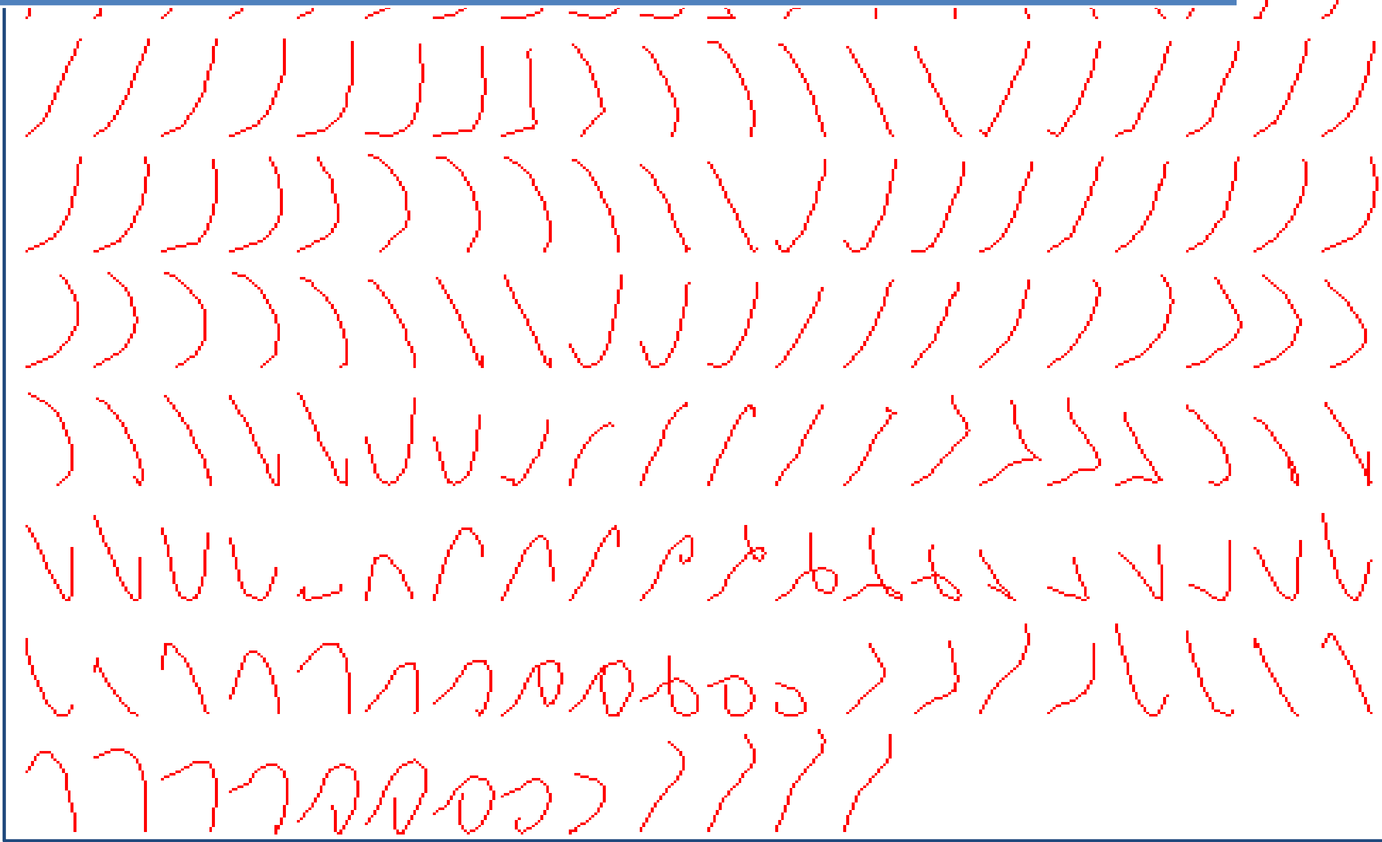
Escola Universitària Politècnica de Mataró

TecnoCampus Mataró-Maresme

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

AN ONLINE WRITER RECOGNITION SYSTEM BASED ON IN-AIR AND ON-SURFACE TRAJECTORIES



A doctoral dissertation submitted by **Enric Sesa i Nogueras** at Universitat Politècnica de Catalunya in partial fulfilment of the requirements for the PhD degree.

Advisor (Director): **Dr. Marcos Faúndez Zanuy**

Spring term 2012

Aquesta tesi està dedicada a tots aquells que tenen un lloc en el meu cor, i molt especialment al meu estimat pare, de qui sempre vaig rebre tot el recolzament imaginable.

Agraïments

Si bé una tesi doctoral té una component individual considerable, mai no podré dir que aquest sigui un treball que hagi hagut de realitzar tot sol. Ans al contrari, afortunadament. Moltes persones han (m'han) aportat alguna cosa, més gran o més petita, que ha contribuït, d'una manera o altra, al treball que ara teniu a les mans. A totes elles, el meu agraïment sincer. En particular:

A tots els qui són i sento com a *la meva família*, perquè ells són *l'entorn de desenvolupament* on bona part d'aquesta tesi ha estat realitzada: els meus pares, el Rafael i la Margarita; la meva esposa, la Jezabel, que mai no ha deixat d'insistir, sempre optimista, en que *jo podia* i que em fa veure el camí, en molts àmbits de la vida, fins i tot quan jo estic poc predisposat a veure'l; la meva germana Mercè; la Montserrat, la meva sogra; els meus cunyats i cunyades, el David, la Ethel, l'Héctor, l'Àlex M. i l'Alex G.; i les nebodes i nebots, l'Ènia, l'Eva, la Fiona, el Nil, l'Ona i el Dídac; i el Vicens, que també hi és, i molt.

Al Dr. Marcos Faúndez, el *meu Director*, que quan vaig trucar a la seva porta per dir-li que potser m'interessaria de fer una tesi em va rebre amb els braços oberts i amb més entusiasme del que potser jo mateix tenia. Un entusiasme que ha mantingut durant tot el període de recerca i que ha sabut combinar amb uns consells que s'han demostrat encertadíssims. I encara que m'ha deixat treballar al meu aire, la seva supervisió ha fet que aquest treball arribi a bon port.

A la Núria, la meva *amiga de despatx*, que ha estat testimoni del dia a dia d'aquesta tesi. A ella li haig d'agrair la seva alegria sincera *quan els resultats eren bons* o quan *l'article era acceptat*, el seu optimisme, el seu recolzament des de l'altre costat de la taula i, sobretot, la facilitat per gestionar el dia a dia de la nostra convivència laboral.

Al Dr. Josep Roure que és una mica *culpable* de tot plegat perquè va ser ell qui em va dir *"el Marcos té en el cap un tema que podria donar lloc a una tesi doctoral..."* i això em va fer decidir a *tornar-ho a provar*. A ell no només li haig d'agrair aquesta empenta inicial sinó també, i molt especialment, la seva bonhomia que, en omplir tot l'espai disponible, no deixa lloc als sentiments negatius. Ell ja sap de què parlo.

Al responsable docent de l'Escola Universitària Politècnica de Mataró, l'Eduard de Bru de Sala, Sotsdirector Cap d'Estudis durant el període de temps en què aquesta tesi s'ha anat elaborant; i als coordinadors dels Graus, el Dr. Josep Roure, l'Albert Monté i el Joan Jou. Sense haver de concedir-me cap mena privilegi especial, sempre m'han proposat una assignació docent raonable que m'ha permès de *trobar forat* per a la recerca.

Al Juan Garcia, que sempre ha estat disposat ha donar-me un cop de mà per tal que *el meu anglès* sigui entenedor i soni el millor possible. També a la Colm Bremner a qui algunes vegades he abordat amb els meus dubtes lingüístics i qui mai no ha deixat d'atendre'm. *Nevertheless, any remaining mistakes are my sole responsibility*.

A tots els companys i companyes de l'Escola Universitària Politècnica de Mataró i en especial a tots aquells que en algun moment o altre se m'han atansat amb la pregunta *"i què, com va la tesi?"* Les seves mostres d'interès sempre han estat positivament esperonadores. I d'entre els companys de l'Escola no puc deixar d'esmentar el Dr. Santiago Ortego (el Santi) amb qui vaig tenir el privilegi de compartir despatx fins que la malaltia li va guanyar la batalla. Tot el que passa de bo a l'Escola té la seva llavor en el seu treball incansable.

A tots els amics i amigues que heu hagut de patir la llauna de *"estic fent una tesi sobre reconeixement de persones a partir de l'escriptura"*. Potser poc us ho pensàveu, però mentre us ho explicava m'anaven passant coses pel cap, algunes de les quals m'il·luminaven el camí.

I, finalment, als qui van endegar el projecte que va donar lloc a la base de dades BiosecurID i a tots els donants anònims que van contribuir-hi amb les seves mostres. Sense aquesta base de dades el més probable és que aquest treball no hagués existit mai.

Abstract

Biometric writer recognition is a very active field of research that has produced a considerable body of scientific work and papers over the last decades. Signature verification has been one of the most favoured approaches while others, such as text-dependent recognition, have attracted very little attention due to a perceived lack of practical applicability. This dissertation is circumscribed to the particular field of online text-dependent writer recognition, based on short sequences of text, a field where the number of relevant reference is really scarce. In this context, *online* means that not only the static images of the handwriting are available but also their dynamic dimension, that is, data is available as a function of time. What is more, thanks to modern acquisition devices, other time-dependent information is also available: the pressure exerted while writing, the angles of the writing device with respect to the horizontal plane and with respect to the vertical axis, and even the trajectories described while not exerting any pressure on the writing surface, when the hand moves in the air while transitioning from one stroke to the next (in-air trajectories made up of pen-up strokes).

The main motivation of this dissertation is the exploration of the aforementioned field of online text-dependent writer recognition, in order to provide evidence of the usefulness of short sequences of text to perform identification and verification, which are the two tasks involved in recognition. From this motivation stem its main goals and contributions: an exploration performed from a practical perspective, thus requiring the development of a recognition system, and the gathering of evidence concerning the discriminative power of in-air trajectories, i.e. their ability to discriminate among writers.

In-air and on-surface trajectories have been analyzed from the perspective of information theory and the results yielded by this analysis show that, except for pressure, they contain virtually equal amounts of information and are notably non-redundant. This suggests that in-air trajectories may have a considerable discriminative power and that they may help improve the overall recognition performance when combined with on-surface trajectories.

An innovative writer recognition system that fulfils the abovementioned practical goal has been devised. It follows an allographic approach, that is, it does not take into account the global characteristics of the text but focuses on character and character-fragment shapes. Strokes are considered the structural units of handwriting and any piece of text is regarded as two separate sequences, one of pen-up and one of pen-down strokes. The system relies on a pair of catalogues of strokes, built in an unsupervised manner by means of self-organizing maps, which allow mapping sequences of strokes into sequences of integers. The latter sequences, much simpler than the original ones, can be effectively compared by means of dynamic time warping, which takes advantage of the neighbouring properties exhibited by self-organizing maps. Measures obtained from each sequence can be combined in a later step.

The recognition system has been experimentally tested using 16 uppercase words from the BiosecurID database, which contains 4 executions of each word donated by 400 writers. The experimental results obtained clearly sustain the claim that online words have a notable

recognition potential and show the suitability of the allographic approach to perform writer recognition in the online text-dependent context. Regarding identification, the system compares positively to other word-based identification schemas. As for verification, the accuracy levels attained do not lie much below the accuracies reported for today's state-of-the-art signature verification methods. Furthermore, the results obtained from in-air trajectories have substantiated what the information analysis had already suggested: their considerable recognition power and their notable non-redundancy with respect to on-surface trajectories.

Finally, a new method to generate synthetic samples of online words from real ones has been proposed. This method is based on the recognition system previously described, takes advantage of its main characteristics and can be seamlessly integrated into it. Synthetic samples are used to enlarge the enrolment sets, which has the effect of substantially improving the recognition accuracy of the system.

Resum

El reconeixement biomètric de persones basat en l'escriptura (el reconeixement d'escriptors) és un camp de recerca molt actiu que ha generat, al llarg de les darreres dècades, un volum considerable de treballs i articles científics. La verificació de signatures ha estat la modalitat que ha atret més interès mentre que d'altres, com ara el reconeixement basat en el text, han rebut molta menys atenció perquè no se'n percebien les possibles aplicacions pràctiques. Aquesta tesi doctoral es circumscriu en el camp del reconeixement d'escriptors en la modalitat *online*, depenent del text (*text-dependent*) i basat en seqüències curtes, un camp en què el número de referències rellevants és veritablement escàs. En aquest context, *online* significa que, a més de disposar de les imatges estàtiques de l'escriptura, també es disposa de la seva dimensió dinàmica, això és, les dades són accessibles com a funció del temps. També, gràcies als dispositius d'adquisició moderns, altres informacions que depenen del temps són accessibles: la pressió que s'exerceix mentre s'escriu, els angles de l'instrument d'escriptura respecte del pla horitzontal i de l'eix vertical i, fins i tot, les trajectòries descrites quan no s'exerceix cap mena de pressió sobre la superfície d'escriptura, en els intervals en què la mà es desplaça d'un traç al següent (trajectòries *en l'aire* constituïdes per *traços elevats*).

La principal motivació d'aquesta dissertació és la investigació en el camp del reconeixement d'escriptors en la modalitat *online* depenent del text, amb intenció de proporcionar evidències que avalin la utilitat de les seqüències curtes per a la identificació i la verificació, que són les dues tasques compreses en el reconeixement. D'aquesta motivació se'n deriven els seus objectius més rellevants: una exploració feta des d'una perspectiva pràctica que requereix, doncs, el desenvolupament d'un sistema de reconeixement; i la recerca d'evidència relacionada amb la potència discriminant de les trajectòries *en l'aire*, això és, la seva capacitat per a reconèixer escriptors.

Les trajectòries *en l'aire* i *sobre la superfície* han estat analitzades des de la perspectiva de la teoria de la informació. Els resultats obtinguts d'aquesta anàlisi mostren que, llevat de la pressió, ambdós tipus de trajectòries contenen quantitats d'informació pràcticament idèntiques, amb un nivell notable de no redundància. Això suggereix que les trajectòries *en l'aire* potser posseeixen una potència discriminant considerable i que la capacitat global de reconeixement pot millorar si es combinen amb les trajectòries sobre la superfície.

S'ha desenvolupat un sistema de reconeixement innovador que representa l'assoliment de l'objectiu pràctic. Aquest sistema està basat en una aproximació al·logràfica, això és, no té en compte les característiques globals del text sinó que està focalitzat en les formes dels caràcters i dels seus fragments. Els traços són considerats la unitat estructural bàsica de l'escriptura i qualsevol fragment de text és entès com un parell de seqüències separades, una de traços *en superfície* i una de *traços elevats*. El sistema treballa en base a un parell de catàlegs de traços, construïts de manera no supervisada amb l'ajut de mapes autoorganitzats, que li permeten transformar les seqüències de traços en seqüències de números enters. Aquestes darreres seqüències, molt més simples que no pas les originals, poden ser comparades, de manera efectiva, mitjançant el *dynamic time warping* (alineament temporal dinàmic) el qual treu profit

de les propietats de veïnatge característiques dels mapes autoorganitzats. Les mesures que s'obtenen de cada seqüència poden ser combinades en un pas posterior.

El sistema de reconeixement ha estat provat experimentalment fent ús de les 16 paraules en majúscules de la base de dades BiosecurID, la qual en conté 4 realitzacions de cadascuna donades per 400 persones. Els resultats experimentals que s'han obtingut recolzen clarament l'afirmació que les paraules *online* presenten una potència discriminant notable i avalen l'adequació de l'aproximació al·logràfica per a dur a terme reconeixement d'escriptors en el context *online* depenent del text. Quant a la identificació, el sistema es compara favorablement amb altres mètodes basats en paraules. I, pel que fa a la verificació, els nivells de precisió obtinguts no es troben gaire lluny dels nivells assolits pels mètodes de verificació de signatura representatius de l'estat de l'art actual. És més, els resultats que s'obtenen de les trajectòries en l'aire han corroborat allò que havia estat suggerit per l'anàlisi de la informació: la seva considerable potència discriminant i la seva substancial manca de redundància respecte de les trajectòries *sobre la superfície*.

Finalment, s'ha proposat un nou sistema de generació de mostres sintètiques de paraules *online*. Aquest mètode està basat en el sistema de reconeixement abans descrit, n'aprofita les característiques principals i s'hi pot integrar amb facilitat. Les mostres sintètiques s'utilitzen per engrandir els conjunts d'inscripció (*enrolment sets*), la qual cosa té com a efecte una millora substancial de la precisió del sistema.

Contents

List of Tables

List of Figures

1

INTRODUCTION

This chapter is the starting point of this doctoral dissertation. First of all, the main motivation that led to the research that will be expounded upon in the forthcoming chapters is stated. From this motivation originate the objectives of the thesis, which are formulated in the second section, together with the contributions brought by this dissertation. The third section is intended to explain the main approach that will be followed and the experimental context in which the reported research is set. The fourth section provides the bibliographic references of the papers produced during the research period culminating in this dissertation. Finally, the fifth and last section outlines the contents of this document.

1.1 MOTIVATION

In today's society, with security being a matter of growing concern, biometrics has conquered an important role in applications where the identification of individuals is deemed as an issue of paramount importance.

Biometrics is not a single field of research but a variety of not necessarily tightly coupled fields that share a similar goal (i.e. the accurate identification of individuals) and similar methods and techniques (e.g. pattern matching). Some physiological modalities, such as fingerprint, iris or face geometry, are currently being used in practical scenarios (e.g. airports), while behavioural modalities, such us handwriting, still lack the recognition provided by the deployment in real environments. Nevertheless, handwriting-based recognition is a very active field of research and the results reported in scientific literature are quite promising.

Handwriting itself cannot be regarded as a monolithic field since at least two different focuses of attention can be considered: signature-based recognition and text-based recognition. Although both focuses share a common foundation, they may require different approaches. What is more, it is not even clear whether signature and text are the same or different biometric modalities, with some authors opting for the latter. Regardless of whether they are a single or two different modalities, signature has received much more attention than text. Does this mean that (handwriting of) text is not worth being explored as a biometric modality? The motivation of this dissertation is to give an answer to this question. To put it another way, **this dissertation has as its main motivation to provide evidence that online text, actually short sequences of online text, has a recognition potential not far from that of signatures, and that it deserves to be considered on its own.**

1.2 OBJECTIVES AND CONTRIBUTIONS

The abovementioned motivation crystallises as the two main objectives of this dissertation:

1. The first objective can be straightforwardly stated: **to explore the issue of the recognition potential of short sequences of online text from a practical perspective.**

2. Since modern acquisition devices such as digitizing tablets and other pressure-sensitive devices (e.g. tablets) are becoming more popular and widely used, and since these online devices can not only gather *invisible* information (e.g. pressure and writing angles) but even gather it when the writer is not pressing the writing device against the writing surface, the following question arises: to what extent is this *invisible* information meaningful and useful for writer-recognition purposes? Thus, the second objective of this dissertation can be formulated as follows: **to gather evidence of the recognition power (discriminative power) of the in-air trajectories performed when handwriting**.

During the research period leading to the main results reported in this dissertation, when it was already clear that the two main objectives would be fulfilled, a third objective sprang up:

3. The accuracy of biometric recognition systems is highly dependent on the size of the set of samples used to model the users. A simple, although quite naïve, solution to this problem is just to get more samples from each user. In the context of the research that was being conducted, this solution was absolutely out of the question since it was virtually impossible to acquire any single extra sample. But, is it possible to synthetically enlarge a set of samples in a way that increases recognition accuracy? Thus, the third objective of this dissertation is **to explore the possibility of synthetically enlarging the sets of samples, aiming at improving recognition performance**.

Each objective has led to one or more contributions:

(a) In order to explore the recognition potential of short sequences of online text, **a system to perform online text-based writer recognition has been proposed and experimentally evaluated**. This system is based on an innovative idea: the combined use of Self-Organizing Maps and Dynamic Time Warping, and has achieved an accuracy level that exceeds the initial expectations.

(b) The second objective has been fulfilled twofold:
   i. **The information contained in the in-air trajectories of the handwriting has been analyzed from the perspective of the information theory**. The results firmly suggest that in-air trajectories are rich in information (almost as rich as on-surface trajectories) and that this information is, to a considerable extent, non-redundant.
   ii. **The proposed system is capable of effectively performing recognition using only in-air trajectories**. The results not only corroborate their potential but also their non-redundant nature.

(c) Regarding the third objective, **a method to generate synthetic samples from real ones has been proposed**. This method does not only take advantage of the characteristics of the writer recognition system that constitutes the first contribution, but can also be seamlessly integrated into it. The experimental evaluation of this method has yielded promising results.

1.3 APPROACH AND EXPERIMENTAL CONTEXT

Writer recognition can be considered from different perspectives:

1. Regarding the information available, writer recognition can be performed online, when spatiotemporal (dynamic) data is available, or offline when only static images are available

2. As for the contents of the handwriting itself, it can be performed in a text-dependent way, when it is mandatory that the writer executes a *predefined* text, or it can be performed in a text-independent way, when it is not necessary to execute any predefined text. To some extent, signature-based recognition can be considered as a text-dependent approach.

3. With respect to the qualities of the handwriting that are taken into account, it can follow a structural approach, when attention is focused on global characteristics of the text (e.g. curvature), or it can follow an allographic approach, when attention is focused on character shapes or on character-fragment shapes

The work reported in this dissertation pertains to the online, text-dependent, allographic approach (see Fig. 1.1). To the best of the author's knowledge, no relevant scientific references exist that report research in this specific area. Therefore, this dissertation explores a *terra incognita* area in the writer-recognition field.

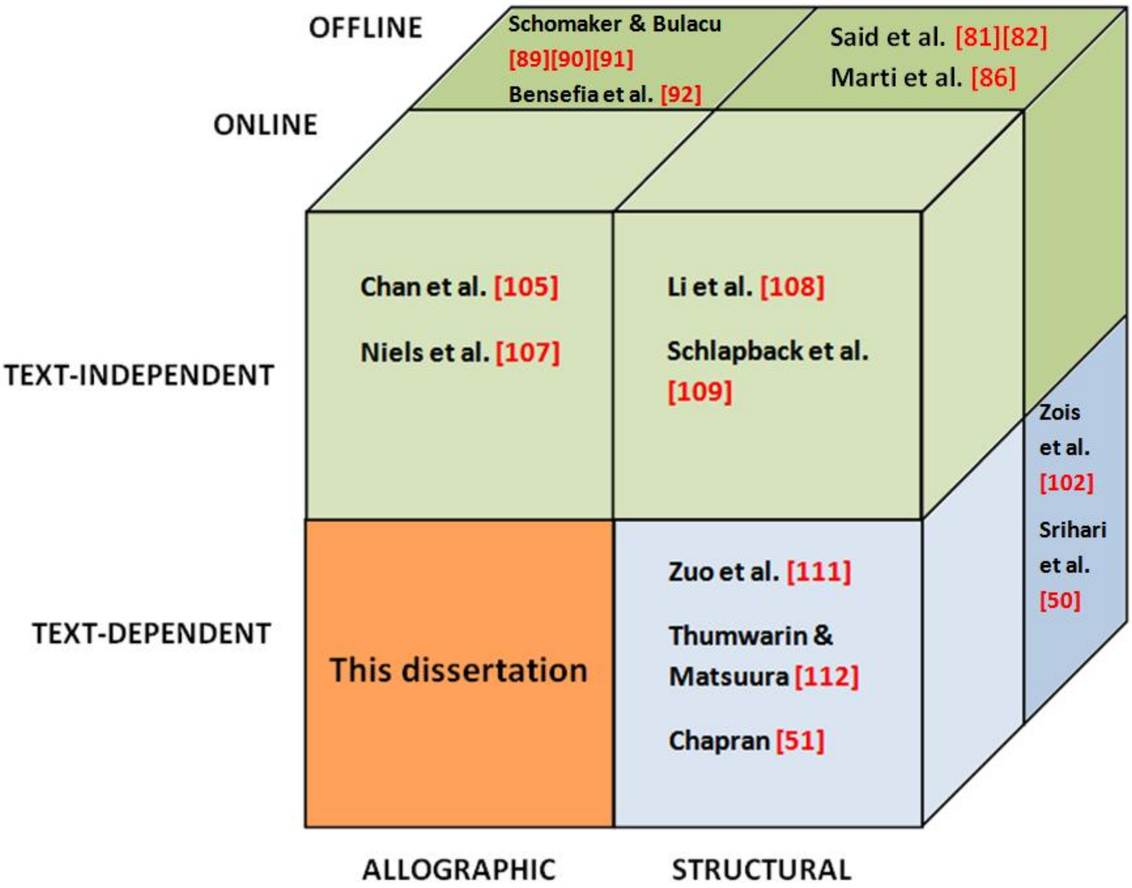


Figure 1.1: Approach of this dissertation and references to relevant works following other approaches. To the best of the author's knowledge, the invisible octant (offline text-dependent allographic approaches) contains no relevant references.

To be fulfilled, the objectives of this dissertation require an experimental context where the devised system and the hypothesis postulated can be put to the test. Quite often, research in biometrics depends upon the existence of databases that contain samples from a large enough number of donors to experiment on. Regarding handwriting other than signature, very few publicly available databases exist that can straightforwardly be put to use in the specific field where the research reported in this dissertation takes place. Some handwriting databases do exist, mainly offline though, but they were collected to be used in the recognition of the handwriting, not of the writer that had produced it. Fortunately, a joint effort of several Spanish universities led to the acquisition of the BiosecurID database [1]. This multimodal database, which will be reviewed in a forthcoming chapter, is the true experimental context of this dissertation. All the experiments that will be reported upon have been conducted using data from the BiosecurID.

Among others, the BiosecurID database contains 4 repetitions of 16 uppercase Spanish words (Fig. 1.2) donated by 400 writers.

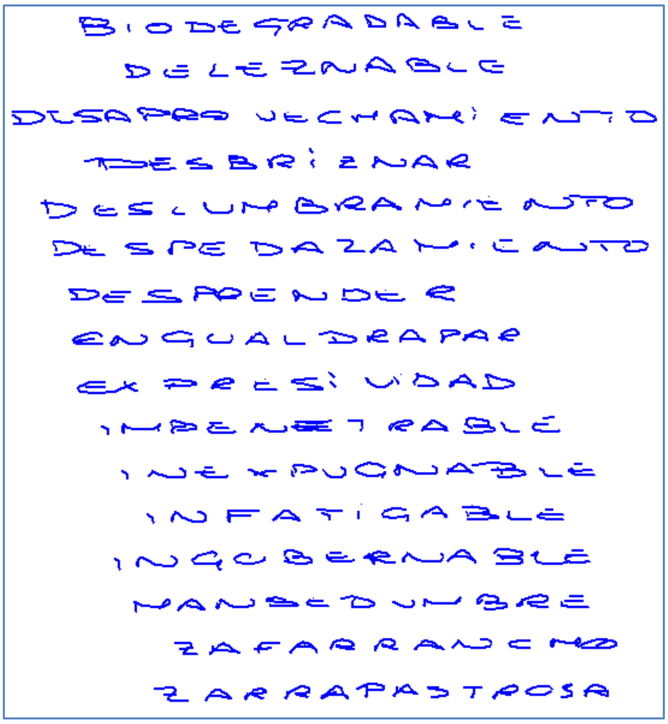


Figure 1.2: The 16 uppercase Spanish words in the BiosecurID database, written by one of the donors (see section 3.4.2 for further details).

These 16 uppercase Spanish words will be the *short sequences of online text* referred to when stating the objectives of this dissertation.

The fact that only uppercase words are considered in the experiments should not be regarded as a negative issue. Two executions of the same word by two different writers may be more similar in uppercase than in lowercase. Thus, contrary to some intuitions, uppercase-based writer recognition seems to pose a more challenging problem, since the less personal the style the more difficult to distinguish one writer from the rest because there are fewer differences among different writers. What is more, uppercase handwriting seems to be more resilient to changes due to aging and mental conditions. Also, uppercase is very often the preferred writing method in online-gathering capable devices such as tablets and PDAs.

1.4 PUBLICATIONS

During the research period leading to this dissertation, several papers have been submitted to journals and conferences. The following ones have been published or are already accepted for publication.

1.4.1 INDEXED JOURNALS

- SESA-NOGUERAS, E. AND FAUNDEZ-ZANUY, M. 2012. Biometric recognition using online uppercase handwritten text. ***Pattern Recognition*** 45, 128-144. (Available online at: http://dx.doi.org/10.1016/j.patcog.2011.06.002)

- SESA-NOGUERAS, E., FAUNDEZ-ZANUY, M. AND MEKYSKA, J. 2012. An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting. ***Cognitive Computation*** 4, 195-205 (available online at http://dx.doi.org/10.1007/s12559-011-9119-y). Doi: 10.1007/s12559-011-9119-y.

- SESA-NOGUERAS, E. AND FAUNDEZ-ZANUY, M. Writer recognition enhancement by means of synthetically generated handwritten text. ***Engineering Applications of Artificial Intelligence***
(In press, available online at http://dx.doi.org/10.1016/j.engappai.2012.03.010). Doi: 10.1016/j.engappai.2012.03.010

1.4.2 CONFERENCES AND WORKSHOPS

- SESA NOGUERAS, E. 2011. Discriminative power of online handwritten words for writer recognition. In Proceedings of the *2011* ***IEEE International Carnahan Conference on Security Technology*** *(ICCST),* OES Publications, Lexington, Ky. (Available online at: http://dx.doi.org/10.1109/CCST.2011.6095953)

- SESA-NOGUERAS, E. AND FAUNDEZ-ZANUY, M. 2011 Writer recognition by means of stroke categorization based on self-organizing maps. In *Proceedings of the* ***21th Italian Workshop on Neural Networks (WIRN 2011)****,* IOS Press.

1.5 THESIS OUTLINE

From this chapter on, this dissertation has the following structure:

- Chapter 2 is an **introduction to biometrics**. Its purpose is to provide the reader with a general overview of the topic, emphasizing aspects such as the existence of multiple modalities, the parameters usually considered to assess their suitability for identification purposes, the architecture of biometric recognition systems and the accuracy metrics that measure their performances.

- **Handwriting**, a specific biometric modality, is the subject of chapter 3. First, the topic is briefly introduced from two complementary perspectives: regarding it as a process and as the result of that process, that is, a sequence of graphemes on a writing surface. Then, two important issues are introduced: the discriminative power of short sequences of text and the comparison between text and signature. Online handwriting, that is, handwriting performed with capturing devices able to acquire the temporal dimension of the process, is also introduced in this chapter. Finally, a brief survey of handwriting databases is provided, paying special attention to the BiosecurID database, the database containing all the handwriting samples used in the experiments reported in this dissertation.

- Chapter 4 is entirely devoted to **the state of the art in writer recognition**. Two main approaches are considered: text-based and signature-based writer recognition. Regarding the former, works addressing the issue of individuality are reviewed first, and, after that, relevant references pertaining to the online and offline fields are commented upon. The section dealing with text-based approaches is concluded with a comprehensive summary that includes most of the references discussed. As for signature-based methods, the most successful approaches are considered and special attention is paid to the results obtained in signature competitions since they are quite often considered the state-of-the-art results on which comparisons should be made.

- In chapter 5, **the contents of in-air and on-surface trajectories are compared and analysed from the perspective of information theory**. First, for each type of trajectory and recorded feature, the amount of information (entropy) is computed. As the results strongly suggest that both types of trajectories contain equivalent amounts of information, the degree of redundancy is also analysed. Finally, the inter- and intra-writer differences are also scrutinized.

- **The recognition system** that materialises one of the main objectives of this dissertation is thoroughly presented in chapter 6. First section provides a general overview, emphasizing the stroke-based approach, the use of automatically generated catalogues of strokes and how Dynamic Time Warping is efficiently utilized to compare very short sequences of integers that represent executions of online words. The second section of this chapter provides a much more detailed view of the system, emphasizing relevant aspects such as the segmentation and pre-processing of the strokes, the automatic categorization of strokes by means of self-organizing maps, the encoding and subsequent comparison of words and the combination of dissimilarity measures to obtain word-level measures that encompass partial measures from in-air and on-surface trajectories.

- Chapter 7 reports on all the **experiments performed in order to assess the performance of the system** presented in the previous chapter, along with the performances related to the in-air and on surface trajectories and their combinations. First, the reference experiment, the results of which will be considered as the baseline for comparisons, is presented. Later on, the following issues are closely scrutinized: the impact of the origin and the number of samples used for building the catalogues, the impact the size of the catalogues and the impact of the overall number of words considered.

- Chapter 8 addresses the issue of the **recognition enhancement achieved when the enrolment sets are enlarged with synthetically generated samples**. Firstly, the topic is motivated, then related work is briefly reviewed, and next the proposed generation method is presented. The experimental results obtained are shown and discussed in the last sections of the chapter.

- Chapter 9 **concludes this dissertation**. First, the motivation, the objectives and the main contributions are briefly reviewed so that they can be linked to main conclusions. Then, these main conclusions are explained in detail and some other conclusions are enumerated and commented upon. A glimpse to some research lines, steaming from the work just presented puts an end to the chapter and the dissertation.

2

INTRODUCTION TO BIOMETRICS

This chapter is intended as a brief and general overview of the field of biometrics, to which this dissertation is circumscribed. It is organized as follows: the first section provides a definition of the term *biometrics,* introduces its physiological and behavioural modalities and compares biometric and non-biometric authentication methods. Biometric recognition systems, for identification and verification, are introduced in the second section. In the third section the main biometric modalities are enumerated and concisely introduced. The fourth and last section presents the metrics more commonly used for measuring the accuracy of biometric systems.

2.1 BIOMETRICS TODAY

The online version of the Merriam-Webster dictionary [2] provides the following definition for the term biometrics:

> *The measurement and analysis of unique physical or behavioural characteristics (as fingerprint or voice patterns) especially as a means of verifying personal identity*

From an etymological perspective, the noun biometrics originates from the Greek words βίος (***bios***, life) and μέτρον (***metron***, measure). In its earlier usages (circa 1900) it meant *application of mathematics to biology*. The very closely related term *biometry* (coined circa 1830) meant *calculation of life expectancy* [3]. Since its coinage, the term biometrics has acquired a *verification of the identity of individuals* flavour[1]. Indeed, biometrics has an increasing importance in the field of security applications. As this thesis is circumscribed to this particular field, this sense is going to be adopted. Nevertheless, it must be said that biometrics is starting to get relevance beyond the field of security applications (e.g. in health applications [4].) The term *biometric* (noun) is very often used with the following sense: *Information about someone's traits which can be used to prove their identity.* In order to avoid confusions between biometrics and biometric, *biometric modality* will be used instead of the latter.

[1] In [170], it is defined as

> *The science of establishing the identity of an individual based on the physical, chemical or behavioral attributes of the person* [Sic].

David Zhang, in [171], states that

> *[…] we are usually concerned with technologies that analyze human characteristics for automatically recognizing or verifying identity […]* [Sic].

Depending on the traits under consideration, different biometric modalities exist. Typically, two categories are considered: ***physiological biometric modalities*** and ***behavioural biometric modalities***. Physiological modalities are based on direct measurements of parts of the human body. Among them: face, iris, retina, hand-palm, hand-veins, wrist-veins and friction-ridges of the fingers. Behavioural modalities are based on measurements and data derived from actions performed by a person. Handwriting, typewriting (key stroking), speech, gait and gesture are biometric modalities belonging to this category. Notice that behavioural modalities depend on indirect measures of some characteristics of the human body, thus all behavioural modalities have a physiological component.

It is the growing concern about security in today's society what is giving biometrics its important role in personal identification. In the later years, the demand of strong authentication methods and technologies has notably increased. As biometrics is regarded as one such strong authentication technology [5], it has become an important and very active field of research.

The **identity of an individual** can be defined as the information associated with that person in a given context, while an **identifier** is something that points to an individual (or to their identity). **Identity authentication** is the process of establishing an understood level of confidence that an identifier refers to a specific identity whereas **individual authentication** is the process of establishing an understood level of confidence that an identifier refers to a specific individual. In the biometrics literature, individual authentication of an identifier claimed by an individual is very often called **verification** [6]. Identifiers can be based on something material that a person possesses, on something that they know, on something that the person is or on something that the person can do. The two last types of identifiers belong to the category of biometric identifiers since physiological modalities involve something that a person is, while behavioural modalities involve something that a person can do. Table 2.1 summarizes and compares biometric and non-biometric authentication methods.

METHODS	BASED ON	EXAMPLES	GENERAL ADVANTAGES	GENERAL DRAWBACKS
Non biometric	Something material possessed (token)	Id-cards, passports, keys	Quite standard. Well accepted. Can be (re)issued.	Can be lost, stolen, faked, or shared. One individual, multiple identities
	Something known	Passwords, pin-numbers	Simple and economical. Can be changed if compromised.	Can be guessed and/or cracked. Can be forgotten. Can be shared. One individual multiple identities
Biometric	Something an individual is	Physiological traits: fingerprint, iris structure	Difficult or impossible to lose, steal, forget , share or fake	Cannot be changed or replaced. May vary over time (e.g. aging)
	Something an individual can do	Behavioural traits: handwriting, gait	Difficult or impossible to lose, steal or share	Can be faked. Difficult or impossible to change or replace. May vary over time (e.g. aging)

Table 2.1: Comparison of biometric and non-biometric authentication methods.

2.2 BIOMETRIC RECOGNITION SYSTEMS

Biometric recognition systems can be regarded as pattern recognition systems that compare acquired biometric data with models (or templates) stored in a database. Depending on the biometric modality and on the design of the system itself, the comparison may be based on different features extracted from the data. Usually, biometric systems operate in two phases: the **enrolment phase**, the first one, and the **testing phase**, the second. During the enrolment phase the user provides one or more samples of their biometric trait. These samples are pre-processed and the required features are extracted from the resulting data. With this information a model of the user is built and stored within the system. During the testing phase, the user provides a sample that undergoes a similar pre-processing, feature extraction and model computation procedure. A matching engine then compares the newly obtained model with one or all of the previously stored models. The context of application determines whether the system performs identification (Fig. 2.1) or verification (Fig. 2.2). Identification involves a one-to-many search where, given a sample of unknown origin, the goal is to determine whose sample it is. Identification can be used for negative recognition purposes [7] when it is necessary to establish whether a person is who he/she denies to be. Identification can also be used for positive recognition when, for convenience, the user is not required to claim an identity [8]. It is worth noticing that negative recognition can only be performed through biometric methods. Verification is a one-to-one comparison where, given two samples, the objective is to determine whether they originated from the same individual or not. Verification is usually performed for positive recognition purposes, aiming at preventing multiple people from using the same identity [7] [8].

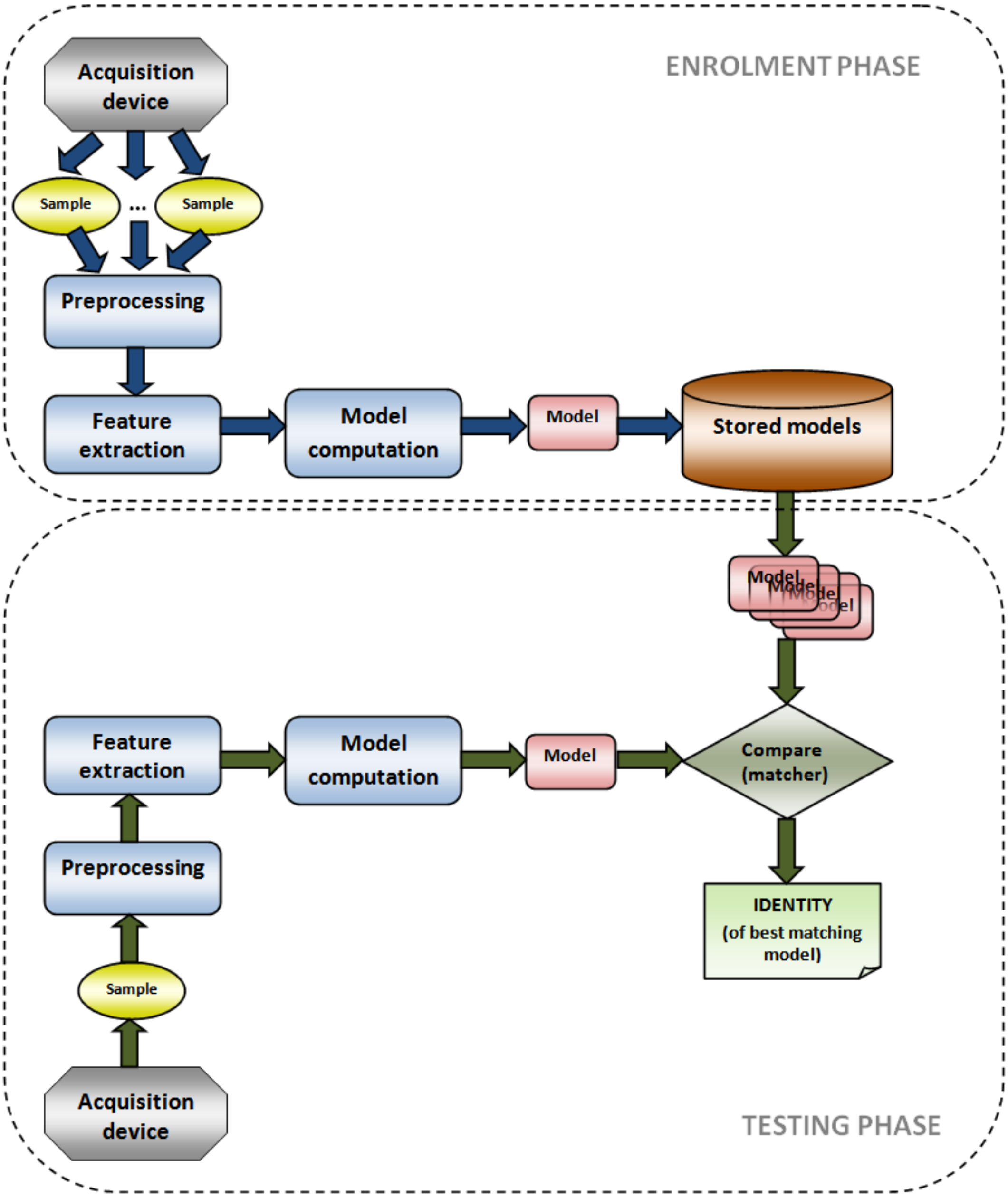


Figure 2.1: Graphical depiction of biometric system to perform identification.

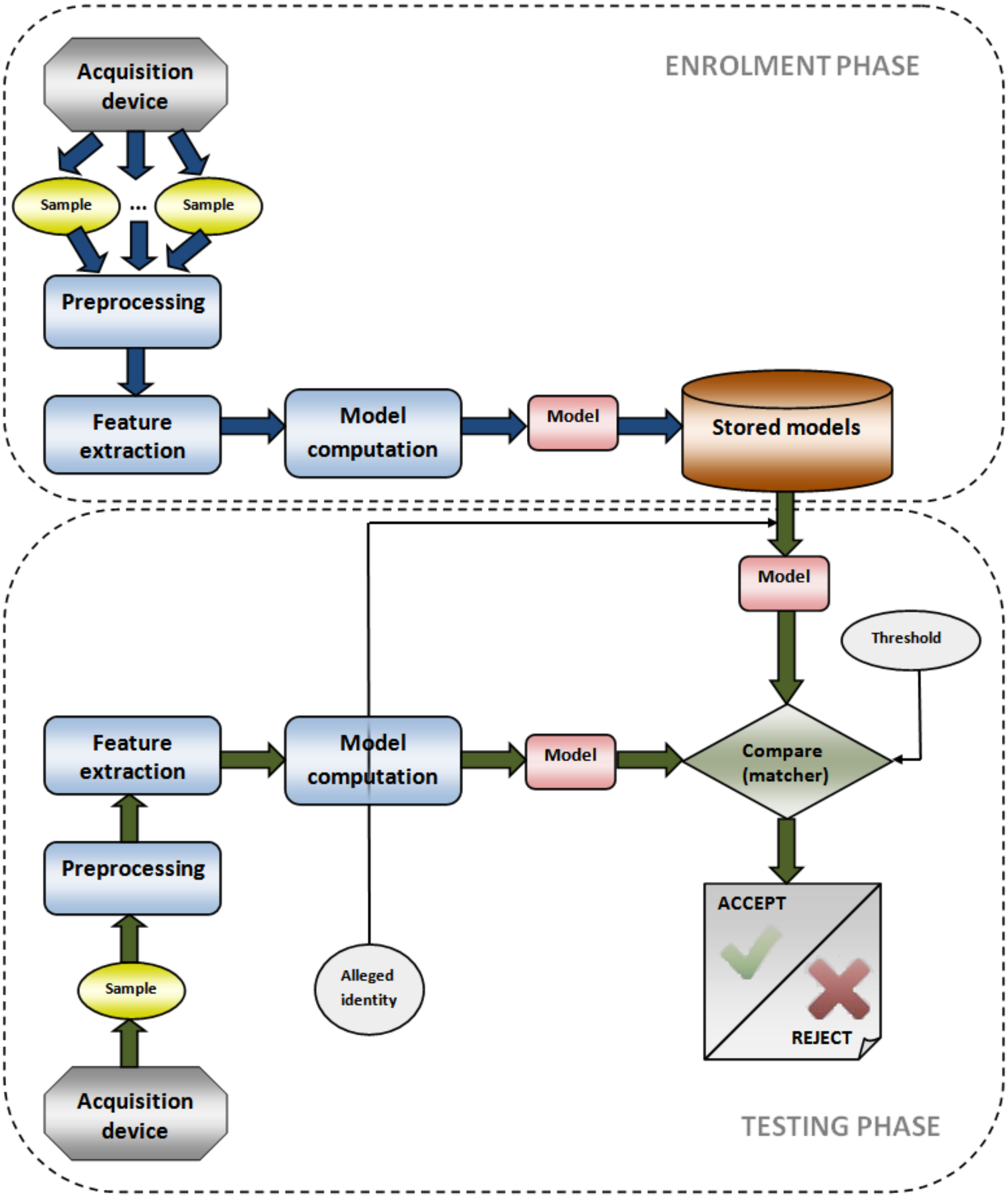


Figure 2.2: Graphical depiction of biometric system to perform verification.

2.3 BIOMETRIC MODALITIES

2.3.1 MAIN BIOMETRIC MODALITIES

FACE. Face recognition is a task that humans perform routinely and almost effortlessly. The increased availability of computing power has put face recognition in the arenas of security systems, human-computer interaction and others [9]. The distinctive features of the human face (eyes, mouth nose, eyebrows ...) extracted from still or moving images are used to recognize individuals. Face-recognition systems can be based and integrated with existing surveillance systems (e.g. CCTV systems). Usually, user cooperation is required during the enrolment stage but not necessarily afterwards, during the testing stage, when most systems can operate without user involvement. These systems arise some concerns due to privacy considerations and their acceptance is not always high. Facial recognition technologies have some notable weaknesses regarding acquisition conditions, especially lighting and angle. In poor conditions, accuracy can diminish drastically.

VOICE. Speaker recognition systems (or voice-scan systems) aim at extracting, characterizing and recognizing the information in the speech signal conveying speaker identity [10]. Thus, they recognize the individual who is speaking based on their vocal characteristics. They should not be confused with speech recognition systems, the main goal of which is to translate what a person is saying. Speech recognition may be used for purposes other than the authentication of the speaker or in conjunction with speaker recognition. This modality has some important drawbacks: Illness (cold, flu, aphonia) and the emotional state can drastically change a person's voice. Voice also changes with aging. Finally, some acquisition conditions, especially ambient noise, can lead to poor performances.

FINGERPRINT. Finger-scan systems are the most commonly deployed biometric systems. They are based on the comparison of the patterns of ridges and valleys (known as minutiae) found in the tips of the human fingers [11]. Fingerprints have a long tradition in forensic applications and have been successfully used for decades, possessing some advantages over other biometric modalities: they are quite stable, highly discriminative (even monozygotic –identical-twins do not share their fingerprints) and easily acquired. They have some drawbacks too: some (very rare) medical conditions involve the absence of fingerprints, and manual workers manipulating certain substances (acid, cement) may have they fingerprints altered or erased. Due to the extended use of fingerprints in forensics and their linking to law enforcement agencies and crime scene investigations (popularized by some TV series), finger-scan systems may arise some concerns regarding privacy and misuse.

Figure 2.3: Detail of the structure of ridges and valleys in a fingertip.

HAND-GEOMETRY. The distinctive aspects of the geometry of the hand, which include the height and width of the back, the height and width of the fingers, the area and the perimeter, have proven to be useful to perform recognition [12]. Hand-scan systems based on hand-geometry are generally used for verification purposes only, since they lack the accuracy required for identification. Hand-geometry is affected by development, aging, and some medical conditions (e.g. arthritis). As an advantage, it is not affected by environmental factors that may have an effect on other biometric modalities. It is generally perceived as non-intrusive.

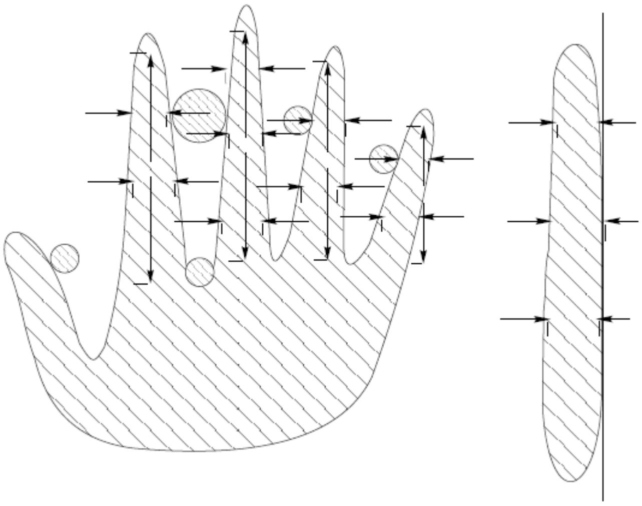


Figure 2.4: Typical hand geometry measurements (taken from [12]).

The geometry of the hand is not the only hand-related biometric modality. The patterns of the ridges and valleys found in the handprints (similar to those in the fingerprints) and the patterns of the veins in the back of the hand have also been considered as biometric modalities [13] .

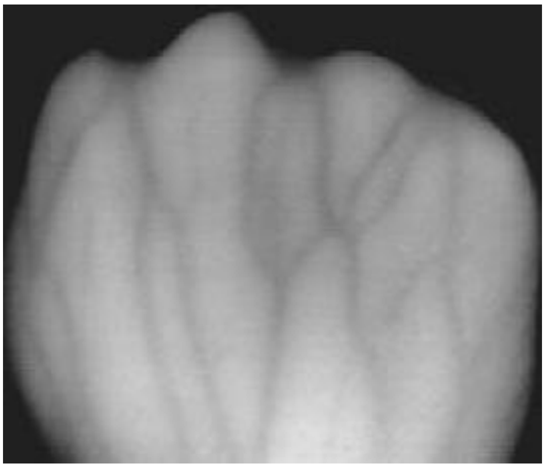


Figure 2.5: An infrared (IR) image of the vein patterns in the back of a hand (taken from [13]).

IRIS. individuals can be recognized based on the pattern of the iris, the elastic, pigmented, connective tissue responsible for controlling the diameter and size of the pupil. Iris-scan systems can achieve high levels of accuracy because the complex textures in this tissue are highly distinctive [14]. Iris patterns are very stable since they do not change over a lifetime. Despite its accuracy and stability, this biometric modality presents some important weaknesses: iris-based systems are difficult to use, tend to yield high rates of false negatives and a considerable number of users find them intimidating, even arising visceral reactions.

RETINA. Iris is not the only eye-based modality. Retina is also a biometric modality [5] . The retina is the light-sensitive tissue found on the surface of the back of the eye, whose blood vessels form a unique and highly discriminative pattern. Like iris-based systems, retina-based systems are difficult to use and many users find them quite intimidating.

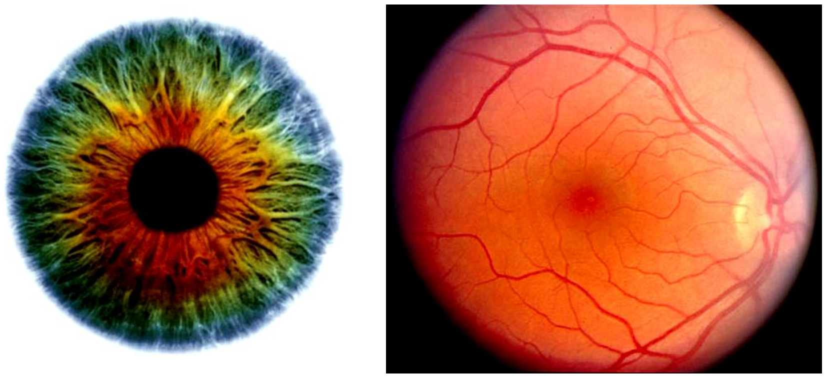


Figure 2.6: An image of the iris (left) and of the retina (right) (taken from [16] and [17] respectively).

SIGNATURE. Signature-based systems analyze the way a user signs their name, considering features such as the shape, the speed and the pressure exerted on the writing surface [5]. A set of signatures provided during the enrolment phase are compared against a newly executed one. The new signature is deemed authentic or not depending on the degree of similitude found. As people are quite used to signature as a way to prove one's identity, this biometric modality is generally well accepted and regarded as non-intrusive. Some drawbacks exist: the fact that people may not always sign in a consistent manner and the relative easiness to produce good fakes are among the most important. Signature is not the only biometric modality that relies on handwriting. Handwritten text is also a biometric modality, leading to the field of writer recognition. Although most authors use the term handwriting to refer to both text and signature, there is some evidence suggesting they may be different biometric modalities [18]. The state of the art on text-based recognition and on signature verification is reviewed in chapter 4.

2.3.2 OTHER BIOMETRIC MODALITIES AND MULTIBIOMETRICS

Other biometric modalities exist. Among them are key-stroking dynamics, where an individual is recognized by means of statistical information related to the manner and rhythm they type characters on a keyboard or keypad [19]; gait, where spatial-temporal and kinematic parameters related to locomotion can be used to recognize individuals [20][21] and ear, where the shape of the outer part of this organ provides discriminative information [22]. Also DNA can be considered a biometric modality although as of this writing it is almost exclusively used in the context of forensic applications [23].

Some of the limitations found in systems based on a single biometric modality can be overcome by the combination of two or more traits or by the combination of different features extracted from a single trait. Systems that rely on such combinations are known as multibiometric systems [24]. Multibiometrics relies on fusing information from various sources, with Biometric information fusion [25] being an active field of research.

2.3.3 COMPARISON OF BIOMETRIC MODALITIES

The variety and diversity of biometric modalities makes it difficult to compare them. In fact, no modality can be claimed to be the best to perform recognition since all have advantages and drawbacks. In order to assess the suitability of a biometric modality for security purposes, several parameters are usually considered [23].

- **Universality**: the trait on which the biometric modality is based should be possessed by every person.
- **Uniqueness** (or distinctiveness): given any two persons, the trait should be sufficiently different in terms of the features under consideration. Uniqueness is a parameter directly related to the discriminative power of the features extracted for any particular biometric modality.
- **Performance**: which refers to the recognition accuracy a particular modality may yield and also to the resources (time, memory, storage...) required to attain that accuracy, and all the factors that may have an impact on accuracy and speed. When it comes to recognition accuracy, several measures are considered: rate of false rejections, rate of false acceptances, rate of failures to enrol, rate of failures to acquire, etc. These will be considered in more detail in a further section.
- **Permanence** (or stability): referred to the degree of invariance over time of the trait considered, with respect to the comparison criterion used to decide whether two samples belong to the same individual or not. The notion of permanence can be extended to invariance with respect to other factors that may have a significant impact on the traits, such as health or weather conditions.
- **Non-circumventability**: which refers to the resistance to fraudulent usages, such as forging or supplantation (spoofing).
- **Collectability**: the degree of acquisition simplicity.
- **Acceptability:** which indicates the extent to which the target population is willing to accept the use of a particular biometric modality in their daily life. Acceptability depends on many different issues such as privacy concerns, intrusiveness, tradition or ease of use and also on collectability.

Although some of the aforementioned parameters related to biometric modalities mainly apply to the traits themselves (e.g. universality) others also apply and are strongly dependent on the system which uses them (e.g. performance.) Table 2.2 compares the biometric modalities previously discussed according to the seven aforementioned parameters.

MODALITY	PARAMETER						
	UNIVERSALITY	DISTINCTIVENESS	PERMANENCE	COLLECTABILITY	PERFORMANCE	ACCEPTABILITY	NON-CIRCUMVENTABILITY
Face	High	Low	Medium	High	Low	High	Low
Voice	Medium	Low	Low	Medium	Low	High	Low
Fingerprint	Medium	High	High	Medium	High	Medium	Medium
Hand-geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	High	Medium	High	Low	High
Retina	High	High	Medium	Low	High	Low	High
Signature	Low	Low	Low	High	Low	High	Low
Key-stroking	Low	Low	Low	Medium	Low	Medium	Medium
Gait	Medium	Low	Low	High	Low	High	Medium
DNA	High	High	High	Low	High	Low	High

Table 2.2: Comparison of biometric modalities (adapted from [8]).

2.4 PERFORMANCE OF BIOMETRIC SYSTEMS: ACCURACY METRICS

Given two models, the matching engine of a biometric system yields a **matching score**, a measure that quantifies the degree of similarity between them. The higher the score the more certain the system is that both models have been produced by the same individual. In verification, the decision whether the two models will be deemed from the same individual or not depends on a **threshold** *t*: models scoring over *t* will be considered from the same individual, while they will be considered from different individuals if their score is below *t*.

Due to a number of factors, two biometric samples from the same individual, even if acquired with a very short time difference between them, may not be identical. Thus, in biometric systems, **perfect matches**, that is, two models achieving the highest theoretical score, are unlikely to occur[2]. The variability between samples from the same individual is termed **intra-user (or intra-class) variability**, while the variability occurring between samples of different users is termed **inter-user (or inter-class) variability**. To be useful, a biometric modality should allow extracting features exhibiting small intra-user and large inter-user variabilities. Intra-user variability is responsible for **false non-matches**, a situation arising when individuals are not correctly identified as who they are. Inter-user variability leads to **false matches**, where an individual is successfully verified as someone else.

False non-matches, also known as false mismatches or false rejections, are type-I errors while false matches, also known as false acceptances, are type-II errors. All biometric systems are prone to these two types of errors. Type-I errors are measured by means of the false rejection rate (FRR) while type-II errors are measured by means of the false acceptance rate (FAR). The

[2] In fact, a perfect match between two models might indicate that a reply attack has taken place [170].

values of FAR and FRR can be changed by regulating the value of the threshold *t*. Nevertheless, it is not possible to decrease both error rates simultaneously: with a low value of t, the system is more tolerant to individual variations, meaning that FRR decrease but FAR increases. Higher values of t make the system more secure, that is, less tolerant to false acceptances, decreasing the value of FAR but increasing that of FRR (see Fig. 2.7).

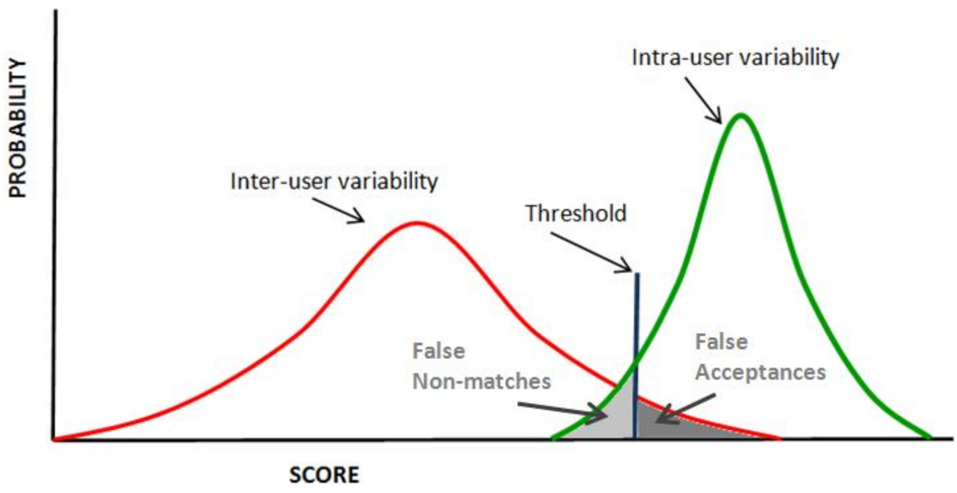


Figure 2.7: The two types of score variability. The point where the threshold is located determines the behaviour of the system with respect to false non-matches and false acceptances.

Systems with a low FAR can be considered secure and thus suitable to be deployed in environments where security is the main concern. When the value of the FAR is low, very few unauthorized individuals are incorrectly given access. Systems exhibiting a low FRR are considered comfortable for the (righteous) user since very rarely an authorized user is rejected. As FAR and FRR are interdependent, so are security and comfort.

The receiver operating characteristic (ROC) curve is a plot of the false acceptance rate (FAR) against the true acceptance rate (1-FRR) for all the operating points (values of the threshold t) [26]. It is a very usual way to depict the performance of biometric systems. In some case, values of FAR are plotted against those of FRR. Fig. 2.8 shows both types of ROC curves.

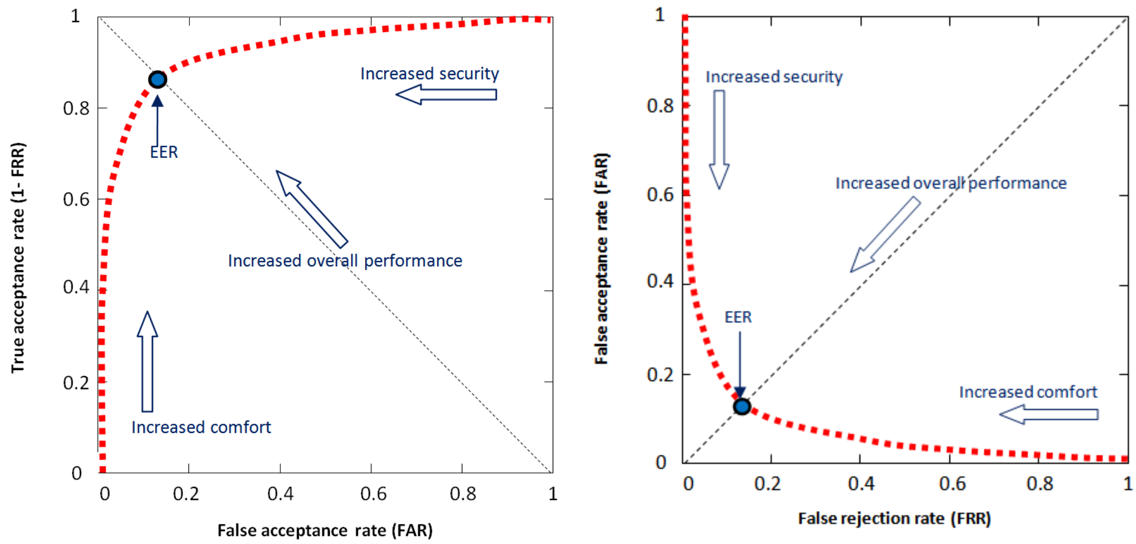


Figure 2.8: A ROC curve plotting true acceptance vs. false acceptance rates (left). An equivalent ROC curve plotting false rejection vs. false acceptance rates (right).

Detection Error Tradeoff (DET) curves are another usual way to depict the performance of biometric systems [27]. In DET-curves the values of FAR are plotted against the values of FRR in a normal deviate scale (Fig. 2.9).

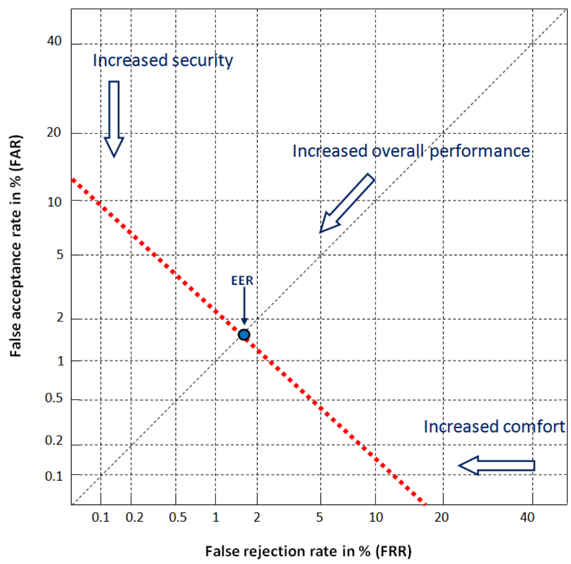


Figure 2.9: A DET-curve plotting false rejection vs. false acceptance rates.

Several measures have been proposed to summarize in a single value the performance of a biometric system. These measures allow, to a certain extent, the comparison of different biometric systems. When it comes to verification, the most commonly used are the equal error rate (EER) and the minimum of the detection cost function (min of DCF).

- The **equal error rate** (**EER**), also known as **crossover error rate** is the value that satisfies FAR = FRR. The lowest this value, the better the performance (overall accuracy) of the system. In ROC curves plotting FAR against FRR and in DET curves, the value of the EER is on the point where the curve cuts the diagonal.

- The **detection cost function** (**DCF**) is the tradeoff between false acceptances and false rejections when this two types of errors do not have the same importance. It is defined as

$$DCF = C_{fa} \cdot P_{fa} \cdot P_t + C_{fr} \cdot P_{fr} \cdot (1 - P_t) \qquad (2.1)$$

where C_{fa} and C_{fr} are the costs of a false acceptance and a false rejection respectively, P_{fa} and P_{fr} the probabilities of a false acceptance and a false rejection and P_t the a-priori probability that the user presented to the system is the one who claims to be (not an impostor). The values of P_{fa} and P_{fr} vary as a function of the rejection/acceptance threshold *t.* The minimum of DCF is the best possible tradeoff. When C_{fa} and C_{fr} are set to 1 and P_t is set to 0.5 then minimum of DCF is also known as the **half total error rate** (HTER) or simply the **verification error rate** (VE).

$$HTER = VE = {}^{1}\!/_{2}(P_{fa} + P_{fr}) \qquad (2.2)$$

Usually, the verification error rate is higher than the EER. Often, when a *positive view* is preferred, **verification accuracy** (**VA**) is used:

$$VA = 1 - {}^{1}\!/_{2}(P_{fa} + P_{fr}) = 1 - VE \qquad (2.3)$$

Quite often, VE and VA are expressed as percentages.

For identification, the ratio between the number of well-identified users and the number of enrolled users, the **identification ratio** (**IDR**), expressed as a percentage, is commonly used. It should be noticed that while FAR and FRR may be considered independent of the number of enrolled users, IDR is not independent of this number since IDR tends to decrease as the number of enrolled users increases.

As biometric systems are mainly intended to be used in security-related scenarios, FAR, FRR and values derived from these two are the measures commonly used to assess their performance. In forensic scenarios, IDR is also considered. Nevertheless, other accuracy metrics may be taken into account: the **failure to enrol rate** (FTE), which measures the proportion of potential users for whom it is not possible to build a model, and the **failure to acquire rate** (FTA), that measures the proportion of users for who it is not possible to extract the required features. The reasons for FTE and FTA mainly depend on the biometric modality under consideration. For instance, in a fingerprint-based system they may be caused by scarred fingers or difficulties in the placement of the fingers, in face-based systems they may be caused by objects worn by the subjects (glasses, hats, scarves, etc.) or by the lighting conditions, while in handwriting-based systems they may be due to inconsistent handwriting, often caused by poor training in the usage of the writing device [5].

3

HANDWRITING

The main purpose of this chapter is to introduce the topic of handwriting, not from a general perspective but emphasizing those aspects that are more relevant in the context of this dissertation. It is organized as follows: the first section gives a general idea about the complexity of the process involved in handwriting, briefly reviewing some popular models. The second section considers the outcome of the handwriting process, the handwriting itself, pointing out the differences and similitudes between text and signature. The third section is devoted to online handwriting, the modality on which the recognition method presented in this dissertation is based, underlining the distinction between on-surface and in-air trajectories. Finally, the fourth section describes some relevant databases containing handwriting samples, most of them cited in the chapter on the state of the art, and the BiosecurID database, the database that has been used in all the experiments reported in this dissertation.

3.1 HANDWRITING AS A PROCESS

The term handwriting may refer to the complex movements performed by the hand while writing a text or to the results of this process, that is, a piece of text written by hand. As a process, handwriting is a complex perceptual-motor task, a skill usually learnt at school. The hand is a very complex structure that contains 27 bones (including the wrist), more than 40 different muscles and that is innervated by 3 nerves each of which performs sensory and motor functions [28]. Different types of factors exert influence on the production of handwriting: the muscular movements involved in the process are controlled by the central nervous system and are partly outside the conscious control of the writer. Thus, the long and short term conditions within the central nervous system have an effect on the handwriting. Biomechanical factors such as the structure and size of the hand, the arm and the shoulder or the state of the muscles (stiffness, elasticity) also influence the produced handwriting. The cultural background not only determines the kind of characters written (Asian, Western, Arabic ...) but also influences directional preferences (are strokes mainly performed left-to-right or right-to-left? From top to down or bottom up?) and other factors, such as slant, that affect the shape of the produced graphemes.

3.1.1 MODELS OF HANDWRITING

Several theoretical models of handwriting generation have been postulated, seeking applicability in different, although related, fields: the automated generation of (realistic) cursive handwriting, the automated recognition of the handwriting and, to a lesser extent, the automated recognition of the writer. Although it is out of the scope of this dissertation to fully review the different models that have been proposed, the forthcoming paragraphs will provide some general insights on this matter. The interested reader can find comprehensive overviews elsewhere (e.g. [29], [30]) .

Cybernetics [31] contributed a system model based on a first-order feedback system (Fig. 3.1). These systems consist of three components (a comparator, an effector and a sensor) and a feedback loop. By means of a first-order feedback system, the handwriting process can be described as follows [32]: a target letter shape enters the system at a comparator, a subsystem that continuously produces a signal based on the difference between the target shape and the obtained (written) shape. The signal produced by the comparator activates the effector, the subsystem that generates the output. The sensor subsystem gathers information about the state of the effector (e.g. velocity), and the already produced shape (visual perception) and these data are fed to the comparator through the feedback-loop.

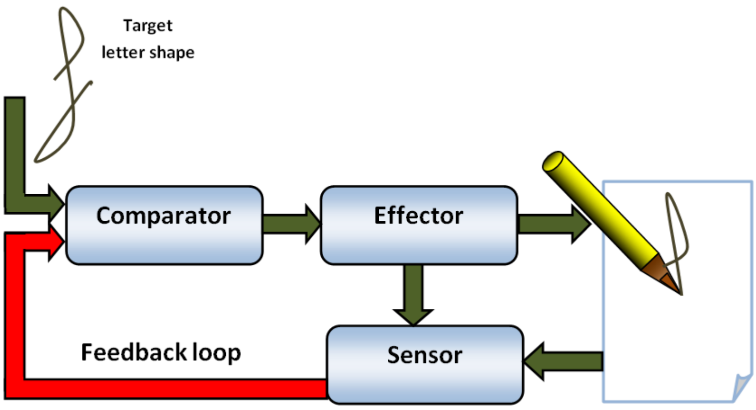


Figure 3.1: Graphical depiction of the first-order feedback system model for handwriting

This initial model, which provided a first systematic explanation of the handwriting process, presents some limitations. Because of timing constraints, it does not seem correct to assume that a continuous feedback signal can really exist in the nervous central system. In order to overcome this limitation the cognitive motor theory has been introduced. This newer approach is based on an open-loop. Only short sequences of simple movements, aimed at producing a limited number of shapes (strokes) are determined in advance. The feedback information does not affect the stroke under production but the production of future strokes, with learning allowing the adjustment of the process [33].

In 1981, Hollerbach proposed a model based on the assumption that handwriting is produced by two oscillatory processes which generate horizontal and vertical signals respectively [34]. These signals are combined and superimposed to a constant-velocity left-to-right movement of the hand over the writing surface. From a mechanical point of view, such signals can be produced by a mass-spring system consisting of a mass (the pen) attached to orthogonal antagonistic spring pairs (Fig. 3.2). The human motor system would be responsible for controlling the whole system, providing the required stimuli.

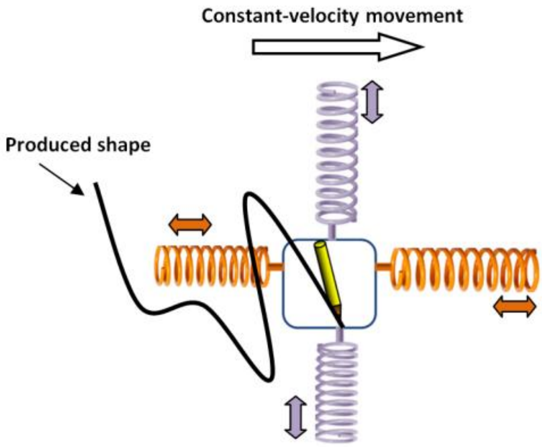


Figure 3.2: Graphical depiction of the model for handwriting generation proposed in [34].

In [35] Schomaker, Thomassen and Teulings propose a multi-level model in which the generation of strokes (in the lowest level of the model) is performed following Hollerbach's model.

At the stroke-level, the kinematic theory of rapid human movements [36,37] has produced the **sigma-lognormal model**. According to this theory, a stroke is the result of a synergy involving agonist and antagonist neuromuscular systems and can be characterized by a lognormal velocity profile (Fig. 3.3).

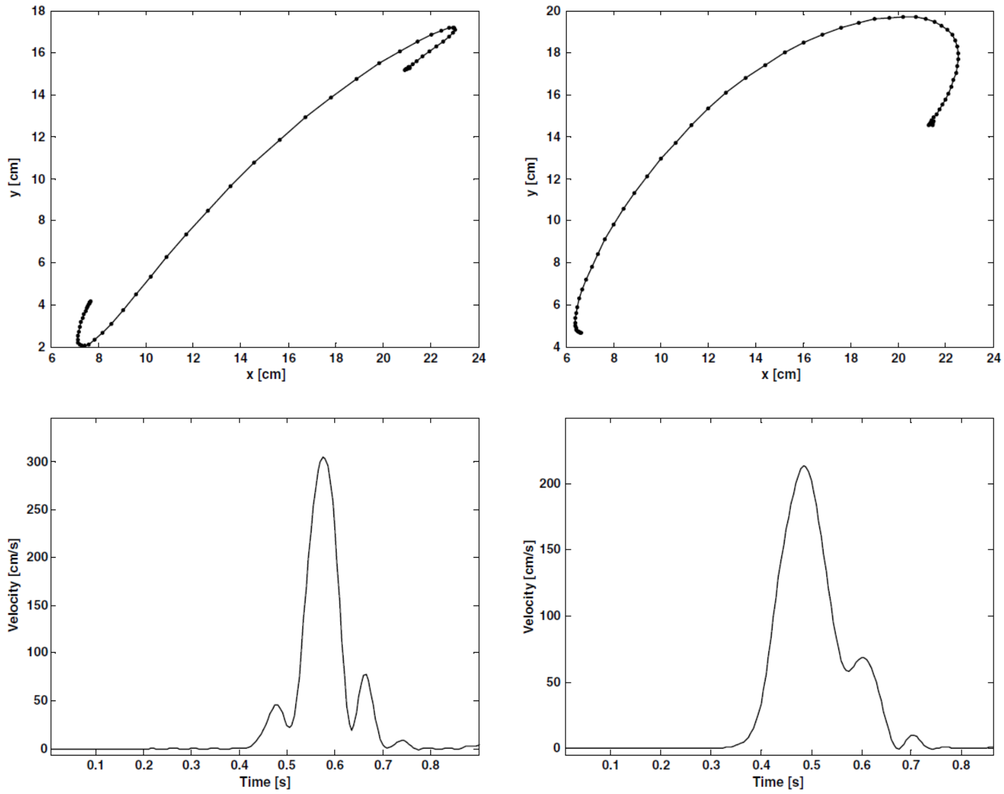


Figure 3.3: Two strokes (up) and their velocity profiles (down). Image taken from [38].

Each stroke is described by its velocity module

$$|\vec{v}(t)| = \frac{D}{\sigma\sqrt{2\pi}(t-t_0)} e^{\frac{(\ln(t-t_o)-\mu)^2}{-2\sigma^2}} \qquad (3.1)$$

and its direction

$$\varphi(t) = \theta_s + \frac{\theta_e - \theta_s}{2}(1 + erf(\frac{\ln(t-t_o)-\mu}{\sigma\sqrt{2}})) \qquad (3.2)$$

Where D is the amplitude of the impulse, t_0 is the time occurrence of the input command initiating the stroke, μ is the delay of the neuromuscular response expressed in a logarithmic time scale, σ is the response time of the neuromuscular system also expressed in a logarithmic time scale, θ_s and θ_e are the starting and ending directions of the stroke, respectively and *erf* is the error function. The set of parameters $\{D, t_0, \mu, \sigma, \theta_s, \theta_e\}$ fully describes a single stroke.

Any handwriting involving a number of strokes could be described by the vectorial summation of the delayed sequence of the individual strokes.

$$\vec{v}(t) = \sum_{i=1}^{N} \vec{v}_i(t) \qquad (3.3)$$

Where $\vec{v}_i(t)$ is the velocity profile of the i-th stroke in the sequence and N is the length of the sequence.

3.2 HANDWRITING AS GRAPHEMES ON A WRITING SURFACE

The outcome of the handwriting process, the resulting graphemes on a writing surface, can be measured and analyzed. This analysis has proven useful in different areas. In the forensic field, questioned document examination (QDE) aims at answering questions about documents in dispute in a court of law [39]. These questions are mainly related to authorship. In the medical field, the study of handwriting has proven to be an aid to diagnose and track some diseases of the nervous system. For instance, handwriting skill degradation and Alzheimer's disease appear to be significantly correlated [40] and some handwriting aspects can be good indicators for its diagnosis [41] or help differentiate between mild Alzheimer's disease and mild cognitive impairment [42]. Also, the analysis of handwriting has proven useful to assess the effects of substances such alcohol [43] [44], marijuana [45] or caffeine [46]. Aided by modern acquisition devices, the field of psychology has also benefitted from the analysis of handwriting. For instance, in [47], Rosenblum et al. link the proficiency of the writers to the length of the in-air trajectories of their handwritings. In a more controversial field, graphology aims at drawing conclusions about different psychological traits of the writer based on traits of their handwriting [48] [49].

In the field of security-related biometrics, the term handwriting is very often used to designate the production of signatures, not to designate the production of text. As signature has been used for centuries as an authentication method, practical approaches have favoured signature over any other type of handwriting.

3.2.1 DISCRIMINATIVE POWER OF (SHORT SEQUENCES OF) ONLINE TEXT

One of the goals of this dissertation is to explore the extent to which handwriting other than signature, and more precisely words or short sequences of text, can be used to recognize (identify and verify) writers. Do words and short sequences have enough *discriminative power*? In the context of this dissertation, discriminative power refers to the ability to distinguish, in a set of writers, a particular writer from the rest, given a sample of their handwriting. Thus defined, discriminative power arises from *writer individuality*. Writer individuality refers to the hypothesis that each individual has consistent handwriting which is distinct from the handwriting of other individuals. In the past, several authors have scrutinized this hypothesis in the particular case of short sequences of text (isolated characters, isolated words and sentences comprising a small number of words) and reached the conclusion that handwriting is an individual trait, both in the offline case (e.g. [50]) and in the online one (e.g.[51]).

In the forthcoming chapters, it will be shown that isolated handwritten words can indeed be used to recognize writers. It will be shown that a single isolated word exhibits a remarkable discriminative power, probably not far from that of signatures. The fact that a single word can effectively be used to discriminate between writers may lead to a somehow deeper question: signature and handwritten text other than signature are one or two different biometric modalities? Although it is not the purpose of this dissertation to give an answer to this issue, it is worth noting that there seems to be evidence pointing towards the second answer: they may well be regarded as different biometric modalities. In [18] Boulétreau, Vincent, Sabourin and Emptoz concluded that signature and text handwriting may be not one but two *personality identifiers*. In that paper, they claim that the signing and handwriting behaviours of a writer are independent. Therefore, signature and handwriting can be sources of complementary information. Boulétreau et al. reached this conclusion after having compared 20 signatures and 20 samples of text donated by 48 writers. This comparison showed that the intra-class (within-writer) and the inter-class (between-writer) variances for the signature and the handwriting were substantially different. Also, they found that the linear correlations between the values of the features measured for the signature and for the handwriting were quite low ($r<0.35$). Nevertheless, it must be said that whether this results can be generalized to different sets of features or not is an issue not contemplated in the paper.

3.2.2 TEXT VS. SIGNATURE

Even if sequences of text and signatures were different biometric modalities, would there be any advantage in using the former instead of the latter? The answer to this question is positive since sequences of text exhibit some remarkable properties that signatures lack:

- A compromised word, like a compromised password, can be easily changed. When a signature is compromised, it is quite unrealistic to think that it can be easily and consistently changed.
- Some signatures are too short or too simple to be considered appropriate for security purposes. Sequences of text, on the other hand, can be made shorter or longer

depending on the intended accuracy. Our experiments show that the accuracy attained when combining two words is greater than their individual accuracies.

When these two properties are taken into account, recognition based on sequences of text somehow resembles speech-based recognition: text-length and text-contents can be easily changed.

Although signature is a biometric modality quite well accepted, especially for its non-intrusiveness, it still arises some concerns regarding privacy and the possibility of being *stolen* [52]. As a biometric identifier, words will probably arise fewer concerns. Thus, in some environments its acceptability could be greater than that of signatures. Nevertheless, it is not in the scope of this dissertation to provide scientific proof of this intuition.

There is another difference between sequences of text and signature that is worth being mentioned: the degree of *accepted* inter-writer variability. While the inter-writer variability of text is heavily constrained by legibility, signature puts no constraints to dissimilarities among signers. Although by definition a signature is a written name[3] (hence a piece of text), some signatures are so stylized and embellished that they cannot be effectively read. A considerable percentage of signatures are closer to a drawing than to a text and signatures of different signers tend to be quite different from each other. On the other hand, different writers' executions of the same word tend to be much more similar (Fig. 3.4). They have to be similar since otherwise they would not be considered the same word. Hence, user verification based on text could be combined with text recognition, while this combination, merging a biometric schema with a traditional password-like approach, does not seem plausible in the case of signature-based verification. Again, text-based recognition is closer to speech-based recognition than to signature-based recognition.

Figure 3.4: Signature and word INFATIGABLE by two different writers. Notice that dissimilarity between signatures is much higher than between words.

[3] Signature: your name written by yourself, always in the same way, usually to show that something has been written or agreed by you. Definition taken from [172].

3.3 ONLINE HANDWRITING

Traditionally, the analysis of handwriting has been divided into two categories, depending on the available information. When only the writing itself (strokes on a paper) is available then the analysis is said to be performed in an ***offline*** manner. In this category, a great deal of the temporal dimension of handwriting is lost (e.g. in what order have the strokes been produced?). When spatiotemporal information is available and taken into account, its analysis is referred as **online**. Typical modern digitizing devices, such as the one seen in Fig. 3.5, can gather the x-y coordinates that describe the movement of the writing device as it changes its position. Furthermore, they can gather other time-dependent data, mainly the pressure exerted by the writing device on the writing surface and also two writing angles: azimuth, the angle of the pen in the horizontal plane, and altitude, the angle of the pen with respect to the vertical axis (Figs. 3.6 and 3.7).

Figure 3.5: A modern digitizing tablet and pen by WACOM® (image from [53]).

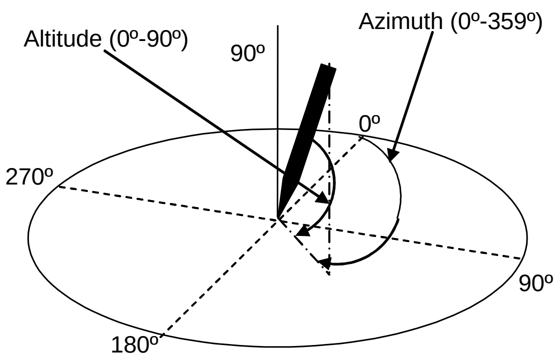


Figure 3.6: Azimuth and altitude angles of the writing device with respect to the plane of the writing surface.

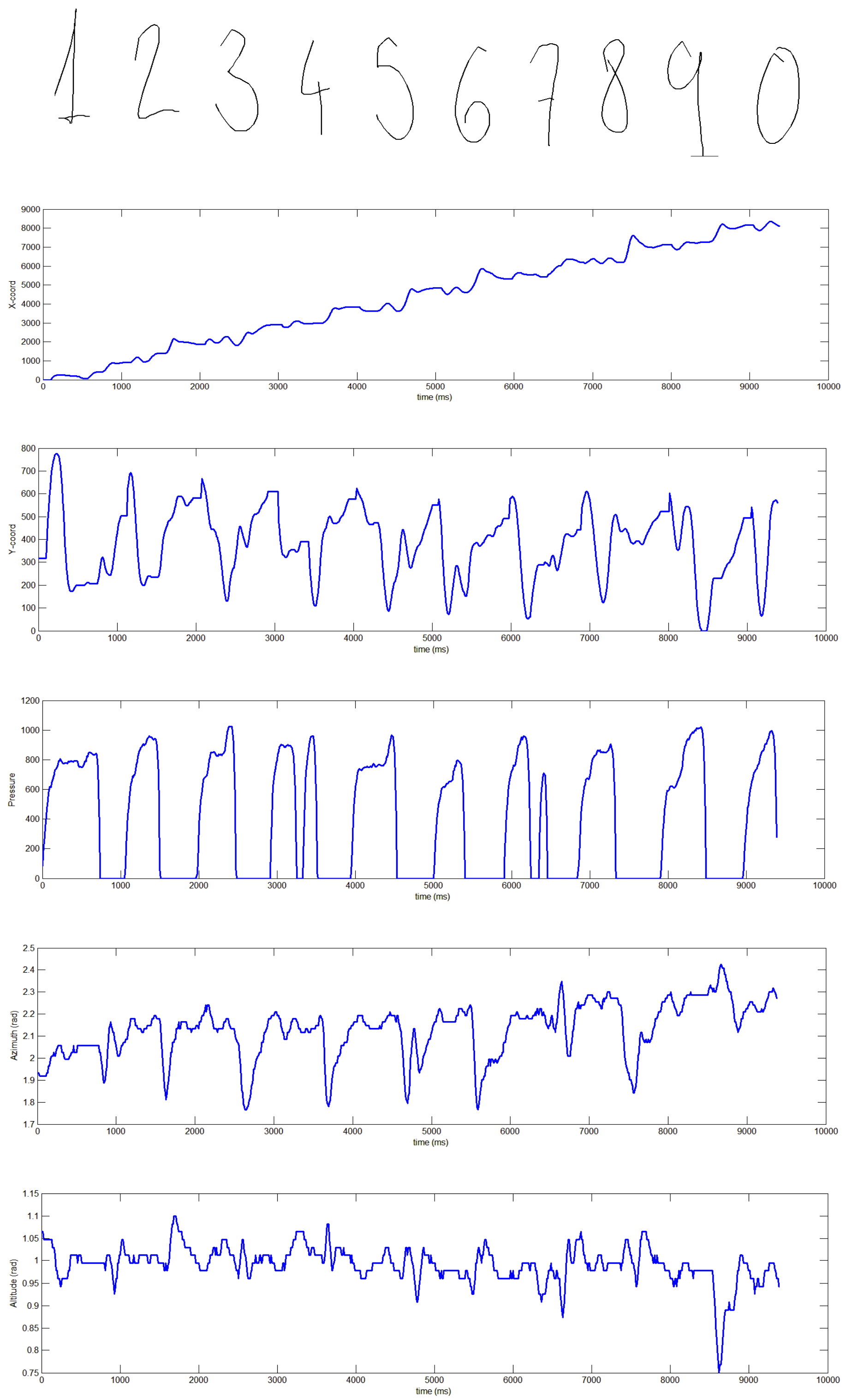


Figure 3.7: A sample of handwriting and five time-dependent signals recorded by the digitizing tablet (from top to bottom: the sample, X-coordinate, Y-coordinate, pressure, azimuth and altitude).

A very interesting aspect of the modern online analysis of handwriting is that it can take into account information gathered when the writing device is not exerting pressure on the writing surface. Thus, the movements performed by the hand while writing a text can be split into two classes: **on surface-trajectories**, the visible part of the writing; and **in-air trajectories**, the invisible part. In chapter 5, both types of trajectories will be analyzed, and in chapter 6 a recognition schema that gives in-air and on-surface trajectories the same importance is presented.

3.4 HANDWRITING DATABASES

Databases that provide samples of handwriting from a number of different writers are of paramount importance for the research community. These databases provide the necessary data to train and test new methods and algorithms and to perform benchmarks of the proposed solutions. *Standard* databases allow that different research teams share a common ground, making their measures if not fully comparable at least more comparable than they would be without such a common ground. The possibility to compare results is indispensable to measure the progress in any scientific field. Some publicly available databases exist that contain samples that can be used in the field of text-based writer recognition.

The first part of the rest of this section will provide a very brief description of some databases that are relevant in the field of writer recognition, in both the offline and the online approach. We have chosen to comment mainly on the databases containing western script that are mentioned in the chapter on the state of the art. The reader may notice that several of these databases were not initially intended to be used in writer recognition research but in handwriting recognition, although they have later on been successfully applied to writer recognition.

The last part of this section is entirely devoted to the BiosecurID database since all our experiments are based on data contained in it.

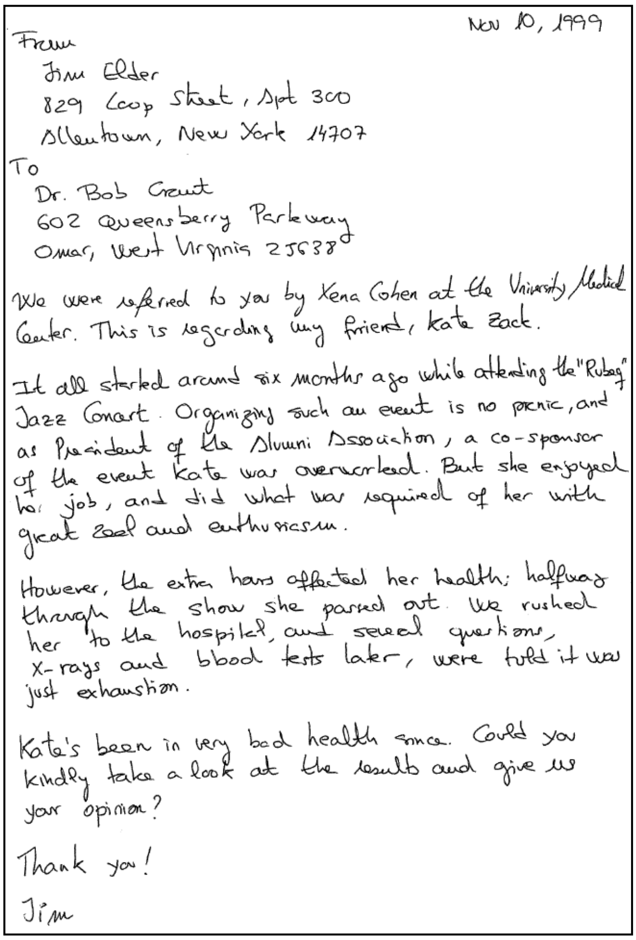
3.4.1 RELEVANT DATABASES

IAM. This offline database was first described in 1999 [54] and its contents have grown since then. A much more detailed description was published in 2002 [55]. In its newest version (the IAM Handwriting Database 3.0 [56]) it contains 13353 labelled lines of handwritten text extracted from 1539 scanned pages produced by 657 writers. It also contains 5685 labelled sentences and 115320 isolated and labelled words. The original purpose of the database was to be used in handwriting recognition, actually to train and test word recognizers, but it has also been extensively used to perform writer recognition experiments.

CEDAR letter dataset. This is a huge offline database containing 4701 documents produced by 1567 writers selected to be representative of the US population [50]. Each writer was asked to copy the *CEDAR*[4] *letter* three times. This letter is a document containing 156 words, from a lexicon of 124, which includes all characters (letters and numerals). The document was carefully designed so that it contained each letter of the alphabet in uppercase at the initial position of a word and in lowercase in the initial, middle and terminal positions of a word (see Fig. 3.8). Each document was digitized using a desktop scanner and images of paragraphs, lines

[4] Center of Excellence for Document analysis and Recognition (CEDAR) at the University of Buffalo (state of New York, USA).

and isolated words (***Cohen*** and ***referred***) were manually extracted from each document. The eight characters of the word *referred* were also segmented.


Nov 10, 1999

From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubey" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, X-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!

Jim

Figure 3.8: The CEDAR letter (Manuscript is by Enric Sesa).

FIREMAKER. This offline database [57] contains images of handwriting donated by 250 Dutch subjects, mostly students. Each donor was requested to write 4 different A4 pages. On page 1 they were asked to copy a given text, on page 2 they had to write, in their own words, the description of a given cartoon. On pages 3 and 4, the subjects were asked to write uppercase forged-style samples. The sheets were scanned and no further segmentation was applied.

UNIPEN. This online database is the result of the collaborative effort of several research institutes and industrial firms that generated the UNIPEN standard and database [58][59]. Since 1999, the International UNIPEN Foundation (iUF) hosts the data. Only a portion of the database is publicly available: the *Train-R01/V07* distribution. The whole database contains, among others, thousands of isolated digits, isolated lower and uppercase letters, lower and uppercase words, and fragments of texts comprising two or more words.

RIMES. This is another offline database that contains 12723 pages of fully labelled text [60] [61] This huge amount of text was produced by 1300 writers who were asked to write 5 different letters each (in French). From these pages, 100.000 isolated characters, 250.000 isolated words and 6500 blocks of words were obtained.

PSI. This is a non-publicly available database collected at the *Laboratoire PSI* (University of Rouen, France) [62]. It is built out of handwritten letters (in French) donated by 88 writers. Each writer was asked to choose from two different letters containing 107 and 98 words respectively.

IRONOFF. The IRONOFF (IRESTE ON OFF) is a public dual (both online and offline) handwriting database, collected at the *École Polytechnique de l'Université de Nantes* (France), formerly known as IRESTE [63] [64]. For the same writing it gives access to the digital image (offline) and to the pen trajectories (online). Its original purpose was handwriting recognition. Its contents include 4086 isolated digits, 10685 isolated lower case letters, 10679 isolated upper case letters and 31346 isolated words (28657 French, 2689 English) from a 197 word lexicon.

All the aforementioned databases contain western script. Other databases exist that contain non-western script (e.g. the MADCAT offline database [65] contains Arabic script, the HCL2000 offline database [66] contains Chinese script, the PE92 offline database [67] contains Korean scripts and the ISI [68] offline database contains three different Indian scripts).

3.4.2 THE BiosecurID DATABASE

BiosecurID [1] is a multimodal biometric database acquired in the framework of the BiosecurID project, a project conducted by a consortium of six Spanish universities (Universidad Autónoma de Madrid, Universidad Politécnica de Madrid, Universitat Politècnica de Catalunya, Universidad de Zaragoza, Universidad de Valladolid and Euskal Herriko Unibertsitatea). The main goal of the project, as stated in [1], was:

> *the acquisition of a realistic multimodal and multisession database, statistically representative of the potential users of biometric applications, and large enough in order to infer valid results from its usage* [Sic].

Eight different biometric modalities (traits) are represented in the database, namely **speech**, **iris**, **face** (still images and videos of talking faces), **fingerprints**, **hand**, **keystrokes**, **handwritten signature** and **handwritten text** (online dynamic signals and offline scanned images). All data was collected in 4 sessions distributed in a time span of 4 months. The number of donors was 400 with a balanced gender distribution. Table 3.1 summarizes the most relevant statistics regarding the participating donors.

GENDER DISTRIBUTION		*AGE DISTRIBUTION*				*HANDEDNESS*	
MALE	*FEMALE*	*FROM 18 TO 25*	*FROM 25 TO 35*	*FROM 35 TO 45*	*FROM 45 ONWARDS*	*RIGHT*	*LEFT*
54%	*46%*	*42%*	*22%*	*16%*	*20%*	*93%*	*7%*

Table 3.1: Relevant statistics regarding the donors in the BiosecurID database.

With respect to online handwritten text, BiosecurID provides data gathered from 3 different tasks: a Spanish text handwritten in lowercase, the sequence of digits (from 1 to 9 and 0 in the last position) written in a single line, and 16 Spanish words, in uppercase, written each in a line

(Fig. 3.9).The writers were asked not to perform neither corrections nor crossing outs. The acquisition of these data was carried out with a WACOM® INTOUS3 A4 pen tablet, capturing seven dynamic features at 100Hz (x-coordinate, y-coordinate, button status (up or down), azimuth, altitude and pressure). Data was saved in the SVC format [69]. Table 3.2 gives some relevant technical details.

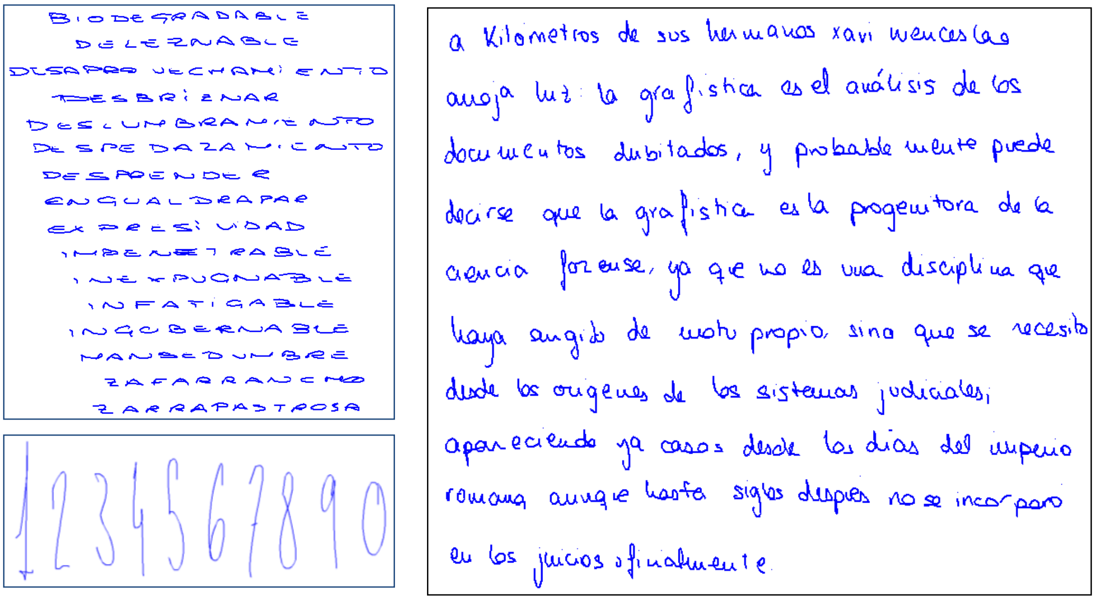


Figure 3.9: Samples of the 3 tasks in the BiosecurID database involving handwritten text.

FEATURE	VALUE
Active area (WxD)	304.8 x 228.6 mm (12.0 x 9.0 in)
Coordinate resolution	200 lines per mm (5080 lpi)
Accuracy	±0.25 mm (±0.010 in)
Maximum sampling rate	200 Hz
Maximum. reading height	6.0 mm (0.25 in)
X-coord range of values	[0-30480]
Y-coord range of values	[0-22860]
Pressure range of values	[0-1023]
Azimuth range of values	[0-3600]
Altitude range of values	[0-900]

Table 3.2: Relevant technical details regarding the device used to acquire the handwritten data in the BiosecurID database.

4

HIGHLIGHTS OF THE STATE OF THE ART IN WRITER RECOGNITION

In this chapter the state of the art in writer recognition is reviewed. Two recognition modalities are considered: text-based modalities, reviewed in the first section, and signature-based modalities, reviewed in the second. The first section starts with the review of the works on the individuality of handwriting published by professor Sargur N. Srihari and his colleagues. Up to date, these works are the most comprehensive effort to assess the individuality of handwriting and its suitability for recognition purposes. After that, relevant references pertaining to the offline, first, and the online field, later, are surveyed. A table, comprehensively summarizing all the references in text-based writer recognition previously considered, concludes the first section. The second and last section is entirely devoted to review some relevant references from the signature-based recognition field, emphasizing the results achieved in several online signature-verification competitions.

4.1 TEXT-BASED WRITER RECOGNITION

Text-based writer recognition has mainly addressed issues related to forensic applications where the main purpose is to match a document to a set of samples belonging to an individual (e.g. a suspect). Historical document analysis (computerized palaeographic analysis of ancient manuscripts) has also received some attention (see, for instance, [70], devoted to writer identification of ancient Hebrew documents, [71] devoted to the identification of the contributors of dried plants in the Botanic Museum of Berlin from the handwriting on herbarium sheets, or [72], devoted to the recognition of the author in the context of music scores).

Different criteria can be taken into account to classify the different approaches that have been presented. Regarding the kind of available information, the recognition can be performed in an ***offline*** manner or in an ***online*** manner (see chapter 3). Although an accurate analysis of offline images may reveal some relevant information on pressure and writing speed [73], it is commonly acknowledged that the online approach is bound to produce more accurate results due to the increased amount of data available.

When attention is focused in the text used to perform the recognition then the approaches can be ***text-dependent*** or ***text-independent***. Text dependent approaches assume that the text used to train the recognition system is exactly the same that will be presented to the system when it is required to ascertain a writer's identity. Text-independent approaches do not make such assumption since their goal is to identify the authorship of a given document based on other documents written by the same person and, probably, having different contents.

Text-independency tends to require considerable amounts of text. In [74], Brink et al., experimenting with the IAM and the Firemaker databases, concluded that when well-performing features are considered, the minimum number of characters required is about 100, both in models –samples- and questioned documents –queries-. This figure may increase, up to 200 characters, for less well-performing features. Any text above these minima does not significantly increase accuracy. The authors also concluded that for a fixed number of characters in the questioned documents, the recognition accuracy tends to increase when the number of characters in the samples grows.

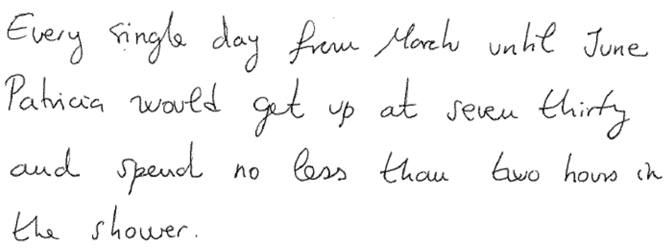


Figure 4.1: A fragment of text containing exactly 100 characters.

And finally, when the focus of attention lies in the qualities of the writing itself, then two categories can be considered: ***structural*** approaches and ***allographic*** approaches. The former approaches focus on structural (sometimes also referred as textural) features of the image of the writing, such as slant and curvature. According to [15], these particular features indirectly measure *pen-grip* and *pen-attitude* preferences. The latter approaches, the allographic, focus on the shape of the characters or on the shapes of fragments of characters. From this point on, we are going to use the term *structural* in its broader extent, that is, to refer to global (i.e. statistical) features extracted from whole blocks of text.

4.1.1 INDIVIDUALITY OF HANDWRITING: THE WORKS OF SRIHARI ET AL.

Most of the works dealing with writer recognition implicitly or explicitly assume that handwriting do bear enough individuality as to allow performing recognition, provided this individuality can be conveniently *extracted* from the handwriting samples. In order to tackle the issue of weather this assumption is correct or not, motivated by several rulings in the US courts concerning expert testimony in the field of handwriting, Srihari, Cha, Arora and Lee undertook a study aiming at the validation of the hypothesis that handwriting is individual [50]. All their works were based on the *CEDAR letter* dataset (see section 3.4.1) .

Two sets of features were considered: *macro-features*, pertaining globally to each handwritten sample, and *micro-features*, which were locally extracted. The following eleven macro features were considered: *entropy of grey values*, *grey-level threshold*, *number of black pixels*, *number of interior and exterior contours*, *number of vertical, horizontal, negative and positive slope components*, *slant* and *height*. *Gradient, Structural and Concavity* (GSC) *attributes* were used as micro-features. Macro-features were extracted at different levels: from the whole handwritten document (document-level), from a particular paragraph (paragraph level), from a particular-line (line-level), from a particular word (word-level) and from a particular character (character-level). Micro-features were only obtained at the character level.

Identification was performed by finding the closest match: each questioned sample was assigned to the writer that produced the prototype yielding the smallest distance to the sample (nearest neighbour rule). Accuracy depended on the level, the features used and the number of writers considered.

When only macro-features were considered, the identification accuracies shown in Table 4.1 were obtained. The word level corresponds to the words ***Cohen*** and ***referred***; the paragraph level corresponds to the address block (comprising 11 words); the document level corresponds to the entire document. As it could be expected, obtained accuracies were higher at the paragraph and document levels.

NUMBER OF WRITERS	CORRECTLY IDENTIFIED WRITERS		
	WORD LEVEL	PARAGRAPH LEVEL	DOCUMENT LEVEL
2	88%	97%	99%
10	~ 62%	~ 81%	~ 96%
100	~ 28%	~ 49%	~ 81%
900	~ 9%	~ 36%	~ 59%

Table 4.1: Identification accuracies (IDRs in percentage) reported in [50] for macro-features.

Identification accuracy related to micro-features was tested considering the following characters: *r*, *e*, *f*, *e*, *r*, *r*, *e*, *d*, *b* and *h*. In this case, the combination of micro and macro-features was also tested. The reported results are shown in Table 4.2.

NUMBER OF WRITERS	CORRECTLY IDENTIFIED WRITERS	
	MICRO-FEATURES ONLY	MICRO AND MACRO-FEATURES
2	99%	99%
10	~ 98%	~ 98%
100	~ 90%	~ 95%
900	~ 82%	~ 88%

Table 4.2: Identification accuracies (IDRs in percentage) reported in [50] for the set of characters {*r, e, f, e, r, r, e, d, b, h*}.

To address the issue of verification, Srihari et al. used three-layered artificial neural networks (ANN). Regarding macro-features, each AAN (one for the word level, one for the paragraph level and another one for the document level) was trained with data (pairs of distances) coming from 250 writes. Later on, each ANN was tested with data coming from two disjoint sets of 250 users not previously used. At the paragraph level data was obtained from the address block, while at the word level data was obtained from the word *referred*. With regard to micro-features, the ANN was trained with data obtained from the set of characters {*r, e, f, e, r, r, e, d, b, h*}, acquired from 964 writers. Table 4.3 summarizes the best reported results for both types of features.

MACRO-FEATURES			MICRO-FEATURES
WORD LEVEL	PARAGRAPH LEVEL	DOCUMENT LEVEL	{ *r, e, f, e, r, r, e, d, b, h* }
83.1% (FRR=14.5%; FAR=19.3%)	89.1% (FRR=14.2%; FAR=7.6%)	95.95% (FRR=4.5; FAR=3.6%)	91.75% (FRR=10.0; FAR=6.5%)

Table 4.3: Highest verification accuracies (VAs) reported in [50] for macro and micro features.

Eventually, Srihari and his colleagues concluded that the hypothesis of handwriting individuality could be accepted with a 95% confidence. They also pointed out that finer features could have allowed reaching the same conclusion with a higher, near 100%, confidence.

In [75], Zhang et al. extended the research in [50] analyzing the discriminative powers of individual characters. Actually the characters ['*0*'-'*9*'], ['*a*'-'*z*'] and ['*A*'-'*Z*'] were scrutinized. Experiments regarding identification and verification were performed, using, in both cases, a simple nearest neighbour classifier. The same micro-features reported in [50] were considered. The testing sets comprised documents from 500 different writers. The identification results showed very different performances depending on the character. Thus, the character '1', the worst performing, yielded a 1.6% correct identification ratio, the character '*0*', the second worst performing, a 4.88%; while the characters '*G*' and '*M*', the first and the second best performing, yielded identification ratios about 35%. Regarding verification, similar results were observed. Character '1' had an accuracy of 61.79%[5] (the lowest) while characters '*I*' and '*G*' had accuracies about 79% (the highest). Other experiments also reported in the same paper, showed that combinations of characters outperformed single characters and that longer combinations outperformed shorter ones. In order to assess the discriminability measure of individuality of each character in a classifier independent way, Zhang et al. proposed a measuring method based on the receiver operating characteristic (ROC) curve. Applying this method, the 62 characters considered in the aforesaid experiments were ranked according to their discriminative power. Table 4.4 shows the ranking.

[5] Notice that a verification accuracy (VA) of 61.79% accounts for a FAR+FRR=76.42%.

RANK	CHARACTER	RANK	CHARACTER	RANK	CHARACTER	RANK	CHARACTER
1	***G***	17	***n***	33	***4***	49	***e***
2	***b***	18	***V***	34	***5***	50	***p***
3	***N***	19	***A***	35	***t***	51	***1***
4	***I***	20	***w***	36	***k***	52	***T***
5	***K***	21	***L***	37	***Q***	53	***x***
6	***J***	22	***v***	38	***2***	54	***7***
7	***W***	23	***y***	39	***z***	55	***3***
8	***D***	24	***S***	40	***g***	56	***i***
9	***h***	25	***E***	41	***o***	57	***c***
10	***F***	26	***R***	42	***6***	58	***O***
11	***r***	27	***s***	43	***q***	59	***C***
12	***H***	28	***f***	44	***9***	60	***X***
13	***B***	29	***U***	45	***a***	61	***0***
14	***M***	30	***Z***	46	***P***	62	***1***
15	***m***	31	***u***	47	***8***		
16	***d***	32	***Y***	48	***J***		

Table 4.4: Handwritten characters ranked in descending order of discriminative power, as reported in [75]. Digits appear in red.

The research in [50] was also extended in [76]. The particular purpose of this paper was to assess the discriminative power of digits. Using the micro-features already described in [50], the authors found identification rates ranging from 1.8% (digit 1) to 13.8% (digit 5) and verification accuracies ranging from 70.6% (digit 1) to 79.6% (digits 2 and 4). When it comes to the discriminative power of each digit, the authors used the Bhattacharya distance to measure the separation between the probability density function (pdf) of the intra-user distances and the probability density function of the inter-user distances. According to this criterion, the digits were ranked as shown in Table 4.5, which also shows the achieved identification rates and verification accuracies.

RANK	1	2	3	4	5	6	7	8	9	10
DIGIT	***2***	***5***	***3***	***8***	***6***	***1***	***4***	***7***	***9***	***0***
IDENTIFICATION (IDR)	12%	13.8%	11.4%	10.9%	10.3%	1.8%	12.9%	11.2%	9.9%	6.8%
VERIFICATION (VA)	79.6%	79.8%	77.5%	76.8%	78.4%	70.6%	79.6%	76.7%	78.9%	74.8%

Table 4.5: Handwritten digits ranked in descending order of discriminative power, as reported in [76].

The reader may notice that digits are ranked differently in [75] and in [76]. This might be due to the fact that in [75] the ranking criterion is based on the ROC curve, while in [76] it is based on the Bhattacharya distance.

As in [75], the effect of the combination on the recognition accuracies was also explored. From this exploration it was concluded that, although longer combinations tend to perform more accurately, the addition of some digits may have a negative impact on the verification accuracy.

There exist other studies that address the issue of the discriminative power of digits. In [77], Leedham and Chachra used a different set of features to assess the identification and verification performance of isolated digits and even considered the case where forged digits were present. Unfortunately, the small number of writers, only 15, prevents a fair comparison with [76] (1000 writers).

Another study that complements the research in [50] was carried out by Zhang and Srihari and reported in [78]. The purpose of this new study was that of becoming a first step towards establishing objective measurements of individuality of handwritten words. The following four words were considered: ***been***, ***Cohen***, ***Medical*** and ***referred***. The words *Cohen* and *referred* had already been used in [50]. Experimentation was carried out using GSC micro-features as described in [50], represented with a 1024-bit vector. Identification relied in a nearest neighbour classifier while verification relied in a 6-nearest neighbour classifier. The testing phase comprised 875 writers. Table 4.6 summarizes the best reported values for identification and verification.

WORD	*been*	*Cohen*	*Medical*	*referred*	COMBINATION OF THE FOUR WORDS
IDENTIFICATION (IDR)	~ 46%	~ 44%	~ 47%	~ 49%	~ 83%
VERIFICATION (VA)	~ 41%	~ 43%	~ 47%	~ 49%	90.94%

Table 4.6: Best recognition performances for the words *been*, *Cohen*, *Medical*, *referred* and their combination, as reported in [78].

In 2004, yet another study on the discriminative power of words was published by Tomasi, Zhang and Srihari [79]. This time, the first 25 words from the CEDAR letter were considered. Four different feature sets were used, three of which relied on the segmentation of words into characters. Words were tested individually and, in a second phase, the ten best performing were combined. Authors concluded that segmentation did not seem to make any substantial difference, that longer words tended to perform better and that words containing highly discriminative characters (e.g. G) performed well. Table 4.7 summarizes the best reported results for single words.

WORD	IDENTIFICATION (IDR)	VERIFICATION (VA)	WORD	IDENTIFICATION (IDR)	VERIFICATION (VA)
From	64%	34%	*York*		
Nov			*14707*		
10			*To*		
1999			*Dr*		
Jim			*Bob*		33%
Elder	65%	35%	*Grant*	67%	43%
829			*602*		
Loop			*Queensberry*	66%	34%
Street			*Parkway*	66%	35%
Apt			*Omar*		
300			*West*		33%
Allentown	66%	29%	*Virginia*		58%
New					

Table 4.7: Words considered in [79] and their best recognition performances, as reported by the authors. Overall highest performances appear in red. Blank cells correspond to non-reported results.

4.1.2 OFFLINE APPROACHES

4.1.2.1 Text-independent structural approaches

An early attempt to perform writer identification was made in 1977 by B. Arazi [80]. Arazi proposed to perform writer identification by means of histograms of vertical and horizontal run-lengths of background intensity values [80]. Accuracy reached 100% although experiments were carried out with 13 writers only.

In 2000 Said, Tan and Baker, building upon a previously published work [81], presented a paper that has become a classic in the field of writer recognition, being one of the most cited [82]. Each individual's handwriting was regarded as a potentially different texture so that texture recognition algorithms could be applied. More precisely, they used a multi-channel Gabor filtering technique and grey-scale co-occurrence matrixes to extract features from the images of the text. Experimenting with a small set of 20 writers and 25 blocks of text from each one, the identification accuracy reaches 96% when the Gabor-related features are considered. Gabor filters have also been applied to writer recognition of Chinese texts ([83]). The textural approach of Said et al. based on the classification of the texture has inspired many other researchers who have proposed alternative ways to represent and recognize textures (e.g. [84], for Chinese handwriting and [85] for Arabic handwriting).

One year later, in 2001, Marti, Messerli and Bunke, proposed a recognition system [86] based on twelve line-level features that mainly correspond to *visible* characteristics of the writing (Table 4.8).

LINE-LEVEL FEATURE	
f1	Height of the upper zone of the line (see Fig. 4.2)
f2	Height of the middle zone of the line
f3	Height of the lower zone of the line
f4	Ratio of height of upper zone to height of middle zone (f1/f2)
f5	Ratio of height of upper zone to height of lower zone (f1/f3)
f6	Ratio of height of middle zone to height of lower zone (f2/f3)
f7	Writing width (The median of the lengths of the runs of white pixels between two black runs in the row with more black to white and white to black transitions)
f8	Ratio of height of the middle zone to writing width (f2/f7)
f9	Mean value of the slant angle
f10	Standard deviation of the slant angle
f11	Fractal dimension of writing [6]
f12	Second dimension of writing

Table 4.8: Line-level features used in [86] .

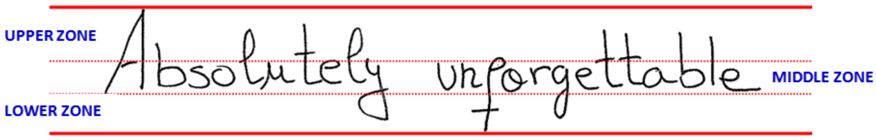


Figure 4.2: An example of the three zones of a handwritten line considered in [86]. (Representation is approximate).

An image of text is described as a sequence of twelve-dimensional vectors, each vector corresponding to a line. Two different classifiers were considered: a k-nearest neighbour classifier and a feed-forward neural network. In both cases, experimentation was carried out using samples from 20 different writers. Different experiments were performed considering different subsets of the twelve features previously described, mainly to overcome potential problems deriving from the redundancy between features f1, f2, f3 and f7 and the ratios computed from them (f4, f5, f6 and f8). Eventually, the best identification performance was achieved by the neural network when all twelve features were considered simultaneously. In that case, identification rate reached 90.7%.

The set of twelve features proposed by Marti et al. was enlarged by Hertel and Bunke [87]. They considered features (a) extracted from the connected components found in the handwriting images (average distance from consecutive bounding boxes, average distance between components of the same word, average distance between words, etc) (b) extracted from enclosed regions –blobs enclosed from loops- (e.g. average roundness of the blobs) (c) features extracted from the upper and lower contour (slant, frequency of local maxima and of local minima etc.) and (d) fractal features similar to those presented in [88] and used in [86]. The combination of this huge set of features yielded a 90.7% identification rate when tested with 50 writers and a single line of text. With a whole page of text and the same number of writers, the accuracy reached 99.6%.

[6] High values of both f11 and f12 typically correspond to legible handwritings while low values usually correspond to badly formed writings. For an accurate explanation of the meanings of f11 and f12 refer to [88].

4.1.2.2 *Text-independent allographic approaches*

Lambert Schomaker and Marius Bulacu show a preeminent position among the most cited researchers in the field of writer recognition. In a 2004 paper [89], they proposed to use a histogram of connected component contours (CO^3) usage as a feature. According to this proposal, a writer would be modelled as a discrete probability function (PFD) of CO^3 usage. A connected component contour is the shape of a character or of part of a character, obtained by segmentation. A codebook of prototypical CO^3s is computed by means of a self-organizing map (see Fig 4.3). This codebook, obtained from samples coming from writers not included in the testing phase, is used to compute all the required PDFs. Using only this feature, identification rates ranging from 72% to 85% are attained when experimenting with 150 writers producing uppercase handwriting. The authors also consider other non-allographic features based on edge directions and the effect of combining them with the aforementioned allographic feature (95% IDR). This combination of features was enhanced and further explored in [90], a paper published in 2007 that will be reviewed in a forthcoming section. In another paper [91], also published in 2004, the use of CO3s is extended to mixed-style (uppercase and lowercase) handwriting achieving a 97% identification accuracy when experimenting with 150 writers.

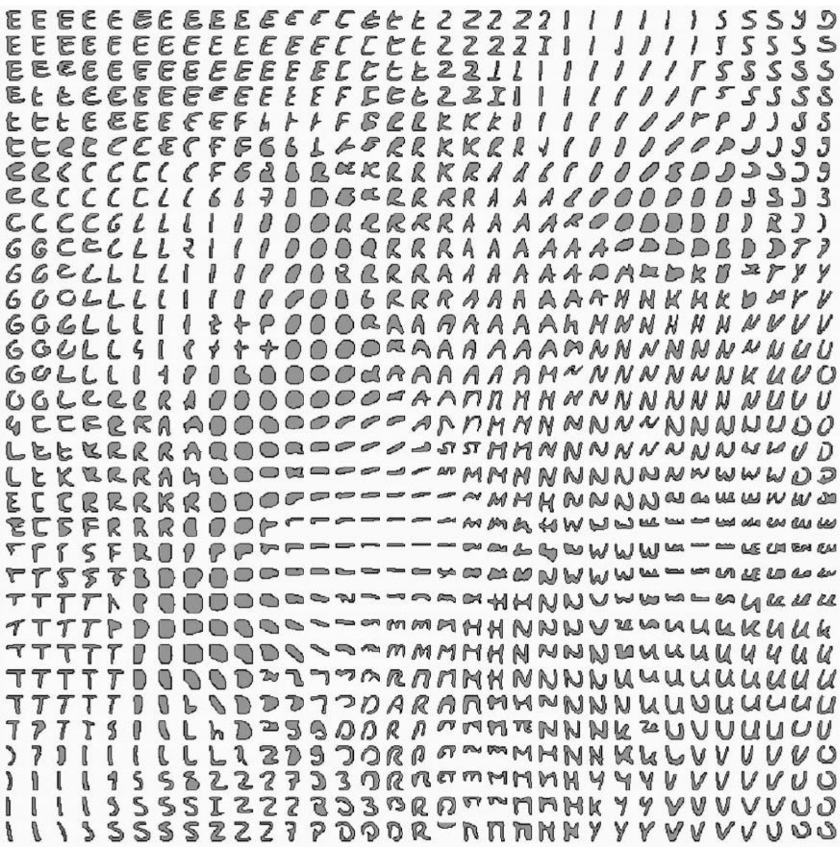


Figure 4.3: A codebook of 1089 CO^3s prototypes derived from two paragraphs of Dutch text written by 100 writers. Clustering was achieved by means of a self-organizing map (Taken from [89]).

Another frequently cited allograph-based recognition system was proposed in 2005 by Bensefia, Paquet and Heutte [92], [93]. Like in the previously described works by Schomaker et al., Bensefia and his colleagues based their proposal on the construction of a codebook of prototypical shapes (graphemes). In this case, graphemes are obtained by a segmentation algorithm based on the analysis of the minima of the upper contour. The codebook of prototypical shapes is obtained by repeatedly applying sequential clustering. Those shapes that are always clustered together are regarded as members of an invariant clusters and each invariant cluster constitutes a prototype (Fig. 4.4).

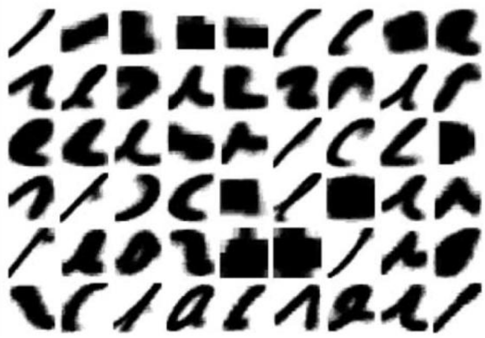


Figure 4.4: A set of invariant clusters (prototypes) obtained from handwriting samples (Taken from [92]).

Each one of the prototypes in the codebook constitutes a binary feature. Documents are characterized using a vector space model (VSM) and recognition is performed following an information retrieval (IR) approach [94]. Each document is described by means of a vector containing, for each feature, a weight calculated as the product of the frequency of that feature in that particular document by the inverse document frequency of that feature (the inverse of the number of documents in the database that contain the feature). Similarity between documents is defined as the normalized inner product of the vectors that define them. Identification is carried out by obtaining a similarity-sorted list of documents matching the questioned one. The reported accuracy ranges from 86% (experimenting with 150 writers) to 95% (experimenting with 88 writers). In order to cope with the possibility that a writer/document under identification does not belong to the reference database, a verification approach based on a mutual information criterion is proposed: two documents are deemed written by the same author if their features show a strong independence on the writer (low mutual information between features and writers). Reported verification accuracy is about 96%.

Jain and Doermann presented, in the 2011 edition of the International Conference on Document Analysis and Recognition (ICDAR), a paper [95] with yet another allographic proposal, based on K-adjacent segments (KAS). A KAS is an ordered sequence of K lines where each pair shares an endpoint (see Fig. 4.5). As in most allographic approaches, a codebook is required as a starting point to represent documents as vectors of prototype frequencies. Some experiments were conducted to determine the best performing values for K (finally set to 3) and the size of the codebook (finally set to 300). The approach was tested against two databases: the IAM database (650 writers, English text) and the MADCAT database (325 writers, Arabic script). The following identification rates were reported: 99.8% with 127 writers from the IAM database, 3 training samples and an IAM-based codebook; about 90% with 302 writers from the MADSET database, 7 pages as training samples and a MADSET-based codebook, and 92.1% with the 650 writers from the IAM database and a MADSET-based codebook. This last result suggests that KAS codebooks may be, to some extent, independent of the language.

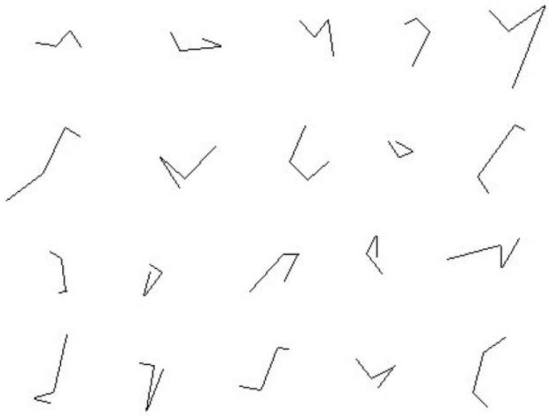


Figure 4.5: Some prototypical K-adjacent segments (K=3) extracted from a codebook (Taken from [95]).

4.1.2.3 Text-independent mixed approaches

Mixed approaches are the ones that consider features of both types and combine them in order to obtain an enhanced performance.

In [90] Bulacu and Schomaker, presented an approach that combines structural and allographic features (see Table 4.9). As structural features they consider the probability density functions (PDFs) of several angles, the PDFs of the horizontal and vertical run-lengths on background and the autocorrelation in the horizontal raster. A single allographic feature is considered: the PDF of the graphemes found in the writing. Here a grapheme is an allographic fragment (a character, part of a character or part of more than one character) segmented from a line of text (see fig 4.7). In order to compute this PDF, a codebook of 400 graphemes generated by k-means clustering is used. The authors perform several sets of experiments. In one of them they establish the recognition performance of each feature. Feature f2, the hinge of the contour, turns out to be the most discriminative one (80% identification rate and 4.8% EER experimenting with 900 writers); f4, the allographic feature, is the second best performing (76% identification rate and 5.8% EER). In another set of experiments, the performances of different combinations of features are considered. The combination of f2 and f4 is one of the best-performing combinations (86% identification rate and 2.9% EER with 900 writers). The best performance, with 900 writers is achieved when features f2, f4, f5h and f5v are combined (87% identification rate and 2.6% EER). **Up to date, the results reported in this paper are considered the best in the field of offline text-independent writer recognition**.

	FEATURE	TYPE
f1	One-dimensional PDF of the contour direction –slant- (angle ϕ in Fig 4.6-a)	Structural
f2	Two-dimensional PDF of the hinge of the contour –curvature- (angles ϕ1 and ϕ2 in Fig 4.6-b)	Structural
f3h	Two-dimensional PDF of the horizontal co-occurrences at both ends of a background run (angles ϕ1 and ϕ3 in Fig 4.6-c)	Structural
f3v	Two-dimensional PDF of the vertical co-occurrences at both ends of a background run (angles ϕ1 and ϕ3 in Fig 4.6-d)	Structural
f4	PDF of the graphemes found in the writing.	Allographic
f5h	One-dimensional PDF of the horizontal run-lengths on background	Structural
f5v	One-dimensional PDF of the vertical run-lengths on background	Structural
f6	The autocorrelation in the horizontal raster.	Structural

Table 4.9: Set of structural and allographic features considered in [90]. Names of features are those given by the authors.

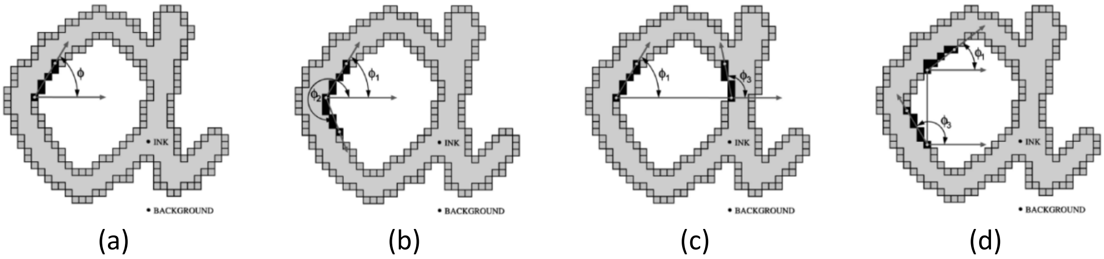


Figure 4.6: Examples of contour angles. Slant (a), hinge (b), horizontal co-occurrences (c) and vertical co-occurrences (d). Images taken from [90].

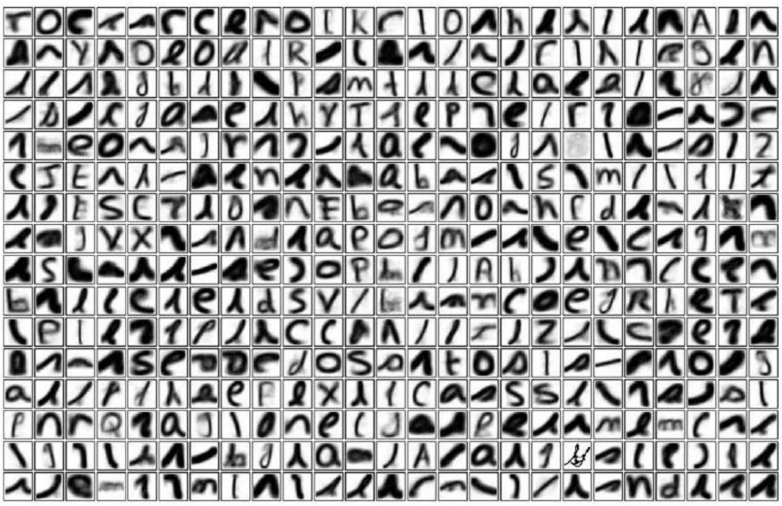


Figure 4.7: A codebook of 400 graphemes obtained by K-means clustering. Taken from [90].

The mixed approach proposed by Bulacu and Schomaker in [90] was also successfully applied to Arabic handwriting [96] (IDR=88% and EER=6% with 350 writers) thus suggesting its script-independent nature.

In a paper published in 2010 [97], Siddiqi and Vincent proposed another mixed approach that combines allographic and structural features. The allographic feature considered is the PDF of prototypic writing shapes. A codebook of very simple writing shapes (see Fig. 4.8) is built by clustering writing fragments with a k-means clustering algorithm (K set to 100). The allographic feature achieves identification rates of 84% (650 writers from the IAM database) and 74% (375 writers from the RIMES database). Verification performances are EER=4.49% (IAM) and EER=10.57 (RIMES). Regarding structural features, these are based on contours. Two different representations of the contours are considered: one based on Freeman chain codes [98] and another one based on approximating polygons. The former representation is aimed at grabbing details at the pixel-level whereas the latter is aimed at grabbing more coarse-grained details. Using these two representations, a set of 14 features is defined. When combined, structural (contour-based) features alone yield identification ratios of 89% (650 writers from the IAM database) and 85% (375 writers from the RIMES database). In this case, verification performances are EER=2.46% (IAM) and EER=4.87 (RIMES). Finally, the combination of both allographic and structural features produces the following results: IDR = 91% (IAM), IDR=84% (RIMES), IDR = 88% (IAM+RIMES); EER=2.23% (IAM) EER=4.90% (RIMES), EER=2.86% (IAM+RIMES).

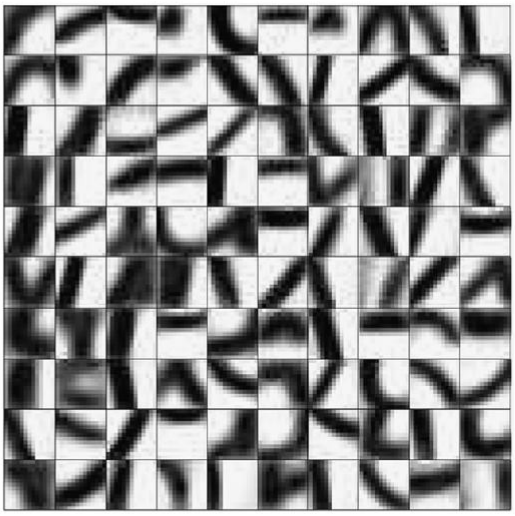


Figure 4.8: A codebook of 100 simple written shapes (Taken from [97]).

4.1.2.4 Other text-independent approaches

Some approaches can neither be considered structural nor allographic. For instance, in a paper published in 2004 [99], Schlapbach and Bunke describe a system that performs writer recognition by means of handwriting recognition. They use, for each writer, an HMM trained to transcript (recognize) the handwriting of that particular writer. Transcriptions also include a log-likelihood score. More accurate transcriptions are expected to have higher scores. To recognize a writer a line of text produced by them is submitted to each HMM and the authorship is granted to the writer whose HMM produces a better transcription (i.e. with the highest score). The experiments yield an identification rate of 96.56% (with 100 writers) and an EER of 2.5% (120 writers). In 2006, the same authors presented a paper [100] that somehow complemented the work in [99]. They used, for each writer, a Gaussian mixture model (GMM) which is computed based on nine pixel-level features extracted using a sliding window.

The interested reader may find a comprehensive summary of less recent text-independent offline identification methods, dating from 1977 until 1983, in [101].

4.1.2.5 Text-dependent approaches

When it comes to text-dependent offline approaches, it is worth mentioning the works of Zois and Anastassopoulos. In [102] a writer recognition method based on a single word (‘*characteristic*’ written both in English and Greek) is proposed. The morphologically processed horizontal profiles of the word are used as features. Experimenting with a database consisting of 45 repetitions of the aforementioned word written by 50 users, the proposed method achieves an identification rate higher than 96% and a verification error smaller than 2.5% In [103] the same authors propose a method based on the fusion of word-level measures taken from the different words of a short sentence. Each word in the sentence is used separately from the rest in order to derive a decision about the authenticity of the writer. At the word level, the method is similar to that reported in [102]. The performance of their method is tested with a database containing 4800 sentences from 20 different writers. The experimental results show a verification error of less than 1% when five-word sentences are considered.

Bensefia, Paquet and Heute, in 2004, also proposed a text-dependent approach based on a single word [104]. Their method, belonging to the *allographic family*, segments words into sequences of graphemes (not necessarily characters) and then compares them by means of the Levenshtein distance. Experimenting with 5 repetitions of the French word *manuscript* donated by 20 different writers the approach yields a false acceptance rate of about 15% (no other performance metrics are reported).

The works of Srihari et al. presented in section 4.1.1 can, to some extent, also be considered text-dependent approaches.

4.1.3 ONLINE APPROACHES

The online field has almost completely been monopolized by signature verification and, therefore, the number of relevant scientific references dealing with non-signature approaches is quite small. What is more, when the topic of online writer recognition is circumscribed to the use of single words or short sentences, then the number of available references is really very scarce.

Contrary to what happens in the offline field, where most approaches are text-independent, in the online field, text-independent approaches do not seem to have attracted much attention from the scientific community. The reason might lie in its perceived lack of applicability both in the forensic field (e.g. *suspects* are not likely to produce several samples of online handwriting) and in the security field since a system relying on not previously know text to validate a user may be impractical due to the amount of text required. Nonetheless, some approaches exist and they deserve some attention. Moreover, the current situation may change in the future as tablet-like devices allowing handwritten input modalities become more and more popular, as the trend seems to be.

4.1.3.1 Text-independent allographic approaches

In 2007, Chan, Tay and Viard-Gaudin presented an allographic approach [105] based on the shapes of the letters (and not in the shapes of portions of letters or fragments of the handwriting as is common in offline approaches due to the difficulty of obtaining more precise segmentations). A set of 10 different prototypes of each of the 26 letters in the English alphabet is used. Each prototype, built by a slightly modified version of the k-means clustering algorithm, is intended to model a specific allographic variation of a letter. Because prototyping at the character level requires a segmentation process involving character recognition, Chan and his colleagues use a language-aware industrial handwriting recognition software engine (MyScript SDK). The following seven time-dependent features are taken into account: x and y coordinates, curvature of x and y coordinates, direction of x and y coordinates and the status of pen-up or pen-down. Each writer is modelled as the frequency distribution of their usage of the letter prototypes. Recognition is performed by following an information retrieval (*IR*) approach [94] (somehow similar to the one proposed in [92]) where the allograph (prototype) frequency (*tf*) and the inverse document frequency (*idf*) are used to compute a weighted Euclidean distance between the vectors representing the models and the query. The identification accuracy of the proposed method is about 95% when experimenting with 82 writers.

In a paper published in 2009 [106], Tan et al. present an improved version of the framework described in the previously reviewed paper. The same language-aware recognition software is used for segmentation purposes and the same seven features are considered. The novelty in [106] lies in the fact that each character, in a model or in a query, is not assigned to a single class (prototype) but instead a fuzzy membership is considered thus allowing a more accurate representation of the variability regarding writing style. Experimenting with 120 writers, a maximum identification accuracy of 99.2% is attained when χ^2 is used as distance metric. Another approach similar to that of Chan et al. ([105]) was proposed, in 2008, by Niels, Grootjen and Vuurpijl [107]. The achieved accuracy reaches 100% when more than 30 characters are used in the queries. Experimentation was performed with a reduced database comprising samples from 43 writers. Nevertheless and contrarily to the approach of Chan et al., this approach assumes that samples have been pre-segmented into characters.

4.1.3.2 Text-independent structural approaches

In a paper by Li, Sun and Tan published in 2007 [108], the idea of using discrete probability distribution functions (PDFs) as features is extended to the online field. Actually their method relies in the comparison of histograms of pressure levels, velocity, azimuth and altitude, extracted from segmented strokes. 12 different types of strokes are considered (primary stroke types) so that for each writer (document) and feature, 12 histograms can be obtained, one for each type of stroke. This method was used to identify the writer of Chinese texts. Experiments involving 55 individuals who donated three pages of handwriting, two used for training, one for testing, yielded a maximum accuracy about 90%. The aforementioned paper also reports on the individual performance of each feature (Table 4.10).

FEATURE	IDENTIFICATION RATE (IDR)
PDF of velocity	~ 82%
PDF of pressure	~77%
PDF of azimuth	~68%
PDF of Altitude	~62%

Table 4.10: Reported performances of the different features considered in [108].

Schlapbach, Liwicki and Bunke presented, in 2008, a work where online data is not obtained from a *conventional* digitizing tablet but from a whiteboard [109]. Data is acquired by means of an *eBeam*® whiteboard system [110], which consists of a normal pen in a special casing that sends infrared signals to a receiver located in one of the corners of the whiteboard (See Fig. 4.9).

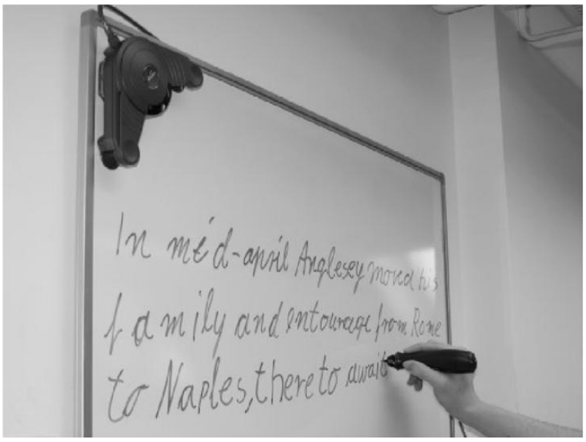


Figure 4.9: Data acquisition from a whiteboard and a capturing device positioned in the upper left corner (Taken from [109]).

In their work, Schlapbach et al. consider five different sets of features: a *point-based feature set*, containing features captured at the point level (writing speed, writing direction and curvature); an *extended point-based feature set*, containing the features of the previous set plus a considerable number of other more complex features; a *point-based offline feature* set containing features computed from a two-dimensional matrix representing an offline version of the data; an *all point-based feature set*, containing all the features in the three previously described sets; and a *stroke-based feature set* containing features captured from whole on-surface trajectories (in-air trajectories are not taken into account). Each writer is modelled by a Gaussian Mixture Model (GMM). First a Universal Background Model (UBM) is obtained using all the available training data (coming from all writers) and, afterwards, a specific GMM is adapted to each writer using their own training data. Recognition is performed as follows: a text line (or paragraph) of unknown origin is presented to each one of the models. Then, each model returns a log-likelihood score for the given text and authorship is granted to the writer whose model produces the highest score. Table 4.11 summarizes the identification performances achieved by this method.

FEATURE SET	IDENTIFICATION RATE	
	LINE LEVEL	PARAGRAPH LEVEL
Point-based set	48.67%	88.56%
Extended point-based set	71.84%	95.75%
Point-based offline set	71.64%	96.44%
All point-based set	88.96%	98.56%
Stroke based feature set	62.55%	92.56%

Table 4.11: Identification performance achieved by each set of features at the line and paragraph level, as reported in [109]. The highest performances are achieved by the *all point-based feature set*.

4.1.3.3 Text dependent approaches

Zuo, Wang and Tan, proposed in 2002, a two-phase text-dependent identification schema for Chinese handwriting [111]. In the first phase, Principal Component Analysis (PCA) is used in order to determine, for each writer, a set of words that best characterize their writing. In the second phase recognition is carried out only taking into account the writer's characterizing words. The writer's identity is established measuring the distances between words that pertain to their characterizing set. In these schema each word is represented as a grey-scale image with the grey level obtained from the pressure signal. Accuracy depends on the number of words in the set. Experimenting with 40 writers that donated 10 repetitions of a Chinese text containing 40 words, the identification rate ranged from 86.5%, with only one word, to 97.5% with ten words.

Another non-western script based method falling in the category of the online text dependent approaches was presented in 2004 by Thumwarin and Matsuura [112]. They proposed a writer recognition in which the trajectory and velocity of the baricenter of the pen movement are computed and their Fourier coefficients obtained. These coefficients are considered the input (velocity) and the output (trajectory) of a Finite Impulse Response system (FIR system). Then, the impulse response is regarded as the feature that characterizes the handwriting. Experimentation is carried out with a database of Thai numerals and scripts obtained from 81 writers. The reported verification accuracy is about 98.85% (FRR=1.50, FAR=0.80).

In 2006 Joulia Chapran presented her research on writer identification [51]. **Excluding our own papers ([113],[114]) this is, to the best of the author's knowledge, the only paper that, up to the date, has been published in an indexed journal and that is devoted to online writer recognition based on isolated short words from western script**. Chapran describes a writer identification system based on the serial combination of a Bayes classifier and a minimum distance classifier. A set of 11 features is chosen (4 static, 7 dynamic) from a larger set of 46 (28 dynamic, 18 static) after an analysis aimed at identifying the most discriminative subset of features, that is, those features that exhibit the highest separation between the intra-writer and the inter-writer distributions. Table 4.12 contains the set of 28 dynamic features considered. In the pre-processing phase, the invisible in-air trajectories are eliminated from the samples. Experimentation is carried out using a database of 45 writers who produced 25 repetitions of each of the following five English words: *February*, *January*, *November*, *October* and *September*. An identification rate of 95% is achieved with at least one of the classifiers trained and tested with the whole set of samples.

ORIGIN	DYNAMIC FEATURE
Time	Relative duration of writing
Pressure	Average pressure
	Amplitude of pressure
	Average displacement in pressure
	Amplitude of displacement in pressure
	Average pressure acceleration
	Amplitude of pressure acceleration
	Number of pen-ups
x-coordinate and y-coordinate	Average horizontal displacement
	Amplitude of horizontal displacement
	Average vertical displacement
	Amplitude of vertical displacement
	Average Cartesian displacement
	Amplitude of Cartesian displacement
	Average horizontal acceleration
	Amplitude of horizontal acceleration
	Average vertical acceleration
	Amplitude of vertical acceleration
	Average Cartesian acceleration
	Amplitude of Cartesian acceleration
Altitude	Average altitude
	Amplitude of altitude
Azimuth	Average azimuth
	Amplitude of the azimuth
	Average angular displacement in azimuth
	Amplitude of angular displacement in azimuth
	Average angular acceleration in azimuth
	Amplitude of angular acceleration in azimuth

Table 4.12: Dynamic features considered in [51]. In red the most discriminative.

4.1.4 SUMMARY OF TEXT-BASED RECOGNITION APPROACHES

The following table comprehensively summarizes the most relevant references regarding text-based writer recognition surveyed in the preceding sections.

Author(s) and reference(s)	Year(s)	Domain	Approach	# of writers	Sample size (aprox.)	Best reported Accuracies	
						Identification (IDR)	verification
Arazi [80]	1977	**OFFLINE**, TI	**STRUCTURAL** Histograms of run-lengths of background intensity	13	9 lines of text	**100%**	
Said et al. [81] [82]	1998, 2000	**OFFLINE**, TI	**STRUCTURAL** Texture analysis (Gabor filters and greyscale co-ocurrence)	20	Small blocks of text (~8 lines)	**96%**	**EER = 0.57%**
Zois et al. [102]	2000	**OFFLINE**, TD	**STRUCTURAL** Morphological features	50	45 repetitions of the word *characteristic*	**96.5%** (English) **97%** (Greek)	**VA=97.7%** (English) **VA=98.6** (Greek)
Zois et al. [103]	2001	**OFFLINE** TD	**STRUCTURAL** Morphological features	20	Sentence with five words		**VA=99%**
Marti et al. [86]	2001	**OFFLINE**, TI	**STRUCTURAL** Features based on *visible* characteristics of the writing	20	5 text pages containing from 5 to 11 lines each	**90.7%**	
Srihari et al. [115][50]	2001-2002	**OFFLINE** TD	**STRUCTURAL** Macro and micro (GSC) features	Differed numbers from a database of 1500		**9%** (words Cohen and referred, 900 writers, macro features). **88%** (set of characters, 900 writers, micro and macro features)	**VA=83.1%** (word referred, 250 users, macro features). **VA=91.75%** (set of characters, 964 writers, micro features)
Zuo et al. [111]	2002	**ONLINE**, TD	**STRUCTURAL** Distances between writer dependent characteristic words obtained by PCA	40	40 words (Chinese). For each writer 10 words are selected	**86.5%** (1 word) **97.5%** (10 words)	
Zhang and Srihari [78]	2003	**OFFLINE** TD	**STRUCTURAL** Micro (GSC) features	875	From 1 to 4 words	**49%** (word *referred*); **83%** (combination of four words)	**VA= 49%** (word *referred*); **VA=91%** (combination of four words)

Author(s) and reference(s)	Year(s)	Domain	Approach	# of writers	Sample size (aprox.)	Best reported Accuracies	
						Identification (IDR)	verification
Hertel and Bunke [87]	2003	**OFFLINE,** TI	**STRUCTURAL** Features based on *visible* characteristics of the writing	50	One line or one page of text	**90.7%** (one line) **99.6%** (one page)	
Tomasi et al. [79]	2004	**OFFLINE** TD	**STRUCTURAL** (different sets of features)	1000	One word	**67%** (word *Grant*)	**VA=43%** (word *Grant*)
Bensefia et al. [104]	2004	**OFFLINE** TD	**ALLOGRAPHIC** Levenshtein distance	20	French word *manuscrit*		**FAR=15%**
Thumwarin et al. [112]	2004	**ONLINE** TD	**STRUCTURAL** FIR system from Fourier coefficients of trajectory and velocity of baricenter of pen movement	81	5 scripts (Thai numerals and/or words)		**FRR=1.50%** **FAR = 0.80%**
Schlapbach et al. [99]	2004	**OFFLINE,** TI	Recognition based on transcription (uses HMMs)	100 (identification) 120 (verification)	27 to 54 text lines	**96.56%**	**EER =2.5%**
Schomaker et al. [89]	2004	**OFFLINE,** TI	**STRUCTURAL** and **ALLOGRAPHIC** Connected component contours and edge-based features	150	Half paragraph (Uppercase)	**85%** (allographic approach only) **95%** (mixed approach)	
Schomaker et al. [91]	2004	**OFFLINE,** TI	**ALLOGRAPHIC** Fragmented component contours	150 (Firemaker database) up to 210 (Unipen[7] database)	Half paragraph (Firemaker) One paragraph (Unipen)	**97%** firemaker **83%** Unipen with 210 writers	
Bensefia et al. [92] [93]	2005	**OFFLINE,** TI	**ALLOGRAPHIC** Grapheme clustering and information retrieval	88 from the PSI database (French) 150 from the IAM database (English)	One page of text (107 words) in the PSI database.	**95%** with the PSI database **86%** with the IAM database	**VA=96%**
Schlapbach et al. [100]	2006	**OFFLINE** TI	**STRUCTURAL** Gaussian mixture model (GMM)	100	27 to 54 text lines	**98.46%**	

[7] Unipen is an online database. *The on-line x_k, y_k coordinates were transformed to a simulated 300-dpi image using a Bresenham line generator and an appropriate brushing function* [Sic]

Author(s) and reference(s)	Year(s)	Domain	Approach	# of writers	Sample size (aprox.)	Best reported Accuracies	
						Identification (IDR)	verification
Chapran [51]	2006	**ONLINE** TD	STRUCTURAL Minimum distance and Bayes classifiers, using selected dynamic and static features	45	25 repetitions of one word	**95%**	
Bulacu et al. [90]	2007	**OFFLINE**, TI	**STRUCTURAL** and **ALLOGRAPHIC** Textural characterization. Allograph characterization.	900 (combining writers from 2 databases)	From 3 lines to a full page	**87%**	**EER = 2.6%**
Chan et al. [105]	2007	**ONLINE**, TI	**ALLOGRAPHIC** Distribution of character prototypes and information retrieval	82	Not reported	**95%**	
Li et al. [108]	2007	**ONLINE** TI	**STRUCTURAL** Histograms of pressure, velocity and angles	55	One page of Chinese text.	**90%**	
Schlapbach et al. [109]	2008	**ONLINE**, TI	**STRUCTURAL** GMM	200	8 paragraphs of 8 lines (average) each (captured from a whiteboard)	**88.96%** (line) **98.56%** (paragraph)	
Niels et al. [107]	2008	**ONLINE**, TI	**ALLOGRAPHIC** Distribution of allograph prototypes and information retrieval	43	From 10 to 100 characters	**100%** (with 30 or more characters)	
Tan et al. [106]	2009	**ONLINE**, TI	**ALLOGRAPHIC** Fuzzy Distribution of character prototypes and information retrieval	120	3 lines of text (~ 160 characters)	**99.2%**	
Siddiqi et al. [97]	2010	**OFFLINE**, TI	**STRUCTURAL** and **ALLOGRAPHIC** Distribution of elementary shapes and contour-based features	375 (RIMES database) And 650 (IAM database)	Not reported	**85%** (RIMES) **91%** (IAM) **88%** (IAM + RIMES)	**EER=4.87%** (RIMES); **EER=2.23%**(IAM) **EER = 2.86%** (IAM+RIMES)
Jain at al. [95]	2011	**OFFLINE**, TI	**ALLOGRAPHIC** Distribution of K-Adjacent Segments	127	Two or three sentences	**99.8%**	

Table 4.13: Summary of text-based recognition approaches (TI and TD stand for text-independent and text-dependent, respectively).

4.2 SIGNATURE-BASED WRITER RECOGNITION

Signature has a long tradition as a method to prove one's identity (legal documents, banking transactions) and it is one of the most widespread means of personal verification, if not the most widespread. Therefore, it should not come as a surprise that signature verification is the handwriting modality that has attracted more research efforts and that has produced more scientific publications. Works until 1993 have been extensively surveyed in [101] and in [116]. Subsequent works up to 2000 are surveyed in [117]. More recently (2008), Impedovo and Pirlo have presented a very comprehensive survey containing more than 350 references [118].

Like handwriting in general, signature verification can be performed offline, when only the scanned images of the signatures are available, or online, when time-dependent data are available. Although some offline approaches exhibit a remarkably good performance ([119], [120]), online methods tend to outperform them.

A wide variety of methods has been proposed to tackle signature verification [118]. Among the most relevant are Hidden Markov Models (HMM) and Dynamic Time Warping (DTW). In the SVC2004 competition [69] (see section 4.2.5), the first place was for a DTW-based system and the second was for a HMM-based system.

4.2.1 HMM-BASED APPROACHES

HMMs are one of the most popular methods belonging to the category of statistical methods. An HMM is a finite state machine where a probability density function (PDF) is associated with each state. States are connected by transition probabilities. Training is carried out using the Baum-Welch algorithm. The likelihood that a sequence of feature vectors was generated by a given model can be computed by the Viterbi algorithm. HMMs have become very successful in speech recognition ([121]) and handwriting recognition. They can manage signals of different time duration. Usually a left-to-right topology is used. Like other statistical methods, HMMs require a considerable number of training samples to achieve an acceptable performance.

In [122] Yang, Widjaja and Prasad proposed a method where signatures were described by the normalized directional angle function of the distance along the signature trajectory. A 6-state left-to-right HMM was used to model each signer. Although pen-up information was not explicitly recorded, when the distance between two consecutive points was greater than a threshold, a pen-up was assumed. Under certain circumstances, the performance was improved if those assumed pen-ups were considered. Experimenting with 31 signers, a FRR=1.75% and a FAR=4.44% were achieved.

Yoon, Lee and Yang, reported an EER of 2.2% also using an HMM but representing signatures by velocity and trajectory in polar space [123]. Experimentation was carried out with signatures from 100 signers.

In 2003, Ortega-García, Fiérrez-Aguilar, Martín-Rello and González-Rodríguez proposed a system [124] where each signer is modelled by a left-to-right 4-state HMM. Signatures are represented using 24 time-sequences: x and y coordinates, pressure, azimuth, altitude, path-tangent angle, path velocity magnitude, log curvature radius and the first and second-order derivatives of all the aforementioned. Very low verification errors are achieved when considering *a posteriori* signer-dependent thresholds: EER<1% for both random and skilled forgeries (experiments were performed with signatures from 50 different signers). This work was extended in 2007 by Fiérrez-Aguilar, Ortega-García, Ramos and González-Rodríguez [125]. This time, each signer is modelled by a left-to-right 2-state HMM and signatures are represented using time sequences similar to those reported in the previous work, although azimuth and altitude are discarded because, according to the authors, they may degrade the verification performance. Again, very low verification errors are achieved: EER=0.05% with random forgeries and EER = 0.74% with skilled forgeries (145 signers, *a posteriori* user-dependent thresholds). This system also participated to SVC2004 with excellent results.

4.2.2 DTW-BASED APPROACHES

DTW is a template matching technique well suited to cope with random variations due to the writer's behaviour (pauses, hesitations) [126]. In DTW, the questioned and reference signatures are compared by means of a dynamic programming strategy that can manage the variability in the signatures' lengths. As HMMs, DTW has been very successfully applied to speech and speaker recognition.

In [127] Jain, Griess and Connell proposed a method, based on DTW that considers a single global feature (the number of strokes) and a set of local ones (x and y coordinates, the difference between two consecutive x coordinates, the same for y coordinates, the sine and cosine of the angle with the x-axis, the curvature, the speed and some others). The similarity measure obtained is compared against a threshold that is the combination of a global one and a writer-dependent one. Experimenting on a database containing more than 1000 signatures from 102 signers, their method achieves low error rates (FRR=2.8%, FAR=1.6%).

Another system also based on DTW and that considered a stroke-oriented description of signatures was presented by Bovino, Impedovo, Pirlo and Sarcinella [128] (also in [129] by Impedovo and Pirlo). Segmentation into strokes is not performed according to the pen-up/pen-down status feature provided by the acquisition device. Instead a more sophisticated method is used: in a first step the whole signature is split into segments based on the analysis of the local maxima in the vertical direction. In a second step each segment is split into strokes on the basis of the local minima in the vertical direction. The reader should take into account that some signatures are performed with a single pen-down stroke. To cope with this possibility a segmentation procedure such as the one described and used by the authors is required. Furthermore, this segmentation approach also copes with the problems posed by long strokes as those found in the embellishments of some signatures. The questioned signature and the model signature are aligned according to the points found by the analysis of the local extremes. As this technique splits each test and reference signature into the same number of strokes there is no need for any treatment of the additional or missing strokes. Each pair of matched strokes can be compared within several representation domains. Actually, they are compared according to position, velocity and acceleration and the different resulting distances are combined to obtain a single similarity measure. An EER of 0.4% is achieved when simple averaging is considered (experimentation carried out with 15 signers that produced both real and forged samples).

In [130] Lee, Yoon, Soh, Chun and Chung also presented a system based on DTW that aligns signatures after having segmented them by means of geometric extrema detection. An EER of 0.98% is reported when experimenting with signatures, without skilled forgeries, provided by 271 people. Accuracy experiments a small decrease to an EER of 1.08% when a limited number of forgeries, realized by 5 people, are considered.

4.2.3 OTHER APPROACHES

Other popular approaches to tackle signature verification are neural networks (NN) and support vector machines (SVM) [131]. In [132], Lee compares the performance of three different types of NNs and the best accuracy is attained by a Bayes multilayer perceptron (EER=2.67%). In [133] Justino, Bortolozzi and Sabourin report a FAR=1.68% and a FRR=3.5% when using a SVM and experimenting with offline signatures donated by 40 signers. In [134], Özgündüz, Şentürk and Karsligil, also using a SVM and experimenting with offline signatures from 70 signers report a FAR=0.11% and a FRR=0.02%.

Fàbregas-Peinado and Faúndez-Zanuy presented, in 2009, a system based on the Biometric Dispersion Matcher (BDM) [135].

Mixed approaches also exist. For instance, in [136] Faundez-Zanuy proposed a combination of Vector Quantization (VQ) and DTW. The combination is performed by means of score fusions and the results are satisfactory since the proposed scheme outperforms other algorithms, achieving minimum detection cost function (DCF) values equal to 1.37% and to 5.42% for random and skilled forgeries respectively, when using a database of 330 signers. When compared to HMM and DTW, VQ exhibits a low computational burden. In subsequent works ([137,138]), VQ-based signature verification has been improved.

4.2.4 APPROACHES CONSIDERING PEN-UP INFORMATION

Although the number of references reporting online approaches for signature verification is quite large [118], the number of significant references that explicitly report the use of pen-up information is meagre. In the following paragraphs two works that we have found relevant will be briefly commented.

Brigitte Wirtz, in [139] presented an online signature verification method based on stroke segmentation and a variant of the DTW algorithm. Her approach considered strokes as the natural structural units of the signature. Both pen-down and pen-up strokes are taken into account. According to the *author:*

> *'[...]The near strokes, that are not visible in the signature image, are used in the same way as the visible strokes since they carry shape and dynamic information.'*[Sic].

DTW is used to compute the distance between (position-matching) strokes. In their contribution to the final similarity measure, the distances between different pairs of strokes can be weighted differently according to their robustness. In order to cope with additions or omissions in the stroke sequence, the DTW algorithm was modified so that it was allowed, in a very constrained way, to consider the distance between strokes occupying non-matching positions in the reference and test sequences. According to the author, when compared to other non stroke-based systems, Wirtz's renders better accuracies. Nevertheless, the number of signers and samples used in testing does not allow to draw significant conclusions. We are commenting on this work because our proposal presents some similarities with Wirtz's: stroke-based segmentation and consideration of both pen-down and pen-up strokes. The use of DTW is quite different: in Wirtz's work, DTW is applied to the strokes themselves while we apply it to sequences of integers (stroke indexes).

In [140] Yanikoglu and Kholmatov presented an online signature verification system based on the Fast Fourier Transform (FFT). They propose to use pen-up-related data to improve the verification performance. In their work, pen-up data is not directly gathered by the acquisition device but interpolated from the last and first points of the pen-down strokes that precede and follow a period in time where no data was acquired. These periods are detected using the timestamps of consecutive points. The incorporation of this pen-up-related data helps increase the overall performance when the system is tested with skilled forgeries (From an EER of 9.09% to an EER of 6.20%). Nevertheless, the overall accuracy of the proposed system lies below that of state-of-the-art approaches.

4.2.5 SIGNATURE COMPETITIONS

Since 2004 several competitions have been organized. In these competitions all the contestants are provided with exactly the same datasets to train and test their systems. Furthermore, all competing systems are tested under the same conditions. All this has the virtue of rendering the results fully comparable.

In 2004 took place the First International Signature Verification Competition (SVC2004) [69]. SCV2004 aimed at allowing researchers to evaluate the performance of their verification systems. The contestants were proposed two different verification tasks. In the first task, only positional information was available whereas in the second task positional information was complemented with pen inclination (azimuth and altitude) and pressure. Testing was carried out with signatures from 60 different donors. For each donor, 10 genuine and 20 forged signatures were considered. When skilled forgeries were taken into account, the winning system, for both tasks, was [141], based on DTW and presented by Kholmatov and Yanikoglu, from Sabanci University (Turkey). When only random forgeries were considered the best results were achieved by an anonymous system (task 1) and by the HMM-based system [125] submitted by *Universidad Autónoma de Madrid* (task 2). Although the results from SVC2004 are widely cited and quite often still considered the state-of-the art figures in online signature verification, the following facts should not be disregarded [142]: first of all, the signatures were not true ones but made up signatures since the donors were advised not to use their real signatures. Secondly, the signers had different cultural origins and, finally, the size of the testing database was limited (60 users) with all signatures collected in a single session.

Other public campaigns have been organized for comparing advances in online signature verification. In 2009, the BioSecure evaluation campaign (BSEC'2009) [142] was held. BSEC'2009 had several different goals: to assess the impact of the acquiring device (a digitizing tablet or a PDA touch screen), to assess the impact of time variability (single session vs. multisession) and to assess the impact of signature complexity. Also in 2009, was held the ICDAR 2009 Signature Verification Competition (SigComp2009) [143]. This competition included online and offline modalities, with the online modality clearly outperforming the offline one. In 2011, the BioSecure Signature Evaluation Campaign (ESRA'2011) was held [144]. One of its main goals was to assess the impact of the quality of the skilled forgeries in the accuracy of online verification systems. Table 4.14 summarizes the best results achieved in the aforementioned competitions.

COMPETITION	TESTING CONDITIONS	PERFORMANCE (EER)	
		WITH SKILLED FORGERIES	WITH RANDOM FORGERIES
SVC2004 [69]	Tested with signatures from 60 donors. One session only	**2.84%** with x and y coordinates only	**1.85%** with x and y coordinates only
		2.89% with x and y coordinates, pressure and writing angles	**1.70%** with x and y coordinates, pressure and writing angles
ICDAR 2009 (SigComp2009) [143]	Tested with signatures from 100 donors (online modality)	**2.85%** with x and y coordinates, pressure and writing angles.	
BSEC'2009 [142]	Tested with signatures from 382 donors, acquired from a digitizing tablet and from a PDA	**2.20%** (from tablet) and **4.97%** (from PDA) with x and y coordinates only	**0.51%** (from tablet) and **0.55%** (from PDA) with x and y coordinates only
		1.71% (from tablet) with x and y coordinates, pressure and writing angles.	**0.42%** (from tablet) with x and y coordinates, pressure and writing angles.
		3.48% (from tablet) with x and y coordinates, pressure and writing angles. Multisession	**1.37%** (from tablet) with x and y coordinates, pressure and writing angles. Multisession
ESRA'2011 [144]	Tested with signatures from 382 donors acquired during two sessions. Two devices used: a digitizing tablet and a PDA.	**2.85%** (from tablet) and **7.15%** (from PDA) with x and y coordinates only. Forgeries of good quality	
		2.43% (from tablet) with x and y coordinates, pressure and writing angles. Forgeries of good quality	

Table 4.14: Best results achieved in the online signature verification competitions.

5

AN INFORMATION ANALYSIS OF IN-AIR AND ON-SURFACE TRAJECTORIES

This chapter details the analysis of the in-air and on-surface trajectories that constitute the pen-up and pen-down strokes found in online handwriting. This analysis, performed from the perspective of the information theory, is independent of any recognition system since no assumption is made on how the analyzed data will be used to perform recognition. Three different issues are considered: (a) the amount of information found in each type of trajectory, (b) the level of redundancy between them and (c) the differences, for each type of trajectory, among the intra- and the inter-user cases.

The chapter is organized as follows: the first section (re)introduces the distinction between in-air and on-surface trajectories, showing some graphical examples. The second section provides a very brief introduction to information theory, highlighting the definitions of the measures that are used thereafter. In the third section, the core of the chapter, the results of the analysis are presented. Finally, the fourth section summarizes the most relevant conclusions drawn from the results of the analysis.

5.1 IN-AIR VS. ON SURFACE

In chapter 3, the distinction between on-surface and in-air trajectories has been introduced.

On-surface trajectories (pen-downs) correspond to the movements executed while the writing device is touching the writing surface. Each of these trajectories produces a visible stroke.

In-air trajectories (pen-ups) correspond to the movements performed by the hand while transitioning from one stroke to the next. During these movements the writing device exerts no pressure on the surface.

Fig. 5.1 shows an execution of the word INEXPUGNABLE where the distinction between pen-up and pen-down strokes has been emphasized.

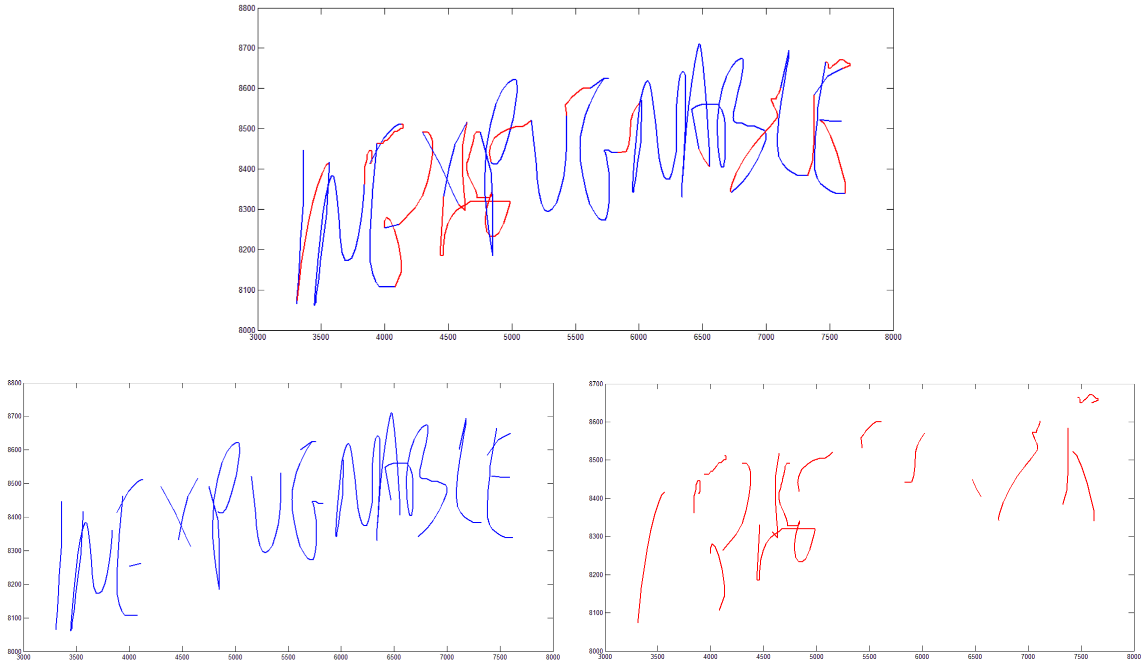


Figure 5.1: Execution of the word INEXPUGNABLE as captured by the acquisition device. Image at the top shows both pen-up and pen-down strokes. Images at the bottom show pen-down (left) and pen-up (right) strokes separately.

Applications based on the analysis of online handwriting (mainly character recognition and online signature verification systems) have paid little or no attention at all to in-air trajectories. Indeed, all the attention has been focused on on-surface trajectories while in-air ones have been disregarded or simply discarded. The results of our research show that in-air trajectories not only contain individual information but that this information is rich enough to perform writer recognition and that it can be combined with information from on-surface trajectories to enhance the recognition accuracy, since both types of information are to a considerable extent non-redundant. Our research extends the reach of the hypothesis of writer individuality to the invisible part of handwriting and gives evidence pointing towards a positive answer: the invisible part of handwriting contains enough personal information to successfully discriminate among writers.

Fig. 5.2 shows two executions of the word BIODEGRADABLE performed by two different writers. Notice how in-air trajectories are quite different from one writer to the other while quite similar when performed by the same writer. This difference may suggest a noticeable inter-writer variability and lower intra-writer variability for this type of trajectories. Also notice that in-air trajectories appear to have a considerable degree of complexity.

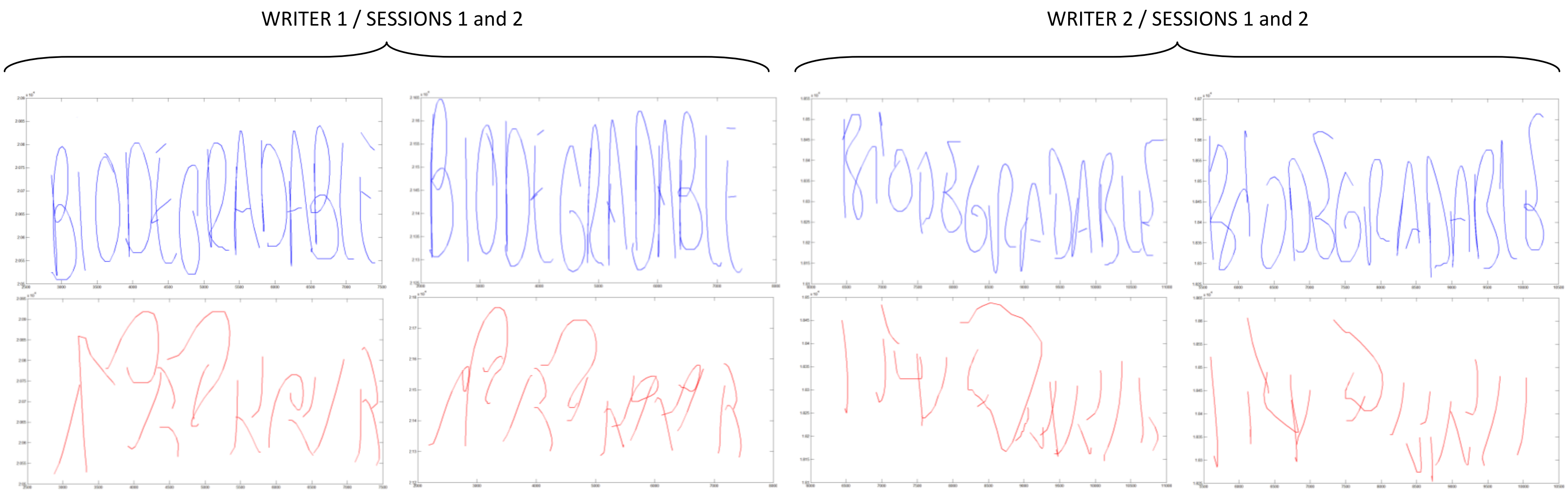


Figure 5.2: On-surface (top) and in-air (bottom) trajectories from different executions of the word BIODEGRADABLE performed by two writers.

5.2 BACKGROUND ON INFORMATION THEORY

In order to facilitate the understanding of the following sections, this section provides a very brief introduction to information theory. The most relevant aspects, connected to the results shown in this dissertation, are highlighted. The interested reader can find in the literature much more in-depth treatments of the topic (e.g. [145] [146])

If X is a random variable with several possible values x and a marginal probability distribution function p(x), the *entropy* of X, measured in bits, is defined as

$$H(X) = -\sum_{x \in X} p(x) \cdot \log_2(p(x)) \tag{5.1}$$

H(X) is a measure of the uncertainty associated with X. If X is a source of data or a message, then H(X) measures the average information content in X. Other equivalent interpretations are also possible. For instance H(X) is the average number of bits (binary symbols) required to encode all the possible outcomes (values) of X.

For two random variables, X and Y, with possible values x and y, a joint probability distribution function p(x,y) and marginal distribution functions p(x) and p(y) respectively, the following measures are often considered:

(a) *Conditional entropy*, often called equivocation in information theory, quantifies the remaining entropy (i.e. the uncertainty) of one of the variables when the value of the other one is know. It is defined as:

$$H(Y \mid X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2(p(y \mid x)) \tag{5.2}$$

(b) *Joint entropy*, a measure of the amount of information in the joint system of X and Y. Its definition is:

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2(p(x, y)) \tag{5.3}$$

(c) *Mutual information*, a measure of the amount of information shared by X and Y. It is defined as:

$$I(X;Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2\left(\frac{p(x, y)}{p(x) \cdot p(y)}\right) \tag{5.4}$$

Intuitively, a low value for $I(X;Y)$ suggests that X and Y provide different, non redundant, information. Notice that $I(X;Y) = 0$ if and only if X and Y are independent (the knowledge of one has no effect whatsoever on the knowledge of the other).

These three measures are tightly related to each other:

$$I(X;Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,Y) \tag{5.5}$$

Fig. 5.3 graphically depicts the relations among conditional entropy, joint entropy and mutual information, as shown in [146]

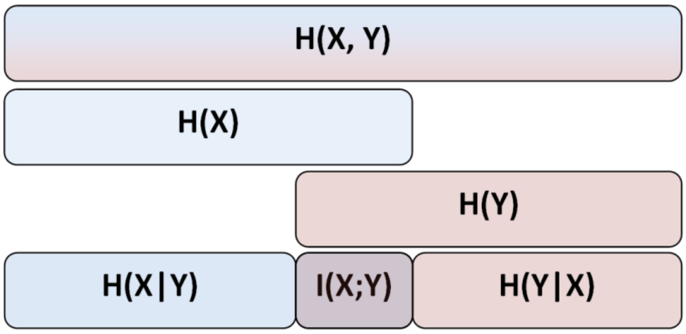


Figure 5.3 Relations among the individual entropies (H(X), H(Y)), the conditional entropies (H(X|Y), H(Y|X)), the joint entropy (H(X,Y)) and the mutual information (I(X;Y))

In order to facilitate the comparison of amounts of mutual information obtained from different pairs of random variables, I(X;Y) can be expressed relative to H(X,Y):

$$I'(X;Y) = I(X;Y) \,/\, H(X,Y) \tag{5.6}$$

Thus, *relative mutual information I'(X;Y)* is the proportion of the joint entropy that is shared by both random variables.

5.3 INFORMATION ANALYSIS OF TRAJECTORIES

In the forthcoming subsections, in-air and on-surface trajectories will be scrutinized from the perspective of the information theory. The following issues will be analyzed: the average amount of information contained in each feature of each type of trajectory, the amount of information they share (redundancy), and the differences between the intra and the inter-writer measures of the mutual information. All the measures have been obtained from the 16 uppercase words in the BiosecurID database (described in 3.4.2). A subset of 100 writers has been considered, each one providing 4 repetitions of each word, thus totalling 400 executions of each word.

5.3.1 ENTROPY OF EACH FEATURE

For each word, type of trajectory and feature, the entropy has been computed. The results are shown in Table 5.1. Each figure was obtained averaging the 400 executions of each word. All entropies are expressed in *bits* (i.e. log_2 is considered when computing *H(x)*). Figure 5.4 provides a summarized view of data in Table 5.1.

WORD	TEXT	LENGTH	PRESSURE				X-COORD				Y-COORD				AZIMUTH				ALTITUDE			
			IN-AIR		ON-SURF.		IN-AIR		ON-SURF.		IN-AIR		ON-SURF.		IN-AIR		ON-SURF.		IN-AIR		ON-SURF.	
			AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
W1	BIODEGRADABLE	12	n/a	n/a	7,7	0,2	7,6	0,4	7,7	0,3	7,1	0,4	7,2	0,3	4,1	0,5	4,0	0,5	3,1	0,6	2,6	0,5
W2	DELEZNABLE	10	n/a	n/a	7,6	0,2	7,3	0,4	7,1	0,3	6,9	0,4	6,9	0,3	3,9	0,5	3,8	0,5	3,0	0,6	2,5	0,5
W3	DESAPROVECHAMIENTO	18	n/a	n/a	7,9	0,2	8,0	0,4	7,9	0,3	7,4	0,3	7,4	0,3	4,3	0,5	4,1	0,5	3,1	0,5	2,6	0,5
W4	DESBRIZNAR	10	n/a	n/a	7,6	0,2	7,1	0,5	7,4	0,3	6,7	0,4	7,1	0,3	3,9	0,6	3,9	0,6	3,0	0,6	2,7	0,5
W5	DESLUMBRAMIENTO	15	n/a	n/a	7,8	0,2	7,6	0,5	7,8	0,3	7,1	0,4	7,4	0,3	4,1	0,5	4,0	0,6	3,1	0,6	2,7	0,5
W6	DESPEDAZAMIENTO	15	n/a	n/a	7,8	0,2	7,8	0,4	7,7	0,3	7,2	0,4	7,3	0,3	4,2	0,6	4,1	0,6	3,0	0,6	2,7	0,5
W7	DESPRENDER	10	n/a	n/a	7,6	0,2	7,1	0,5	7,3	0,3	6,7	0,5	7,0	0,3	3,9	0,5	3,9	0,6	2,8	0,6	2,6	0,5
W8	ENGUALDRAPAR	12	n/a	n/a	7,7	0,2	7,4	0,5	7,6	0,3	6,9	0,4	7,2	0,3	4,1	0,6	4,1	0,5	3,1	0,6	2,6	0,5
W9	EXPRESIVIDAD	12	n/a	n/a	7,6	0,2	7,4	0,4	7,3	0,3	6,9	0,4	7,0	0,3	4,1	0,5	4,1	0,5	3,0	0,5	2,7	0,5
W10	IMPENETRABLE	12	n/a	n/a	7,7	0,2	7,5	0,4	7,5	0,3	7,0	0,4	7,1	0,3	4,2	0,6	4,1	0,6	3,0	0,6	2,7	0,5
W11	INEXPUGNABLE	12	n/a	n/a	7,6	0,2	7,5	0,5	7,4	0,3	7,0	0,4	7,0	0,3	4,2	0,5	4,1	0,5	3,1	0,6	2,7	0,6
W12	INFATIGABLE	11	n/a	n/a	7,6	0,2	7,3	0,4	7,2	0,4	6,9	0,3	6,9	0,4	4,0	0,5	4,0	0,6	3,0	0,6	2,5	0,6
W13	INGOBERNABLE	12	n/a	n/a	7,6	0,2	7,3	0,5	7,6	0,3	6,9	0,4	7,2	0,3	4,2	0,5	4,1	0,5	3,1	0,6	2,6	0,5
W14	MANSEDUMBRE	11	n/a	n/a	7,6	0,2	7,3	0,5	7,5	0,3	6,8	0,4	7,2	0,3	4,2	0,5	4,2	0,5	3,2	0,6	2,7	0,5
W15	ZAFARRANCHO	11	n/a	n/a	7,6	0,2	7,3	0,5	7,5	0,4	6,8	0,6	7,1	0,5	4,2	0,5	4,1	0,6	3,2	0,6	2,6	0,5
W16	ZARRAPASTROSA	13	n/a	n/a	7,7	0,2	7,6	0,4	7,8	0,3	7,0	0,8	7,3	0,8	4,3	0,5	4,2	0,5	3,3	0,6	2,8	0,6
Average over the 16 words			**n/a**	**n/a**	**7.7**	**0.2**	**7.4**	**0.5**	**7.5**	**0.3**	**7.0**	**0.4**	**7.1**	**0.4**	**4.1**	**0.5**	**4.1**	**0.5**	**3.1**	**0.6**	**2.6**	**0.5**

Table 5.1: Entropy in bits (average –avg- and standard deviation –std-) of each feature.

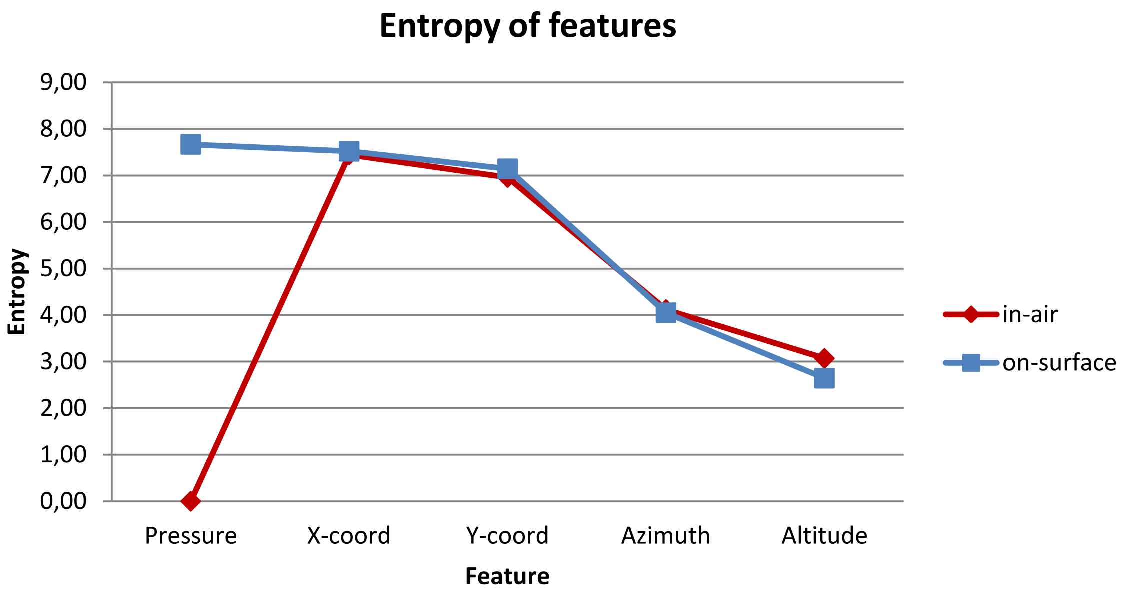


Figure 5.4: Entropy of each feature. The values shown are the averages among the 16 words of the entropies shown in table 5.1.

The following facts are worth noticing:

(a) Pressure and coordinates contain much more information than writing angles. For instance, in the in-air case, the entropy of the X coordinate is about 7.5 bits. This represents a $2^{7.5} \approx 180$ different *states* for this feature. For the azimuth, the entropy is about 4.1 bits, representing $2^{4.1} \approx 17$ different *states*. Nevertheless, these differences among features should not come as a surprise since their ranges of possible values, imposed by the acquisition device, are rather different too (see Table 3.2 in chapter 3).

(b) If pressure is not taken into account, the global amount of information (considering all the other features) is almost the same in both types of trajectories. Figs. 5.1 and 5.2 already suggested that for the X and Y coordinates the amount of information in the in-air case might not be much lower than in the on-surface case, because both cases do not appear to have very different degrees of *complexity*.

(c) X-coordinate and Y-coordinate, especially the latter, tend to have higher entropies in on-surface trajectories

(d) On the other hand, both azimuth and altitude have higher entropies in in-air trajectories

(e) Variability (measured by the standard deviation of the entropy of the 400 executions considered) is low. This means that, in average, there are no great differences among users and sessions.

5.3.2 REDUNDANCY BETWEEN IN-AIR AND ON-SURFACE TRAJECTORIES

The fact that in-air and on-surface trajectories are different does not imply that they convey entirely different information. Mutual and the relative mutual information between pairs of measures of the same feature, taken from in-air and on-surface trajectories, will be used in order to evaluate the degree of redundancy between in-air and on-surface trajectories.

For a given word and feature f, fa_u^e and fs_u^e respectively denote the in-air and on-surface values of that feature for the e-th execution of this word performed by writer u. $Joint^{air-surface}$ denotes the average joint entropy between pairs of measures (one in-air, one on-surface) of that feature taken from the same user and execution. Analogously, $Mutual^{air-surface}$ and $RMutual^{air-surface}$ denote the average mutual information and relative mutual information between pairs of measures of a feature, taken from the same user and execution:

$$Joint^{air-surface} = \underset{\forall e, \forall u}{avg}(H(fa_u^e, fs_u^e)) \qquad (5.7)$$

$$Mutual^{air-surface} = \underset{\forall e, \forall u}{avg}(I(fa_u^e; fs_u^e)) \qquad (5.8)$$

$$RMutual^{air-surface} = \underset{\forall e, \forall u}{avg}(I'(fa_u^e; fs_u^e)) \qquad (5.9)$$

Table 5.2 contains the average values obtained for $Joint^{air-surface}$, $Mutual^{air-surface}$ and $RMutual^{air-surface}$. In the case of $RMutual^{air-surface}$, the standard deviation is also shown. Fig. 5.5 provides a summarized view of the proportion between joint entropy and mutual information.

WORD	TEXT	LENGTH	X-COORD				Y-COORD				AZIMUTH				ALTITUDE			
			JOINT	MUTUAL	RMUTUAL		JOINT	MUTUAL	RMUTUAL		JOINT	MUTUAL	RMUTUAL		JOINT	MUTUAL	RMUTUAL	
			AVG	AVG	AVG	STD	AVG	AVG	AVG	STD	AVG	AVG	AVG	STD	AVG	AVG	AVG	STD
W1	BIODEGRADABLE	12	8,4	6,8	**0,81**	0,03	8,4	5,8	**0,70**	0,04	6,5	1,6	**0,25**	0,09	5,2	0,5	**0,10**	0,06
W2	DELEZNABLE	10	8,0	6,4	**0,80**	0,03	8,0	5,6	**0,71**	0,04	6,2	1,6	**0,25**	0,08	4,9	0,5	**0,10**	0,06
W3	DESAPROVECHAMIENTO	18	8,7	7,1	**0,82**	0,02	8,7	6,0	**0,68**	0,04	6,8	1,6	**0,23**	0,09	5,2	0,5	**0,09**	0,06
W4	DESBRIZNAR	10	8,0	6,4	**0,79**	0,03	8,0	5,6	**0,70**	0,04	6,1	1,6	**0,26**	0,09	5,0	0,6	**0,12**	0,07
W5	DESLUMBRAMIENTO	15	8,5	6,8	**0,80**	0,03	8,5	5,9	**0,69**	0,04	6,5	1,7	**0,26**	0,09	5,1	0,6	**0,11**	0,06
W6	DESPEDAZAMIENTO	15	8,5	6,9	**0,81**	0,02	8,5	5,9	**0,69**	0,04	6,6	1,7	**0,26**	0,09	5,1	0,6	**0,10**	0,06
W7	DESPRENDER	10	8,0	6,4	**0,79**	0,04	8,0	5,6	**0,70**	0,04	6,1	1,7	**0,28**	0,10	4,9	0,6	**0,12**	0,08
W8	ENGUALDRAPAR	12	8,3	6,6	**0,80**	0,03	8,3	5,7	**0,69**	0,04	6,4	1,8	**0,28**	0,10	5,1	0,6	**0,11**	0,06
W9	EXPRESIVIDAD	12	8,1	6,5	**0,80**	0,03	8,1	5,8	**0,71**	0,04	6,3	1,9	**0,29**	0,09	5,0	0,6	**0,12**	0,08
W10	IMPENETRABLE	12	8,3	6,6	**0,80**	0,03	8,3	5,8	**0,70**	0,04	6,4	1,8	**0,28**	0,10	5,1	0,6	**0,12**	0,07
W11	INEXPUGNABLE	12	8,2	6,6	**0,80**	0,03	8,2	5,7	**0,70**	0,04	6,4	1,9	**0,29**	0,09	5,1	0,6	**0,12**	0,07
W12	INFATIGABLE	11	8,0	6,4	**0,80**	0,03	8,0	5,7	**0,71**	0,04	6,2	1,8	**0,29**	0,10	4,9	0,6	**0,13**	0,08
W13	INGOBERNABLE	12	8,3	6,6	**0,80**	0,03	8,3	5,7	**0,69**	0,04	6,4	1,9	**0,30**	0,10	5,1	0,7	**0,13**	0,08
W14	MANSEDUMBRE	11	8,2	6,5	**0,79**	0,03	8,3	5,8	**0,69**	0,04	6,4	1,9	**0,29**	0,09	5,1	0,6	**0,12**	0,07
W15	ZAFARRANCHO	11	8,2	6,5	**0,79**	0,03	8,2	5,7	**0,69**	0,05	6,4	1,9	**0,29**	0,10	5,1	0,6	**0,12**	0,07
W16	ZARRAPASTROSA	13	8,4	6,8	**0,81**	0,03	8,4	5,9	**0,70**	0,04	6,6	1,9	**0,28**	0,09	5,4	0,6	**0,11**	0,06
Average over the 16 words			**8.3**	**6.6**	**0.80**	**0.03**	**8.3**	**5.8**	**0.70**	**0.04**	**6.4**	**1.8**	**0.27**	**0.09**	**5.1**	**0.6**	**0.11**	**0.07**

Table 5.2: Relations between in-air and on-surface trajectories measured by their joint, mutual and relative mutual information. Columns containing values of relative mutual information appear shadowed to facilitate reading.

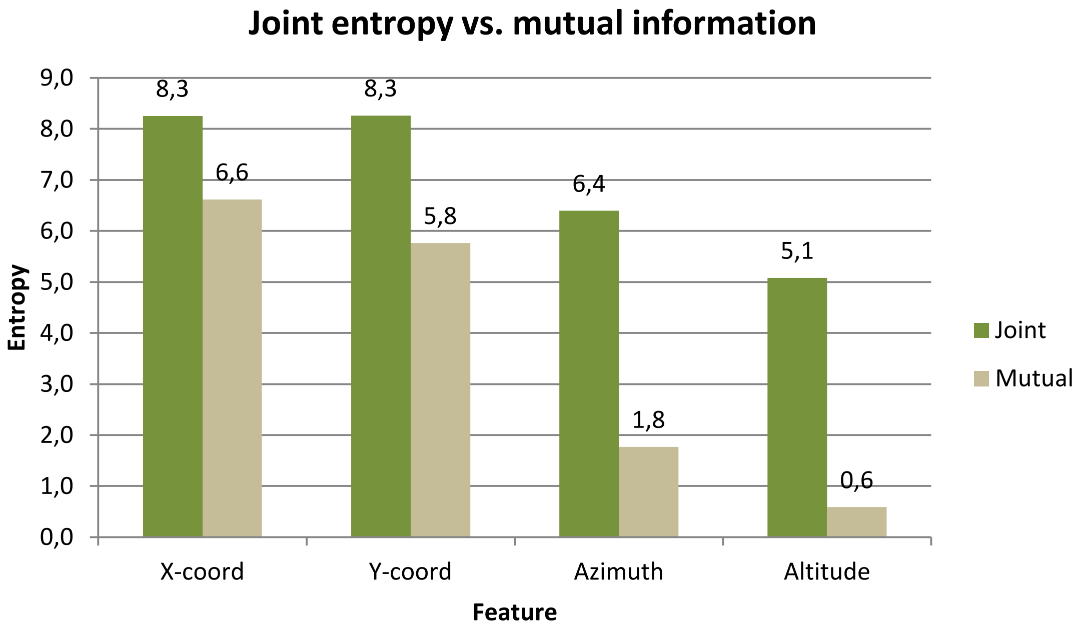


Figure 5.5: Comparison of joint entropy and mutual information. The values shown are the averages among the words of the values shown in table 5.2.

The following facts are worth noticing:

(a) X-coord shows a redundancy of about 6.5 bits (with relative mutual information around 0.8, that is, about 80%) while Y-coord shows a redundancy of slightly less than 6 bits (70%). Although in both cases redundancy is high, there is still a significant amount of non-redundant information (20%-30%). The reader should notice that entropy being a logarithmic measure, one bit of difference amounts to a multiplication factor of 2 in the number of states. Thus, a difference of 1.5 bits between the joint entropy and the mutual entropy (as in the X-coordinate) amounts to a multiplication factor of $2^{1.5} \approx 2.83$ in the number of states.

(b) On the other hand, azimuth and altitude, especially the latter, show a very low redundancy. In the case of azimuth, it is less than 2 bits (25%-30%) and when it comes to altitude, it is less than 1 bit (about 10%)

(c) Variability, measured by the standard deviation, is low (azimuth and altitude) and very low (X-coord, Y-coord). This means that, on average, all users/sessions have a similar behaviour with respect to redundancy.

5.3.3 INTER-WRITER AND INTRA-WRITER DIFFERENCE

From a biometric recognition perspective, the fitness of a feature to perform recognition does not only depend on the amount of information it contains but also on the difference between the intra-writer and the inter-writer case. Given a feature *f*, it is highly desirable that different measures of *f* taken from the same writer are closer to each other than measures taken from different writers. From an information theory perspective, it would be desirable that the amount of mutual information was higher when considering the same writer (intra-writer) than when considering different writers (inter-writer). The average difference between both cases will be used as a mean to evaluate the potential usefulness of a given feature.

As in the previous subsection, for a given word and feature f, fa_u^e and fs_u^e respectively denote the in-air and on-surface values of that feature for the *e*-th execution of this word performed by user *u*. For a given word and feature, $Intra_u^{air}$ and $Intra_u^{surface}$ denote the average values for all measures of the relative mutual information between different executions of this word performed by writer *u*.

$$Intra_u^{air} = \underset{i \neq j}{avg}(I'(fa_u^i; fa_u^j)) \tag{5.10}$$

$$Intra_u^{surface} = \underset{i \neq j}{avg}(I'(fs_u^i; fs_u^j)) \tag{5.11}$$

Analogously, for a given word and feature, $Inter_u^{air}$ and $Inter_u^{surface}$ denote the average value of the relative mutual information between executions of this word performed by writer *u* and any other writer.

$$Inter_u^{air} = \underset{u \neq v}{avg}(I'(fa_u^*; fa_v^*)) \tag{5.12}$$

$$Inter_u^{surface} = \underset{u \neq v}{avg}(I'(fs_u^*; fs_v^*)) \tag{5.13}$$

Where * means any execution.

Finally, $Diff_u^{air}$ and $Diff_u^{surface}$ denote the differences between the inter-writer and the intra-writer measures for the in-air and on-surface cases respectively:

$$Diff_u^{air} = Intra_u^{air} - Inter_u^{air} \tag{5.14}$$

$$Diff_u^{surface} = Intra_u^{surface} - Inter_u^{surface} \tag{5.15}$$

Should $Diff_u^{air} > 0$ and $Diff_u^{surface} > 0$, this would mean that, on average and relative to their joint entropies, the executions from writer u share more information among them than they share with executions from other writers.

Table 5.3 shows, for each word and feature, the average values for $Diff_u^{surface}$ and their standard deviation. The reader will notice that the averages are all positive but quite close to zero. In order to determine whether these average values are significantly positive, they have been put to a Student's unilateral paired t-test with the following parameters: null hypothesis $H_0 : \underset{\forall u}{avg}(Diff_u^{surface}) = 0$; alternative hypothesis $H_1 : \underset{\forall u}{avg}(Diff_u^{surface}) > 0$; degrees of freedom: 99. For each feature, the third column (*p-val*) contains the *p-value* of the test. The *p-value* is the probability of obtaining an average value for $Diff_u^{surface}$ as extreme as the one that was actually obtained, assuming that the null hypothesis is true. Table 5.4 shows the same results for the in-air trajectories.

Prior to being put to the Student's t-test, all the differences have been put to a Kolmogorov-Smirnov test in order to determine whether it can be assumed that they follow a normal distribution (a precondition for the applicability of the t-test). With a significance level of $\alpha=0.05$, the normality hypothesis is only rejected in four cases (out of 144): the differences for the altitude of on-surface trajectories in words 1, 2, 3 and 16. All the other differences are consistent with the normality hypothesis.

Fig. 5.6 summarizes and compares the p-values obtained for both types of trajectories. P-values below 0.05 (significance level $\alpha=95\%$) may be considered statistical evidence of a significant difference.

WORD	TEXT	PRESSURE			X-COORD			Y-COORD			AZIMUTH			ALTITUDE		
		AVG	STD	P-VAL	AVG	STD	P-VAL	AVG	STD	P-VAL	AVG	STD	P-VAL	AVG	STD	P-VAL
W1	BIODEGRADABLE	0,003	0,02	6,4E-02	0,018	0,02	1,2E-13	0,017	0,02	7,1E-12	0,019	0,05	2,3E-04	0,017	0,04	
W2	DELEZNABLE	0,004	0,02	3,1E-02	0,025	0,02	1,7E-19	0,023	0,02	5,9E-17	0,021	0,05	4,1E-05	0,020	0,04	
W3	DESAPROVECHAMIENTO	0,003	0,02	9,4E-02	0,017	0,02	1,1E-15	0,016	0,02	3,4E-12	0,019	0,05	1,2E-04	0,017	0,03	
W4	DESBRIZNAR	0,003	0,02	8,4E-02	0,018	0,02	5,5E-12	0,017	0,02	1,3E-10	0,025	0,06	1,3E-05	0,024	0,04	1,3E-07
W5	DESLUMBRAMIENTO	0,003	0,02	6,1E-02	0,018	0,02	4,9E-14	0,015	0,02	4,7E-09	0,022	0,06	9,4E-05	0,021	0,04	5,3E-06
W6	DESPEDAZAMIENTO	0,003	0,02	1,0E-01	0,016	0,02	3,1E-12	0,016	0,02	1,2E-10	0,022	0,06	4,3E-04	0,020	0,04	8,9E-07
W7	DESPRENDER	0,004	0,02	2,2E-02	0,018	0,02	8,3E-13	0,018	0,02	1,3E-11	0,026	0,07	5,1E-05	0,024	0,04	1,1E-08
W8	ENGUALDRAPAR	0,004	0,02	2,1E-02	0,018	0,02	1,1E-13	0,014	0,02	9,2E-10	0,024	0,06	3,7E-05	0,017	0,04	2,6E-05
W9	EXPRESIVIDAD	0,005	0,02	1,5E-02	0,020	0,02	8,5E-18	0,019	0,03	6,2E-12	0,022	0,05	4,9E-05	0,025	0,05	9,6E-07
W10	IMPENETRABLE	0,003	0,02	8,5E-02	0,018	0,02	5,3E-14	0,017	0,02	3,3E-12	0,023	0,06	1,3E-04	0,021	0,04	4,0E-06
W11	INEXPUGNABLE	0,003	0,02	5,8E-02	0,021	0,02	5,4E-16	0,018	0,02	2,1E-11	0,021	0,05	6,5E-05	0,021	0,05	1,4E-05
W12	INFATIGABLE	0,004	0,02	4,1E-02	0,024	0,02	5,7E-18	0,023	0,03	3,8E-15	0,019	0,05	1,9E-04	0,023	0,04	2,1E-07
W13	INGOBERNABLE	0,003	0,02	6,8E-02	0,019	0,02	1,0E-11	0,017	0,02	2,3E-10	0,024	0,06	3,5E-05	0,020	0,05	8,5E-06
W14	MANSEDUMBRE	0,003	0,02	1,1E-01	0,020	0,02	1,2E-16	0,017	0,02	2,7E-11	0,019	0,05	1,9E-04	0,023	0,05	5,6E-06
W15	ZAFARRANCHO	0,003	0,02	6,1E-02	0,019	0,02	5,8E-16	0,019	0,02	1,9E-13	0,022	0,06	3,5E-04	0,013	0,04	1,1E-03
W16	ZARRAPASTROSA	0,003	0,02	6,7E-02	0,015	0,02	6,6E-12	0,014	0,02	3,2E-08	0,020	0,05	2,6E-04	0,018	0,04	
Average over the 16 words		**0,003**	**0,02**	**6,1E-02**	**0,019**	**0,02**	**1,7E-12**	**0,018**	**0,02**	**2,4E-09**	**0,022**	**0,06**	**1,4E-04**	**0,020**	**0,04**	**9,7E-05**

Table 5.3: Differences in relative mutual information between the inter-writer and intra-writer case for on-surface trajectories. P-values shadowed green correspond to statistically significant differences (α=95%). No p-values have been computed for altitude in words 1, 2, 3 and 16 because differences did not pass the normality test.

WORD	TEXT	X-COORD			Y-COORD			AZIMUTH			ALTITUDE		
		AVG	STD	P-VAL	AVG	STD	P-VAL	AVG	STD	P-VAL	AVG	STD	P-VAL
W1	BIODEGRADABLE	0,011	0,02	4,9E-06	0,011	0,02	3,7E-06	0,017	0,05	1,2E-03	0,015	0,04	1,4E-04
W2	DELEZNABLE	0,010	0,03	2,5E-04	0,010	0,02	5,8E-05	0,015	0,05	2,4E-03	0,014	0,04	4,7E-04
W3	DESAPROVECHAMIENTO	0,009	0,02	2,4E-04	0,009	0,03	2,0E-04	0,015	0,05	1,9E-03	0,010	0,03	2,4E-03
W4	DESBRIZNAR	0,014	0,03	4,3E-06	0,012	0,03	1,1E-04	0,022	0,05	2,5E-05	0,018	0,04	4,3E-05
W5	DESLUMBRAMIENTO	0,013	0,03	2,1E-06	0,013	0,02	4,1E-07	0,018	0,05	6,1E-04	0,013	0,04	6,7E-04
W6	DESPEDAZAMIENTO	0,011	0,03	2,7E-05	0,010	0,02	5,0E-05	0,017	0,06	2,7E-03	0,013	0,04	5,3E-04
W7	DESPRENDER	0,019	0,03	2,1E-09	0,018	0,03	8,1E-11	0,025	0,07	1,2E-04	0,021	0,05	1,0E-05
W8	ENGUALDRAPAR	0,011	0,03	1,5E-04	0,011	0,03	6,4E-05	0,021	0,06	2,2E-04	0,014	0,04	9,1E-05
W9	EXPRESIVIDAD	0,013	0,03	1,0E-06	0,013	0,03	1,5E-06	0,016	0,05	1,1E-03	0,018	0,05	7,7E-05
W10	IMPENETRABLE	0,011	0,02	3,3E-06	0,011	0,02	1,9E-06	0,018	0,06	1,5E-03	0,017	0,04	6,7E-05
W11	INEXPUGNABLE	0,014	0,02	7,0E-08	0,013	0,02	3,4E-07	0,018	0,05	3,3E-04	0,015	0,04	1,4E-04
W12	INFATIGABLE	0,012	0,03	1,5E-05	0,013	0,02	3,0E-07	0,019	0,06	8,3E-04	0,019	0,04	3,1E-06
W13	INGOBERNABLE	0,012	0,03	3,4E-06	0,013	0,03	9,7E-07	0,020	0,06	3,6E-04	0,017	0,05	1,8E-04
W14	MANSEDUMBRE	0,014	0,03	3,8E-06	0,014	0,03	5,3E-07	0,018	0,05	6,3E-04	0,017	0,04	1,1E-04
W15	ZAFARRANCHO	0,012	0,03	1,8E-06	0,013	0,03	1,4E-06	0,021	0,06	5,2E-04	0,015	0,04	1,7E-04
W16	ZARRAPASTROSA	0,010	0,02	1,0E-04	0,010	0,03	9,7E-05	0,016	0,06	6,4E-03	0,010	0,03	1,5E-03
Average over the 16 words		**0,012**	**0,03**	**5,0E-05**	**0,012**	**0,03**	**3,7E-05**	**0,018**	**0,06**	**1,3E-03**	**0,015**	**0,04**	**4,2E-04**

Table 5.4: Differences in relative mutual information between the inter-writer and intra-writer case for in-air trajectories. P-values shadowed green correspond to statistically significant differences (α=95%).

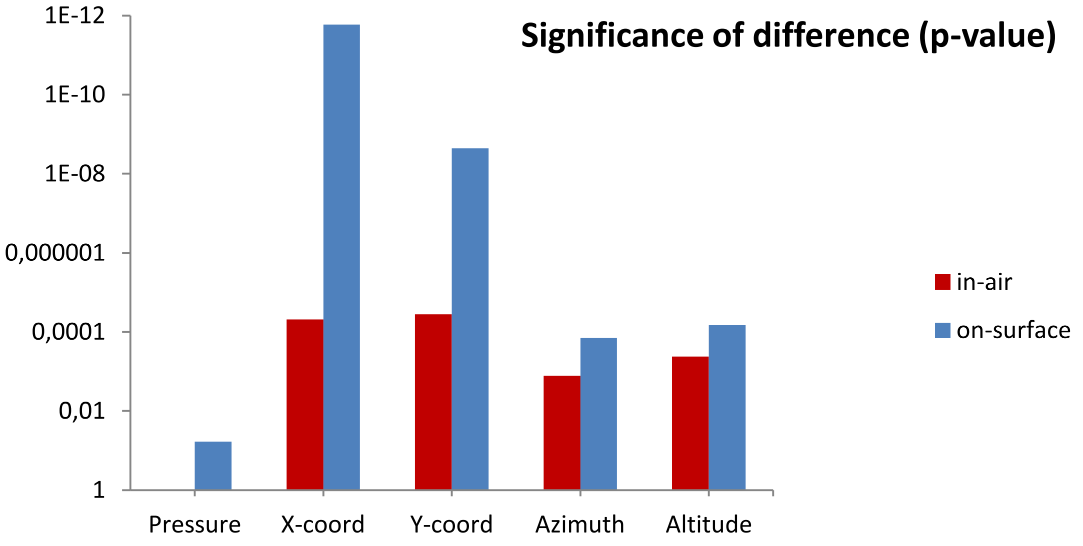


Figure 5.6: P-values for the differences in relative mutual information between inter-writer and intra-writer measures (average values among the words in tables 5.3 and 5.4).

Notice that, for both types of trajectories, with a significance level of α=*0.01* the null hypothesis would be rejected in all cases except for pressure. This means that from a purely statistical point of view, all features but pressure exhibit, in average, a significant difference between the intra-user and the inter-user case. When it comes to pressure, even if the average difference is positive for all words, the variability (standard deviation) is high enough to prevent a clear rejection of the null hypothesis (*p-values* range from 0.01 to 0.1).

5.4 CONCLUSIONS

The experimental results presented in the previous section support the claim that in-air trajectories contain as much information as on-surface trajectories. For four of the five features considered (all except pressure) the difference between the on-surface and the in-air case is lower than a single bit, usually some tenths of a bit (see Table 5.1). In fact, the only substantial difference between the two types of trajectories lies in the information provided by pressure, the fifth feature under consideration. When it comes to redundancy, the results show that although it is noticeable in the case of the X-coord and the Y-coord, it is low and very low in the case of azimuth and altitude respectively (see Table 5.2). From a global perspective it cannot be said that in-air and on-surface trajectories are entirely non-redundant. Nevertheless, although a certain amount of redundancy is present, it is far from seeming to be enough to deem the in-air trajectories as superfluous. Entropy and redundancy, the latter measured by mutual information and relative mutual information, show, for all the analyzed words and features, a low variability. This fact is important because it somehow implies that the obtained results are valid for a great majority of writers since they show a similar behaviour with respect to these measures.

When both aspects, amount of information in each type of trajectory and non-superfluousness of the in-air trajectories, are considered together there seems to be no need to discard the information contained in in-air trajectories, as it is often done in handwriting-based biometric recognition systems. What is more, it may be advisable to gather and process this information separately. Research results presented in chapter 7 will also give support to the notion that in-air trajectories are rich in information.

Regarding the biometric potential of both types of trajectories measured by the difference between the intra-user and the inter-user cases, the results are not conclusive. On the one hand, the differences are always positive and, except for pressure, statistically significant. On the other hand, these differences are very close to zero which may prevent their use as a score of the similitude between different executions of the same word. Nevertheless, this lack of conclusiveness does not imply that handwriting words do not possess a considerable biometric potential; it just means that the information theory and, more precisely the selected measures and the experiments performed, cannot prove the existence of this potential. Fortunately, the recognition schema presented in chapter, 6 along with the experimental results presented in chapter 7, does give evidence that words and short sequences of text can perform well in biometric recognition tasks.

6

A RECOGNITION SYSTEM BASED ON CATALOGUES OF PEN-UP AND PEN-DOWN STROKES AND DYNAMIC TIME WARPING

This chapter is entirely devoted to the recognition system that constitutes one of the major contributions of this dissertation. This system can perform identification and verification and, in both cases, can use strokes from in-air and/or on-surface trajectories. It is based on an innovative allographic approach that relies on catalogues of strokes built in an unsupervised manner by means of Self-Organizing Maps, and on Dynamic Time Warping to compare the sequences of strokes that constitute the sequences of online text. The first section presents a general overview of the recognition method. A much more in-depth view is given in the second and last section. All the experimental results regarding the recognition system will be reported in the next chapter.

6.1 OVERVIEW OF THE PROPOSED METHOD

The recognition approach proposed in this dissertation considers strokes as the structural units of handwriting. Any piece of text is regarded as two separate sequences, one of pen-down (on-surface) and one of pen-up (in-air) strokes. These sequences are obtained from the original sequence of alternated pen-down and pen-up strokes. Each stroke, as gathered by the acquisition device, is itself a sequence of multi-dimensional points.

The core of the proposed method consists of encoding each one of these sequences of strokes as a sequence of class-prototypes, where each class-prototype represents a whole class of strokes. As each class-prototype can be encoded as an integer, a very simple representation is achieved (Fig. 6.1).

Once encoded, sequences of strokes (now sequences of integers) can be compared by means of dynamic time warping (DTW). This comparison yields a measure of the dissimilarity between the compared sequences. The obtained measure of dissimilarity can then be used to perform recognition.

The encoding of strokes requires the existence of a catalogue of class-prototypes. In fact, such a catalogue is a set of classes, each class represented by its prototype. During the encoding stage, each stroke is encoded as the integer number representing the prototype of its class. Several clustering methods can be used to obtain set of classes that are somehow representative of the variability of the strokes found in different writers' executions of a piece of text [147]. Nevertheless, the proposed method demands that prototypes can be easily compared to each other since sequences of class-prototypes are going to be *DTWed* and DTW is itself a time-consuming algorithm. In order to minimize the comparison effort required during dissimilarity computation, a classification method that *preserved* the topology of the space of original strokes was opted for. This way, strokes that are closer to each other are represented by the same prototype or by prototypes that are themselves closer to each other and whose distance is easily computed. Self-organizing maps (SOMs) [148] provide classifications that meet this topology-preservation condition.

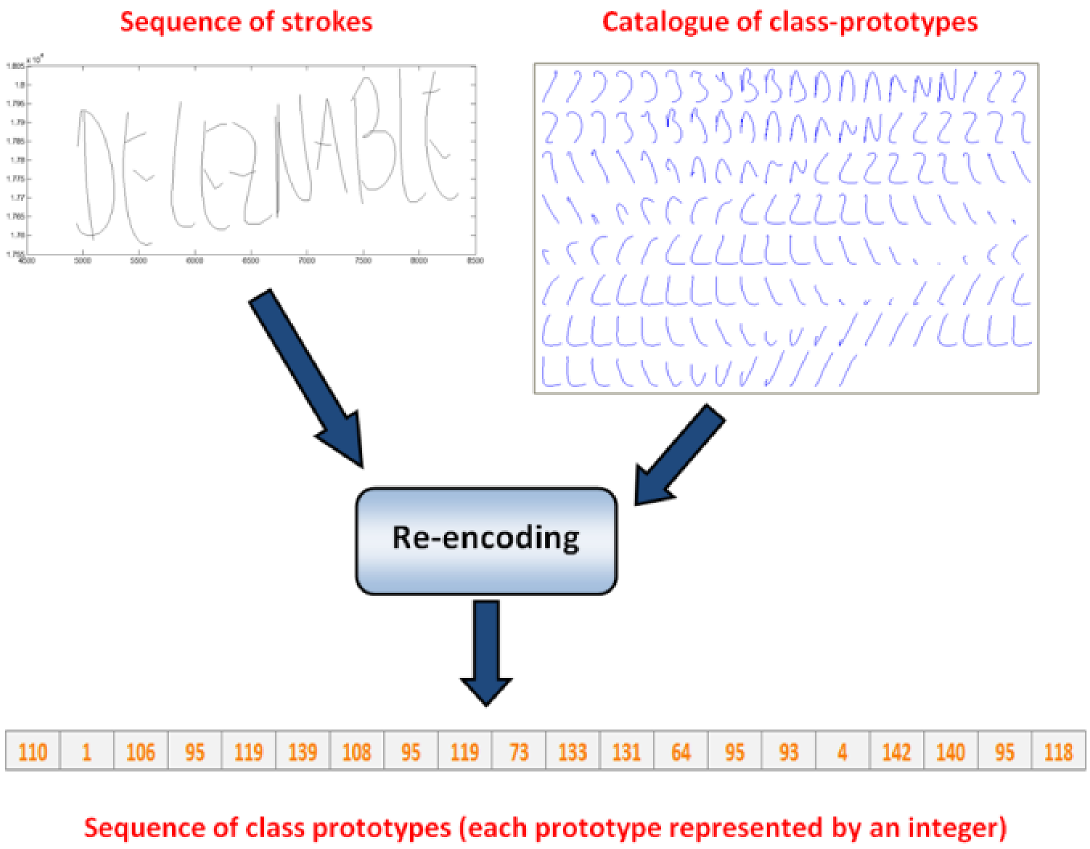


Figure 6.1: Graphical depiction of the re-encoding of sequences of strokes into sequences of prototypes represented by integers.

A number of users provide repetitions of the same word. Each word is decomposed into two sequences, one of pen-up and one of pen-down strokes. All sequences of pen-up strokes are shown to a SOM and as a result a catalogue of pen-up strokes is obtained. This catalogue is an indexed set of classes, each class represented by a prototype. An analogous procedure is followed to obtain the catalogue of pen-down strokes (Fig. 6.2).

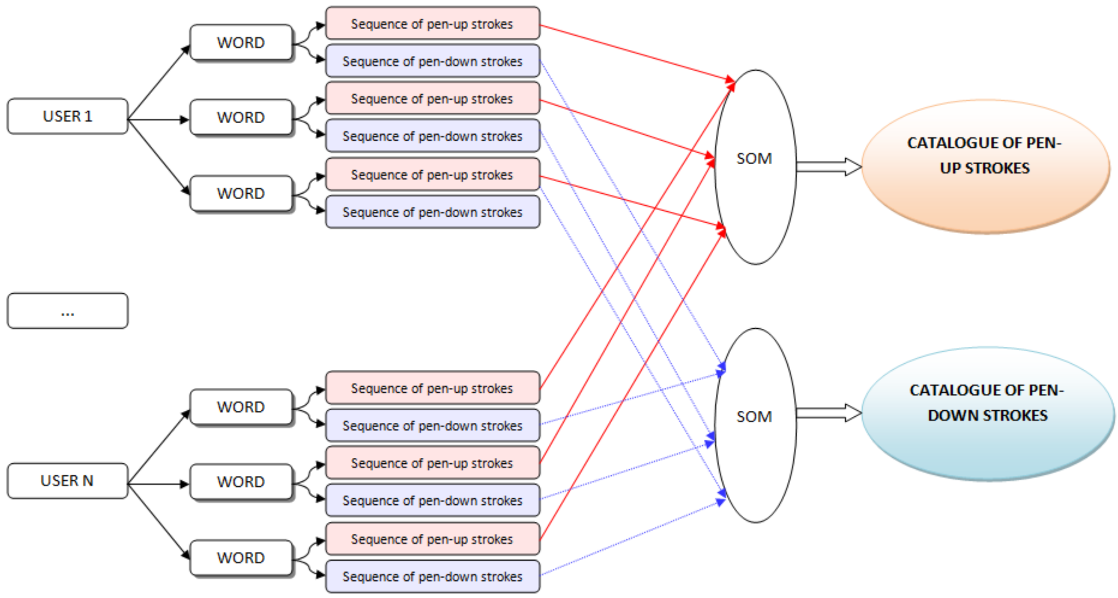


Figure 6.2: Schematic overview of the construction of the catalogues of strokes.

Once a pair of catalogues is available, each writer can be modelled as a set of encoded sequences of pen-up and pen-down strokes. During the enrolment stage, each user provides a number of repetitions of the same word. Each word is decomposed into two sequences of and the corresponding encoded sequences are obtained by substituting each stroke by the index of the nearest stroke in the catalogue (Fig. 6.1). A graphical depiction of the user modelling process is given in Fig. 6.3.

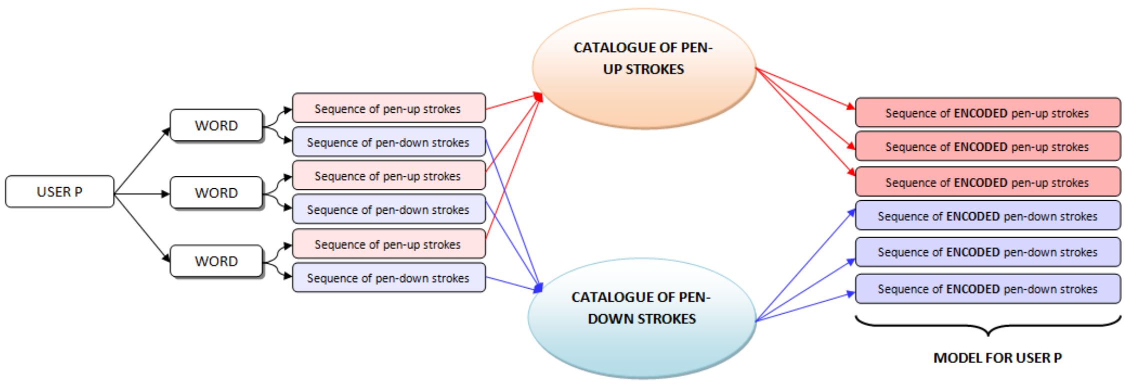


Figure 6.3: Schematic overview of a user's model construction.

During verification, the word the authorship of which is questioned is decomposed into the two sequences of its pen-down and pen-up strokes and each sequence is encoded using the same catalogues that were used to build the models. Then, each sequence is compared (DTW) against the sequences in the alleged user's model and each comparison yields a dissimilarity measure. Pen-up dissimilarity measures are combined into a single measure (w_{up}) and the same is done with the pen-down dissimilarity measures (w_{down}). These two measures are then combined into a single dissimilarity measure of the whole word (w_{word}). As a final step, a score (SC_{word}) is obtained from w_{word}. If this measure is lower than a predefined global threshold, the questioned word is deemed authentic; otherwise it is deemed false (Fig. 6.4). If only one type of stroke is used for verification, the score (SC_{up} or SC_{down}) is directly obtained from the corresponding dissimilarity measure. For identification, the comparison is made against all the known models and the authorship is granted to the user whose model produces the highest score (i.e. the lowest dissimilarity).

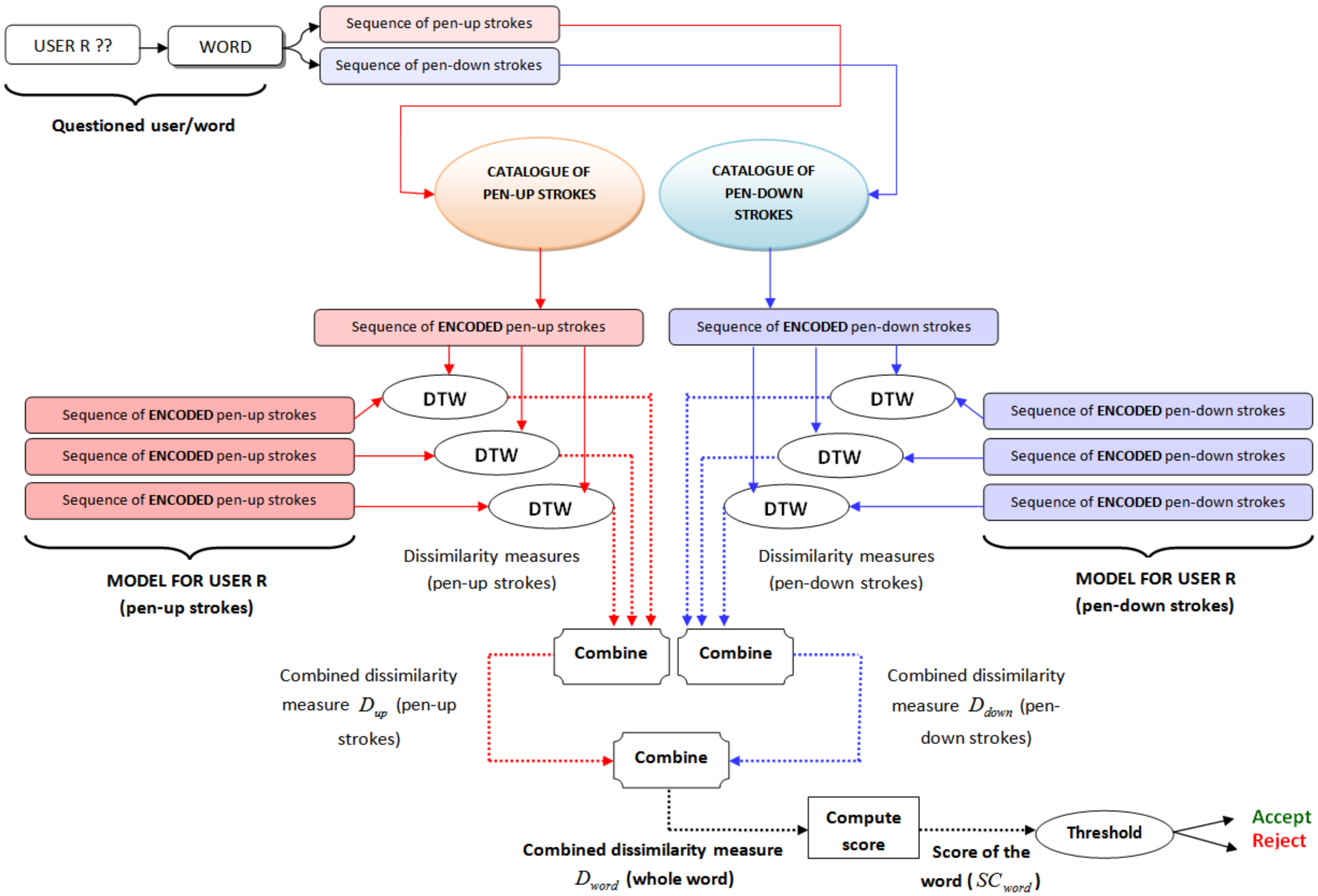


Figure 6.4: Schematic overview of the devised verification process.

6.2 DETAILED VIEW OF THE PROPOSED METHOD

6.2.1 SEGMENTATION AND PRE-PROCESSING OF STROKES

The proposed method relies in data that, prior to any pre-processing adheres to the SVC format [69]. Each execution of a word is given as seven time-sequences (vectors) that hereafter will be called features: *x(t)*, the *x* coordinate; *y(t)*, the y coordinate; *ts(t)*, a time stamp value; *bs(t)*, the button status value (0 for pen-up, 1 for pen-down); *az(t)*, the azimuth; *al(t)*, the altitude and *pr(t)* the pressure. All features have the same length, varying from execution to execution. Thus, the execution of a word can be formally described as a matrix $[x(t), y(t), ts(t), bs(t), az(t), al(t), pr(t)]$ with $t \in [1, N]$ where N is the length (number of sampling units) of the execution.

Segmentation into strokes is straightforwardly achieved thanks to the *bs(t)* feature (*pr(t)* could also have been used). A pen-down stroke starts at a point where *bs(t)* changes from 0 to 1 and ends at a point where *bs(t)* changes from 1 to 0. More formally, a pen-down stroke starts at t_s and ends at t_e ($t_e > t_s$) if $bs(t) = 1, \forall t \in [t_s, t_e]$ and $bs(t_{s-1}) = 0$ (or $t_s = 1$) and $b(t_{e+1}) = 0$ (or $t_e = N$). Pen-up strokes can be characterised analogously.

Thus, a non-pre-processed pen-down stroke can be described as a matrix $[x(t), y(t), ts(t), ts(t), az(t), al(t), pr(t)]$ where $bs(t) = 1\, \forall t$ and whose starting and end-points follow the definition given above. A non-pre-processed pen-up stroke can be described in an analogous manner.

After segmentation, two sequences are obtained from each word: one of pen-down strokes and another one of pen-up strokes. If there is a pen-up stroke at the beginning of the word, then this pen-up stroke is discarded.

Stroke pre-processing has deliberately been kept to a minimum (the impact of a more elaborated pre-processing in the overall performance of the method is left as a possible future line of research). However, some stroke-level pre-processing is required in order to extract and weight the desired features and to equalize the lengths of all strokes, so that they can be presented to a SOM.

After segmentation, each stroke undergoes the following steps:

1. Non-selected features (time-stamp and button-status) are removed

2. Data is resampled to be accommodated in a fixed number of points, equal for all strokes. For long strokes this means downsampling whereas it means upsampling for shorter ones. Resampling is performed by nearest neighbour interpolation and no filter is applied to the data being resampled. The actual resampling procedure is as follows: a vector $V = (v_1, \ldots, v_q)$ of length *q* is resampled into a new vector $VR = (vr_1, \ldots, vr_p)$ of length *p*, with vr_k $(1 \leq k \leq p)$ being:

$$vr_k = \begin{cases} v_{k'} \; if \;\; k' \leq q \\ vr_{k-1}, otherwise \end{cases} \tag{6.1}$$

where $k' = round((k-1)\cdot\frac{q}{p})+1$

After resampling, a pen-down stroke can be formally described as a matrix where each column has been resampled. That is: $[xR(t), yR(t), azR(t), alR(t), prR(t)]$. Similarly, a resampled pen-up stoke can be described as $[xR(t), yR(t), azR(t), alR(t)]$.

The resampling step is required because strokes presented to a SOM must all have the same size (i.e. the same number of points). The actual resampling size (*p*) has been set to the nearest integer greater than or equal to the average size of the strokes to be presented to the SOM.

3. All the selected features: *x(t)*, *y(t)*, *az(t)*, *al(t)* and *pr(t)* are normalized to mean 0 and standard deviation 1.

4. Features are weighted. Some features are more relevant with respect to discrimination than others and thus are weighted accordingly. Our preliminary experiments (conducted with the purpose of having a first assessment of the proposed method) suggested that the relevance of a feature depends on the type of the stroke. For instance, *y(t)* appears to be more relevant than *x(t)* in pen-down strokes but the other way round in pen-up strokes. Furthermore, *al(t)* and *az(t)* seem to be slightly more relevant in pen-down strokes than in pen-up ones.
The actual relative weights of the features were experimentally set to the following values:

- For pen down-strokes, $w_{dx} = 0.27$, $w_{dy} = 0.35$, $w_{dpre} = 0.20$, $w_{daz} = 0.09$, $w_{dal} = 0.09$, for the x coordinate, y coordinate, pressure, azimuth and inclination respectively.
- For pen up-strokes, $w_{ux} = 0.50$, $w_{uy} = 0.37$, $w_{uaz} = 0.05$, $w_{ual} = 0.08$, for the x coordinate, y coordinate , azimuth and inclination respectively.

These values were obtained using a simple brute-force search approach the goal of which was to maximize the identification rate for the word *BIODEGRADABLE* (values for pen-up and pen-down strokes were searched separately). Later on, the same values were tested for the rest of words, introducing slight variations. In all cases, the modified values yielded worst identification rates. Eventually it was decided to apply the same weights to all words.

A fully pre-processed pen-down stroke can be described as $[w_{dx}\cdot x\overline{R}(t),\;\; w_{dy}\cdot y\overline{R}(t),\;\; w_{daz}\cdot az\overline{R}(t),\;\; w_{dal}\cdot al\overline{R}(t),\;\; w_{dpr}\cdot pr\overline{R}(t)]$ and a fully pre-processed pen-up stroke as $[w_{ux}\cdot x\overline{R}(t),\;\; w_{uy}\cdot y\overline{R}(t),\;\; w_{uaz}\cdot az\overline{R}(t),\;\; w_{ual}\cdot al\overline{R}(t)]$ where $\overline{R}$ denotes normalization after resampling and $\cdot$ denotes scalar multiplication.

6.2.2 CONSTRUCTION OF THE CATALOGUES OF STROKES

A set of pre-processed strokes, coming from different users and sessions (but from the same word) is presented to a SOM in order to train it. The purpose of this step is to generate a set of classes (clusters) each one representing a whole set of strokes. After training, each unit in the map is a prototype of a class of strokes.

It is important to guarantee that the catalogues possess the property of *topology preservation* (continuity in the mapping) [149]. Topology preservation means that strokes that are close to each other are mapped close in the catalogues. This property can be measured using the *topographic error* (the proportion of all strokes for which first and second best-matching prototypes are not adjacent units).

In the main experiments, reported in the next chapter, the following settings have been used:

- 150 units in a sheet-shaped two-dimensional lattice with a hexagonal neighbourhood topology (Fig. 6.5). The final number of units may slightly vary because it is automatically optimized to improve the accommodation of the data presented to the SOM.
- The training is performed during 240 epochs: 40 rough-training epochs (high learning rate and high neighbourhood radius) and 200 fine-training epochs (lower learning rate and lower neighbourhood radius). The actual construction of the SOMs is left to the SOM Toolbox[150][151], a specialized Matlab [152] package.

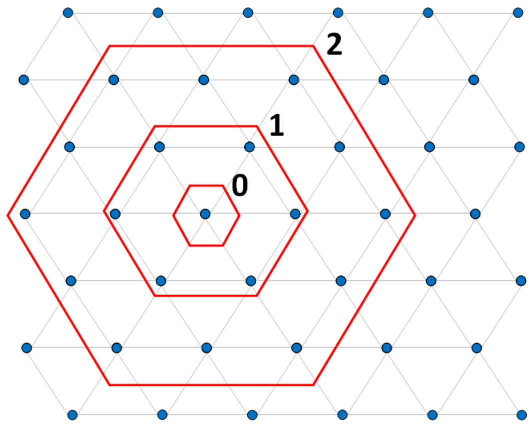


Figure 6.5: Hexagonal neighbourhoods in a two-dimensional lattice.

With the aforementioned settings a topographic error of less than 5% (usually about 3.5%) is achieved. This amount of topographic error has been considered acceptable and no further reduction has been pursued.

The construction of the catalogues of strokes has to be regarded as a one-time initialization step that could be critical for the performance of the recognition process. In a real environment, where users are enrolled as they come and therefore it is unfeasible to wait until all or most of them are known, the catalogue has to be built with just a portion of all the users the system will have and/or using data from writers that will not be enrolled. One of the remarkable properties of the proposed system is that its performance is not critically affected by the origin of the data used to train the SOM (users enrolled or not enrolled in the system).

Furthermore, the system exhibits a good performance even when the catalogues are built from data from few users (see section 7.2 for further details).

Figs. 6.6 and 6.7 show the prototypes in the catalogues built from strokes coming from the executions of words *DESBRIZNAR* and *ZAFARRANCHO*. Notice that in the pen-down catalogues some strokes resemble entire characters (mostly R, S, N, C and Z) while the rest are just parts of characters. In all the catalogues shown, some pairs of prototypes seem almost identical. This is so only apparently because even if they have similar XY trajectories, they may differ in pressure and/or writing angles (Figs. 6.8 and 6.9).

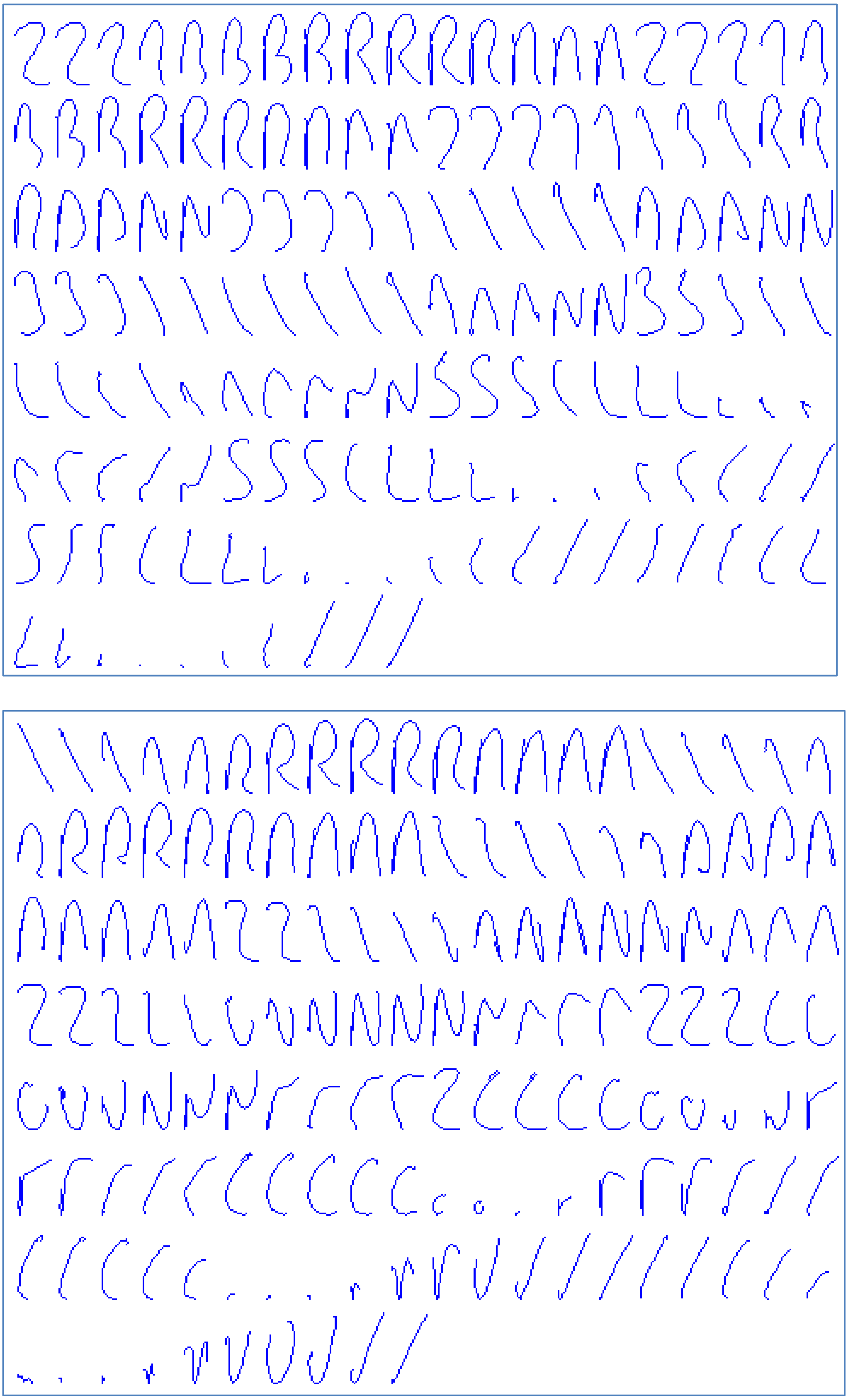


Figure 6.6: Catalogues built out of pen-down strokes from words DESBRIZNAR (up) and ZAFARRANCHO (down).

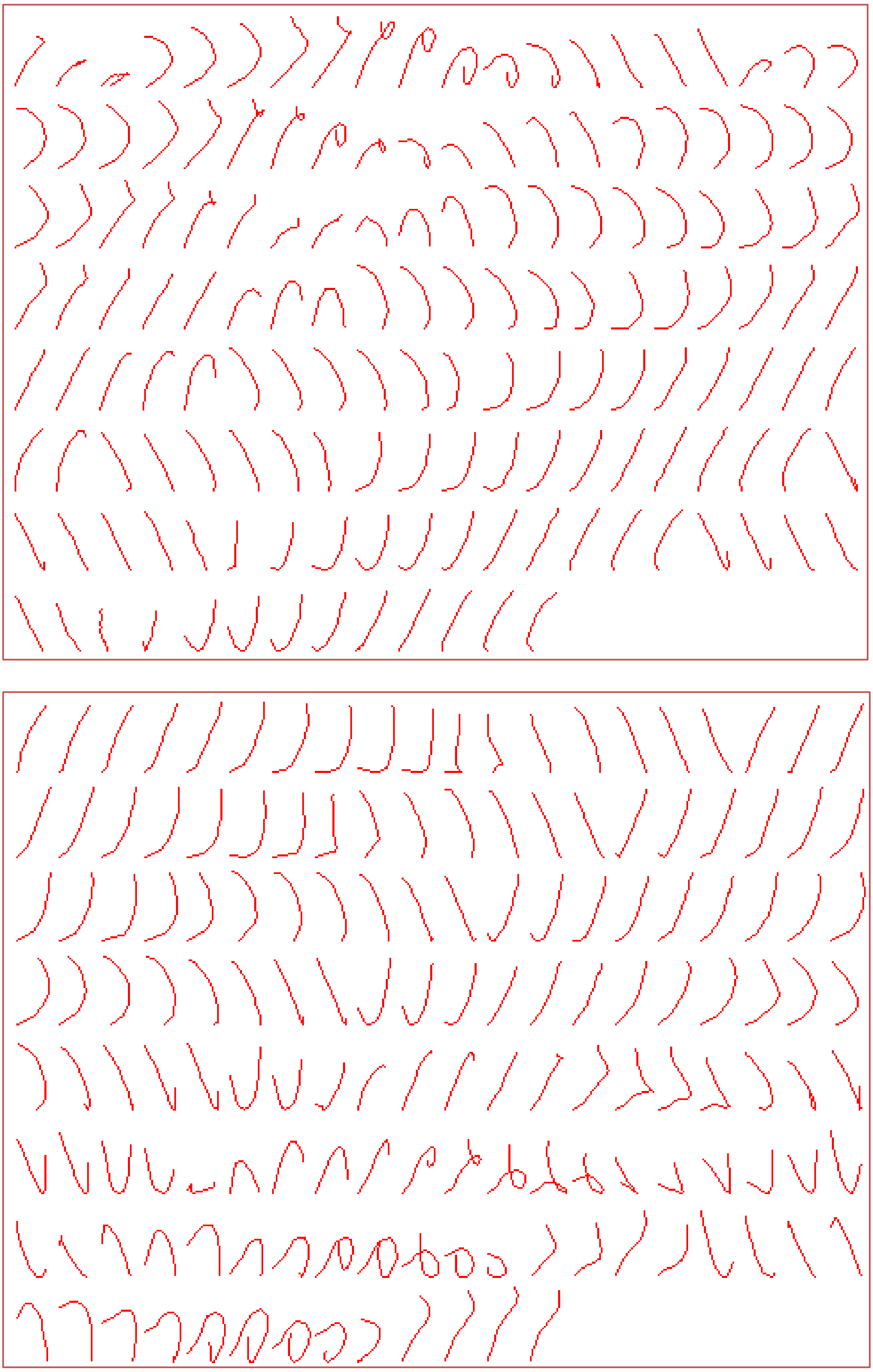


Figure 6.7: Catalogues built out of pen-up strokes from words DESBRIZNAR (up) and ZAFARRANCHO (down).

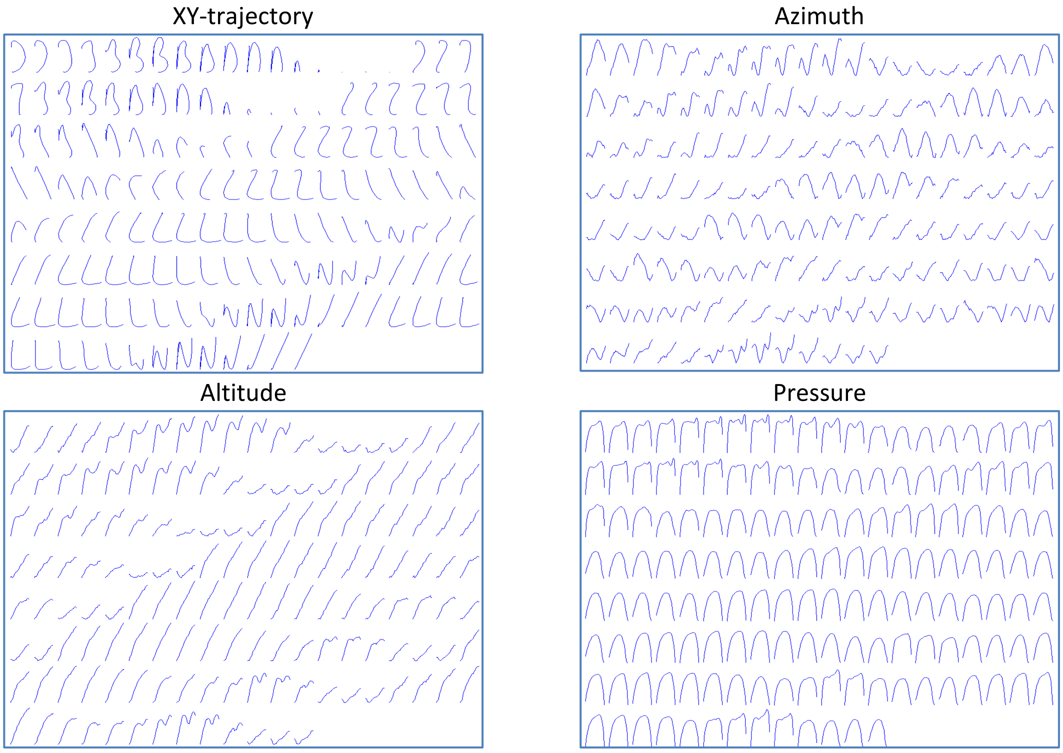


Figure 6.8: Different projections of the catalogue of pen-down strokes built from executions of the word DELEZNABLE (X vs. Y, top-left; Azimuth vs. time, top-right; altitude vs. time, bottom-left and pressure vs. time, bottom-right).

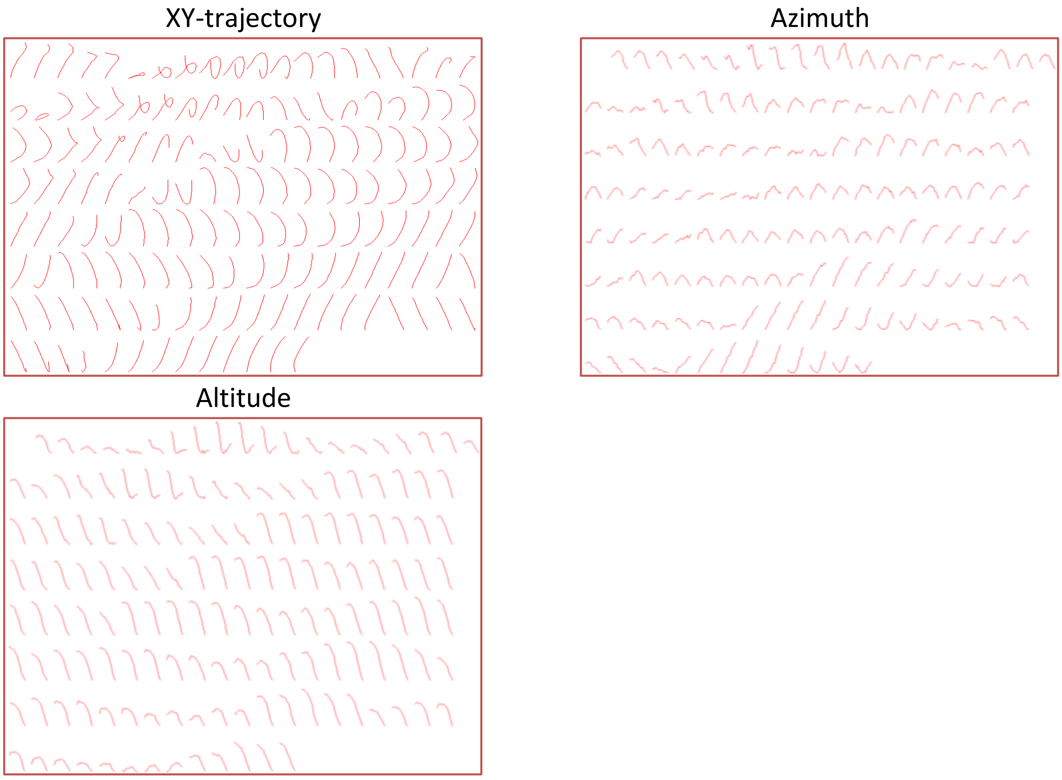


Figure 6.9: Different projections of the catalogue of pen-up strokes built from executions of the word DELEZNABLE (X vs. Y, top-left; Azimuth vs. time, top-right and altitude vs. time, bottom-left).

6.2.3 ENCODING OF STROKES

As each unit in the catalogue is identified by a positive integer (its index) each stroke can be represented by the identifier of the unit whose prototype is nearest to the stroke. That unit is called the Best Matching Unit (BMU) of the stroke. Strokes in each sequence (pen-up and pen-down) are encoded using the corresponding catalogue built in the previous phase.

So, when considering strokes of a given type (up or down), a sequence

$$S = s_1, \ldots, s_n \qquad (6.2)$$

where each s_i is a pre-processed stroke of that type , will be encoded as

$$SE = BMU(s_1), \ldots, BMU(s_n) \qquad (6.3)$$

being *BMU* the function returning the index (in the SOM) of the best matching unit (*BMUIdx*) for the given stroke. Eventually

$$SE = BMUIdx_1, \ldots, BMUIdx_n \qquad (6.4)$$

where $BMUIdx_i$ is the index of best matching unit for the *i-th* stroke in the sequence.

6.2.4 DTW OVER SEQUENCES OF ENCODED STROKES

The dissimilarity between a pair of sequences $SE_1 = BMUIdx_1^1, \ldots, BMUIdx_m^1$ and $SE_2 = BMUIdx_1^2, \ldots, BMUIdx_n^2$ where $BMUIdx_i^j$ is the best matching unit for the *i*-th stroke in the *j*-th sequence is computed by means of straightforward implementation of the standard DTW algorithm [126] (see Fig. 6.10).

Algorithm DTW
Input: encoded sequences of strokes SE_1 of length *m* and SE_2 of length *n*
Output: *DISS*, the dissimilarity between the two input sequences

CM is a (m+1)x(n+1) matrix of real numbers with indexes ranging from (0,0) to (m,n);

Initialize *CM* with all positions equal to +∞, except *CM*(0,0) = 0;

```
for i in 1..m
        for j in 1..n
                c := COST(BMUIdx_i^1, BMUIdx_j^2);
                if CM(i-1,j) < CM(i,j-1) and CM(i-1,j) < CM(i-1,j-1)
                        c := c + CM(i-1,j);
                else if CM(i,j-1) < CM(i-1,j) and CM(i,j-1) < CM(i-1,j-1)
                        c := c + CM(i,j-1);
                else
                        c := c + CM(i-1,j-1);
                end of if
                CM(i,j) := c;
        end of inner for
end of outer for

DISS = CM(m,n) / (m+n);
```

Figure 6.10: DTW algorithm to compute the dissimilarity between a pair of sequences of strokes.

The function *COST(*$Index_1$*,* $Index_2$*)* computes a measure of the *dissimilarity* between two (indexes of) stroke prototypes and is based on the neighbouring properties of the SOM that materializes the catalogue of strokes. More specifically, this dissimilarity is based on the length, in number of units, of the shortest path between the two prototypes in the catalogue of strokes: *USP(Index1, Index2)*. *USP* is suitable for the purpose of computing a distance between strokes because SOMs exhibit topological preservation (see section 6.2.2): strokes close to each other tend to be clustered into neighbouring units. So, in most cases, the closer the best matching units of any two strokes, the closer the strokes themselves. Thanks to this property, the distance between two strokes can be represented by the distance of their best-matching units in the SOM and from this distance a measure of dissimilarity can be obtained.

Several different definitions of *COST* have been considered, among them:

$$COST(Index_1, Index_2) = USP(Index_1, Index_2) \quad (6.5)$$

$$COST(Index_1, Index_2) = \ln(USP(Index_1, Index_2) + 1) \quad (6.6)$$

$$COST(Index_1, Index_2) = e^{USP(Index_1, Index_2)} \quad (6.7)$$

The definition based on the natural logarithm yielded the best results and has been used in all the recognition experiments reported in this dissertation.

6.2.5 COMBINATION OF DISSIMILARITY MEASURES

During verification, each sequence of encoded pen-down (or pen-up) strokes in a model has to be compared with the corresponding sequence in the questioned word (see Fig. 6.4). Each comparison yields a dissimilarity measure. All these measures have to be combined into a single one. To this purpose, an opinion-level fusion with a fixed-rule strategy has been used [153]. In order to decide which function to use, the *maximum* (select the biggest dissimilarity), the *minimum* (select the lowest dissimilarity) and the *average* were tested. The best results, both in terms of identification rate and verification performance, were obtained when using the *minimum* function.

Once a single measure for each kind of stroke is obtained, these measures have to be combined in order to obtain a single word-level measure. This time an opinion-level fusion with a trained-rule strategy is applied [153].

The world-level dissimilarity measure (D_{word}) is the weighed summation of the dissimilarity measures for pen-down (D_{down}) and pen-up (D_{up}) strokes:

$$D_{word} = w_{down} \cdot D_{down} + (1 - w_{down}) \cdot D_{up} \qquad (6.8)$$

The weighting factor (w_{down}) was determined by means of a brute-force search in the interval [0,1]. The goal of this search was to maximize the identification rate for the word *BIODEGRADABLE*. The best value found was $w_{down} = 0.585$. This value has been used in all the experiments described in this dissertation.

6.2.6 FROM DISSIMILARITIES TO SCORES

Dissimilarity measures computed by the DTW algorithm or obtained by the combination of other dissimilarity measures are transformed into scores. A score *SC* is obtained from a dissimilarity measure *ds* by means of the following formula:

$$sc = \frac{1}{e^{\frac{1}{2} \cdot \frac{ds}{ds_{\max}}}} \qquad (6.9)$$

Where ds_{max} is the greatest value expected by a dissimilarity measure. The ultimate purpose of this transformation is to shift dissimilarities into a common range, mapping high dissimilarities into low scores and low dissimilarities into high scores.

Fig. 6.11 graphically depicts the the mapping of relative dissimilarities (values of $ds / ds_{\max}$) into scores.

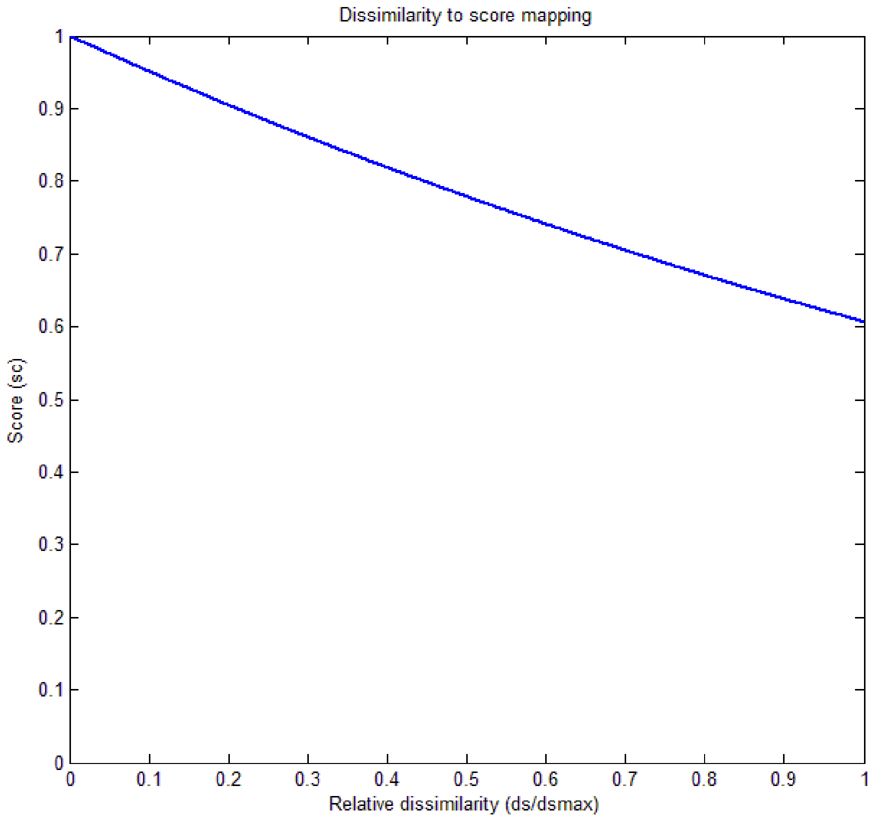


Figure 6.11: Mapping of relative dissimilarities into scores.

7

EXPERIMENTAL RESULTS

This chapter reports on the relevant experimental results yielded by the recognition system presented in the previous chapter, when it is applied to the sixteen uppercase words in the BiosecurID database. It is organized as follows: the first section describes the experimental process and presents the results that thereafter will be considered the baseline for subsequent comparisons. Sections two to five analyze the impact of the origin of the catalogues, of the number of writers that provide samples to build the catalogues, of the number of units in the catalogues and of the number of words used to recognize a writer, respectively. Finally, the main conclusions drawn from the results are summarized in the sixth section.

Several experiments have been devised to serve a double purpose: first to evaluate the performance of the proposed system and the impact of different settings. And second, to assess the effect of combining two or more different words, thus increasing the length of the text-sequences used for recognition.

All the experiments described in this dissertation have been carried out using data from the BiosecurID database (described in section 3.4.2) and, in particular, data of the 16 handwritten words. Although the number of writers that donated their handwriting was 400, a close inspection of the acquired data revealed that 30 of them did not follow the required conditions since their samples contain corrections, crossing-outs or more than one word in a line. These writers have been discarded and therefore the total number of writers available for experimentation is 370.

In order to facilitate the comparison of the results obtained, one of the experiments will be regarded as the reference experiment. The results of this experiment, fully described in the following section, will constitute the baseline on which to make comparisons.

7.1 REFERENCE EXPERIMENTAL PROCESS, REFERENCE EXPERIMENT AND REFERENCE RESULTS

The whole set of 370 writers has been divided into two disjoint subsets of 50 (the *catalogue subset*) and 320 (the *train-and-test subset*) writers.

During the *catalogue-building phase* all data in the *catalogue* subset is used to build the catalogues of pen-up and pen-down strokes. As the BiosecurID database provides 4 executions (obtained in different sessions) of each word, this means that the catalogues have been built with strokes obtained from 50x4=200 executions of each word (See Fig. 6.2 in chapter 6).

The data in the *train-and-test* subset is used in the *train-and-test phase* as follows: for each writer, 3 of the 4 executions of each word are used to build the model for that writer (thus a model contains 3 re-encoded sequences of pen-up strokes and 3 re-encoded sequences of pen-down strokes. See Fig. 6.3 in chapter 6). The fourth execution is used for testing (see Fig. 6.4 in chapter 6). For each type of stroke, each testing execution is compared against the 3x320=960 sequences in the models, yielding 320 dissimilarity measures (each model produces 3 measures, one per sequence, but only 1 is retained). Thus, the total number of dissimilarity measures obtained is 320^2 = 102400 (320 intra-writer measures and 102080 inter-writer measures). From these measures, the identification rate, the verification error and the equal error rate are computed.

The train-and-test phase is repeated four times. Each time a different session is used for testing. Fig. 7.1 provides a graphical description of the experimental process.

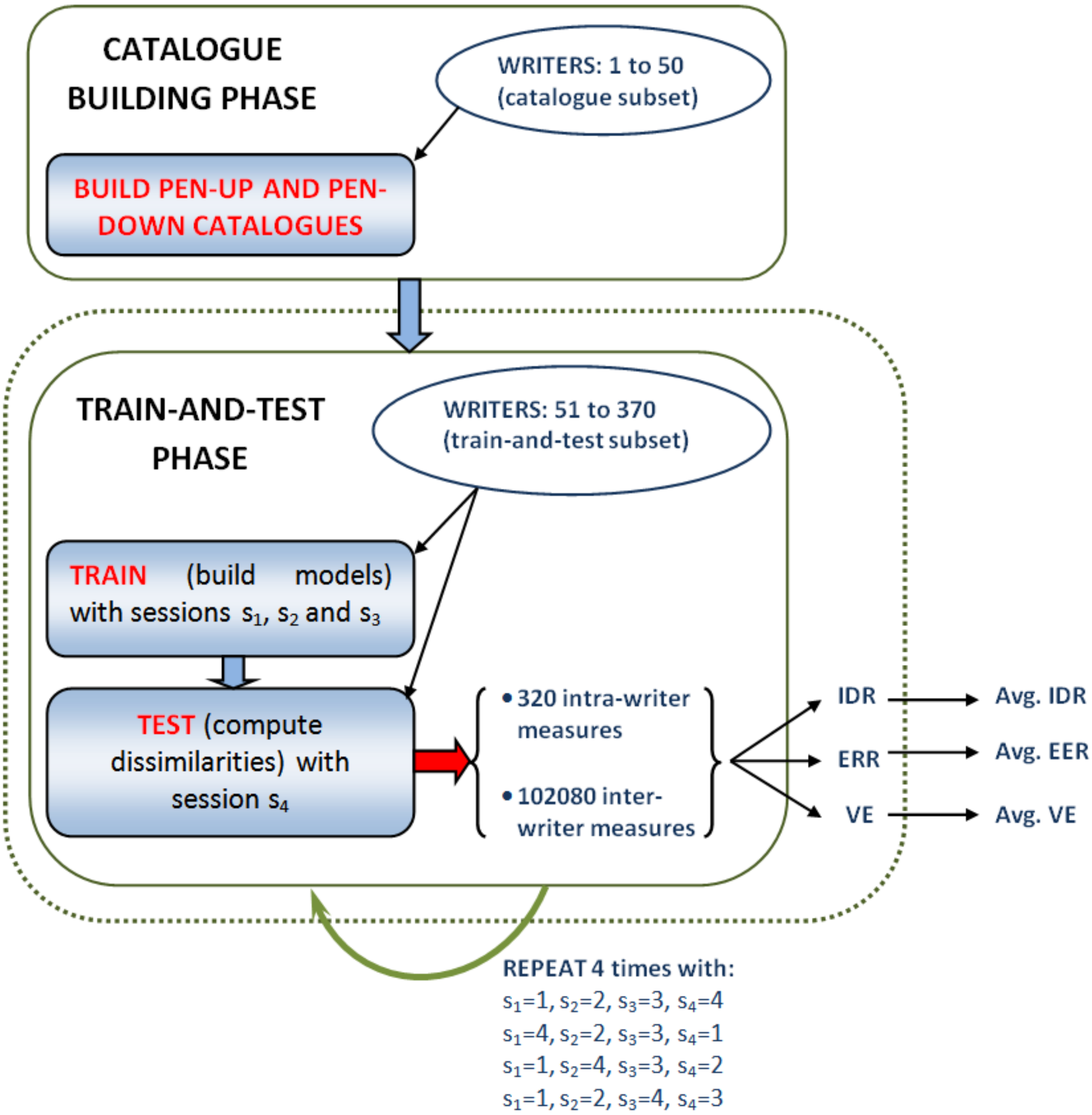


Figure 7.1: Graphical depiction of the experimental process that led to the results shown in Table 7.1.

Table 7.1 summarizes the results obtained in the four repetitions of the test-and-train phase. Values shown are the averages. Standard deviations are given in Table 7.2. Figs. 7.2, 7.3 and 7.4 give a graphical representation of data in Table 7.1. Figs. 7.5 and 7.6 show the DET-curves for the best and the worst performing words.

The following facts are worth noticing:

1. Performance of pen-up and pen-down strokes is quite similar. In some cases pen-up strokes slightly outperform pen-down ones (e.g. word DESPRENDER). The information analysis reported in chapter 5 already showed that, except from pressure, the amount of information in each type of trajectory was quite similar
2. The measures obtained when combining information (dissimilarities) from the two types of strokes always outperform the measures obtained prior to the combination. The information analysis also suggested that the two types of trajectories contained a certain amount of non-redundant information.

WORD	TEXT	LENGTH	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
			Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER
W1	BIODEGRADABLE	12	80,0%	4,35%	4,72%	81,4%	4,29%	4,62%	93,0%	2,67%	2,88%
W2	DELEZNABLE	10	72,1%	5,89%	6,23%	73,8%	4,48%	4,69%	88,4%	2,98%	3,29%
W3	DESAPROVECHAMIENTO	18	89,8%	2,98%	3,28%	89,3%	3,62%	3,91%	96,4%	2,05%	2,50%
W4	DESBRIZNAR	10	74,3%	5,10%	5,61%	71,6%	5,77%	5,99%	86,6%	3,59%	3,75%
W5	DESLUMBRAMIENTO	15	83,9%	3,82%	4,12%	84,2%	3,93%	4,22%	94,7%	2,49%	2,74%
W6	DESPEDAZAMIENTO	15	89,1%	3,39%	3,60%	84,4%	4,47%	4,62%	94,5%	2,84%	3,00%
W7	DESPRENDER	10	71,3%	5,55%	5,69%	67,7%	6,58%	6,88%	84,2%	4,26%	4,59%
W8	ENGUALDRAPAR	12	72,0%	5,76%	6,09%	71,2%	6,33%	6,86%	86,5%	4,18%	4,51%
W9	EXPRESIVIDAD	12	75,5%	5,05%	5,69%	79,3%	4,00%	4,16%	91,9%	2,50%	2,81%
W10	IMPENETRABLE	12	78,4%	4,90%	5,22%	73,1%	5,55%	5,98%	88,5%	3,25%	3,51%
W11	INEXPUGNABLE	12	82,3%	4,05%	4,29%	83,4%	4,18%	4,53%	94,5%	2,45%	2,58%
W12	INFATIGABLE	11	79,2%	4,38%	4,60%	79,5%	4,04%	4,29%	90,9%	2,24%	2,42%
W13	INGOBERNABLE	12	74,8%	5,56%	5,83%	81,3%	5,03%	5,25%	90,9%	3,66%	3,93%
W14	MANSEDUMBRE	11	63,4%	6,81%	7,09%	68,6%	7,87%	8,28%	84,1%	5,60%	5,87%
W15	ZAFARRANCHO	11	63,1%	7,40%	7,58%	65,2%	7,13%	7,42%	81,9%	5,26%	5,69%
W16	ZARRAPASTROSA	13	61,9%	8,22%	8,59%	67,3%	6,81%	7,20%	81,9%	5,37%	5,57%
Average over the 16 words			**75,7%**	**5,20%**	**5,52%**	**76,3%**	**5,25%**	**5,56%**	**89,3%**	**3,46%**	**3,73%**

Table 7.1: Results obtained in the reference experiment.

WORD	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	Std. dev. IDR	Std. dev. VE	Std. dev. EER	Std. dev. IDR	Std. dev. VE	Std. dev. EER	Std. dev. IDR	Std. dev. VE	Std. dev. EER
W1	3,07%	0,35%	0,28%	2,80%	0,23%	0,29%	1,97%	0,25%	0,30%
W2	2,74%	1,19%	1,08%	2,38%	0,70%	0,67%	2,52%	0,29%	0,19%
W3	1,41%	0,27%	0,60%	1,18%	0,37%	0,60%	0,74%	0,46%	0,45%
W4	1,70%	0,20%	0,26%	3,21%	0,65%	0,60%	1,60%	0,33%	0,36%
W5	1,48%	0,28%	0,17%	2,85%	0,29%	0,40%	1,05%	0,45%	0,58%
W6	1,99%	0,74%	0,80%	2,18%	0,49%	0,45%	1,04%	0,70%	0,75%
W7	1,12%	0,32%	0,26%	4,25%	0,75%	0,89%	1,72%	0,18%	0,32%
W8	0,97%	0,82%	0,96%	2,64%	0,32%	0,46%	1,56%	0,38%	0,52%
W9	2,61%	0,90%	1,24%	0,64%	0,57%	0,55%	1,30%	0,90%	1,17%
W10	2,60%	0,82%	0,84%	3,77%	0,73%	0,89%	3,18%	0,67%	0,69%
W11	1,95%	0,44%	0,58%	2,59%	0,37%	0,65%	1,47%	0,22%	0,30%
W12	1,16%	0,18%	0,12%	2,41%	0,15%	0,12%	1,06%	0,24%	0,29%
W13	1,39%	0,29%	0,31%	2,08%	0,35%	0,53%	1,36%	0,22%	0,39%
W14	2,34%	0,53%	0,64%	3,48%	0,52%	0,28%	1,81%	0,36%	0,17%
W15	2,83%	0,37%	0,39%	3,36%	0,39%	0,47%	1,28%	0,24%	0,29%
W16	3,07%	0,84%	0,79%	0,69%	0,40%	0,64%	1,37%	0,47%	0,50%

Table 7.2: Standard deviations corresponding to the results shown in table 7.1

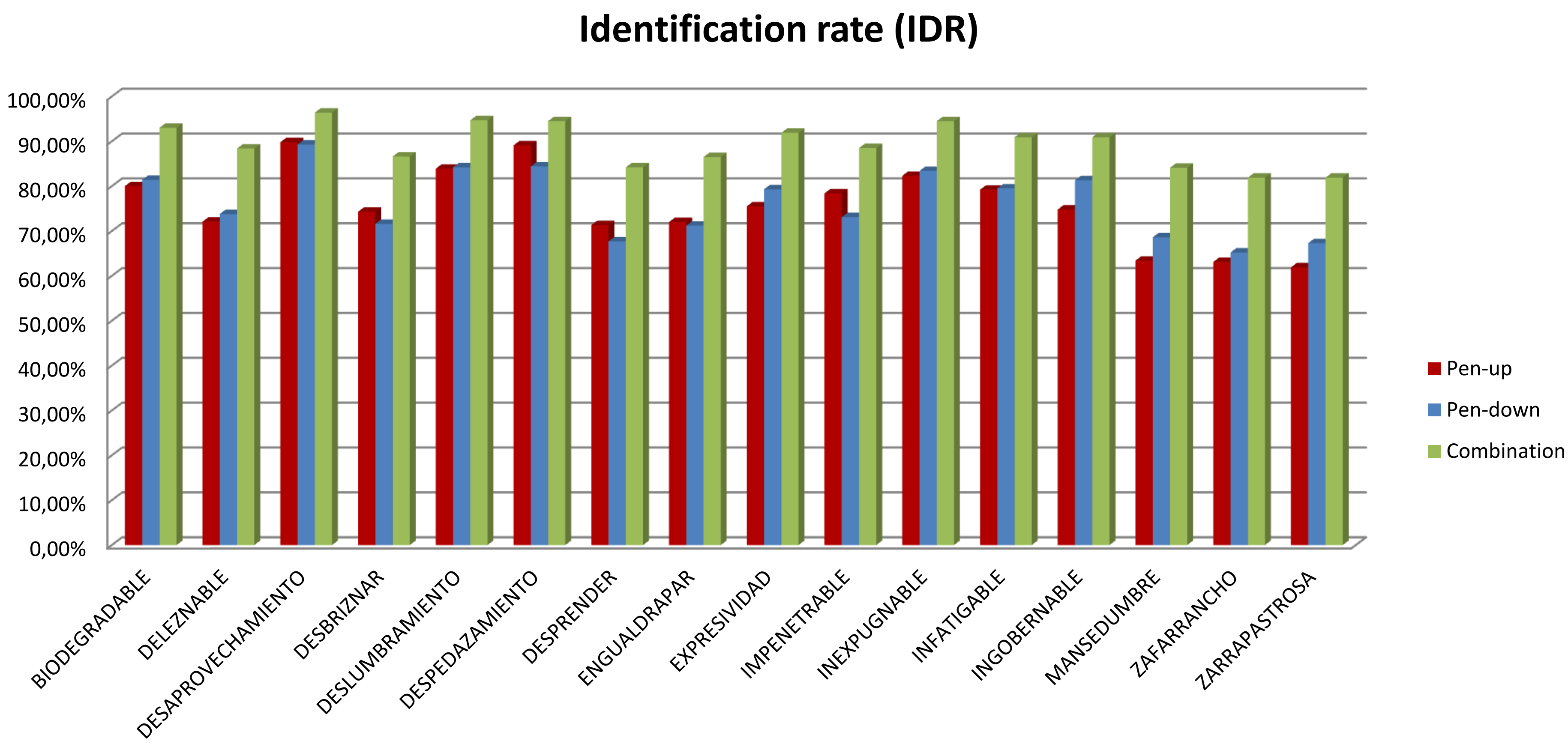


Figure 7.2: Identification rates (IDRs) obtained in the reference experiment.

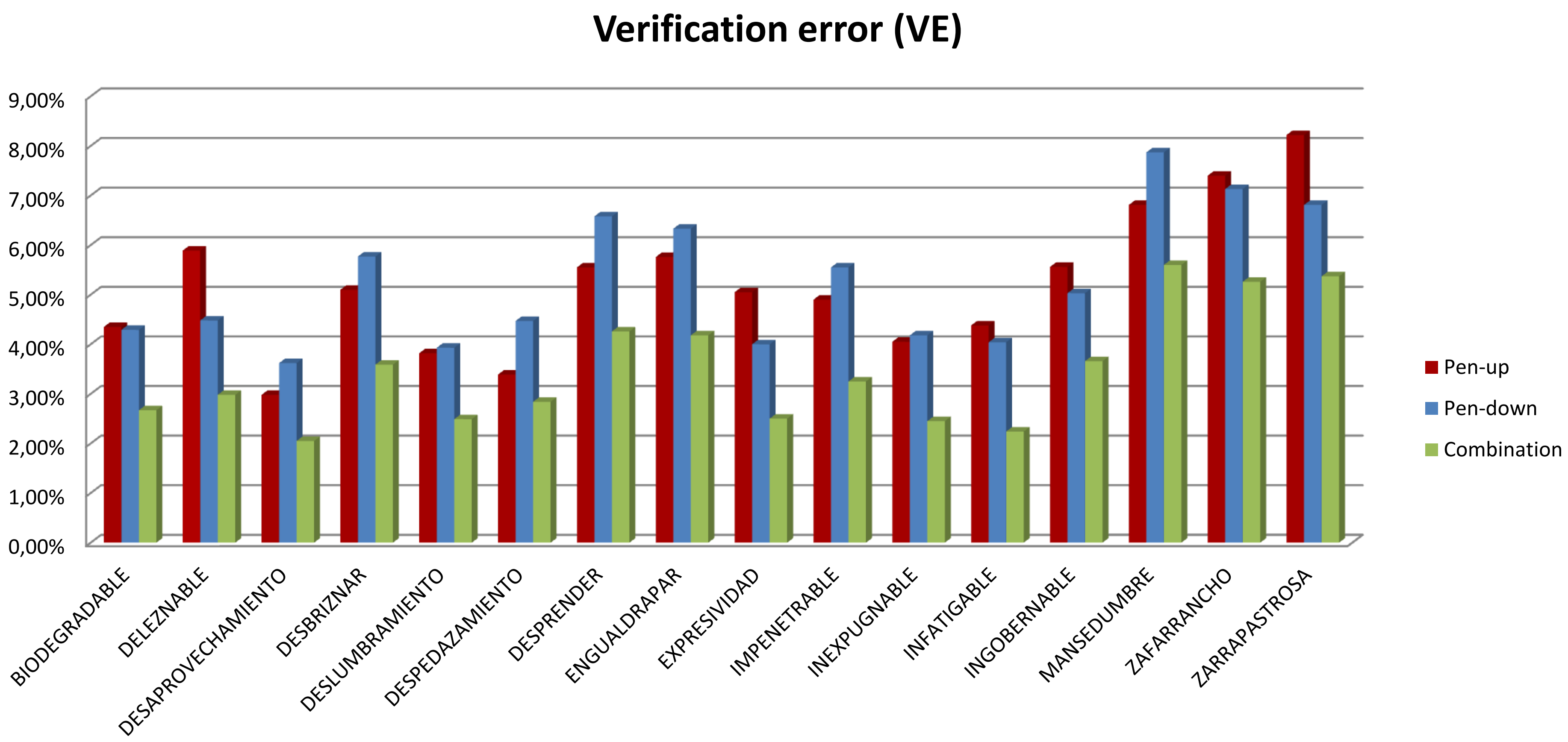


Figure 7.3: Verification Errors (VEs) obtained in the reference experiment.

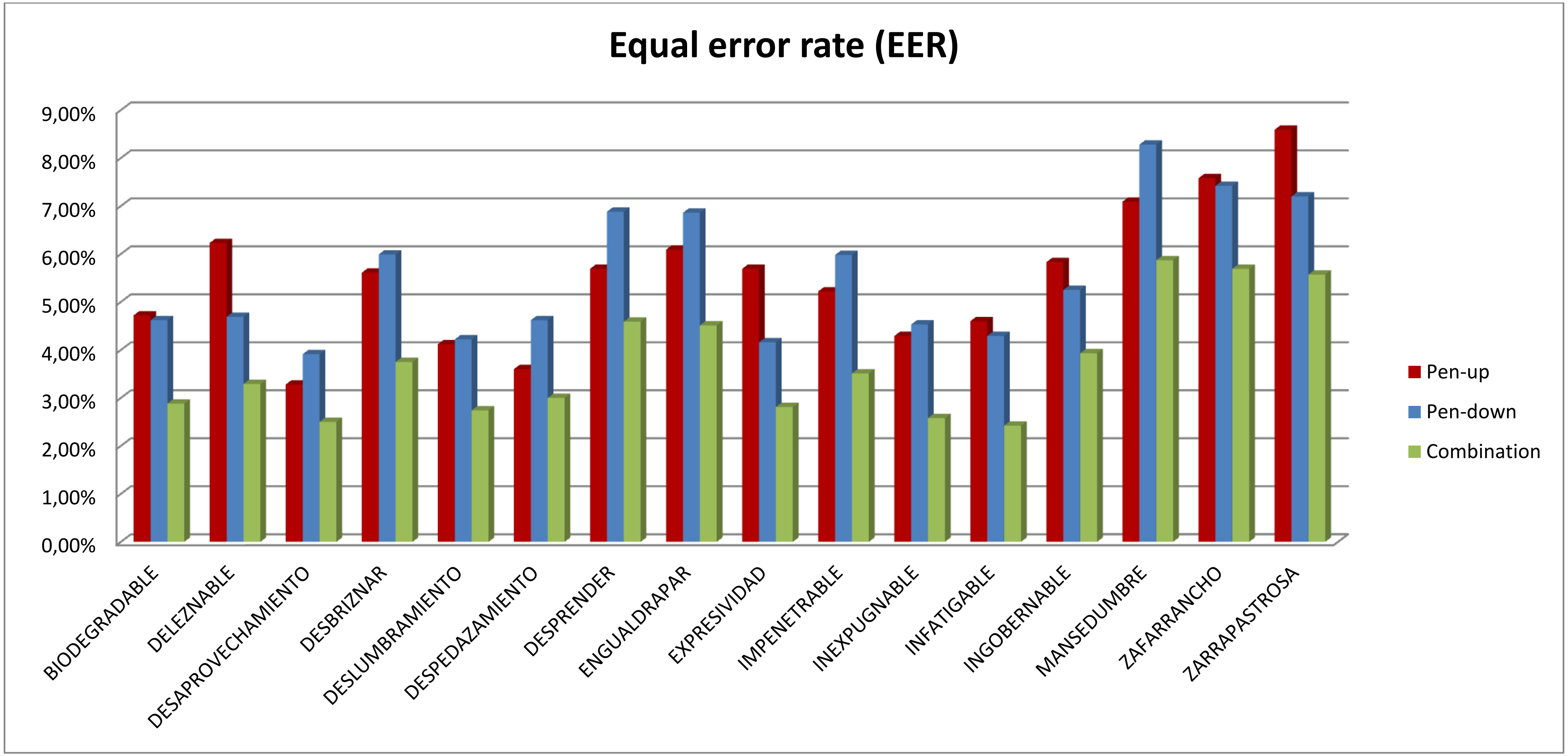


Figure 7.4: Equal Error Rates (EERs) obtained in the reference experiment.

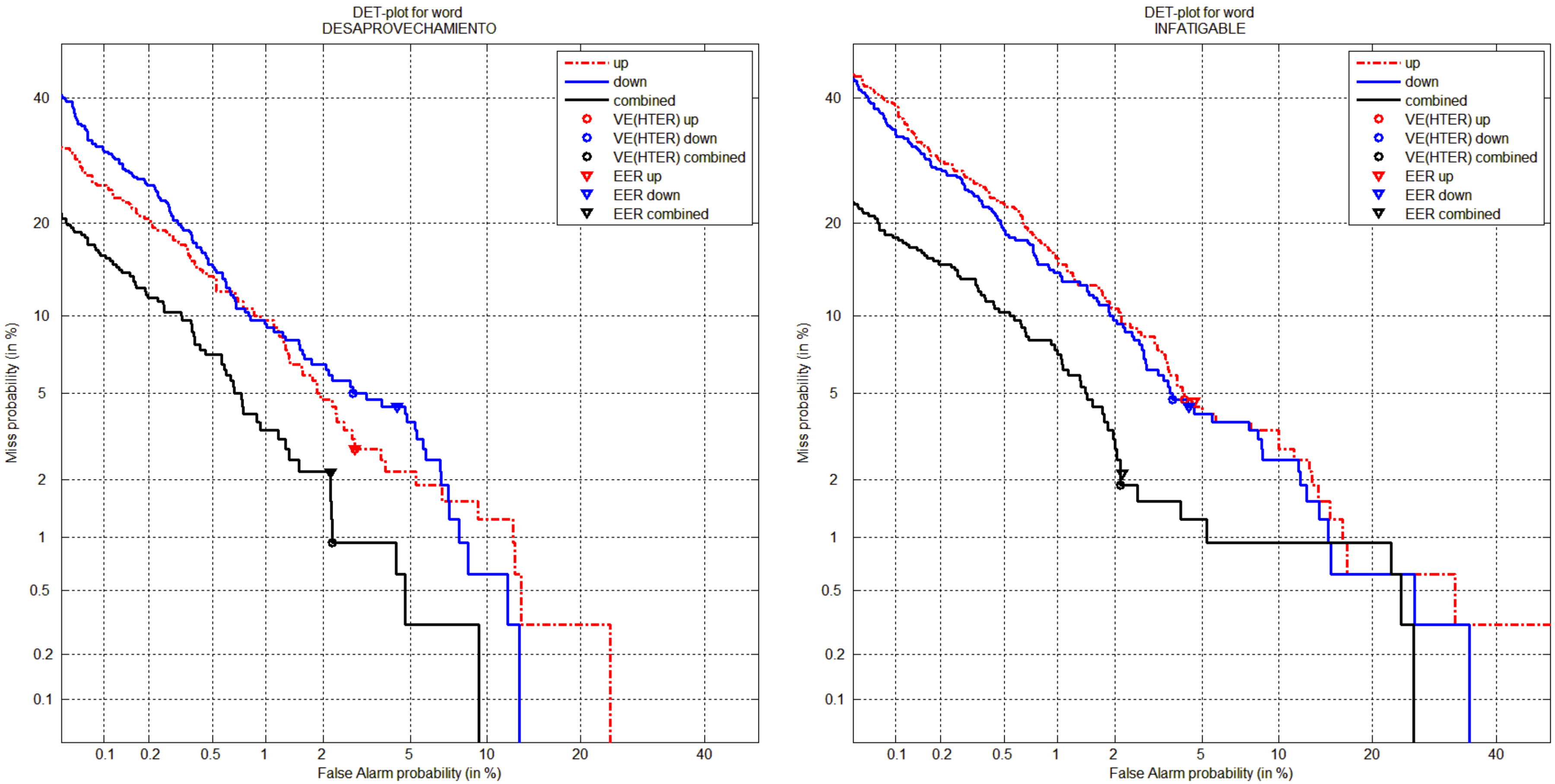


Figure 7.5: DET-plots for the best-performing words in terms of verification error (DESAPROVECHAMIENTO mean VE=2.05%) and equal error rate (INFATIGABLE, mean EER=2.42%).

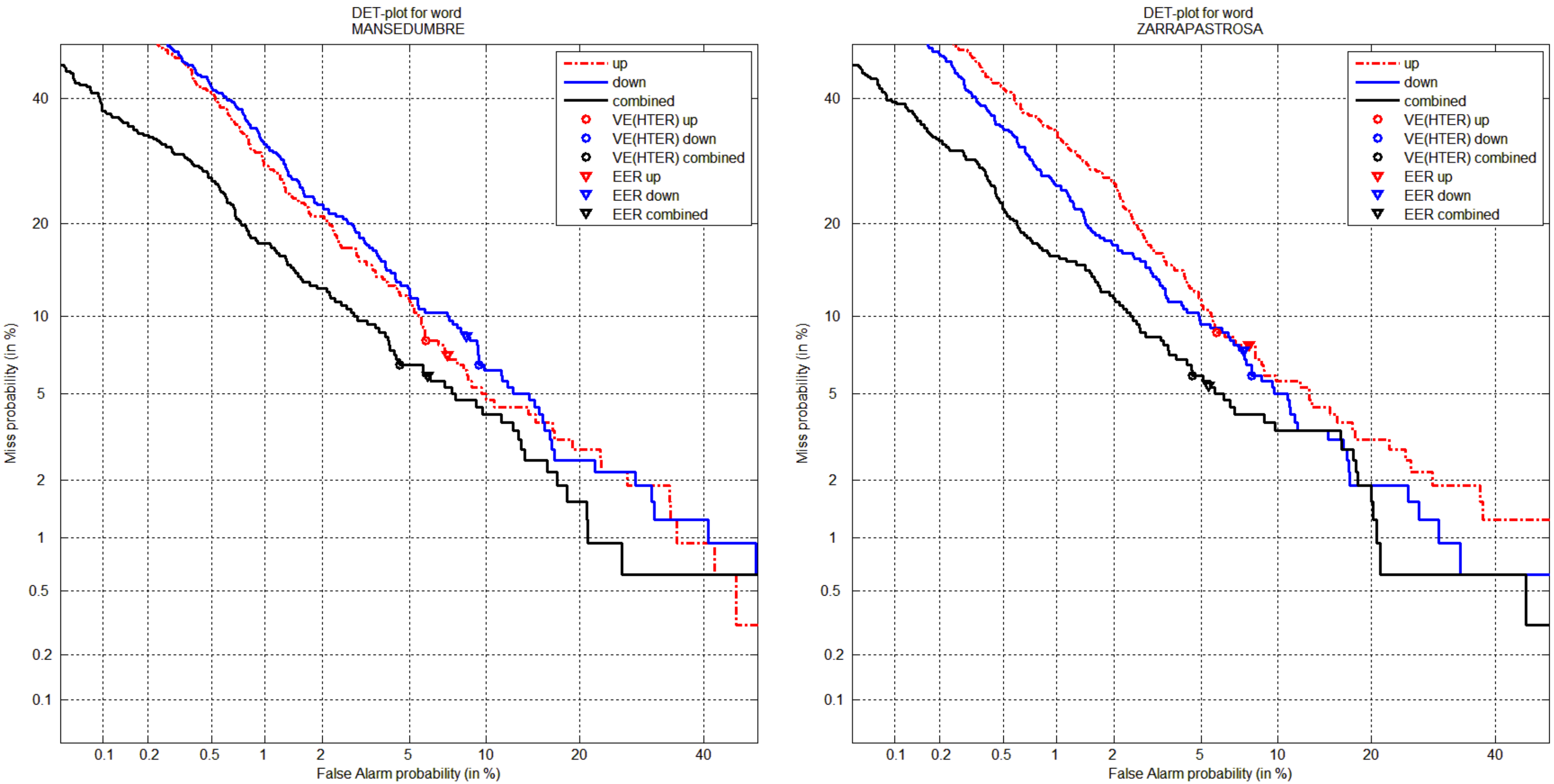


Figure 7.6: DET-plots for the two worst-performing words: MANSEDUMBRE (mean VE=5.60%, mean EER=5.87%) and ZARRAPASTROSA, (mean VE=5.37%, mean EER=5.57%).

7.2 IMPACT OF THE ORIGIN OF THE CATALOGUES. ENDOCATALOGUES AND EXOCATALOGUES

As it has already been stated, the catalogues of strokes are a critical part of the proposed schema. Therefore, it is important to get some knowledge on the impact that some decisions regarding their construction may have on the overall performance of the system. One such decision is the origin of the strokes used to train the SOMs. These strokes could come from the same writers that, later on, would be enrolled, or they could come from different writers that would never be enrolled in the system. The former type of catalogues will be referred as *endocatalogues* while the latter will be referred as *exocatalogues*. Endocatalogues imply that prior to the enrolment phase a certain amount of users is already known. This situation may be, in most practical cases, quite unrealistic since users are enrolled *as they come* and it may make little sense to have to wait until a minimum amount of them is ready to provide their samples. On the other hand, exocatalogues are a more realistic option since it is feasible to obtain data from *anonymous* donors (e.g. writers in a public database), use this data to build the catalogues and later on enrol the real users of the system as they come.

Experiments show that the origin of the writers used to build the catalogues may have little or no impact at all on the performance of the system. Two variations of the reference experimental process have been run. Both variations consider endocatalogues:

- In the first variation, catalogues are built out of executions from 50 users that will also be used in the train-and-test phase. The train-and-test phase is carried out with those 50 users plus other 270 users (totalling 320 users as in the train and test subset of the reference experiment).
- In the second variation the catalogue and the train-and-test subsets contain exactly the same 320 users.

The results obtained in the two aforementioned variations are summarized in Table 7.3. Notice that they are not given in a per-word basis but as the average over the 16 words. Fig. 7.7 provides a graphical view of the same data.

ORIGIN AND NUMBER OF WRITERS TO BUILD THE CATALOGUES	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER
Exocatalogue, 50 (reference experiment)	75,7%	5,20%	5,52%	76,3%	5,25%	5,56%	89,3%	3,46%	3,73%
Endocatalogue, 50	75,9%	5,19%	5,48%	76,9%	5,24%	5,56%	89,4%	3,42%	3,66%
Endocatalogue, 320	76,6%	4,76%	5,07%	77,2%	4,89%	5,15%	90,2%	3,15%	3,01%

Table 7.3: Average over the 16 words of the recognition performance as a function of the origin and number of the writers used to build the catalogues.

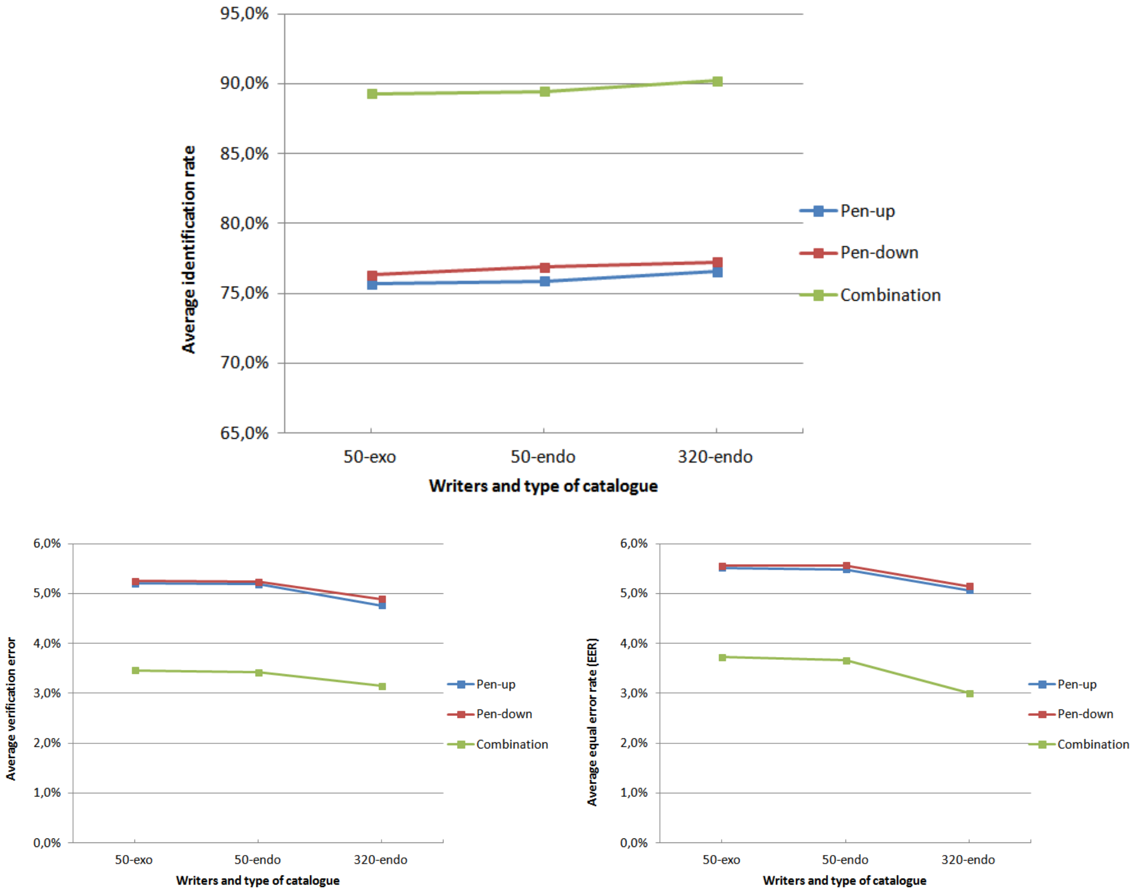


Figure 7.7: Graphical view of the data shown in Table 7.3.

The significance of the difference between the results obtained using an endocatalogue of 50 writers and an exocatalogue of the same number of writers has been studied. For each word and performance metric (IDR, EER, and VE) the difference between one and the other case has been calculated. For each metric, the set of 16 differences has been normalized to mean 0 and standard deviation 1 and put to a Kolmogorov-Smirnov test. In all three cases, the set of differences has been found to be consistent with the hypothesis of normality (H_0: the set of differences follows a normal distribution; H_1: the set of differences does not follow a normal distribution; with a significance level of α=0.05, H_0 is never rejected therefore normality can be assumed). As a final step, each normalized set has been put to a Student's bilateral paired t-test with null hypothesis H_0: the difference is zero; alternative hypothesis H_1: the difference is not zero; degrees of freedom 15. Table 7.4 shows the p-values obtained. Notice that in all cases and for all metrics, the null hypothesis cannot be rejected since the p-values are far from the usual rejection levels (0.05 or less). Therefore, there is no statistical evidence that performance is different when using an endocatalogue.

	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	IDR	VE	EER	IDR	VE	EER	IDR	VE	EER
p-value	0.60	0.88	0.63	0.16	0.84	0.94	0.50	0.45	0.34

Table 7.4: P-values obtained when testing the hypothesis that the difference between the endo-50 and the exo-50 cases is non-null.

7.3 IMPACT OF THE NUMBER OF WRITERS USED TO BUILD THE EXOCATALOGUES

Another issue that deserves some attention is the impact of the amount of data used to build the catalogues. In order to assess this impact the reference experiment has been re-run with catalogues made out of executions coming from 10, 20, 30 and 40 users. Table 7.5 and Fig. 7.8 show the results obtained.

NUMBER OF WRITERS TO BUILD THE EXOCATALOGUES	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER
10	75,2%	5,30%	5,63%	75,3%	5,50%	5,78%	89,0%	3,52%	3,80%
20	76,1%	5,26%	5,56%	76,2%	5,22%	5,52%	89,3%	3,48%	3,71%
30	76,1%	5,28%	5,55%	76,4%	5,15%	5,47%	89,6%	3,35%	3,59%
40	75,9%	5,26%	5,62%	76,2%	5,23%	5,54%	89,2%	3,39%	3,66%
50 (reference experiment)	75,7%	5,20%	5,52%	76,3%	5,25%	5,56%	89,3%	3,46%	3,73%

Table 7.5: Average over the 16 words of the recognition performance as a function of the number of writers used to build the catalogues.

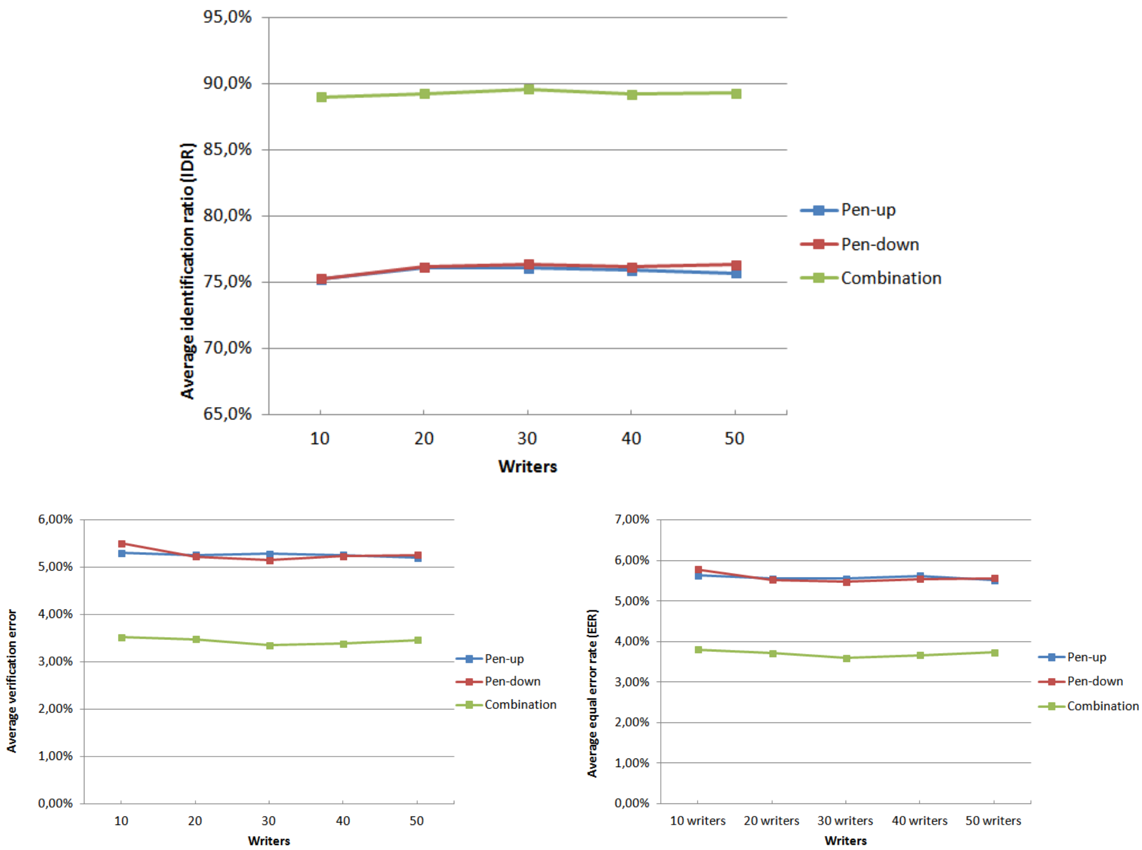


Figure 7.8: Graphical view of the data shown in table 7.5.

The significance of the difference between the results obtained using an exocatalogue of 50 writers (reference experiment) and an exocatalogue of 10 writers has been analyzed. A procedure similar to the procedure described in the previous subsection (impact of the number of writers) has been followed: first the normality of the differences has been tested and the hypothesis has not been rejected. After that, the sets of differences have been put to a Student's bilateral paired t-test with null hypothesis H_0: the difference is zero; alternative hypothesis H_1: the difference is not zero; degrees of freedom 15. Table 7.6 shows the p-values obtained. Only for pen-down strokes, and especially for VE, there seems to be some statistical evidence than an exocatalogue built with strokes from 10 users performs worse than an exocatalogue built out of data from 50 writers.

	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	IDR	VE	EER	IDR	VE	EER	IDR	VE	EER
p-value	0.38	0.34	0.35	0.07	0.013	0.06	0.23	0.42	0.39

Table 7.6: P-values obtained when testing the hypothesis that the difference between the exo-10 and the exo-50 cases is non-null. In green the case where there is statistical evidence of a significant difference (p-value < 0.05). In yellow the cases where there might be some evidence of a significant difference (p-value slightly above 0.05).

7.4 IMPACT OF THE SIZE OF THE CATALOGUES (NUMBER OF UNITS)

Another issue that deserves some attention is the size, in number of units, of the catalogues. Bigger catalogues may require longer training time and will demand larger amounts of storage. The lower these time and space requirements, the wider the number of devices on which the proposed schema could be deployed.

Table 7.7 shows the performances obtained for different sizes of the catalogues. Fig. 7.9 is a graphical view of the data contained in the table.

NUMBER OF UNITS IN THE CATALOGUES	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER	Mean IDR	Mean VE	Mean EER
50	68,3%	6,23%	6,57%	68,5%	6,31%	6,72%	85,1%	4,02%	4,27%
100	74,2%	5,45%	5,79%	74,0%	5,56%	5,88%	88,2%	3,54%	3,84%
150	**75,7%**	**5,20%**	**5,52%**	**76,3%**	**5,25%**	**5,56%**	**89,3%**	**3,46%**	**3,73%**
200	76,1%	5,22%	5,56%	78,0%	4,97%	5,27%	89,9%	3,33%	3,65%
250	76,9%	5,13%	5,44%	78,5%	4,81%	5,11%	90,1%	3,22%	3,52%
300	77,1%	5,16%	5,50%	79,6%	4,74%	5,06%	90,6%	3,23%	3,48%
350	77,1%	5,05%	5,37%	79,7%	4,69%	4,95%	90,9%	3,23%	3,52%
400	76,8%	5,13%	5,43%	80,5%	4,55%	4,84%	90,8%	3,14%	3,42%
450	77,2%	5,13%	5,46%	80,6%	4,55%	4,82%	91,0%	3,12%	3,37%
500	76,8%	5,20%	5,53%	80,7%	4,52%	4,88%	91,0%	3,15%	3,48%

Table 7.7: Average over the 16 words of the recognition performance as a function of the number of units in the catalogues. Results of the reference experiment are in red.

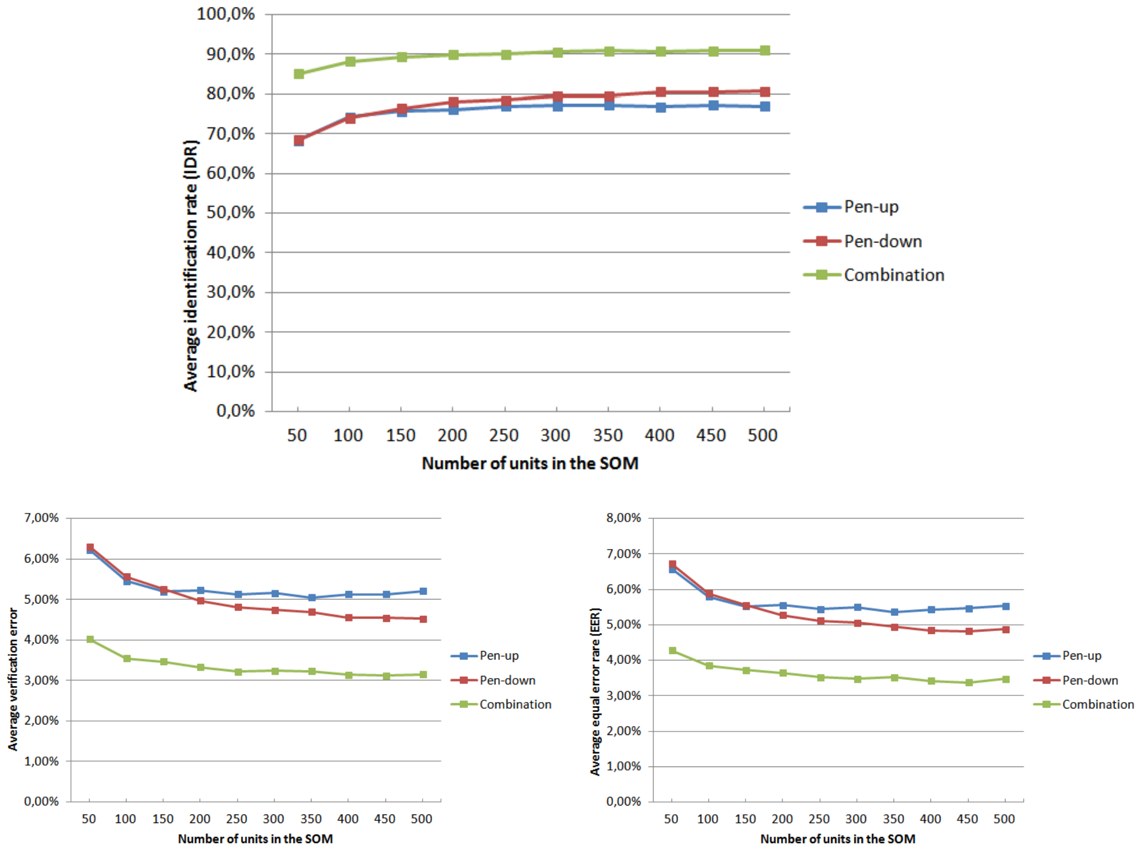


Figure 7.9: Graphical view of the data shown in Table 7.7.

The significance of the difference between the results obtained using an exocatalogue of 50 writers and 150 units (reference experiment) and other exocatalogues of 50 writers and a varying number of units has been studied. In all cases the differences have been found consistent with a normal distribution and have been put to a Student's bilateral paired t-test with exactly the same hypothesis and degrees of freedom than in the preceding cases. Table 7.8 shows the p-values obtained. The following can be noticed:

(a) The difference between 50 and 150 units is always statistically significant (p-value << 0.05).

(b) The difference between 150 and 200 units is significant for pen-down strokes and for the combination of strokes (except for the EER) but no for the pen-up stokes.

(c) For verification (VE and EER) pen-up strokes show no significant difference when going from 150 to 300 or 500 units.

UNITS	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
	IDR	VE	EER	IDR	VE	EER	IDR	VE	EER
50 vs. 150 units	$\sim10^{-10}$	$\sim10^{-9}$	$\sim10^{-8}$	$\sim10^{-9}$	$\sim10^{-6}$	$\sim10^{-7}$	$\sim10^{-9}$	$\sim10^{-7}$	$\sim10^{-7}$
150 vs. 200 units	0.23	0.81	069	$\sim10^{-4}$	$\sim10^{-3}$	0.01	0.02	0.02	0.16
150 vs. 300 units	$\sim10^{-3}$	0.66	0.87	$\sim10^{-8}$	$\sim10^{-5}$	$\sim10^{-5}$	$\sim10^{-4}$	$\sim10^{-3}$	$\sim10^{-3}$
150 vs. 500 units	0.04	0.98	0.88	$\sim10^{-8}$	$\sim10^{-6}$	$\sim10^{-6}$	$\sim10^{-6}$	$\sim10^{-3}$	$\sim10^{-4}$

Table 7.8: P-values obtained when testing the hypothesis of differences among several values for the number of units used in the construction of the catalogues. In green the cases where there is statistical evidence of a significant difference (p-value < 0.05).

From the previously reported statistical tests it seems plausible to infer that pen-up strokes probably require smaller catalogues to achieve an acceptable performance, and that they are less sensitive to the increase in the number of units.

7.5 IMPACT THE OF NUMBER OF WORDS

In chapter 3 it has been asserted that the longer the sequence of text, the greater the accuracy of the recognition. In order to cast some evidence onto this affirmation, the results obtained in the reference experiment have been further combined. This is equivalent to the addition of an extra combination step in the recognition process, in which the dissimilarity measures obtained from two (or more) different words are combined and yield a single dissimilarity measure. Fig. 7.10 shows a graphical depiction of the verifications process when it involves not one but two words.

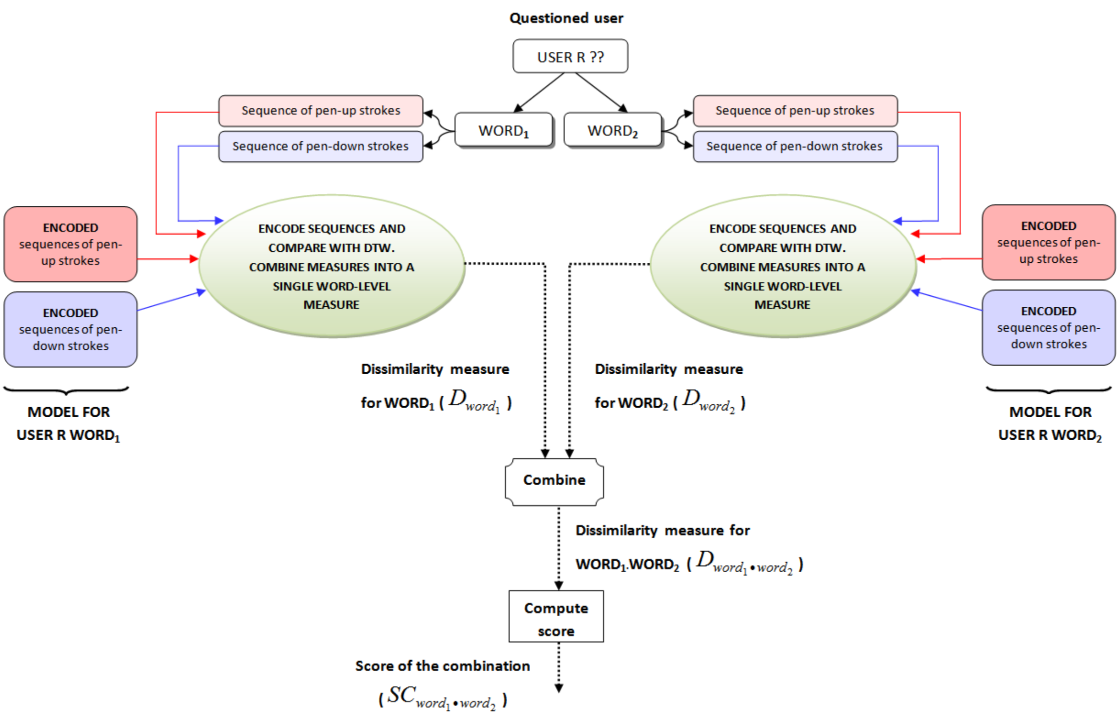


Figure 7.10: Verification process involving two words.

Figs. 7.11 and 7.12 provide a graphical summary of the average performances obtained with one, two, three and four words. Next two subsections give more detailed information.

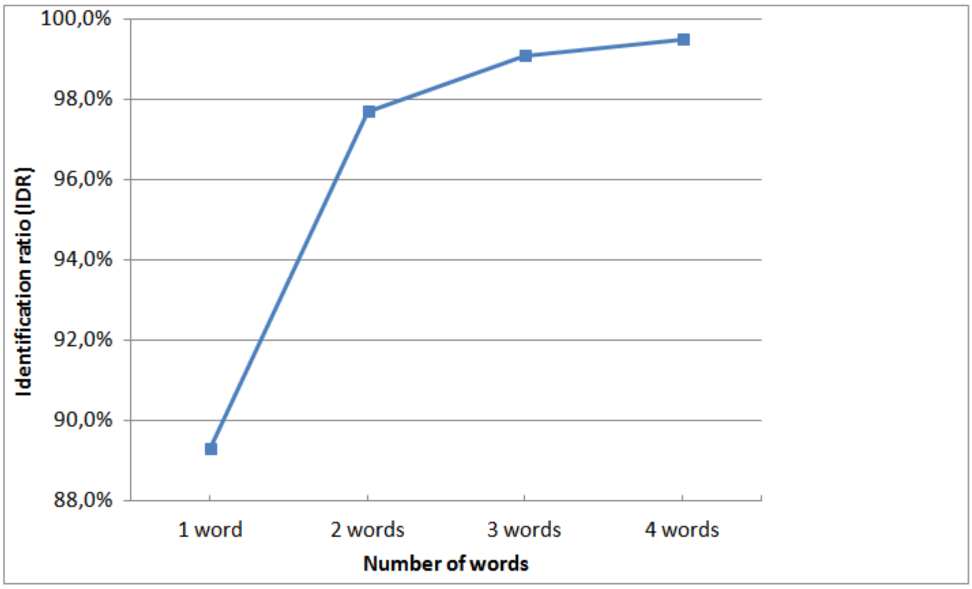


Figure 7.11: Average IDR as a function of the number of words combined.

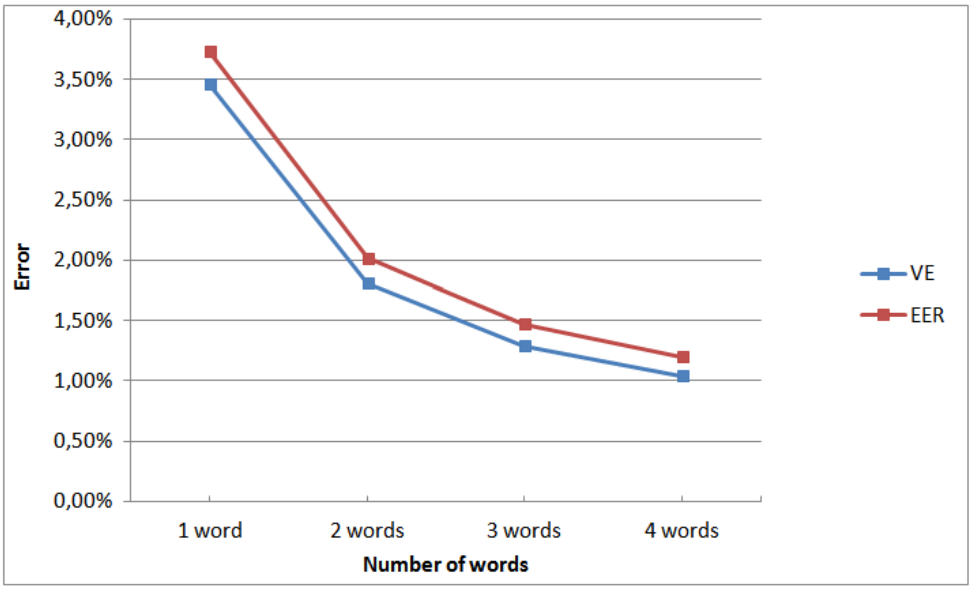


Figure 7.12: Average verification accuracy (VE and EER) as a function of the number of words combined.

7.5.1 COMBINATION OF TWO DIFFERENT WORDS

Tables 7.9, 7.10 and 7.11 contain all the performance metrics obtained with pairs of words from the BiosecurID database. Figs 7.13 and 7.14 show the best and the worst verification enhancements achieved when combining two words. Notice that in all cases, without exception, the combination outperforms the words being combined, both in identification performance (IDR) and verification accuracy (VE and EER).

		DELEZNABLE	DESAPROVECHAMIENTO	DESBRIZNAR	DESLUMBRAMIENTO	DESPEDAZAMIENTO	DESPRENDER	ENGUALDRAPAR	EXPRESIVIDAD	IMPENETRABLE	INEXPUGNABLE	INFATIGABLE	INGOBERNABLE	MANSEDUMBRE	ZAFARRANCHO	ZARRAPASTROSA
		88,4%	96,4%	86,6%	94,7%	94,5%	84,2%	86,5%	91,9%	88,5%	94,5%	90,9%	90,9%	84,1%	81,9%	81,9%
BIODEGRADABLE	93,0%	97,7%	99,1%	97,2%	98,9%	98,6%	97,7%	97,9%	98,4%	98,0%	99,3%	98,4%	98,5%	97,5%	97,6%	98,3%
DELEZNABLE	88,4%		98,3%	98,0%	98,8%	98,1%	96,6%	97,3%	98,4%	97,3%	98,8%	98,7%	98,1%	96,6%	97,7%	97,3%
DESAPROVECHAMIENTO	96,4%			98,6%	98,8%	98,7%	98,1%	98,4%	98,4%	98,2%	99,1%	98,9%	98,9%	98,0%	97,8%	98,3%
DESBRIZNAR	86,6%				98,1%	97,5%	96,1%	97,2%	97,3%	97,3%	98,8%	98,3%	98,4%	96,1%	96,8%	96,7%
DESLUMBRAMIENTO	94,7%					98,3%	97,0%	98,2%	99,1%	98,3%	99,2%	98,8%	98,8%	98,4%	98,5%	98,4%
DESPEDAZAMIENTO	94,5%						97,1%	97,8%	98,3%	98,1%	98,8%	98,6%	98,6%	97,1%	97,3%	98,0%
DESPRENDER	84,2%							96,8%	97,8%	95,7%	97,9%	97,9%	97,3%	93,8%	96,6%	96,8%
ENGUALDRAPAR	86,5%								97,1%	96,8%	98,0%	98,0%	98,0%	95,8%	95,8%	96,8%
EXPRESIVIDAD	91,9%									97,5%	98,8%	97,9%	98,3%	97,0%	97,3%	97,7%
IMPENETRABLE	88,5%										98,1%	97,3%	97,7%	95,9%	97,4%	96,9%
INEXPUGNABLE	94,5%											98,3%	98,4%	97,3%	98,1%	98,4%
INFATIGABLE	90,9%												97,7%	97,2%	96,8%	97,3%
INGOBERNABLE	90,9%													96,3%	97,2%	97,7%
MANSEDUMBRE	84,1%														95,7%	96,6%
ZAFARRANCHO	81,9%															94,4%
Mean over all possible combinations	**97,7%**															

Table 7.9: Results for IDR. Highlighted values are the best (99.3%) and the worst (93.8). Original values for each individual word are shown in red.

		DELEZNABLE	DESAPROVECHAMIENTO	DESBRIZNAR	DESLUMBRAMIENTO	DESPEDAZAMIENTO	DESPRENDER	ENGUALDRAPAR	EXPRESIVIDAD	IMPENETRABLE	INEXPUGNABLE	INFATIGABLE	INGOBERNABLE	MANSEDUMBRE	ZAFARRANCHO	ZARRAPASTROSA
		2,98%	2,05%	3,59%	2,49%	2,84%	4,26%	4,18%	2,50%	3,25%	2,45%	2,24%	3,66%	5,60%	5,26%	5,37%
BIODEGRADABLE	2,67%	1,42%	1,00%	1,58%	1,06%	1,58%	1,60%	1,79%	1,00%	1,32%	1,00%	1,03%	1,35%	1,91%	1,90%	1,90%
DELEZNABLE	2,98%		1,39%	1,66%	1,47%	1,62%	1,97%	1,77%	1,26%	1,82%	1,31%	1,23%	1,59%	2,09%	2,13%	2,09%
DESAPROVECHAMIENTO	2,05%			1,46%	1,32%	1,65%	1,70%	1,68%	1,19%	1,65%	1,11%	1,04%	1,48%	1,96%	2,00%	1,79%
DESBRIZNAR	3,59%				1,56%	1,86%	2,24%	2,13%	1,69%	1,76%	1,35%	1,42%	1,59%	2,52%	2,52%	2,19%
DESLUMBRAMIENTO	2,49%					1,75%	1,84%	1,80%	1,34%	1,74%	1,14%	1,22%	1,53%	2,10%	1,91%	1,71%
DESPEDAZAMIENTO	2,84%						2,13%	2,21%	1,72%	1,70%	1,20%	1,28%	1,63%	2,22%	2,25%	1,94%
DESPRENDER	4,26%							2,34%	1,82%	2,10%	1,62%	1,22%	1,99%	2,72%	2,08%	2,03%
ENGUALDRAPAR	4,18%								1,64%	2,11%	1,72%	1,73%	2,08%	2,79%	3,03%	3,05%
EXPRESIVIDAD	2,50%									1,52%	1,31%	1,32%	1,37%	2,16%	1,48%	1,57%
IMPENETRABLE	3,25%										1,52%	1,54%	2,05%	2,61%	2,34%	2,11%
INEXPUGNABLE	2,45%											1,27%	1,54%	2,27%	1,80%	1,74%
INFATIGABLE	2,24%												1,68%	2,10%	2,12%	2,00%
INGOBERNABLE	3,66%													2,64%	2,43%	2,10%
MANSEDUMBRE	5,60%														3,29%	3,13%
ZAFARRANCHO	5,26%															3,79%
Mean over all possible combinations	**1,81%**															

Table 7.10: Results for VE. Highlighted values are the best (1%) and the worst (3.79%). Original values for each individual word are shown in red.

		DELEZNABLE	DESAPROVECHAMIENTO	DESBRIZNAR	DESLUMBRAMIENTO	DESPEDAZAMIENTO	DESPRENDER	ENGUALDRAPAR	EXPRESIVIDAD	IMPENETRABLE	INEXPUGNABLE	INFATIGABLE	INGOBERNABLE	MANSEDUMBRE	ZAFARRANCHO	ZARRAPASTROSA
		3,29%	2,50%	3,75%	2,74%	3,00%	4,59%	4,51%	2,81%	3,51%	2,58%	2,42%	3,93%	5,87%	5,69%	5,57%
BIODEGRADABLE	2,88%	1,63%	1,18%	1,78%	1,11%	1,73%	1,82%	1,98%	1,11%	1,54%	1,09%	1,24%	1,49%	2,17%	2,13%	2,05%
DELEZNABLE	3,29%		1,56%	1,88%	1,71%	1,82%	2,27%	1,81%	1,41%	1,95%	1,50%	1,47%	1,86%	2,34%	2,52%	2,27%
DESAPROVECHAMIENTO	2,50%			1,64%	1,41%	1,73%	1,95%	1,88%	1,27%	1,80%	1,27%	1,23%	1,72%	2,10%	2,20%	1,96%
DESBRIZNAR	3,75%				1,73%	2,25%	2,41%	2,42%	1,81%	1,96%	1,64%	1,57%	1,79%	2,74%	2,73%	2,49%
DESLUMBRAMIENTO	2,74%					2,03%	2,09%	1,95%	1,35%	1,88%	1,32%	1,34%	1,88%	2,21%	2,19%	2,04%
DESPEDAZAMIENTO	3,00%						2,44%	2,41%	2,02%	1,81%	1,40%	1,51%	1,96%	2,48%	2,48%	2,27%
DESPRENDER	4,59%							2,64%	2,02%	2,33%	1,78%	1,40%	2,28%	2,89%	2,34%	2,26%
ENGUALDRAPAR	4,51%								1,82%	2,46%	1,87%	2,11%	2,26%	2,99%	3,37%	3,36%
EXPRESIVIDAD	2,81%									1,78%	1,48%	1,49%	1,55%	2,29%	1,70%	1,87%
IMPENETRABLE	3,51%										1,70%	1,64%	2,20%	2,80%	2,58%	2,31%
INEXPUGNABLE	2,58%											1,56%	1,79%	2,36%	1,98%	1,95%
INFATIGABLE	2,42%												1,81%	2,21%	2,35%	2,26%
INGOBERNABLE	3,93%													2,97%	2,74%	2,27%
MANSEDUMBRE	5,87%														3,80%	3,44%
ZAFARRANCHO	5,69%															4,00%

Mean over all possible combinations	**2.02%**

Table 7.11: Results for EER. Highlighted values are the best (1.09%) and the worst (4%). Original values for each individual word are shown in red.

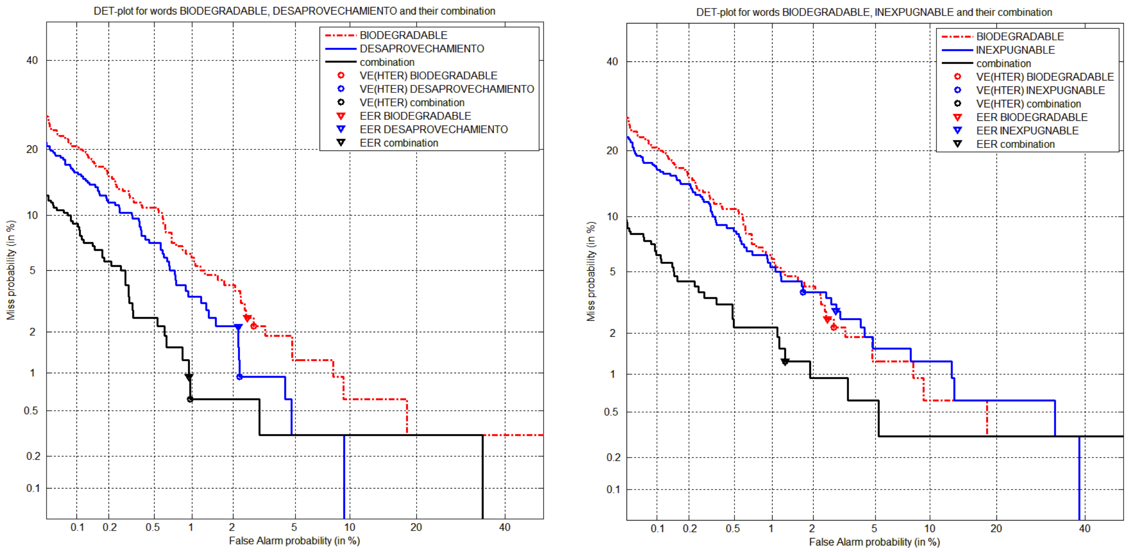


Figure 7.13: DET-plots for the best-performing combinations of words in terms of verification error (BIODEGRABLE+DESAPROVECHAMIENTO, mean VE=1%) and equal error rate (BIODEGRADABLE+INEXPUGNABLE, mean EER=1.09%).

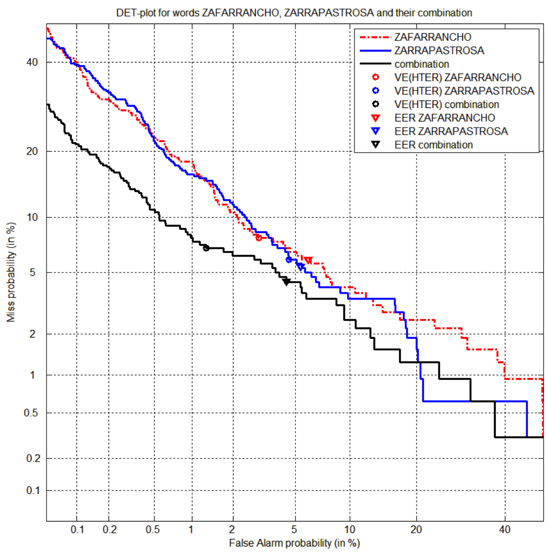


Figure 7.14: DET-plot for the worst-performing combination of words in terms of verification error and equal error rate (ZAFARRANCHO+ZARRAPASTROSA, mean VE=3.79%, mean EER=4%).

7.5.2 COMBINATION OF THREE AND FOUR WORDS

The word-level combination process can be extended to an arbitrary number of words. The cases of three and four words have been scrutinized and the results are summarized in Tables 7.12, 7.13 and 7.14 (three words) and 7.15, 7.16 and 7.17 (four words).

PERFORMANCE	IDR	WORDS	
BEST	99.8%	W1, W5 and W11	BIODEGRADABLE+DESLUMBRAMIENTO+INEXPUGNABLE
		W2, W4, and W12	DELEZNABLE+DESBRIZNAR+INFATIGABLE
		W3, W4, and W12	DESAPROVECHAMIENTO+ DESBRIZNAR+INFATIGABLE
WORST	97.9%	W4, W7 and W14	DESBRIZNAR+DESPRENDER+MANSEDUMBRE
		W7, W10 and W14	DESPRENDER+IMPENETRABLE+MANSEDUMBRE
		W8, W14 and W15	ENGUALDRAPAR+MANSEDUMBRE+ZAFARRANCHO
MEAN	99.1%		

Table 7.12: Summary of IDRs obtained when combining three different words.

PERFORMANCE	VE	WORDS	
BEST	0.61%	W1, W3 and W12	BIODEGRADABLE+DESAPROVECHAMIENTO+INFATIGABLE
WORST	2.78%	W14, W15, W16	MANSEDUMBRE+ZAFARRANCHO+ZARRAPASTROSA
MEAN	1.29%		

Table 7.13: Summary of VEs obtained when combining three different words.

PERFORMANCE	EER	WORDS	
BEST	0.71%	W1, W3 and W12	BIODEGRADABLE+DESAPROVECHAMIENTO+INFATIGABLE
WORST	3.04%	W14, W15, W16	MANSEDUMBRE+ZAFARRANCHO+ZARRAPASTROSA
MEAN	1.47%		

Table 7.14: Summary of EERs obtained when combining three different words.

PERFORMANCE	IDR	WORDS	
BEST	99.9%	W2, W4, W7 and W12	DELEZNABLE+DESBRIZNAR+ +DESPRENDER+INFATIGABLE
		W2, W4, W9 and W12	DELEZNABLE+DESBRIZNAR+ +EXPRESIVIDAD+INFATIGABLE
		W4, W7, W12 and W16	DESBRIZNAR+DESPRENDER+ +INFATIGABLE+ZARRAPASTROSA
WORST	98.6%	W7, W10, W14 and W16	DESPRENDER+IMPENETRABLE+ +MANSEDUMBRE+ZARRAPASTROSA
MEAN	99.5%		

Table 7.15: Summary of IDRs obtained when combining four different words.

PERFORMANCE	VE	WORDS	
BEST	0.51%	W1, W5, W9 and W11	BIODEGRADABLE+DESLUMBRAMIENTO+ +EXPRESIVIDAD+INEXPUGNABLE
WORST	2.22%	W8, W14, W15 and W16	ENGUALDRAPAR+MANSEDUMBRE+ +ZAFARRANCHO+ZARRAPASTROSA
MEAN	1.04%		

Table 7.16: Summary of VEs obtained when combining four different words.

PERFORMANCE	EER	WORDS	
BEST	0.58%	W1, W3, W9 and W11	BIODEGRADABLE+DESAPROVECHAMIENTO+ +EXPRESIVIDAD+INEXPUGNABLE
WORST	2.36%	W8, W14, W15 and W16	ENGUALDRAPAR+MANSEDUMBRE+ +ZAFARRANCHO+ZARRAPASTROSA
MEAN	1.20%		

Table 7.17: Summary of EERs obtained when combining four different words.

7.6 CONCLUSIONS

The results obtained in the experiments and reported in the previous sections show that data gathered from online handwritten words carry enough information to effectively discriminate among writers. Identification rates (IDR) and verification errors (VE and EER) suggest that handwritten words may be of applicability in online biometric recognition systems. Long words (such as *DESAPROVECHAMIENTO*) and combinations of two or more words exhibit a considerable recognition performance. Although much care has to be taken when comparing the performance of different biometric modalities because of their inherent differences and because the conditions under which they are tested may vary greatly (for instance in today's state of the art, methods relying on signatures consider skilled forgeries), we consider that this recognition performance may, in some cases, lie not much below than that of signatures.

It is worth noticing that the experiments described in the previous sections have been carried out with a very limited number of repetitions for training, constrained by the number of repetitions provided by the BiosecurID database (only four repetitions were available and one of them had to be used for testing). From this fact, the following two conclusions may arise:

(a) first, there may be room for further improvement of the identification ratios and verification errors, could more repetitions be available for training and

(b) second, in a real verification scenario, the proposed approach may compare positively with respect to others with regard to users' comfort and acceptance because enrolment is easy, little invasive and can be performed quickly.

The separation of pen-down and pen-up strokes, one of the aspects that contribute to the novelty of the presented approach, has proven to be worthwhile. On the one hand it has given us the possibility to show that the invisible and often discarded or not considered per-se, *in-air* trajectories of the writing device, not only carry a noticeable amount of information but this information is meaningful and useful for biometric recognition purposes. The identification and verification performances of pen-up strokes have been found to be virtually equal to those of pen-down strokes. This result is quite encouraging for it reinforces the idea that pen-up strokes not only are worth being considered but deserve to be analyzed on their own. On the other hand, experiments also show that the dissimilarity measures obtained from both types of strokes can be successfully combined into a single measure, with the latter outperforming the former ones. This result is also relevant because it provides further evidence that pen-up and pen-down strokes contain a certain amount of non-redundant information (see chapter 5), again reinforcing the idea that both types of strokes are worth taking into account.

When it comes to the combination of measures from single words, the experiments show that the combinations outperform the measures being combined. This provides some evidence to the notion that it is possible to improve the overall performance asking the user to write longer sequences or just combing shorter words. Needless to say that the experiments have been constrained by the words provided by the BiosecurID database, which are all quite long (no less than 10 letters). But even so, the fact that the combinations outperform the words being combined may well imply that, in a practical environment, combinations of shorter words may yield reasonable results too, probably dependent on the total length of the sequence.

The experiments have also addressed the issue of the initialization of the system and, more precisely, the issues of the amount of data required to bootstrap the system and of the source of this data, since the proposed method heavily relies on the construction of the catalogues of strokes. This construction must take place before the recognition system is put to work, that is, before the first user is enrolled. It has been shown that these catalogues can be built from sets (of samples) of limited size and that the samples can come from the same users that will later on be enrolled in the system (endocatalogues) or even from users that will not be enrolled in the system (exocatalogues). What is more, the overall performance of the recognition does not seem much affected neither by the number of samples nor by their origin.

8

RECOGNITION ENHANCEMENT BY MEANS OF SYNTHETICALLY GENERATED TEXT

In this chapter a method to generate synthetic executions of online words is presented. This method, inspired and based on the recognition system previously presented in chapter 6, is used to enlarge the enrolment sets, aiming at improving the overall recognition performance. The chapter is organized as follows: the first section introduces the topic of synthetic executions and their use in the enlargement of sets of samples. The second section presents an overview of related works, focusing exclusively on sample duplication, the technique chosen to generate new sequences of text from the available ones. The third section presents the generation method, and the fourth reports on the experimental results achieved, focusing on two issues: the discriminative power of the synthetic samples themselves, and the enhancements obtained when they are used to enlarge the enrolment sets. The last section comments on the conclusions drawn from the reported results.

8.1 WHY SYNTHETIC EXECUTIONS

Recognition based on online handwriting, like all other biometric recognition approaches, involves two stages: enrolment and testing. During the enrolment stage, the user has to produce a set of handwriting samples: the enrolment set (further details have been given in chapters 2 and 6). The number and the quality of the samples acquired during the enrolment stage is an issue of cornerstone importance because a low number leads to poor performance: when the number of samples per user decreases, their ability to effectively discriminate the modelled user from the rest decreases too. In fact, the size of the enrolment set is a critical parameter in recognition systems based on signature [154] and on handwriting in general. Recognition approaches that build models that depend on a larger number of parameters, such as the approaches based on neural networks or statistical methods, tend to require larger enrolment sets in order to attain an accurate enough estimation of those parameters.

Unfortunately, it is not always possible to acquire enrolment sets of the required size and quality. On the one hand, potential donors may be unwilling to donate large numbers of samples of their handwriting due to concerns regarding their future use and possible security compromise. On the other hand, the quality of samples may be compromised due to fatigue if the donor is requested to produce an excessive number of samples in a single session. If the solution to the last issue is to collect one or a little amount of samples in each session during an increased number of sessions, a considerable number of donors may be lost in the way.

The uses of synthetically generated biometric data are manifold. The **synthetic enlargement** of datasets, the **generation of synthetic human-like traits** and the **generation of synthetic individuals** are among the most relevant.

- The synthetic enlargement of datasets aims at increasing the number of samples per user. New samples are obtained transforming and/or combining real samples.
- The synthesis of human-like traits is an approach typically followed in applications that require synthesizing speech or handwriting. Primitive units taken from a pool constructed from real samples are combined to produce the required speech or writing.
- The generation of synthetic individuals is based on the creation of models for a given trait in a population. Once the model of a trait is created, it can be sampled to synthesize *new* individuals. This approach helps overcome the scarcity of biometric-data donors since synthetic individuals can be used to test the performance of new developed systems, even when no real data is available.

In [155], Yanushkevich et al. state that automatically generated data helps creating meaningful sets of data variations that can improve the performance of existing identification systems. Focusing on security issues, they also point out other uses for synthetically generated biometric data such as the improvement of the robustness of the biometric devices thanks to the availability of forged-like data that can help modelling, and thus detecting, forgeries.

The synthetic enlargement of the enrolment set is a mean to increase the size of this set without increasing the burden put on the user during the enrolment stage. Not only the recognition accuracy may benefit from the use of synthetic data: the ability to improve the performance without the need to ask the user for an increased number of samples during the enrolment phase may also have a positive impact on collectability, when compared to a similar improvement achieved by just asking for more real samples. Furthermore, an improved collectability may also lead to a greater degree of acceptability. Thus, synthetically enlarging the enrolment set may have benefits spanning three of the seven factors that impact on the quality of a biometric system (see section 2.3.3): performance, collectability and acceptability.

8.2 RELATED WORK: SAMPLE DUPLICATION

When the main goal of generating synthetic data is to increase the size of the available datasets, the strategy of choice is ***sample duplication*** [156]. Through a series of different transformations, one or several samples from the same individual produce one or more new (different) samples that will be regarded as belonging to that individual. The method proposed in this dissertation belongs to this category. This section highlights some relevant aspects and papers related to the application of sample duplication aimed at the enlargement of the available datasets, first in handwriting recognition and later in writer recognition.

8.2.1 SAMPLE DUPLICATION IN HANDWRITING RECOGNITION

Handwriting recognition, the recognition of written text[8], not of its author, is a discipline closely related to writer recognition since they share some methods and techniques. It can be

[8] The main goal of handwriting recognition is the extraction of a symbolic representation from the spatial forms that constitute a handwritten text.

performed in a writer-dependent way or it can be performed independently of the writer. In the former case, the recognition system is trained with samples from a single writer and its goal is to recognize the writing of that specific user. Writer-dependent recognition has in the later years attracted some research attention thanks to the flourish of pen-based devices such as PDAs and PC-tablets. In writer-independent recognition, the system is trained with samples from multiple writers and its goal includes recognizing the writing of previously unknown individuals. In both cases, the sizes of the training sets are of paramount importance and their enlargement is a question that has been often addressed in the literature.

In [157], Mori et al. report small improvements in the recognition performance of a K-Nearest Neighbour (*K-NN*) classifier when synthetically generated numerals are combined with real samples. Generation is performed by means of a method proposed by the authors. In [158], Cano et al. apply *slant*, *shrink*, *erosion* and *dilation* to images of single characters acquired from different writers. The transformed images are added to the training set of a *K-NN* classifier and a 4% improvement in the recognition rate is reported. Varga and Bunke, in [159], generate synthetic textlines from existing lines of handwriting from different writers by means of geometrical transformations and *thinning*/*thickening* operations. When the synthetic lines are included in the training sets of an HMM-based cursive handwritten sentence recognizer, the recognition rate is improved in 29 out of 33 cases. Using the same HMM-based recognizer, Varga, Kilchhofer and Bunke in [160] assess the impact produced by the enlargement of the training set with lines generated from character templates and the Delta LogNormal handwriting generation model. Some of the enlarged training sets perform better than the non-enlarged. Also Helmers and Bunke, in [161], report some improvements in the performance of the same recognizer when using a generation technique that synthesizes handwritten-like text from ASCII text by means of a dictionary of n-tuples (groups) of characters.

In the online field, much less explored than the offline, it is worth mentioning the work of Mouchère and Anquetil. In [162] they use synthetically generated characters to increase the number of available samples while trying to preserve the user's writing style. Both offline and online transformations are applied to the original online samples. Offline transformations are classical image transformations, actually stretching and slope, while online transformations are speed and curvature changing. With sets of 360 samples (up to 10 original, the rest synthetic ones) of each character they achieve recognition rates that surpass those of reference (86.9% vs. 79.36% with 3 original samples).

8.2.2 SAMPLE DUPLICATION IN WRITER RECOGNITION

To the best of the authors' knowledge no relevant references exist in the scientific literature dealing with sample duplication applied to non-signature-based writer recognition. For this reason, all the references in this section address sample duplication from within the signature based approach.

Duplicated samples are most often produced by distorting, in different ways and with different techniques, one or more real samples. For instance, Huang and Yan in [163] present an offline signature verification method based on a neural network. In order to tackle the issue of the relatively large number of samples required to train the network, they propose to artificially enlarge the enrolment set by means of new samples obtained by applying perturbations to a small set of genuine ones. The perturbations applied are slant distortion, horizontal and vertical size distortion, rotation, and perspective view distortion. Slightly distorted samples are presented to the network as genuine ones, whereas heavily distorted ones are presented as forgeries. Experimenting on a database of over 3000 (real) signatures donated by 21 people they achieve a false acceptance rate of 11.1% and a false rejection rate of 11.8%, when considering skilled forgeries. It is interesting to note the large number of synthetic samples used: regarding reference genuine samples, the network is trained with 8 real and about 300 synthetic ones. Regarding samples of forged signatures, over 600 synthetic samples (plus real ones, up to 3000) are presented to the neural network. Also in the offline field, de Oliveira et al. propose, in [164], a technique for the automatic generation of new signatures obtained from the deformation of real samples (Fig. 8.1). This technique is applied to the reconstruction of the trajectories of the pen obtained from the offline images of the signatures. The signals that represent the trajectories of the pen are deformed by their convolution with deforming functions. Deformations include uniform and non-uniform scaling, uniform and non-uniform rotation and their combination.

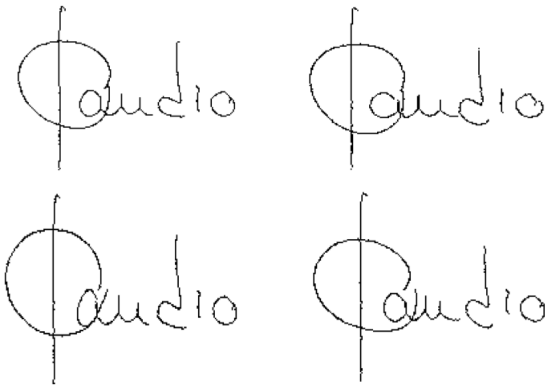


Figure 8.1: An original signature (top left) and three synthetic ones obtained by the deformation procedures presented in [164] (Image taken from [164]).

In [165], Rabasse, Guest and Fairhurst introduce a method for the generation of static signature images from dynamic (online) data. Their method aims at generating the x-y trajectories of synthetic signatures from two real ones, the seed signatures, produced by the same user (Fig 8.2). Seed signatures are first size-normalized and then derivative dynamic time warping (DDTW) is applied to map points in one seed to points in the other. New signatures are generated by interpolating points between mappings. In a final stage, variability is introduced within the synthesized signatures. The added variability is modelled according to the natural variability occurring within the user's signature samples. In order to assess whether the synthesized signatures are representative of the real ones, the authors use a commercial signature verification engine to compare the verification rates obtained when comparing real signatures with other real signatures and the rates obtained when comparing real signatures with synthesized ones. Similar error rates (FAR and FRR) are reported.

Figure 8.2: Seed signatures (leftmost) and four new signatures synthesized from them by means of the method presented in [165] (Taken from [165]).

In the online field, distortion is again the method of choice to obtain synthetic samples. In [166], the methodology introduced in [165] is extended in order to synthesize dynamic data (x-y coordinates plus pressure, time, status, azimuth and altitude). In the performed experiments, the authors found similar enrolment rates but lower verification rates. Even if the verification performance in the latter work was slightly lower for synthetic signatures, the works by Rabasse et al. clearly point out the suitability of synthetically generated signatures to represent real ones.

Munich and Perona, in [167], use duplicated samples in training and in testing. In their system duplicated samples are produced by means of resampling by spline interpolation, and affine deformations (horizontal and vertical affine scaling). In training, the duplicated samples increase the size of the training set. In testing, duplicated samples increase the number of skilled forgeries available. According to the authors, the use of duplicated samples in training and testing helps achieving a better estimation of the statistical performance of the system (measured by equal error rate –EER-). In their experiments, this particular use of duplicated samples leads to a slight increase in error rates.

In [168] Galbally et al. propose a generation method based on three different kinds of distortions sequentially applied to the original samples. These distortions are: (a) addition of noise, (b) resampling by a given factor and (c) amplification/attenuation. The parameters that characterize the distortions are estimated from a dataset not used in testing and are aimed at capturing the intra-user variability. Using a state-of-the-art HMM-based signature verification system, the authors perform some experiments the results of which show that when synthetic samples are added to the real ones, the verification performance is notably increased. They report up to a 70% improvement (the system yields a 23.84% EER when trained with one real sample. This EER decreases to a 7.87% when 19 synthetic samples are added to the real one). The improvement reported by Galbally et al. show that synthetically enlarged enrolment sets can outperform non-enlarged ones.

Figure 8.3: Real signature (leftmost) and synthetic ones obtained from it using the method described in [168] (taken from [168]).

The strategy of duplicated samples also has a potential in the analysis of the behaviour of signature-verification systems to help understand why forged signatures are accepted or genuine ones rejected. For instance, in [169] Djioua et al. present a software tool based on the Kinematic Theory of rapid human movements that can generate modified signatures from real samples by varying some parameters. The authors plan to use those modified signatures to assess how signature-verification systems react to them.

8.3 PROPOSED METHOD

8.3.1 GENERAL OVERVIEW

This subsection concisely summarizes the algorithm that generates synthetic sequences of strokes. The next subsection will provide a more detailed explanation.

1. Two samples from a given user are taken. From each sample, two sequences are obtained, one of pen-up and one of pen-up strokes.
2. Each sequence of strokes is encoded as a sequence of integers, using an existing pair of catalogues.
3. The sequences of encoded pen-down strokes (one from each sample) are aligned using DTW. As a result, a sequence of alignment steps is obtained. Each alignment step is one of {*Match, Insertion, Deletion*}
4. The processing of the sequence of alignment steps yields a new sequence of (non encoded) pen-down strokes. This processing is performed, alignment step by alignment step, as follows:
    a. If the step indicates that the encoded strokes *match*, a new stroke is generated by *averaging* the two corresponding non encoded strokes
    b. If the step indicates an *insertion*, the inserted stroke is added to the end of the sequence under generation.
    c. If the step indicates a *deletion*, nothing is done.

Steps 3 and 4 are repeated for pen-up strokes.

Fig. 8.4 graphically depicts the synthesis process for pen-down strokes. Pen-up strokes follow the very same process.

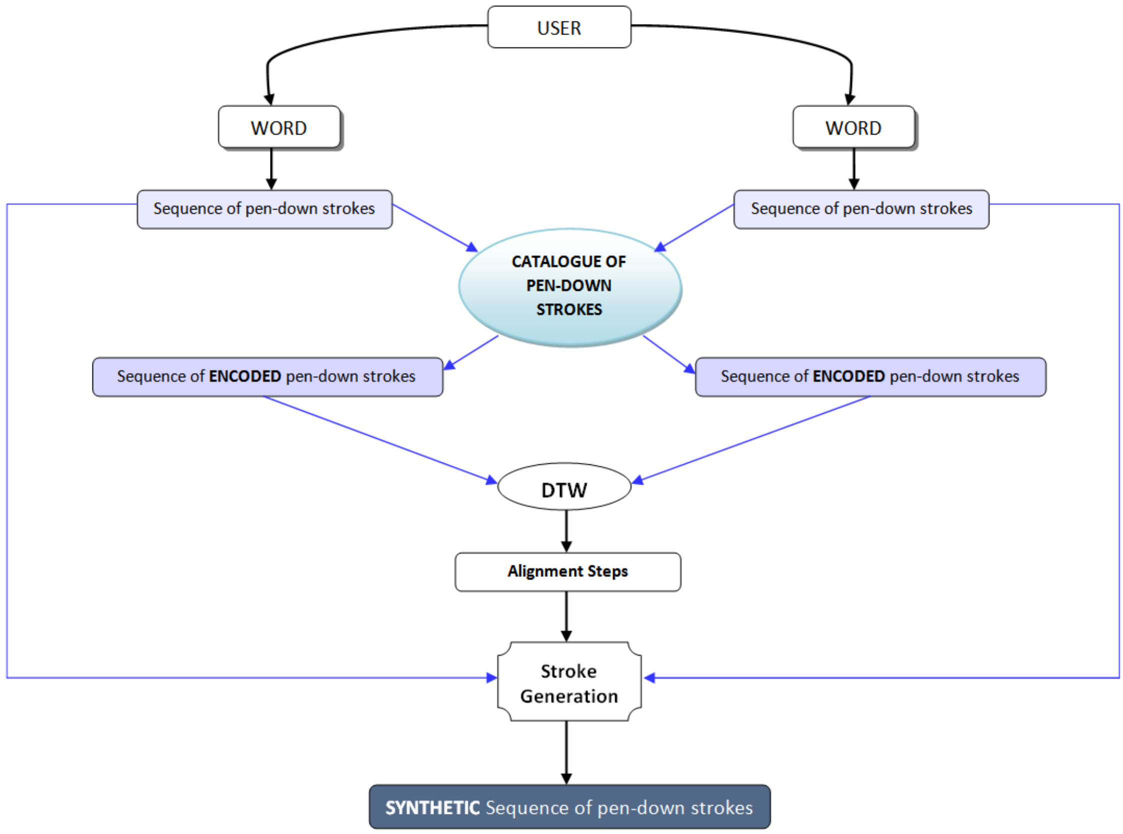


Figure 8.4: Schematic overview of the synthesis process for pen-down strokes.

8.3.2 DETAILED DESCRIPTION

$S_1 = s_1^1,...,s_m^1$ and $S_2 = s_1^2,...,s_n^2$ denote the original sequences of strokes (pen-up or pen-down) that will be used to produce a new synthetic sequence. S_1 and S_2 come from different executions of the same word produced by the same writer. s_i^j denotes the *i*-th stroke of the *j*-th sequence.

As in section 6.2.3, $SE_1 = BMUIdx_1^1,...,BMUIdx_m^1$ and $SE_2 = BMUIdx_1^2,...,BMUIdx_n^2$ denote the encoded versions of S_1 and S_2 respectively. $BMUIdx_i^j$ denotes the index of the stroke prototype corresponding to the *i*-th stroke in the *j*-th original sequence.

$AS = as_1,...,as_p$ denotes the sequence of alignment steps obtained when DTW is applied to SE_1 and SE_2. $as_i \in \{Match, Insertion, Deletion\}$

NS denotes the new synthetic sequence.

The algorithm that generates the synthetic sequences from the real ones is detailed in Figs. 8.5 and 8.6.

Algorithm. Generate synthetic sequence of strokes
Input: original sequences S_1 and S_2; catalogue of strokes; weight W of sequence S_1
Output: NS

Obtain SE_1 and SE_2 from S_1 and S_2 respectively, using the given catalogue;

AS := apply **DTW-alignment** to SE_1 and SE_2 to align SE_1 **with** SE_2

$NS := \varnothing$;

$indexS_1 := 1$;
$indexS_2 := 1$;

for each as_i in AS
    switch on as_i
        case $as_i = Match$
            $ns := average(s^1_{indexS_1}, s^2_{indexS_2}, w)$;
            add ns to the end of NS;
            $indexS_1 := indexS_1 + 1$;
            $indexS_2 := indexS_2 + 1$;

        case $as_i = Insertion$
            add $s^1_{indexS_1}$ to the end of NS;
            $indexS_1 := indexS_1 + 1$;

        case $as_i = Deletion$
            $indexS_2 := indexS_2 + 1$;
    end of switch
end of for each

Figure 8.5: Algorithm used to obtain synthetic sequences from a pair of real ones.

The function *average(stroke1, stroke2, w)* computes a new synthetic stroke that is the weighted *average* of the two parameters. Actually each feature of each stroke is transformed into the frequency domain by a discrete Fourier transform (DFT) and the resulting sequences of coefficients are averaged. First sequence is given weight *w* while the second is given weight *1-w*. The new sequences of coefficients, one per feature, are transformed back to the original time domain applying an inverse discrete Fourier transform (IDFT).

Algorithm. DTW-alignment
Input: encoded sequences of strokes SE_1 of length m and SE_2 of length n
Output: *AS*, the sequence of alignment steps

CM is a (m+1)x(n+1) matrix of real numbers with indexes ranging from (0,0) to (m,n);

MS is a mxn matrix of alignment steps with indexes ranging from (1,1) to (m,n);

Initialize *CM* with all positions equal to +∞, except *CM*(0,0) = 0;

```
for i in 1..m
    for j in 1..n
        c := COST(BMUIdx_i^1, BMUIdx_j^2);
        if CM(i-1,j) < CM(i,j-1) and CM(i-1,j) < CM(i-1,j-1)
            step := Insertion;
            c := c + CM(i-1,j);
        else if CM(i,j-1) < CM(i-1,j) and CM(i,j-1) < CM(i-1,j-1)
            step := Deletion;
            c := c + CM(i,j-1);
        else
            step := Match;
            c := c + CM(i-1,j-1);
        end of if
        CM(i,j) := c;
        MS(i,j) := step;
end of inner for
end of outer for

i := m;
j := n;
AS := ∅;
while i≠1 or j≠1
    step := MS(i,j);
    switch on step
        case step = Insertion
            i=i-1;
        case step = Deletion
            j=j-1;
        case step = Match
            i=i-1; j=j-1;
    end of switch
    Add step to the beginning of AS
end of while
Add MS(i,j) to the beginning of AS
```

Figure 8.6: Version of the DTW algorithm that obtains the sequence of alignment steps used to generate synthetic sequences.

The function *COST(Index1, Index2)* is exactly the same one used in recognition (see section 6.2.4).

Notice that DTW-alignment is non-commutative because it can yield different sequences of alignment steps depending on the order of the input parameters SE_1 and SE_2. An *insertion* is a stroke in the first sequence with no corresponding matching in the second, while a *deletion* is a stroke in the second sequence with no corresponding matching in the first. What will appear as an *insertion* when aligning SE_1 *with* SE_2 will appear as a *deletion* when aligning SE_2 *with* SE_1. As the sequence of alignment steps drives the generation of new sequences of strokes, two different synthetic sequences can be generated from each pair of original sequences.

Figs. 8.7 and 8.8 show a pair of real sequences of strokes and the new sequences synthetically generated from them. Notice that, in the case of pen-down strokes, the generated executions are perfectly legible. Although not shown in the figures, pressure and writing angles have also been synthesized.

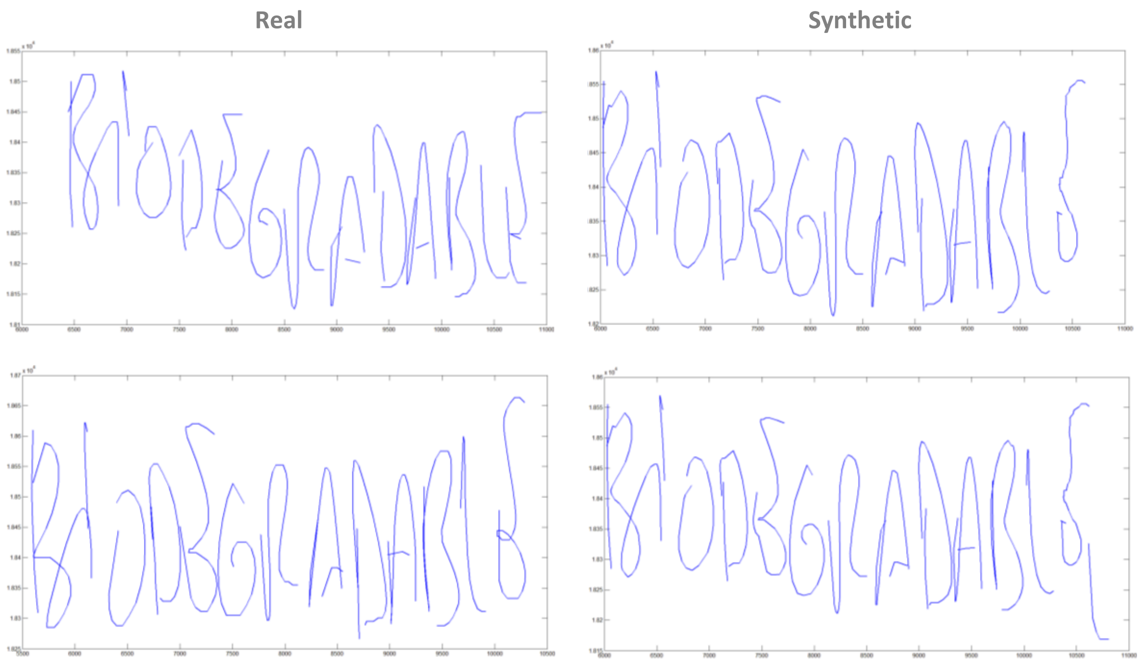


Figure 8.7: Two real executions, pen-down strokes only, of the word BIODEGRADABLE (left) and the synthetic executions generated from them with w=0.5 (right).

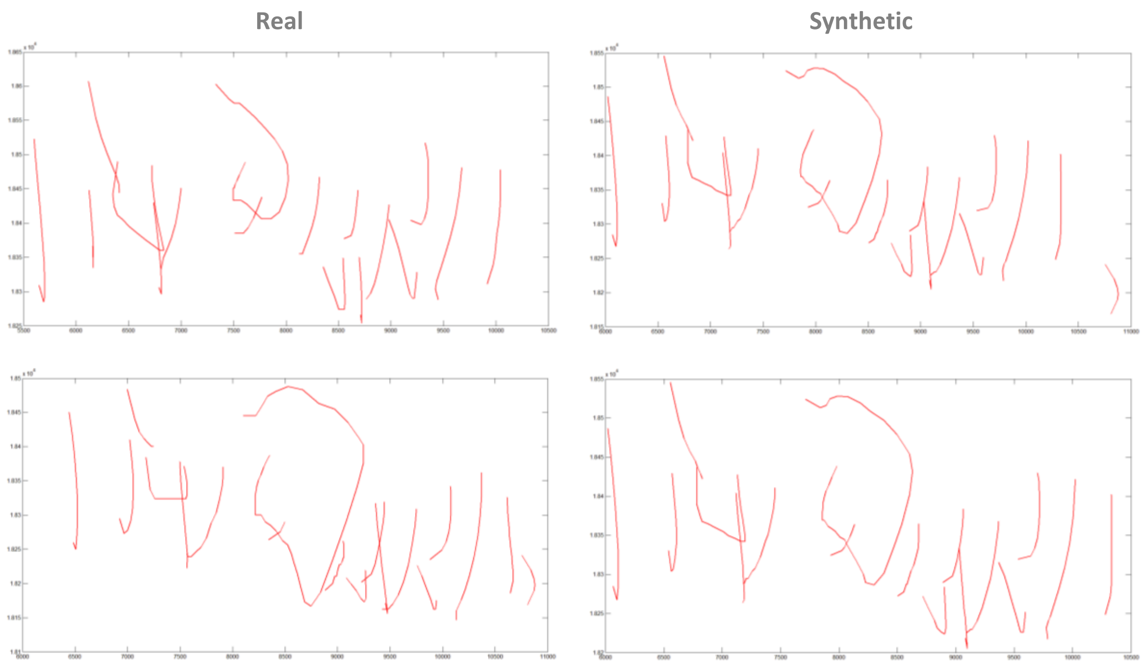


Figure 8.8: Pen-up strokes corresponding to the two real executions of the word BIODEGRABLE shown in the previous figure (left) and the sequences synthetically generated from them with w=0.5 (right).

8.4 EXPERIMENTAL RESULTS

Two different experiments have been carried out:

- The first one considers an enrolment set that only contains synthetically generated executions. The purpose of this experiment is to assess the recognition performance of the synthetic executions when compared to the recognition power of the original executions they derive from, that is, to evaluate how much recognition performance they *retain*.

- The second experiment considers an enlarged enrolment set, that is, an enrolment set that contains original executions and all the synthetic ones generated from them. Its purpose is to reveal the impact of the addition of synthetic executions to the enrolment set.

Fig. 8.9 shows the different types of enrolment sets.

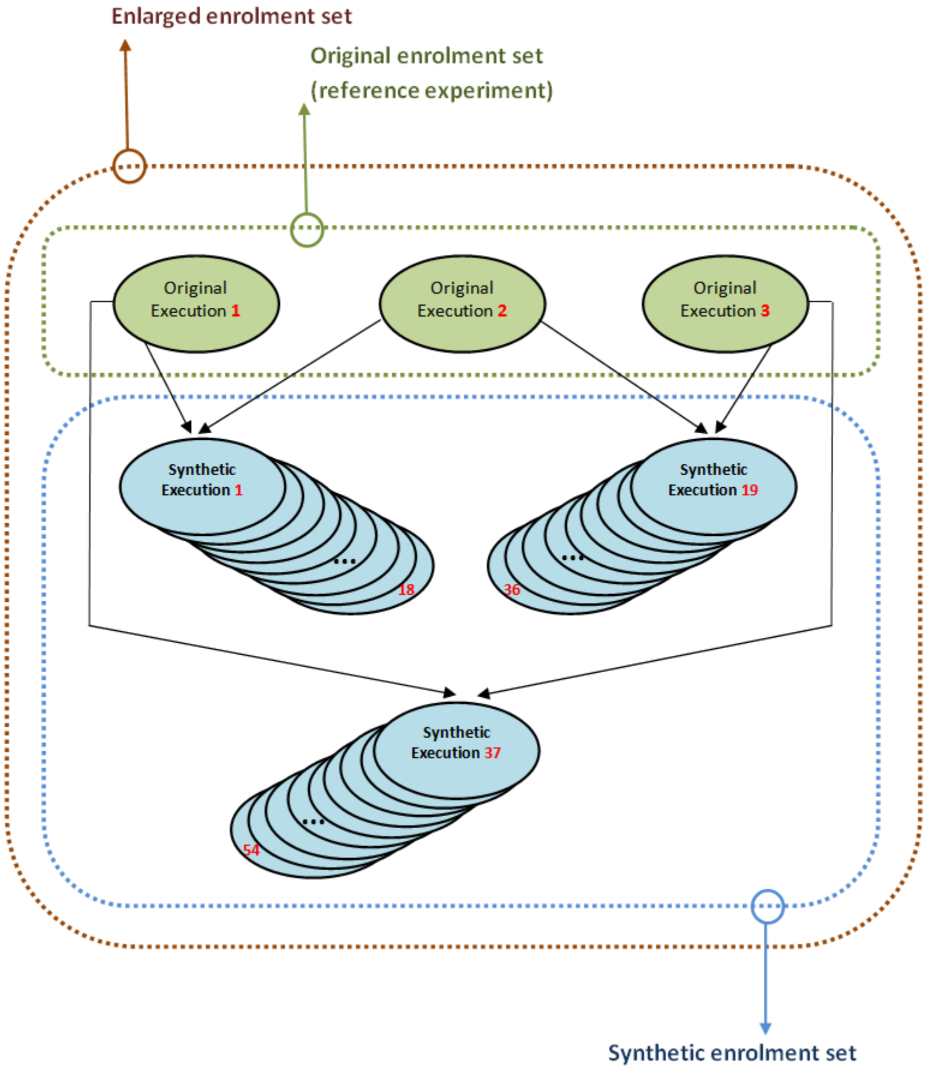


Figure 8.9: The three different types of enrolment sets considered in the experiments.

The experimental process is identical to the reference experimental process described in section 7.1 and depicted in Fig 7.1, except for the enrolment sets used. In the reference experiment, enrolment sets used in traininig (to build models) were always original enrolment sets, that is, they did not contain any synthetic executions, whereas the enrolment sets considered in the experiments reported in this chapter are the *synthetic enrolment sets* (first experiment) and the *enlarged enrolment sets* (second experiment)

Regarding the generation of the synthetic executions, 18 different synthetic executions have been obtained from each pair of original ones, using weights (*W*) ranging from 0.1 to 0.9 with a step of 0.1. As each invocation of the generation algorithm produces 2 sequences (see section 8.3.2 and Fig. 8.5), 18 sequences (new executions) result from the 9 different weights considered. Thus, all synthetic enrolment sets contain 54 executions while all enlarged enrolment sets contain 57 executions.

Notice that synthetic executions are only used in the training stage. No synthetic execution is used neither to build the catalogues of strokes nor for testing: the catalogues are exactly the same ones that were used in the reference experiment while the sample used in testing (i.e. the sample being compared with the models) is always a real execution.

8.4.1 RECOGNITION PERFORMANCE OF THE SYNTHETIC EXECUTIONS

As it has been stated before, the main goal of this experiment is to assess the amount of recognition performance that *remains* in synthetic executions. Tables 8.1, 8.2 and 8.3 comprehensively show the results obtained. For each word and metric (IDR, VE and EER), the result obtained with the synthetic enrolment set is compared to the result yielded by the reference experiment (original enrolment set). The difference is given as the percentage of increase/decrease (*Δ%*) with respect to the value obtained in the reference experiment:

- For IDR, a positive difference means that the synthetic enrolment sets outperform the original ones (more writers correctly identified).

- For VE and EER, a negative difference means that the synthetic enrolment sets outperform the original ones (less verification error, smaller equal error rate).

It is worth noticing that in most cases the synthetic enrolment sets clearly outperform the original ones:

- For IDR, synthetic enrolment sets always outperform the original ones. Moreover, a higher increment is achieved for those words that showed a lower performance (Pearson's correlation coefficient between the original values and the increments is minus 0.93 for pen-up strokes, minus 0.84 for pen-down strokes and minus 0.94 for the combination).

- For VE, only in one case the synthetic enrolment sets do not outperform the originals (word DESAPROVECHAMIENTO, pen-up strokes). Nevertheless, in the case of the combination of both types of strokes synthetic enrolment sets always outperform the original ones.

- For EER, only in two cases the synthetic enrolment sets do not outperform the originals (words DESLUMBRAMIENTO and DESPEDAZAMIENTO, pen-down and pen-up strokes, respectively). In the case of the combination, synthetic enrolment sets always outperform the original ones.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	80,0%	83,1%	3,9%	81,4%	86,3%	6,0%	93,0%	94,8	1,9%
W2	DELEZNABLE	72,1%	78,2%	8,5%	73,8%	78,0%	5,8%	88,4%	92,0	4,2%
W3	DESAPROVECHAMIENTO	89,8%	92,2%	2,6%	89,3%	91,3%	2,2%	96,4%	97,1	0,7%
W4	DESBRIZNAR	74,3%	77,7%	4,5%	71,6%	77,0%	7,4%	86,6%	90,2	4,1%
W5	DESLUMBRAMIENTO	83,9%	88,2%	5,1%	84,2%	89,3%	6,0%	94,7%	96,2	1,6%
W6	DESPEDAZAMIENTO	89,1%	91,3%	2,5%	84,4%	88,7%	5,1%	94,5%	95,9	1,4%
W7	DESPRENDER	71,3%	76,5%	7,2%	67,7%	72,1%	6,5%	84,2%	89,2	5,9%
W8	ENGUALDRAPAR	72,0%	77,0%	6,9%	71,2%	76,7%	7,8%	86,5%	90,2	4,3%
W9	EXPRESIVIDAD	75,5%	81,6%	8,1%	79,3%	84,5%	6,6%	91,9%	94,8	3,1%
W10	IMPENETRABLE	78,4%	83,8%	6,8%	73,1%	79,0%	8,0%	88,5%	91,6	3,5%
W11	INEXPUGNABLE	82,3%	85,7%	4,2%	83,4%	88,0%	5,5%	94,5%	96,2	1,8%
W12	INFATIGABLE	79,2%	82,7%	4,4%	79,5%	82,8%	4,1%	90,9%	93,1	2,5%
W13	INGOBERNABLE	74,8%	79,5%	6,2%	81,3%	85,4%	5,0%	90,9%	93,4	2,8%
W14	MANSEDUMBRE	63,4%	70,2%	10,9%	68,6%	75,4%	9,9%	84,1%	87,9	4,5%
W15	ZAFARRANCHO	63,1%	70,4%	11,5%	65,2%	71,6%	9,8%	81,9%	85,6	4,6%
W16	ZARRAPASTROSA	61,9%	69,9%	13,0%	67,3%	72,8%	8,2%	81,9%	85,9	5,0%
Average over the 16 words		**75.4%**	**80.5%**	**6,6%**	**76.3%**	**81.2%**	**6,5%**	**89.3%**	**92.1%**	**3,2%**

Table 8.1: Comparison of IDRs obtained with original and synthetic enrolment sets. Positive increments (green) mean that the synthetic enrolment sets outperform the original ones.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	4,35%	3,59%	-17,4%	4,29%	3,96%	-7,7%	2,67%	2,31%	-13,5%
W2	DELEZNABLE	5,89%	5,00%	-15,1%	4,48%	4,05%	-9,6%	2,98%	2,73%	-8,4%
W3	DESAPROVECHAMIENTO	2,98%	3,05%	2,1%	3,62%	3,02%	-16,5%	2,05%	1,68%	-18,1%
W4	DESBRIZNAR	5,10%	5,08%	-0,5%	5,77%	5,36%	-7,2%	3,59%	3,10%	-13,6%
W5	DESLUMBRAMIENTO	3,82%	3,20%	-16,2%	3,93%	3,83%	-2,5%	2,49%	2,19%	-12,0%
W6	DESPEDAZAMIENTO	3,39%	3,98%	17,6%	4,47%	3,78%	-15,5%	2,84%	2,36%	-16,9%
W7	DESPRENDER	5,55%	4,06%	-26,8%	6,58%	5,93%	-9,9%	4,26%	3,77%	-11,4%
W8	ENGUALDRAPAR	5,76%	4,92%	-14,5%	6,33%	5,52%	-12,8%	4,18%	3,75%	-10,3%
W9	EXPRESIVIDAD	5,05%	4,06%	-19,5%	4,00%	3,69%	-7,8%	2,50%	2,03%	-18,9%
W10	IMPENETRABLE	4,90%	4,61%	-6,0%	5,55%	4,79%	-13,7%	3,25%	2,68%	-17,5%
W11	INEXPUGNABLE	4,05%	3,44%	-15,2%	4,18%	3,59%	-14,0%	2,45%	1,86%	-23,9%
W12	INFATIGABLE	4,38%	5,08%	15,9%	4,04%	3,85%	-4,7%	2,24%	2,17%	-3,2%
W13	INGOBERNABLE	5,56%	5,16%	-7,2%	5,03%	4,22%	-16,2%	3,66%	3,12%	-14,9%
W14	MANSEDUMBRE	6,81%	6,25%	-8,2%	7,87%	6,93%	-12,0%	5,60%	4,56%	-18,5%
W15	ZAFARRANCHO	7,40%	6,72%	-9,2%	7,13%	6,37%	-10,6%	5,26%	4,60%	-12,6%
W16	ZARRAPASTROSA	8,22%	7,34%	-10,7%	6,81%	5,86%	-14,0%	5,37%	4,36%	-18,9%
Average over the 16 words		**5,20%**	**4,72%**	**-8,17%**	**5,25%**	**4,67%**	**-10,91%**	**3,46%**	**2,95%**	**-14,54%**

Table 8.2: Comparison of VEs obtained with reference and synthetic enrolment sets. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	SYNTHETIC ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	4,72%	3,83%	-18,8%	4,62%	4,37%	-5,4%	2,88%	2,43%	-15,9%
W2	DELEZNABLE	6,23%	5,64%	-9,6%	4,69%	4,23%	-9,8%	3,29%	2,76%	-16,3%
W3	DESAPROVECHAMIENTO	3,28%	3,06%	-6,8%	3,91%	3,19%	-18,4%	2,50%	1,81%	-27,5%
W4	DESBRIZNAR	5,61%	5,15%	-8,1%	5,99%	5,64%	-5,9%	3,75%	3,33%	-11,1%
W5	DESLUMBRAMIENTO	4,12%	3,68%	-10,7%	4,22%	4,28%	1,4%	2,74%	2,35%	-14,1%
W6	DESPEDAZAMIENTO	3,60%	3,65%	1,4%	4,62%	4,06%	-12,0%	3,00%	2,50%	-16,6%
W7	DESPRENDER	5,69%	5,19%	-8,8%	6,88%	6,30%	-8,4%	4,59%	4,14%	-10,0%
W8	ENGUALDRAPAR	6,09%	5,25%	-13,8%	6,86%	5,84%	-14,8%	4,51%	4,22%	-6,5%
W9	EXPRESIVIDAD	5,69%	4,52%	-20,6%	4,16%	3,98%	-4,4%	2,81%	2,19%	-22,2%
W10	IMPENETRABLE	5,22%	4,39%	-15,8%	5,98%	5,01%	-16,3%	3,51%	2,97%	-15,3%
W11	INEXPUGNABLE	4,29%	4,13%	-3,8%	4,53%	3,93%	-13,4%	2,58%	2,11%	-18,5%
W12	INFATIGABLE	4,60%	4,36%	-5,2%	4,29%	4,22%	-1,7%	2,42%	2,33%	-3,7%
W13	INGOBERNABLE	5,83%	5,24%	-10,1%	5,25%	4,69%	-10,7%	3,93%	3,29%	-16,1%
W14	MANSEDUMBRE	7,09%	6,48%	-8,5%	8,28%	7,28%	-12,1%	5,87%	5,00%	-14,8%
W15	ZAFARRANCHO	7,58%	6,51%	-14,2%	7,42%	6,50%	-12,4%	5,69%	4,82%	-15,2%
W16	ZARRAPASTROSA	8,59%	7,59%	-11,7%	7,20%	6,16%	-14,4%	5,57%	4,70%	-15,6%
Average over the 16 words		**5,52%**	**4,92%**	**-10,32%**	**5,56%**	**4.98%**	**-9,93%**	**3,73%**	**3,18%**	**-14,96%**

Table 8.3: Comparison of EERs obtained with reference and synthetic enrolment sets. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

8.4.2 IMPACT OF ENLARGING THE ENROLMENT SET WITH SYNTHETIC EXECUTIONS

This experiment assesses the impact of enlarging the original enrolment set with synthetic executions (enlarged enrolment set). The results obtained are shown in Tables 8.4, 8.5 and 8.6. As in the preceding experiment, for each word and metric, the difference with respect to the reference results is given as the percentage of increase/decrease (*Δ%*). Figs. 8.10 and 8.11 show the DET-plots for the words achieving the maximum and the minimum enhancements.

The reader will notice that although some particular differences exist, the results in this experiment are very close to the results yielded by the previous one. Thus, the enlarged enrolment sets outperform the original ones but virtually in the same measure that the synthetic enrolment sets did. Nevertheless, a difference that may be of importance does exist: there is no case where the enlarged enrolment sets perform worse than the original ones.

At the end of the section, Tables 8.7, 8.8 and 8.9 summarize the results obtained when two or more words are combined. Notice that the enhancements observed for single words persist when more than one word is considered, except for identification where the high rates obtained prevent further improvements.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	80,0%	83,7%	4,6%	81,4%	85,9%	5,6%	93,0%	94,9%	2,0%
W2	DELEZNABLE	72,1%	78,6%	9,0%	73,8%	78,4%	6,3%	88,4%	91,9%	4,0%
W3	DESAPROVECHAMIENTO	89,8%	92,1%	2,5%	89,3%	91,3%	2,2%	96,4%	97,1%	0,7%
W4	DESBRIZNAR	74,3%	77,8%	4,7%	71,6%	76,8%	7,2%	86,6%	90,0%	3,9%
W5	DESLUMBRAMIENTO	83,9%	88,3%	5,2%	84,2%	89,0%	5,7%	94,7%	96,3%	1,7%
W6	DESPEDAZAMIENTO	89,1%	91,2%	2,4%	84,4%	88,9%	5,4%	94,5%	95,8%	1,3%
W7	DESPRENDER	71,3%	76,6%	7,3%	67,7%	72,4%	6,9%	84,2%	89,3%	6,0%
W8	ENGUALDRAPAR	72,0%	76,9%	6,8%	71,2%	76,7%	7,8%	86,5%	90,2%	4,3%
W9	EXPRESIVIDAD	75,5%	81,4%	7,8%	79,3%	84,5%	6,6%	91,9%	94,8%	3,1%
W10	IMPENETRABLE	78,4%	83,8%	6,9%	73,1%	79,0%	8,0%	88,5%	91,6%	3,4%
W11	INEXPUGNABLE	82,3%	85,9%	4,4%	83,4%	87,7%	5,1%	94,5%	96,2%	1,8%
W12	INFATIGABLE	79,2%	82,8%	4,5%	79,5%	82,7%	3,9%	90,9%	93,2%	2,6%
W13	INGOBERNABLE	74,8%	79,6%	6,4%	81,3%	85,7%	5,4%	90,9%	93,5%	2,9%
W14	MANSEDUMBRE	63,4%	70,2%	10,7%	68,6%	75,2%	9,7%	84,1%	88,1%	4,7%
W15	ZAFARRANCHO	63,1%	70,2%	11,3%	65,2%	71,5%	9,7%	81,9%	85,6%	4,6%
W16	ZARRAPASTROSA	61,9%	70,4%	13,8%	67,3%	72,3%	7,5%	81,9%	85,7%	4,7%
Average over the 16 words		**75.4%**	**80,6%**	**6,8%**	**76.3%**	**81,1%**	**6,4%**	**89.3%**	**92,1%**	**3,2%**

Table 8.4: Comparison of IDRs obtained with original and enlarged enrolment sets. Positive increments (green) mean that the synthetic enrolment sets outperform the original ones.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	4,35%	3,53%	-18,9%	4,29%	3,95%	-8,0%	2,67%	2,31%	-13,5%
W2	DELEZNABLE	5,89%	5,39%	-8,5%	4,48%	3,99%	-11,0%	2,98%	2,65%	-11,1%
W3	DESAPROVECHAMIENTO	2,98%	2,84%	-4,9%	3,62%	3,07%	-15,1%	2,05%	1,70%	-17,0%
W4	DESBRIZNAR	5,10%	4,85%	-5,0%	5,77%	5,26%	-8,9%	3,59%	3,11%	-13,3%
W5	DESLUMBRAMIENTO	3,82%	3,36%	-12,0%	3,93%	3,78%	-3,8%	2,49%	2,16%	-13,3%
W6	DESPEDAZAMIENTO	3,39%	3,36%	-0,7%	4,47%	3,73%	-16,6%	2,84%	2,34%	-17,6%
W7	DESPRENDER	5,55%	5,01%	-9,8%	6,58%	5,92%	-10,0%	4,26%	3,72%	-12,7%
W8	ENGUALDRAPAR	5,76%	4,89%	-15,0%	6,33%	5,60%	-11,6%	4,18%	3,78%	-9,5%
W9	EXPRESIVIDAD	5,05%	4,20%	-16,7%	4,00%	3,69%	-7,6%	2,50%	2,04%	-18,4%
W10	IMPENETRABLE	4,90%	4,29%	-12,4%	5,55%	4,71%	-15,2%	3,25%	2,64%	-18,8%
W11	INEXPUGNABLE	4,05%	3,72%	-8,3%	4,18%	3,49%	-16,6%	2,45%	1,82%	-25,8%
W12	INFATIGABLE	4,38%	3,92%	-10,5%	4,04%	3,73%	-7,6%	2,24%	2,13%	-4,7%
W13	INGOBERNABLE	5,56%	4,97%	-10,6%	5,03%	4,13%	-18,0%	3,66%	3,02%	-17,4%
W14	MANSEDUMBRE	6,81%	6,05%	-11,2%	7,87%	6,80%	-13,6%	5,60%	4,55%	-18,7%
W15	ZAFARRANCHO	7,40%	6,20%	-16,1%	7,13%	6,28%	-11,8%	5,26%	4,52%	-14,2%
W16	ZARRAPASTROSA	8,22%	7,10%	-13,6%	6,81%	5,82%	-14,5%	5,37%	4,35%	-19,0%
Average over the 16 words		**5,20%**	**4,60%**	**-10,90%**	**5,25%**	**4,62%**	**-11,87%**	**3,46%**	**2,93%**	**-15,31%**

Table 8.5: Comparison of VEs obtained with reference and enlarged enrolment sets. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

WORD	TEXT	IN-AIR TRAJECTORIES (PEN-UP STROKES)			ON-SURFACE TRAJECTORIES (PEN-DOWN STROKES)			COMBINATION		
		REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
W1	BIODEGRADABLE	4,72%	3,89%	-17,6%	4,62%	4,37%	-5,5%	2,88%	2,41%	-16,3%
W2	DELEZNABLE	6,23%	5,57%	-10,6%	4,69%	4,14%	-11,8%	3,29%	2,80%	-14,9%
W3	DESAPROVECHAMIENTO	3,28%	3,06%	-6,7%	3,91%	3,26%	-16,5%	2,50%	1,82%	-27,1%
W4	DESBRIZNAR	5,61%	5,06%	-9,8%	5,99%	5,55%	-7,4%	3,75%	3,31%	-11,8%
W5	DESLUMBRAMIENTO	4,12%	3,67%	-10,8%	4,22%	4,08%	-3,4%	2,74%	2,33%	-15,1%
W6	DESPEDAZAMIENTO	3,60%	3,59%	-0,2%	4,62%	3,91%	-15,4%	3,00%	2,50%	-16,6%
W7	DESPRENDER	5,69%	5,22%	-8,2%	6,88%	6,33%	-7,9%	4,59%	4,14%	-9,8%
W8	ENGUALDRAPAR	6,09%	5,22%	-14,3%	6,86%	5,80%	-15,4%	4,51%	4,14%	-8,2%
W9	EXPRESIVIDAD	5,69%	4,45%	-21,8%	4,16%	3,82%	-8,2%	2,81%	2,18%	-22,3%
W10	IMPENETRABLE	5,22%	4,39%	-15,8%	5,98%	4,85%	-19,0%	3,51%	2,97%	-15,3%
W11	INEXPUGNABLE	4,29%	4,14%	-3,5%	4,53%	3,75%	-17,3%	2,58%	2,10%	-18,7%
W12	INFATIGABLE	4,60%	4,22%	-8,3%	4,29%	4,00%	-6,8%	2,42%	2,33%	-3,9%
W13	INGOBERNABLE	5,83%	5,25%	-10,0%	5,25%	4,61%	-12,0%	3,93%	3,34%	-14,8%
W14	MANSEDUMBRE	7,09%	6,43%	-9,2%	8,28%	7,19%	-13,1%	5,87%	4,99%	-15,0%
W15	ZAFARRANCHO	7,58%	6,62%	-12,7%	7,42%	6,42%	-13,4%	5,69%	4,91%	-13,6%
W16	ZARRAPASTROSA	8,59%	7,58%	-11,8%	7,20%	6,13%	-14,9%	5,57%	4,61%	-17,2%
Average over the 16 words		**5,52%**	**4,90%**	**-10,71%**	**5,56%**	**4,89%**	**-11,77%**	**3,73%**	**3,18%**	**-15,04%**

Table 8.6: Comparison of EERs obtained with reference and enlarged enrolment sets. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

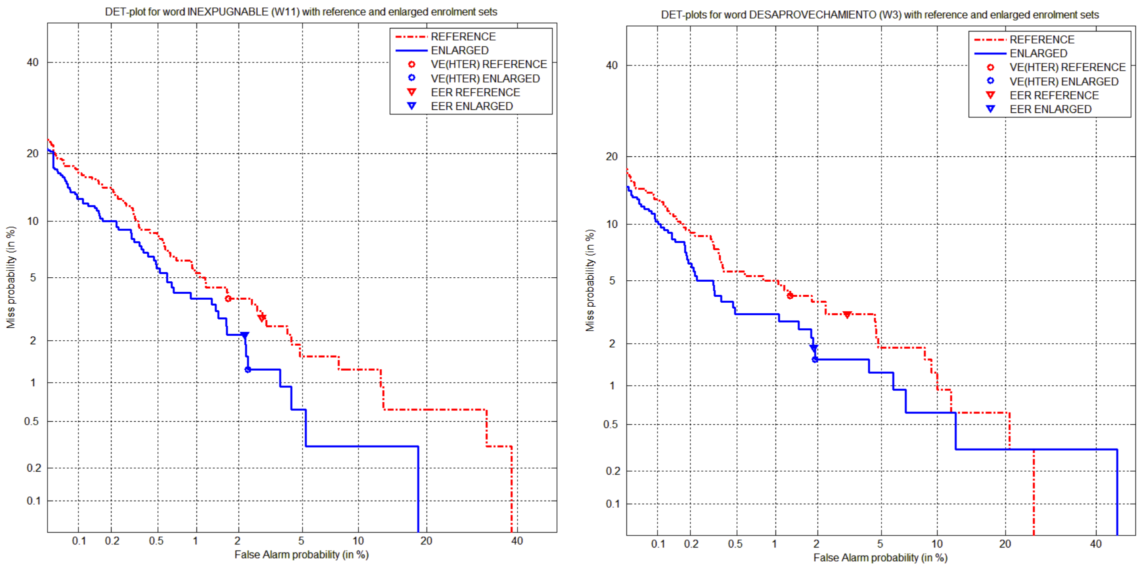


Figure 8.10: DET-plots for the words achieving the maximum enhancements when enlarged enrolment sets are considered (INEXPUGNABLE, -25.8% in VE, and DESAPROVECHAMIENTO, -27.1% in EER).

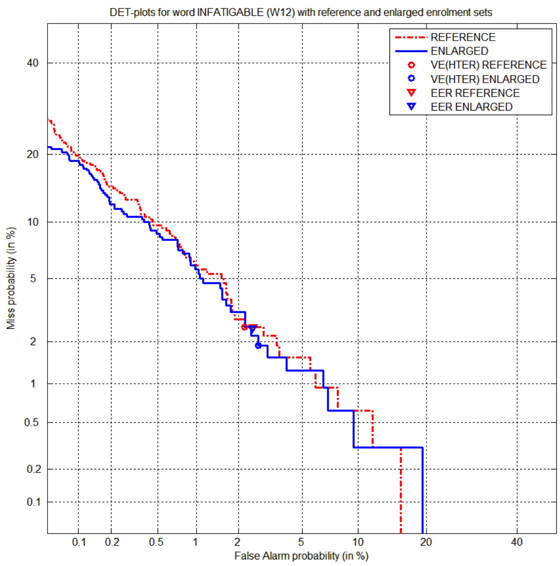


Figure 8.11: DET-plot for the word achieving the minimum enhancement when enlarged enrolment sets are considered (INFATIGABLE, -4.7% in VE, and –3.9% in EER).

NUMBER OF WORDS	WORST			AVERAGE			BEST		
	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
1	81.9%	85.9%	4,9%	89.3%	92.1%	3,2%	96.4%	97.1%	0,7%
2	93.8%	95.6%	1,9%	97.7%	98.6%	0,9%	99.3%	99.5%	0,2%
3	97.9%	98.5%	0,6%	99.1%	99.4%	0,3%	99.8%	99.8%	0,0%
4	98.6%	99.1%	0,5%	99.5%	99.6%	0,1%	99.9%	99.9%	0,0%

Table 8.7: Comparison of IDRs obtained with reference and enlarged enrolment sets when different words are combined. Positive increments (green) mean that the synthetic enrolment sets outperform the original ones.

NUMBER OF WORDS	WORST			AVERAGE			BEST		
	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
1	5.60%	4.60%	-17,9%	3.46%	2.95%	-14,5%	2.05%	1.68%	-18,0%
2	3.79%	3.05%	-19,5%	1.81%	1.44%	-20,4%	1.00%	0.80%	-20,0%
3	2.78%	2.22%	-20,1%	1.29%	1.00%	-22,5%	0.61%	0.50%	-18,0%
4	2.22%	1.57%	-29,3%	1.04%	0.79%	-24,0%	0.51%	0.38%	-25,5%

Table 8.8: Comparison of VEs obtained with reference and enlarged enrolment sets when different words are combined. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

NUMBER OF WORDS	WORST			AVERAGE			BEST		
	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)	REFERENCE ENROLMENT SET	ENLARGED ENROLMENT SET	Δ (IN % WRP. REFERENCE)
1	5.87%	5.00%	-14,8%	3.73%	3.18%	-14,9%	2.42%	1.81%	-25,2%
2	4.00%	3.20%	-20,0%	2.02%	1.65%	-18,3%	1.09%	0.92%	-15,6%
3	3.04%	2.47%	-18,8%	1.47%	1.17%	-20,4%	0.71%	0.61%	-14,1%
4	2.36%	1.88%	-20,3%	1.20%	0.95%	-20,8%	0.58%	0.47%	-19,0%

Table 8.9: Comparison of EERs obtained with reference and enlarged enrolment sets when different words are combined. Negative increments (green) mean that the synthetic enrolment sets outperform the original ones.

8.5 CONCLUSIONS

The results yielded by the experiments performed to assess the impact of enlarging the enrolment sets with synthetic samples generated with the proposed method may be deemed positive and very promising.

Firstly, it has been shown that the synthetic enlargement of the enrolment sets of a writer recognition system based on handwritten short sequences of text can lead to a noticeable improvement in its recognition performance, especially in the verification task. Although an improvement of this kind has already been reported in the signature field (e.g. [168]), this is the first time, to the authors' best knowledge, that such a possibility is reported regarding a non-signature-based schema.

Secondly the results also show the effectiveness of the proposed method. Regarding the verification task, they are quite promising: in all cases the impact of enlarging the enrolment sets is positive (lower VE and EER), with the improvement ranging from a humble -3.9% (EER, word INFATIGBLE) to a noticeable -27.1% (EER, word DESAPROVECHAMIENTO). The average improvement is about -15%, both for VE and EER. With regard to the identification task, the impact is also always positive but of a more humble magnitude: from 0.7% to 6% (words DESAPROVECHAMIENTO and DESPRENDER) with an average of 3.2%. When 2 or more words are combined, similar enhancements are observed. Fig. 8.12 concisely summarizes the results.

The enhancement rates reported in the experiments have to be analyzed within the context in which they have been obtained. It has to be taken into account that the recognition system already yielded high identification rates and low verification errors without the addition of synthetic samples. It is the magnitude of the improvements along with the fact that in some cases there was little room for improvement what make us think that the results are quite encouraging.

The fact that the performance of the recognition system is improved by the synthetic enlargement of the enrolment sets may also have a positive impact on collectability and acceptability, two features to be taken into account when evaluating a biometric system. This possibility would help in keeping enrolment easy, little invasive and performed quickly.

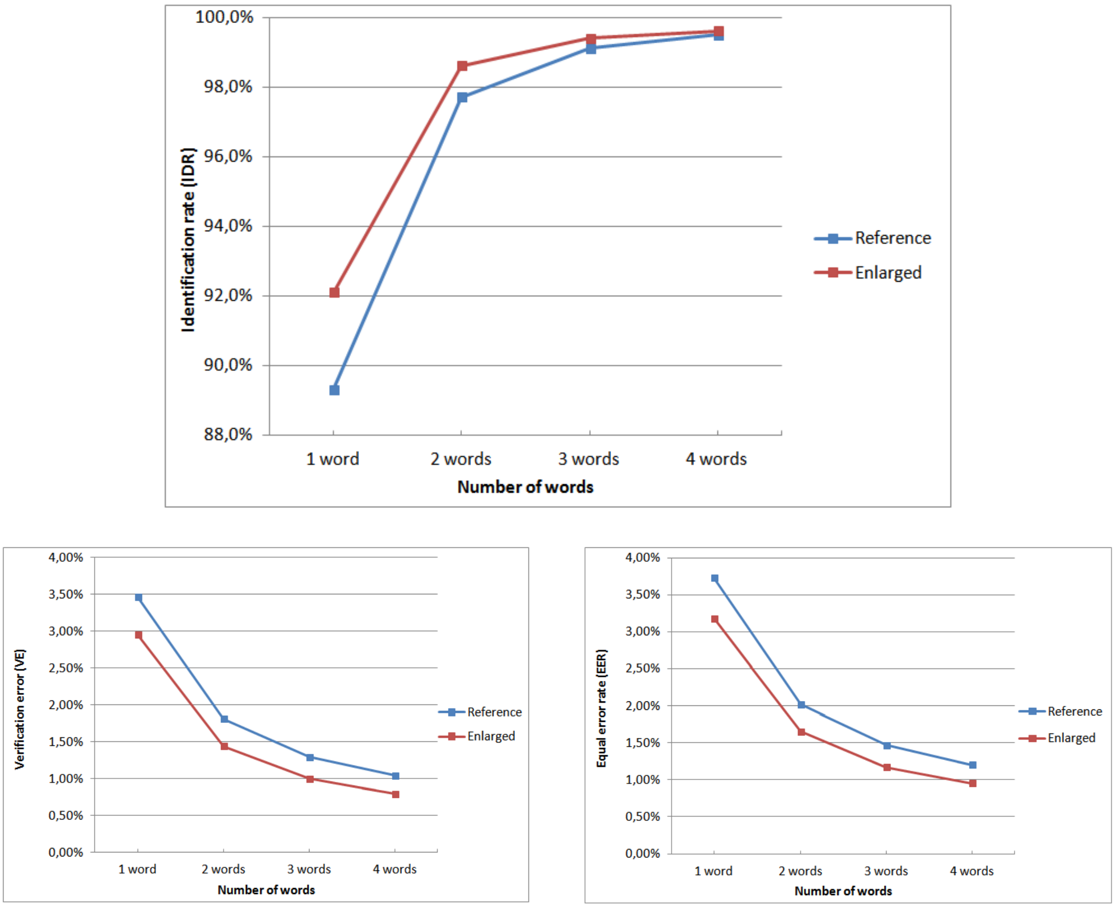


Figure 8.12: Comparison of identification (top) and verification (bottom) performances for the reference and the enlarged enrolment sets, as a function of the number of words considered.

9

CONCLUSIONS AND A BRIEF LOOK TO THE FUTURE

This is the final chapter of this doctoral dissertation. It contains two sections devoted to the main conclusions drawn from the research reported on in the previous chapters, and to the future lines of research that may stem from the work done.

9.1 CONCLUSIONS

Several conclusions can be drawn from the results presented in the preceding chapters. All main conclusions are related to the objectives enumerated in the introduction and to the contributions that materialize the accomplishment of those objectives. Fig. 9.1 is a concise summary of these aspects of the work: its motivation, objectives, contributions and main conclusions. In the forthcoming subsections each main conclusion will be elaborated in more detail, together with other relevant conclusions.

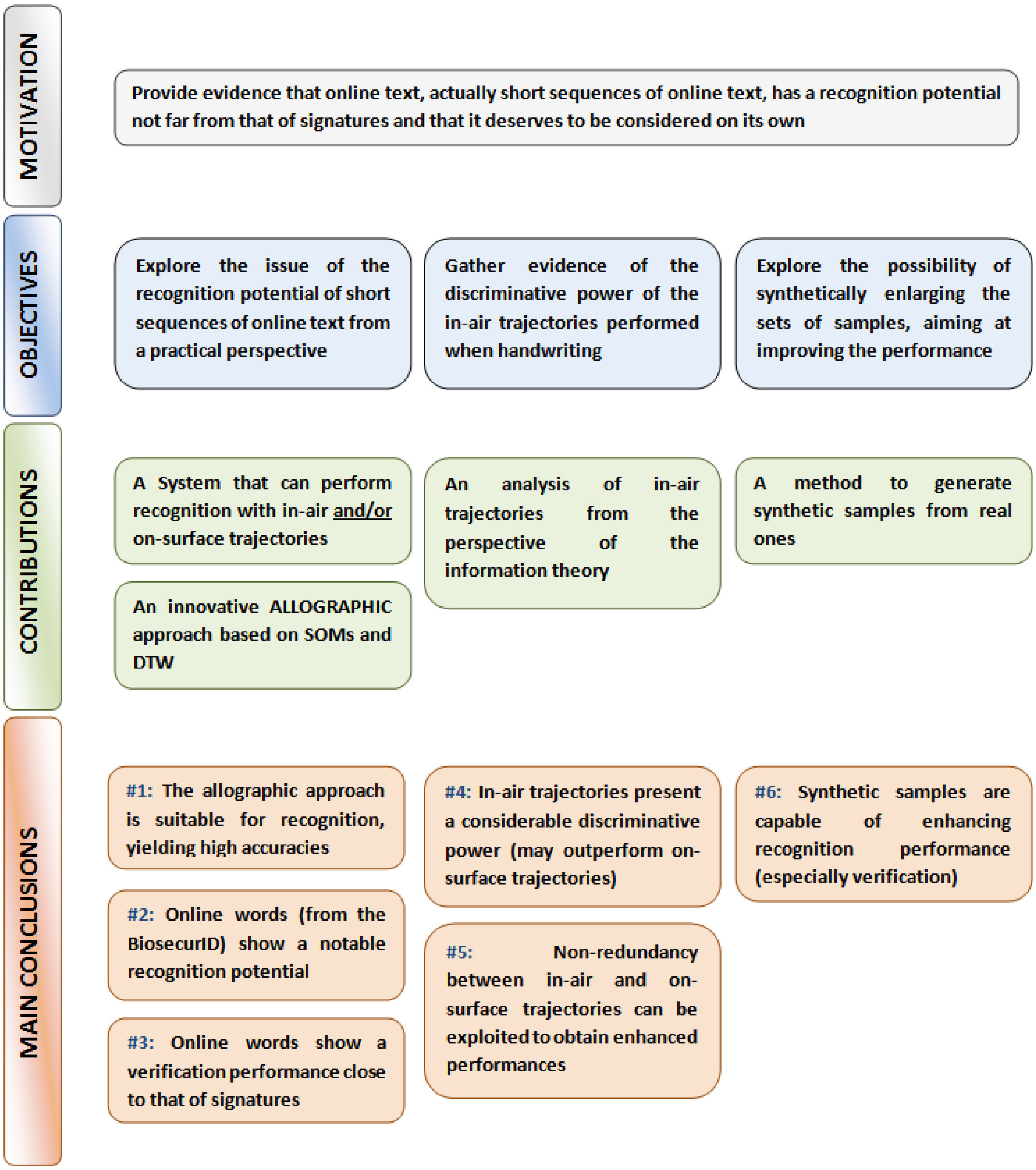


Figure 9.1: Motivation, objectives, contributions and conclusions of this dissertation.

9.1.1 REGARDING THE DEVELOPED SYSTEM AND THE APPROACH FOLLOWED

The objectives of this dissertation required the existence of a recognition system capable of using the words in the BiosecurID as its *raw material*. During the first stages of the research, some approaches that had obtained good results in the signature-verification field were considered (VQ, DTW, HMMs) and experimentally assessed but the performances obtained were not even slightly comparable to those attained with signatures. Thus it was decided to design a new recognition system *from scratch*. This new recognition system had but a few significant requirements: it had to be capable of using both in-air and on-surface trajectories in order to draw conclusions about their relative recognition potentials, and it had to attain reasonable accuracies.

The recognition system eventually developed fulfils the aforementioned requirements. It addresses the issue of text-based writer recognition from an allographic perspective (see Fig.

1.1 in chapter 1), a perspective that to the author's best knowledge, had never before been considered for text-dependent online approaches. The following characteristics of the system can be highlighted:

- It is **based on an innovative idea: the combined use of topology-preserving catalogues of strokes (codebooks), obtained from Self-Organizing Maps, and Dynamic Time Warping**. Thanks to the use of these catalogues, Dynamic Time Warping is applied to very short and simple sequences, substantially reducing the overall computational burden[9].

- It can perform the two tasks involved in recognition: identification and verification.

- In-air and on-surface trajectories are considered separately therefore their recognition potentials can be assessed independently. The yielded results (dissimilarity measures) can be combined in order to obtain full-word measures.

- Finally, **the system successfully addresses the issue of text-based writer recognition from an allographic perspective**.

When it comes to the accuracy level achieved when experimenting with words from the BiosecurID database, it must be said that it exceeds the initial expectations:

MAIN CONCLUSION #1: The allographic approach has proven suitable to perform writer recognition in an online and text-dependent context, achieving a high accuracy level.

Regarding identification, the system compares positively to other word-based recognition systems (see Table 9.1). As for verification, the experiments have shown that, with single words, a performance not far from that of signature-based system can be attained. A more elaborated comparison with signature verification approaches is provided in the next subsection.

[9] Dynamic Time Warping has a quadratic time and space complexity ($O(N^2)$ where N is the length of the sequences to be compared).

FIELD	SOURCE	RECOGNITION PERFORMANCE (IDR)		
		WORST	AVERAGE	BEST
ONLINE	**This dissertation**. Original enrolment set. One Word	81.9% (ZAFARRANCHO)	89.3%	96.4% (DESAPROVECHAMIENTO)
	This dissertation. Extended enrolment set. One word	85.6% (ZAFARRANCHO)	92.1%	97.1% (DESAPROVECHAMIENTO)
	Zuo et al. [111]	86.5% with 1 Chinese word.		
	Chapran [51]	95% (with 25 repetitions of 1 word)		
OFFLINE	Zois et al. [102]	97% (with 45 repetitions of Greek word *χαρακτηριστικό*)		
	Zhang and Srihari [78]	49% (Word *referred*)		
	Tomasi et al. [79]	67% (Word *Grant*)		

Table 9.1: Identification performances obtained in the experiments reported in this dissertation and performances reported in relevant references.

From the experiments reported in chapter 7 other relevant conclusions regarding the proposed recognition system can be drawn:

(a) **The system can be bootstrapped with data from a limited number of writers** (e.g. just 10 writers).

(b) **The system can be bootstrapped with data not originating from the users that will be enrolled (exocatalogues), without any remarkable loss in accuracy**, neither in identification nor in verification.

(c) **Although accuracy is affected by the size of the catalogues, not especially large ones (e.g. 100 or 150 units) achieve good performances**.

(d) **The limited number of samples required to achieved the reported performances, makes the enrolment phase, for this particular system, easily and quickly performed, thus benefitting acceptability and collectability.**

9.1.2 REGARDING THE RECOGNITION POTENTIAL OF SHORT SEQUENCES OF ONLINE TEXT AND ITS COMPARISON WITH THE RECOGNITION POTENTIAL OF SIGNATURES

In the opening chapter, the main motivation of this dissertation was stated as that of providing evidence that short sequences of online text had a recognition potential not far from that of signatures and that they deserved to be considered on their own. This motivation translated into an objective of practical nature: a recognition system had to be built in order to explore the recognition potential of online text. Conclusions regarding the system itself have centred the preceding subsection; when it comes to its particular application to the online words contained in the BiosecurID database, the following main conclusion can be drawn:

MAIN CONCLUSION #2: **reported experimental results clearly sustain the claim that the online words in the BiosecurID do have a notable recognition potential.**

Nevertheless, the original question still remains to be answered: how do short sequences of text compare to signatures? Although much care has to be taken when comparing the performance of different biometric modalities because of their inherent differences and the great variability of the conditions under which they are tested (number of writers, number of samples per writer, number of sessions during which the samples were collected, features, availability and use of in-air information, presence of skilled forgeries, global or local threshold, etc.) [118], the figures shown in Table 9.2 substantiate the assertion that constitutes the third main conclusion:

MAIN CONCLUSION #3: **the online words in the BiosecurID database show a verification performance that do not lie much below the performances reported for today's state of the art signature verifications methods.**

SOURCE	VERIFICATION PERFORMANCE (EER)		
	WORST	AVERAGE	BEST
This dissertation. Original enrolment set. One Word	5.87% (MANSEDUMBRE)	3.73%	2.42% (INFATIGABLE)
This dissertation. Extended enrolment set. One word	4.99% (MANSEDUMBRE)	3.18%	1.82% (DESAPROVECHAMIENTO)
This dissertation. Original enrolment set. Two words	4.00%	2.02%	1.09
This dissertation. Original enrolment set. Three words	3.04%	1.47%	0.71%
This dissertation. Original enrolment set. Four words	2.36%	1.20%	0.58%
SVC2004 with random forgeries [69]	1.70%		
BSEC'2009 with random forgeries, multisession [142]	1.37%		

Table 9.2: Verification performances obtained in the experiments reported in this dissertation and state-of-the-art performances achieved in signature competitions.

Words such as DESAPROVECHAMIENTO or INFATIGABLE and the combination of two or more words yield verification errors that may compare positively to verification accuracies reported for signature-based systems.

Another relevant conclusion accompanies the last two main ones:

(e) **Combining two or more words results in enhanced performances:** with four words almost 100% mean identification rate is achieved. Also with four words, an outstanding best EER of 0.58% has been obtained.

It is worth noticing that the performance experiments described in this dissertation have been carried out with a very limited number of repetitions for training, constrained by the number of repetitions provided by the database: only four repetitions were available and one of them had to be used for testing. Should there have been an ampler number of repetitions, the accuracies might have been higher. It is also remarkable that these results have been attained in the few years devoted to this particular research, a short period of time when compared to the time (and research teams) required to attain today's state-of-the-art signature verification accuracies.

Taking into account the particular context where the experimentation took place, other interesting questions may arise: to what extent can the results be extrapolated to different contexts such as other databases and/or to lowercase sequences? Although for the time being it is not possible to give an assertive answer to such questions, (and therefore no relevant conclusions can be drawn) the following reflections may cast some light onto the issue: first, the proposed recognition system makes very few assumptions regarding the information on which it operates, other than it must adhere to the SVC format [69]. Also, it is worth noticing that many modern acquisition devices can gather the same information (x,y, pressure and writing angles). More *humble* devices may not gather the writing angles, but their overall weight is not high (18% for pen-down strokes and 13% for pen-up ones). Even if the information regarding the in-air trajectories were not available, the degradation of the recognition performance could still be bearable. Secondly, neither the system has been specifically tuned to work with uppercase sequences, nor lowercase sequences cannot be processed the same way uppercase ones have been processed. Therefore nothing prevents the application of the system to lowercase, or even mixed-case, sequences. Whether the performance would be similar or not, it would be expected that under similar conditions, lowercase sequences would yield comparable or even better results since it is believed that recognition based on uppercase text is a more challenging problem.

9.1.3 REGARDING THE RECOGNITION POTENTIAL OF IN-AIR TRAJECTORIES

The theoretical approach reported in chapter 5 firmly suggested that in-air trajectories contained a considerable amount of information and that a significant portion of it might be non-redundant with respect to the information contained in on-surface trajectories. The information analysis of both types of trajectories rendered the following conclusions:

(f) **Except for pressure, the amount of information in each one of the features considered is virtually the same in on-surface and in-air trajectories.**

(g) **Although both types of trajectories share information, there is a significant amount of non-redundancy.**

The recognition performances achieved by the proposed system (see chapter 7) substantiated from a practical perspective what the theoretical approach had suggested:

MAIN CONCLUSION #4: **In-air trajectories have a considerable discriminative power**, sometimes matching or even outperforming the discriminative power of on-surface trajectories.

MAIN CONCLUSION #5: **Non-redundancy between in-air and on surface trajectories can be exploited in order to obtain enhanced performances when both types of trajectories are considered**.

9.1.4 REGARDING THE METHOD TO GENERATE SYNTHETIC SAMPLES AND ITS USE TO ENLARGE THE ENROLMENT SETS

The third objective of this dissertation, ensuing from the fact that the number of samples per writer provided by the BiosecurID database is rather limited, was that of exploring the possibility of enhancing the recognition performance of the proposed system by enlarging the enrolment sets with synthetically generated samples. **The method proposed in chapter 8 fulfils this objective and yields promising results**: with one word, the enhancement in verification ranges from a humble 3.7% to a noteworthy 27.5% with an average of 15%. In the case of identification, enhancement ranges from 0.7% to 5.9% with an average of 3.2%. As identification rates were already quite high, room for improvement was scarce. As far as the author knows, this is the first attempt to apply sample duplication ever reported in the field of text-based writer recognition.

MAIN CONCLUSION #6: ***synthetically generated samples are capable of enhancing the recognition performance, especially in verification.***

From this main conclusion, another one can be drawn:

(h) **Having a method that improves the performance without requiring the writer to provide an extra amount of samples may have a positive impact on collectability and acceptability.**

It is also important to notice that the enhancement achieved thanks to synthetic samples does not interfere with the enhancement achieved when combining two or more words. On the contrary:

(i) **The enlargement of the enrolment sets and the combination of two or more words strengthen each other providing a higher degree of performance enhancement.**

Although the proposed synthesis method has proved its effectiveness in a particular recognition system, it could also be applied to other text-based online recognition schemas or adapted to work with them, since **it is based on simple yet effective and easy to apply ideas: stroke separation, stroke alignment and weighted stroke averaging**.

9.2 FUTURE WORK

Except for signature-based systems, text-dependent online writer recognition, the field in which this dissertation is circumscribed, has not favoured the attention and the efforts of a considerable number of researchers. Therefore, the research on which we are reporting is but a starting point from which several other lines of research may stem. The following is a non-exhaustive list of issues that may deserve some attention in the short- or mid-term:

- The proposed system could be applied to short uppercase sequences and to sequences of digits. Also, the recognition potential of the in-air trajectories within these sequences could be analyzed, from a theoretical and from a practical perspective.

- To present, only the features directly gathered by the acquisition device have been taken into account. But new features can be computed from the aforementioned and their relevance could be assessed. Velocity and acceleration (first and second order derivatives), often used in online signature verification, could be a starting point. Others may be, for instance, the initial set of features considered in [51] (shown in Table 4.12). The set of features to consider could be further extended with the addition of features of global nature, extracted from the whole sequence or in a per-stroke basis. Also, other pre-processing schemas (see section 6.2.1) could be considered.

- The work presented in this dissertation gives evidence of the usefulness of in-air trajectories since their combination with on-surface information consistently renders improved performances. Is in-air information as relevant for signature as it appears to be for text? If so, could in-air trajectories be used, in a similar way, to improve the performance of online signature verification systems? Most of these systems already exploit the information within the in-air trajectories since it is considered that, being invisible to the human eye, these trajectories are difficult to forge therefore giving an extra resilience to the methods that do not discard them [118]. Some authors just use them as a way to perform segmentation, while others take them fully into account, but do not separate them from on-surface ones. Nevertheless, the real impact of in-air trajectories on the performance of signature verification systems is an issue that has not received much consideration and that is the reason why it could become an interesting line of future research. Finally it is important to recall that not all signatures provide in-air information since it is not unusual for some signers to perform their signatures without ever lifting the writing device. What is more, there may be fewer pen-up strokes in a signature than in a long word (such as the ones in the BiosecurID database). Obviously, the potential of in-air trajectories may be highly conditioned by these two issues.

And finally, there is a more ambitious line of research on which we have already started working: health. Security issues have virtually monopolized all research on biometrics. But biometrics can go beyond security and land on other fields, with health being one of such fields. Preliminary results on the analysis of online handwriting and drawings, aiming at the characterization and diagnose of neurodegenerative conditions such as Alzheimer's disease and other forms of dementia, are quite encouraging though far from conclusive. In-air trajectories may end up playing an important role in the aforementioned analysis since they seem to be particularly sensitive to the writer's health state.

References

[1] J Fierrez-Aguilar, J Galbally, J Ortega-Garcia, M Freire, F Alonso-Fernandez, D Ramos, et al. BiosecurID: a multimodal biometric database, Pattern Analysis & Applications. 13 (2010).

[2] Merriam-Webster incorporated, Dictionary and Thesaurus - Merriam-Webster Online, URL: http://www.merriam-webster.com/ (last accessed: 15-march-2012).

[3] D Harper, Online Etymology Dictionary, URL: http://www.etymonline.com/ (last accessed: 15-march-2012).

[4] M Faundez-Zanuy, E Sesa-Nogueras, J Roure-Alcobé, E Monte-Moreno, J Garré-Olmo, K Lopez-de-Ipiña, et al., Aplicaciones biométricas más allá de la seguridad, Actas de las VI Jornadas de Reconocimiento Biometrico de Personas (JRBP 2012). (2012) 25-43.

[5] S Nanavaty, M Thiene, R Nanavati, Biometrics: Identify Verification in a Networked World, Wiley Computer Publising (John Wiley & Sons, Inc.) 2002.

[6] ST Kent, LI Millett, Who goes there?: authentication through the lens of privacy, The National Academies Press, Washington, DC, 2003.

[7] LW James. Fundamentals of biometric authentication technologies, International Journal of Image and Graphics (IIIG). 1 (2001) 93-113.

[8] AK Jain, A Ross, S Prabhakar. An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology. 14 (2004) 4-20.

[9] SZ Li, AK Jain, Handbook of face recognition, Springer 2011.

[10] DA Reynolds, An overview of automatic speaker recognition technology, 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 4 (2002) 4072-4075.

[11] D Maltoni, D Maio, AK Jain, S Prabhakar, Handbook of fingerprint recognition, Springer-Verlag New York Inc 2009.

[12] RL Zunkel, Hand Geometry Based Verification, in: Jain AK, Bolle R, Pankanti S (Eds.), Biometrics, Springer US, 2002, pp. 87-101.

[13] A Badawi. Hand vein biometric verification prototype: A testing performance and patterns similarity, International Journal of Biomedical Sciences. (2007) 141-148.

[14] J Daugman. Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons, Proceedings of the IEEE. 94 (2006) 1927-1935.

[15] L Schomaker, Advances in Writer Identification and Verification, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). 2 (2007) 1268-1273.

[16] Amazingdata.com, A closer look at the eye iris and pupil pics pictures photos, URL: http://amazingdata.com/a-closer-look-at-the-eye-iris-and-pupil-pics-pictures-photos/ (last accessed: 15-march-2012).

[17] Webvision, Simple Anatomy of the Retina – Webvision, URL: http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/ (last accessed: 15-march-2012).

[18] V Bouletreau, N Vincent, R Sabourin, H Emptoz, Handwriting and signature: one or two personality identifiers? Proceedings of the 14th International Conference onPattern Recognition. 2 (1998) 1758-1760 vol.2.

[19] F Monrose, AD Rubin. Keystroke dynamics as a biometric for authentication, Future Generation Computer Systems. 16 (2000) 351-359.

[20] JE Boyd, JJ Little. Biometric gait recognition, Advanced Studies in Biometrics. (2005) 15-26.

[21] D Gafurov, A survey of biometric gait recognition: Approaches, security and challenges, Preceedings of the NIK-2007 conference. (2007).

[22] P Yan, KW Bowyer. Biometric recognition using 3D ear shape, IEEE Transactions on Pattern Analysis and Machine Intelligence. 29 (2007) 1297-1308.

[23] AK Jain, RM Bolle, S Pankanti, Biometrics: Personal Identification in a Networked Society, Springer 2005.

[24] AK Jain, A Ross. Multibiometric systems, Communications of the ACM. 47 (2004) 34-40.

[25] A Ross, A Jain. Information fusion in biometrics, Pattern Recognition Letters. 24 (2003) 2115-2125.

[26] T Fawcett. An introduction to ROC analysis, Pattern Recognition Letters. 27 (2006) 861-874.

[27] A Martin, G Doddington, T Kamm, M Ordowski, M Przybocki, The DET curve in assessment of detection task performance, European Speech Processing Conference Eurospeech. 4 (1997) 1895-1898.

[28] Medscape Reference, Hand Anatomy, URL: http://emedicine.medscape.com/article/1285060-overview#a1 (last accessed: 15-march-2012).

[29] R Plamondon, FJ Maarse. An evaluation of motor models of handwriting, IEEE Transactions on Systems, Man and Cybernetics. 19 (1989) 1060-1072.

[30] L Schomaker, G Van Galen. Computer models of handwriting, Computational psycholinguistics. (1996) 386-420.

[31] N Wiener, Cybernetics or control and communication in the animal and the machine, MIT press 1973.

[32] L Schomaker, Simulation and recognition of handwriting movements, Doctoral Dissertation at Nijmegen University, The Netherlands, (1991).

[33] C Vielhauer, Biometric user authentication for IT security: from fundamentals to handwriting, Springer-Verlag New York Inc 2005.

[34] J Hollerbach. An oscillation-theory of handwriting, Biological cybernetics. 39 (1981) 139-156.

[35] LRB Schomaker, A Thomassen, H Teulings, A computational model of cursive handwriting, in: Plamondon R, Suen CY, Simmer ML (Eds.), Computer recognition and human production of handwriting, World Scientific, 1989, pp. 153-177.

[36] R Plamondon. A kinematic theory of rapid human movements Part I. Movement representation and generation, Biological Cybernetics. 72 (1995) 295-307.

[37] R Plamondon. A kinematic theory of rapid human movements part II. Movement time and control, Biological Cybernetics. 72 (1995) 309-320.

[38] R Plamondon, M Djioua. A multi-level representation paradigm for handwriting stroke generation, Human Movement Science. 25 (2006) 586-607.

[39] RA Huber, AM Headrick, Handwriting identification: facts and fundamentals, CRC Press 1999.

[40] KE Forbes, MF Shanks, A Venneri. The evolution of dysgraphia in Alzheimer’s disease, Brain research bulletin. 63 (2004) 19-24.

[41] J Neils-Strunjas, K Groves-Wright, P Mashima, S Harnish. Dysgraphia in Alzheimer's disease: a review for clinical and research purposes, Journal of speech, language, and hearing research (JSLHR). 49 (2006) 1313-1330.

[42] P Werner, S Rosenblum, G Bar-On, J Heinik, A Korczyn. Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment, The journals of gerontology.Series B, Psychological sciences and social sciences. 61 (2006) 228-236.

[43] F Asıcıoglu, N Turan. Handwriting changes under the effect of alcohol, Forensic science international. 132 (2003) 201-210.

[44] JG Phillips, RP Ogeil, F Müller. Alcohol consumption and handwriting: A kinematic analysis, Human Movement Science. 28 (2009) 619-632.

[45] RG Foley, A Lamar Miller. The effects of marijuana and alcohol usage on handwriting, Forensic science international. 14 (1979) 159-164.

[46] O Tucha, S Walitza, L Mecklinger, D Stasik, T Sontag, KW Lange. The effect of caffeine on handwriting movements in skilled writers, Human Movement Science. 25 (2006) 523-535.

[47] S Rosenblum, S Parush, P Weiss. The in Air Phenomenon: Temporal and Spatial Correlates of the Handwriting Process, Perceptual & Motor Skills. 96 (2003) 933.

[48] F Viñals Carrera, ML Puente Balsells, Grafología criminal, Herder, Barcelona, 2009.

[49] ML Puente Balsells, F Viñals Carrera, Grafología y ciencia: validación con ciento cincuenta tesis doctorales, Editorial UOC, Barcelona, 2010.

[50] SN Srihari, SH Cha, H Arora, S Lee. Individuality of handwriting, Journal of forensic sciences. 47 (2002) 856-872.

[51] J Chapran. Biometric Writer Identification: Feature Analysis and Classification, International Journal of Pattern Recognition & Artificial Intelligence. 20 (2006) 483-503.

[52] A Morales, MA Ferrer, CM Travieso, JB Alonso, About user acceptance in hand, face and signature biometric systems, Proceedings of the Workshop de tecnologías multibiométricas para la identificación de personas (1er WTM-IP). (2010) 14-17.

[53] Wacom, Intuos4 Small Pen Tablet | Wacom Americas, URL: http://www.wacom.com/en/Products/Intuos/Intuos4Small.aspx (last accessed: 15-march-2012).

[54] U- Marti, H Bunke, A full English sentence database for off-line handwriting recognition, Proceedings of the Fifth International Conference on Document Analysis and Recognition ICDAR '99. (1999) 705-708.

[55] U- Marti, H Bunke. The IAM-database: an English sentence database for offline handwriting recognition, International Journal on Document Analysis and Recognition. 5 (2002) 39-46.

[56] FKI: Research Group on Computer Vision and Artificial Intelligence IAM, University of Bern, IAM Handwriting Database — Computer Vision and Artificial Intelligence 2012.

[57] L Schomaker, L Vuurpijl, Forensic writer identification: A benchmark data set and a comparison of two systems (internal report for the Netherlands Forensic Institute), (2000).

[58] I Guyon, L Schomaker, R Plamondon, M Liberman, S Janet, UNIPEN project of on-line data exchange and recognizer benchmarks, Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 2-Conference B: Computer Vision & Image Processing. 2 (1994) 29-33 vol. 2.

[59] The UNIPEN Project, The UNIPEN Project, URL: http://hwr.nici.kun.nl/unipen/index-noframe.html (last accessed: 15-march-2012).

[60] E Augustin, M Carré, E Grosicki, JM Brodin, E Geoffrois, F Preteux, RIMES evaluation campaign for handwritten mail processing, 231 (2006) 235.

[61] The RIMES database, The RIMES database, URL: http://www.rimes-database.fr (last accessed: 15-march-2012).

[62] A Bensefia, A Nosary, T Paquet, L Heutte, Writer identification by writer's invariants, Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition. (2002) 274-279.

[63] C Viard-Gaudin, PM Lallican, S Knerr, P Binter, The IRESTE On/Off (IRONOFF) dual handwriting database, Proceedings of the Fifth International Conference on Document Analysis and Recognition ICDAR '99. (1999) 455-458.

[64] Ecole Polytechnique de l'Université de Nantes, Database IRONOFF URL: http://www.irccyn.ec-nantes.fr/~viardgau/IRONOFF/IRONOFF.html (last accessed: 15-march-2012).

[65] S Strassel, Linguistic resources for Arabic handwriting recognition, Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt. (2009).

[66] Honggang Zhang, Jun Guo, Guang Chen, Chunguang Li, HCL2000 - A Large-scale Handwritten Chinese Character Database for Handwritten Character Recognition, 10th International Conference on Document Analysis and Recognition, ICDAR '09. (2009) 286-290.

[67] DH Kim, YS Hwang, ST Park, EJ Kim, SH Paek, SY Bang, Handwritten Korean character image database PE92, Proceedings of the Second International Conference on Document Analysis and Recognition. (1993) 470-473.

[68] U Bhattacharya, BB Chaudhuri, Databases for research on recognition of handwritten characters of Indian scripts, Eighth International Conference on Document Analysis and Recognition, 2005. Proceedings. (2005) 789-793 Vol. 2.

[69] D Yeung, H Chang, Y Xiong, S George, R Kashi, T Matsumoto, et al., SVC2004: First International Signature Verification Competition, In Proceedings of the International Conference on Biometric Authentication (ICBA), Hong Kong. (2004) 16-22.

[70] B Itay, B Isaac, K Klara, D Itshak. Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. International journal on document analysis and recognition. 9 (2007) 89-99.

[71] K Steinke, M Gehrke, R Dzido, Recognition of Humboldt's Handwriting in Complex Surroundings, International Conference on Frontiers in Handwriting Recognition (ICFHR 2010). (2010) 553-558.

[72] A Fornes, J Llados, G Sanchez, X Otazu, H Bunke. A combination of features for symbol-independent writer identification in old music scores. International journal on document analysis and recognition. 13 (2010) 243-259.

[73] M Wirotius, A Seropian, N Vincent, Writer identification from gray level distribution, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003). (2003) 1168-1172.

[74] A Brink, M Bulacu, L Schomaker, How much handwritten text is needed for text-independent writer verification and identification, 19th International Conference on Pattern Recognition, ICPR 2008. (2008) 1-4.

[75] B Zhang, SN Srihari, S Lee, Individuality of Handwritten Characters, Proceedings of the Seventh International Conference on Document Analysis and Recognition. 2 (2003) 1086.

[76] SN Srihari, CI Tomai, B Zhang, S Lee, Individuality of numerals, Proceedings of the Seventh International Conference on Document Analysis and Recognition. (2003) 1096.

[77] G Leedham, S Chachra, Writer identification using innovative binarised features of handwritten numerals, Seventh International Conference on Document Analysis and Recognition. (2003) 413-417.

[78] B Zhang, S Srihari, Analysis of handwriting individuality using word features, Proceedings of the Seventh International Conference on Document Analysis and Recognition. (2003) 1142-1146.

[79] CI Tomai, B Zhang, SN Srihari, Discriminatory power of handwritten words for writer recognition, Proceedings of the 17th International Conference on Document Analysis and Recognition. 2 (2004) 638-641.

[80] B Arazi. Handwriting Identification by Means of Run-Length Measurements, IEEE Transactions on Systems, Man and Cybernetics. 7 (1977) 878-881.

[81] H Said, G Peake, T Tan, KD Baker, Writer identification from non-uniformly skewed handwriting images, Proceedings of the ninth Brithish Machine Vision Conference. (1998) 478-487.

[82] H Said, T Tan, K Baker. Personal identification based on handwriting, Pattern Recognition. 33 (2000) 149-160.

[83] Y Zhu, T Tan, Y Wang, Biometric personal identification based on handwriting, Proceedings of the 15th International Conference on Pattern Recognition. 2 (2000) 797-800.

[84] Z HE, X YOU, YY TANG, B FANG, J DU. Handwriting-Based Personal Identification, International Journal of Pattern Recognition & Artificial Intelligence. 20 (2006) 209-225.

[85] S Gazzah, NE Ben Amara, Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting Scheme, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). 2 (2007) 1133-1137.

[86] U- Marti, R Messerli, H Bunke, Writer identification using text line based features, Proceedings of the Sixth International Conference on Document Analysis and Recognition. (2001) 101-105.

[87] C Hertel, H Bunke, A Set of Novel Features for Writer Identification, Audio- and Video-Based Biometric Person Authentication, Lecture Notes in Computer Science. 2688 (2003) 1058.

[88] V Bouletreau, N Vincent, R Sabourin, H Emptoz, Synthetic parameters for handwriting classification, Proceedings of the Fourth International Conference on Document Analysis and Recognition. 1 (1997) 102-106 vol.1.

[89] L Schomaker, M Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase Western script, IEEE Transactions on Pattern Analysis and Machine Intelligence. 26 (2004) 787-798.

[90] M Bulacu, L Schomaker. Text-independent writer identification and verification using textural and allographic features, IEEE Transactions on Pattern Analysis and Machine Intelligence. 29 (2007) 701-717.

[91] L Schomaker, M Bulacu, K Franke, Automatic writer identification using fragmented connected-component contours, Ninth International Workshop onFrontiers in Handwriting Recognition (IWFHR-9 2004). (2004) 185-190.

[92] A Bensefia, T Paquet, L Heutte. A writer identification and verification system, Pattern Recognition Letters. 26 (2005) 2080-2092.

[93] A Bensefia, T Paquet, L Heutte. Handwritten document analysis for automatic writer recognition, Electronic letters on computer vision and image analysis. 5 (2005) 72-86.

[94] G Salton, A Wong, CS Yang. A vector space model for automatic indexing, Communications of the ACM. 18 (1975) 613-620.

[95] R Jain, D Doermann, Offline Writer Identification using K-Adjacent Segments, Eleventh International Conference on Document Analysis and Recognition (ICDAR). (2011) 769-773.

[96] M Bulacu, L Schomaker, A Brink, Text-independent writer identification and verification on offline Arabic handwriting, Proceedings of the 9th International Conference on Document Analysis and Recognition. (2007) 769-773.

[97] I Siddiqi, N Vincent. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features, Pattern Recognition. 43 (2010) 3853-3865.

[98] H Freeman. On the Encoding of Arbitrary Geometric Configurations, IRE Transactions on Electronic Computers. EC-10 (1961) 260-268.

[99] A Schlapbach, H Bunke. Using HMM Based Recognizers for Writer Identification and Verification, International Workshop on Frontiers in Handwriting Recognition. (2004) 167-172.

[100] A Schlapbach, H Bunke, Off-lineWriter Identification Using Gaussian Mixture Models, 18th International Conference on Pattern Recognition (ICPR 2006). 3 (2006) 992-995.

[101] R Plamondon, G Lorette. Automatic signature verification and writer identification — the state of the art, Pattern Recognition. 22 (1989) 107-131.

[102] E Zois, V Anastassopoulos. Morphological waveform coding for writer identification, Pattern Recognition. 33 (2000) 385-398.

[103] E Zois, V Anastassopoulos. Fusion of correlated decisions for writer verification, Pattern Recognition. 34 (2001) 47-61.

[104] A Bensefia, T Paquet, L Heutte, Handwriting analysis for writer verification, Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004). (2004) 196-201.

[105] Siew Keng Chan, Yong Haur Tay, C Viard-Gaudin, Online text independent writer identification using character prototypes distribution, 6th International Conference on Information, Communications & Signal Processing. (2007) 1-5.

[106] GX Tan, C Viard-Gaudin, AC Kot. Automatic writer identification framework for online handwritten documents using character prototypes, Pattern Recognition. 42 (2009) 3313-3323.

[107] R Niels, F Grootjen, L Vuurpijl, Writer identification through information retrieval: the allograph weight vector, International Conference on Frontiers in Handwriting Recognition. (2008) 481-486.

[108] B Li, Z Sun, T Tan. Online Text-Independent Writer Identification Based on Strokes Probability Distribution Function, Lecture Notes in Computer Science. 4642 (2007) 201-210.

[109] A Schlapbach, M Liwicki, H Bunke. A writer identification system for on-line whiteboard data, Pattern Recognition. 41 (2008) 2381-2397.

[110] Luidia Inc., eBeam Edge Whiteboard, URL: http://www.e-beam.com/products/ebeam-whiteboards.html (last accessed: 15-march-2012).

[111] L Zuo, Y Wang, T Tan, Personal handwriting identification based on pca, Proceedings of SPIE Second International Conference on Image and Graphics. (2002) 766-771.

[112] P Thumwarin, T Matsuura, On-line writer recognition for Thai based on velocity of barycenter of pen-point movement, International Conference on Image Processing ICIP '04. 2 (2004) 889-892 Vol.2.

[113] E Sesa-Nogueras, M Faundez-Zanuy. Biometric recognition using online uppercase handwritten text, Pattern Recognition. 45 (2012) 128-144.

[114] E Sesa-Nogueras, M Faundez-Zanuy, J Mekyska. An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting, Cognitive Computation. Doi: 10.1007/s12559-011-9119-y (2012).

[115] S Srihari, C Sung-Hyuk, L Sangjik, Establishing handwriting individuality using pattern recognition techniques, Proceedings of the Sixth International Conference on Document Analysis and Recognition. (2001) 1195-1204.

[116] F Leclerc, R Plamondon. Automatic Signature Verification: The State of the Art - 1989-1993, International Journal of Pattern Recognition & Artificial Intelligence. 8 (1994) 643.

[117] R Plamondon, S Srihari, On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence. 22 (2000) 63-84.

[118] D Impedovo, G Pirlo. Automatic Signature Verification: The State of the Art, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 38 (2008) 609-635.

[119] H Baltzakis, N Papamarkos. A new signature verification technique based on a two-stage neural network classifier, Engineering Applications of Artificial Intelligence. 14 (2001) 95-103.

[120] M Ferrer, J Alonso, C Travieso. Offline Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic. IEEE Transactions on Pattern Analysis and Machine Intelligence. 27 (2005) 993-997.

[121] LR Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE. 77 (1989) 257-286.

[122] L Yang, B Widjaja, R Prasad. Application of hidden Markov models for signature verification, Pattern Recognition. 28 (1995) 161-170.

[123] HS Yoon, JY Lee, HS Yang, An online signature verification system using hidden Markov model in polar space, Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition. (2002) 329-333.

[124] J Ortega-Garcia, J Fierrez-Aguilar, J Martin-Rello, J Gonzalez-Rodriguez, Complete Signal Modeling and Score Normalization for Function-Based Dynamic Signature Verification, in: Kittler J, Nixon M (Eds.), Audio- and Video-Based Biometric Person Authentication, Springer Berlin / Heidelberg, 2003.

[125] J Fierrez-Aguilar, J Ortega-Garcia, D Ramos, J Gonzalez-Rodriguez. HMM-based on-line signature verification: Feature extraction and signature modeling, Pattern Recognition Letters. 28 (2007) 2325-2334.

[126] JR Deller, JG Proakis, JHL Hansen, Dynamic Time Warping, Discrete-time processing of speech signals, Macmillan Publishing Co., New York, 1993.

[127] A Jain, F Griess, S Connell. On-line signature verification, Pattern Recognition. 35 (2002) 2963-2972.

[128] L Bovino, S Impedovo, G Pirlo, L Sarcinella, Multi-expert verification of hand-written signatures, Proceedings of the Seventh International Conference on Document Analysis and Recognition. (2003) 932-936.

[129] D Impedovo, G Pirlo, On-line Signature Verification by Stroke-Dependent Representation Domains, International Conference on Frontiers in Handwriting Recognition (ICFHR). (2010) 623-627.

[130] J Lee, H Yoon, J Soh, BT Chun, YK Chung. Using geometric extrema for segment-to-segment characteristics comparison in online signature verification, Pattern Recognition. 37 (2004) 93-103.

[131] CJC Burges. A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery. 2 (1998) 121-167.

[132] Ling-Lee L., Neural approaches for human signature verification, Proceedings of the Third International Conference Document Analysis and Recognition. 2 (1995) 1055-1058 vol.2.

[133] EJR Justino, F Bortolozzi, R Sabourin, An off-line signature verification method based on SVM classifier and graphometric features, Fifth International Conference on Advances in Pattern Recogniton (ICAPR 2003). (2003) 487-490.

[134] E Özgündüz, T Sentürk, M Karsligil, Off-line signature verification and recognition by support vector machine, Proceedings of the 2005 European Signal Processing Conference (EUSIPCO 2005). (2005).

[135] J Fàbregas-Peinadoi, M Faundez-Zanuy. On-line signature verification system with failure to enrol management, Pattern Recognition. 42 (2009) 2117-2126.

[136] M Faundez-Zanuy. On-line signature recognition based on VQ-DTW, Pattern Recognition. 40 (2007) 981-992.

[137] JM Pascual-Gaspar, M Faundez-Zanuy, C Vivaracho. Fast on-line signature recognition based on VQ with time modeling, Engineering Applications of Artificial Intelligence. 24 (2010) 368-377.

[138] M Faundez-Zanuy, JM Pascual-Gaspar. Efficient on-line signature recognition based on multi-section vector quantization, Pattern Analysis & Applications. 14 (2011) 37-45.

[139] B Wirtz, Stroke-based time warping for signature verification, Proceedings of the Third International Conference on Document Analysis and Recognition. 1 (1995) 179-182.

[140] B Yanikoglu, A Kholmatov. Online Signature Verification Using Fourier Descriptors, EURASIP Journal on Advances in Signal Processing. 2009 (2009).

[141] A Kholmatov, B Yanikoglu. Identity authentication using improved online signature verification method. Pattern Recognition Letters. 26 (2005) 2400-2408.

[142] N Houmani, A Mayoue, S Garcia-Salicetti, B Dorizzi, MI Khalil, MN Moustafa, et al. BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures, Pattern Recognition. 45 (2012) 993-1003.

[143] VL Blankers, C Heuvel, KY Franke, LG Vuurpijl, Icdar 2009 signature verification competition, Proceedings of the10th International Conference onDocument Analysis and Recognition. (2009) 1403-1407.

[144] N Houmani, S Garcia-Salicetti, B Dorizzi, J Montalvão, J Canuto, M Andrade, et al., BioSecure Signature Evaluation Campaign (ESRA'2011): Evaluating Systems on Quality-based categories of Skilled Forgeries, IJCB'2011 International Joint Conference on Biometrics. (2011).

[145] TM Cover , JA Thomas, Elements of information theory, Wiley, New York, 1991.

[146] DJC MacKay, Information theory, inference, and learning algorithms, Cambridge University Press 2003.

[147] AK Jain, MN Murty, PJ Flynn. Data clustering: a review, ACM computing surveys (CSUR). 31 (1999) 264-323.

[148] T Kohonen. The self-organizing map, Proceedings of the IEEE. 78 (1990) 1464-1480.

[149] K Kiviluoto, Topology preservation in self-organizing maps, IEEE International Conference on Neural Networks, 1996. 1 (1996) 294-299 vol.1.

[150] Helsinki University of Technology. Laboratory of Computer and Information Science, SOM Toolbox, URL: http://www.cis.hut.fi/somtoolbox/ (last accessed: 15-march-2012).

[151] J Vesanto, J Himberg, E Alhoniemi, J Parhankangas, Self-organizing map in Matlab: the SOM toolbox, Proceedings of the Matlab DSP Conference 1999. (1999) 35-40.

[152] I The MathWorks, MATLAB - The Language Of Technical Computing, URL: http://www.mathworks.com/products/matlab/ (last accessed: 15-march-2012).

[153] M Faundez-Zanuy. Data fusion in biometrics, Aerospace and Electronic Systems Magazine, IEEE. 20 (2005) 34-38.

[154] J Fierrez, J Ortega-Garcia, On-Line Signature Verification, in: Jain AK, Flynn P, Ross AA (Eds.), Handbook of Biometrics, Springer, 2008.

[155] SN Yanushkevich, V Shmerko, A Stoica, P Wang, S Srihari, Introduction to Synthesis in Biometrics, in: Yanushkevich SN, Wang PS, Gavrilova ML, Srihari SN (Eds.), Image Pattern Recognition - Synthesis and Analysis in Biometrics, World Scientific, Singapore ; Hackensack, NJ, 2007, pp. 5-29.

[156] J Galbally, Vulnerabilities and Attack Protection in Security Systems Based on Biometric Recognition. PhD Thesis. Universidad Autonoma de Madrid, (2009).

[157] M Mori, A Suzuki, A Shio, S Ohtsuka, Generating New Samples from Handwritten Numerals Based on Point Correspondence, Proceedings of the 7th International Workshop on Frontiers of Handwriting Recognition. (2000) 281-290.

[158] J Cano, J Perez-Cortes, J Arlandis, R Llobet, Training Set Expansion in Handwritten Character Recognition, Structural, Syntactic and Statistical Pattern Recognition, LNCS 2396. (2002) 548-556.

[159] T Varga, H Bunke, Generation of synthetic training data for an HMM-based handwriting recognition system, Proceedings of the Seventh International Conference on Document Analysis and Recognition. (2003) 618-622 vol.1.

[160] T Varga, D Kilchhofer, H Bunke, Template-based synthetic handwriting generation for the training of recognition systems, Proceedings of the 12th Conference of the International Graphonomics Society. (2005) 206-211.

[161] M Helmers, H Bunke, Generation and Use of Synthetic Training Data in Cursive Handwriting Recognition, in: Perales F, Campilho A, de la Blanca N, Sanfeliu A (Eds.), Pattern Recognition and Image Analysis, Springer Berlin / Heidelberg, 2003, pp. 336-345.

[162] H Mouchere, E Anquetil, Synthèse de caractères manuscrits en-ligne pour la reconnaissance de l'écriture, Actes du Colloque International Francophone sur l'Ecrit et le Document (CIFED'06). (2006) 187-192.

[163] K Huang, H Yan. Off-line signature verification based on geometric feature extraction and neural network classification, Pattern Recognition. 30 (1997) 9-17.

[164] Cd Oliveira, CAA Kaestner, F Bortolozzi, R Sabourin, Generation of Signatures by Deformations, Proceedings of the BSDIA97. (1997) 283-298.

[165] C Rabasse, RM Guest, MC Fairhurst. A New Method for the Synthesis of Signature Data With Natural Variability. IEEE Transactions on Systems, Man, and Cybernetics, Part B. 38 (2008) 691-699.

[166] C Rabasse, RM Guest, MC Fairhurst, A Method for the Synthesis of Dynamic Biometric Signature Data, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). 1 (2007) 168-172.

[167] ME Munich, P Perona. Visual identification by signature tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence. 25 (2003) 200-217.

[168] J Galbally, J Fierrez, M Martinez-Diaz, J Ortega-Garcia, Improving the Enrollment in Dynamic Signature Verfication with Synthetic Samples, 10th International Conference on Document Analysis and Recognition, 2009. ICDAR '09. (2009) 1295-1299.

[169] M Djioua, C O'Reilly, R Plamondon, An interactive trajectory synthesizer to study outlier patterns in handwriting recognition and signature verification, 18th International Conference on Pattern Recognition ICPR 2006. 1 (2006) 1124-1127.

[170] PJ Flynn, AK Jain, AA Ross, Handbook of biometrics, Springer, New York, N.Y., 2008.

[171] D Zhang, Automated biometrics: Technologies and systems, Kluwer Academic Publishers 2000.

[172] Cambridge University Press, Cambridge Dictionary Online URL: http://dictionary.cambridge.org/ (last accessed: 15-march-2012).