final score, and s_g and s_l are the match scores of the global and local systems

Table 5.5: Exp. 5: *System performance results in terms of EER (%) on the development dataset for global and local systems, and final fusion of them.*

	Skilled forgeries			*Random forgeries*		
Training vs Testing	Global	Local	Fusion	Global	Local	Fusion
DS2 - DS2	8.3	9.3	5.2	4.0	0.9	0.4
DS3 - DS3	20.5	18.1	12.7	8.6	1.5	1.9
DS2 - DS3	20.0	22.9	18.3	13.7	4.3	4.3
DS3 - DS2	21.9	15.7	11.5	13.2	2.9	3.2

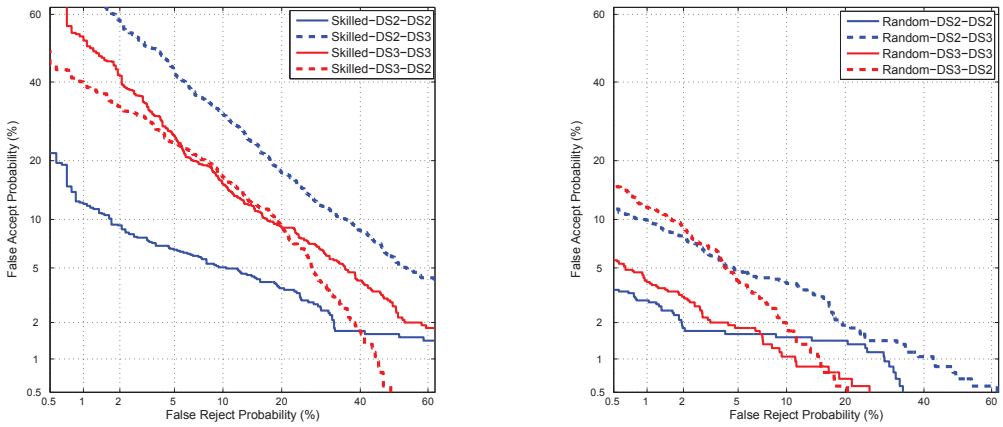
Table 5.6: Exp. 6: *System performance results in terms of EER (%) on the evaluation dataset for the fusion of the optimal global and local systems via weighted sum of scores. Comparison of the results obtained by baseline and proposed systems choosing a k value of 0.3 for the fusion.*

Fusion of Systems	*Skilled forgeries*		*Random forgeries*	
Training vs Testing	Baseline	Proposed	Baseline	Proposed
DS2 - DS2	7.1	6.2	3.4	2.0
DS3 - DS3	11.3	12.8	2.6	2.7
DS2 - DS3	21.5	18.9	10.6	4.9
DS3 - DS2	14.8	13.4	4.7	4.7

respectively. The fusion weighting coefficient k has been heuristically set by observing the performance of the system in terms of the EER and taking into account all the cases at the same time. Fig. 5.5 depicts the performance of the fusion system for different values of k. In general, the system performance gets worse for both random and skilled forgeries when we choose a high value of k, whereas for a low value of k the performance of the system also gets worse for skilled forgeries. For this reason, a k value of 0.3 has been finally selected as it provides a good performance for all cases at the same time. Therefore, the local system outweighs the global system in the final score. Table 5.5 shows the individual performance of global and local systems (see Sec. 5.4.1.4) and the final fusion performance. The performance of the proposed fusion system is much better compared to the individual performance of the systems in most cases, especially for skilled forgeries where the proposed fusion system provides an average relative improvement of 27.7% EER compared to the local system.

5.4.1.6. Experiment 6: Validation Experimental Results

Our final proposed system based on the fusion of the optimal global and local systems are now tested using the evaluation dataset composed of the remaining 70 users of the Biosecure datasets. Fig. 5.6 shows the system performance results in terms of the DET curve. The EERs achieved for both baseline and proposed systems using a fusion weighting coefficient $k = 0.3$ are shown in Table 5.6. Analysing the device interoperability cases, the proposed system provides an average relative improvement of 11.0% EER for skilled forgeries and 37.3% EER for random forgeries compared to the baseline system. Therefore, these results are similar compared to the same ones obtained in the development phase, proving the robustness of the proposed scheme.



(a) Skilled forgery cases

(b) Random forgery cases

Figure 5.6: Exp. 6: *DET curves for the final signature recognition system based on fusion of the proposed global and local systems on the evaluation dataset.*

5.4.1.7. Biosecure Database Conclusions

The experimental work carried out using the Biosecure database has revealed a high degradation of the signature verification systems on device interoperability scenarios. Our proposed two-stage approach has proved to be very useful, alleviating the degradation of the system performance on this novel scenario. However, despite the improvement achieved, the system performance is still highly affected by this challenging scenario with results around 20% EER for skilled forgeries when the quality of the training device is higher than the testing device (DS2 - DS3).

For this reason, in order to assess the real impact of device interoperability scenarios on current devices and applications, the new e-BioSign database, which comprises a total of 5 COTS devices (3 Wacom devices specifically designed for capturing handwriting and 2 Samsung general purpose tablets), is now used in order to evaluate the system performance on current scenarios after applying our two-stage proposed approach.

5.4.2. e-BioSign Database

Two main experiments are carried out using the e-BioSign database in order to analyse device interoperability scenarios for both stylus and finger cases. Sec. 5.4.2.1 analyses the system performance of each individual device considering the scenario of having signatures from the same device and writing input. Both global and local systems are considered on this analysis. Then, in Sec. 5.4.2.2 the scenario of considering signatures from different devices but same writing input for training and testing the system is studied.

Table 5.7: *Global features considered in the global baseline system. Global feature # taken from Tables 4.1 and 4.2.*

#	**Feature description**
1	Signature total duration Ts
2	N(pen-ups)
36	$(x_{max}$ - $x_{min})$ / $x_{acquisition\ range}$
67	$(y_{max}$ - $y_{min})$ / $y_{acquisition\ range}$
101	Average pressure $\overline{p}$

5.4.2.1. Experiment 1: Intra-Device Analysis

Two different approaches are considered in this experiment. First, baseline systems whose local and global features are fixed from previous works (see Tables 5.1 and 5.7). Second, the proposed systems are adjusted using our proposed two-stage approach described in Sec. 5.1 over the development dataset in order to improve the system performance. The SFFS algorithm has been individually applied to each device and writing input in order to select the optimal feature subsets for each specific case. The first approach is considered as Baseline (B) whereas the second one is considered as Proposed (P).

Tables 5.8 and 5.9 show the results for both Baseline and Proposed approaches considering the 35 users of the evaluation dataset. The global system is only considered for the case of using the stylus (Table 5.9) as information related to pen ups and pressure is not available in the W5 device when signatures are acquired using the finger.

Analysing Table 5.8 for the case of using the stylus as the writing input, the baseline systems achieve an average EER of 11.7% and 1.9% for skilled and random forgery cases respectively whereas for the proposed systems the average EER improves to 10.1% and 1.1% for skilled and random forgeries respectively. These results show the benefits of using the SFFS algorithm over a development dataset in order to select the optimal local features. In addition, two important observations can be highlighted from the results. First, very similar system performance has been achieved for general purpose devices (i.e., W4 and W5) compared to devices specifically designed to capture on-line handwriting and signatures (i.e., W1, W2 and W3) for both skilled and random forgery cases when the stylus is used as the writing input. This shows the feasibility of general purpose devices in real banking and commercial applications. Second, it is worth noting a lower performance of the W3 device compared to the other devices despite being a high quality Wacom device. We attribute this fact to the the user interface as a cross shape marker was included on the display during the acquisition process in contrast to the other devices, and this could have been uncomfortable for the users.

Analysing Table 5.8 for the case of using the finger as the writing input, the baseline systems achieve an average EER of 25.5% and 1.9% for skilled and random forgery cases respectively whereas for the proposed systems the average EER improves to 24.3% and 0.7% for skilled and random forgeries respectively. A high degradation of the system performance is produced for skilled forgery cases when using the finger as the writing input. A thorough analysis of this novel finger scenario has been already performed in Sec. 5.3.

Table 5.8: Intra-device scenario: *System performance results in terms of EER (%) on the evaluation dataset for the* ***local systems*** *when signatures are acquired using stylus and finger. B = Baseline and P = Proposed.*

	STYLUS										FINGER			
	W1		W2		W3		W4		W5		W4		W5	
	B	P	B	P	B	P	B	P	B	P	B	P	B	P
Skilled	10.0	8.3	10.0	10.0	15.7	13.6	10.0	7.9	12.9	10.7	24.0	22.1	27.0	26.4
Random	1.4	0.0	1.1	0.7	4.3	2.9	0.8	0.7	2.1	1.0	1.4	0.3	2.3	1.0

Table 5.9: Intra-device scenario: *System performance results in terms of EER (%) on the evaluation dataset for the* ***global systems*** *when signatures are acquired using stylus. B = Baseline and P = Proposed.*

	STYLUS									
	W1		W2		W3		W4		W5	
	B	P	B	P	B	P	B	P	B	P
Skilled	13.6	13.5	12.9	16.4	22.6	19.3	14.3	17.9	19.3	10.0
Random	12.1	10.7	12.1	13.6	20.8	17.9	11.4	12.1	17.9	6.4

Analysing Table 5.9, the average EER for the baseline systems is 16.5% and 14.8% for skilled and random forgery cases respectively whereas for the proposed systems the average EER is 15.4% and 12.1% for skilled and random forgery cases respectively. Therefore, global systems do not achieve as good results as local systems. In particular, the results obtained for random forgery cases are very critical. For this reason, the global system has been discarded in the remaining experiments.

Finally, we compare the best system performance results obtained on the e-BioSign database to the state-of-the-art results obtained using other existing on-line signature databases (see Table 3.1). However, note that most algorithms and experimental conditions vary between the listed studies, e.g., the amount and type of training and testing data. For the case of using the stylus as the writing input, the best results obtained in the present study are 7.9% and 0.0% EER for skilled and random forgery cases respectively. The result obtained for skilled forgeries is a bit higher in terms of EER compared to the results obtained in the two largest databases (i.e., 6.20% and 4.77% EER for Biosecure and BiosecurID databases, respectively). One of the possible reasons for this effect could be the high quality of the forgeries considered on e-BioSign database as forgers could even place on the screen device a paper with the image of the signatures to forge. For the random forgery case, the best result achieved here is 0.0% EER, which is the best one compared to other databases. Analysing the case of using the finger as the writing input, the best results obtained are 17.9% and 0.3% EER for skilled and random forgeries, respectively. For the skilled forgery case, the result obtained using the e-BioSign database has outperformed the preliminary results obtained in [Martinez-Diaz *et al.*, 2016], in which users were asked to perform a simplified version of their signature (a.k.a. pseudo-signatures) based on their initials or part of their signature flourish. For the random forgery case, the result obtained is very close to zero, similar to the result obtained in [Blanco-Gonzalo *et al.*, 2014].

Table 5.10: Inter-device scenario: *System performance results in terms of EER (%) for the proposed local system when signatures are acquired using the* ***stylus****. Skilled and random forgery results are shown on top and bottom of each cell respectively.*

		Test				
		W1	W2	W3	W4	W5
Train	W1	10.7 0.7	7.9 0.8	15.7 5.0	10.7 0.7	10.7 2.1
	W2	11.4 1.1	10.0 0.7	16.4 5.7	14.3 0.7	11.4 1.6
	W3	9.3 0.3	8.6 0.7	13.6 2.1	11.2 0.0	11.4 1.4
	W4	10.0 0.7	9.3 0.9	17.1 5.0	10.7 0.7	11.4 1.4
	W5	12.7 1.4	10.0 1.1	16.9 5.0	12.1 0.7	11.2 1.4

Table 5.11: Inter-device scenario: *System performance results in terms of EER (%) for the time functions-based system when signatures are acquired using the* ***finger****. Skilled and random forgery results are shown on top and bottom of each cell respectively.*

		Test	
		W4	W5
Train	W4	19.3 0.7	23.5 0.2
	W5	24.2 0.7	22.9 0.3

5.4.2.2. Experiment 2: Inter-Device Analysis

In this experiment we focus on device interoperability scenarios on COTS devices. The idea is to evaluate the potential of our proposed two-stage approach on current devices as the initial analysis carried out in Sec. 5.4.1 for the Biosecure database considered old-fashion devices with a high different quality level. This could be the reason for the high degradation of the system performance obtained on device interoperability scenarios (i.e., DS2 - DS3 case).

Two different systems are developed for all five devices (one for signatures acquired using the stylus and another one for the finger) using the 30 users of the development dataset. In order to select the optimal feature subsets for this novel device interoperability scenario, the SFFS algorithm has been applied. For the stylus case, a total of 5 genuine signatures (1 signature per device) of the first session are considered as training signatures whereas a total of 20 genuine signatures (4 signatures per device) of the second session are left for testing. For the finger case, a total of 4 genuine signatures (2 signatures per device) of the first session are considered as training signatures whereas the 8 genuine signatures (4 signatures per device) of the second session are left for testing. Then the systems developed were tested on the 35 users of the evaluation dataset. The results achieved are depicted in Tables 5.10 and 5.11.

Table 5.10 shows all possible device combinations for training and testing the systems when the stylus is used as the writing input. The diagonal of Table 5.10 (highlighted in darker colour) contains all results without device interoperability. The proposed system developed for this scenario achieves an average EER for device interoperability cases of 11.9% and 1.8% for skilled

and random forgery cases respectively whereas for intra-device cases the average EER is 11.2% and 1.1% for skilled and random forgery cases respectively. Therefore, very similar results have been achieved for both intra- and inter-device scenarios when the stylus is considered as the writing input. These results show the importance of applying our proposed two-stage approach in order to compensate device interoperability scenarios. It is also important to note the system performance obtained when W3 device is used for testing the system. For these cases, the average EER increases up to 15.9% and 4.6% for skilled and random forgery cases respectively. This degradation of the system performance could have been produced due to the same reasons explained in the previous experiment.

Table 5.11 shows all possible device combinations for training and testing the system when the finger is considered as the writing input. The proposed system developed for this scenario achieves an average EER for the device interoperability cases of 23.9% and 0.5% for skilled and random forgeries respectively whereas for the intra-device scenario the average EER is 21.1% and 0.5% for skilled and random forgeries respectively. Therefore, the same observations previously extracted for the stylus case can be also applied here when writing with the finger in a general mobile device, but with a worsening of the system performance due to the higher variability on this finger input scenario.

Finally, some important conclusions can be extracted from the results obtained using the e-BioSign and Biosecure databases. The high technological evolution and sensor quality improvement together with our proposed two-stage approach for dealing with device interoperability lead to very competitive signature verification systems on this novel scenario.

5.5. Mixed Writing-Input Scenarios

In this section we explore a new scenario where on-line signature verification systems are trained and tested using signatures from the same device but different acquisition tools (i.e., stylus and finger). This scenario can be very useful for many real applications where the user first register in the system using the stylus and then in posterior usages they could make use of their personal smartphone or tablet devices using the finger as the writing input. Two different local systems are trained in this section, one for the W4 device and another one for the W5 device. In order to select the optimal feature subsets for the mixed writing-input scenario, the SFFS algorithm has been applied on the development dataset. SFFS has been individually applied to W4 and W5 devices considering a total of 4 genuine signatures (2 signatures per writing tool) as training signatures and 8 genuine signatures (4 signatures per writing tool) of the second session as testing signatures. Table 5.12 shows the results obtained for each device in this mixed writing-input scenario considering the 35 users of the evaluation dataset.

Analysing the skilled forgery cases, results show in general a system performance degradation when signatures acquired using the finger are considered for training or testing the systems. For this case, an average 19.0% EER is achieved when the finger is considered for both training and testing the system whereas for the mixed writing-input scenario the average EER is 21.8%.

Table 5.12: Mixed writing-input scenario: *System performance results in terms of EER (%) for the proposed local systems. Skilled and random forgery results are shown on top and bottom of each cell respectively.*

		Test			
		W4		W5	
		Stylus	Finger	Stylus	Finger
Train	Stylus	12.9 0.7	22.9 0.7	12.9 0.1	17.9 0.2
	Finger	27.9 0.7	20.0 0.7	18.6 0.7	17.9 0.5

Therefore, although the system performance is slightly worse for the mixed writing-input scenarios, the main problem resides in the signatures acquired with the finger. In addition, it is important to note that for both devices the worst mixed writing-input case seems to be when the system is trained using signatures acquired through the finger and testing with the stylus but this would not be a common application scenario.

Analysing the random forgery cases, similar results can be observed for both devices when using the same or different writing input for training and testing. Therefore, the deployment of signature verification in real applications on this new and challenging finger and mixed writing-input scenarios seems to be feasible when random forgeries are considered.

5.6. Chapter Summary and Conclusions

This chapter has evaluated the functioning of traditional on-line signature verification systems on the following emerging scenarios: *i*) finger input, *ii*) device interoperability, and *iii*) mixed writing-input. For the analysis, both Biosecure and e-BioSign databases have been considered in the experimental work in order to perform a complete analysis of these novel scenarios using traditional and COTS devices.

In order to alleviate the degradation of the system performance on these novel scenarios, we have proposed a two-stage approach based on a first data preprocessing stage applied to achieve a high similarity between signatures coming from different devices and writing tools, and second, a new criterion in order to extract and select the optimal features for each specific scenario.

Our proposed approach has proved to be decisive for the system performance on these novel scenarios, especially for device interoperability and mixed writing-input.

We have first studied the new finger input scenario, showing a higher EER for skilled forgeries compared to the case of using the stylus as the writing input. We have observed that this degradation is produced due to two main factors. First, the higher variability as in general users who perform their signatures using closed letters (i.e., a, e, o, l, p, q, etc.) tended to perform much larger writing executions in comparison with other letters due to the lower precision they are able to achieve using the finger. Besides, users whose signatures are composed of a long name and surname (or two surnames) tend to simplify some parts of their signatures on the finger scenario. Finally, it is important to remark the very challenging scenario considered in

this experiment as forgers had access to the dynamic realization of the signatures to forge. A recommendation for the usage of signature recognition on mobile devices would be for the users to protect themselves from other people that could be watching while signing, as this is more feasible to do in a mobile scenario compared to an office scenario. This way skilled forgers might have access to the global shape of the signature but not to the dynamic information. Therefore, the higher variability together with the challenging forgeries considered in this experiment conclude that these new finger input scenario can be applied to those scenarios where the knowledge of the impostor about the user to forge is scarce, e.g., random forgeries.

Then, we have focused on device interoperability scenarios considering both traditional and COTS devices. We have first applied our proposed two-stage approach on the Biosecure database, proving to be very useful against skilled forgeries with average relative improvements of 40.5% and 14.0% EER for global and local systems, respectively. We have also proposed a final fusion of the systems with an average relative improvement of 27.7% EER. However, and despite the improvements achieved, the system performance was still highly affected with results around 20% EER when the quality of the device considered for training is better than the testing device (DS2 - DS3). The new e-BioSign database has then considered in order to analyse our proposed two-stage approach on COTS devices. The high technological evolution and the sensor quality improvement together with our proposed two-stage approach have led to very competitive signature verification systems on device interoperability scenarios with an average EER of 11.9% and 1.8% EER for skilled and random forgeries, respectively.

Finally, we have also evaluated a mixed writing-input scenario as it can be very useful for many real applications where the user first register in the system using the stylus and then in posterior usages they could make use of their personal smartphone or tablet devices using the finger as the writing input. In general, a high degradation of the system performance has been produced compared to the traditional stylus scenario. An average 19.0% EER is achieved when the finger is considered for both training and testing the system whereas for the mixed writing-input scenario the average EER is 21.8%. Therefore, although the system performance is slightly worse for the mixed writing-input scenarios, the main problem resides in the signatures acquired with the finger. In addition, it is important to note that for both devices the worst mixed writing-input case seems to be when the system is trained using signatures acquired through the finger and testing with the stylus but this would not be a common application scenario. Therefore, an exhaustive analysis of the finger scenario must be carried out in order to understand and propose systems that are better adapted to this thriving scenario.

Chapter 6

Long-Term Multi-Session Acquisition Scenarios

In this chapter, we explore a novel scenario where the number of stored samples or templates per user can grow very fast, making it possible to train more robust statistical user models, improving the performance of the biometric systems and in particular reducing the template aging effect. This chapter carries out an exhaustive experimental analysis of template update strategies for three popular on-line signature verification approaches, extracts various practical findings related to the template aging effect in signature biometrics, and configures time-adaptive improved versions of the considered baseline approaches overcoming to some extent the template aging. Fig. 6.1 illustrates the concept of template update as studied here. The Traditional Approach on top only uses for enrolment an initial collection of genuine signatures. In the Present Study (Fig. 6.1 bottom), we explore ways to incorporate additional enrolment data coming across time.

Based on the main scenario in our mind mentioned before (in-branch banking operations), we assume that signatures coming across time are all genuine (See Fig. 6.1) after some kind of human validation (typical when collecting signatures in bank branches, e.g., checking ID cards or knowing the customer). The problem is therefore authenticating a new signature ($t = T + 1$ in Fig. 6.1) based on a collection of past genuine signatures.

This chapter is structured as follows. Sec. 6.1 describes the methods studied in this work in order to reduce the template aging effect. Sec. 6.2 describes the three signature systems and the experimental work carried out. Finally, Sec. 6.3 draws the final conclusions.

This chapter is based on the following publications: [Tolosana *et al.*, 2018f, 2015e].

6.1. Methods

6.1.1. Template Update Strategies

Two different approaches are analysed in the experimental work for template update:

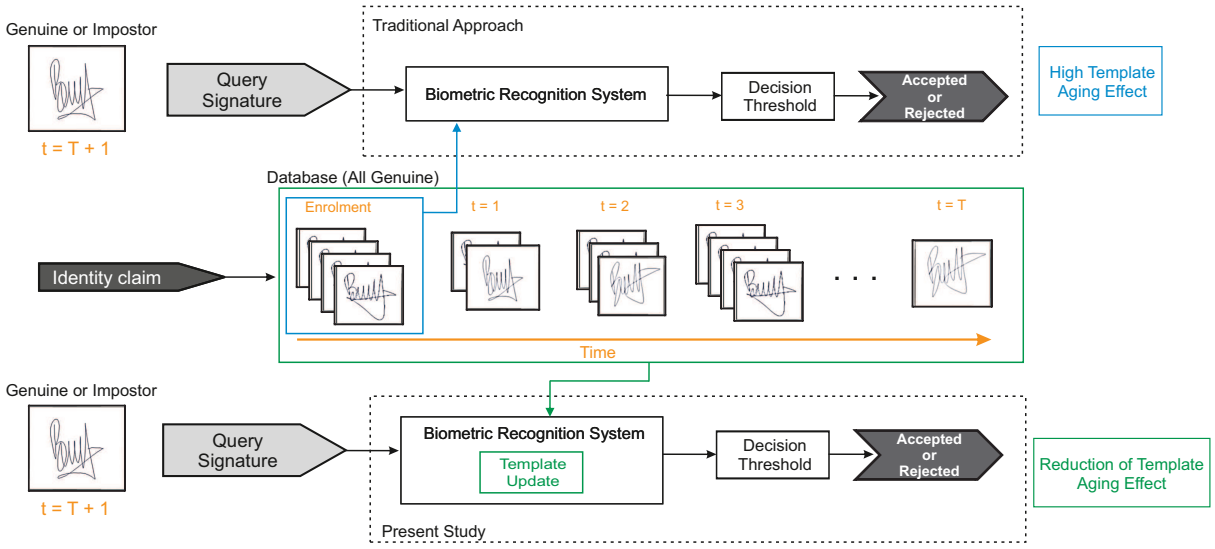


Figure 6.1: *Template update concept compared to the traditional one based only in an initial collection of enrolment signatures.*

1. Adding all the available signatures of the subject at hand across time to the ones from the enrolment session.

2. Not considering old signatures from the subject at hand for updating the user model.

Several intermediate configurations are also analysed in order to study the performance evolution of the three systems considered. Besides, it is important to highlight that many additional factors are considered in this stage (i.e., computational cost, resources, etc.) as they are very important for practical applications.

6.1.2. System Complexity Configuration

Finding an optimal system configuration for a given task can provide significant improvements of performance. In [Fierrez *et al.*, 2007], a preliminary analysis of the system performance was carried out considering different system configuration parameters for an HMM-based online signature verification system. The main limitation of that work was that only 5 training signatures were considered in the user models to study the optimal system configuration parameters, so a broader study including different number of training signatures per user model will be helpful.

An exhaustive analysis of the system performance considering different system configuration parameters was first carried out in this Thesis for HMM- and GMM-based systems in scenarios where the number of available training signatures per user increases with time. Table 6.1 summarises the optimal system configuration parameters proposed, which are considered in this chapter as a starting point in order to perform a more comprehensive analysis of template update strategies for the HMM and GMM systems. In Table 6.1 we see that when the number of available training signatures is small, the optimal system configuration for the HMM system

Table 6.1: *Optimal system configuration parameters regarding the number of available training signatures. N denotes the number of hidden states and M the number of Gaussian mixtures per state.*

	HMM		GMM
# Training Signatures	N	M	M
<15	2	16	32
16 to 31	32	2	128
>31	64	2	512

is based on a small number of hidden states ($N = 2$) and a medium number of Gaussian mixtures per state ($M = 16$). On the other hand, as the number of available signatures increases (between 16 and 31 signatures), then the number of optimal hidden states increases ($N = 32$) and the number of mixtures per state decreases to $M = 2$. Finally, the number of hidden states increases up to $N = 64$ for the case of having more than 31 available signatures. For a GMM-based system, as the number of training signatures increases, the number of Gaussian mixtures also increases ($M = 512$ for 41 training signatures). For a thorough analysis of these selection parameters, we encourage the reader to see [Tolosana *et al.*, 2015e].

6.1.3. Statistical Analysis

For interpreting our results we have applied a statistical analysis similar to [Sae-Bae and Memon, 2015]. In that work the authors proposed a metric to measure the quality of an on-line signature template derived from a set of enroled signature samples in terms of its distinctiveness against random forgeries. The use of random and not skilled forgeries for measuring our proposed template quality (Q) is motivated due to the lack of skilled forgeries in real scenarios for training.

Let (μ_g,σ_g) and (μ_r,σ_r) be the mean and standard deviation of the genuine and random matching score distributions provided by the on-line signature verifier, then the template quality for these two distributions is defined as follows:

$$Q = \frac{\|\mu_g - \mu_r\|}{\sqrt{(\sigma_g^2 + \sigma_r^2)/2}} \tag{6.1}$$

The goal of this template quality metric Q is to measure how separated are the genuine from the random matching score distributions. The larger the separation between the score distributions, the higher is Q (note that Q is equivalent to the metric d defined in [Daugman, 2000]). In this work, we compute both the EER and this Q metric in order to analyse the different template update strategies proposed.

6.2. Experiments

6.2.1. On-Line Signature Verification Systems

Three well-known local systems described in Sec. 4.1.2 are considered here: HMM, GMM and DTW. In all of them, signals captured by the digitizer (only X and Y coordinates and

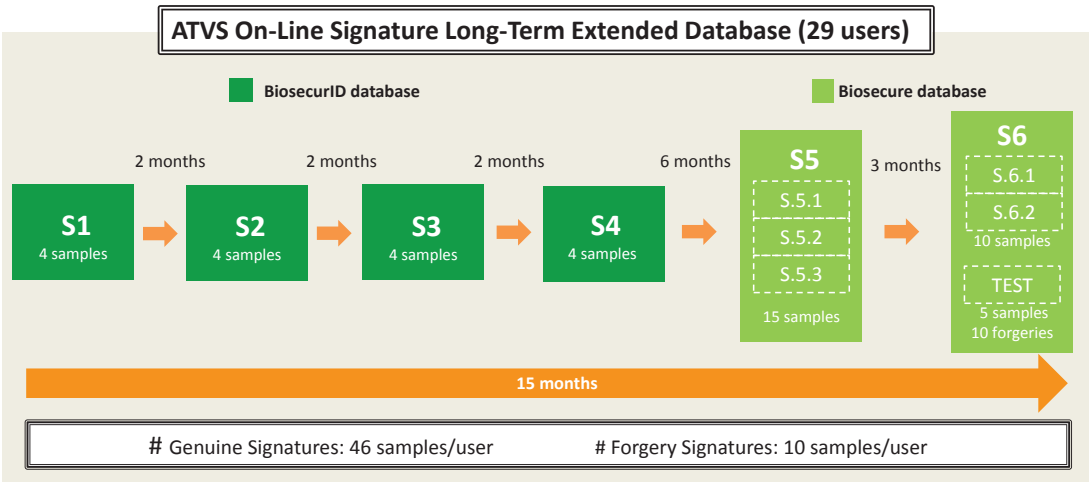


Figure 6.2: *General time diagram of the different acquisition sessions and number of genuine signatures per user that form the ATVS On-Line Signature Long-Term Extended Database.*

pressure) are used to extract a set of 23 local features for each signature [Tolosana *et al.*, 2015e]. Information related to pen angular orientation (azimuth and altitude angles) was discarded in order to consider the same set of local features that we would be able to use in general purpose devices such as tablets and smartphones. For the HMM- and GMM-based systems considered in this work, the optimal subset of local features used in the experiments is based on [Martinez-Diaz *et al.*, 2014] whereas for the DTW-based system is based on [Tolosana *et al.*, 2015d]. Both subsets comprise 9 local features and were generated using the SFFS algorithm described in Sec. 2.1.5.

A final fusion of the three systems after applying template update is also performed computing the sum of the matching scores. The sum rule fusion algorithm is one of the most successful and easiest approaches used in many related works [Fierrez *et al.*, 2018a; Kittler *et al.*, 1998]. Before applying the fusion, the scores from each system are normalised to a common range [0,1] using tanh-estimators [Jain *et al.*, 2005].

6.2.2. Experimental Protocol

The experimental protocol has been designed to enable the study of the template aging effect and template update strategies for on-line signature authentication. For this, the extended version of the ATVS On-Line Signature Long-Term database, which has been described in Sec. 3.2.2, is divided into several training sets (in order to analyse different cases and obtain optimal strategies for each one), but only one test set composed of the last 5 genuine samples and 10 skilled forgeries samples (i.e., TEST block in Fig. 6.2). This way, fair comparative analysis can be carried out as all experiments use the same signatures for test. Skilled forgery scores are obtained by comparing training signatures against the 10 available skilled forgeries for the same user whereas random or zero-effort forgery scores are obtained by comparing the training signatures to one genuine signature of the remaining users.

Table 6.2: *Experimental protocol designed to study the template aging effect (Sec. 6.2.3.1), and template update strategies (Sec. 6.2.3.2 and 6.2.3.3). p/s indicates de number of signatures used per session.*

Experiments	Training	# Signatures	# Sessions	Aging Analysis	Template Update
A	S1	4	1	X	
B	S2	4	1	X	
C	S3	4	1	X	
D	S4	4	1	X	
E	S.5.2	4	1	X	
F	S.6.2	4	1	X	
G	S1	4	1		X
H	S1-S4	16	4		X
I	S1-S5	31	5		X
J	S1-S4, S.5.2	20 (4 p/s)	5		X
K	S2-S4, S.5.2	16 (4 p/s)	4		X
L	S3, S4, S.5.2	12 (4 p/s)	3		X
M	S4, S.5.2	8 (4 p/s)	2		X
N	S.5.2	4	1		X
O	S5	15	1		X

First, Sec. 6.2.3.1 performs an analysis of the template aging effect in on-line signature verification by comparing sets of training data from different sessions with the test set. Second, Sec. 6.2.3.2 and 6.2.3.3 carry out an exhaustive search of combinations of training data in order to find an optimal template update strategy for each of the systems studied. All experiments considered in this work for studying the aging effect and template update strategies are depicted in Table 6.2, which details the number of signatures used for training and the session(s) they come from. A final fusion of the three optimal systems is carried out in Sec. 6.2.3.4 in order to provide an improved system performance and reduce the template aging effect.

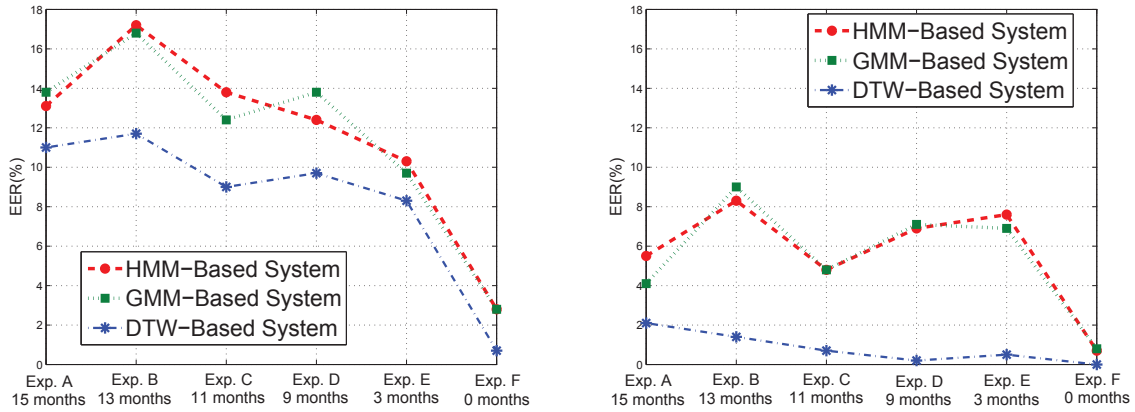
6.2.3. Experimental Results

6.2.3.1. Template Aging Analysis

The aim of this section is to analyse the template aging effect for on-line signature verification. Thus, six different experiments (Exp. A to Exp. F, as depicted in Table 6.2) have been considered. In all cases four signatures from different sessions are used for training. Exp. A contains training signatures from the first session (S1) with a 15-month time gap with the test session. For the following experiments the time gap (in months) between the training and test data are 13, 11, 9 and 3 months for Exp. B, C, D, and E, respectively. Finally in Exp. F signatures from the same session (S6) are used for training and test, so the time gap in this case is just a few minutes.

Experiments have been conducted for the three systems considered (i.e., DTW, HMM and GMM). The system configuration parameters of the HMM-based system chosen for this aging analysis are $N = 2$ and $M = 16$, whereas for the GMM-based system we use $M = 32$, being these system configuration parameters the optimal ones depicted in Table 6.1.

Fig. 6.3 shows the performance of the systems for all the experiments and for both skilled and random forgeries. Analysing the skilled forgery cases in Fig. 6.3(a), a general improvement



(a) Skilled forgery cases

(b) Random forgery cases

Figure 6.3: Template Aging Analysis*. Below each experiment is included the time gap between the training and testing signatures.*

of the performance is achieved for the three systems considered when the elapsed time between the testing and training signatures is reduced. For example, the average performance in terms of EER (%) of the three systems taking into account Exp. A is 12.6% whereas for Exp. E is 9.4%.

However, when analysing random forgery cases (Fig. 6.3(b)) the effect of the elapsed time does not affect in the same way the performance of the three systems. The performance of the DTW-based system keeps improving as the time between training and testing signatures is reduced (i.e., from 2.1% EER for Exp. A to 0.5% EER for Exp. E). However, for the HMM and GMM, the performance does not improve as the time gap is reduced, only showing a very significant improvement of performance for Exp. F, in which the data used for training and testing comes from the same session.

In addition, it is important to highlight that DTW achieves much better performance than HMM and GMM for all the experiments in these conditions. This is due to the fact that DTW is an elastic technique whereas HMM and GMM are statistical algorithms. Therefore, as the number of training signatures considered in these experiments is small (i.e., 4 signatures), it makes sense that DTW works better than HMM and GMM systems, which agrees with previous works [Fierrez-Aguilar *et al.*, 2005a]. On the other hand, for an increasing volume of enrolment data, as happens in some of the setups explored in Sec. 6.2.3.2, we will see that the statistical models HMM and GMM are superior to DTW.

Finally, it is also worth noting the results of Exp. F. This experiment does not consider inter-session variability as training and testing signatures come from the same session. In this experiment, the performance for the three considered systems is much better compared to the previous experiments. However, it is important to highlight this is an unusual case as it would only happen in a real application during the enrolment day.

As a general conclusion, we can confirm that on-line signature verification is significantly

affected by aging. These trends coincide with previous experiments performed in [Galbally *et al.*, 2013] where signatures from S1 were considered as training signatures and the rest of sessions were used as test, and in [Sae-Bae and Memon, 2014] where the degradation of the system performance increased when the time lapse between training and test signatures also increased. The goal of the following sections is to reduce the effect of the template aging for on-line signature verification considering different template update strategies regarding the number of available training signatures and the elapsed time between training and testing.

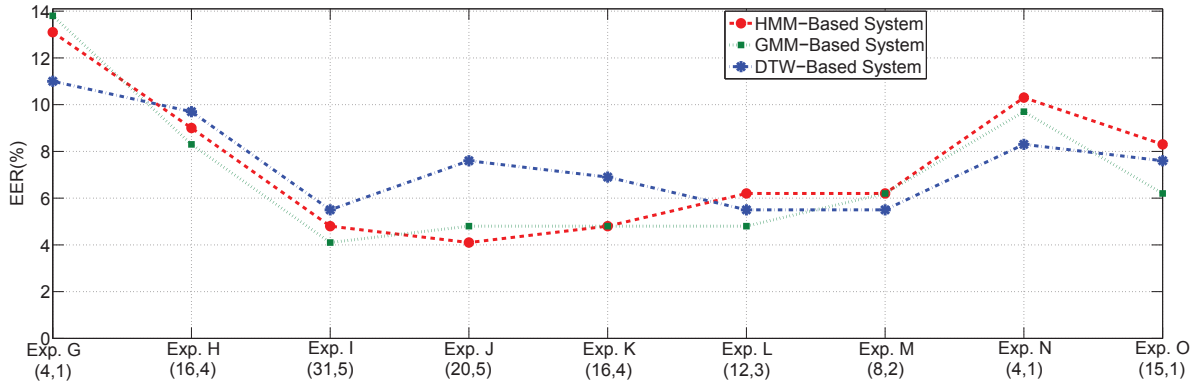
6.2.3.2. Template Update Strategies

This section focuses on template update strategies given a set of training signatures per user acquired at different sessions, with the final goal of reducing the aging effect. Two different methods are analysed: 1) Adding newer signatures to the enrolment ones; and 2) removing signatures from the older sessions from all the available training signatures. All experiments considered are depicted in Table 6.2 (Exp. G to Exp. O). The template update strategy followed starts by considering only the enrolment signatures (Exp. G) and adds newer signatures to the enroled ones (Exp. G to I), this way the time gap between training and test signatures is reduced, and also as the size of the training data increases, better system performance is expected. Then, when all available signatures are considered, we follow the strategy to remove signatures from older sessions (Exp. J to N), in order to analyse whether these signatures with a large time gap with the test ones can still contribute to obtain optimal system performances or not.

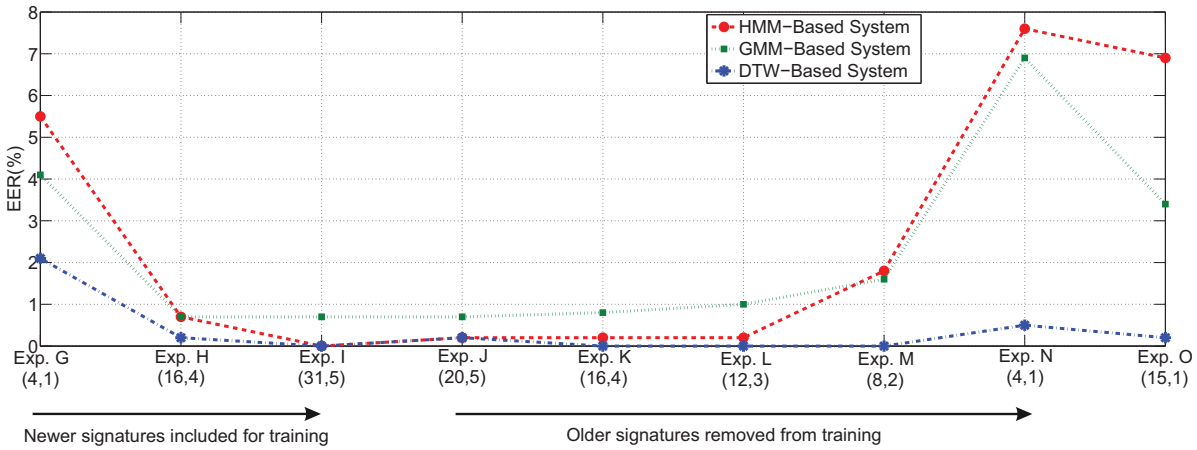
It is worth noting that for HMM and GMM, the optimal system configuration parameters described in Sec. 6.1.2 regarding the number of available training signatures have been taken into account in these experiments in order to study the proposed template update strategies properly.

Fig. 6.4 shows the performance of the three systems for all the experiments considered in this section, obtaining this way a global figure to analyse the different possibilities for template update. Results are obtained for both skilled and random forgeries.

Analysing HMM and GMM systems for both skilled and random forgery cases, the performance improves when increasing the size of the templates with newer signatures (from Exp. G to Exp. I). Then, when we remove the older signatures (from Exp. J to Exp. N), the EER increases slowly achieving significantly worse results for Exp. N. Therefore for both HMM and GMM the best performance is achieved for Exp. I with 4.8% and 4.1% EER for HMM and GMM systems respectively for skilled forgery cases, and 0.01% and 0.7% EER respectively for random forgery cases. However, in Exp. I, we are considering 15 signatures from Session S5, which is very unlikely that this happens in a realistic scenario. Therefore, we consider the case of Exp. J, where only four signatures from S5 are used (S.5.2). In this experiment (Exp. J), a few training signatures from different sessions are considered, achieving very similar results compared to Exp. I. Anyway, the process of model training for both HMM and GMM is performed off-line, so the score computation time would not be very affected in the case of having a much larger database with higher number of sessions and signatures.



(a) Skilled forgery cases



(b) Random forgery cases

Figure 6.4: Template Update Strategies. *Below each experiment in brackets the first number indicates the number of training signatures, and the second the number of sessions they come from. Exp. G corresponds to using 4 training signatures from the enrolment session. From Exp. G to Exp. I we add training signatures from more recent sessions. Exp. J has 4 training signatures from each of the 5 sessions. Then, from Exp. J to Exp. N, we remove signatures from older sessions. Exp. O is included for completeness and contains 15 signatures from the closest session to the test.*

Another finding worth highlighting is that it is better to build the user's template considering training signatures coming from different sessions (i.e., Exp. K) instead of using all of them from only one session closer in time (i.e., Exp. O), as there is a significant worsening of performance in this last case. Therefore, the best strategy for both HMM and GMM systems for template update would be to take into account all available training signatures or at least a few training signatures but from several sessions in order to generate a more reliable user's template. This conclusion agrees with the results obtained in [Sae-Bae and Memon, 2014], in which the on-line signature verification system further improved for an increasing number of training signatures.

On the other hand, the optimal template update strategy for the DTW system is different compared to the HMM and GMM as it can be seen in Fig. 6.4. Analysing the performance of DTW for both skilled and random forgery cases, the best configurations correspond to Exp. I, L and M with 5.5% EER for skilled forgery cases and 0.01% EER for random forgeries. The first

Table 6.3: *Comparison of the system performance in terms of EER(%) for Baseline, and Proposed Systems. S stands for Skilled forgeries and R for Random forgeries.*

ATVS Signature Long-Term Extended DB						
	HMM		GMM		DTW	
	S	R	S	R	S	R
Baseline	13.1	5.5	13.8	4.1	11.0	2.1
Proposed	4.1	0.2	4.8	0.7	5.5	0.01
BiosecurID DB (+10,000 signatures from 371 users)						
Baseline	10.0	3.8	11.1	4.1	5.8	0.7
Proposed	5.9	2.9	6.4	1.7	3.6	0.2

case considers all available training signatures, but the other two just consider 12 and 8 training signatures respectively from sessions closer to the test. In this case the trends would suggest to choose Exp. L and M as the EER increases slowly when we add older training signatures (i.e., Exp. J and K). It is important to highlight that DTW-based systems carry out one to one comparisons of the signatures, so as the number of training signatures increases, the number of DTW comparisons also increases. Thus, it is necessary to establish a limit of comparisons in order to make this system feasible for real time scenarios. As a conclusion for the DTW-based system, the optimal template update strategy would be to consider a few training signatures (i.e., between 8 and 12) from the last sessions closer in time to the test in order to achieve both optimal performance and feasible computation cost.

Finally, in order to quantify the reduction of the aging effect achieved, our proposed template update strategy is evaluated on two different databases: *i)* the ATVS Signature Long-Term Extended database presented in this study, acquired in a 15-month total time span, and *ii)* the remaining 371 users of the BiosecurID database [Fierrez *et al.*, 2010], which is composed of four different acquisition sessions (from S1 to S4 in Fig. 3.5) with a total 6-month time span. It is important to remark that the 371 users of the BiosecurID database have not been used during the analysis of our proposed template update strategy. Table 6.3 shows the performance of our Proposed Systems incorporating template update strategies for both ATVS Signature Long-Term Extended database and BiosecurID database, respectively. We also include the performance of the traditional case in signature verification (i.e., Baseline) just using the enrolment data from the first acquisition session (S1 in Fig. 3.5), which would be the case where the aging effect is more pronounced.

Analysing in Table 6.3 (top) the results obtained for the ATVS Signature Long-Term Extended database, the systems proposed in this work achieve a significant improvement of performance, hence a significant reduction of the template aging effect with an average relative improvement in comparison to the baseline system of 62.0% and 92.2% EER for skilled and random forgeries, respectively.

Results on the unseen users of the BiosecurID database. Results are depicted in Table 6.3 (bottom). Regarding the experimental protocol, the 4 genuine signatures from the last session (i.e. S4) are always used as test signatures. For the Baseline system, we just use the 4 genuine

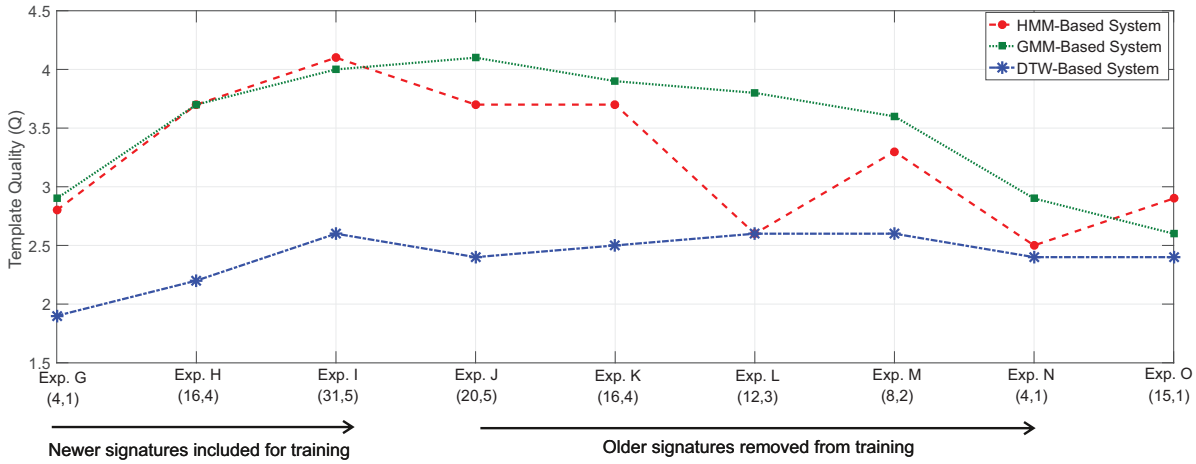


Figure 6.5: Statistical Analysis. *The template quality metric Q is computed for all experiments from Sect. IV.C.2. Below each experiment in brackets the first number indicates the number of training signatures, and the second the number of sessions they come from. Exp. G corresponds to using 4 training signatures from the enrolment session. From Exp. G to Exp. I we add for training signatures from more recent sessions. Exp. J has 4 training signatures from each of the 5 sessions. Then from Exp. J to Exp. N we remove signatures from older sessions. Exp. O is included for completeness and contains 15 signatures from the closest session to the test.*

signatures from the enrolment session (i.e. S1) for training. However, following our proposed template update strategy, for the HMM and GMM proposed systems we would select all available genuine signatures from S1 to S3 (i.e. 12) for building the user models whereas for the DTW system we would select up to 12 genuine signatures from the last sessions in time, i.e., all available genuine signatures from S1 to S3 as well. Our proposed template update strategy has proved to be very effective against the aging effect, achieving an average relative improvement in comparison to the baseline system of 40.9% and 44.2% EER for skilled and random forgeries, respectively.

6.2.3.3. Statistical Analysis

Here we explore the template quality metric Q defined in Sec. 6.1.3. It is important to highlight that only random forgeries are considered in this statistical analysis due to the lack of skilled forgeries in real scenarios. Fig. 6.5 shows the template quality Q of the three systems for the same experiments considered in Sec. 6.2.3.2, obtaining this way a global figure to support the conclusions extracted in Sec. 6.2.3.2. It is important to remark that the higher the Q value is, the better the template update strategy will be.

Analysing the HMM and GMM systems, the trend of the Q value is to increase with the number of training samples (from Exp. G to Exp. I) and then to decrease when we remove the older signatures (from Exp. J to Exp. N). These results make sense as both HMM and GMM are statistical approaches and they are able to better model the intra-user variability when increasing the number of training signatures and sessions. Therefore, for both HMM and GMM systems the highest value of Q is obtained for Exp. I and J respectively when

training signatures from 5 different sessions are considered. This statistical analysis agrees with the template update strategies proposed in Sec. 6.2.3.2 where the best system performance is obtained when all available training signatures or at least a few training signatures from several sessions are considered.

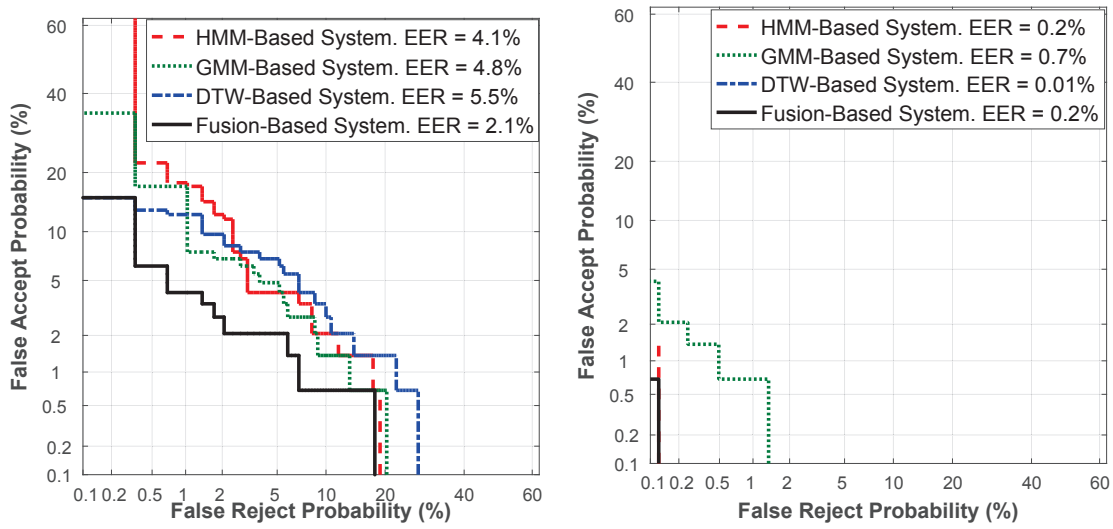
Analysing the DTW system, the best value of Q is obtained for Exp. I, L and M. In this case the trend would suggest to choose Exp. L or M as the best template update strategies as the DTW system carries out one to one comparisons of the signatures and the larger the number of training signatures is, the higher the computational cost. These statistical results also support the template update strategies proposed in Sec. 6.2.3.2 for the DTW-based system being the best approach to select a few training signatures (i.e., between 8 and 12) from the last sessions closer in time to the test.

In summary, this statistical analysis based on a template quality metric agrees with the results achieved in the previous sections.

6.2.3.4. Fusion of the Proposed Systems

Fusion of biometric systems has been considered in many different related works [Alonso-Fernandez *et al.*, 2010; Kittler *et al.*, 1998] as an easy and reliable way of achieving a further system performance improvement. In this section, the main goal is to carry out the fusion of the three systems studied in order to achieve an improvement of recognition performance and to reduce the template aging effect even further, especially for the challenging case of skilled forgeries. The final fusion is carried out at the score level with the sum rule after normalising the scores from the three optimal systems to a common range as described in Sec. 6.2.1. For both HMM and GMM systems, Exp. J has been selected as the optimal template update strategy as it achieves good performance in a realistic set up. In this case, there is a total of 20 training signatures coming from five different sessions. The optimal parameters for the HMM system are $N = 32$ and $M = 2$ whereas for the GMM system, $M = 128$. For the case of the DTW-based system, the optimal template update strategy considered corresponds to Exp. L, which uses a total of 12 training signatures from the last three sessions closer in time to the test. In this case, the final score is the average of all comparisons. The performance of the three systems and the fusion of all of them is represented using DET plots in Fig. 6.6.

As shown in the figures, the proposed fusion achieves a significant improvement of performance, especially for skilled forgery cases. In this case, the performance of the Fusion System achieves a significant absolute improvement of 2.0% EER compared to the best individual system whereas for the random forgery cases, the proposed fusion does not improve the best system (DTW), which resulted to be almost perfect (i.e., 0.01%). In this case, as the systems being combined behave quite differently, a weighted sum would be more adequate. Anyway, the fused performance is still very competitive (0.2% EER).



(a) Skilled forgery cases

(b) Random forgery cases

Figure 6.6: Fusion of the Proposed Systems. *DET curves for the three optimal systems after applying the proposed template update approach and fusion of all of them via sum rule of scores.*

6.3. Chapter Summary and Conclusions

This chapter reports the first significant experimental results regarding the effect of template aging and template update strategies for on-line signature authentication considering both random and skilled forgery cases. For this, we have created an extension of the ATVS On-Line Signature Long-Term database, in which skilled forgeries are included. The complete signature database is publicly available at `https://github.com/BiDAlab/xLongSignDB`.

Experiments have been carried out using three well-known systems based on local features: HMM, GMM, and DTW. First, the effect of template aging in on-line signatures has been analysed, concluding that it has a significant impact in the system performance. In order to compensate for this aging effect, an exhaustive experimental analysis of various template update strategies has been carried out. For the case of HMM and GMM systems the optimal template update strategy would be to select all available training signatures or at least a few of them from several sessions in order to generate a more reliable user's template. For the DTW system the optimal would be to consider a few training signatures (i.e., between 8 and 12) from sessions closer in time to the test. By incorporating the considered template update techniques, we have demonstrated a significant improvement of performance of the three baseline systems, hence a significant reduction of the template aging effect with similar results to the ideal case for random forgeries, and an average relative improvement of 61.9% EER for skilled forgeries.

Finally, a fusion of the three individual systems after applying the best resulting template update approach has been carried out in order to further improve the recognition performance achieving an EER of 2.1% and 0.2% for skilled and random forgeries respectively.

Part III

Towards the Near Future

Chapter 7

Deep Learning

In this chapter we evaluate the potential of our proposed RNN on-line signature verification systems described in Chapter 4.2. A thorough analysis has been carried out considering multiple RNN approaches and architectures. In addition, different types of impostors (i.e., skilled and random forgeries) have been considered in our study.

This chapter is structured as follows. Sec. 7.1 summarises our proposed RNN signature verification system. Sec. 7.2 describes the experimental protocol considered in this chapter. The results achieved are presented in Sec. 7.3. Sec. 7.4 aims to provide the last new advancements obtained in this thriving topic. Finally, the conclusions of this chapter are summarised in Sec. 7.5.

This chapter is based on the following publications: [Tolosana *et al.*, 2019a, 2017b, 2018c].

7.1. Proposed RNN On-Line Signature Verification Systems

This section summarises our proposed end-to-end writer-independent on-line signature verification system. For a complete understanding of our proposed approach, we recommend the reader to see Sec. 4.2. Fig. 7.1 graphically summarises our proposed end-to-end signature verification approach. This system has been obtained after carrying out an exhaustive analysis in terms of the number of time functions used to feed the network and the complexity level of the RNN system (i.e., the number of hidden layers and memory blocks per hidden layer). All details are described in this experimental chapter.

We propose both LSTM and GRU systems with a Siamese architecture. In addition, a bidirectional scheme is considered for both LSTM- and GRU-based systems in order to be able to access both past and future context. For the input of the RNN system, we extract a set of 23 local features (Table 4.3) per signature from signals related to X and Y spatial coordinates and pressure. The first layer is composed of two LSTM/GRU hidden layers with 46 memory blocks each, sharing the weights between them. The outputs of the first two parallel LSTM/GRU hidden layers are concatenated and serve as input to the second layer, which corresponds to a LSTM/GRU hidden layer with 23 memory blocks. Finally, a feed-forward neural network layer

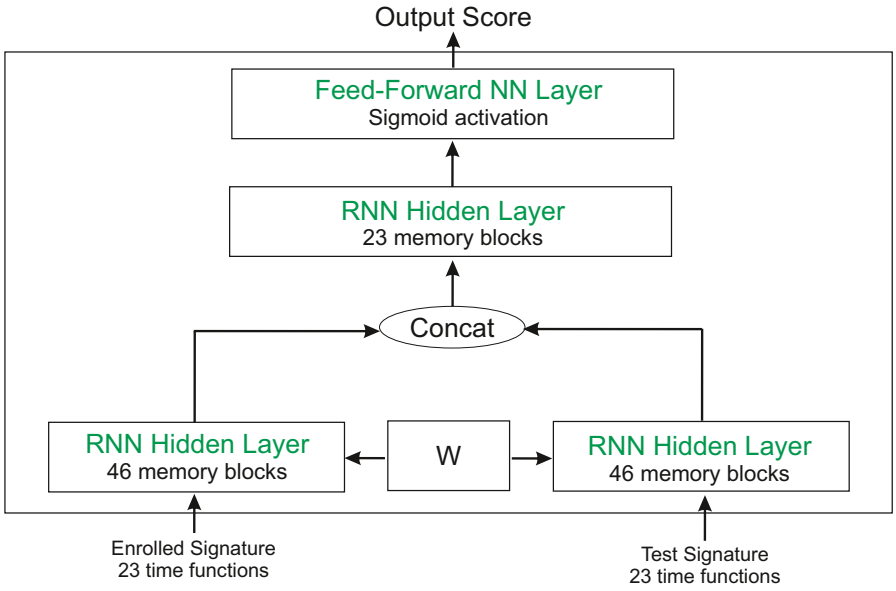


Figure 7.1: *End-to-end writer-independent on-line signature verification system proposed in this Thesis based on the use of LSTM and GRU RNNs with a Siamese architecture.*

with a sigmoid activation is considered, providing an output score between 0 and 1 for each pair of signatures.

7.2. Experimental Protocol

The experimental protocol has been designed in order to analyse and prove the feasibility of both LSTM and GRU RNNs for on-line signature verification in practical scenarios. Therefore, different users and signatures are considered for the two main stages, i.e., development of the RNNs system (Sec. 7.3.1) and the final evaluation of them (Sec. 7.3.2). Additionally, the two most common types of impostors are considered here: skilled and random forgeries.

The first 300 users of the BiosecurID database are used for the development of the system, while the remaining 100 users are considered for the evaluation. For both stages, the 4 genuine signatures of the first session are used as training signatures, whereas the 12 genuine signatures of the remaining sessions are left for testing. Therefore, inter-session variability is considered in our experiments. Skilled forgery scores are obtained by comparing training signatures against the 12 available skilled forgery signatures for each user whereas random forgery scores are obtained by comparing the training signatures with one genuine signature of 12 other random users.

Finally, three different scenarios are analysed regarding the type of forgery considered for training the RNN systems: *i)* **"skilled"**, the case which considers only pairs of genuine and skilled forgery signatures, *ii)* **"random"**, the case which considers only pairs of genuine and random forgery signatures, and *iii)* **"skilled + random"**, the case which considers pairs of both genuine/skilled and also genuine/random signatures in order to train just one system for both types of forgeries.

7.3. Results

7.3.1. Development Results

This section describes the development and training of our proposed LSTM and GRU RNN systems with a Siamese architecture considering the 300 users of the development dataset. Three different types of pairs of signatures can be used as inputs of the RNN systems: *i)* two genuine signatures performed by the same user, *ii)* one genuine signature from the claimed user and one skilled forgery signature performed by an impostor, and *iii)* one genuine signature from the claimed user and one random forgery signature. For each of these three cases there are a total of $4 \times 12 \times 300 = 14,400$ comparisons, having the same number of genuine and impostor signatures for testing. Our RNN systems are implemented under Theano [Bastien *et al.*, 2012] with a NVIDIA GeForce GTX 1080 GPU.

In order to find the most suitable RNN system architecture we explored different configurations regarding the number of local features used as inputs and the complexity level of the RNN system (i.e., number of hidden layers and memory blocks per hidden layer). In all cases, we considered our proposed Siamese architecture in order to learn a dissimilarity from pair of signatures. Our first attempt was based on the use of the 11 most commonly used local features from a total of 23 (i.e., x_n, y_n, z_n, θ_n, v_n, ρ_n, a_n, $\dot{x_n}$, $\dot{y_n}$, $\ddot{x_n}$, and $\ddot{y_n}$) with a RNN system based on two RNN hidden layers (with 22 and 11 memory blocks, respectively), and finally a feed-forward neural network layer with a sigmoid activation. Both input-to-hidden and hidden-to-hidden layers are fully-connected. The initial system performance obtained with this configuration over the evaluation dataset was 8.25% EER. Then, we decided to increase the complexity of the RNN system in order to achieve better results over the evaluation dataset. First, we added a new RNN layer composed of 6 memory blocks on top of the second RNN layer providing a 20.00% EER over the evaluation dataset, so this configuration was discarded. Another approach was based on the use of the original configuration based on two RNN hidden layers but increasing the number of memory blocks (44 and 22 per RNN hidden layer, respectively) achieving a final 10.00% EER, being this result worse compared to the 8.25% EER of the original configuration. We concluded that increasing the complexity of the RNN system always ended up with a worse generalization over the evaluation dataset (i.e., overfitting). Then we decided to feed the RNN system with as much information as possible and let the network to select the most important information for the task, i.e., all 23 available local features described in Table 4.3.

After repeating the same previous exploration, the best topology obtained for both LSTM and GRU proposed RNNs is based on the use of two RNN hidden layers, and finally a feed-forward neural network layer with a sigmoid activation. This is the final architecture of our proposed system described in Sec. 4.2.2. The first layer is composed of two LSTM/GRU hidden layers with 46 memory blocks each and sharing the weights between them. The outputs provided for each LSTM/GRU hidden layer of the first layer are then concatenated and serve as input to the second layer which corresponds to a LSTM/GRU hidden layer with 23 memory blocks.

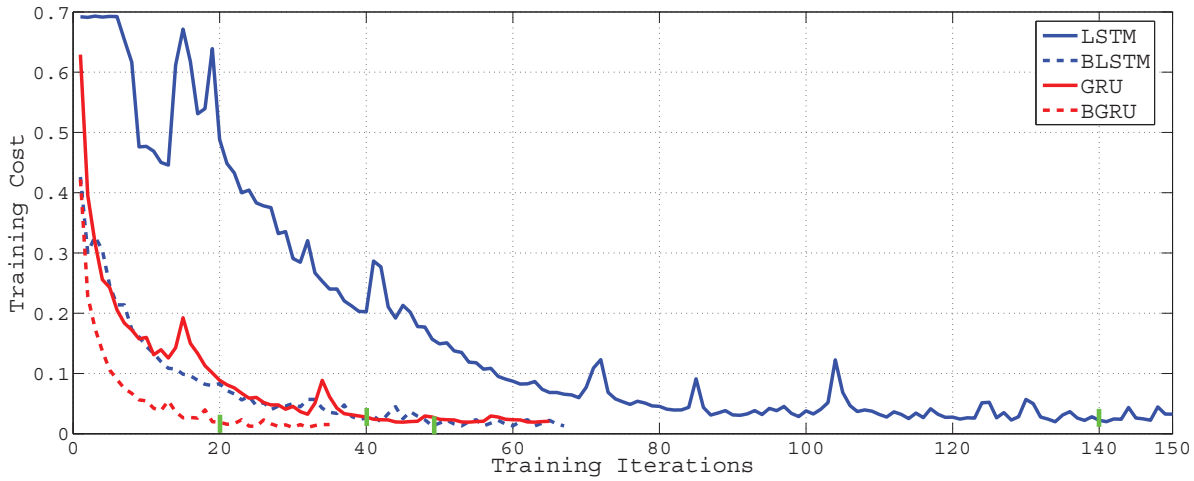


Figure 7.2: *Considered RNNs cost during training for the "skilled" scenario. A small green vertical line indicates for each proposed RNN system the training iteration which provides the best system performance over the evaluation dataset.*

Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score for each pair of signatures. Fig. 4.7 visually represents the architecture of our proposed end-to-end on-line signature verification system.

Fig. 7.2 shows the training cost of the considered RNNs with the number of training iterations for the "skilled" scenario. Four different RNN-based systems are considered, i.e., LSTM, GRU and their bidirectional schemes (i.e., BLSTM and BGRU). A small green vertical line is included in the figure for each proposed RNN system indicating the training iteration which provides the best system performance over the evaluation dataset, with a training cost value very close to zero. Similar results were obtained for both "random" and "skilled + random" scenarios as well. It is important to remark two different aspects of the figure. First, the difference in the number of training iterations needed between normal and bidirectional schemes. For example, the best LSTM configuration is obtained after 140 training iterations whereas only around 50 iterations are needed for the BLSTM RNN system. This shows the importance of considering both past and future contexts in order to train RNNs faster and also with a lower value of training cost. Additionally, it is important to highlight the difference in the number of training iterations between both LSTM and GRU RNN systems. As the GRU memory block is a simplified version of the LSTM memory block (see Sec. 4.2.1.3) the number of parameters to train are lower and therefore, we are able to get similar and even better values of training cost with fewer number of training iterations compared to the LSTM RNN system.

7.3.2. Evaluation Results

This section analyses the performance of the proposed RNN systems trained in the previous section for the three different training scenarios considered (i.e., "skilled", "random" and "skilled + random"). The remaining 100 users (not used for development) are considered here. Regarding the system performance, two different cases are considered. First, the evaluation of

Table 7.1: 1vs1 Evaluation Results: *System performance in terms of EER(%) for the three different training scenarios considered, i.e., "skilled", "random" and "skilled + random".*

	Train: "skilled"		Train: "random"		Train: "skilled + random"	
	Skilled	Random	Skilled	Random	Skilled	Random
LSTM	6.44	24.48	13.31	5.38	7.94	6.22
GRU	7.69	29.42	15.63	6.92	7.67	5.98
BLSTM	**5.60**	24.48	15.31	**5.28**	**6.83**	**5.38**
BGRU	6.31	19.14	12.56	5.33	7.88	5.52

Table 7.2: 4vs1 Evaluation Results: *System performance in terms of EER(%) for the three different training scenarios considered, i.e., "skilled", "random" and "skilled + random".*

	Train: "skilled"		Train: "random"		Train: "skilled + random"	
	Skilled	Random	Skilled	Random	Skilled	Random
LSTM	5.58	24.03	15.17	4.08	6.17	3.67
GRU	6.25	28.69	13.92	4.25	5.58	3.63
BLSTM	**4.75**	24.03	15.58	3.89	**5.50**	3.00
BGRU	4.92	19.69	12.33	**3.25**	5.92	**2.92**

Table 7.3: 1vs1 and 4vs1 DTW-based Evaluation Results: *System performance in terms of EER(%).*

	1vs1	4vs1
Skilled	10.17	7.75
Random	0.94	0.50

the system performance considering scores directly from all pairs of signatures (i.e., 1vs1) and second, the case of performing the average score of the four one-to-one comparisons (i.e., 4vs1) as there are four genuine training signatures per user. In order to make comparable our approach to related works, we have considered a highly competitive system based on the popular DTW approach [Gomez-Barrero *et al.*, 2015] with a total of 9 out of 27 different local features selected using the SFFS algorithm.

Tables 7.1 and 7.2 show the system performance in terms of EER(%) for our Proposed RNN-based Systems for both 1vs1 and 4vs1 cases, respectively. In addition, Table 7.3 shows the system performance in terms of EER(%) for the DTW-based System [Gomez-Barrero *et al.*, 2015] for both 1vs1 and 4vs1 cases, over the same evaluation set of Tables 7.1 and 7.2. We now analyse the results obtained for each of the three different training scenarios considered.

Skilled training scenario: First, we analyse in Tables 7.1 and 7.2 the case in which only pairs of genuine and skilled forgery signatures are used for the development of the systems (i.e., "skilled"). Overall, very good results have been obtained for all Proposed Systems when skilled forgeries are considered. Bidirectional schemes (i.e., BLSTM and BGRU) have outperformed normal schemes, highlighting the importance of considering both past and future contexts. In addition, both LSTM and GRU RNN systems have achieved very similar results proving their feasibility for handwritten signature verification. Analysing the results obtained in Tables 7.1 and 7.3 for the 1vs1 case, our Proposed BLSTM System has achieved the best results with a 5.60% EER, which corresponds to an absolute improvement of 4.57% EER compared to the 10.17% EER achieved for the DTW-based System. This result (i.e., 5.60% EER) outperforms

related state-of-the-art results for the case of considering just one signature for training [Diaz *et al.*, 2016b]. Analysing the results obtained in Tables 7.2 and 7.3 for the 4vs1 case, our Proposed BLSTM System achieves a 4.75% EER, which corresponds to an absolute improvement of 3.00% EER compared to the 7.75% EER achieved for the DTW-based System. Moreover, it is worth noting that the result obtained with our Proposed BLSTM System for the case of using just one training signature (1vs1) outperforms the result obtained with the DTW-based System (i.e., 5.60% vs 7.75% EER) for the 4vs1 case. Additionally, our Proposed BLSTM System outperforms other state-of-the-art signature verification systems such as the one proposed in [Galbally *et al.*, 2015] based on fusion of a local system with DTW algorithm and a global system with Mahalanobis distance (i.e., 4.75% vs 4.91% EER) for the case of considering 4 training signatures. These results show the high ability of our proposed approach for learning even with small amounts of signatures. However, the results obtained in Tables 7.1 and 7.2 for our Proposed RNN Systems when random forgeries are considered are far away from the state-of-the-art results. The best result has been obtained using our Proposed BGRU System with a value of 19.14% EER whereas a 0.50% EER is obtained in Table 7.3 for the DTW-based System. These bad results obtained for the random forgery case make sense as only skilled and not random forgeries were used for training the RNNs.

Random training scenario: In order to see the ability of the RNN systems to detect different types of forgeries, Tables 7.1 and 7.2 also show the system performance in terms of EER(%) for the scenario in which our Proposed RNN Systems are trained using only pairs of genuine and random forgery signatures (i.e., "random"). Overall, a high improvement of the system performance is achieved for the case of random forgeries compared to the results previously analysed in the "skilled" training scenario. The best result corresponds to our Proposed BGRU System with a 3.25% EER. However, as it happened for the "skilled" training scenario previously commented, bad results are achieved for the task in which the RNN system is not trained (i.e., skilled forgeries in this "random" training scenario).

Skilled+random training scenario: Finally, Tables 7.1 and 7.2 show the system performance in terms of EER(%) for the case in which our Proposed RNN Systems are trained using pairs of genuine and skilled forgery signatures and also pairs of genuine and random forgery signatures (i.e., "skilled + random"). Analysing the results obtained for skilled forgeries, the best system performance has been obtained using our Proposed BLSTM System with a value of 5.50% EER. Moreover, the result obtained with our Proposed BLTM System for the case of using just one training signature (1vs1) still outperforms the result obtained with the DTW-based System for the 4vs1 case (i.e., 6.83% vs 7.75% EER), showing the high ability of our proposed approach for learning even with small amounts of signatures. Analysing the results obtained for random forgeries, our Proposed BLSTM System has achieved a 3.00% EER. These results prove the ability of RNN-based systems to detect two different types of forgeries using just one system. Despite of the high improvements achieved when both skilled and random forgeries are used for training the RNNs, the 3.00% EER obtained using our Proposed BLSTM System can not outperform the 0.5% EER obtained using the DTW-based System against random forg-

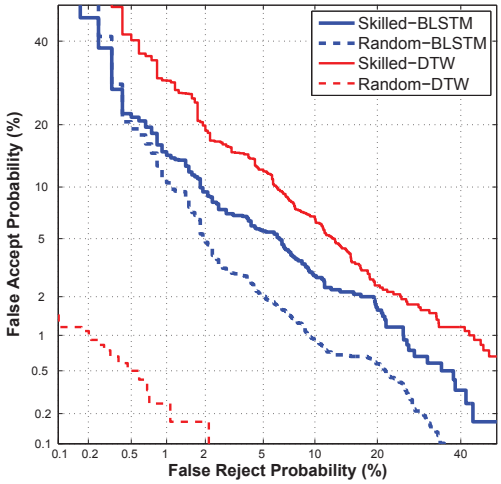


Figure 7.3: *System performance results obtained using our Proposed BLSTM System for the 4vs1 case and "skilled + random" training scenario over the BiosecurID evaluation dataset.*

eries. Fig. 7.3 shows the DET curve of both Proposed BLSTM and DTW-based Systems for the 4vs1 case and "skilled + random" training scenario for completeness. In order to achieve state-of-the-art results for both skilled and random forgeries, a possible solution is to perform two consecutive stages similar to [Gomez-Barrero *et al.*, 2015]: 1) first stage based on DTW optimised for rejecting random forgeries, and 2) our Proposed RNN Systems in order to reject the remaining skilled forgeries. Another recent example of multiple classifier contribution for signature is [Tolosana *et al.*, 2015d].

7.4. New Advancements

This final section aims to provide the last new advancements obtained in this thriving topic. So far this chapter, only the 300 first users of the BiosecurID database were considered for training our DL models. In order to analyse the potential of DL technology when having more available training data, we have created the novel DeepSignDB on-line handwritten signature database. Fig. 7.4 graphically summarises the design, acquisition devices, and writing tools considered in the DeepSignDB database. This database is obtained through the combination of some of the most well-known databases, and a novel dataset not presented yet. It comprises more than 70K signatures acquired using both stylus and finger inputs from a total 1526 users. Two acquisition scenarios are considered, office and mobile, with a total of 8 different devices. Additionally, different types of impostors and number of acquisition sessions are considered along the database.

Regarding the experimental protocol, the DeepSignDB database has been divided into two different datasets, one for the development and training of the system and the other one for the final evaluation. The development dataset comprises around 70% of the users of each database

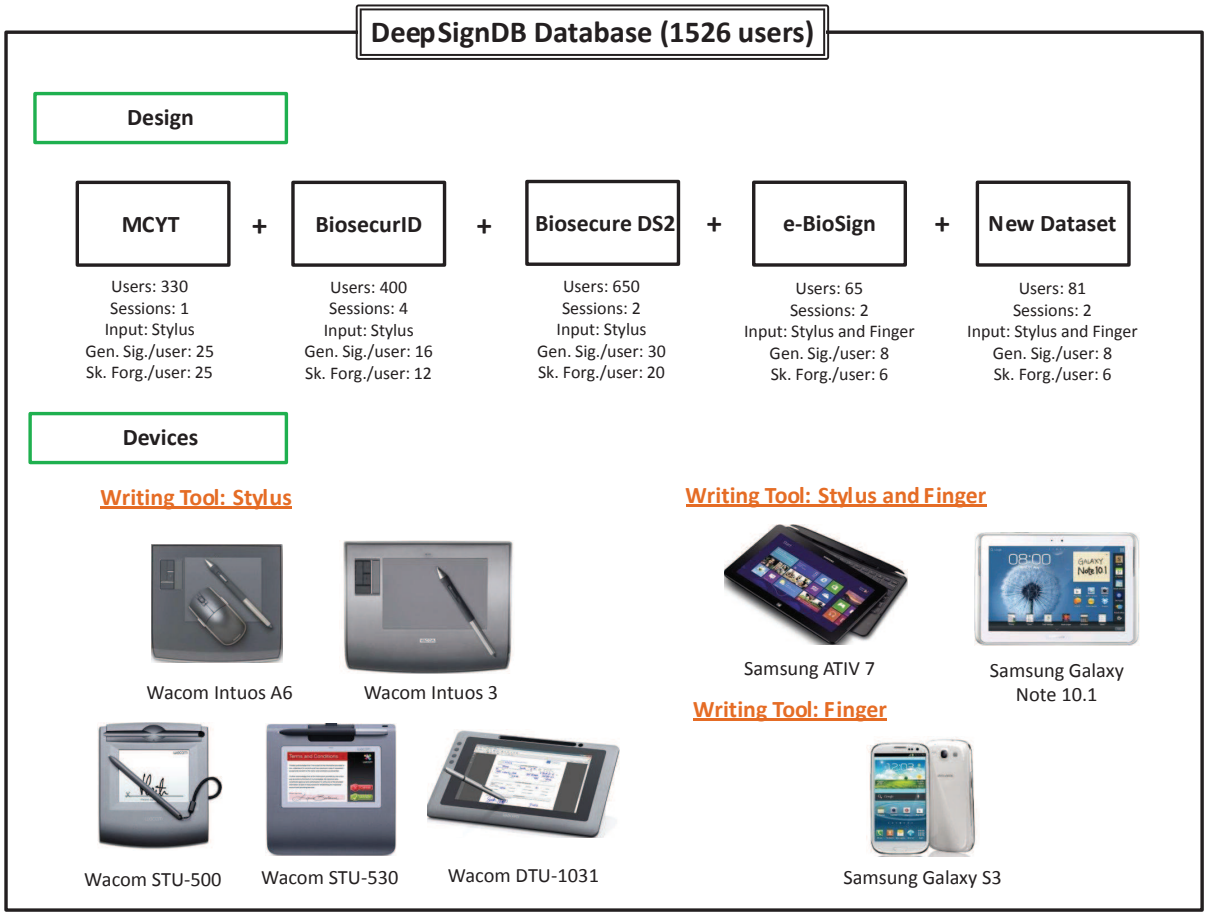


Figure 7.4: *Description of the design, acquisition devices, and writing tools considered in the new DeepSignDB database. A total of 1526 users and 8 different captured devices are used (5 Wacom and 3 Samsung general purpose devices). For the Samsung devices, signatures are also collected using the finger. Gen. Sig. = Genuine Signatures, and Sk. Forg. = Skilled Forgeries.*

whereas the remaining 30% is included in the evaluation dataset. Therefore, we use 1084 users in total for training the RNN systems, 3.6 times more users compared to the 300 initial users considered in the BiosecurID database. Signatures acquired using the stylus are only considered in this preliminary study, ending up with around 309K genuine and impostor comparisons (247K and 62K for training and validation, respectively). It is important to remark: *i)* the same number of genuine and impostor comparisons are used to train the networks in order to avoid bias, and *ii)* only skilled forgeries are used as impostors (the DTW is in charge of detecting the random forgeries as explained at the end of the previous section).

Table 7.4 describes the system performance results over the same BiosecurID evaluation dataset considered in the previous sections. Two different approaches are considered: *i)* training the networks using only the first 300 users of the BiosecurID development dataset, and *ii)* training the networks using the 1084 users of the DeepSignDB development dataset. The same BGRU RNN system described in the previous sections is considered in this analysis. DTW system is also included for completeness. Therefore, Table 7.4 intends to compare the improvements

Table 7.4: *System performance results over the BiosecurID evaluation dataset.*

	1vs1	4vs1
DTW	10.17	7.75
BGRU (300 users)	5.60	4.75
BGRU (1084 users)	3.90	3.40

achieved when more data is available to train the neural network models.

Analysing the 1vs1 case, the best system performance result is obtained using the BGRU system trained with the 1084 users of the DeepSignDB development dataset. This system achieves an absolute improvement of 1.7% EER compared to the same BGRU system but trained with only the 300 users of the BiosecurID development dataset. This improvement is much higher if we compare to the DTW system, achieving an absolute improvement of 6.27% EER.

Analysing the 4vs1 case, similar conclusions are obtained compared to the 1vs1 case. The BGRU system trained using the 1084 users achieves an absolute improvement of 1.35% and 4.35% EER for the BGRU system trained using the 300 users and the DTW system, respectively.

Finally, it is also interesting to remark that even for the case of using just a single training signature per user (1vs1), the BGRU system trained using the 1084 users achieves almost the double system performance improvement compared to the DTW system for the case of using 4 training signatures per user (4vs1). These results prove how importance the amount of data is for training more robust neural network models.

The DeepSignDB database, experimental protocol proposal and benchmark evaluation of it has been submitted to the International Conference on Document Analysis and Recognition (ICDAR) 2019. All this information (included the database) will be available in GitHub very soon. Finally, we would also like to highlight that the application of DeepSignDB extends from the improvement of signature verification systems via deep learning to many other potential research lines, e.g., studying: *i)* user-dependent effects, and development of user-dependent methods in signature biometrics, and handwriting recognition at large [Yager and Dunstone, 2010b], *ii)* the neuromotor processes involved in signature biometrics, and handwriting in general [Ferrer *et al.*, 2018], *iii)* sensing factors in obtaining representative and clean handwriting and touch interaction signals [Tolosana *et al.*, 2015d], *iv)* human-device interaction factors involving handwriting and touchscreen signals, and development of improved interaction methods [Harbach *et al.*, 2016], and *v)* population statistics around handwriting and touch interaction signals, and development of new methods aimed at recognising or serving particular population groups.

7.5. Chapter Summary and Conclusions

In this chapter we have assessed the feasibility of different RNN systems in combination with a Siamese architecture [Chopra *et al.*, 2005] for the task of on-line handwritten signature verification. As far as we know, this study has provided the first complete and successful framework on the use of multiple RNN systems (i.e., LSTM and GRU) for on-line handwritten signature

verification considering both skilled and random forgery cases. The BiosecurID database composed of 400 users and 4 separated acquisition sessions has been considered in the experimental work, using the first 300 users for development and the remaining 100 users for evaluation. Three different impostor scenarios are considered for training the RNN systems (i.e., "skilled", "random", "skilled + random"). Additionally, we have considered two different cases regarding the number of available training signatures per user. First, the evaluation of the system performance considering scores directly from all pairs of signatures (i.e., 1vs1) and second, the case of performing the average score of the four one-to-one comparisons (i.e., 4vs1) as there are 4 genuine training signatures per user (from the first session).

Regarding the development of our Proposed RNN Systems, it is important to remark the different number of training iterations needed between normal (i.e., LSTM and GRU) and bidirectional schemes (i.e., BLSTM and BGRU). This shows the importance of considering both past and future contexts in order to train RNNs faster and also with a lower value of training cost. In addition, it is important to highlight the different number of training iterations between both LSTM and GRU RNNs as the GRU memory block is a simplified version of the LSTM memory block with fewer parameters to train.

Analysing the results obtained using the 100 users of the evaluation dataset, our Proposed BLSTM System has achieved for the "skilled + random" train scenario and 4vs1 case values of 5.50% and 3.00% EER for skilled and random forgeries, respectively. Moreover, the result obtained with our Proposed BLSTM System for the case of using just one training signature (1vs1) still outperforms the result obtained with the highly competitive system based on the popular DTW approach for the 4vs1 case (i.e., 6.83% vs 7.75% EER), showing the high ability of our proposed approach for learning even with small amounts of signatures. It is important to highlight the results obtained in this work compared to the ones obtained by Otte *et al.* in [Otte *et al.*, 2014] where all experiments failed obtaining for the best case a 23.75% EER as systems were based on standard LSTM architectures.

Finally, the preliminary results obtained here using the novel DeepSignDB database prove how importance the amount of data is for training more robust neural network models, as an absolute improvement of 1.7% EER has been obtained when training the models with 1084 total users instead of the 300 initial users of the BiosecurID development dataset.

Chapter 8

Signature Complexity

In this chapter we propose an on-line signature verification system adapted to the signature complexity level of the user. Despite all the studies performed in the on-line signature trait, none of them have exploited, as far as we know, the complexity concept for the development of more robust and accurate on-line signature verification systems. This chapter further investigates this line considering both stylus- and finger-based scenarios.

The chapter is structured as follows. Sec. 8.1 introduces our proposed complexity-based on-line signature verification system. Then, Sec. 8.2 presents the experimental protocol followed in this chapter. The results achieved are described in Sec. 8.3. Conclusions are finally drawn in Sec. 8.4.

This chapter is based on the following publications: [Tolosana *et al.*, 2017c; Vera-Rodriguez *et al.*, 2019, 2018].

8.1. Proposed Approach

Our proposed complexity-based signature verification system, which is composed of two main modules, is depicted in Fig. 8.1. The first module is the signature complexity detector, which considers as features the number of lognormals from the Sigma LogNormal writing generation model. For the second one, we propose a separate local feature extraction module adapted to each signature complexity level. Both modules are further described in Sec. 8.1.1 and 8.1.2.

8.1.1. Signature Complexity Detector

We propose a signature complexity detector based on the number of lognormals extracted from the Sigma LogNormal writing generation model, which was first introduced to on-line signature in [Reilly and Plamondon, 2009], and it has been widely used in many different tasks such as signature verification [Fischer and Plamondon, 2017; Gomez-Barrero *et al.*, 2015], recovering on-line signatures from image-based specimens [Diaz *et al.*, 2017b] and to monitor a range of neuromuscular diseases [Impedovo *et al.*, 2013; Stefano *et al.*, 2017], among many others.

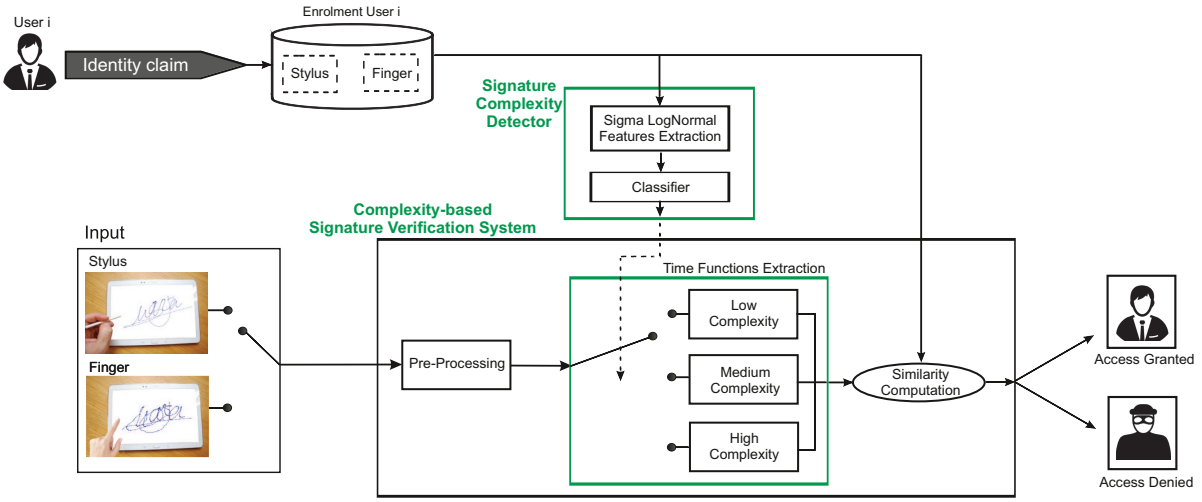


Figure 8.1: *Architecture of our proposed methodology focused on the development of an on-line signature verification system adapted to the signature complexity level. The proposed approach is analysed for the stylus, finger and mixed writing-input scenarios considering e-BioSign and BiosecurID databases.*

The model emulates the physiological human movement production for the generation of signatures. The idea is based on the fact that one signature can be decomposed into strokes in which each stroke i follows a lognormal velocity distribution $\vec{v_i}(t)$:

$$|v_i(t)| = \frac{D_i}{\sqrt{2\pi\sigma_i(t-t_{0i})}} \exp(-\frac{(\ln(t-t_{0i})-\mu_i)^2}{2\sigma_i^2}) \tag{8.1}$$

where t_{0i} is the starting time of the stroke, D_i its length, μ_i the logtime delay and σ_i the logresponse time. In addition, the angular position of each stroke along a pivot direction is expressed through the start angle θ_s and the end angle θ_e. Thus, each stroke is represented by (D_i, t_{0i}, μ_i, σ_i, θ_{si}, θ_{ei}). The complete velocity profile of one signature can be modelled as a sum of the different individual stroke velocity profiles as:

$$\vec{v}(t) = \sum_{i=1}^{N} \vec{v_i}(t) \tag{8.2}$$

where N represents the number of strokes involved in the generation of a given signature. Fig. 8.2 shows the lognormal velocity profiles extracted for each stroke of one example signature.

We propose to use the number of lognormals (N) that models each signature as a measure of the complexity level of the signature. Once this parameter is extracted for all available enrolment signatures of a particular user, that user is classified into a complexity level using the majority voting algorithm (i.e., the signature complexity level of the majority of the enrolment signatures of that user). At the test stage, we consider the complexity level of the claimed user (see Fig. 8.1). In the case that there is no claimed identity, e.g., in signature identification, the complexity level of the identity being compared with the test signature would be used. The advantage of this approach is that the signature complexity detector can be trained and developed as a previous

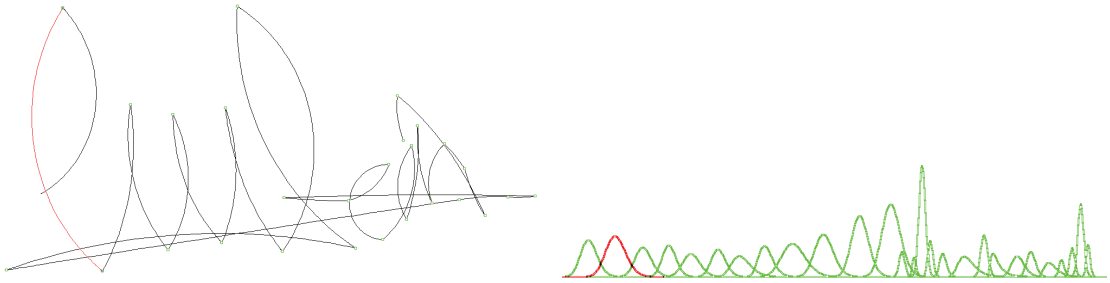


Figure 8.2: *Trace and velocity profile of one reconstructed on-line signature using the Sigma LogNormal model. A single stroke of the signature and its corresponding lognormal profile are highlighted in red colour. Individual strokes are segmented within the LogNormal algorithm [Reilly and Plamondon, 2009].*

off-line process thereby avoiding time consuming delays and making it feasible to be applied at the verification stage in real time scenarios.

8.1.2. Complexity-based Signature Verification System

Once the user is classified into a signature complexity level, we propose to develop a separate local feature extraction module adapted to the signature complexity level as it is depicted in Fig. 8.1.

First, for each signature acquired using the stylus or the finger, signals related to X and Y spatial coordinates are used to extract an initial set of 21 local features (see Table 4.3). In addition, the same two-stage approach proposed in Sec. 5.1 is first considered in order to mitigate the degradation performance on mixed writing-input scenarios.

Second, the SFFS algorithm described in Sec. 2.1.5 is applied here in order to select the optimal local feature subset for each complexity level. This way we can increase the robustness of the features selected and improve the final system performance.

Finally, DTW algorithm is used to compute the similarity between the local features extracted from the query input signature and the training signatures of the claimed user. Therefore, it is important to highlight that for the signature verification stage, the same DTW algorithm is always considered for obtaining the similarity score but different subsets of local features are selected for each complexity level and database.

8.2. Experimental Protocol

The experimental protocol is designed in order to allow the development and evaluation of the following modules: *i)* signature complexity detector, and *ii)* a separate local feature extraction module for each signature complexity level. Both BiosecurID and e-BioSign databases are divided into development (40% of the users) and evaluation (60% of the remaining users) datasets.

For the evaluation of each module, the 4 genuine signatures of the first session of each database are used as training signatures, whereas the remaining genuine signatures (i.e., 4 and

12 for the e-BioSign and BiosecurID databases, respectively) are used for testing. Skilled forgery scores are obtained by comparing the training signatures against the available skilled forgeries for each user (i.e., 6 and 12 for the e-BioSign and BiosecurID databases, respectively) whereas random (zero-effort) forgery scores are obtained by comparing the training signatures with one genuine signature of each of the remaining users. The final score is obtained after performing the average score of the four one-to-one comparisons.

Finally, the following nomenclature is proposed in order to facilitate the readability and understanding of the paper about the different input scenarios considered: "training-testing", where "training" and "testing" mean the writing tool considered for the training and testing signatures, respectively. For example, the case "stylus-finger" means that signatures considered for training are acquired using the stylus whereas signatures considered for testing are acquired using the finger as input.

8.3. Results

8.3.1. Signature Complexity Detector

The signature complexity detector was developed in two different stages. First, each user of the BiosecurID database was manually labelled in one of the signature complexity levels (low, medium, high). This process was carried out by visualising the image of just one genuine signature per user and was performed by two annotators twice each in order to keep consistency on the results. Three different complexity levels were considered based on previous works [Houmani and Garcia-Salicetti, 2016]. Users with signatures with a longer writing time and with an appearance more similar to handwriting were labelled as high-complexity users whereas those users with signatures shorter in time and with generally simple flourish with no legible information were labelled as low-complexity users. This first stage served as a ground truth. Following this stage, the number of lognormals N form the Sigma LogNormal model was extracted from each available genuine signature of the BiosecurID database (i.e., a total of $400 \times 16 = 6400$ genuine signatures). Then, we represented for each complexity level their corresponding distribution of lognormals according to the ground truth performed during the first stage. Fig. 8.3 shows the distributions of the number of lognormals obtained for each complexity level using all genuine signatures of the BiosecurID database. The three proposed complexity-dependent decision thresholds are highlighted by black dashed lines. They were selected in order to minimise the number of misclassifications between different signature complexity levels. Signatures with lognormal values equal or less than 17 are classified as low-complexity signatures whereas those signatures with more than 27 lognormals are classified into the high-complexity group. Otherwise, signatures are categorised into medium-complexity. Additionally, an analysis of the stability regarding the number of lognormals for different signatures of the same user is carried out in order to assess the feasibility of our proposed signature complexity detector. In general, low standard deviation values are obtained. Users with a low signature complexity level provide

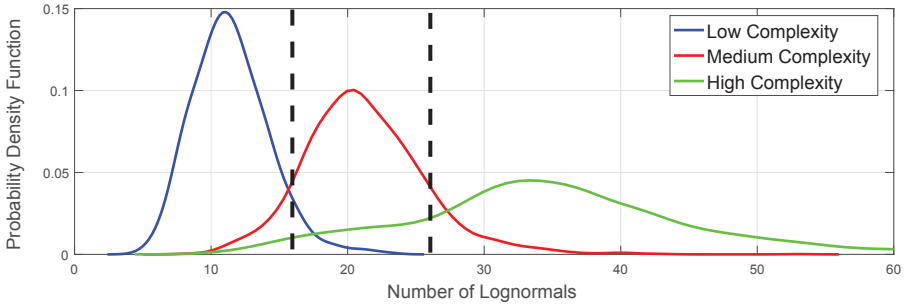
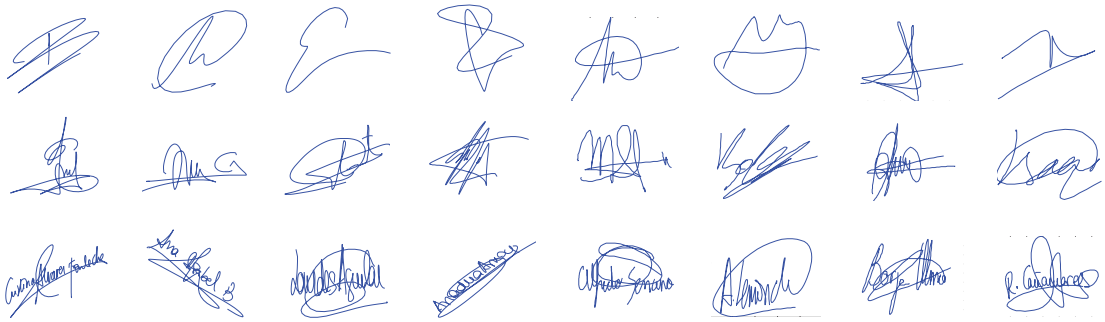


Figure 8.3: *Probability density function of the number of lognormals for each manually annotated complexity level using all genuine signatures of the BiosecurID database. The three proposed complexity-dependent decision thresholds are highlighted by black dashed lines.*



(a) BiosecurID (b) e-BioSign

Figure 8.4: *Signatures categorised into each complexity level using our proposed signature complexity detector. From top to bottom: low, medium and high complexity.*

an average number of 12.5 lognormals and a standard deviation of 1.3 whereas medium and high signature complexity levels achieve averages of 21.1 and 31.3 lognormals with standard deviations of 2.6 and 3.9, respectively. These results make sense as the intra-user variability increases with the signature complexity level. The same thresholds are extrapolated to the e-BioSign database. Fig. 8.4 shows some of the signatures classified into each complexity level for both BiosecurID and e-BioSign databases.

We now evaluate our proposed signature complexity detector following the same procedure carried out in [Houmani and Garcia-Salicetti, 2016], analysing the system performance for different complexity groups considering state-of-the-art signature verification systems as Baseline Systems [Gomez-Barrero *et al.*, 2015; Tolosana *et al.*, 2017a]. These Baseline Systems are based on DTW and a selection of the best local features through SFFS for each database regardless of the signature complexity level.

Table 8.1 shows the system performance in terms of EER(%) for each complexity level using the evaluation datasets and the stylus scenario. It is important to remark that each user is classified into a complexity level applying the majority voting algorithm to the 4 training signatures of the user.

Results show different system performance regarding the signature complexity level. Users

Table 8.1: *Signature complexity detector: System performance results (EER in %) of each complexity level using the BiosecurID and e-BioSign evaluation datasets for the stylus scenario. Skilled and random forgeries results are shown on top and bottom of each cell respectively.*

	Low C.	Medium C.	High C.
BiosecurID	13.8 1.5	7.5 0.7	6.2 0.9
e-BioSign	11.1 0.1	8.3 0.1	5.6 0.1

with a high complexity level have achieved an absolute improvement of 7.6% and 5.5% EER compared to users categorised into a low complexity level for the BiosecurID and e-BioSign databases, respectively. Similar results were obtained in previous studies using other approaches [Houmani and Garcia-Salicetti, 2016]. In that work, users categorised into a high complexity level achieved an absolute improvement of 8.5% EER compared to users categorised into a low complexity level for the MCYT database. These results prove the effectiveness of our proposed signature complexity detector based on the number of lognormals and the capacity to be applicable to other databases and scenarios.

In the following sections we analyse the idea of considering an on-line signature verification system adapted to the signature complexity level so as to further reduce the system performance.

8.3.2. Complexity-based Signature Verification System

This section aims to analyse which are the most discriminative and robust local features for each signature complexity level applying the SFFS over the development datasets. It is important to highlight that for the signature verification stage, the same DTW is always considered for obtaining the similarity score but different subsets of local features are selected for each complexity level and database.

For the BiosecurID database, a total of 4 genuine signatures from the first session and 12 genuine signatures from the remaining sessions are considered as training and testing signatures, respectively.

For the e-BioSign database, we obtain a separate optimal feature vector for each complexity level regardless of the writing input used while signing. This approach is achieved using training and testing signatures acquired by means of both stylus and finger inputs in order to select the best discriminative local features for all scenarios together. A total of 4 genuine signatures from the first session (2 signatures per writing input) and 8 genuine signatures from the second session (4 signatures per writing input) are considered as training and testing signatures, respectively.

The following three cases are analysed after applying the SFFS to each signature complexity level using the development dataset:

1. Local features selected for all three signature complexity levels.
2. Local features selected only for medium and high signature complexity levels.
3. Local features selected only for low and medium signature complexity levels.

Table 8.2: *Local features selected for each case and database using SFFS.*

	Case 1)	Case 2)	Case 3)
BiosecurID	$\dot{a_n}$, v_n^r	$\dot{v_n}$, $\ddot{y_n}$, $\dot{\alpha_n}$	c_n
e-BioSign	y_n, $\dot{x_n}$	$\dot{\theta_n}$, $\dot{y_n}$, v_n^r	x_n, s_n

Table 8.3: Stylus scenario: *System performance results (EER in %) on the BiosecurID and e-BioSign evaluation datasets for each complexity level. Skilled and random forgery results are shown on top and bottom of each cell respectively.*

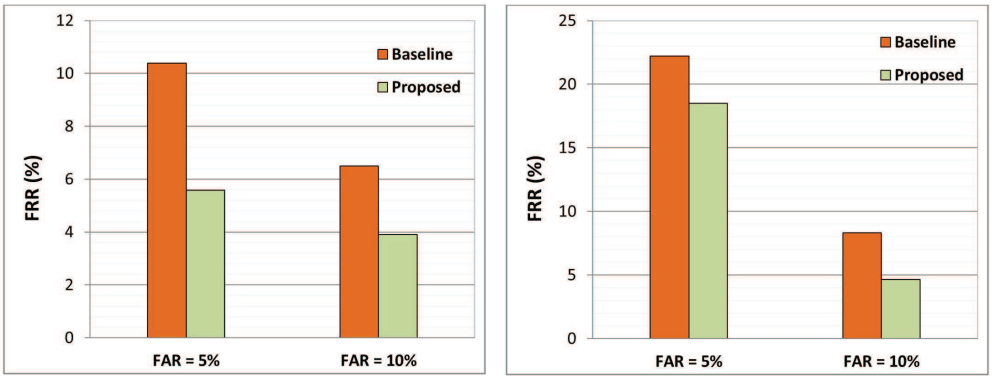
	Low C.		Medium C.		High C.	
	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
BiosecurID. Stylus-Stylus	13.8 1.5	10.1 1.3	7.5 0.7	5.2 0.5	6.2 0.9	4.6 0.9
e-BioSign. Stylus-Stylus	11.1 0.1	8.3 0.1	8.3 0.1	10.2 0.1	5.6 0.1	5.6 0.1

Table 8.2 shows the local features selected for each case and database. For the first case, the time functions $\dot{a_n}$ and v_n^r are selected in all systems of the BiosecurID database as robust local features regardless of the signature complexity level whereas for the e-BioSign database the local features selected are y_n and $\dot{x_n}$. While for the BiosecurID database the local features selected are more related to the acceleration and speed of the users performing their signatures, for the e-BioSign database local features related to the position of the writing tool (i.e., X and Y spatial coordinates) are more stable for all complexity levels. The reason why local features related to the acceleration and speed are not selected for the e-BioSign database is due to the fact that both stylus and finger writing tools are considered during training, and therefore, the way subjects sign on each input scenario is much more different than the local features related to the spatial position of the signature. For the second case, very similar local features have been selected for BiosecurID and e-BioSign databases for both medium and high signature complexity levels. These local features provide information related to the variation of the velocity, vertical acceleration and variation of angle, local features more related to the geometry of characters and therefore, to handwriting. Finally, local features such as c_n and s_n are selected for the third case and provide information related to the angles as signatures with low and medium complexity level are usually categorised for having simple flourishes with no legible information.

8.3.3. Stylus Scenario

This section evaluates our proposed complexity-based signature verification system for the case of using the stylus as input (i.e., Stylus-Stylus). Table 8.3 shows the results achieved for both BiosecurID and e-BioSign evaluation datasets. The same Baseline System described and used in Sec. 8.3.1 are considered here in order to make comparable our proposed approach. The only two differences between the Proposed and Baseline Systems are: *i)* the signature complexity detector, and *ii)* selection of the local features for each complexity level.

Analysing the results obtained for the BiosecurID database, our Proposed System achieves an average absolute improvement of 2.5% EER compared to the Baseline System for the skilled



(a) BiosecurID (b) e-BioSign

Figure 8.5: Stylus scenario: *False Rejection Rates (FRR) at different values of False Acceptance Rate (FAR) for both Proposed and Baseline Systems on the evaluation dataset.*

forgery case. It is important to remark that for the most challenging users (users with low complexity), our proposed approach achieves an absolute improvement of 3.7% EER compared to the Baseline System. Analysing the results obtained for random forgeries, our Proposed System also achieves improvements. For this case, the improvement is lower compared to the skilled forgery case due to SFFS is developed focusing on the most challenging impostor scenarios (i.e., skilled forgeries).

Analysing the results obtained for the e-BioSign database, our Proposed System also achieves the same trends. The improvement is slightly lower compared to the BiosecurID database due to the fact that a single system is developed for each complexity level considering not only the stylus, but also the finger and mixed writing-input scenarios.

Finally, Fig. 8.5 shows the performance of the Baseline and Proposed Systems considering all complexity levels together in terms of the FRR at different values of FAR. Our Proposed System achieves an average absolute improvement of 3.7% FRR for both BiosecurID and e-BioSign databases with a final value of 3.9% and 4.6% FRR for a value of FAR = 10.0% for BiosecurID and e-BioSign databases, respectively. These results show the importance of considering an on-line signature verification system adapted to the signature complexity level so as to increase the robustness of the system with more discriminative local features.

We now compare our proposed complexity-based signature verification system with other existing state-of-the-art approaches that have been evaluated in related publications using the BiosecurID database. The comparison is not straightforward as different experimental protocols are considered in each of the studies. This is something worth highlighting, not only for this comparison, but also for future experiments as results can vary significantly depending on the particular protocol used. For this reason, in order to perform a fair comparison to other studies, Table 8.4 depicts not only the FAR and FRR values achieved for each approach but also other very important features that affect the final system performance such as the complexity level of the considered users or the effect of the inter-session variability when testing. Our Proposed System outperforms the results achieved in previous works using a baseline system based on

Table 8.4: Stylus scenario: *System performance results (FAR and FRR in %) on the BiosecurID database. Comparison to previous works. It is worth noting that the % of users of different complexity levels shown for the different approaches have been computed by the complexity detector system proposed in this work.*

Work	Algorithm	Inter-Session Variability	# Training Signatures	% Users Low C.	% Users Medium C.	% Users High C.	FAR	FRR
[Ferrer *et al.*, 2017b]	DTW-based	No	5	7.6	39.4	53.0	3.1	3.1
[Diaz *et al.*, 2017b]	Manhattan-based	No	5	8.0	36.0	56.0	3.2	3.2
[Galbally *et al.*, 2015]	DTW-based	Yes	4	7.6	39.4	53.0	6.9	6.9
[Gomez-Barrero *et al.*, 2015]	Impostor Detector + DTW-based	Yes	4	7.2	38.0	54.8	4.8	4.8
Baseline System	DTW-based	Yes	4	9.6	37.9	52.5	5.0	10.4
Proposed Approach	**Complexity-based DTW**	**Yes**	**4**	**9.6**	**37.9**	**52.5**	**5.0**	**5.8**

the DTW algorithm, but without considering the signature complexity concept [Galbally *et al.*, 2015]. Besides, very similar results are achieved compared to [Gomez-Barrero *et al.*, 2015], in which a skilled forgery detector was incorporated to an already competitive baseline system. Finally, our proposed approach is also compared to other approaches based on Manhattan distance [Diaz *et al.*, 2017b], producing worse results due to a different number of training signatures, percentages of users in the complexity levels, and mainly due to the inter-session variability effect was not considered. This critical effect can be observed in [Ferrer *et al.*, 2017b] as well, where better results are achieved when applying a simple DTW approach based only on X and Y coordinates and their derivatives.

8.3.4. Finger and Mixed Writing-Input Scenarios

This section evaluates our proposed complexity-based approach considering COTS devices on two different scenarios: 1) the case of using only the finger as input for acquiring signatures (i.e., Finger-Finger), and 2) mixed writing-input (i.e., Stylus-Finger and Finger-Stylus) where signatures acquired using different inputs (i.e., stylus or finger) are independently considered for training and testing the system. Therefore, only the e-BioSign evaluation dataset is used in this section as signatures acquired using the finger are not available for the BiosecurID database. The same Baseline and Proposed Systems considered in the previous section are analysed here across both scenarios.

First, we analyse the results obtained for the case of using only the finger as input (i.e. Finger-Finger). Analysing the skilled forgery results depicted in Table 8.5, our Proposed System achieves an average absolute improvement of 3.4% EER compared to the Baseline System. Similar to the stylus scenario, the highest improvement is achieved for the most challenging users (i.e., users with low complexity level) with an absolute improvement of 5.6% EER compared to the Baseline System. Regarding random forgeries, the same very good results (close to 0.0% EER) are achieved with our proposed approach.

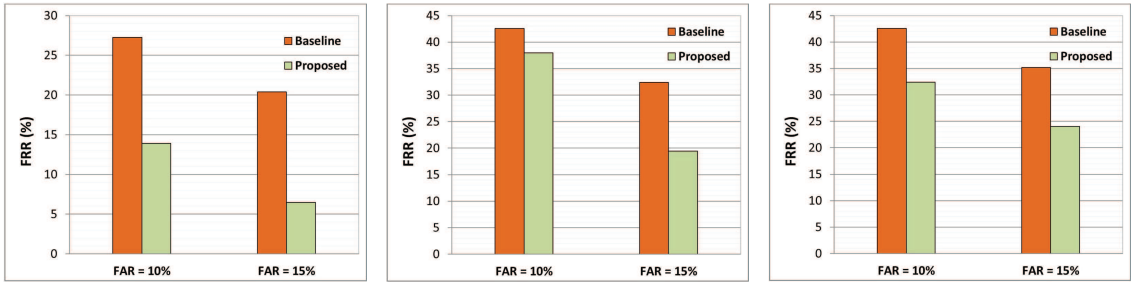
Despite the high improvement achieved in the finger scenario using our proposed approach, there is still a high difference in the system performance between both stylus and finger scenarios

Table 8.5: Finger and mixed writing-input scenarios: *System performance results (EER in %) on the e-BioSign evaluation dataset for each complexity level and scenario. Skilled and random forgery results are shown on top and bottom of each cell respectively.*

	Low C.		Medium C.		High C.	
	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
Stylus-Stylus	11.1 0.1	8.3 0.1	8.3 0.1	10.2 0.1	5.6 0.1	5.6 0.1
Finger-Finger	16.7 0.1	11.1 0.1	19.4 0.1	15.7 0.1	11.1 0.1	10.2 0.1
Stylus-Finger	30.6 0.1	27.8 0.1	22.2 0.1	16.7 0.1	11.1 0.1	11.1 0.1
Finger-Stylus	27.8 0.1	25.0 0.1	19.4 0.1	16.7 0.1	25.0 0.1	11.1 0.1

(Stylus-Stylus vs Finger-Finger). The results obtained using our Proposed System on the finger scenario show an absolute worsening of 4.3% EER compared to the stylus scenario. The reasons for this worsening of the system performance when using the finger were already explained in Chapter 5, and was mainly based due to the high variability of the users while signing in this novel scenario (e.g., closed letters such as a, e, and o tend to be much larger writing executions in comparison with other letters due to the lower precision users are able to achieve using the finger). Also, it is important to remark the challenging finger scenario considered in this work as forgers had access to the dynamic realization of the signatures to forge. A recommendation for the usage of signature recognition on mobile devices would be for the users to protect themselves from other people that could be watching while signing, as this is more feasible to do in a mobile scenario compared to an office scenario. This way skilled forgers might have access to the global shape of a signature but not to the dynamic information.

Now we describe the results obtained for the mixed writing-input scenarios (i.e., Stylus-Finger and Finger-Stylus), where signatures acquired using stylus and finger inputs are independently considered for training and testing the system. Analysing the results obtained in Table 8.5 for skilled forgeries, our Proposed System achieves an average absolute improvement of 2.8% and 6.5% EER compared to the Baseline System for the Stylus-Finger and Finger-Stylus scenarios, respectively. For the case of skilled forgeries, it is important to remark the significant worsening of the system performance for those users with a low complexity level with results around 25.0% EER. These results are much higher compared to the case of using the same writing input for testing. However, for users with medium and high complexity levels, the system performance on mixed writing-input scenarios are very close to the Finger-Finger scenario with results of 16.7% and 11.1% EER for medium and high complexity levels, respectively. Therefore, two very important conclusions can be extracted from our analysis on mixed writing-input scenarios and skilled forgery cases. The first is that mixed writing-input scenarios are feasible in practical applications for users with medium and high complexity levels. Users categorised into low complexity level should perform a more robust signature in order to be able to use these mixed writing-input scenarios. The second is that the degradation of the system performance on mixed writing-input scenarios seems to almost disappear for those users with medium and high complexity levels after applying our proposed approach based on the use of the signature complexity



(a) Finger-Finger (b) Stylus-Finger (c) Finger-Stylus

Figure 8.6: Finger and mixed writing-input scenarios: *FRR at different values of FAR for both Proposed and Baseline Systems on the evaluation dataset.*

detector and the selection of the most discriminative local features for each complexity level, obtaining similar results to the Finger-Finger scenario.

Finally, Fig. 8.6 shows the performance of both Baseline and Proposed Systems for the finger and mixed writing-input scenarios considering all complexity levels together in terms of FRR at different values of FAR. For the Finger-Finger scenario, our Proposed System achieves an average absolute improvement of 13.6% FRR compared to the Baseline System, with a final value of 13.9% FRR for a value of FAR = 10.0%. For the mixed writing-input scenarios, our Proposed System achieves an average absolute improvement of 8.8% and 10.7% FRR for the Stylus-Finger and Finger-Stylus scenarios, respectively. It is important to note the higher improvements achieved on the finger and mixed writing-input scenarios compared to the stylus scenario after applying our proposed approach proving the importance of exploiting the concept of complexity on these new challenging scenarios. Final values of 19.4% and 24.0% FRR are achieved for the Stylus-Finger and Finger-Stylus scenarios for a value of FAR = 15.0%. Therefore, the deployment of real applications on the Stylus-Finger scenario seems to be more feasible with rates below 20.0% of FRR and FAR. However, a possible recommendation for real applications could be to ask clients to perform their signatures using both stylus and finger writing tools during the enrolment stage in order to obtain better results, or at least for those users with low complexity level to avoid modifications of their signatures.

8.4. Chapter Summary and Conclusions

In this chapter we have proposed the first methodology focused on the development of on-line signature verification systems adapted to the signature complexity level of the user. This approach comprises two main modules: *i)* a new signature complexity detector based on the number of lognormals from the Sigma LogNormal writing generation module, and *ii)* a separate local feature extraction module adapted to each signature complexity level.

Our proposed approach has been tested on traditional and emerging scenarios, e.g., finger and mixed writing-input. Two well-known databases have been used in the experimental work of this Chapter: *i)* BiosecurID, which is considered for the traditional stylus scenario and comprises

a total of 400 users, and *ii)* e-BioSign, a new database that comprises signatures acquired using COTS devices on stylus, finger and mixed writing-input scenarios for a total of 65 users.

The proposed signature complexity detector has shown to be very effective despite of being based only on the number of lognormals. Signatures longer in time and with an appearance more similar to handwriting were labelled as high-complexity signatures whereas signatures shorter in time and with generally simple flourish with no legible information were labelled as low-complexity signatures. Additionally, an analysis of the stability regarding the number of lognormals for different signatures of the same user has been carried out in order to assess the feasibility of our proposed signature complexity detector. This simple approach has proven to be as useful and applicable to other databases and scenarios as other more sophisticated approaches.

Finally, our proposed complexity-based signature verification system has outperformed previous studies through the selection of the optimal subset of local features for each complexity level and input scenario. Analysing the results obtained for the stylus scenario, our Proposed System has achieved for the BiosecurID database an average absolute improvement of 2.5% EER for skilled forgeries compared to the Baseline System (the case where the local features are fixed to all complexity levels). Analysing the results obtained for the finger scenario, our Proposed System has achieved an absolute improvement of 5.6% EER for the most challenging users (i.e., users with low complexity level). We have concluded giving some recommendations in order to improve the performance of the on-line signature verification systems on real scenarios, and increase the security of the users against impostors.

Part IV

Handwritten Passwords for Touchscreen Biometrics

Chapter 9

Handwritten Passwords for Touchscreen Biometrics

This chapter evaluates the advantages and potential of incorporating biometrics to password-based mobile authentication systems, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual. This way, the traditional authentication systems are enhanced by incorporating dynamic handwritten biometric information. One example of use that motivates our proposed approach is on internet payments with credit cards. Banks usually send a numerical password (typically between 6 and 8 digits) to the user's mobile device. This numerical password must be inserted by the user in the security platform in order to complete the payment. Our proposed approach enhances such scenario by including a second authentication factor based on the user biometric information while drawing the digits. Fig. 9.1 shows a general architecture of our proposed password-based mobile authentication approach. The three following main modules are analysed in this study: *i)* enrolment set, *ii)* password generation, and *iii)* touch biometric system. Depending on the final application (i.e., PIN or OTP), the handwritten digits can be first recognised using for example an Optical Character Recognition (OCR) system in order to verify the authenticity of the password. After this first authentication stage, the biometric information of the handwritten digits is compared in a second authentication stage to the enrolment data of the claimed user, comparing each digit one by one. In this study we focus on the second authentication stage based on the behavioral information of the user while performing the handwritten digits as the recognition of numerical digits has already shown to be an almost solved problem with errors close to 0% [Liang and Hu, 2015; Wan *et al.*, 2013]. Therefore, in this study we make the assumption that impostors pass the first stage of the security system (i.e., they know the password of the user to attack) and thus, the attack would have 100% success rate if our proposed approach was not present.

This chapter is organised as follows. Sec. 9.1 describes our proposed touch biometric system. Sec. 9.2 and 9.3 describe the experimental protocol and results achieved using our proposed approach, respectively. Sec. 9.4 discusses specific details for the deployment of our proposed ap-

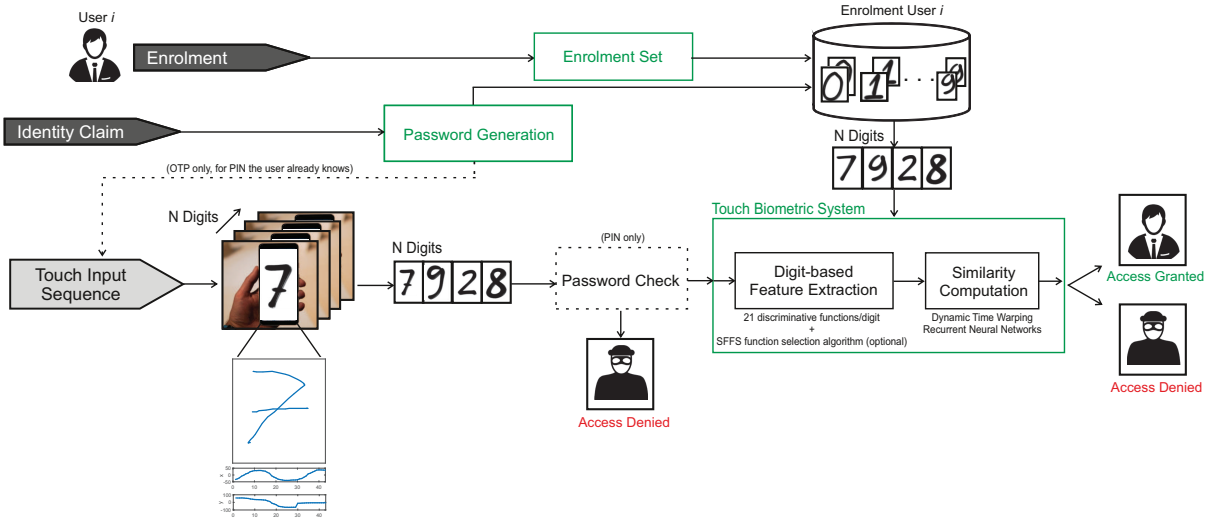


Figure 9.1: *Architecture of our proposed password-based mobile authentication approach including handwritten touch biometrics in a two-factor authentication scheme applicable both to user-generated PIN and OTP systems.*

proach on current PIN- and OTP-based authentication systems, including password generation strategies. Sec. 9.5 aims to provide the last new advancements in the topic. Finally, Sec. 9.6 draws the final conclusions.

This chapter is based on the following publications: [Tolosana *et al.*, 2018b,d, 2019b].

9.1. Touch Biometric System

9.1.1. Digit-based Feature Extraction

In this chapter we evaluate the potential of touch biometric verification systems based on local features. Signals captured by the digitizer (i.e., X and Y spatial coordinates) are used to extract the same set of 21 local features described in Chapter 4 for each numerical digit sample (see Table 4.3). Information related to pressure, pen angular orientations or pen ups broadly used in other biometric traits such as handwriting and handwritten signature is not considered in this chapter as this information is not available in all mobile devices when using the finger touch as input.

SFFS algorithm is used for the DTW algorithm in some of the experiments in order to select the best subsets of local features for each handwritten digit and improve the system performance in terms of EER (%).

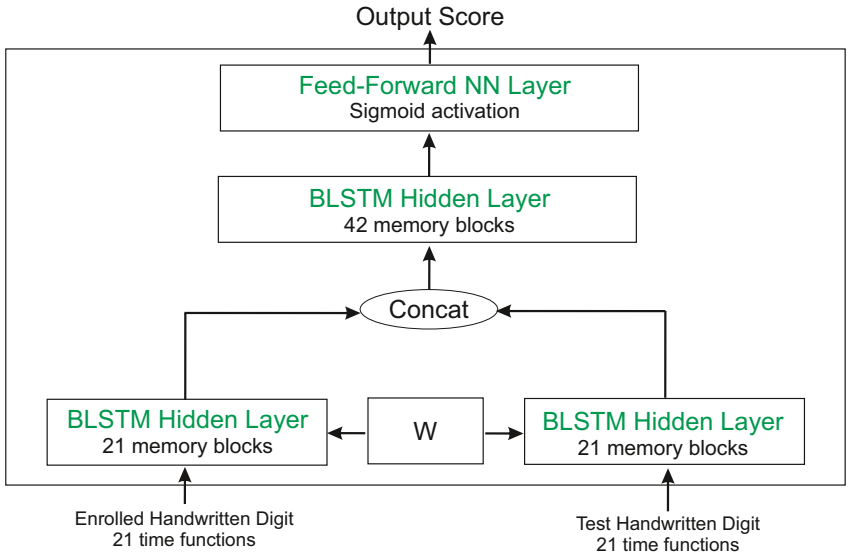


Figure 9.2: *Proposed end-to-end writer-independent BLSTM touch biometric system based on a Siamese architecture.*

9.1.2. Similarity Computation

9.1.2.1. Dynamic Time Warping

The same DTW configuration described in Chapter 4 is used here to compare the similarity between genuine and query input samples. Scores are obtained as $score = e^{-D/K}$, where D and K represent respectively the minimal accumulated distance and the length of the warping path [Martinez-Diaz *et al.*, 2015b].

9.1.2.2. Recurrent Neural Networks

In this study we adapt the original BLSTM system proposed in Chapter 4 for handwritten signature verification to handwritten passwords for touchscreen biometrics. To the best of our knowledge, this is the first study to date that studies recurrent Siamese networks to model handwritten password authentication systems. Fig. 9.2 shows our proposed end-to-end writer-independent BLSTM touch biometric system based on a Siamese architecture. For the input of the system, we feed the network with as much information as possible, i.e., all 21 local features per digit. The first layer is composed of two BLSTM hidden layers with 21 memory blocks each, sharing the weights between them. The outputs of the first two parallel BLSTM hidden layers are concatenated and serve as input to the second layer, which corresponds to a BLSTM hidden layer with 42 memory blocks. Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score for each pair of digits. It is important to highlight that our approach is trained to distinguish between genuine and impostor patterns from all numerical digits and users. Thus, we just train one writer-independent system for all digits and users through a development dataset.

9.2. Experimental Protocol

The experimental protocol designed in this study intends to cover all details of the two following main modules of our proposed password-based touch biometric system (see Fig. 9.1):

- **Enrolment Set:** When designing biometric authentication systems for real applications, there are usually two conflicting factors: *i)* the amount of data requested to the user during the enrolment, and *ii)* the security level provided by the biometric system. From the point of view of the security system, it seems clear that the ideal case would be to have as much information of the user as possible. However, in most real scenarios, the feasibility and success depend on the development of user-friendly applications.

  This aspect has shown to be crucial for different tasks such as the handwritten signature. In Chapter 6, we evaluated this effect using statistical systems based on HMM and GMM, achieving an absolute improvement of 11.7% EER when training the user models with 41 genuine signatures instead of just 4. In this chapter, we analyse the intra-user variability on this new authentication scenario and perform a complete analysis of how the biometric system performance changes with the number of enrolment samples acquired per digit.

- **Password Generation:** The selection of a password that is robust enough for a specific application is a key factor. The number of digits that comprise the password depends on the scenario and level of security considered in the final application. For example, for everyday applications such as Facebook or Gmail, it is not reasonable from the point of view of the users to memorise passwords composed of 12 digits. Additionally, OTP-based systems could request longer passwords compared to PIN-based systems as users do not have to memorise them, i.e., the security system is in charge of selecting and providing the password to the user.

  In this experimental chapter we evaluate the robustness of handwritten passwords regarding the three following features: *i)* which digits better discriminate users, *ii)* whether repetitions of the same numerical digits in a password can help to discriminate users or not, and *iii)* the length of the password. For short passwords (i.e., fewer than 6 digits), this analysis is carried out performing all possible digit combinations, whereas for longer passwords, the SFFS algorithm is used to select the best digit combinations due to the high cost of performing all possible comparisons.

In order to perform a complete analysis of these two modules, the e-BioDigit database described in Chapter 3 is divided into development (the first 50 users) and evaluation (the remaining 43 users) datasets.

For the development of our proposed handwritten touch biometric systems, N genuine signatures (up to 4) from the first session can be used as enrolment samples, whereas the 4 remaining genuine samples from the second session are used for testing. Impostor scores are obtained by

Table 9.1: *Local features for the Baseline System.*

#	Feature description
1	X-coordinate: x_n
2	Y-coordinate: y_n
7-8	First-order derivate of features 1-2: $\dot{x_n}, \dot{y_n}$
13-14	Second-order derivate of features 1-2: $\ddot{x_n}, \ddot{y_n}$

comparing the N enrolment samples with one genuine sample of each of the remaining users (simulating this way the imitation attack in which the impostor knows the password).

For the evaluation of our proposed touch biometric system, different scenarios are generally considered regarding the number of available enrolment samples per user (i.e., Nvs1), in which the final score is performed as the average score of N one-to-one comparisons. In addition, in case of using passwords composed of several digits, the final score is produced after averaging the different one by one digit score comparisons.

It is important to highlight that the inter-session variability problem is also considered in the experimental protocol carried out in this study as genuine digit samples from different sessions are used as enrolment and testing samples respectively. This effect has proven to be very important for many behavioral biometric traits such as the case of the handwritten signature [Galbally *et al.*, 2013].

9.3. Experimental Results

9.3.1. One-Digit Analysis

This section analyses the potential of each numerical digit (i.e., from 0 to 9) for the task of user authentication. We consider three different systems: *i)* a baseline DTW system, *ii)* an adapted DTW considering feature selection, and *iii)* a system based on RNNs.

Experimental results on the evaluation dataset for these three systems are shown in Table 9.2 and 9.3 in terms of EER (%) for the cases of 1vs1 and 4vs1 comparisons, respectively.

9.3.1.1. DTW Baseline System

In order to provide an easily reproducible framework, we first consider a baseline system based on DTW with the same fixed local features for all numerical digits. Table 9.1 shows the local features selected, which are commonly used as baseline in other biometric traits such as the handwritten signature [Blanco-Gonzalo *et al.*, 2014; Tolosana *et al.*, 2017a].

Analysing the first rows of Tables 9.2 and 9.3 we can see how very good authentication results are obtained by the DTW Baseline System taking into account that we only consider one digit and the same local features for all numerical digits.

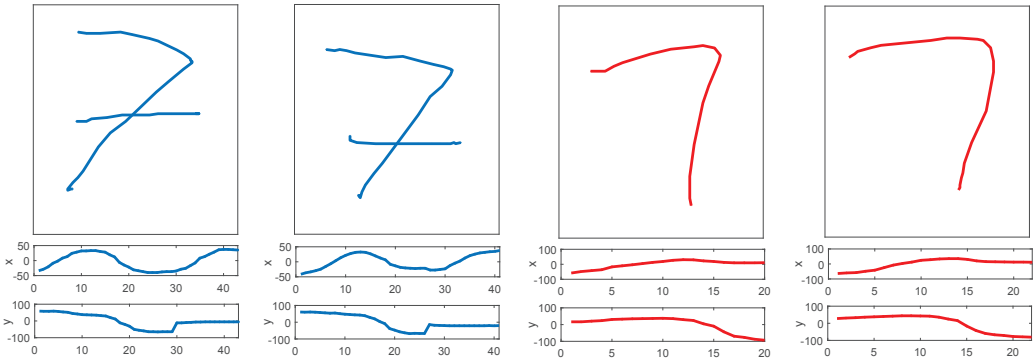
Analysing in Table 9.2 the extreme scenario of having just one available digit sample during the enrolment (1vs1), the numerical digit 7 achieves the best result with 22.5% EER. In addition, other numerical digits such as 4 or 5 achieve similar results with EERs below 25.0%. This first

Table 9.2: *System performance as EER(%) of each numerical digit for the* ***1vs1*** *case on the evaluation dataset.*

	Numerical Digit									
	0	1	2	3	4	5	6	7	8	9
DTW Baseline System	34.9	32.3	32.8	35.0	23.5	24.4	36.9	**22.5**	26.0	29.6
DTW Adapted System	33.0	34.0	30.9	32.3	22.0	**21.7**	33.6	21.8	21.8	27.0
BLSTM System	32.8	30.8	32.8	32.3	26.2	**19.6**	35.2	28.5	21.7	23.8

Table 9.3: *System performance as EER(%) of each numerical digit for the* ***4vs1*** *case on the evaluation dataset.*

	Numerical Digit									
	0	1	2	3	4	5	6	7	8	9
DTW Baseline System	33.1	28.5	30.2	32.6	**18.0**	20.3	36.6	19.2	22.7	25.0
DTW Adapted System	31.4	33.1	27.9	29.7	19.2	**16.9**	29.7	20.3	18.6	23.3
BLSTM System	31.4	27.9	31.4	26.2	24.4	**17.4**	35.4	24.4	18.0	20.9



(a) User A, sample 1 (b) User A, sample 2 (c) User B, sample 1 (d) User B, sample 2

Figure 9.3: *Examples of the numerical digit 7 performed by two different users.*

experiment puts in evidence the discriminative power of each handwritten digit. Fig. 9.3 shows examples of the digit 7 performed by two different users in order to observe the low intra- and high inter-user variability of this number. This effect is produced as different users tend to perform a specific digit in a different way, i.e., starting from a different stroke of the digit or even removing some of them such as the crossed horizontal stroke of the number 7.

Analysing in Table 9.3 the scenario of using four enrolment samples (4vs1), an average absolute improvement of 3.2% EER is achieved compared to the 1vs1 scenario showing the importance of acquiring more than one sample during the enrolment stage, if possible. For this scenario, the digit 4 achieves the best result with 18.0% EER.

9.3.1.2. DTW Adapted System

We now apply SFFS over the development dataset in order to enhance the DTW touch biometric system through the selection of specific local features for each handwritten digit. Fig. 9.4 shows the number of times each local feature is selected in our DTW Adapted System from the 21 total local features described in Chapter 4, Table 4.3. In general, we can highlight

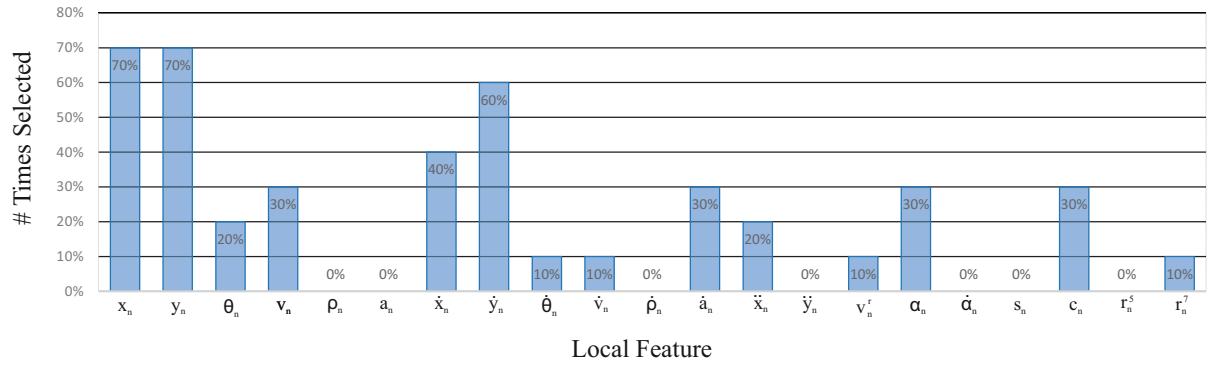


Figure 9.4: *Histogram of local features selected by SFFS for our DTW Adapted System. Local features described in Table 4.3.*

the importance of x_n, y_n local features as they are selected for 70% of the numerical digits. In addition, local features $\dot{x_n}, \dot{y_n}$ related to X and Y time derivatives seem to be very important as they are selected for near half of the digits. Other local features such as ρ_n, $\dot{\rho_n}$, $\dot{\alpha_n}$ and s_n related to geometrical aspects of the numerical digits are proven not to be very useful to discriminate between genuine and impostor users.

The second rows of Tables 9.2 and 9.3 show the results achieved for each digit using our DTW Adapted System over the evaluation dataset for both 1vs1 and 4vs1 cases, respectively. In general, better results are achieved compared to the DTW Baseline System. Analysing the 1vs1 scenario, our DTW Adapted System achieves an average absolute improvement of 2.0% EER, being the numerical digit 5 the one that provides the best result with a 21.7% EER. Analysing the 4vs1 scenario, our DTW Adapted System achieves an average absolute improvement of 1.6% EER, being again the numerical digit 5 the one that achieves the best result with a 16.9% EER. These results put in evidence the importance of considering different local features for each digit in order to develop more robust biometric authentication systems against attacks.

9.3.1.3. BLSTM System

We now explore the potential of state-of-the-art deep learning technology applied to our touch biometric data. Our proposed end-to-end writer-independent BLSTM system is trained using only the 50 users of the development dataset. Samples from all numerical digits (i.e., from 0 to 9) and development users are considered together during training as we intend to distinguish between genuine and impostor handwritten digit samples regardless of the user and the numerical digit. This approach resulted in better generalisation results compared to the case of training one system per numerical digit. Therefore, our BLSTM system is trained considering two different cases: *i)* pairs of genuine digit samples drawn by the same user, and *ii)* pairs of genuine and impostor digit samples, one performed by the claimed user and the other one by an impostor. For each case there are a total of 4 train samples $\times$ 4 test samples $\times$ 10 numerical digits $\times$ 50 users $\simeq$ 8,000 comparisons, having the same number of genuine and impostor comparisons. Our BLSTM System has been implemented under Keras using Tensorflow as back-end, with

a NVIDIA GeForce RTX 2080 Ti GPU. Adam optimizer is considered with a learning rate of 0.001 and a loss function based on binary cross-entropy.

The third rows of Tables 9.2 and 9.3 show the results achieved for each digit using our BLSTM System over the evaluation dataset for both 1vs1 and 4vs1 cases, respectively. In general, better results are achieved compared to the DTW Baseline System. Analysing the 1vs1 scenario, our BLSTM System achieves an average absolute improvement of 1.4% EER, being the numerical digit 5 the one that provides the best result with a 19.6% EER. Analysing the 4vs1 scenario, our BLSTM System achieves an average absolute improvement of 0.9% EER, being again the numerical digit 5 the one that achieves the best result with a 17.4% EER.

Finally, we compare our BLSTM System to the DTW Adapted System. In general, very similar results have been achieved for both authentication systems. The BLSTM System has outperformed the DTW Adapted System for some numerical digits (e.g., numerical digits 1 and 9 in Tables 9.2 and 9.3), proving the potential of deep learning technologies even in the scenario considered here where only a small amount of data is available during the training process. Despite these improvements, the DTW Adapted System outperforms slightly the BLSTM System in general, achieving an average absolute improvement of 0.5% and 0.7% EER for the 1vs1 and 4vs1 cases, respectively.

9.3.2. Digit Combinations

This section explores the robustness of our proposed approach when increasing the length of the password and also the number of available enrolment samples. The DTW Adapted System is considered in this analysis as it has outperformed the other systems studied. Regarding the type of digits that comprises the password, repetitions of the same numerical digits are allowed. However, the number of repetitions is restricted to 4, e.g., "2 5 8 8 8 8". The reason for this limitation is motivated due to only 4 samples were acquired per digit during the second session of the e-BioDigit database. Table 9.4 shows the evolution of the system performance in terms of EER (%) on the evaluation dataset when increasing the length of the handwritten password (from 1 to 8 digits) and also the number of available enrolment samples (from 1 to 4).

First, we analyse how the length of the handwritten password affects the system performance. In general, a considerable system performance improvement is achieved when adding more handwritten digits to the password. For example, for the case of having just one enrolment sample per user (1vs1), a password that is composed of just two handwritten digits achieves a 14.0% EER, an absolute improvement of 7.7% EER compared to the case of using a password with just one digit. This result is further improved when increasing the number of handwritten digits of the password with a final 8.5% EER for the case of considering a 6-digit password. However, there seems to exist a limit in the system performance improvement with the number of digits that comprise the password. In our experiments, the best results are obtained for passwords with a length of 6 and 7 digits.

Now, we analyse the effect of the number of available enrolment samples on the system performance. In general, the system performance improves with the number of enrolment sam-

Table 9.4: *Evolution of the system performance in terms of EER (%) on the evaluation dataset. The best system performance achieved and the corresponding handwritten digits selected are shown on top and bottom of each cell respectively.*

		# Digits that comprise the password							
		1	2	3	4	5	6	7	8
# Enrolment samples	1	21.7 [5]	14.0 [5, 8]	11.6 [5, 7, 9]	11.6 [1, 5, 7, 9]	9.3 [2, 5, 6, 7, 8]	8.5 [2, 3, 5, 6, 7, 8]	8.5 [1, 2, 3, 5, 6, 7, 8]	8.5 [2, 3, 4, 5, 6, 7, 8, 9]
	2	18.6 [5]	11.6 [5, 8]	9.3 [2, 5, 8]	7.4 [2, 5, 8, 9]	7.3 [1, 2, 5, 7, 9]	4.6 [2, 5, 6, 7, 8, 9]	4.6 [1, 2, 3, 5, 7, 8, 9]	4.6 [1, 2, 3, 4, 5, 6, 7, 8]
	3	16.3 [5]	9.5 [2, 8]	7.4 [1, 2, 8]	5.9 [2, 5, 8, 9]	4.7 [1, 2, 5, 8, 9]	4.6 [1, 2, 3, 5, 8, 9]	**3.8** **[1, 2, 3, 4, 5, 8, 9]**	4.6 [0, 1, 2, 3, 4, 5, 7, 8]
	4	16.9 [5]	11.6 [5, 8]	7.0 [7, 8, 9]	6.1 [5, 7, 8, 9]	4.7 [1, 5, 7, 8, 9]	4.6 [1, 2, 5, 7, 8, 9]	4.3 [1, 2, 3, 5, 7, 8, 9]	4.8 [0, 1, 2, 3, 4, 5, 7, 8]

ples. For example, for the case of having just one enrolment sample and a password composed of just one digit, the biometric system achieves a 21.7% EER. This result is further improved when increasing the number of enrolment samples to 4, achieving a final value of 16.9% EER, an absolute improvement of 4.8% EER. However, there seems to exist a limit in the system performance improvement with the number of enrolment samples. In our experiment, very similar results are obtained when considering 3 or 4 enrolment samples, achieving a final value of 3.8% EER when considering 3 enrolment samples and a handwritten password of 7 digits. This interesting finding is different compared to other behavioral biometric traits such as the handwritten signature as the system performance keeps improving even with large number of enrolment samples (see Chapter 6). This effect may be due to the lower intra-user variability of our proposed touch biometric approach compared to other behavioral biometrics as well as the DTW similarity computation algorithm considered.

Finally, we pay attention to the content and the number of possible combinations of the best handwritten passwords using our proposed touch biometric system so as to achieve the best system performance. Table 9.4 indicates in the bottom of each cell the best handwritten digits selected but not their order, as the final score of our proposed touch biometric system is produced after averaging the different one by one digit score comparisons. Therefore, for the case of having a password comprised of n digits, there are a total of $n!$ possible password combinations (note that in our experiments we did not have any case of repetitions of digits achieving the best results).

9.3.3. Comparison to the State of the Art

Our proposed approach is now compared to other state-of-the-art biometric authentication approaches described in Table 2.1. In order to perform a fair analysis, we compare our proposed approach to all studies that consider the same type of impostors, i.e., imitation attacks.

In general, our proposed approach achieves better results than other touch biometric approaches. For the case of lock pattern dynamic systems [Angulo and Wastlund, 2011; Lacharme and Rosenberger, 2016], the best system performance reported was an average 10.39% EER. Our proposed approach also outperforms other biometric methods such as the handwritten signature

or graphical passwords. In Chapter 5 of this Thesis, we have analysed handwritten signature verification systems adapted to mobile scenarios, i.e., using mobile devices such as smartphones and tablets with the finger as input, achieving EERs around 20.0%. In [Martinez-Diaz *et al.*, 2016], the authors proposed the use of graphical doodles and pseudosignatures (i.e. simplified versions of the signatures drawn with the finger). EERs above 20.0% were obtained in both cases for skilled forgeries.

Finally, our proposed approach has been compared to other state-of-the-art authentication systems based on handwritten passwords. In [Kutzner *et al.*, 2015], the authors proposed the use of handwritten passwords with a fixed length of 8 characters, achieving a final False Acceptance Rate (FAR) of 10.42% when using a total of 12 training samples per user (the False Rejection Rate FRR was not provided by the authors). Nguyen *et al.* [2017a] evaluated the potential of drawing each digit of a 4-digit PIN one by one, achieving a final result of 4.84% EER when considering a total of 5 enrolment samples. Our proposed approach achieves a final value of 3.8% EER and it is able to mitigate the limitations of [Kutzner *et al.*, 2015] about the size of the touchscreen, as users perform numerical digits one at a time. Additionally, we only consider 3 enrolment samples and not 5 as in [Nguyen *et al.*, 2017a] in order to improve the usability of our approach.

9.4. Password Generation and System Setup

In this section we discuss specific details for the deployment of our proposed approach in real scenarios considering the same experimental protocol described in Sec. 9.2. The DTW Adapted System has been considered for this analysis.

First, we focus on PIN-based systems. For this scenario, we propose to use passwords based on 4 digits as users have to memorise them and it is not feasible from the point of view of the user to consider longer passwords. Regarding the enrolment stage, we propose to request 3 enrolment samples per digit to each user. We consider this as something feasible for real applications as users would have to perform a total of 4 digits $\times$ 3 samples/digit = 12 samples, i.e., 12 samples $\times$ 2 seconds/sample $\simeq$ 25 seconds.

Once we have fixed the number of enrolment samples and digits parameters, we design what type of passwords we let users to use (i.e., we design the Password Generation module in Fig. 9.1). The following cases are considered regarding both the system performance and number of possible combinations: *i)* ALL password combinations are allowed, and *ii)* only combinations using the BEST 4 digits selected in Table 9.4 and with no repetitions (recall in Sect. 9.3.2 we obtained that the most discriminative password combinations in terms of touch biometric information didn't include repeated digits). Fig. 9.5 shows the EER distribution values obtained for all possible password combinations. On the box, the central mark indicates the median, and the left and right edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually. In general, we can see that the 75% of password combinations provide

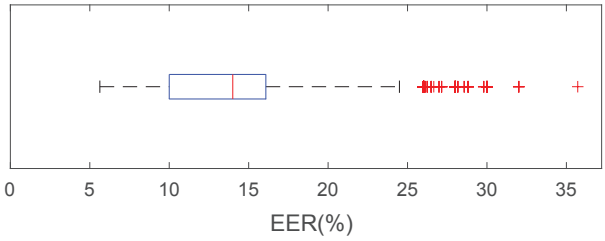


***Figure 9.5: PIN System**: Boxplot for the case of considering all 4-digit password combinations. On the box, the central mark indicates the median, and the left and right edges of the box indicate the 25th and 75th percentiles, respectively.*

***Table 9.5: OTP System**: number of 7-digit possible combinations and system performance results.*

	# Password Combinations	EER(%)
Case ALL	10^7	3.8 to 14.0
Case BEST	5,040	3.8

results below 16.2% EER. Analysing the case ALL, the system performance results achieved are between 5.9% and 35.7% EER with a total of 10^4 combinations. The performance is improved in the case BEST with a 5.9% EER for all considered combinations. However, users would be able to choose only among 4! combinations (i.e., 24). Besides, the security level of the first authentication stage would decrease as fewer password combinations would be possible. Therefore, a good choice could be to select all possible passwords that provide results in a range of EERs. For example, permitting between 5.9% and 10.0% EER. This approach would allow users to choose among 2,956 different 4-digit passwords.

Now, we analyse the OTP-based system. For this scenario, we propose to use passwords composed of 7 digits, similar to current OTP-based applications, as users do not have to memorise the password, i.e., the system is in charge of selecting and providing different passwords to the user each time is required. Regarding the enrolment stage, we also propose to request 3 enrolment samples per digit so users would have to perform a total of 10 digits $\times$ 3 samples/digit $=$ 30 samples, i.e., 30 samples $\times$ 2 seconds/sample $\simeq$ 1 minute.

Once we have fixed both the number of enrolment samples and the length of the password, we analyse the content of the passwords. For this scenario, the following cases are considered: *i)* ALL digit combinations are allowed, and *ii)* only combinations using the BEST 7 digits selected in Table 9.4 with no repetitions. Table 9.5 depicts the number of possible combinations as well as the EER (%) for both cases. Analysing the case in which users can choose any possible combination, the system performance results achieved are between 3.8% and 14.0% EER. However, it is important to remark that for this case (longer passwords) results were obtained due to experimental restrictions using the SFFS algorithm and limiting the maximum number of digit repetitions to 4, so the final 14.0% EER might get a bit worse in practice when considering all possible digit combinations. This approach is further improved in the case BEST with a final 3.8% EER. For this scenario we propose to use this second case as there would be a total of 7! (i.e., 5,040) combinations that provide the best system performance for our proposed

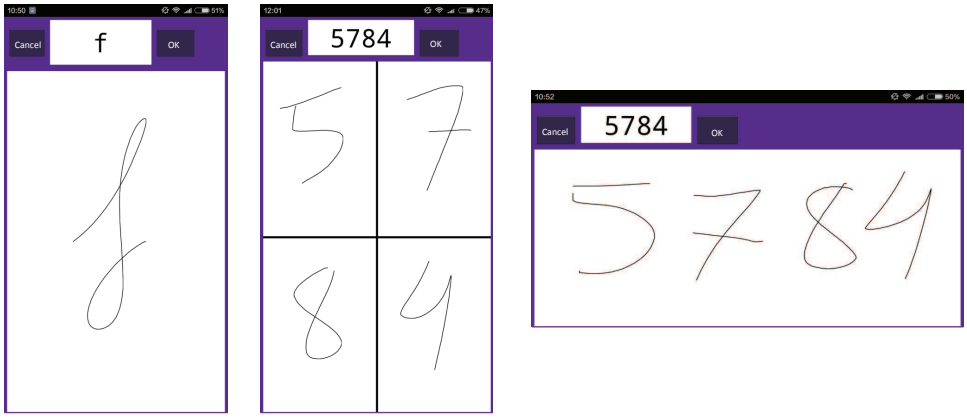


Figure 9.6: *Different interfaces designed for the acquisition app. Both portrait and landscape orientations are considered in order to analyse different user experiences while drawing.*

touch biometric approach.

9.5. New Advancements

This final section aims to provide the last new advancements obtained in this interesting approach. So far this chapter, we have analysed the potential of drawing each digit of the password on the touchscreen instead of typing them as usual. Very good results have been obtained taking into account that the impostor would have 100% success rate if our proposed approach was not present. Therefore, and after performing this first study, we decided to extend this approach to more practical scenarios through the acquisition of a novel mobile touch on-line database named MobileTouchDB. The database contains more than 64K on-line character samples performed by 218 users, using 94 different smartphone models, with an average of 314 samples per user. In each acquisition session, users had to draw all numbers (from 0 to 9), upper- and lower-case letters (54), different symbols (8), and passwords composed of 4 numbers (6). Regarding the acquisition protocol, MobileTouchDB comprises a maximum of 6 captured sessions per subject with a time gap between them of at least 2 days. This database studies an unsupervised mobile scenario with no restrictions in terms of position, posture, and devices.

Regarding the acquisition, we implemented an Android application. Fig. 9.6 represents the different interfaces designed for the acquisition. All interfaces are composed of: *i)* the character to draw (top, middle) and two buttons "OK" (top, right) and "Cancel" (top, left) to press after drawing if the sample was good or bad respectively. If the sample was not good, then it was repeated. And *ii)* a rectangular area to perform the character or password. In order to study an unsupervised mobile scenario, the acquisition app was uploaded to the Google Play Store. This way all participants could download and use the app on their own devices without any kind of supervision, simulating a practical scenario in which users can generate handwritten information in any possible scenario, e.g., standing, sitting, walking, indoors, outdoors, etc. As

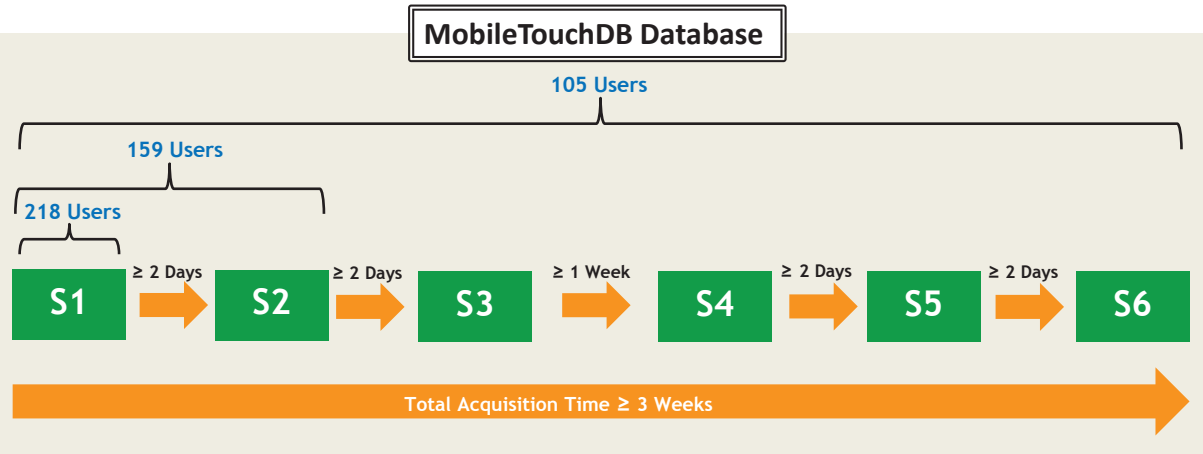


Figure 9.7: *Description of the design and number of available users of the new MobileTouchDB.*

a result, 94 different models from the following 16 brands were used during the acquisition: *Alcatel, Blackberry, BQ, Coolpad, Doogee, Google, Huawei, LeTV, LG, Motorola, OnePlus, Samsung, Sony, UMIDIGI, Xiaomi, and ZTE*. The acquisition app was designed to capture the following time signals: X and Y spatial coordinates, the area covered by the finger, timestamp, accelerometer, and gyroscope. However, information related to the area covered by the finger, accelerometer, and gyroscope was not available in some cases depending on how old was the acquisition device.

The acquisition protocol considered in the MobileTouchDB database is depicted in Fig. 9.7. It comprises a total of 6 sessions (i.e., S1-S6) with different time gaps among them. It is important to highlight that in all sessions, the time gap refers to the minimum time between one user finishes a session and the following session is available. However, participants usually performed their corresponding sessions later on thanks to notifications sent automatically by the acquisition app to the users. Regarding the data acquired, each session comprises 8 different capturing blocks (i.e., from Block1 to Block8). Fig. 9.8 shows some examples of each of the eight acquisition blocks for two different users (indicated in blue and red colours). The green dashed lines indicate pen ups trajectories between strokes. In Block1, we asked users to draw all numbers (from 0 to 9). Block2 and Block3 comprise upper- and lower-case letters respectively, with a total of 27 letters each. Block4 is composed of 8 different symbols (i.e., *"?"*, *"#"*, *"*"*, *"@"*, *"%"*, *"="*, *"ϵ"*, *and "α"*). It is important to remark that inside each block, characters were randomised before asking users to draw them. This way, each user performs a different character sequence in each session. From Block1 to Block4, the acquisition interface was designed as portrait to provide a better user experience (see Fig. 9.6, left). After finishing the first 4 blocks focused on performing one single character at a time (one sample per character), we asked users to draw passwords composed of 4 numbers (always *"5 7 8 4"*) in different ways (6 samples in total). In Block5, users performed the password twice using a landscape orientation interface (see Fig. 9.6, right). We provided the users with a graphical visualization of the numbers while

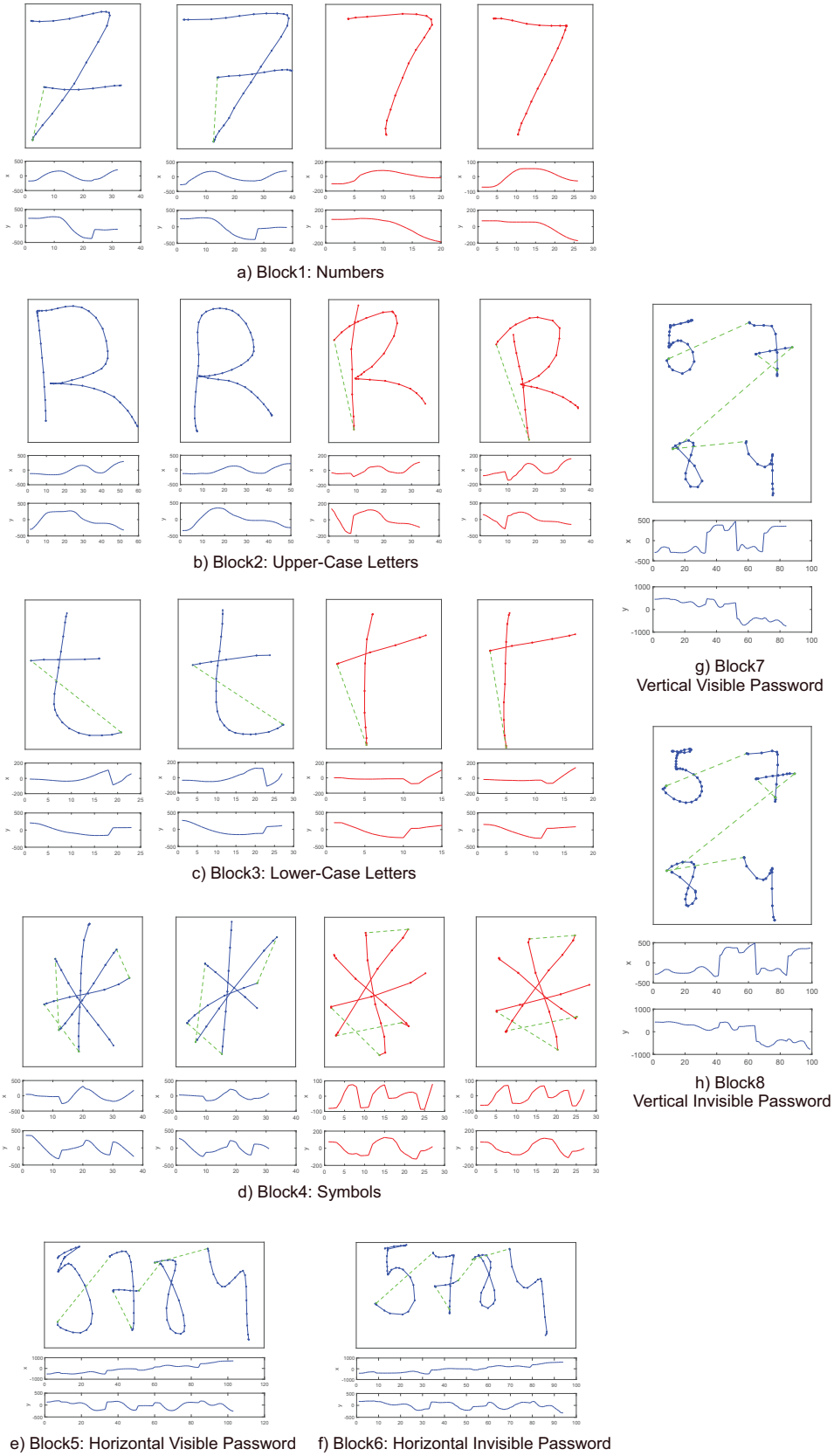


Figure 9.8: *Example of the data collected in MobileTouchDB database. Blue and red colours represents samples drawn by different users. The green dashed lines indicate pen ups trajectories between strokes. Curves under each character represent X and Y trajectories over time.*

drawing them (i.e., visible mode). Then, in Block6, users had to repeat once the same task considered in Block5 but this time in an invisible mode, i.e., we did not provide to the users any visualization of the numbers while drawing them. The main motivation of this novel acquisition scenario is to protect us against shoulder surfing attacks, as commented in [Nguyen *et al.*, 2017b]. In Block 7, users had to draw each number of the password inside of each of the four available boxes (two times), considering first a visible mode (see Fig. 9.6, middle). Finally, in Block8 users had to repeat once the same task considered in Block7 but this time in an invisible mode. In both Block7 and Block8 the acquisition interface was kept portrait to analyse the user experience in different settings.

Regarding the MobileTouchDB population statistics, 218 users completed the S1 acquisition session. S1 and S2 were completed by 159 users. Finally, a total of 109 users completed the six acquisition sessions. This participant reduction between S1 and S6 sessions is produced due to the challenging acquisition scenario considered in this study as it was completely unsupervised and comprised several acquisition sessions along time. Regarding the age distribution, 36.2% of the participants are younger than 22 years old, 31.9% are between 22 and 27 years old, and the remaining 31.9% are older than 27 years old. Regarding the gender, 63% of the participants were males, and 37% females. 96% of the population was righthanded.

Preliminary results carried out using a baseline system based on DTW and the same fixed time functions for all characters (i.e., X and Y coordinates over time and their first- and second-order derivatives) prove the discriminative power of lower- and -upper case letters, and symbols compared to the initial results obtained using only the numbers. Lower-case letters achieve an average absolute improvement of 1.9% EER compared to numbers whereas for symbols and upper-case letters the average absolute improvement is 1.8% and 0.9% EER, respectively. For future work, we expect to further reduce the EER through more advanced techniques based on feature selection and deep learning. Additionally, we will study the discriminative power of new features acquired in the database such as the area covered by the finger, accelerometer, and gyroscope in order to further improve the system performance. Finally, we will also analyse the user experience in different acquisition settings through the analysis of the information acquired from Block5 to Block8 of the MobileTouchDB.

The MobileTouchDB database, experimental protocol, and benchmark evaluation of it has been submitted to the Conference on Computer Vision and Pattern Recognition Workshops (CVPRw) [Tolosana *et al.*, 2019b]. All these information will be available in GitHub very soon. Finally, we would also like to highlight that the application of MobileTouchDB can be also useful for other research lines beyond touchscreen biometric authentication, e.g.: *i)* user-dependent effects [Yager and Dunstone, 2010b], and development of user-dependent methods for handwriting recognition, *ii)* the neuromotor processes involved in writing over touchscreens [Ferrer *et al.*, 2018], *iii)* sensing factors in obtaining representative and clean touch interaction signals [Tolosana *et al.*, 2015d], *iv)* human-device interaction factors involving touchscreen signals, and development of improved interaction methods [Harbach *et al.*, 2016], and *v)* population statistics around touch interaction signals, and development of new methods aimed at recognis-

ing or serving particular population groups.

9.6. Chapter Summary and Conclusions

In this chapter we have evaluated the advantages and potential of incorporating touch biometrics to password-based mobile authentication systems. The new e-BioDigit database, which is described in Sec. 3.3.1, is used in the experiments reported in this chapter. Data was collected in two sessions with a time gap of at least three weeks between them for a total of 93 subjects. Handwritten numerical digits were acquired using the finger as the writing input on a Samsung Galaxy Note 10.1 general purpose tablet device.

For the new e-BioDigit database, we report a benchmark evaluation using two different state-of-the-art approaches: *i)* DTW in combination with the SFFS function selection algorithm, and *ii)* RNN deep learning technology. In addition, we perform a complete analysis of the touch biometric system regarding the discriminative power of each handwritten digit, and the robustness of our proposed approach when increasing the length of the password and the number of enrolment samples per user.

Our proposed approach achieves remarkable results with EERs ca. 4.0% when considering skilled forgeries, outperforming other traditional biometric verification traits such as the handwritten signature or graphical passwords on similar mobile scenarios. Additionally, we discuss specific details for the deployment or our proposed approach on current PIN- and OTP-based authentication systems.

Finally, the preliminary results obtained here using the novel MobileTouchDB database over a DTW baseline system prove the higher discriminative power of lower- and upper case letters, and symbols compared to numbers, with average absolute improvements ranging from 1-2% EERs. For future work, we expect to further reduce the EER through more advanced techniques based on feature selection and deep learning.

Part V

Conclusions

Chapter 10

Conclusions and Future Work

This final chapter brings together and summarises the main points and important results presented in this Dissertation with reference to the research objectives of Chapter 1. This Thesis is divided into five main parts. *Part I* concentrates on the problem statement and main contributions of this Dissertation. This part comprises Chapters 1, 2, 3, and 4. There are three experimental parts: *Part II*, *Part III*, and *Part IV*. *Part II* focuses on the new challenging and current signature verification scenarios, comprising Chapters 5, and 6. *Part III* describes the experimental work carried out in order to enhance traditional signature verification systems. This part is composed of Chapters 7 and 8. *Part IV* addresses the experimental work carried out for incorporating handwriting biometric information to traditional password-based authentication systems, including Chapter 9. Lastly, *Part V* concludes the Dissertation.

The **major contributions** made in this Thesis are:

- Analysis and adaptation of on-line signature verification systems to emerging scenarios such as finger input, device interoperability and mixed writing-input through robust preprocessing and feature selection techniques.
- An exhaustive experimental analysis of template update strategies for three very popular on-line signature verification approaches, extracting various practical findings related to the template aging effect in signature biometrics, and configuring time-adaptive improved versions of the considered baseline approaches overcoming to some extent the template aging.
- Exploring the potential of Deep Learning approaches for on-line signature and handwriting verification. We have proposed a novel end-to-end writer-independent on-line signature verification system based on Recurrent Neural Networks with a Siamese architecture, which has outperformed other state-of-the-art systems.
- Improvement of traditional signature verification systems through the incorporation of the signature complexity concept.

- Enhancement of traditional PIN and OTP authentication systems through the incorporation of handwriting biometric information as a second level of user authentication.
- Acquisition of new unprecedented handwriting and signature databases, making them freely available to the research community.
- Part of the research presented in this Thesis has been deployed successfully in a pilot project in which on-line signature verification will be used massively in the Spanish banking sector.

In what follows, we proceed to describe in more detail the particular findings corresponding to the major contributions.

10.1. Conclusions

Chapter 1 first introduced the basics of biometric systems, including properties, and biometric traits. Then, we focused on handwritten signature biometrics, which is the main topic of study in this Thesis, and the challenges and opportunities for it on emerging scenarios. Later on we summarised the success of DL techniques in many different biometric applications, which motivated the exploration carried out in this Dissertation. Then, we motivated the incorporation of handwriting biometric information in traditional password-based authentication approaches. We finished the chapter by stating the Thesis, giving an outline of the Dissertation, and summarising the research contributions originated from this work.

Chapter 2 first described each module of a traditional on-line signature verification systems as well as the two modalities considered, global systems (a.k.a. feature-based systems), and local systems (a.k.a. time functions-based systems). Then, the chapter reviewed related works of this Thesis: *i*) emerging scenarios for signature biometrics such as finger input, device interoperability, mixed writing-input, and signature template aging, *ii*) the importance and reasons for the success of DL together with a brief overview of the most famous DL architectures nowadays, and *iii*) advantages and limitations of recent touchscreen biometrics approaches.

Chapter 3 first gave an overview of the most relevant features of existing on-line signature databases, making special emphasis on the databases used in the experimental work of this Thesis. Then, we presented the new e-BioSign database as well as the extension of the ATVS On-Line Signature Long-Term database, acquired in this Thesis. Finally, we introduced the new e-BioDigit database acquired for the analysis of handwriting biometric information on password-based scenarios.

Chapter 4 first focused on traditional signature verification systems, describing the specific features and matching algorithm configurations considered in the experimental parts of the Thesis. Then, we concentrated on novel signature verification systems based on deep learning architectures. We first explained the basics of RNN systems and gave an overview of the main relevant studies. Then, the specific details of our proposed end-to-end writer-independent RNN signature verification systems were described.

Chapter 5 is the first experimental chapter of *Part II*, which focused on the new challenging and current signature verification scenarios. It first analysed the system performance of traditional signature verification systems on emerging scenarios such as finger input, device interoperability and mixed writing-input. Both Biosecure and e-BioSign databases were considered in the experimental work. Then, we proposed a two-stage approach based on robust preprocessing and feature selection techniques in order to alleviate the degradation of the system performance on these novel scenarios. The key findings of this chapter were:

- The high technological evolution and sensor quality improvement together with our proposed two-stage approach led to very competitive signature verification systems on device interoperability scenarios with an average EER of 11.9% and 1.8% EER for skilled and random forgery cases, respectively.

- A high variability is produced when using the finger as input. This is due to two main reasons: *i*) users who performed their signatures using closed letters (i.e., a, e, o, l, p, q, etc.) tended to perform much larger writing executions in comparison with other letters due to the lower precision they were able to achieve using the finger, and *ii*) users whose signatures were composed of a long name and surname (or two surnames) tended to simplify some parts of their signatures. As a result, a high degradation of the system performance was produced compared to the stylus case, with an average EER of 25.5% and 1.9% for skilled and random forgeries, respectively.

- A recommendation for the usage of signature recognition on mobile devices would be for the users to protect themselves from other people that could be watching while signing, as this is more feasible to do in a mobile scenario compared to an office scenario. This way skilled forgers might have access to the global shape of the signature but not to the dynamic information.

- An analysis of mixed writing-input scenarios, concluding that the main problem resides in the signatures acquired with the finger.

Part II of the Thesis also comprises Chapter 6. This chapter studied the novel scenario where the number of stored samples or templates per user can grow very fast, making it possible to train more robust statistical user models, improving the performance of the biometric systems, and in particular, reducing the template aging effect. We first introduced the methods studied in this work in order to reduce the template aging effect. Then, we described the three popular signature systems (DTW, HMM, and GMM) and the experimental protocol and results achieved using the ATVS Signature Long-Term Extended Database. The main contributions of this chapter were:

- We analysed the effect of template aging in on-line signature biometrics concluding that it has a significant impact in the system performance.

- In order to compensate for this template aging effect, an exhaustive experimental analysis of various template update strategies were carried out. For the case of HMM and GMM systems the optimal template update strategy would be to select all available training signatures or at least a few of them from several sessions in order to generate a more reliable user's template. For the DTW system the optimal one would be to consider a few training signatures (i.e., between 8 and 12) from sessions closer in time to the test.
- By incorporating the considered template update techniques, we demonstrated a significant improvement of performance with respect to the three baseline systems, hence we achieved a significant reduction of the template aging effect with similar results to the ideal case for random forgeries, and an average relative improvement of 61.9% EER for skilled forgeries.
- A final fusion of the three individual systems after applying the best resulting template update approach was carried out in order to further improve the recognition performance achieving an EER of 2.1% and 0.2% for skilled and random forgeries, respectively.

Chapter 7 and 8 in *Part III* proposed new ways to improve traditional signature verification systems. Concretely, Chapter 7 evaluated the potential of our proposed end-to-end writer-independent on-line signature verification system based on RNNs with a Siamese architecture. We first described the experimental protocol considered based on the BiosecurID database, and the results achieved on the development and evaluation datasets for both skilled and random forgery cases. The key contributions were:

- The first complete and successful framework on the use of multiple RNN systems (i.e., LSTM and GRU) for on-line handwritten signature verification considering both skilled and random forgery cases.
- Regarding the development stage, it is important to remark the different number of training iterations needed between normal (i.e., LSTM and GRU) and bidirectional schemes (i.e., BLSTM and BGRU). This showed the importance of considering both past and future contexts in order to train RNNs faster and also with a lower value of training cost. In addition, it is important to highlight the different number of training iterations between both LSTM and GRU RNNs as the GRU memory block is a simplified version of the LSTM memory block with fewer parameters to train.
- For the scenario of using just one training signature per user, our Proposed BLSTM System achieved a 5.60% EER for skilled forgeries, which corresponded to an absolute improvement of 4.57% EER compared to the 10.17% EER achieved for the highly competitive DTW System. It is important to remark that our proposed system outperformed the result obtained with the DTW-based System for the case of using 4 training signatures even just using one signature (i.e., 5.60% vs 7.75% EER).
- When training for both skilled and random forgery cases, our Proposed BLSTM System achieved for the case of using 4 training signatures values of 5.50% and 3.00% EER for

skilled and random forgeries, respectively. Despite the very good results achieved for skilled forgeries, the 3.00% EER obtained for random forgeries could not outperform the 0.5% EER obtained using the DTW-based System. A possible solution is to perform two consecutive stages similar to [Gomez-Barrero *et al.*, 2015]: *i*) first stage based on DTW optimised for rejecting random forgeries, and *ii*) our Proposed RNN System in order to reject the remaining skilled forgeries.

Chapter 8 proposed on-line signature verification systems adapted to the signature complexity level of the user. Despite all the studies performed in the on-line signature trait, none of them exploited, as far as we know, the concept of complexity for the development of more robust and accurate on-line signature verification systems. This chapter further investigated this line considering both stylus- and finger-based scenarios. We first presented our proposed complexity-based on-line signature verification system. Then, we introduced the experimental protocol based on both e-BioSign and BiosecurID databases. Finally, we conducted the experiments for the evaluation of our proposed signature complexity detector, and the final proposed signature verification system which selected the optimal feature subset for each complexity level. The main findings attained in this chapter were:

- The proposed signature complexity detector showed to be very effective despite of its simplicity. Signatures longer in time and with an appearance more similar to handwriting were labelled as high-complexity signatures whereas signatures shorter in time and with generally simple flourish with no legible information were labelled as low-complexity signatures.
- Our proposed complexity-based signature verification system outperformed previous studies through the selection of the optimal subset of features for each complexity level and input scenario (i.e., stylus and finger). Analysing the results obtained for the stylus scenario, our Proposed System achieved for the BiosecurID database an average absolute improvement of 2.5% EER for skilled forgeries compared to the Baseline System (the case where the local features are fixed to all complexity levels). Analysing the results obtained for the finger scenario, our Proposed System achieved an absolute improvement of 5.6% EER for the most challenging users (i.e., users with low complexity level).

Finally, Chapter 9 in *Part IV* evaluated the incorporation of handwriting biometric information to traditional authentication systems based on passwords, asking the users to draw each digit of the password on the touchscreen instead of typing them as usual. This chapter performed a complete analysis of our proposed biometric system in terms of the discriminative power of each handwritten digit and the robustness when increasing the length of the password and the number of enrolment samples. The new e-BioDigit database was considered in the experimental work of the chapter. Concretely, the major contributions of this chapter were:

- A baseline system composed of a set of simple and fixed local features for all numerical digits in order to make our experimental work easily reproducible.

- An study of the best features for each handwritten numerical digit through the SFFS algorithm on the e-BioDigit development dataset.
- A complete analysis of our proposed touch biometric system regarding the most discriminative handwritten digits and how robust the system is when increasing the length of the password and the number of enrolment samples per user.
- Our proposed approach achieved remarkable results with EERs ca. 4.0% when considering skilled forgeries, outperforming other traditional biometric verification traits such as the handwritten signature or graphical passwords on similar mobile scenarios.
- Specific details for the deployment or our proposed approach on current PIN- and OTP-based authentication systems.

10.2. Future Work

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

- The emerging finger input scenario allows a high deployment of signature authentication technology on a daily basis. However, the preliminary analysis carried out in this Dissertation has shown the challenge of this new scenario. A more in-depth analysis of the finger input scenario needs to be done in order to understand the higher variability observed in this Thesis compared to the traditional stylus scenario. A long-term study should be conducted focusing on how the technology advancements and age population sectors are affected on this emerging scenario. Finally, new core matchers unlike the traditional ones should be designed in order to better adapt the higher variability observed.
- The effect of template aging has shown to have a significant impact in the system performance. The analysis carried out in this Dissertation should be extended to other state-of-the-art signature verification systems based on, for example, deep learning [Tolosana *et al.*, 2018c; Zhang *et al.*, 2017]. In addition, more efficient techniques should be studied in order to retrain user models with new samples instead of generating them from scratch every time new samples are available. This aspect it is very important for capacity storage as the number of enrolled samples can further increase with time.
- Deep learning approaches have outperformed several state-of-the-art signature verification systems for many different cases, such as skilled forgeries with low number of training signatures. However, the study carried out in this Dissertation has been only a small demonstration of the potential of this groundbreaking technology. New DL architectures should be proposed in order to *i*) better generalise against different levels of forgeries, and *ii*) feeding the network with a set of training signatures instead of a single one in order to better model the user and take the final decision. This can be carried out using our novel

DeepSignDB on-line handwritten signature database composed of 1526 different users and more than 70K signatures [Tolosana *et al.*, 2019a].

- The signature complexity concept has proved to be very important for the system performance. Very good results have been achieved using our proposed approach. However, the study conducted on this Thesis was very basic. A more in-depth study should be carried out proposing more robust signature complexity detectors [Houmani and Garcia-Salicetti, 2016; Houmani *et al.*, 2008; Lim and Yuen, 2016], and also studying the signature complexity over others state-of-the-art signature systems.

- The incorporation of handwriting biometric information to traditional password-based authentication systems has proved to further increase the security against impostors through a user-friendly interface. However, the study carried out in this Dissertation should be further investigated through the following lines: *i*) considering not only numbers but also lower- and upper-case letters, and especial symbols, *ii*) proposing different configurations for introducing the passwords in order to provide more robust systems against shoulder-surfing and smudge attacks, and *iii*) enhancing the authentication systems through the incorporation of deep learning techniques [Tolosana *et al.*, 2018c; Zhang *et al.*, 2017]. This can be carried out using our novel unsupervised MobileTouchDB database [Tolosana *et al.*, 2019b], which contains more than 64K on-line character samples performed by 218 users, using 94 different smartphone models, with an average of 314 samples per user.

Apéndice A

Resumen Extendido de la Tesis

Aproximaciones Disruptivas para la Mejora de
Sistemas de Autenticación basados en Firma y Escritura Manuscrita

A.1. Resumen

La firma manuscrita es uno de los rasgos biométricos más aceptados en la sociedad debido a su amplio uso en el ámbito legal y de las finanzas. Sin embargo, ¿está la tecnología de firma realmente adaptada a los escenarios modernos? Con el despliegue masivo de los dispositivos móviles tales como smartphones y tablets, nuevos e interesantes escenarios han surgido más allá del tradicional escenario de oficina bancaria en el que las firmas son capturadas en condiciones muy controladas. Además, a pesar de la gran evolución tecnológica producida en los últimos años, y en concreto el gran éxito de las tecnologías basadas en deep learning gracias al uso de GPUs, el nucleo algorítmico de la mayoría de los sistemas de verificación de firma manuscrita dinámica sigue siendo el mismo que el de hace 20 años. La pregunta es, ¿por qué las aproximaciones basadas en deep learning no han mejorado por el momento a las tradicionales como ocurre en otros campos de investigación?

La última motivación de la Tesis está relacionada con los sistemas basados en contraseña. Tradicionalmente, los dos enfoques de autenticación de usuarios más frecuentes han sido PIN y OTP. Sin embargo, y a pesar de la gran popularidad y despliegue de los mismos en escenarios de la vida diaria, muchos estudios han resaltado las debilidades de estos enfoques al ser muy fáciles de adivinar o robar (i.e., mediante ataques del tipo shoulder-surfing y smudge). Sin embargo, ¿es posible aumentar la seguridad de estos sistemas que utilizamos en el día a día al mismo tiempo que brindamos una buena experiencia a los usuarios?

Con el objetivo de encontrar las respuestas a estas preguntas, esta Tesis se centra principalmente en el análisis de las nuevas oportunidades que presentan estos nuevos escenarios y tecnologías en el ámbito de la firma manuscrita, así como los retos que deben ser abordados para lograr resultados en el estado del arte.

Esta Disertación consta de cinco partes diferentes. La primera parte se centra en la declara-

ción del problema y de las principales contribuciones de la Tesis, así como un amplio resumen del estado del arte. Los capítulos experimentales se dividen en tres partes, *Parte II*, *Parte III* y *Parte IV*. Por último, la *Parte V* concluye la Tesis.

La *Parte I* primero introduce los conceptos básicos de la biometría, centrándose en la firma manuscrita dinámica, que es el tema principal de estudio en esta Tesis, y los desafíos y oportunidades que ofrece este rasgo biometríco a lo largo de una visión exhaustiva del estado del arte. Posteriormente, se describen las características más relevantes de las bases de datos de firma manuscrita dinámica, haciendo especial hincapié en todas las bases de datos adquiridas durante la realización de la Tesis. Finalmente, la *Parte I* concluye explicando primero los detalles más específicos de los sistemas tradicionales de verificación de firma considerados en las partes experimentales de esta Tesis, y finalmente, las nuevas aproximaciones propuestas en esta Tesis (i.e., deep learning y complejidad de la firma).

La primera parte experimental (*Parte II* de esta Disertación) comienza analizando el rendimiento de los sistemas de verificación de firma tradicionales en escenarios emergentes como la adquisición de firmas con el dedo, la interoperabilidad de dispositivos y el uso de múltiples útiles de escritura. Debido a la alta degradación del rendimiento del sistema en estos nuevos escenarios, se propone en esta Tesis un enfoque basado en técnicas de preprocesado de los datos y selección de las características más robustas. Posteriormente se estudia el nuevo escenario en el que el número de muestras o patrones biométricos almacenados por usuario puede aumentar con el paso del tiempo a través de múltiples sesiones de captura, lo que permite entrenar modelos estadísticos más robustos, mejorar el rendimiento, y en particular, reducir el efecto del envejecimiento de los patrones biométricos. La investigación realizada en esta parte tiene como objetivo responder a las siguientes preguntas: ¿Cómo se ve afectado el rendimiento del sistema en estos nuevos escenarios? ¿Qué enfoque debemos considerar para superar estos retos?

En la segunda parte experimental (*Parte III* de esta Disertación) se proponen nuevas formas de mejorar los sistemas tradicionales de firma manuscrita. Concretamente, primero se evalúa el potencial de deep learning a través del diseño de una nueva arquitectura (Siamesa) más adaptada a la tarea de verificación de firma. Finalmente se analiza el concepto de complejidad en el ámbito de la firma, proponiendo mejoras respecto a los sistemas tradicionales a través de la selección de las características más robustas para cada nivel de complejidad de la firma.

Finalmente, la *Parte IV* de esta Disertación evalúa la incorporación de información biométrica de escritura en los sistemas de autenticación tradicionales basados en contraseñas, solicitando al usuario que dibuje cada dígito de la contraseña en la pantalla táctil del dispositivo, en lugar de teclearlos como de costumbre.

La investigación llevada a cabo en esta Disertación ha logrado las siguientes contribuciones: *i*) el análisis y adaptación de los sistemas de verificación de firma dinámica a los escenarios emergentes, tales como la adquisición de firmas con el dedo, la interoperabilidad de dispositivos y útiles de escritura gracias a las técnicas de preprocesado de los datos y selección de características más robustas, *ii*) un análisis experimental exhaustivo de las estrategias de actualización de los modelos o patrones biométricos para tres sistemas populares en el mundo de la firma manus-

crita, extrayendo varias conclusiones prácticas relacionados con el efecto del envejecimiento de los modelos y patrones biométricos, y proponiendo configuraciones adaptadas en el tiempo que permitan superar dicho envejecimiento de los sistemas, *iii*) explorar el potencial de los enfoques basados en deep learning en el ámbito de la firma manuscrita dinámica. El sistema propuesto en esta Tesis está basado en una arquitectura Siamesa, más adaptada a la tarea de autenticación de firmas, y que ha superado en rendimiento a otros sistemas en el estado del arte, *iv*) mejora de los sistemas tradicionales de verificación de firma a través de la incorporación del concepto de complejidad de la firma, *v*) mejora de los sistemas tradicionales de autenticación de PIN y OTP mediante la incorporación de información biométrica de escritura como un segundo nivel de autenticación del usuario, *vi*) adquisición de nuevas bases de datos de escritura y firma manuscrita en escenarios sin precedentes, así como su libre disposición para la comunidad científica, y *vii*) gran parte de la investigación presentada en esta Tesis se ha implementado con éxito en un proyecto piloto. Dicha tecnología será utilizada de forma masiva en el sector bancario español en un futuro cercano.

A.2. Conclusiones

Esta Tesis se ha centrado en el análisis y adaptación de los sistemas de firma manuscrita dinámica a los escenarios actuales, las nuevas oportunidades que surgen para este rasgo biométrico con la creciente cantidad de datos y recursos hardware disponibles, así como la mejora de los sistemas de autenticación tradicionales basados en PIN y OTP por medio de la incorporación de información biométrica de escritura. En concreto, esta Disertación se encuentra estructurada en 5 partes principales. La primera parte (*Parte I*) se centra en la definición del problema a estudiar y su contexto. Esta primera parte engloba los Capítulos 1, 2, 3, y 4. Posteriormente, se definen las 3 partes experimentales de esta Tesis: *Parte II*, *Parte III* y *Parte IV*. La *Parte II* estudia el comportamiento de los sistemas de firma manuscrita en los nuevos escenarios emergentes, y engloba los Capítulos 5 y 6. La *Parte III* describe las nuevas arquitecturas propuestas con el objetivo de mejorar los sistemas tradicionales de firma manuscrita dinámica. Esta parte se compone de los Capítulos 7 y 8. La última parte experimental (*Parte IV*) aborda la mejora de los sistemas tradicionales de autenticación basados en PIN y OTP por medio de la incorporación de información biométrica de escritura sobre los dispositivos móviles. Esta última parte experimental engloba el Capítulo 9. Finalmente, la *Parte V* concluye la Disertación.

El Capítulo 1 introdujo los fundamentos básicos de los sistemas biométricos, sus propiedades y algunos de los rasgos biométricos más utilizados hoy en día. Posteriormente, se estudió los sistemas biométricos basados en firma manuscrita dinámica, que constituye el tema principal de estudio de esta Tesis. Más tarde, se hizo hincapié en el gran éxito actual de las tecnologías basadas en deep learning, que sirvió de motivación para explorar su potencial en el ámbito de la firma manuscrita dinámica. Tras esto, se destacaron las razones que motivaron la incorporación de información biométrica de escritura en los sistemas tradicionales de autenticación basados en contraseñas. El capítulo concluye con la motivación de la Tesis, y las contribuciones fruto de

esta Disertación.

El Capítulo 2 describió en primer lugar cada uno de los módulos de los que se compone los sistemas tradicionales de firma manuscrita dinámica, así como las dos grandes vertientes existentes: *i*) sistemas basados en características (a.k.a. sistemas globales) y *ii*) sistemas basados en funciones temporales (a.k.a. sistemas locales). Posteriormente, se resumió los trabajos en el estado del arte en las temáticas de estudio de esta Tesis.

El Capítulo 3 resumió las principales características de las bases de datos de firma manuscrita dinámica más utilizadas hoy en día, haciendo especial hincapié en las bases de datos utilizadas en las partes experimentales de esta Tesis. Posteriormente, se describieron las nuevas bases de datos capturadas durante el periodo de realización de la Tesis: e-BioSign, así como la extensión de la base de datos ATVS On-Line Signature Long-Term. Finalmente, se presentó la base de datos e-BioDigit, adquirida con el objetivo de estudiar las ventajas de la incorporación de escritura manuscrita en los sistemas tradicionales de autenticación basados en contraseñas.

El Capítulo 4 describió los fundamentos básicos de cada uno de los módulos de los que se componen los sistemas basados en redes neuronales recurrentes (RNNs), así como los principales trabajos que hacen uso de ellas. Finalmente, se introdujo el sistema de firma manuscrita dinámica propuesto en esta Tesis basado en RNNs con una arquitectura Siamesa cuyo principal objetivo es aprender las similitudes y diferencias entre pares de firmas genuinas e impostoras.

El Capítulo 5 constituye el primer capítulo experimental de la *Parte II*, y se centra en el análisis de los nuevos escenarios de firma manuscrita. Concretamente, en primer lugar se estudió el rendimiento de los sistemas en escenarios donde la firma es capturada utilizando el dedo como útil de escritura, múltiples dispositivos de capturada y útiles de escritura (dedo/stylus) en las fases de entrenamiento y testeo de los mismos. Para ello, las bases de datos Biosecure y e-BioSign fueron utilizadas en la parte experimental. Tras el análisis preliminar de los sistemas en estos nuevos escenarios de captura, se propuso una aproximación basada en 2 etapas con el objetivo de reducir la degradación del rendimiento de los sistemas. En primer lugar, una etapa de preprocesado robusta frente a los cambios producidos en la adquisición, a la que siguió una etapa basada en la selección óptima de las características ante los nuevos escenarios descritos. Las conclusiones más importante de este capítulo fueron:

- El gran desarrollo tecnológico y mejora de la calidad de los sensores junto con el enfoque propuesto en esta Tesis han producido como resultado sistemas de verificación de firma muy competitivos en escenarios de interoperabilidad de dispositivos con un EER promedio de 11.9 % y 1.8 % para falsificaciones de tipo skilled y random, respectivamente.

- Se ha observado una gran variabilidad en los escenarios en los que el dedo se ha utilizado como útil de escritura. Esto se debe principalmente a dos razones: *i*) los usuarios que realizaron sus firmas con letras compuestas por trazos cerrados (i.e., a, e, o, l, p, q, etc.) tendían a realizar ejecuciones de escritura mucho más grandes en comparación con el resto de letras debido a la menor precisión que podían lograr usando el dedo, y *ii*) los usuarios cuyas firmas estaban compuestas de un nombre largo y un apellido (o dos apellidos) tendían

a simplificar algunas partes de sus firmas debido a las limitaciones de espacio. Esto ha supuesto una alta degradación del rendimiento del sistema en comparación con el caso tradicional de utilizar el stylus como útil de escritura, con un EER promedio de 25.5 % y 1.9 % para falsificaciones skilled y random, respectivamente.

- Una posible recomendación para el usuario final de cara a mejorar la seguridad frente a ataques sería la de intentar protegerse del resto de usuarios que puedan observar o incluso grabar el proceso de captura de su firma, ya que esto es más factible en un escenario móvil en comparación con un escenario de oficina. De esta manera, los falsificadores expertos podrían tener acceso a la imagen de la firma pero no a la información dinámica.

- El análisis de los escenarios con múltiples útiles de escritura en el entrenamiento y testeo de los sistemas concluye que el principal problema reside en la captura de la firma con el dedo, más que en el propio escenario en sí.

La *Parte II* de la Tesis se compone también del Capítulo 6. Este capítulo estudió el novedoso escenario en el que el número de muestras o patrones biométricos por usuario puede crecer con el paso del tiempo a través de múltiples sesiones de captura, lo que hace posible entrenar modelos estadísticos más robustos, permitiendo una mejora del rendimiento, y en particular, reduciendo el efecto de envejecimiento de los modelos. En este capítulo se introdujo en primer lugar las distintas aproximaciones estudiadas en esta Tesis para reducir el efecto de envejecimiento de los modelos biométricos. A continuación, se describieron los tres sistemas de verificación de firma propuestos (DTW, HMM y GMM), así como el protocolo experimental y los resultados obtenidos a través del uso de la base de datos ATVS Signature Long-Term Extended Database. Las principales aportaciones de este capítulo fueron:

- Análisis del efecto de envejecimiento de los modelos y patrones biométricos en los sistemas de firma manuscrita dinámica, concluyendo que tiene un elevado impacto en el rendimiento del sistema.

- Con el objetivo de compensar el efecto de envejecimiento, se realizó un exhaustivo análisis experimental de varias estrategias de actualización de los modelos. Para el caso de los sistemas basados en HMM y GMM, la estrategia optima de actualización consistió en la selección de todas las muestras o patrones biométricos disponibles, o al menos algunas de ellas pero procedentes de al menos varias sesiones de captura con el objetivo de modelar mejor la variabilidad del usuario. Para el caso del sistema basado en DTW, la estrategia óptima consistió en utilizar entre 8 y 12 muestras o patrones biométricos procedentes de sesiones de captura cercanas en el tiempo a la muestra de test.

- Aplicando las actualizaciones propuestas, se consiguió una mejora significativa del rendimiento en comparación con los sistemas tradicionales basados únicamente en el uso de las muestras o patrones biométricos adquiridos en la primera etapa de registro. En concreto, para el caso de falsificaciones de tipo random, los resultados conseguidos fueron muy

similares al caso ideal (i.e., el caso de considerar muestras de la misma sesión en el entrenamiento y testeo de los sistemas), mientras que para el caso de falsificaciones de tipo skilled, se consiguió una mejorar relativa en media del 61.9 % EER.

- Finalmente se realizó una fusión de los tres sistemas estudiados tras aplicar las políticas de actualización propuestas con el objetivo de mejorar el rendimiento de los sistemas, alcanzando un 2.1 % y 0.2 % EER para falsificaciones de tipo skilled y random, respectivamente.

Los Capítulos 7 y 8 de la *Parte III* se centran en las nuevas aproximaciones estudiadas para mejorar el rendimiento de los sistemas tradicionales de firma manuscrita dinámica. En concreto, el Capítulo 7 estudia el potencial de nuestro sistema propuesto basado en redes neuronales recurrentes con una arquitectura Siamesa. En primer lugar se describió el protocolo experimental propuesto, en el que se utilizó la base de datos BiosecurID, para finalizar analizando los resultados obtenidos tanto en la etapa de desarrollo de los sistemas, como en su posterior evaluación utilizando usuarios totalmente distintos. Ambos tipos de falsificaciones, skilled y random, fueron considerados en los experimentos. Las contribuciones claves fueron:

- El primer estudio detallado y completo que demuestra el gran potencial de múltiples sistemas basados en RNNs (i.e., LSTM y GRU) en el campo de verificación de firma manuscrita dinámica, tanto para falsificaciones de tipo skilled como random.

- Con respecto a la etapa de desarrollo de los sistemas, es importante destacar la gran diferencia existente en el número de épocas de entrenamiento entre los esquemas normales (i.e., LSTM y GRU) y bidireccionales (i.e., BLSTM y BGRU). Estos resultados demuestran la importancia de considerar tanto el contexto pasado como futuro para entrenar los sistemas RNN de manera más rápida al mismo tiempo que se consigue un valor inferior de la función de coste. Además, es importante resaltar la variación que existe en el número de épocas de entrenamiento entre los sistemas LSTM y GRU, ya que el bloque de memoria GRU es una versión simplificada del bloque de memoria LSTM, con menos parámetros que entrenar.

- Para el escenario en el que solamente se dispone de una firma de entrenamiento por usuario, nuestro sistema BLSTM propuesto logró un EER de 5,60 % para falsificaciones de tipo skilled, consiguiendo así una mejora absoluta de 4,57 % EER en comparación con el 10,17 % EER logrado para el sistema DTW. Es importante destacar que nuestro sistema propuesto incluso superó el resultado obtenido por el sistema DTW para el caso de utilizar 4 firmas de entrenamiento en lugar de solo una (i.e., 5.60 % EER frente a 7.75 % EER).

- Para el caso de entrenar los sistemas para ambos tipos de falsificaciones, skilled y random, nuestro sistema BLSTM propuesto logró alcanzar resultados de 5.50 % y 3.00 % EER para falsificaciones de tipo skilled y random, respectivamente. A pesar de los buenos resultados logrados para las falsificaciones de tipo skilled, el EER del 3.00 % obtenido para las falsificaciones de tipo random no pudo superar el 0.5 % EER obtenido por el sistema DTW.

Una posible solución sería considerar un sistema compuesto por dos etapas consecutivas, similar a [Gomez-Barrero et al., 2015]: *i*) primera etapa basada en un sistema DTW optimizado para rechazar falsificaciones de tipo random, y *ii*) nuestro sistema RNN propuesto para rechazar las falsificaciones de tipo skilled.

El Capítulo 8 propuso sistemas de verificación de firma dinámica adaptados al nivel de complejidad de la firma. A pesar de todos los estudios realizados en firma manuscrita dinámica, ninguno de ellos ha explotado, hasta donde sabemos, el concepto de complejidad de la firma para el desarrollo de sistemas de verificación de firma más robustos y precisos. Este capítulo investigó el efecto de la complejidad en escenarios de captura basados en el stylus y el dedo. En primer lugar, se presentó nuestro sistema de verificación de firma manuscrita dinámica propuesto basado en complejidad. Posteriormente, se describió el protocolo experimental utilizado, basado en las bases de datos e-BioSign y BiosecurID. Finalmente, se realizaron los experimentos que permitieron la evaluación de nuestro detector de complejidad de firma propuesto y el nuevo sistema de verificación de firma propuesto, en el que se selecciona para cada nivel de complejidad el subconjunto de características más óptimo. Los principales hallazgos alcanzados en este capítulo fueron:

- El detector de complejidad de firma propuesto demostró ser muy efectivo a pesar de su simplicidad. Las firmas más largas en el tiempo y con una apariencia más similar a la escritura fueron etiquetadas como firmas de alta complejidad, mientras que las firmas más cortas en el tiempo y con una apariencia generalmente más simple y sin información legible fueron etiquetadas como firmas de baja complejidad.
- El sistema de verificación de firma propuesto basado en la complejidad consiguió mejorar los resultados obtenidos en los estudios previos. La selección de un subconjunto óptimo de características para cada nivel de complejidad y escenario de captura (i.e., stylus y dedo) demostró proporcionar sistemas más robustos frente ataques. En el caso de utilizar el stylus como útil de escritura, nuestro sistema propuesto logró para la base de datos BiosecurID una mejora absoluta promedio de 2.5 % EER para falsificaciones de tipo skilled en comparación con el sistema tradicional (el caso donde el mismo vector de características es utilizado para todos los niveles de complejidad). Para el escenario de captura con el dedo, nuestro sistema propuesto logró una mejora absoluta de 5.6 % EER para los usuarios más sencillos de falsificar (i.e., usuarios con un nivel de complejidad bajo).

Finalmente, la *Parte IV* de esta Tesis doctoral se compone del Capítulo 9, en el que se evalúa la incorporación de información biométrica de escritura manuscrita en los sistemas tradicionales de autenticación basados en contraseñas sobre escenarios móviles. Para ello, se le solicita al usuario que dibuje cada dígito de la contraseña en la pantalla táctil de su dispositivo móvil, en lugar de teclearlos como de costumbre. Este capítulo realizó un profundo análisis de nuestro sistema biométrico propuesto, tanto en términos del poder discriminativo de cada dígito, como de la robustez del sistema a medida que se aumenta la longitud de la contraseña y el número de

muestras o patrones biométricos de entrenamiento disponibles. La nueva base de datos e-BioDigit fue considerada en el trabajo experimental de este capítulo. Las principales aportaciones de este capítulo fueron:

- Un sistema de referencia compuesto por un conjunto de características simple y fijo para todos los dígitos con el fin de hacer que nuestro trabajo experimental sea fácilmente reproducible.
- Un estudio de las mejores características seleccionadas para cada dígito a través del algoritmo de selección de características SFFS en el conjunto de datos de desarrollo.
- Un análisis exhaustivo del sistema biométrico táctil propuesto con respecto a qué dígitos son más discriminativos y cómo de robusto es el sistema a medida que se incrementa la longitud de la contraseña y el número de muestras de entrenamiento disponibles por usuario.
- Nuestro enfoque propuesto logró muy buenos resultados con EERs sobre el 4.0 % para falsificaciónes de tipo skilled, superando así a otros rasgos biométricos tradicionales, tales como la firma manuscrita o las contraseñas gráficas en escenarios móviles similares.
- Detalles específicos para la implementación de nuestro enfoque propuesto en escenarios reales de autenticación basados en PIN y OTP.

A.3. Líneas de Trabajo Futuro

Se proponen las siguientes líneas de trabajo futuro relacionadas con el trabajo desarrollado en esta Tesis:

- El escenario de adquisición de firmas en el que se utiliza nuestro propio dedo como útil de escritura hace posible un alto despliegue de la tecnología de autenticación de firma en la vida diaria. Sin embargo, el análisis preliminar realizado en esta Disertación ha puesto en evidencia el reto que supone este nuevo escenario de cara a mantener las buenas tasas de reconocimiento. Por este motivo, se debe realizar un análisis más exhaustivo de este nuevo escenario de captura, comprendiendo así los verdaderos motivos que originan esta alta variabilidad intra-usuario en comparación con el escenario tradicional de usar el stylus como útil de escritura. Por lo tanto, pensamos que se debe realizar un estudio a largo plazo centrado en dos objetivos fundamentales: *i*) el rápido avance tecnológico que permita mejorar la sensación de usabilidad por parte del usuario, y *ii*) cómo se ve afectado el rendimiento de los sistemas para distintos sectores de edad de la población, ya que la mayor parte de los jovenes hoy en día han aprendido a interactuar desde pequeños con los dispositivos móviles. Finalmente, es necesario desarrollar nuevos nucleos de comparación distintos de los tradicionales que sean capaces de adaptarse mejor a las variaciones intra-usuario observadas.

- El efecto del envejecimiento de los patrones biométricos ha demostrado tener un elevado impacto en el rendimiento de los sistemas de firma manuscrita dinámica. El análisis realizado en esta Disertación debe extenderse a otros sistemas de verificación de firma en el estado del arte, por ejemplo, en sistemas basados en deep learning [Tolosana *et al.*, 2018c; Zhang *et al.*, 2017]. Además, deben estudiarse técnicas más eficientes que permitan reentrenar el modelo de un usuario con nuevos patrones biométricos disponibles sin necesidad de disponer de todos los patrones biométricos anteriores. Este aspecto es muy importante en términos de capacidad de almacenamiento para entornos reales con millones de usuarios.

- Los enfoques basados en deep learning han superado ya algunos de los sistemas de verificación de firma y escritura más potentes para algunos escenarios, por ejemplo, para falsificaciones de tipo skilled con un bajo número de firmas de entrenamiento. Sin embargo, el estudio realizado en esta Disertación ha sido sólo una pequeña demostración del potencial de esta tecnología. Nuevas arquitecturas deben ser estudiadas con el objetivo de: *i*) generalizar mejor contra diferentes niveles de falsificaciones, y *ii*) alimentar los sistemas con un conjunto de firmas de entrenamiento en lugar de con una sola firma para modelar mejor al usuario y tomar una decisión final más robusta. Estas mejoras pueden llevarse a cabo gracias a la reciente base de datos construida al final de esta Tesis, DeepSignDB, compuesta de un total de 1526 usuarios y más de 70 mil firmas [Tolosana *et al.*, 2019a].

- El concepto de complejidad de la firma ha demostrado ser muy importante para el rendimiento del sistema. Se han logrado muy buenos resultados utilizando el enfoque propuesto. Sin embargo, el estudio realizado en esta Tesis es sólo el comienzo de un largo camino. Todavía queda mucho por mejorar, concretamente, se debe realizar un estudio más exhaustivo que proponga detectores de complejidad de firma más robustos [Houmani and Garcia-Salicetti, 2016; Houmani *et al.*, 2008; Lim and Yuen, 2016], y que evalue también el efecto de la complejidad en sistemas de última generación, por ejemplo basados en deep learning.

- La incorporación de información biométrica de escritura en los sistemas tradicionales de autenticación basados en contraseñas ha demostrado una mejora considerable de la seguridad frente a ataques a través del uso de una interfaz fácil de usar. Sin embargo, el estudio realizado en esta Disertación debe investigarse más a fondo a través de las siguientes líneas de trabajo: *i*) incorporando a las contraseñas no solo números sino también letras mayúsculas y minúsculas, y símbolos especiales, *ii*) proponiendo diferentes configuraciones para introducir las contraseñas con el fin de proporcionar sistemas más robustos contra los ataques shoulder-surfing y smudge, y *iii*) mejorar los sistemas de autenticación mediante la incorporación de técnicas de deep learning [Tolosana *et al.*, 2018c; Zhang *et al.*, 2017]. Estas mejoras pueden llevarse a cabo gracias a la reciente base de datos capturada al final de esta Tesis, MobileTouchDB, compuesta de más de 64 mil muestras de caracteres procedentes de 218 usuarios distintos, y utilizando hasta 94 dispositivos móviles diferentes [Tolosana *et al.*, 2019b].

References

Eab CiTeR European Cooperative Identification Technology Research Consortium, 2015. 5

European Association for Biometrics (EAB), 2017. http://eab.org. 6

K. Ahrabian and B. Babaali. On Usage of Autoencoders and Siamese Networks for Online Handwritten Signature Verification. *arXiv preprint arXiv:1712.02781*, 2017. 39, 45

A. M. Alimi. Beta neuro-fuzzy systems. *TASK Quarterly Journal, Special Issue on"Neural Networks*, 7(1):23–41, 2003. 8

F. Alonso-Fernandez, M. Fairhurst, J. Fierrez, and J. Ortega-Garcia. Impact of Signature Legibility and Signature Type in Off-Line Signature Verification. In *Proc. Biometrics Symposium, BSYM*, pages 1–6, 2007. 35

F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez. Quality-Based Conditional Processing in Multi-Biometrics: application to Sensor Interoperability. *IEEE Transactions on System, Man, and Cybernetics: Part A*, 40(6):1168–1179, 2010. 25, 99

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. Intl. Workshop on Biometric Recognition Systems, IWBRS*, volume 3781, pages 180–187, 2005. 13, 32

J. Angulo and E. Wastlund. Exploring Touch-Screen Biometrics for User Identification on Smart Phones. *J. Camenisch, B. Crispo, S. Fischer-Hubner, R. Leenes, G. Russello (Eds.), Privacy and Identity Management for Life, Springer*, pages 130–143, 2011. 12, 39, 40, 135

ANSI/NIST. NIST ITL American National Standards for Biometrics, 2009. http://fingerprint.nist.gov/standard/. 6

M. Antal and A. Bandi. Finger or Stylus: Their Impact on the Performance of On-line Signature Verification Systems. *MACRo 2015*, 2(1):11–22, 2017. 23, 33, 44, 45

A. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. Smith. Smudge Attacks on Smartphone Touch Screens. In *Proc. of the 4th USENIX Conference on Offensive Technologies*, pages 1–7, 2010. 12

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba. Learning Deep Features for Discriminative Localization. In *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 4

F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio. Theano: New Features and Speed Improvements. *In Proc. Advances in Neural Information Processing Systems*, 2012. 105

BBfor2, 2010. BBfor2: Bayesian Biometrics for Forensics, FP7-ITN-2008-238803. (http://www.bbfor2.net/). 5

BC. Biometrics Consortium, 2009. (http://www.biometrics.org/). 6

R. Beveridge, J. Phillips, D. Bolme, B. Draper, G. Givens, Y. Lui, M. Teli, H. Zhang, W. Scruggs, K. Bowyer, P. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras (PaSC). In *Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013. 5

BF. The Biometric Foundation, 2009. (http://www.biometricfoundation.org/). 6

B. Bhanu and A. Kumar, editors. *Deep Learning for Biometrics*. Springer, 2017. 4, 11

BI. Biometrics Institute, 2009. (http://www.biometricsinstitute.org/). 6

BioAPI. The BioAPI Consortium, 2009. http://www.bioapi.org. 6

BioSec, 2004. Biometrics and Security, FP6 IP IST-2002-001766. (http://www.biosec.org/). 5

Biosecure, 2004. Biometrics for Secure Authentication, FP6 NoE IST-2002-507634. (http://www.biosecure.info/). 5

R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez. Performance Evaluation of Handwritten Signature Recognition in Mobile Environments. *IET Biometrics*, 3:139–146(7), 2014. 23, 32, 44, 45, 76, 83, 131

J. Bonneau, C. Herley, P. Oorschot, and F. Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *Proc. IEEE Symposium on Security and Privacy*, pages 553–567, 2012. 3, 12, 14

T. Boult, D. Maltoni, S. Li, and M. Vatsa, editors. *Proceedings of IEEE International Joint Conference on Biometrics (IJCB)*, 2014. 5

K. Bowyerin and M. Burge. *Handbook of Iris Recognition*. Springer, 2016. 6

J. Brault and R. Plamondon. A Complexity Measure of Handwritten Curves: Modeling of Dynamic Signature Forgery. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:400–413, 1993. 35

J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, and R. Shah. Signature Verification Using a Siamese Time Delay Neural Network. *In Proc. Advances in Neural Information Processing Systems*, 1993. 25, 61

D. Buschek, A. D. Luca, and F. Alt. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proc. of the CHI Conference on Human Factors in Computing Systems*, pages 1393–1402, 2015a. 40, 41

D. Buschek, A. D. Luca, and F. Alt. There is more to Typing than Speed: Expressive Mobile Touch Keyboards via Dynamic Font Personalisation. In *Proc. of the International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 125–130, 2015b. 40, 41

BWG. Communications-electronics Security Group (CESG) – Biometric Working Group (BWG) (UK government), 2009. http://www.cesg.gov.uk/policy_technologies/biometrics/index.shtml. 6

K. Cao and A. Jain. Automated Latent Fingerprint Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 6

M. Castrillon-Santana, J. Lorenzo-Navarro, and E. Ramon-Balmaseda. Multi-scale score level fusion of local descriptors for gender classification in the wild. *Multimedia Tools and Applications*, pages 1–17, 2016. 25

C. H. Chan, X. Zou, N. Poh, and J. Kittler. Illumination invariant face recognition: a survey. In *Face Recognition in Adverse Conditions*, pages 147–166. IGI Global, 2014. 8

C. Chen, A. Dantcheva, and A. Ross. Impact of facial cosmetics on automatic gender and age estimation algorithms. In *Proceedings of IEEE International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 182–190, 2014. 8

L. Chen, Z. Mu, B. Zhang, and Y. Zhang. Ear recognition from one sample per person. *PloS one*, 10(5):e0129505, 2015. 6

W. Chen and J. Hays. Sketchygan: Towards diverse and realistic sketch to image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9416–9425, 2018. 37

K. Cho, B. Merrienboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv:1409.1259*, 2014a. 37, 63

K. Cho, B. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv:1406.1078*, 2014b. 37, 63

Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *arXiv preprint*, 1711, 2017. 37

S. Chopra, R. Hadsell, and Y. LeCun. Learning a Similarity Metric Discriminatively, With Application to Face Verification. *In Proc. Computer Vision and Pattern Recognition*, 2005. 61, 111

T. Chugh, K. Cao, and A. Jain. Fingerprint Spoof Buster: Use of Minutiae-centered Patches. *IEEE Transactions on Information Forensics and Security*, 2018. 11

J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555*, 2014. 37

S. 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. 33

COST. COST 2101: Biometrics for Identity Documents and Smart Cards, 2007. http://cost2101.org/. 5

A. Dantcheva, P. Elia, and A. Ross. What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3):441–467, 2016. 8

J. Daugman. Biometric Decision Landscapes. Technical Report UCAM-CL-TR-482, University of Cambridge, Computer Laboratory, 2000. 91

J. Daugman. The Importance of Being Random: Statistical Principles of Iris Recognition. *Pattern Recognition*, 36:279–291, 2003. 35

D. Deb, N. Nain, and A. Jain. Longitudinal Study of Child Face Recognition. In *Proc. International Conference on Biometrics*, pages 225–232, 2018. 33

D. Dessimoz and *et al.* Multimodal Biometrics for Identity Documents (MBioID). *Forensic Science International*, (167):154–159, 2007. 45

S. Dey, A. Dutta, J. Toledo, S. Ghosh, J. Llados, and U. Pal. SigNet: Convolutional Siamese network for writer independent offline signature verification. *arXiv preprint arXiv:1707.02131*, 2017. 9

M. Diaz, M. Ferrer, G. Eskander, and R. Sabourin. Generation of Duplicated Off-Line Signature Images for Verification Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5):951–964, 2017a. 9

M. Diaz, M. Ferrer, and J. Hernandez. Anthropomorphic Features for On-Line Signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018a. 9, 26

M. Diaz, M. Ferrer, D. Impedovo, M. Malik, G. Pirlo, and R. Plamondon. A Perspective Analysis of Handwritten Signature Technology. *ACM Computing Surveys*, pages 103–112, 2018b. 5, 6, 9, 14, 23, 25, 26, 28, 41, 61, 74

M. Diaz, M. Ferrer, A. Parziale, and A. Marcelli. Recovering Western On-line Signatures From Image-Based Specimens. *Proc. Int. Conf. on Document Analysis and Recognition, ICDAR*, 2017b. 113, 121

M. Diaz, A. Fischer, M. Ferrer, and R. Plamondon. Dynamic Signature Verification System based on One Real Signature. *IEEE Transactions on Cybernetics*, 2016a. 9, 27

M. Diaz, A. Fischer, M. Ferrer, and R. Plamondon. Dynamic Signature Verification System Based on One Real Signature. *IEEE Transactions on Cybernetics*, pages 1–12, 2016b. 28, 45, 61, 108

C. Ding and D. Tao. Pose-Invariant Face Recognition with Homography-Based Normalization. *Pattern Recognition*, 66:144–152, 2017. 8

DoD, 2005. Biometrics Management Office, Department of Defense, USA. (http://www.biometrics.dod.mil/). 6

J. Dolfing, E. Aarts, and J. V. Oosterhout. On-Line Signature Verification with Hidden Markov Models. In *Proc. Int. Conference on Pattern Recognition, ICPR*, page 1309, 1998. 28

N. Drempt, A. McCluskey, and N. Lannin. A Review of Factors that Influence Adult Handwriting Performance. *Australian Occupational Therapy Journal*, 58(5):321–328, 2011. 33

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001. 17

B. Dumas and *et al.* MyIDEA: Multimodal Biometrics Database, Description of Acquisition Protocols. *CT-2OS75 Workshop Biometrics on the Internet.*, 2005. 45

EBF, 2009. European Biometrics Forum. (http://www.eubiometricforum.com/). 6

D. Erhan, Y. Bengio, A. Courville, P. Manzagol, P. Vincent, and S. Bengio. Why Does Unsupervised Pre-Training Help Deep Learning? *Journal of Machine Learning Research*, 11:625–660, 2010. 38

M. Fairhurst and E. Kaplani. Strategies for Exploiting Signature Verification Based on Complexity Estimates. In *In University of Kent, Canterbury*, 1998. 34

M. Faundez-Zanuy. On-Line Signature Recognition based on VQ-DTW. *Pattern Recognition*, 40(3):981–992, 2007. 28

M. Ferrer, S. Chanda, M. Diaz, C. Banerjee, A. Majumdar, C. Carmona-Duarte, P. Acharya, and U. Pal. Static and Dynamic Synthesis of Bengali and Devanagari Signatures. *IEEE transactions on cybernetics*, 2017a. 9

M. Ferrer, M. Diaz, C. Carmona, and R. Plamondon. iDeLog: Iterative Dual Spatial and Kinematic Extraction of Sigma-Lognormal Parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 111, 141

M. Ferrer, M. Diaz, C. Carmona-Duarte, and A. Morales. A behavioral handwriting model for static and dynamic signature synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1041–1053, 2017b. 45, 121

M. A. Ferrer, M. Diaz-Cabrera, and A. Morales. Static signature synthesis: A neuromotor inspired approach for biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):667–680, 2015. 8, 9

J. Fierrez, J. Galbally, J. Ortega-Garcia, M. Freire, F. Alonso-Fernandez, D. Ramos, D. Toledano, J. Gonzalez-Rodriguez, J. Siguenza, J. Garrido-Salas, *et al.* BiosecurID: A Multimodal Biometric Database. *Pattern Analysis and Applications*, 13(2):235–246, 2010. 13, 43, 45, 46, 51, 97

J. Fierrez, A. Kumar, M. Vatsa, R. Veldhuis, and J. Ortega-Garcia, editors. *Proceedings of IAPR International Conference on Biometrics (ICB)*, 2013. IEEE. 5

J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho. Multiple Classifiers in Biometrics. Part 1: Fundamentals and Review. *Information Fusion*, 44:57–64, 2018a. 92

J. Fierrez and J. Ortega-Garcia. On-Line Signature Verification. *A.K. Jain, A. Ross and P.Flynn (Eds.), Handbook of Biometrics, Springer*, pages 189–209, 2008. 23, 26, 41

J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez. HMM-based On-Line Signature Verification: Feature Extraction and Signature Modeling. *Pattern Recognition Letters*, 28(16):2325–2334, 2007. 25, 27, 28, 29, 58, 90

J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales. Benchmarking Touchscreen Biometrics for Mobile Authentication. *IEEE Trans. on Information Forensics and Security*, 13, 2018b. 39, 40, 41

J. Fierrez-Aguilar. *Adapted Fusion Schemes for Multimodal Biometric Authentication.* PhD thesis, Universidad Politecnica de Madrid, May 2006. 57, 58, 59

J. Fierrez-Aguilar, S. Krawczyk, J. Ortega-Garcia, and A. Jain. Fusion of Local and Regional Approaches for On-Line Signature Verification. In *Proc. Intl. W. on Biometric Recognition Systems*, volume 3781, pages 188–196, 2005a. 27, 30, 94

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. 5th IAPR Intl. Conf. on Audio- and Video-based Biometric Person Authentication, AVBPA*, volume 3546, pages 523–532, 2005b. xxvii, 26, 27, 55, 56

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, Special Issue on Biometric Systems*, 35(3):418–425, 2005c. 25, 46

A. Fischer and R. Plamondon. Signature Verification based on the Kinematic Theory of Rapid Human Movements. *IEEE Transactions on Human-Machine Systems*, 47(2):169–180, 2017. 27, 28, 113

J. Galbally, I. Coisel, and I. Sanchez. A New Multimodal Approach for Password Strength Estimation â€" Part I: Theory and Algorithms. *IEEE Transactions on Information Forensics and Security*, 12:2829–2844, 2017. 3, 12, 14

J. Galbally, M. Diaz-Cabrera, M. Ferrer, M. Gomez-Barrero, A. Morales, and J. Fierrez. On-Line Signature Recognition through the Combination of Real Dynamic Data and Synthetically Generated Static Data. *Pattern Recognition*, 48(9):2921–2934, 2015. 10, 26, 108, 121

J. Galbally, J. Fierrez, M. Freire, and J. Ortega-Garcia. Feature Selection based on Genetic Algorithms for On-Line Signature Verification. In *Proc. IEEE Workshop on Automatic Identification Advanced Technologies*, pages 198–203, 2007. 30

J. Galbally, J. Fierrez, M. Martinez-Diaz, and J. Ortega-Garcia. Improving the Enrollment in Dynamic Signature Verfication with Synthetic Samples. In *Proc. International Conference on Document Analysis and Recognition*, pages 1295–1299, 2009. 33

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. 9

J. Galbally, R. Haraksim, and J. Beslay. Fingerprint Quality: a Lifetime Story. In *Proc. International Conference of the Biometrics Special Interest Group*, 2018. 33

J. Galbally, S. Marcel, and J. Fierrez. Biometric Antispoofing Methods: A Survey in Face Recognition. *IEEE Access*, 2:1530–1552, 2014. 6

J. Galbally, M. Martinez-Diaz, and J. Fierrez. Aging in Biometrics: An Experimental Analysis on On-Line Signature. *PLOS ONE*, 8(7), 2013. 25, 28, 30, 34, 44, 45, 51, 52, 95, 131

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. 9

S. Garcia-Salicetti and *et al. BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities*, pages 845–853. 2003. 45

O. Ghahabi and J. Hernando. Deep Learning Backend for Single and Multisession i-vector Speaker Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(4):807–817, 2017. 6

M. Gomez-Barrero, J. Galbally, J. Fierrez, J. Ortega-Garcia, and R. Plamondon. Enhanced On-Line Signature Verification Based on Skilled Forgery Detection Using Sigma-LogNormal Features. *In Proc. IEEE/IAPR Int. Conf. on Biometrics, ICB*, 2015. 26, 45, 61, 107, 109, 113, 117, 121, 149

E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez. Facial soft biometrics for recognition in the wild: Recent works, annotation, and cots evaluation. *IEEE Transactions on Information Forensics and Security*, 13(8):2001–2014, 2018. 8

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`. XXIV, 4, 5, 11, 17, 35, 36, 37, 64

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 37

A. Graves and N. Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In *International Conference on Machine Learning*, pages 1764–1772, 2014. 11, 38, 64

A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):855–868, 2009. 11, 13, 37, 64

A. Graves and J. Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. *In Proc. Advances in Neural Information Processing Systems*, pages 545–552, 2009. 11, 13

R. Guest. Age Dependency in Handwritten Dynamic Signature Verification Systems. *Pattern Recognition Letters*, 27(10):1098–1104, 2006. 11, 33, 44, 45

D. Guru and H. Prakash. Online Signature Verification and Recognition: An Approach based on Symbolic Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1059–1073, 2009. 26

M. Harbach, A. D. Luca, and S. Egelman. The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens. In *Proc. Conference on Human Factors in Computing Systems*, pages 4806–4817, 2016. 111, 141

C. M. Harris and D. M. Wolpert. Signal-Dependent Noise Determines Motor Planning. *Nature*, 394(6695):780, 1998. 9

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. 37, 61

E. Hoffer and N. Ailon. Deep Metric Learning Using Triplet Network. In *Proc. International Workshop on Similarity-Based Pattern Recognition*, pages 84–92, 2015. 39

N. Hogan. An Organizing Principle for a Class of Voluntary Movements. *Journal of Neuroscience*, 4(11):2745–2754, 1984. 8

C. Hong, J. Yu, J. Wan, D. Tao, and M. Wang. Multimodal Deep Autoencoder for Human Pose Recovery. *IEEE Transactions on Image Processing*, 24(12):5659–5670, 2015. 37

M. E. Hoque, V. M. Patel, N. Saxena, and M. Vatsa, editors. *Proc. IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2017. 5

W. Hou, X. Ye, and K. Wang. A Survey of Off-Line Signature Verification. In *Proc. Int. Conf. on Intelligent Mechatronics and Automation*, pages 536–541, 2004. 9

N. Houmani and S. Garcia-Salicetti. On Hunting Animals of the Biometric Menagerie for Online Signature. *PLOS ONE*, 11:1–26, 2016. 35, 116, 117, 118, 151, 161

N. Houmani, S. Garcia-Salicetti, and B. Dorizzi. A Novel Personal Entropy Measure Confronted to Online Signature Verification Systems Performance. In *In Proc. BTAS*, pages 1–6, 2008. 35, 151, 161

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. Intl. Conf. on Biometrics: Theory, Applications and Systems, BTAS*, pages 1–6, 2009. 33

N. Houmani, S. Garcia-Salicetti, B. Dorizzi, J. Montalvão, J. Canuto, M. Andrade, Y. Qiao, X. Wang, T. Scheidat, A. Makrushin, *et al.* BioSecure Signature Evaluation Campaign (ESRA'2011): Evaluating Systems on Quality-based Categories of Skilled Forgeries. In *International Joint Conference on Biometrics, IJCB*, pages 1–10, 2011. 25, 28

A. Humm, J. Hennebert, and R. Ingold. Combined Handwriting and Speech Modalities for User Authentication. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(1):25–35, 2009. 45

I. Sutskever, O. Vinyals and Q.V. Le. Sequence to Sequence Learning with Neural Networks. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014. 4

IBIA. International Biometric Industry Association, 2009. (http://www.ibia.org/association/). 6

D. Impedovo and G. Pirlo. Automatic Signature Verification: The State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 38(5):609–635, 2008. 5, 8, 9, 14, 23

D. Impedovo, G. Pirlo, F. Mangini, D. Barbuzzi, A. Rollo, A. Balestrucci, S. Impedovo, L. Sarcinella, C. Reilly, and R. Plamondon. Writing Generation Model for Health Care Neuromuscular System Investigation. In *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, pages 137–148. Springer, 2013. 9, 113

International Biometric Group. Biometrics market and industry report 2006-2010, 2006. (http://www.biometricgroup.com/). 6

ISO/IEC JTC 1/SC 27 . IT Security Techniques, 2009. http://www.jtc1.org/sc27/. 6

ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Information Technology - Biometric presentation attack detection - Part 1: Framework*. International Organization for Standardization, 2016. 6

A. Jain, P. Flynn, and A. Ross, editors. *Handbook of Biometrics*. Springer, 2008. 5

A. Jain, K. Nandakumar, and A. Ross. Score Normalization in Multimodal Biometric Systems. *Pattern Recognition*, 38(12):2270–2285, 2005. 25, 79, 92

A. Jain and D. Zongker. Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997. 26, 30

A. K. Jain, F. D. Griess, and S. D. Connell. On-Line Signature Verification. *Pattern Recognition*, 35(12):2963–2972, 2002. 10

A. K. Jain, K. Nandakumar, and A. Ross. 50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities. *Pattern Recognition Letters*, 79:80–105, 2016. 3, 5, 6

A. K. Jain, A. Ross, and S. Pankanti. Biometrics: A Tool for Information Security. *IEEE Transactions on Information Forensics and Security (TIFS)*, 1(2):125–143, 2006. 5

A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004. 5

A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer Science+Business Media, LLC, 2011. 5, 8

J. Jenkins, J. Shelton, and K. Roy. One-Time Password for Biometric Systems: Disposable Feature Templates. In *Proc. SoutheastCon*, 2017. 12

R. Jonas and D. Andrzej. Gaussian Mixture Models for On-Line Signature Verification. In *Proc. of the ACM SIGMM workshop on Biometrics methods and applications*, pages 115–122, 2003. 25, 27, 28

R. Jozefowicz, W. Zaremba, and I. Sutskever. An Empirical Exploration of Recurrent Network Architectures. *Journal of Machine Learning Research*, 2015. 63

A. Kareem, E. IM, M. Rashad, and O. Nomir. On-Line Signature Verification based on PCA Feature Reduction and Statistical Analysis. In *Proc. International Conference on Computer Engineering and Systems (ICCES)*, pages 3–8, 2010. 25, 26

P. Kasprowski and K. Harezlak. Fusion of Eye Movement and Mouse Dynamics for Reliable Behavioral Biometrics. *Pattern Analysis and Applications*, 21(1):91–103, 2018. 8

I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The Megaface Benchmark: 1 million Faces for Recognition at Scale. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, 2016. 5

W. Khan, M. Aalsalem, and Y. Xiang. A Graphical Password Based System for Small Mobile Devices. *International Journal of Computer Science*, 5(2):145–154, 2011. 40, 41

A. Kholmatov and B. Yanikoglu. Identity Authentication Using Improved Online Signature Verification Method. *Pattern Recognition Letters*, 26(15):2400–2408, 2005. 28

A. Kholmatov and B. Yanikoglu. SUSIG: an On-Line Signature Database, Associated Protocols and Benchmark Results. *Pattern Analysis and Applications*, 12(3):227–236, 2009. 45

J. Kittler, M. Hatef, R. Duin, and J. Matas. On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998. 25, 79, 92, 99

T. Kutzner, F. Ye, I. Bonninger, C. Travieso, M. Dutta, and A. Singh. User Verification Using Safe Handwritten Passwords on Smartphones. In *Proc. 8th International Conference on Contemporary Computing, IC3*, 2015. 40, 41, 42, 136

P. Lacharme and C. Rosenberger. Synchronous One Time Biometrics With Pattern Based Authentication. In *Proc. 11th Int. Conf. on Availability, Reliability and Security, ARES*, 2016. 12, 39, 40, 135

S. Lai, L. Jin, and W. Yang. Online Signature Verification Using Recurrent Neural Network and Length-Normalized Path Signature. *arXiv preprint arXiv:1705.06849*, 2017. 39

F. Leclerc and R. Plamondon. Automatic Signature Verification: The State of the Art. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(3):643–660, 1994. 5, 9, 14

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015. 11

L. Lee, T. Berger, and E. Aviczer. Reliable On-Line Human Signature Verification Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):643–647, 1996. 26, 29, 55

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. 25, 58

L. Li, X. Zhao, and G. Xue. Unobservable Reauthentication for Smartphones. In *Proc. 20th Network and Distributed System Security Symposium, NDSS*, 2013. 40, 41

M. Liang and X. Hu. Recurrent Convolutional Neural Network for Object Recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3367–3375, 2015. 127

M. Lim and P. Yuen. Entropy Measurement for Biometric Verification Systems. *IEEE Transactions on Cybernetics*, 46:1065–1077, 2016. 35, 151, 161

H. Ling, S. Soatto, R. Ramanathan, and D. Jacobs. A Study of Face Recognition as People Age. In *Proc. International Conference on Computer Vision*, pages 1–8, 2007. 33

Y. Liu, Z. Yang, , and L. Yang. Online Signature Verification Based on DCT and Sparse Representation. *IEEE Transactions on Cybernetics*, 45(11):2498–2511, 2014. 25, 26, 61

A. Lozano-Diez, R. Zazo, D. T. Toledano, and J. Gonzalez-Rodriguez. An Analysis of the Influence of Deep Neural Network (DNN) Topology in Bottleneck Feature based Language Recognition. *PloS one*, 12(8):e0182580, 2017. 38

A. Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch Me Once and I Know It is You! Implicit Authentication based on Touch Screen Patterns. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996, 2012. 12

Y. Mahajan and S. Sondur. Aging Face Recognition Using Deep Learning. *International Journal of Engineering and Applied Sciences*, 2018. 33

M. Malik, S. Ahmed, A. Marcelli, U. Pal, M. Blumenstein, L. Alewijns, and M. Liwicki. ICDAR2015 competition on Signature Verification and Writer Identification for On-and Off-Line Skilled Forgeries (SigWIcomp2015). In *13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1186–1190, 2015. 24, 28, 45

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. 12th International Conference on Document Analysis and Recognition*, pages 1477–1483, 2013. 28

E. Marasco and A. Ross. A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, 47(2):28, 2015. 6

S. Marcel, M. S. Nixon, and S. Z. Li, editors. *Handbook of Biometric Anti-Spoofing*. Springer, 2014. 6

A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The det curve in assessment of detection task performance. Technical report, National Inst of Standards and Technology Gaithersburg MD, 1997. 10

M. Martinez-Diaz. *Automatic Signature and Graphical Password Verification: Discriminant Features and New Application Scenarios*. PhD thesis, Universidad Autonoma de Madrid, June 2015. xxiii, xxiii, xxvii, 56, 57, 58, 59, 60

M. Martinez-Diaz and J. Fierrez. *Signature Databases and Evaluation*, pages 1367–1375. Springer, 2015. 43

M. Martinez-Diaz, J. Fierrez, M. Freire, and J. Ortega-Garcia. On the Effects of Sampling Rate and Interpolation in HMM-Based Dynamic Signature Verification. In *Proc. Intl. Conf. on Document Analysis and Recognition, ICDAR*, volume 2, pages 1113–1117, 2007. 24, 28, 70, 75

M. Martinez-Diaz, J. Fierrez, and J. Galbally. The DooDB Graphical Password Database: Data Analysis and Benchmark Results. *IEEE Access*, 1:596–605, 2013. 13, 32, 44, 45

M. Martinez-Diaz, J. Fierrez, and J. Galbally. Graphical password-based user authentication with free-form doodles. *IEEE Trans. on Human-Machine Systems*, 46(4):607–614, 2016. 40, 41, 45, 83, 136

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. Int. Conference on Pattern Recognition, ICPR*, pages 1–5, 2008. 30

M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Signature Features*, pages 1375–1382. Springer, 2015a. ISBN 978-1-4899-7487-7, re-edited from 2009. 25

M. Martinez-Diaz, J. Fierrez, and S. Hangai. *Signature Matching*, pages 1382–1387. Springer, 2015b. ISBN 978-1-4899-7487-7, re-edited from 2009. 25, 129

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. 25, 26, 27, 30, 61, 74, 92

W. Meng, D. Wong, S. Furnell, and J. Zhou. Surveying the Development of Biometric User Authentication on Mobile Phones. *IEEE Communications Surveys Tutorials*, 17(3):1268–1293, 2015. 3, 12

R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. Deep Learning for Healthcare: Review, Opportunities and Challenges. *Briefings in Bioinformatics*, 2017. 11

S. Modi and S. Elliott. Impact of Image Quality on Performance: Comparison of Young and Elderly Fingerprints. In *Proc. International Conference on Recent Advances in Soft Computing*, pages 10–12, 2006. 33

S. Modi, S. Elliott, and K. Hakil. Impact of Age Groups on Fingerprint Recognition Performance. In *Proc. IEEE Workshop on Automatic Identification Advanced Technologies*, pages 19–23, 2007. 33

A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos, and S. Marcel. Keystroke Biometrics Ongoing Competition. *IEEE Access*, 4:7736–7746, 2016. 6

MTIT. Minutiae Template Interoperability Testing, FP6-2004-IST-4, 2009. http://www.mtitproject.com/index.html. 5

D. Muramatsu and T. Matsumoto. Effectiveness of Pen Pressure, Azimuth, and Altitude Features for Online Signature Verification. In *Proc. International Conference on Biometrics, ICB*, pages 503–512, 2007. 25

S. Nam, C. Seo, and D. Choi. Mobile Finger Signature Verification Robust to Skilled Forgery. *Journal of the Korea Institute of Information Security and Cryptology*, 26:1161–1170, 2016. 33

L. Nanni and A. Lumini. Ensemble of Parzen Window Classifiers for On-Line Signature Verification. *Neurocomputing*, 68:217–224, 2005. 25

P. D. Neilson. The Problem of Redundancy in Movement Control: The Adaptive Model Theory Approach. *Psychological research*, 55(2):99–106, 1993. 9

W. Nelson and E. Kishon. Use of Dynamic Features for Signature Verification. In *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics*, pages 201–205, 1991. 55

W. Nelson, W. Turin, and T. Hastie. Statistical Methods for On-Line Signature Verification. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(3):749–770, 1994. 55

J. Neves and H. Proença. ICB-RW 2016: International Challenge on Biometric Recognition in the Wild. In *Proc. of IAPR International Conference on Biometrics (ICB)*, pages 1–6, 2016. 5

T. Nguyen, N. Sae-Bae, and N. Memon. DRAW-A-PIN: Authentication Using Finger-Drawn PIN on Touch Devices. *Computers and Security*, 66:115–128, 2017a. 40, 42, 136

T. Nguyen, N. Sae-Bae, and N. Memon. DRAW-A-PIN: Authentication Using Finger-Drawn PIN on Touch Devices. *Computers & Security*, 66:115–128, 2017b. 141

V. Nguyen, M. Blumenstein, and G. Leedham. Global Features for the Off-Line Signature Verification Problem. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 1300–1304, 2009. 9

R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. Fingerprint Liveness Detection Using Convolutional Neural Networks. *IEEE Trans. Information Forensics and Security*, 11(6):1206–1213, 2016. 8, 38

J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. Sukno, S. Pavani, A. Frangi, L. Akarun, and A. Savran. The Multi-Scenario Multi-Environment BioSecure Multimodal Database (BMDB. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(6):1097–1111, 2010. 23, 43, 45, 46, 51

J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J. Igarza, C. Vivaracho, *et al.* MCYT Baseline Corpus: A Bimodal Biometric Database. *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6):395–401, 2003. 13, 23, 43, 45

S. Otte, M. Liwicki, and D. Krechel. Investigating Long Short-Term Memory Networks for Various Pattern Recognition Problems. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 484–497. Springer, 2014. 12, 38, 112

B. Paltridge. Thesis and dissertation writing: An examination of published advice and actual practice. *English for Scientific Purposes*, 21:125–143, 2002. 15

O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.* Deep face recognition. In *BMVC*, volume 1, page 6, 2015. 11

M. Parodi and J. G. L. Alewijnse. Automatic Online Signature Verification based Only on FHE Features: An Oxymoron? In *Proc. 14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 73–78, 2014. 25, 26

M. Parodi and J. Gomez. Legendre Polynomials based Feature Extraction for Online Signature Verification. Consistency Analysis of Feature Combinations. *Pattern Recognition*, 47(1):128–140, 2014. 30

A. Parziale, S. Fuschetto, and A. Marcelli. Exploiting Stability Regions for Online Signature Verification. In *International Conference on Image Analysis and Processing*, pages 112–121, 2013. 26

R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio. How to Construct Deep Recurrent Neutal Networks. *arXiv*, 1312.6026, 2014. 61

V. Patel, R. Chellappa, D. Chandra, and B. Barbello. Continuous User Authentication on Mobile Devices: Recent Progress and Remaining Challenges. *IEEE Signal Processing Magazine*, 33:49–61, 2016. 41

A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch. Recurrent neural network based prediction of epileptic seizures in intra- and extracranial eeg. *Neurocomputing*, 30:201–218, 2000. 61

J. P. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(10):1090–1104, 2000. 5

P. J. Phillips. Face and Iris Evaluations at NIST. In *CardTech/SecurTech*, May 2006. 5

P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer. An Introduction to the Good, the Bad, and the Ugly Face Recognition Challenge Problem. In *Proc. of IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG)*, pages 346 –353, march 2011. 5

P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, Iii, and S. Weimer. Overview of the multiple biometrics grand challenge. In *Proceedings of International Conference on Advances in Biometrics*, pages 705–714, Berlin, Heidelberg, 2009a. Springer-Verlag. ISBN 978-3-642-01792-6. 5

P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 Large-Scale Experimental Results. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 99, 2009b. 5

G. Pirlo, V. Cuccovillo, D. Impedovo, and P. Mignone. On-Line Signature Verification by Multi-Domain Classification. In *Proc. 14th International Conference on Frontiers in Handwriting Recognition, ICFHR*, pages 67–72, 2014. 27

R. Plamondon. A Kinematic Theory of Rapid Human Movements. *Biological Cybernetics*, 72(4):295–307, 1995. 8

R. Plamondon and G. Lorette. Automatic Signature Verification and Writer Identification â€" The State of the Art. *Pattern recognition*, 22(2):107–131, 1989. 5, 9, 10, 14, 23, 24

R. Plamondon and M. Parizeau. Signature Verification from Position, Velocity and Acceleration Signals: A Comparative Study. In *9th International Conference on Pattern Recognition*, pages 260–265. IEEE, 1988. 9

R. Plamondon and S. N. Srihari. Online and Off-Line Handwriting Recognition: a Comprehensive Survey. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):63–84, 2000. 5, 9, 14, 23, 26, 41, 74

N. Poh, J. Kittler, and T. Bourlai. Improving Biometric Device Interoperability by Likelihood Ratio-Based Quality Dependent Score Normalization. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–5, 2007. 8, 25

M. Przybocki and A. Martin. NIST Speaker Recognition Evaluation chronicles. In J. Ortega-Garcia *et al.*, editors, *ISCA Workshop on Speaker and Language Recognition (ODYSSEY)*, pages 15–22, 2004. 5

P. Pudil, J. Novovičová, and J. Kittler. Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994. 26, 30

R. Plamondon and G. Pirlo and D. Impedovo. Online Signature Verification. *D. Doermann and K. Tombre (Eds.), Handbook of Document Image Processing and Recognition, Springer*, pages 917–947, 2014. 41

L. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. 28, 60

A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1511.06434*, 2015. 38

N. Ramanathan and R. Chellappa. Face Verification Across Age Progression. *IEEE Transactions on Image Processing*, 15(11):3349–3361, 2006. 33

N. Ratha, J. Connell, and R. Bolle. Enhancing Security and Privacy in Biometrics-Based Authentication Systems. *IBM Systems Journal*, 40(3):614–634, 2001. 6

N. K. Ratha and V. Govindaraju, editors. *Advances in biometrics: Sensors, algorithms and systems*. Springer, 2008. 5

C. Reilly and R. Plamondon. Development of a Sigma-Lognormal Representation for On-Line Signatures. *Pattern Recognition*, 42(12):3324–3337, 2009. xxvi, 9, 113, 115

C. Reilly and R. Plamondon. Design of a Neuromuscular Disorders Diagnostic System Using Human Movement Analysis. In *Proc. Int. Conf. on Information Science, Signal Processing and Their Applications, ISSPA*, pages 787–792, 2012. 33

J. Richiardi, H. Ketabdar, and A. Drygajlo. Local and Global Feature Selection for On-Line Signature Verification. In *Proc. 8th International Conference on Document Analysis and Recognition, ICDAR*, pages 625–629, 2005. 26, 30, 58

J. Robertson and R. Guest. A Feature Based Comparison of Pen and Swipe based Signature Characteristics. *Human movement science*, 43:169–182, 2015. 13, 32

A. Ross and A. Jain. Biometric sensor interoperability: A case study in fingerprints. In *International Workshop on Biometric Authentication*, pages 134–145, 2004. 8

A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer, 2006. 5, 8

E. Rúa and J. Castro. Online Signature Verification based on Generative Models. *IEEE Transactions on System, Man, Cybernetics. Part B*, 42(4):1231–1242, 2012. 30

N. Sae-Bae and N. Memon. Online Signature Verification on Mobile Devices. *IEEE Transactions on Information Forensics and Security*, 9(6):933–947, 2014. 25, 26, 32, 33, 34, 40, 41, 44, 45, 95, 96

N. Sae-Bae and N. Memon. Quality of Online Signature Templates. In *Proc. of the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–8, 2015. 91

N. Sae-Bae, N. Memon, K. Isbister, and K. Ahmed. Multitouch Gesture-Based Authentication. *IEEE Transactions on Information Forensics and Security*, 9(4):568–582, 2014. 40, 41

M. Salehan and A. Negahban. Social Networking on Smartphones: When Mobile Phones Become Addictive. *Computers in Human Behavior*, 29(6):2632–2639, 2013. 3, 11, 13

SC37, 2005. ISO/IEC JTC 1/SC 37 . (http://www.jtc1.org/sc37/). 6

J. Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural networks*, 61:85–117, 2015. 5, 11, 17, 36

M. Schuster and K. Paliwal. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Processing*, 45: 2673–2681, 1997. 37, 64

A. Serwadda, V. Phoha, and Z. Wang. Which Verifiers Work?: A Benchmark Evaluation of Touch-based Authentication Algorithms. In *Proc. of the Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1–8, 2013. 41

A. Sharma and S. Sundaram. An Enhanced Contextual DTW based System for Online Signature Verification Using Vector Quantization. *Pattern Recognition Letters*, 84:22–28, 2016. 26

A. Sharma and S. Sundaram. A Novel Online Signature Verification System Based on GMM Features in a DTW Framework. *IEEE Trans. Information Forensics and Security*, 12(3):705–718, 2017. 27, 28

C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion. Performance Analysis of Multi-Motion Sensor Behavior for Active Smartphone Authentication. *IEEE Transactions on Information Forensics and Security*, 13(1):48–62, 2018. 6

C. Shen, Y. Zhang, X. Guan, and R. Maxion. Performance Analysis of Touch-Interaction Behavior for Active Smartphone Authentication. *IEEE Transactions on Information Forensics and Security*, 11(3):498–513, 2016. 40, 41

D. Shukla, R. Kumar, A. Serwadda, and V. Phoha. Beware, Your Hands Reveal Your Secrets! In *Proc. of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014. 12

R. Singh, K. Nandakumar, and D. Maltoni, editors. *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2016. IEEE. 5

V. Smejkal, J. Kodl, L. Sieger, F. Hortai, and P. Tesar. Stability of a Dynamic Biometric Signature Created on Various Devices. In *Proc. Int. Carnahan Conf. on Security Technology*, 2017. 32

C. Sousedik and C. Busch. Presentation Attack Detection Methods for Fingerprint Recognition Systems: a Survey. *IET Biometrics*, 3(4):219–233, 2014. 6

C. Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. Freca. A Brief Overview on Handwriting Analysis for Neurodegenerative Disease Diagnosis. In *Proc. Workshop on Artificial Intelligence with Application in Health*, 2017. 113

K. Sundararajan and D. L. Woodard. Deep Learning for Biometrics: A Survey. *ACM Computing Surveys (CSUR)*, 51(3):65, 2018. 4, 11

I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112, 2014. 11

J. Svoboda, J. Masci, and M. Bronstein. Palmprint Recognition Via Discriminative Index Learning. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 4232–4237, 2016. 6

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 11

Tabula Rasa, 2010. TABULA RASA: Trusted Biometrics under Spoofing Attacks, FP7-ICT-257289. (http://www.tabularasa-euproject.org). 5

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 6

L. Tang, Y. Fang, Q. Wu, W. Kang, and J. Zhao. Online Finger-Writing Signature Verification on Mobile Device for Local Authentication. In *Proc. Chinese Conference on Biometric Recognition*, pages 409–416, 2016. 33

S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, 2008. 17, 29, 30, 57

C. Tiflin and C. Omlin. LSTM Recurrent Neural Networks for Signature Verification. *In Proc. Southern African Telecommunication Networks and Applications Conference*, 2003. 12, 38

M. Tistareli, S. Z. Li, and R. Chellappa, editors. *Handbook of Remote Biometrics for Surveillance and Security*. Springer, 2009. 5

R. Tolosana, M. Gomez-Barrero, J. Kolberg, A. Morales, C. Busch, and J. Ortega-Garcia. Towards Fingerprint Presentation Attack Detection Based on Convolutional Neural Networks and Short Wave Infrared Imaging. In *Proc. 17th International Conference of the Biometrics Special Interest Group, BIOSIG*, 2018a. 6

R. Tolosana, R. Vera-Rodriguez, and J. Fierrez. BioTouchPass: Handwritten Passwords for Touchscreen Biometrics. *IEEE Transactions on Mobile Computing*, 2018b. under review. 23, 43, 128

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database. *PLOS ONE*, 5(12), 2017a. 23, 39, 40, 41, 43, 45, 69, 117, 131

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Do You Need More Data? The DeepSignDB On-Line Handwritten Signature Biometric Database. In *Proc. 15th Int. Conference on Document Analysis and Recognition, ICDAR*, 2019a. under review. 103, 151, 161

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Feature-Based Dynamic Signature Verification under Forensic Scenarios. In *Proc. 3rd International Workshop on Biometrics and Forensics, IWBF*, 2015a. 25, 26, 55

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Biometric Signature Verification Using Recurrent Neural Networks. In *Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, 2017b. 55, 103

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics. *IEEE Access*, pages 5128–5138, 2018c. 23, 25, 55, 103, 150, 151, 161

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Incorporating Touch Biometrics to Mobile One-Time Passwords: Exploration of Digits. In *Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, 2018d. 43, 128

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Presentation Attacks in Signature Biometrics: Types and Introduction to Attack Detection. *S. Marcel, M.S. Nixon, J. Fierrez and N. Evans (Eds.), Handbook of Biometric Anti-Spoofing (2nd Edition), Springer*, 2018e. 6, 39, 69

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Reducing the Template Aging Effect in On-Line Signature Biometrics. *IET Biometrics*, 2018f. under review. 55, 89

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. MobileTouchDB: Mobile Touch Character Database in the Wild and Biometric Benchmark. In *Proc. Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, 2019b. under review. 128, 141, 151, 161

R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, and J. Ortega-Garcia. Complexity-based Biometric Signature Verification. In *Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, 2017c. 113

R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. Increasing the Robustness of Biometric Templates for Dynamic Signature Biometric Systems. In *Proc. 49th Annual Int. Carnahan Conf. on Security Technology*, 2015b. 30

R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. Optimal Feature Selection and Inter-Operability Compensation for On-Line Biometric Signature Authentication. In *Proc. IEEE/IAPR Int. Conf. on Biometrics, ICB*, pages 163–168, 2015c. 24, 69

R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. Preprocessing and Feature Selection for Improved Sensor Interoperability in Online Biometric Signature Verification. *IEEE Access*, 3:478–489, 2015d. 23, 24, 25, 45, 55, 69, 92, 109, 111, 141

R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. Update Strategies for HMM-Based Dynamic Signature Biometric Systems. In *Proc. 7th IEEE International Workshop on Information Forensics and Security, WIFS*, 2015e. 25, 43, 45, 55, 89, 91, 92

R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, and J. Fierrez. Assessment of Using the Finger in On-Line Handwritten Signature Veri
fication Systems. In *Proc. 18th International Graphonomics Society Conference, IGS*, 2017d. 69

M. Trauring. Automatic comparison of finger ridge patterns. *Nature*, 197:938–940, 1963. 5

L. Van, S. Garcia-Salicetti, and B. Dorizzi. On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1237–1247, 2007. 25, 28

J. Vargas, M. Ferrer, C. Travieso, and J. B. Alonso. Off-line signature verification based on grey level information using texture features. *Pattern Recognition*, 44(2):375–385, 2011. 9

R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Acien, A. Morales, J. Fierrez, and J. Ortega-Garcia. *Modeling the Complexity of Signature and Touch-Screen Biometrics using the Lognormality Principle*, pages 1–22. World Scientific, 2019. 113

R. Vera-Rodriguez, R. Tolosana, J. Hernandez-Ortega, A. Morales, J. Fierrez, and J. Ortega-Garcia. Modeling the Complexity of Biomechanical Tasks using the Lognormality Principle: Applications to Signature Recognition and Touch-Screen Children Detection. In *Proc. IAPR Intl. Conf. on Pattern Recognition and Artificial Intelligence, ICPRAI*, 2018. 113

R. Vera-Rodriguez, R. Tolosana, J. Ortega-Garcia, and J. Fierrez. e-biosign: Stylus- and finger-input multi-device database for dynamic signature recognition. In *Proc. 3rd International Workshop on Biometrics and Forensics (IWBF)*, 2015. 43, 69

L. Wan, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus. Regularization of Neural Networks using DropConnect. In *Proc. of the 30th International Conference on Machine Learning*, pages 1058–1066, 2013. 127

T. Wang, D. J. Wu, A. Coates, and A. Y. Ng. End-to-end text recognition with convolutional neural networks. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3304–3308. IEEE, 2012. 38

T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. *arXiv preprint arXiv:1711.11585*, 2017. 37

W. Wang, Y. Xu, J. Shen, and S.-C. Zhu. Attentive fashion grammar network for fashion landmark detection and clothing category classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4271–4280, 2018. 11

Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A Discriminative Feature Learning Approach for Deep Face Recognition. In *Proc. European Conference on Computer Vision*, pages 499–515, 2016. 39

D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. Are arm trajectories planned in kinematic or dynamic coordinates? an adaptation study. *Experimental brain research*, 103(3):460–470, 1995. 9

X. Xia, X. Song, F. Luan, J. Zheng, Z. Chen, and X. Ma. Discriminative Feature Selection for On-Line Signature Verification. *Pattern Recognition*, 74:422–433, 2018. 28

N. Yager and T. Dunstone. The Biometric Menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:220–230, 2010a. 35

N. Yager and T. Dunstone. The Biometric Menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2):220–230, 2010b. 111, 141

H. Yang, D. Huang, Y. Wang, and A. K. Jain. Learning Face Age Progression: A Pyramid Architecture of GANS. *arXiv preprint arXiv:1711.10352*, 2017. 37

L. Yang, B. Widjaja, and R. Prasad. Application of Hidden Markov Models for Signature Verification. *Pattern Recognition*, 28(2):161–170, 1995. 28

M. Yasuhara and M. Oka. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. volume 26, pages 43–49. 1977. 27

M. Yasuhara and M. Oka. Signature Verification Experiment based on Nonlinear Time Alignment: a Feasibility Study. volume 12, pages 212–216. 1978. 27

D. Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First International Signature Verification Competition. In D. Zhang and A. K. Jain, editors, *Biometric Authentication*, pages 16–22. Springer LNCS-3072, 2004. 5, 46

D. Y. Yeung and *et al. SVC2004: First International Signature Verification Competition*, pages 16–22. Springer Berlin Heidelberg, 2004. 45

Q. Yue, Z. Ling, X. Fu, B. Liu, W. Yu, and W. Zhao. My Google Glass Sees Your Passwords! In *Black Hat USA*, 2014. 12

R. Zazo, A. Lozano-Diez, J. Gonzalez-Dominguez, D. Toledano, and J. Gonzalez-Rodriguez. Language Identification in Short Utterances Using Long Short-Term Memory (LSTM) Recurrent Neural Networks. *PloS one*, 11(1), 2016. 61

H. Zeinali, B. BabaAli, and H. Hadian. Online Signature Verification Using i-Vector Representation. *IET Biometrics*, 2017. 26

N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie. Facial Expression Recognition Via Learning Deep Sparse Autoencoders. *Neurocomputing*, 273:643–649, 2018. 37

E. Zezschwitz, M. Eiband, D. Buschek, S. Oberhuber, A. D. Luca, F. Alt, and H. Hussmann. On Quantifying the Effective Password Space of Grid-based Unlock Gestures. In *Proc. of the International Conference on Mobile and Ubiquitous Multimedia*, pages 201–212, 2016. 39, 40

X. Zhang, G. Xie, C. Liu, and Y. Bengio. End-to-end Online Writer Identification with Recurrent Neural Network. *IEEE Transactions on Human-Machine Systems*, 47(2):285–292, 2017. 11, 13, 150, 151, 161