Automatic Signature Verification: Bridging the Gap between Existing Pattern Recognition Methods and Forensic Science

Dissertation

approved by the

Department of Computer Science

Technische Universität Kaiserslautern

for the award of the Doctoral Degree

Doctor of Engineering

(Dr.-Ing.)

by

Muhammad Imran Malik

Thesis supervisors:
Prof. Dr. Prof. h.c. Andreas Dengel, TU Kaiserslautern
apl.-Prof. Dr. habil. Marcus Liwicki, TU Kaiserslautern

Chair of the supervisory committee:
Prof. Dr. Karsten Berns, TU Kaiserslautern

Kaiserslautern, October 19, 2015

D 386

Muhammad Imran Malik
Trippstadter Str. 121
67663 Kaiserslautern

Kaiserslautern, den December 14, 2015

Erklärung

Ich versichere hiermit, dass ich die vorliegende Promotionarbeit mit dem Thema "*Automatic Signature Verification: Bridging the Gap between Existing Pattern Recognition Methods and Forensic Science*" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, habe ich durch die Angabe der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht.

(Muhammad Imran Malik)

To my dearest Abbu Ji, who couldn't be here to see this day.

Acknowledgement

First of all, my greatest gratitude to the Almighty who has enabled me complete the thesis at hand and has made this day possible for me to write this acknowledgment.

I would like to thank apl.-Prof. Dr. habil. Marcus Liwicki, who has always remained a firm support for me during this entire study endeavor. His advice, correction, and encouragement have been a great source of motivation for me. He always came up with the ideas when I needed them, thanks Marcus.

A special thanks to Prof. Dr. Prof. h.c. Andreas Dengel, I am really honored to have an opportunity to work in his group to complete this thesis. This has been one of the best experiences of my life.

I would like to thank my buddy, Sheraz. His continuous support has always been my asset. His on and off discussions helped me a lot in understanding various issues concerned with me, my life, and my studies. Special thanks to my friends, Saad, Mohsin, Sajid, Abdullah, Adeel, Zain, and Gulzar. Their company and support has helped me greatly throughout my PhD.

I would also like to say thanks all my colleagues at the KM group, DFKI Kaiserslautern.

In the last but not the least, I really really admire the support of my family. Without this support it was never possible for me to think of this day. A huge bouquet of love, thanks, and gratitude for my family.

Executive Summary

The main goal of this thesis is twofold. First, the thesis aims at bridging the gap between existing Pattern Recognition (PR) methods of automatic signature verification and the requirements for their application in forensic science. This gap, attributed by various factors ranging from system definition to evaluation, prevents automatic methods from being used by Forensic Handwriting Examiners (FHEs). Second, the thesis presents novel signature verification methods developed particularly considering the implications of forensic casework, and outperforming the state-of-the-art PR methods.

The first goal of the thesis is attributed by four important factors, i.e., data, terminology, output reporting, and how evaluation of automatic systems is carried out today. It is argued that traditionally the signature data used in PR are not actual/close representative of the real world data (especially that available in forensic cases). The systems trained on such data are, therefore, not suitable for forensic environments. This situation can be tackled by providing more realistic data to PR researchers. To this end, various signature and handwriting datasets are gathered in collaboration with FHEs and are made publicly available through the course of this thesis. A special attention is given to disguised signatures–where authentic authors purposefully make their signatures look like a forgery. This genre was at large neglected in PR research previously.

The terminology used, in the two communities - PR and FHEs, differ greatly. In fact, even in PR, there is no standard terminology and people often differ in the usage of various terms particularly related to various types of forged signatures/handwriting. The thesis presents a new terminology that is equally useful for both forensic scientists and PR researchers. The proposed terminology is hoped to increase the general acceptability of automatic signature analysis systems in forensic science.

The outputs reported by general signature verification systems are not acceptable for FHEs and courts as they are either binary (yes/no) or score (raw evidence) based on similarity/difference. The thesis describes that automatic systems should rather report the probability of observing the evidence (e.g., a certain similarity/difference score) given the signature belongs to the acclaimed identity, and the probability of observing the same evidence given the signature does not belong to the acclaimed identity. This will take automatic systems from hard decisions to soft decisions, thereby enabling them to report likelihood ratios that actually represent the evidential value of the score rather than the raw score (evidence).

When automatic systems report soft decisions (as in the form of likelihood ratios), the thesis argues that there must be some methods to evaluate such systems. This thesis presents one such adaptation. The thesis argues that the state-of-the-art evaluation methods, like equal error rate and area under curve, do not address the needs of forensic science. These needs require an assessment of the evidential value of signature verification, rather than a hard/pure classification (accept/reject binary decision). The thesis demonstrates and validates a relatively simple adaptation of the current verification methods based on the Bayesian inference dependent calibration of continuous scores rather than hard classifications (binary and/or score based classification).

The second goal of this thesis is to introduce various local features based techniques which are capable of performing signature verification in forensic cases and reporting results as anticipated by FHEs and courts. This is an important contribution of the thesis because of the following two reasons. First, to the best of author's knowledge, local feature descriptors are for the first time used for development of signature verification systems for forensic environments (particularly considering disguised signatures). Previously, such methods have been heavily used for recognition tasks, rather than verification of writing behaviors, such as character and digit recognition. Second, the proposed methods not only report the more traditional decisions (like scores-usually reported in PR) but also the Bayesian inference based likelihood ratios (suitable for courts and forensic cases).

Furthermore, the thesis also provides a detailed man vs. machine comparison for signature verification tasks. The men, in this comparison, are forensic scientists serving as forensic handwriting examiners and having experience of varying number of years. The machines are the local features based methods proposed in this thesis, along with various other state-of-the-art signature verification systems. The proposed methods clearly outperform the state-of-the-art systems, and sometimes the human experts.

Finally, the thesis details various tasks that have been performed in the areas closely related to signature verification and its application in forensic casework. These include, developing novel local feature based methods for extraction of signatures/handwritten text from document images, hyper-spectral image analysis for extraction of signatures from forensic documents, and analysis of on-line signatures acquired through specialized pens equipped with Accelerometer and Gyroscope. These tasks are important as they enable the thesis to take PR systems one step further close to direct application in forensic cases.

Contents

List of Figures

List of Tables

1 Introduction

Handwritten signatures have been used as a means of establishing authorship since a long time [7, 8]. Right from the beginning, there has been a demand of verifying whether the acclaimed signatures were authentic (belonging to the claiming person) or not. This verification was performed by human experts who used various tools, e.g., scaling devices and optics, for verifying the authenticity of signatures/handwriting. With the evolution of modern computing technologies and an ever increasing demand for signature verification, there was a natural interest (of computer science/pattern recognition researchers at least) towards developing automatic methods for signature verification. This interest has been reinforced with the concernment shown by industry (e.g., banking) and sciences (e.g., forensic science). Various automatic signature verification systems have been reported over the last few decades [9, 10, 11, 12]. However, most of these systems are developed for general purpose signature verification and/or considering the requirements of only specific industries, mostly banking [13].

Forensic signature/handwriting analysis has, until recent times, often carried out without actual application of automatic systems, except for some limited tools [14] like, CEDAR-FOX [15], FISH [16], WANDA project framework [17], and FLASH ID [18]. The underlying issue is that most of the state-of-the-art signature/handwriting analysis systems cannot be directly applied to forensic cases. This is because of, at least, the following four major issues which are addressed in detail in this thesis and solutions are presented to their effect.

- Difference in data: The data currently used in pattern recognition (PR) research are not actual/close representative of data available in forensic cases.

- Difference in terminology: Standard terminology that is equally interpretable/usable for both the communities, PR and forensic handwriting examiners (FHEs), is currently not available.

- Output reporting: The current PR methods report results in the form of either hard decisions (yes/no), or score (raw evidence) based on similarity/difference. This is not suitable for forensic casework as FHEs and courts require outcomes in the form of likelihood ratios, i.e., ratios indicating the evidential value of indicia found in the signatures under examination [19].

- Evaluation: PR researchers generally evaluate/rank systems on the basis of metrics relying on hard decisions (e.g., accuracy, false acceptance rate, false rejection rate, equal error rate, area under curve, etc.). FHES, however, would like to also consider the severity of errors made by an automatic system and therefore rank the systems accordingly [20].

This thesis presents various novel automatic signature verification methods to fulfill the application needs of forensic casework. These methods are subjected on local features based techniques (SURF [21], FAST [4], and FREAK [5], etc.) and are capable of performing signature verification in forensic cases. This is an important contribution of the thesis because of the following two reasons.

- First, to the best of authors knowledge, local feature descriptors are for the first time used for development of signature verification systems for forensic environments (particularly considering cases where authentic authors tried to forge their own signatures, i.e., disguised signatures).

- Second, the proposed methods not only report the more traditional decisions (like scores-usually reported in PR) but also the Bayesian inference based likelihood ratios (suitable for forensic experts and the courts of law).

An important issue concerning the application of automatic systems in forensic signature/handwriting analysis cases is whether such system possess potential to show nearly similar performance than human experts? This thesis provides a detailed Man vs. Machine comparison for forensic signature verification tasks. The automatic methods proposed in this thesis clearly outperform the state-of-the-art PR systems, and sometimes the human experts on various publicly available datasets.

1.1 Motivation

The main motivation of this thesis is to take automatic signature verification closer to forensic science by considering the inherent implications of forensic casework and developing state-of-the-art signature verification systems that are readily usable in forensic casework. Although handwriting/signature analysis makes an important portion of a forensic document examiner's work, forensic casework is dealt without or very little application of automatic signature/handwriting verification systems. This is quite in contrast with the heavy usage of automatic systems in forensic cases considering other biometric modalities like, face, iris, fingerprints, and so on [22, 23]. The aim of this thesis is to explore these issues and to present possible solutions which are adoptable for both the communities, PR and FHEs.

Forensic handwriting examiners often encounter cases where they deal with the signature categories that are either completely or mostly neglected in the PR research, e.g., disguised signatures (when an authentic author forges her/his own signatures). The aim is to take such categories to the PR community at large so that to facilitate the development of automatic system capable of dealing with such novel/forgotten categories as well with the previously considered categories, e.g., genuine and forged signatures. Another motivation of the thesis at hand is to evaluate whether the state-of-the-art signature verification systems are capable of assisting human experts? The aim is not to replace the humans, but to increase their overall productivity (efficiency and effectiveness).

1.2 Problem Statement

This thesis aims at solving the following main problems.

- Developing a shared conceptualization, for signature verification that is standard across communities–particularly PR and FHEs.

- Developing a scientific foundation where the state-of-the-art PR signature verification systems can provide objective results for forensic applications and in the courts.

- Developing novel signature verification methods that: (i) Outperform the state-of-the-art methods, (ii) are comparable to human experts so that to assist them, and (iii) can produce results as anticipated by FHEs and the courts.

1.3 Signature Verification

Signature verification is the process of establishing whether or not a signature belongs to the person who claims to be the owner of the signature [9]. It is considered as a classical 1 *to* 1 pattern classification problem: either a signature is verified to belong to a given person or it is not. More specifically, the PR community defines signature verification as [24, 25]: Given an input feature vector F_s, obtained from a signature s and a claimed identity I, determine whether (I, F_s) belongs to class k_1 or to class k_2. Here, class k_1 indicates that the signature is from the original author, while k_2 indicates that the signature is not from the original author. The feature vector F_s is compared with M_I, i.e., the model corresponding to the author I, to find whether to assign it to class k_1 or to class k_2. Therefore, we can write

$$(I, F_s) \epsilon \begin{cases} k_1, & \text{if } f\left(M_I, F_s\right) \geq \theta \\ k_2, & \text{otherwise} \end{cases}$$

where the function $f(M_I, F_s)$ returns a similarity score $sc \epsilon \mathbb{R}$ indicating the similarity of F_s and model M_I of author I and where $\theta \epsilon \mathbb{R}$ is a predefined threshold. As we increase θ, the chances of the input signature being classified as not belonging to the authentic author increase.

If a signature is verified to belong to the claiming person, the signature is said to be genuine/authentic and the person is called a *client* or *authentic/target/specimen author/writer/signer*. Otherwise, the signature is called a forgery or simulation and the person who wrote it an *impostor/forger* or *non-specimen writer* [26, 24]. In PR research, a forgery/simulation is usually categorized in any of the following types [27].

- Random Forgery/Simple Forgery: where the forger has no idea about and has not seen the genuine signatures of authentic author.

- Unskilled Forgery: where the forger has seen the original signatures of a person but has not practiced to forge the signatures.

- Skilled Forgery: where the forger has seen and practiced to forge the genuine signature of a person. The forger may make use of various devices for observing the signature and practice forging it.

The signature data can be present in on-line or off-line format. In the case of on-line verification, a time ordered sequence of coordinates, representing the pen movement, is

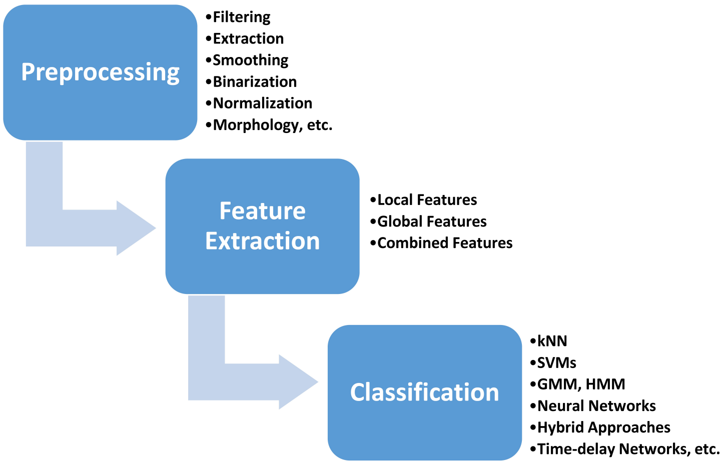


Figure 1.1: General signature verification steps

available. This can be produced by any electronic sensing device, such as, electronic pens or digitizing tablets. In the case of off-line verification only the image of the signature is available, which usually is scanned or photographed from paper. The temporal information adjunct with the writing style of a person, his/her angle of grasping the pen, direction of movement etc., simplifies the task verification to some extent. Whether verification is performed on off-line or on-line data, the basic architecture of a verification system is similar. The following section provides an overview of a general signature verification system.

1.3.1 General Signature Verification System

A signature verification system is usually divided into consecutive units which iteratively process the input data to finally accept or reject a signature. The main units are illustrated in Fig. 1.1 and summarized in this subsection. Note that there are differences between off-line and on-line processing, but the underlying principles are the same. Only the methodology of performing the individual steps differs.

The first unit is the preprocessing unit, where the input is raw signature data and the output consists of signatures that have undergone some initial processing like binarization and size/skew normalization. The amount of effort that has to be invested into preprocessing depends on the given data. If the data have been acquired from a system that does not produce much noise there is not much to do in this step. However, the data usually contain noise, which has to be removed to improve the quality of the data, or are available with some background information, e.g., on a bank check where we need to extract the signature from its background. Though there is an argument about the requirement of preprocessing and how it affects the possible outcomes, especially in forensic cases where sometimes lives are at stake [28]. Over the last few years, different systems capable of performing signature preprocessing have been reported. Note that often signature preprocessing is fused with verification, but sometimes it is considered as a completely separate task due to its difficult nature, e.g., extraction of signatures from historic documents [12, 29]. A detailed review of the state-of-the-art signature preprocessing techniques is provided in Chapter 2.

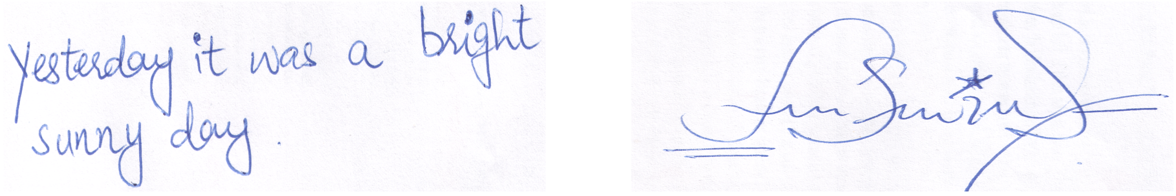
The second unit, feature extraction, is essential for any signature verification system, for any classifier needs numeric data as input [30]. Over the last four decades, a large number of features and feature extraction methods have been reported in the literature [12]. These features are specialized to glean local, global and combined information from signatures. Once features are extracted, the same methods can be used for off-line and on-line verification. A detailed review of some of the important off-line and on-line feature extraction and verification techniques is provided in Chapter 2.

The next unit classifies the input features. Various classifiers have been considered in the literature, e.g., for a fixed length feature vector, neural networks [31, 32], support vector machines [33], or nearest neighbor classifiers can be applied [34]. For sequences of feature vectors, hidden Markov models [35], time delay neural networks [36], or hybrid approaches [37, 38] are generally used for classification [30]. The classification unit generally reports the final classification results, e.g., a binary decision or a similarity based score/probability. Nonetheless, result reporting can be performed by a separate unit, e.g., when reporting evidential values of similarities in the form of likelihood ratios. This issue is discussed in detail in Chapter 4.

1.4 Related Topics

Closely related to signature "verification" is signature "identification". Unlike verification, identification is the task is to assign a signature to an author out of a set of possible authors. It is a 1 *to* $N + 1$ classification problem where N is the number of possible authors and $N+1$ represents the case where the signature does not belong to any person in the N set [9, 11, 12].

More specifically, identification is defined as [24, 25]: Given an input feature vector F_t, extracted from a signature t, determine the identity I_k, $k \epsilon \{1, 2, 3, ..., N, N+1\}$. Here $I_1, I_2, ..., I_N$ are the identities of authors represented in the system by a corresponding



(a) (b)

Figure 1.2: (a) Free-hand writing of a person, (b) Signatures by the same person

model M_k, $k \epsilon \{1, 2, 3, ..., N\}$. The model I_{N+1} indicates the reject case where no suitable identity can be established for the user. Given k authors, the signature identification problem is a $k + 1$ class classification problem. Therefore,

$$F_t \; \epsilon \begin{cases} I_k, & \text{if } max_k \left\{ s\left(M_k, F_t\right)\right\} \geq \theta, \quad k = 1, 2, 3, ..., N \\ I_{N+1}, & otherwise \end{cases}$$

where the function $s(M_k, F_t)$ returns a score $s \epsilon \mathbb{R}$ indicating that F_t corresponds to the model M_I of the user I and where $\theta \epsilon \mathbb{R}$ is a predefined threshold. If this threshold is increased then there is a possibility of non-identification, i.e., rejection.

Writer identification and verification are also closely related to signature identification and verification, as a signature is an over-practiced piece of handwriting [39, 40]. However unlike signature verification, in writer identification and verification the systems are usually presented with pieces of text lines and sometimes complete pages of text. Having a large corpus of text per writer generally provides a bundle of features to be extracted and used for the prediction of potential writer. This is usually not the situation with signatures. Signatures are sometimes more graphical than usual text. Furthermore, some people deliberately make their signatures more graphical by exaggerating the slants or curves of the letters they write. Sometimes people also like to add special symbols, other than normal letters, to increase the visual plausibility of their signatures, e.g., using a tiny star instead of the "dot" of an "i" and etc. Figure 1.2 shows the writing and signatures of a person. Notice a "*" used as the dot of "i" and slant differences in characters are also evident.

Moreover, people usually deform the letters of their signatures or even simply ignore writing some letters. This is because people write text with the purpose of being legible but they use signatures merely as a symbol or insignia of their identity. Nonetheless, a majority of the techniques developed for writer identification/verification are also applicable for signature identification/verification but with varying results.

Another closely related area, but with a principle difference, is Handwriting Recognition (HWR). In signature/handwriting identification or verification, the emphasis is on maintaining and even enhancing the writer specific information so that to positively identify/verify the original author. In contrast, HWR tries to remove the writer specific variations so that to increase the recognition performance [24].

1.5 Contribution

The main contribution of this thesis is to bridge the gap between the state-of-the-art automatic signature verification and forensic science. More specifically, this thesis contributes the following novelties to the field.

1. A theoretical framework is presented on the basis of which the PR community in general is facilitated to develop signature verification solutions particular to the needs of forensic science.

2. A terminology, equally informative for both PR and FHEs, for various categories of signature forgeries is presented.

3. Through the course of this work, five novel signature datasets have been developed in collaboration with forensic experts and made publicly available. These datsets contain specific cases similar to the real-world cases faced by forensic experts, e.g., cases involving categories of signatures that were available with FHEs but were generally neglected by PR researchers..

4. Various signature verification competitions and evaluations have been carried out which enabled succinct testing of the theoretical framework presented in this thesis and validated its suitability to automatic signature verification.

5. Various novel signature verification methods have been developed which not only outperform the state-of-the-art but also are optimized to cater the needs of forensic examination.

6. The performance of the novel methods has been thoroughly evaluated against that of human forensic experts and results are reported in order to present automatic signature verification systems as good assistants for forensic experts.

7. Related application areas, such as extraction of signatures from documents have been considered and solutions are presented to their effect.

1.6 Outline

The remainder of this thesis is structured as follows. Chapter 2 discusses automatic signature verification and typical modules of signature verifiers (preprocessing, feature extraction, verification) in detail. The state-of-the-art of off-line, on-line, and combined/hybrid signature verification systems along with a summary of various signature verification competitions organized in the past, and the state-of-the-art metrics available for evaluating automatic signature verifiers, is provided.

Chapter 3 provides an insight into the state-of-the-art of how signature examination is performed in forensic scenarios. It elaborates on how forensic experts perform signature analysis, how they report results, what instruments they use for forensic comparison of signatures, and finally which automatic systems are available currently to assist forensic experts.

Chapter 4 presents the author's approach for bridging the gap between PR-researchers and FHEs. This chapter marks the major differences currently present in the terminology and its usage between PR and FHEs, and presents a novel common terminology useful for both the communities. It highlights the need of having data for training and testing PR systems similar to what are available in forensic cases. Furthermore, this chapter argues that the current PR output reporting schemes do not address the needs of forensic science. Finally, the use of a relatively simple adaptation of the current comparison methods is proposed to successfully address the needs of FHEs and PR.

Chapter 5 validates the postulates of this thesis. This chapter describes how the author has validated the postulates laid in this thesis (especially in Chapter 4). The author achieved the purpose by organizing various international signature verification competitions (co-joined with leading PR researchers and forensic scientists) in order to validate the working of the proposed standard terminology and the generic evaluation scheme discussed in this thesis. By these competitions, the author has also made various forensic like signature and handwriting datasets publicly available.

Chapter 6 presents various local features/part-based systems (based on local feature approaches, SURF, FAST, and FREAK) proposed through the course of this thesis in order to perform signature verification particularly considering the implications of forensic casework. Chapter 7 presents a novel analysis of stability of signatures on the basis of signature's local / part-based features (SURF).

Chapter 8 provides a direct performance comparison between the state-of-the-art signature verification systems and FHEs against the local feature based methods proposed in this thesis. The proposed methods clearly outperform other automatic methods and also FHEs in some cases thereby justifying the potential use of the methods, reported in this thesis, in forensic signature verification cases.

Chapters 9, and 10, detail various tasks that have been performed in areas closely related to signature verification and its application in forensic casework. Note that in almost all of the existing verification systems, it is assumed that signatures are already segmented/extracted from documents. However, in real world scenarios, signatures are a part of documents which also contain other information, e.g., bank checks, invoices, contracts, credit card pay slip, wills, etc. In these scenarios, the existing signature verification and identification systems cannot be used 'as is' because of the lack of segmentation. Chapter 9 presents a novel local features based signature segmentation system and further focuses on the challenges which must be tackled for the development of a complete automatic document analysis system capable of performing signature segmentation/extraction from documents and then performing signature verification. Chapter 10 further explores the research questions raised in Chapter 9 by applying hyper-spectral imaging (HSI) for automatic signature segmentation.

Chapter 11 describes some further tasks, associated with the core signature verification, that have been performed during the due course of this thesis. The chapter describes various experiments that have been performed in order to look into the various dimensions and implications of on-line data. First, a generic framework based on the use of a digitizing pen (Anoto digital pen) for various real-time on-line signature verification scenarios is presented. Second, this chapter compared the expedience of two easy to use and cheap devices for on-line signature verification. Both devices are ballpoint pens with a sensor attached. One of them measures pen acceleration with an accelerometer, while the other sensor is a gyroscope that measures pen angular momentum. Third, the chapter investigates the suitability of a nonparametric test (the Kolmogorov-Smirnov distribution) for comparison of on-line features and verification of signatures.

Finally, Chapter 12 summarizes the main conclusions of this thesis, elicits the potential limitations of the presented approaches along with that of the scenarios of their applications, and provides an outlook on future research.

2

Automatic Signature Verification

This chapter provides detailed insights into the various important aspects and working of the state-of-the-art automatic signature verification systems. First, the major building blocks for any signature verification system are discussed. Later, the recent on-line, off-line, and combined/hybrid signature verification systems are reviewed. Finally, a detailed commentary on the state-of-the-art methods generally adopted in pattern recognition for evaluating the results of automatic signature verification systems is provided.

2.1 The Building Blocks

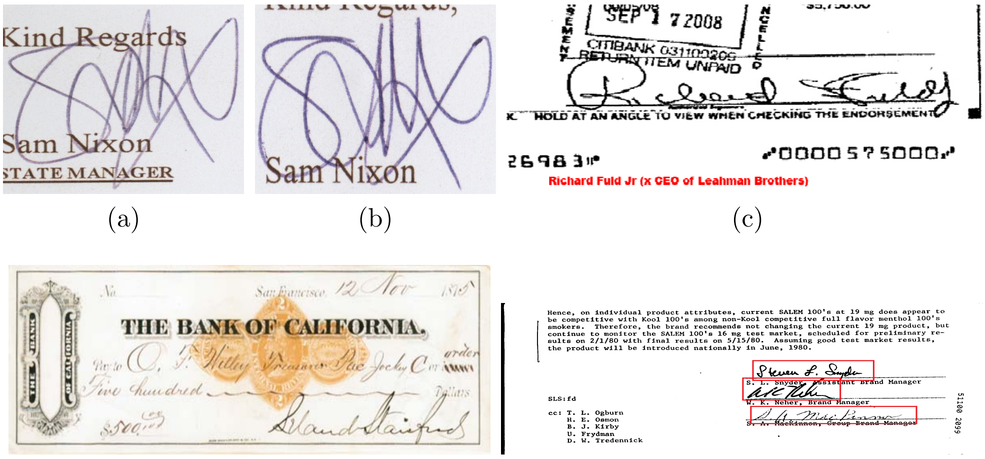
Most of the state-of-the-art automatic signature verification systems are made up of the following three major components:

- Signature preprocessing
- Feature extraction
- Verification

The following sections describe the state-of-the-art of each in detail.

2.1.1 Signature Preprocessing

Generally the aim of signature preprocessing is to reduce as much noise as possible and to get to the actual signature for later performing verification. The common preprocessing approaches are smoothing, filtering, normalizing, and segmenting. Preprocessing can affect all the successive phases in signature verification [12]. There is an argument that


Kind Regards
Sam Nixon
STATE MANAGER
Sam Nixon
SEP 17 2008
CITIBANK
RETURN ITEM UNPAID
HOLD AT AN ANGLE TO VIEW WHEN CHECKING THE ENDORSEMENT
Richard Fuld Jr (x CEO of Leahman Brothers)
THE BANK OF CALIFORNIA.
San Francisco 12 Nov 1875
Five hundred
Dollars
$500
SLS:fd
S. L. Snyder, Assistant Brand Manager
W. K. Neher, Brand Manager
S. A. MacKinnon, Group Brand Manager

(a) (b) (c)

(d) (e)

Figure 2.1: Examples depicting signatures overlapping with the background.

preprocessing is a lossy step– which may potentially lose important information that could have been used as a clue in verification, especially in forensic cases [41]. Nonetheless, it is sometimes a mandatory step as signatures are overlapping, to varying degrees, with other objects present in the document. Depending upon difficulty, signature preprocessing is sometimes considered as a complete separate task [28], especially when the system has to extract/segment signatures from complex backgrounds, e.g., bank checks, legal documents, etc. Figure 2.1[1][2] shows such examples.

The most common off-line signature extraction/segmentation techniques are based on analysis of connected components, analysis of tree structures obtained from projection profiles (horizontal and vertical), and local features (SURF, SIFT). Furthermore, statistics of directional data characterized by local orientation of gradients are also considered to perform signature segmentation [12]. Djeziri et al. [42] proposed an approach to extract signatures from bank checks. This approach is inspired from human visual perception and is based on filiformity criteria whose specialized task is to extract lines. Based on the filiformity measure, the contour lines of objects are differentiated from the handwritten lines. Madasu et al., [43] proposed an approach based on sliding window to calculate the entropy and finally fit the window to signature block on bank checks. Zhu et al. [44] proposed an approach based on multi-scale structural saliency map for signature detection in document images. Zhu et al. [45] further improved their signature detection method by combining it with signature matching to provide a complete framework for

[1] http://bartbaggett.com/blog/lehman-brothers-ceo-handwriting-analysis/

[2] http://www.glabarre.com/item/Leland_Stanford_Check_Bank_of_California_SOLD/2192/p6c41

document verification based on signatures. Mandal et al. [46] proposed an approach using conditional random field for segmentation of signatures from machine printed documents.

For on-line signatures, segmentation techniques could be devised directly from the acquired signals. The simplest approaches get to the blocks of signatures by considering signatures as sequences of writing (pen-down) and interruptions (pen-up). Advanced techniques also take into account the velocity of writing and changes in angles [12]. Dynamic Time Warping (DTW) is one of the most common techniques used for dynamic signature segmentation: when to segment two or more signatures with equal number of perfectly corresponding segments. The underlying approach is to split the first signature based on the positions of geometric extremes/uniform spatial criteria and then finding the corresponding points in other specimen signatures [47, 12]. Furthermore, DTW could also be applied to segment on-line signatures on the basis of reference models generated from the specimen signatures [12, 48, 49, 50].

2.1.2 Feature Extraction

Feature extraction for automatic signature verification pivots around two diverse types of features, i.e., functions and parameters. Function features are generally considered in case of on-line signature verification. Examples of on-line function features include, velocity, acceleration, position, pressure, force, inclination, and direction of pen movement. Parameter features for on-line signature verification include, number of pen ups/downs, orientation of signatures, and various statistics computed by applying different mathematical transforms, e.g., Fourier, Hadamard, Cosine, and Wavelet transforms [51, 52, 53, 54]. These parameters can be computed locally or globally. In general the features describing the overall aspects of a signature are considered to be global. In contrast, the features describing specific portions of a signature are considered to be local (computed by components, e.g., grids, windows; or on point to point basis). An observation about global and local features is that for global features (even when they are extracted using grids), the number of total features remains constant but local features may get affected by the nature of handwritten input resulting in different number of overall extractions [41].

The typical parameter features for off-line signature verification include -but are not limited to - image area, signature height to width ratio, centroid, medoid, histogram of oriented gradients, local binary patterns, edge-hinge features, run length features, geometrical features (based on number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures, chain codes and edge based directional features), and various texture features based on intensity. Morphological shape descriptors, e.g., pecstrum–obtained by successive morphological openings [55], are also used for performing signature verification.

Recently, various local features (Speeded Up Robust Features, Scale Invariant Feature Transform, Features from Accelerated Segment Test, Fast Retina Keypoints, etc.) which have previously shown very good results for different computer vision, and recognition task, e.g., character recognition [56], are applied to signature verification [57, 58, 40, 59, 60, 61]. These local feature based approaches have even outperformed various state-of-the-art global features based systems for different tasks [62, 41, 63]. Franke et al. [64] highlighted the importance of local features by presenting an empirical study on the kinematics of signatures and formalizing that the signing behavior of writers mostly differs in terms of local characteristics, and global characteristics are comparatively easily perceived and forged. Nonetheless, global features have also shown some convincing results [65], however further investigation is required regarding suitability of global and local features for different types of signature data. In addition, some approaches combining local and global features, e.g., by shadow-codes [66], for signature verification have also been presented [67, 68].

Associated with feature extraction, a very important aspect is feature selection (which features to consider to get to the succinct and optimal feature set). Various feature selection approaches have been proposed primarily based on Principal Component Analysis [69], Self-Organizing Feature Maps [70], Sequential Forward/Backward Search [71], Inter-Intra Class Distance Ratios [72], Genetic Algorithms [73], and Neural Networks [74].

In the last few years, special attention has been given to analyze stability of signatures. Humans show intra-writer or within-writer variations when they write signatures. It is a common observation that if a person writes her/his signatures “X” times, even by keeping the writing positions, paper, pen, postures, and etc., similar, the signatures vary to a lesser or a greater extent. The analysis of signatures so that to identify the areas where the signatures were almost always very similar/stable and selection of features from those areas has been shown to deliver very good results both for off-line [75, 76, 77, 78, 40] and on-line verification [79, 80, 81, 82, 83].

Most recently, the signature verification research has been largely influenced by the work of Forensic Handwriting Examiners (FHEs). Various signature verification workshops, e.g., AFHA-2011, AFHA-2013, AFHA-2014, have been organized in order to make PR researchers aware about the work of FHEs and the signature features which are exclusively considered by FHEs. Note that it is not always easy to represent every feature considered by FHEs with corresponding features analyzed by machines and vice versa; yet, close representatives could be found [12, 14, 84, 85, 86, 87]. However, it remains an open question that how closely an extracted feature is a representative of the actual feature considered by a forensic expert and how the relationship can be mathematically grounded.

2.1.3 Verification

After feature extraction, the next step is to perform verification, i.e., establishing whether a questioned signature belongs to the claiming person or not. In case of on-line verification, the most common approaches are based on Dynamic Time Warping (DTW). This is because the signature matching is usually effected by random variations, like writer hesitation or pauses, which create additions and deletions in the signature signal being recorded. The DTW allows expansion and compression of the time axis of signature signal sequences to obtain the minimum of a given distance measure, which makes it suitable for on-line verification [88]. Various data reduction techniques have also been used in the verification step as to ideally remove the less important/noisy signals, such as Principal Component Analysis (PCA), Minor Component Analysis (MCA), Linear Regression (LR), Polygonal Approximation (PA), Extreme Points (EP), and Random Selection [12]. Further approaches like, fuzzy logic, relaxation matching, and split and merge mechanisms are also used [89, 79, 90]. Distance measures like Mahalanobis and Euclidean are commonly used for verification when parameters are used as features [12].

Verification strategies based on the use of Support Vector Machines (SVM) have shown potential for both off-line and on-line verification [91]. Hidden Markov models (HMM) are used quite often since they are largely conformable to intra-writer variations [35]. Neural Networks (NN) are used for signature verification as they show high generalization capabilities, some examples include, multi-layer perceptron, Bayesian networks, Time-delay neural networks, ARTMAP neural networks, and fuzzy neural networks [36, 92, 93]. It is worth noting that most of the researchers in PR use various models, e.g., NN, SVM, by training them on the positive (genuine) and negative (forgery) samples at the same time [94, 12]. According to FHEs [95], this is unrealistic as in the real world, one can never limit the forgery set since every signature other than the genuine signature written naturally by an authentic author can be a forgery. In real cases, a verification system can have only genuine specimen samples and one or more questioned signatures. If forgeries are used for training, there is always a chance that an automatic system may learn to declare signatures as forgeries when they are coming from the forgers on whom the system is trained [39, 62]. Henceforth, the best approach while using such models is to train them only on genuine specimen signatures. This can be done by following different one class classification approaches like, SVM/NN for one class classification [96, 97, 98, 99, 100, 101, 102, 103, 104, 105].

Furthermore, the NN based approaches are highly influenced by the amount of data available for training as they usually require a large amount of training data. This problem is somewhat tacked by the use of “synthetic data”. Two main approaches for synthetic data generation are,

- Signature generation: Transformations are applied on real signatures and then the transformed versions are fed along with the real signatures for training.
- Individual generation: Models are developed from real signatures of individuals, and then new individuals are created using combination of these models.

The use of synthetic data has improved the results of verification both in terms of error rates and resistance to brute force attacks [106, 107, 108, 109, 110]. However, the use of synthetic signatures for improving the results of verification is usually criticized by forensic experts [39] as such synthetic signatures seem to miss or artificially over-generate realistic artifacts available in real genuine signatures and forgeries, e.g., hesitation, look-backs, etc. Nevertheless, it is an open debate and further research is needed in this direction.

Various Multi-Classifier Systems (MCS), combining different classifiers, have also been proposed for on- and off-line verification [111]. A majority of such systems combine classifiers based on global and local features, which sometimes outperform the individual local and global feature based verifiers. Systems based exclusively on global features seem to neglect individual stroke level information, and local features based systems lack in grasping a holistic view of the signature– and therefore combining the two optimally could improve verification results. However, finding an optimal combination sometimes posses problems as it is data dependent to a large extent.

2.2 State-of-the-Art Off-line Verification Systems

A large number of systems for off-line signature verification have been proposed in the literature. These systems use a wide range of features; global, local, and combined, and have been evaluated over different datasets using different evaluation protocols. Therefore, the verification results are not directly comparable and are presented for completion.

Bajaj et al. [32] presented a system based on global features, particularly–horizontal and vertical projection profiles and contour analysis. Classification is performed by combining decisions with feed forward NN (ADALINE). The use of global projection profiles with DTW has been investigated by Fang et al. [112]. Fang et al. [113] further discuss the sparse data problem and present two methods based on peripheral features and Mahalanobis distance classifier. The system developed artificial training samples by elastic matching, and the results have been shown to improve. Armand et al. [114] considered contour-based global features, a combination of Modified Directional Features (MDF), centroid, skew, surface area, and length. Similarly, Papamarkos et al. [31] presented a system based on global grid and texture features. Deng et al. [115] used global curvature data in a multi-resolution approach where the data are decomposed into signals using the Wavelet transform. Statistical measures are then applied to get to the most stable and writer dependent data (contours and frequencies), that are used for final classification. Another global feature based system is presented by Drouhard et al. [116]. This system analyzes global shape factors (directional probability function) and uses back propagation neural networks for classification. Wavelet based global features in conjunction with geometrical and statistical features are used by Fadhel et al. [117]. The wavelets were opted as they helped in data reduction and feature selection. Classification was performed with NN.

Ferrer et al. [118] performed a comparison of different systems based on Euclidean distance, SVM, and HMM using geometric features extracted from signature contours and strokes, where HMM outperformed other approaches on a dataset containing random and simple forgeries[3]. Huang et al. [119] extracted various geometric parameters by scaling signatures at different magnitudes and then using MLP for final classification. Prashanth et al. [120] propose to use pixel density at several geometric locations for verification. Some methods, like the one proposed by Ramesh et al. [121], combined various geometric, global, and wavelet features and applied genetic approaches to reach final classification. Pattern matching by extracting, thinning, and blurring signature strokes has been tested for off-line signature verification by Ueda [122]. Ferrer et al. [123] proposed using the pseudo-cepstral coefficients extracted from global histograms of grey-scale signature images.

[3] Note that the term random forgery should not be used as it is misleading. Various types of forgeries are discussed in detail in Chapter 4

Using fixed grids to divide signatures into n number of regions is a common method of extracting global features [124]. Such features extracted from different regions/cells can be fed to various learning mechanisms like NN [125]. Final classification can be performed by majority voting based on the results obtained from each cell. Further grid based feature selection approaches are presented where feature extraction is performed using circular grid, in contrast to the most widely used rectangular grid [126]. Parodi et al. [126] extract graphometric features, inspired by the work of forensic document examiners, from circular grids based on the discrete Fourier transform and apply SVM for classification. Santos et al. [127] use graphometric features in combination with MLP to perform off-line verification. Sabourin et al. [66] propose using the extended shadow codes as global features and perform classification with the nearest neighbors based classifier.

In addition to the global, geometric, and grid based features some systems based particularly on the signatures' local information have been reported, e.g., systems implying local keypoint detectors and descriptors like, SIFT [128], SURF [21], FAST [4], FREAK [5], etc. These local feature detectors/descriptors were previously used for writer retrieval and identification [60], object and character recognition [129, 130, 131, 56, 132], multiple object tracking [133], object recognition for smart phone platforms [134], and recognition of degraded handwritten characters [135]. All of the above mentioned tasks (including writer/signature identification and retrieval) differ from signature verification in their essence. Through the course of this thesis, the author (Malik et al. [136]) applied various local feature based approaches for off-line signature verification. One such approach based on the use of SURF successfully outperformed the state-of-the-art systems (mostly based on global features) for a forensic like dataset, 4NSigComp2010 [137], containing disguised signatures (where authentic authors imitated their own signatures). The results re-affirmed the power of local features for signature verification, investigated earlier by Franke et al. [64], Richiardi et al. [63], Liwicki et al. [62], and various other researchers [41]. Malik et al. [58, 57] implied combination of local features from SURF, FAST, and FREAK and further improved the results on the same dataset, i.e., the 4NSigComp2010 data [137]. The improvements involved, intelligent feature selection, improved reference model building per writer, and clustering of most consistent features. Srikanta et al. [61] also used SURF and proposed GSURF (amalgamation of Gabor filter based features and SURF features) for off-line verification using SVM.

Stability of signatures has been extensively studied in order to improve the results of signature verification [138, 139, 90]. Sacha et al. [140] emphasized that signature samples from some individuals result in significantly worse performance than other samples from the same individuals. Thus, selection of the most stable specimen samples is important. Pirlo et al. [75, 76] selected signatures' stable regions (from upper and lower contour only) by a multiple matching procedure based on Hamming Distance and used for obtaining local verification responses. Majority Voting is used to combine decisions achieved at regional level and experiments were performed on publicly available GPDS signature dataset where the presented system gave performance on par with the case when complete signatures were used. Pirlo et al. [77] further used an equi-mass segmentation approach to analyze stability and non-uniformly to split signatures into a standard number of regions. Successively, a multiple matching technique is adopted to estimate stability of each region, based on cosine similarity. Xiao et al. [141] used stroke direction-based and grid features with an MLP classifier where the classifier was optimized to focus exclusively the local stable parts of signatures by weighting the corresponding node responses higher. Pirlo et al. [78] exclusively consider local stability in off-line signatures by extracting shape-based features (five equally spaced parallel segments for the horizontal, vertical, + 45, 45 directions, and the total number of black pixels) from each region in the signature image. Local stability was estimated by averaging the degree of similarity between corresponding regions extracted from diverse genuine specimens. During this thesis, the author used the Speeded Up Local features (SURF) and K-nearest neighbors approach for local stability analysis of signatures ([40]) so that to reach out the potential areas from whom the features should be exclusively considered while performing signature verification. Final classification is based on Euclidean distance, and the proposed system outperforms the state-of-the-art by a large margin.

A general observation about signature verification is that the performance of a verification system decreases with the decreasing complexity as the signatures are then comparatively easy to forge, but at the same time if complexity increases greatly– the performance of verifiers may also decrease due to potential unstable signing behavior. So, the question is what should be the complexity of signatures so that to make them optimally difficult to forge but be optimally stable at the same time to increase the performance of verification systems [142, 143, 144]. Found et al. [145] theorized that the success or failure of a forgery attempt is related to the complexity of a signature. They related complexity to the total amount of concatenated strokes, the likelihood of how similar signature characteristics are possessed by different authors, and to the difficulty of producing a forgery. They further presented a model based on complexity using objective measures such as the number of turning points, pen lifts, and intersections found in a signature [146, 147]. Erwin et al. [148, 149] investigated the theorized relationship between signature complexity and the quality of forgeries as postulated by Found et al. [145], and provided some baseline proofs confirming the theory. Avni et al. [150, 151] performed detailed experiments on the complexity of signatures using specialized hardware for tracking the gaze movement of eyes. They studied the eye movements and handwriting dynamics during subjects attempts at forging two model signatures of different complexities. For both the low and high complexity model signatures, the subjects gazed more on the model signatures than on the forged signatures, but more fixations–and of longer durations–were invested on the higher complexity model signature. Results of the study suggest that the complexity classification can be modeled to forgers perceptions of difficulty when forging signatures.

Related to complexity is "legibility" of signatures, which is also expected to have an effect on the quality of a forgery. A highly legible signature somehow gives a forger the advantage to rely on her/his highly practiced and automated handwriting motor skills, assuming the forger practices writing on a regular base. Forging an illegible signature implicates copying a rhythmic motor pattern that is unfamiliar to the motoric repertoire of the forger [152, 153]. Note that, however while forging, a forger also leaves indicia of her/his writing in a signature, so if a forger is mostly relying on her/his own motor repertoire–there is a good chance of getting caught at the end. Fairhurst et al. [152] investigated this issue using two machine experts; first performed global image analysis and used statistical distance measures, the second performed local image analysis and used left-to-right hidden Markov models. In both the cases, the signatures with medium legibility (some portions of signatures were legible while other not) posed the maximum difficulty. Nevertheless, more research is needed to quantify the relationship.

2.3 State-of-the-Art On-line Verification Systems

The most common features in case of on-line data (both spatial and temporal information available) are, position, velocity, pressure, acceleration, pen-ups/downs, inclination angle, etc. Bovino et al. [89] analyzed each stroke for position, velocity, and acceleration and combined results at two levels with soft-hard decisions and various voting strategies. Guru and Prakash [154] performed global analysis of signature signals by considering various statistics like, pen-ups/downs, total duration of signatures, maxima in x-direction, respective maxima in y-direction, and verified signatures based on writer dependent thresholds. Lee et al. [49] focused on getting the global geometric extremes to segment the signatures, and applied Dynamic Programming to perform string matching with features like, average velocity, total number of pen-ups/downs, etc.

Ibrahim et al. [155] proposed an on-line directional analysis of signatures decomposing the signals into directional stroke bands on the basis of relative angles. The approach was tested with both the user-dependent and user-independent cases where the proposed method successfully exploited the inter-feature dependencies. Ibrahim et al. [156] further tested the idea of directional analysis with a focus on the velocity profiles in horizontal and vertical trajectories. Final classification between genuine and forged signatures was performed using SVM.

Jain et al. [157] used local velocity and curvature information to perform verification. Maiorana et al. [158] presented a convolution-based non-invertible transformation based template projection method in connection with hidden Markov models (HMM) for template matching. HMMs are also used by B. Ly-Van et al. [159], where 25 features are extracted from each point in a signature. These features are writer-dependent and are largely related to gesture and local shape. The system was tested on various on-line signature datasets; PHILIPS, MCYT, BIOMET and SVC2004. J. Fierrez [35] proposed the use of HMMs with features based on x-y trajectories, pressure, angle, velocity, curvature, acceleration, and their first-order time derivatives. Shafiei and Rabiee [160] also used HMMs for signature verification where four dynamic and three scale and displacement invariant static features were extracted for all perceptually important points in the input signature. Wu et al. [161] used direction of pen movements and performed experiments with various topologies of HMM–where HMM-LR outperformed the others in capturing the individual features and resisting to variability. Wu et al. [162] used cepstral-coefficients in combination with the Fourier-transform. Wang at al. [163] proposed the use of three dimensional forces applied to all points in the signature signal that are detected by special sensors. Classification was performed using dynamic time warping.

Richiardi et al. [63] studied the effect of using global and local features for on-line signature verification. They extracted global and local features from a 50-users dataset. A large number of global features (more than 150) were reviewed and the most commonly used 46 global features were selected. As for local features, 39 features (13 base local features, with their first and second order derivatives approximated by regression) were considered. In the experiments, the local features outperformed global features.

Some systems, like by Kashi et al. [164], use both the global and local features from the signature signals. The global features provide the holistic temporal and spatial information while local features contribute the dynamics of the signature writing process. Nyssen et al. [165] used global and local information at three different levels of verification. At the first level, various parameter features are extracted and Mahalanobis distance is used for finding dissimilarities between signatures. At the second level, signature segmentation is performed by corner extraction and matching. At the third level, point-to-point elastic matching is performed thereby ensuring high level of security. Pippin [166] performed verification by using global features at the first stage and local features at the second. Stroke segmentation is performed by velocity minima and compared by using dynamic time warping and writer-specific thresholds. Fierrez et al. [167] presented a system capable of using both local and global information by decision-level fusion. The global information is recognized by using Parzen Windows Classifiers and local information by using hidden Markov models. The method based on local analysis outperforme the method based on global analysis when enough training data are available, however for smaller training sets, the method based on global analysis performs better. The two systems show complementary results when combined at the decision level.

Stability of on-line signatures is analyzed by various researchers [79, 80, 82, 83]. Huang et al. [79] model signatures by weighted structure description graph of the handwriting components and consider the most stable strokes with the use of dynamic time warping–operating on signature components to extract correspondences and stability information and calculating the match similarity. A signature is accepted if it is close to a permissible path within the weighted graph. Antonio et al. [80] segment the signature signal into a sequence of strokes, and label them as ascender, descender, or normal, and detect the most stable regions by an ink matcher that finds the longest common sequences of strokes with similar shapes between the ink traces of a pair of signatures. The concept of saliency (proposed to account for attentional gaze shift in primate visual system [168]) is used to decide on the match. Pirlo et al. [82] present a stability analysis approach for signature verification based on DTW algorithm applied to signatures modeled by speed profiles extracted from appropriate parameters of Sigma-Lognormal distributions according to the Kinematic theory. In another work, Pirlo et al. [83] present an experimental investigation on stability of dynamic signatures where they used DTW for estimating local stability in different representation domains and to select the most profitable domain for automatic signature verification.

Brault et al. [142] present an approach to consider complexity and quantify a priori about the possible difficulty that would be faced by a forger when forging a given on-line signature. Gruber at al. [169] used force and inclination features to develop a longest common sub-sequence algorithm which allows considering local variability and then reach a conclusion about the genre of a signature, genuine or forged. Katrin et al. [64] presented an empirical study on the kinematics of signatures considering ink-trace characteristics for different signatures belonging to the three graphical complexity levels–low, medium, and high [170]. Ahmed et al. [171] presented an analysis of handwritten signature dynamics based on signatures complexity and legibility as perceived by human examiners. They investigated a diverse pool of dynamic features extracted separately for groups of genuine authors and forgers. The results demonstrated that pen pressure is the most distinctive and significantly different for genuine and forged signatures. Galbally et al. [172] presented a study of "name legibility" for on-line handwritten signatures. They divided the signatures into two different classes namely: legible and non-legible and performed experiments using global features (Mahalanobis distance was used to rank the features for feature selection) and a multilayer Perceptron. Experimental results demonstrated the feasibility of using legibility for automatic on-line signature classification.

2.4 Hybrid and Combined Approaches

In the last few years, some systems have been proposed which combine or rely on both the on-line and off-line data for performing verification. Mostly, both off-line and on-line data are recorded while training/registering a user with a system [173]. While testing, only off-line data are available, e.g., as in banks while processing a check. Mayank et al. [174] combined off-line (texture and topological) and on-line (pressure, breakpoints, and the time taken to create a signature) features for performing verification. Weighted sum rule was applied to achieve the final matching result. McCabe et al. [175] presented a system that analyzes both the off-line features (e.g., shape, slant, size), and on-line features (e.g., velocity, pen-tip pressure, timing) of signatures and perform verification using HMMs and NNs. In addition to this, a few attempts have been made to glean some representative on-line information from off-line signatures using the so called pseudo-dynamic features mostly using the direction of potential ink flow and pseudo-cepstral coefficients [123]. Some PR researchers are also focusing on related topics like identifying the forger out of a set of suspected forgers who might has forged a genuine signature (the idea is already long studied in forensic science [8]). The underlying assumption is that while forging a signature, all the forgers–no matter how skilled–also leave some clue of their own original handwriting (governed by their neuro-motoric activity). This information can then be used to identify the possible forgers [176].

2.5 Signature Verification Competitions

A common problem faced by many PR applications, including signature verification, is the non-availability or scarcity of data. A lot of research in signature verification is done on the data that are proprietary and are not publicly available as in most cases the contributing authors like to not make their signatures openly available at mass forums. In such cases the experiments are not repeatable and results cannot be authenticated or compared with the other state-of-the-art methods. Nonetheless, various well known signature datasets are publicly available, e.g., Ministerio de Ciencia y Tecnologia (MCYT) [177], Grupo De Procesado Digital De Senales (GPDS) [178], Sabanci University Signatures database (SUSIG) [179], and so on, but the problem is that the machine learning methods not only evolve but sometimes are developed exclusively for particular datasets. In this case after a certain amount of time such datasets are not very helpful for the research community, and thus it is argued that there should be expiration dates for such data. A solution to this is to design benchmarks, but benchmarks also have limited utility and could be stale [39]. A better approach, therefore, is to organize signature verification competitions on regular basis and provide previously unseen data every time and follow the same evaluation protocol. In this way, the performance of verifiers could be judged in a comparatively balanced manner.

Over the last decade, several evaluation procedures [180], standards [181], databases, and signature verification competitions (for both on-line and off-line data) are organized. Note that the author of this thesis has remained actively involved in organizing (in collaboration with forensic scientists) the most recent signature verification competitions and they make an important contribution of this dissertation since they verify the various postulates laid in this thesis. The details of the competitions namely; ICFHR Forensic Signature Verification Competition (4NSigComp 2010), ICDAR Signature Verification Compatition (SigComp 2011), ICFHR Forensic Signature Verification Competition (4NsigComp 2012), ICDAR Signature Verification and Writer Identification Competition (SigWiComp 2013), and ICDAR Signature Verification and Writer Identification Competition (SigWIcomp 2015) is provided in Chapter 5. This section, therefore, provides brief details about other competitions. An overview of the results of previous competitions is provided in Table 2.1.

2.5.1 First Int. Signature Verification Competition (SVC) 2004

This competition focused on on-line data only [179]. Two verification tasks were defined. The data for the first task contained only coordinate information and for the second task additional information like, orientation and pressure were provided. The dataset consisted of 100 sets of signatures with 20 genuine signatures and 20 skilled forgeries, each. The random forgeries[4] were just genuine signatures of other subjects in the dataset. The signatures were in English and Chinese. Performance Evaluation was done by Equal Error Rates (EER), calculated when the False Reject Rate (rate at which genuine signatures are misclassified as forged by a system) is same as the False Accept Rate (rate at which forged signatures are misclassified as genuine by a system). The results of the winning systems are provided in Table 2.1.

2.5.2 BioSecure Multimodal Evaluation Campaign 2007 & BioSecure Signature Evaluation Campaign 2009

In 2007 the BioSecure Multimodal Evaluation Campaign (BMEC)–including the signature modality–was organized. Its objective was to evaluate several biometric algorithms. It had a large scale multi-modal database acquired in 11 sites under 3 different acquisition conditions corresponding to different application scenarios [182].

The BioSecure Signature Evaluation Campaign (BSEC) 2009 used the BioSecure Signature Corpus DS3-382 and DS2-382 databases. Those databases were acquired in a mobile scenario on a PDA for DS3, and on a digitizing tablet for DS2 [183]. Both datasets comprised on-line data from 382 people. The aim of BSEC 2009 was to assess the influence of a mobile platform on the algorithms performance on large databases, to measure the algorithms performances according to information content in signatures and to evaluate the impact of time variability[5]. Again, the EER was used to report the results. The results of the winning systems are provided in Table 2.1.

2.5.3 BioSecure Signature Evaluation Campaign, Evaluation of Signature Resistance to Attacks (ESRA) 2011

In this competition, on-line signature verification systems were evaluated in terms of resistance to time variability, resistance to changes of platform (fixed vs. mobile platform)

[4] The use of so-called random forgeries is quite contrary to the requirements posed by forensic science. This issue is discussed in detail in Chapter 4

[5] http://biometrics.it-sudparis.eu/BSEC2009/

and finally in terms of quality of the genuine signatures [184]. The DS2-382 and DS3-382 datasets were used. Systems were tested for two tasks. Task 1 focused on assessing the impact of forgery quality (good or bad) on the systems' performances. Task 2 focused on testing systems on the appropriate categorization of skilled forgeries, namely that depending on the chosen representation of signatures of each participant [184]. The results of the best systems are provided in Table 2.1.

Note that FHEs observe various issues related to these signature verification competitions. For example, the data used in SVC 2004 were gathered in a way where participants/data contributors gave data by providing their pseudo signatures, i.e., they developed/practiced a new signature and then gave it as their genuine signature. Since the level of practice may vary from one individual to another these data may not be representative of subjects under question. In addition, all competitions mentioned in this sections exclusively focused on the on-line data. However, until now off-line signature verification is the common type of criminal casework for a forensic expert [137]. Furthermore, these competitions did not look into different genres/categories of signatures as encountered by forensic examiners, and the results were not reported the way required in forensic casework. In summary, the underlying motivation of organizing these competitions differed from FHEs' requirements. For this purpose, in 2009 FHEs joined with PR-experts for the very first time and organized a signature verification competition (ICDAR SigComp2009) targeting the very basic signature verification requirements of FHEs.

2.5.4 Signature Verification Competition (SigComp) 2009

The SigComp 2009 focused both the on-line and off-line skilled forgery detection. Two unpublished signature datasets were used. The training set was composed of $1,920$ images from 12 authentic writers (5 authentic signatures per writer) and 31 forging writers (5 forgeries per authentic signature). The test set consisted of authentic signatures from 100 newly introduced writers (each writer wrote his signature 12 times) and forged signatures from 33 writers (6 forgeries per signature). Each authentic signature was forged by 4 writers. Altogether it had $1,953$ signatures available in on-line and off-line format. The systems were expected to calculate a similarity score between one questioned and one reference signature. The systems were evaluated based on their Equal Error Rate (EER) and the conclusions were presented in a probabilistic way [185]. The results of the best systems for different tasks–on-line, off-line, and combined, are provided in Table 2.1.

Table 2.1: Statistics of some previous competitions. Abbreviations used: Signature (S), Genuine Signatures (G) , Forgeries (F), Random Forgeries (RF), Simple Forgeries (SF), Skilled Forgeries (SK), Regain Fogeries (RgF), Number of Authors (A), Bad Quality Forgery (BQ), Good Quality Forgery (GQ). For definitions consult concerned publications.

Competition	**Datasets**	**Task**	**Results (% EER)**
SVC 2004	English and Chinese: 80 (A), 100 (S) sets, 20 (G)/(A), 20 (F)	T1: Coordinates info. T2: Additional info. e.g., orientation, pressure	T1: 1.85(RF), 2.84(SF) T2: 1.70(RF), 2.89(SF)
BMEC 2007	One of the 4 monomodal Data Sets: DS3 (PDA), 430 subjects	T1: Random forgeries T2: Skilled forgeries T3: Regain forgeries	T1: 4.03 T2: 13.43 T3: 10.71
BSEC 2009	DS2-382 (digitizing tablet), DS3-382 (PDA)	T1: On DS2-382 T2: On DS3-382	T1: 0.51(RF), 2.20(SF) T2: 0.55(RF), 4.97(SF)
SigComp 2009	100 (A), 1920 (S) , 12(G)/(A), 6 (F)/(A)	T1: On-line verification T2: Off-line verification	T1: 2.85(SK) T2: 9.15(SK)
BSEC 2011	DS2-382 (digitizing tablet), DS3-382 (PDA)	T1: Impact of forgery quality T2: SK categorization	T1: DS2: 2.73(BQ), 0.85(GQ) DS3: 6.05(BQ), 7.15(GQ) T2: DS2: 3.32(BQ), 4.31(GQ) DS3: 1.67(BQ), 2.43(GQ)

2.6 State-of-the-Art Evaluation Methods

The state-of-the-art evaluation metrics for judging the performance of automatic signature verification systems are primarily based on the number of correct/incorrect judgments passed by the system [186]. The most common metric is the “accuracy” which simply measures the percentage of correct judgments with respect to all the judgments passed by a verifier, as given in Equation 2.1.

$$Accuracy = \frac{Number\ of\ correct\ judgments}{Number\ of\ total\ judgments} \tag{2.1}$$

Though accuracy can be useful in some cases, e.g., when the test set contains equal number of genuine and forged signatures, but it is not optimally informative in other cases. For example, with unequal number of genuine and forged signatures available in the test set–lets say total test cases: 100, out of which 90 are forged and 10 are genuine. Now if accuracy is used as an evaluation metric, and a system only reports all the signatures as forged without actual analysis–even then its accuracy would be 90%, which is not the correct representative of the performance of this system.

To deal with this problem, the performance of automatic verification systems is usually

Table 2.2: Confusion Matrix depicting different types of errors and correct outcomes that appear while evaluating verification results.

	(H_0) is Valid	**(H_0) is Invalid**
Reject (H_0)	$Type - I$ error (False positive)	Correct outcome (True positive)
Accept (H_0)	Correct outcome (True negative)	$Type - II$ error (False negative)

represented in terms of $Type - I$ error and $Type - II$ error (see Table 2.2). Here the outputs of a verifier are tested with respect to a null hypothesis (H_0), i.e., the questioned signature is genuine.

The $Type - I$ error, also termed as False Rejection Rate (FRR) or miss probability, occurs when a genuine signature is rejected by a verification system. More specifically it reflects the ratio between genuine signatures misclassified as forgery and all the genuine signatures available in the test set (see Equation 2.2).

$$Type - I\ error = \frac{Number\ of\ misclassified\ genuine\ signatures}{Number\ of\ total\ genuine\ signatures} \tag{2.2}$$

The $Type - II$ error, also termed as False Acceptance Rate (FAR) or false alarm probability, occurs when a forged signature is accepted by a verification system. It reflects the ratio between the forged signatures misclassified as genuine and all the forged signatures available in the test set (see Equation 2.3).

$$Type - II\ error = \frac{Number\ of\ misclassified\ forged\ signatures}{Number\ of\ total\ forged\ signatures} \tag{2.3}$$

The advantage of calculating the FAR and FRR separately is that one can put different weights on their importance depending on the application, e.g., in some areas–like banking–false rejection potentially does not cause a major financial loss as compared to false acceptance, so here a system could have slightly more false rejection but with very little (ideally zero) false acceptance.

Note that for any verification system, both the errors ($Type - I$ and $Type - II$) can be computed at any decision threshold (operating point of a system). However, computing these errors at any single point can not reflect the actual performance of a verifier and therefore we need complete performance curves. The total performance of a verification system is usually reflected by a 2D-Receiver Operating Characteristic (ROC) curve. Traditionally, an ROC curve represents correct detection probability (1-FRR) versus false acceptance rate at all possible decision thresholds on whom a verifier can operate. Since an ROC curve represents the complete performance behavior of a verifier, often a single scalar value is required to rank the systems in accordance with their overall performance, e.g., in competitions. The Area Under the ROC Curve (AUC) is often used for this purpose. The AUC is equivalent to the probability that a verification system will rank a randomly chosen genuine signature higher than a randomly chosen forged signature [186]. The value of AUC lies between 0 and 1, where 0 is the worst and 1 is the best possible AUC value. The AUC value of a curve representing random guess (a diagonal between (0,0) and (1,1)) is 0.5, any realistic system should have an AUC value greater than 0.5.

A famous variant of the ROC curve is the Detection Error Tradeoff (DET) curve. The DET curves are generated by plotting FRR versus FAR with non-linearly warped axis (warped by standard normal deviates or just by logarithmic transformation). The warping is such that if the score distributions for genuine and simulated signatures are Gaussian, their DET curve will be a straight line [187]. Thus, more linear curves providing a detailed view (highlighting the performance differences) in the region of interest, as compared to the ROC curves, are generated. Some example DET graphs from SigComp2009 signature verification competition [1] are provided in Figure 2.2. This dissertation also represents the performance of various proposed systems using DET curves provided in appropriate chapters.

To give a single value representing the performance of a system in the ROC or the DET space, the Equal Error Rate (EER) is defined as the specific point where the FAR and FRR coincide. The lower the EER, the better the system (EER varies between [0, 100]). Figure 2.2 shows the EER on the diagonal. For the cases when the decision threshold can not be adjusted, e.g., on a final testing set when trying to assess the performance without adjusting any parameter, the Average Error Rate (AER) can be used, which is the mean of the FAR and the FRR.

Along with the above mentioned metrics, various other evaluation metrics are used when automatic signature systems work as identifier instead of verifier (e.g., precision, recall, F-measure, Cumulative Match Characteristic (CMC) curves). However, since these are not directly used for evaluating verification systems, their details are not provided here. For further details, refer to Fawcett et al. [186].

Furthermore, there are certain considerations regarding the data used for evaluation of automatic signature verification systems, for example whether the evaluation data were from the same population from whom the training samples were collected–false identity claims only within the set can occur (closed-set methodology) or the more realistic open

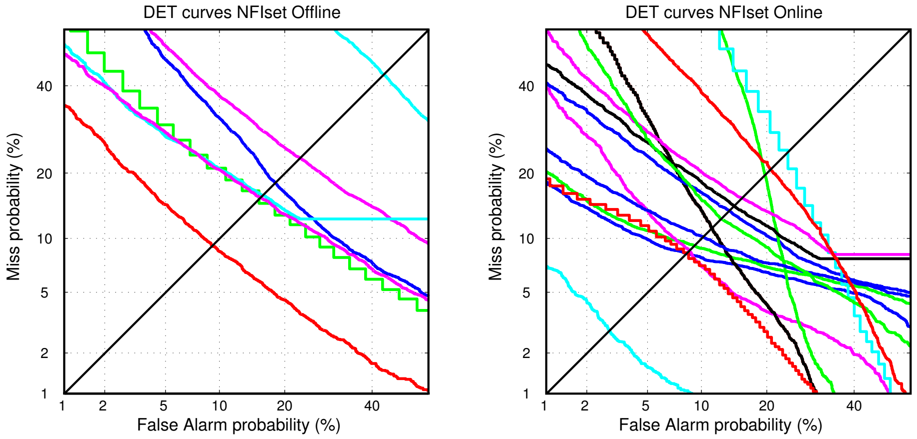


Figure 2.2: DET curves of performance reported by the Signature Verification Competition SigComp2009 for off-line systems (left) and on-line systems (right) [1].

set data collection methodology was followed. Since the data used in evaluation have direct impact on the performance of a verifier, measures such as the entropy are presented [188, 189] which allow to compare the performance of a given system on different subsets of the test data. Another problem related to the data used for evaluating automatic verifiers is to determine what size of test set will ensure statistically significant error results as a function of the expected error rate [190]. Note that the relation between the evaluation set size and the error rates claimed is a topic to be considered because, e.g., an EER of 0.5% has not the same statistical relevance when the size of comparisons is 1000 and when 100000. To cater this problem, Guyon et al. [190] suggest that an expected error rate of around P percent requires at least $N \approx 10000/P$ statistically independent samples to guarantee, with 95 percent confidence, that the expected value of the error rate is not worse than $1.25E$, where E is the empirical error rate of the best system calculated on the test set and a difference of $0.3E$ between the error rates of two recognizers is significant. Mansfield et al. [191] provide further test size estimators which aim at addressing the question: "What is the lowest error rate that can be statistically established with a given number N of comparisons?" However, an important downside of such test size estimators is that they assume the test samples to be IID (independent and identically distributed) which is not true in general signature verification scenarios. Nevertheless, it is an open research topic and involves various aspects that are out of the scope of this chapter. For further details, refer to Mansfield et al. [191].

3

Forensic Signature Comparison: The State-of-the-Art

This chapter provides detailed insights into the various important aspects and the state-of-the-art of signature comparison as performed by Forensic Handwriting Examiners (FHEs). Note that FHEs are more likely to speak about signature *comparison* than about signature *verification*. This is because apart from verification they might also be asked to address the identity of the writer of a forged signature, and both tasks will be approached by a signature comparison. However, as this thesis only concerns verification; the terms comparison and verification are used interchangeably meaning the same, i.e., verification of signatures. The first section of this chapter, Section 3.1, describes how trained FHEs currently perform signature examination in the real world forensic scenarios. In fact "forensic scenario" is an abstract term in itself. Such scenarios can vary from cases where signatures are available directly from the documents (see Fig. 3.1) to cases where signatures/handwriting are available only partially, or are secluded or corrupted by other substances, such as mud or blood (see Fig. 3.2[1]). This thesis deals with the examination of signatures by disregarding the difficulty of obtaining them in particular cases, and focuses on examination and evaluation once the so called "clean" signatures are available. Section 3.2 describes the state-of-the-art instruments that FHEs use in order perform signature and handwriting examination. Finally, Section 3.3 focuses the current PR methods available for FHEs and discusses some of the limitations of these methods.

[1] http://bartbaggett.com/blog/lehman-brothers-ceo-handwriting-analysis/
http://norvalmorrisseau.blogspot.com/2014_02_01_archive.html
http://www.richardfrasermd.com/jfk-marilynmonroe.html
http://www.dfki.de/ liwicki/sigTutorial/
http://www.dfki.de/ liwicki/sigTutorial/ICFHR%Heuvel%20%28high%20res%29.pdf

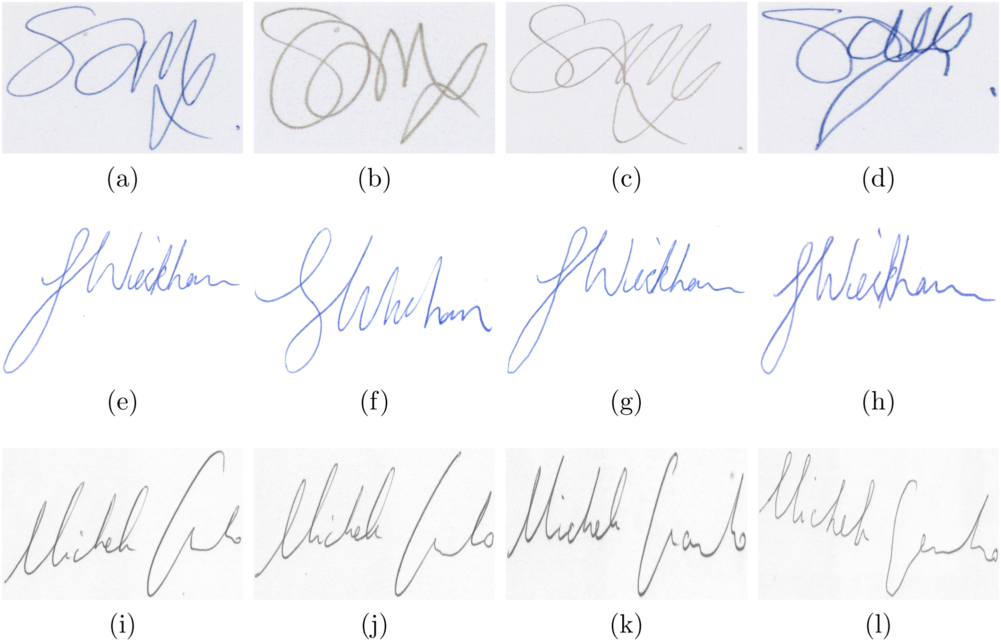
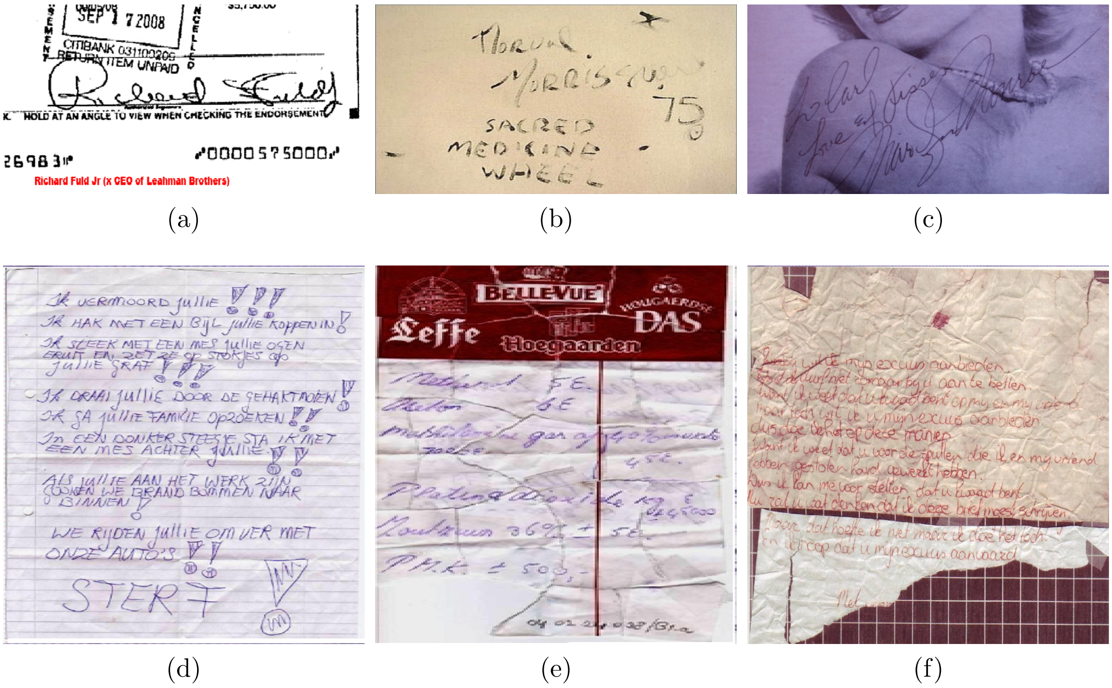


Figure 3.1: Examples showing signatures from different authors, readily available for FHEs so that to attribute authorship.



Ik VERMOORD jullie !!!
Ik HAK MET EEN BIJL jullie koppen IN!
Ik STEEK MET EEN MES jullie OGEN ERUIT EN ZET ZE op STOKJES op jullie GRAF !!!
Ik DRAAI jullie DOOR DE GEHAKTMOLEN!
Ik GA jullie FAMILIE OPZOEKEN!!
In EEN DONKER STEEGJE STA IK MET EEN MES ACHTER jullie. !!
ALS jullie AAN HET WERK ZIJN GOOIEN WE BRAND BOMMEN NAAR BINNEN!
WE RIJDEN jullie OMVER MET ONZE AUTO'S !!
STERF !

Hierbij wil ik mijn excuus aanbieden.
Want ik durf niet zomaar bij u aan te bellen.
Want ik weet dat u kwaad bent op my en my vriend
maar toch wil ik u mijn excuus aanbieden
dus doe ik het op deze manier.
Want ik weet dat u voor de spullen die ik en my vriend
hebben gestolen hard gewerkt hebben.
Dus ik kan me voor stellen dat u kwaad bent
Nu zal u wat denken dat ik deze brief moest schrijven.
Maar dat hoefte ik niet maar ik doe het toch!
En ik hoop dat u mijn excuus aanvaard
Met

Figure 3.2: Examples depicting signature/handwriting that could appear in forensic cases.

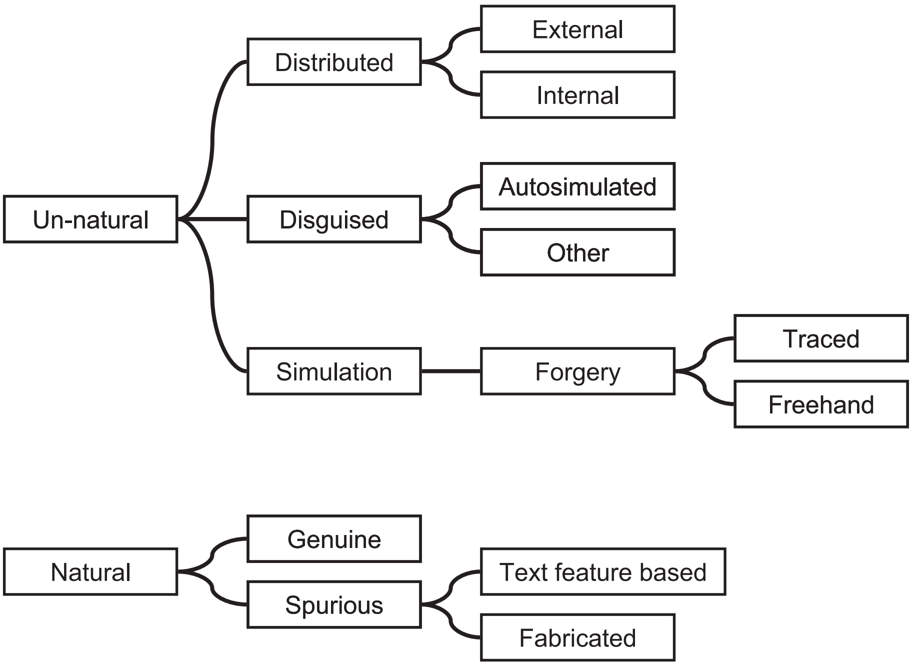


Figure 3.3: Types of Signing Behaviors

3.1 Forensic Signature Comparison

Contrary to pattern recognition, FHEs do not see signature verification as a two-class classification problem [137], rather they consider possibilities of various natural and unnatural handwriting behaviors, as shown in Figure 3.3 [192]. In fact, an FHE conducts a forensic comparison of signatures to express an opinion as to which of the following hypotheses the examination findings support:

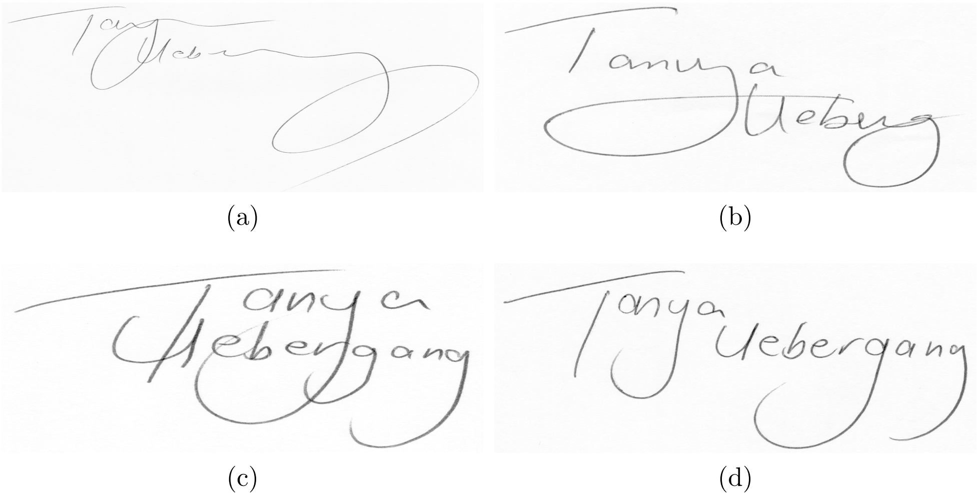
- The questioned signature is genuine, i.e., naturally written by the writer of the reference signatures (the specimen writer).
- The questioned signature is spurious, i.e., naturally written by a person in his/her own style but claiming to be some other person (fabricated–not seen other person signatures, text-feature based–seen other person's signatures).
- The questioned signature is not genuine, i.e., unnaturally written and is a product of simulation (forgery) behavior by a writer other than the specimen writer. The forgery can be made free-hand or can be traced from the actual signatures of a person.
- The questioned signature is not genuine, i.e., unnaturally written and is the product of disguise behavior by the specimen writer, i.e., the specimen writer tries to imitate

a forgery attempt. The disguised behavior is referred to as "auto-simulation" if the writer imitated a forgery (made his/her signatures look like a forgery) with the purpose of later denial, i.e., if a later denial is intended, the disguised behavior falls into the category of auto-simulation.

- The questioned signature is not genuine and has been unnaturally written by the sample writer under the influence of internal or external factors. Note that in this case the intention of the sample writer was to produce a genuine signature, but the appearance of the signature is not genuine anymore (see below).

Note that both the genuine and disguised signatures are written by the same authentic author but with different intention. A genuine signature is written by an authentic author with the intention of being positively identified when needed. A disguised signature, on the other hand, is written by an authentic author with the purpose of making it look like a forgery. The purpose of making disguised signatures can be hundreds, e.g., a person trying to withdraw money from his/her own bank account via signatures on the bank check and trying to deny the check signatures (the so called auto-simulation) after some time, or even making a false copy of his/her will etc. Potentially whatever the reason is, disguised signatures appear in the real world and FHEs do face them [137, 193]. Figure 3.4 shows some example signatures. Figure 3.4a shows the genuine signatures of a person and Figure 3.4b shows a forgery attempt for the same signatures. Here even with the naked eye (FHEs have the additional possibility to view the signatures under various optical devices etc.) the symptoms of a forgery, like., hesitation, feedback, and uneven ink spills are visible. The Figures 3.4c, and 3.4d show two disguised attempts by the original author where the disguised signatures appear more close to the forged signatures (Figure 3.4c to Figure 3.4b, respectively) as compared to genuine signatures (to Figure 3.4a). Such disguised attempts thus pose problems in classification for FHEs [86]. Similarly, consider Figure 3.5 where genuine authors have tried to disguise their signatures and also provided the information how they tried to disguise. The first author wrote the strokes in blocks and paused before he changed the stroke direction while disguising. The second author deliberately separated the initial to the last name and looped the "k" while disguising.

During examination, an FHE compares a questioned signature against a set of reference specimen signatures and separately accounts for the similarities and dissimilarities found. If the examination results in discovery of significant similarities, and no or very less dissimilarities; an FHE usually makes the following three hypotheses to explain the observations:



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

Figure 3.4: (a) Genuine signatures of an author, (b) Forged by a skilled forger, (c) and (d) Disguised signatures by the original author.

- The questioned signature is a result of genuine signing behavior / normal signing behavior of the specimen writer.

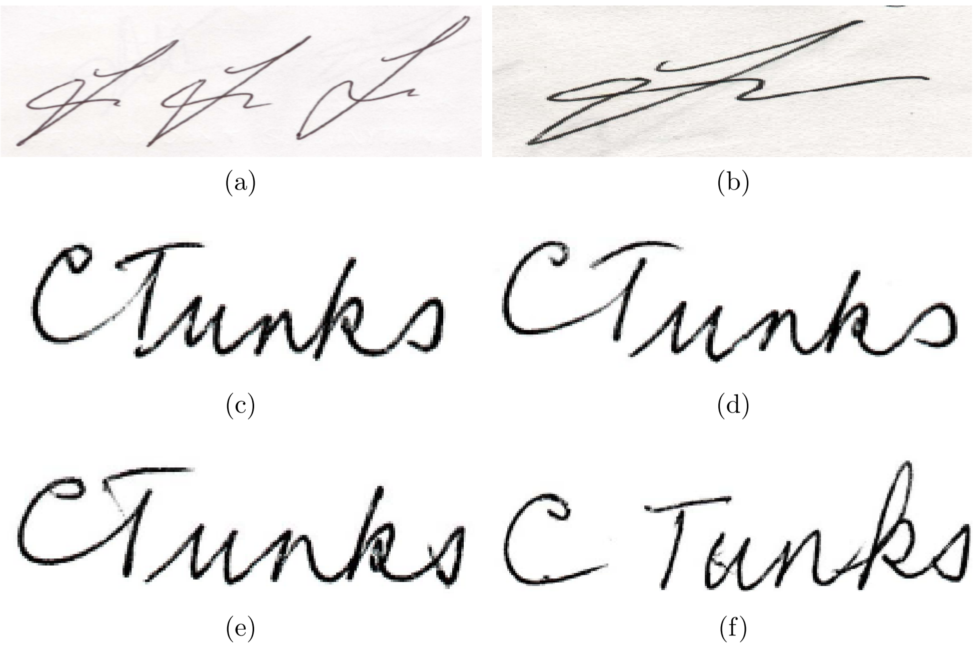
- The questioned signature is a product of "chance-match". This can happen when some writer other than the specimen writer has written the questioned signature where (s)he may have tried to forge the specimen signatures but her/his own natural writing style was "too close" to the specimen writer's style so that s(he) has produced arguably indistinguishable signatures.

- The questioned signature is a product of simulation (forgery) behavior. Although, the FHE was able to find out significant similarities between the questioned signature and the specimen signatures but that was only due to the expertise of the forger, who faithfully and successfully forged the specimen signatures. Furthermore, the striking similarity can be a result of tracing.

If an FHE finds significant dissimilarities along with some similarities, four hypotheses can be generated to explain the observations;

- The questioned signature is a product of simulation (forgery) behavior: The forger might have been inexperienced, or have not practiced enough so that make a forgery significantly similar to the specimen signatures.

- The questioned signature is a product of disguise behavior: The questioned signature was written by the specimen writer but s(he) purposefully tried to make their



(a) (b)

(c) (d)

(e) (f)

Figure 3.5: (a) Genuine signatures of an author, (b) Disguised signatures (wrote the strokes in blocks and paused before changing the stroke direction), (c), (d), and (e) Genuine signatures of another author, (f) Disguised signatures (separated the initial to the last name and looped the “k”).

signatures look like a forgery, i.e., made then significantly different.

- The questioned signature is a product of internal disturbances that resulted in altered signing behavior: Such behaviors are a result of any derangement to the motor control system of the writer, e.g., illness affecting human motor control, or a person under the influence of liquor, etc.

- The questioned signature is a product of internal disturbances that resulted in altered signing behavior: Such behaviors are a result of disturbances due to factors such as the stability of the supporting medium, the texture of the backing surface over which the signature is being executed, or a significantly unaccustomed or awkward writing position [192].

On finding significant dissimilarities along with no similarities, an FHE can suspect a fabricated signing behavior: Such signatures have no similarity/similar features compared to the specimen signatures except that which may have arisen by chance. These signatures can be a result of a disguise attempt by the specimen writer, or a forgery attempt by a forger where the forger has opted not to follow the model/specimen signatures that might be available to her/him, or simply the forger has no access to genuine specimen signatures for training before forging.

During examination, an FHE compares a questioned signature submitted for examination against various known items of established origin (known specimen signatures) associated with the case under investigation. The foremost requirement for such a comparison is that the writing / signatures (both the questioned and the specimen) must be of the same type, i.e., handwritten/cursive versus hand printed [194]. The most common type of such comparisons are carried out in order to investigate the possibility of a forgery (an attempt to imitate/duplicate the writing of another person) or a disguise (an attempt to avoid identification as the writer) [195]. The purpose of a comparison is to examine writing/signing features/characteristics. These characteristics could be those taught to the writer, e.g., in a class when s(he) learned writing, her/his own personal writing attributes developed over time, or subtle characteristics. Comparisons are made on the assumption that there is a very high likelihood that no two writers write the very same way, along with that all the writers suffer from the so called intra-writer or within writer variations when they write/sign. Thus, an FHE takes into account the inter-writer variations–how handwriting features vary across a population of possible writers and intra-writer variations–how the writing of a writer vary from sample to sample. In order to show that the two signatures were written by the same writer, an FHE must show that the degree of variation of the samples under examination is more consistent with the intra-writer variations then with the inter-writer variations, and vice versa. Though, in some forgery cases, e.g., tracing, intra-writer variations are so low that an FHE can consider high likelihood of a forgery. After accessing the different hypotheses under the light of her/his examination, an FHE reports her/his opinion regarding the extent to which the evidence, found in the signatures, supports one of the competing hypotheses.

The results of a signature comparison, carried out by an FHE, are subjective. However, terminology has been developed so that to express such subjective conclusions in a meaningful manner. Several scales, such as a five-point scale and a nine-point scale, are used by FHEs worldwide [194, 196, 197, 2].

The nine-point scale is as follows:

1. Identification (a definite conclusion that the questioned signature matches the specimen signatures)

2. Strong probability (major evidence of a match is found, yet some critical quality is missing)

3. Probable (strong steering towards identification)
4. Indications–weak opinion (a few significant features are found which hint that the same person might have created the specimen and the questioned signatures)
5. No conclusion (reached when lack of material, e.g., very limited number of specimen signatures, or limiting factors, e.g., very bad quality, disguise behavior etc., are involved)
6. Indications (a few significant features are found which hint that the same person did not create the specimen and the questioned signatures– same weight as indications with a weak opinion)
7. Probably did not (quite strong evidence that the same person did not write the specimen and questioned signatures)
8. Strong probably did not (virtual certainty)
9. Elimination (shows highest degree of confidence of an FHE that the same person has certainly not written the specimen and questioned signatures).

Furthermore, it is also known that *the conclusion of an examination should express the degree of support provided by the forensic findings for one proposition versus the specified alternative(s) depending upon the magnitude of the likelihood ratio (LR), which can be expressed by a verbal equivalent according to a scale of conclusions [197].* An example is provided in Figure 3.6 for illustration purposes only, please refer to [2, 197] for further discussions and details which are beyond the scope of this thesis.

For examination of signatures and handwriting, FHEs use various lighting and magnification instruments. The following section provides a brief overview of such instruments.

3.2 Instruments for Forensic Signature Examination

FHEs use a broad range of lighting and optical devices for examining signatures and handwriting. Some important such devices are [198];

- Stereo Microscope: Binocular stereo microscopes, with a magnification of $6\times - 50\times$ and a fiber optic based two goose neck light sources illumination mechanism, are widely used for examining the disputed handwriting, signatures, hand printing, and commercial printing cases.

- Magnifiers: Hand-held and framed magnifiers having capable of zooming signatures from $3.5\times - 7\times$ are in common use.

- Light Panels: Specialized light panels or light boxes having strong light emitting sources are often used for dating the documents, detecting watermarks on the paper, and checking consistency of the paper when some portion of the paper is deranged.

- MiScope: A small hand-held microscope, connected to computer via USB port and having a magnification power of $12\times - 140\times$. The images can be saved, analyzed, and printed with the software often provided along with.

- Video Spectral Comparators: Cameras with specialized light sources and filters (visible, ultraviolet, and near infrared) are frequently used for differentiating different ink traces and detecting alterations in handwriting (e.g., adding characters, strokes).

- Multi-spectral and Hyper-spectral Imaging: Multi-spectral ($4 - 20$ color channels) and Hyper-spectral (more than 20 color channels) imaging devices are used to discriminate different inks, same inks-with different aging, alterations to signatures, and for many other related applications. A downside of such devices is that they produce files of enormous sizes (as they have many color channels, e.g., 240) and they are usually quite expensive.

- Electrostatic Detection Devices: Such devices are used to analyze indented impressions that are formed when a writing instrument, e.g., pen, forms impressions on a sheet below the document. Particular applications include, anonymous letters, bank robbery notes, and medical records [198].

- Photography and Imaging: Specialized high resolution cameras mounted on quadrapods are often used for recording the case specific documents which are later used for further analysis and demonstration.

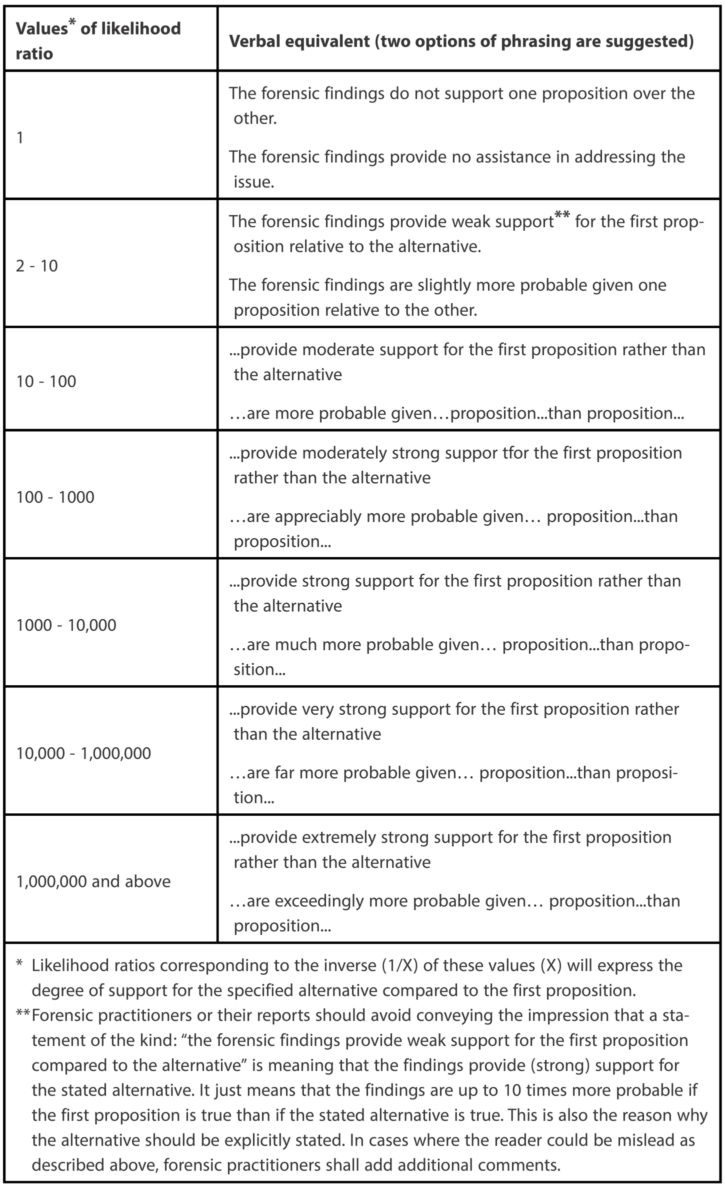
- Specialized Grids: FHEs generally use computer generated grids printed on transparent films for detecting insertions and alignment defects in typewritten or computer generated documents.

FHEs tend to use one or the other above mentioned instruments in their routine casework, however use of these devices requires expertise and also FHEs need quite a fair amount of time to perform such examination. It is thus supposed that the productivity of FHEs can be enhanced by providing them automatic software/programs that would assist them in forensic examination. The following section discusses this issue briefly.

3.3 Automatic Systems for Forensic Examination

Some automatic tools exist that have been designed specifically for supporting FHEs in their daily case work, e.g., CEDAR-FOX [15], FISH [16], WANDA project framework [17], FLASH ID [18], iFOX [199], and D-Scribe [200]. It is, however, interesting that with respect to signature authentication, FHEs have traditionally made very limited use of such automated tools. This may be because many of these tools have been designed to perform comparison tasks and present results in a form that FHEs are not comfortable using. Some tools have certain drawbacks which affect the possibility of their usage in the real world forensic casework. For example, CEDAR-FOX, introduced by the Center of Excellence for Document Analysis and Recognition for supporting a semi-automated analysis for handwritten input, has a drawback that its output is based on the assumption that the appearance of each occurrence of a character is independent of the next. If this assumption were true there would be no point in comparing handwriting with the purpose of source attribution. The behavior of the system differs from that which would be expected in many forensically relevant scenarios [14].

It is also the case that sometimes automated tools (e.g., Forensic Information System for Handwriting (FISH)–which automates several measurement tasks usually done by FHEs) are not available to FHEs outside of the agencies or specific organizations. Further, the FISH toolkit does not include the recent PR methods and comes with an old-fashioned graphical user interface. To overcome these issues, the WANDA project [17] aims to realize a framework for writer identification and handwriting examination. This framework incorporates already several state-of-the-art PR methods and can be used by FHEs. The Forensic Language-Independent Analysis System for Handwriting Identification (FLASH ID) identifies writers by analyzing graphemes topology and geometric features. The system iFOX (Interactive FOrensic eXamination) and D-Scribe are introduced most recently to convert and assist in handwriting interpretation through automation. Unfortunately, none of these systems is in such wide spread use so that its utility could be fully authenticated.



Values* of likelihood ratio	**Verbal equivalent (two options of phrasing are suggested)**
1	The forensic findings do not support one proposition over the other. The forensic findings provide no assistance in addressing the issue.
2 - 10	The forensic findings provide weak support** for the first proposition relative to the alternative. The forensic findings are slightly more probable given one proposition relative to the other.
10 - 100	...provide moderate support for the first proposition rather than the alternative …are more probable given…proposition...than proposition...
100 - 1000	...provide moderately strong suppor tfor the first proposition rather than the alternative …are appreciably more probable given… proposition...than proposition...
1000 - 10,000	...provide strong support for the first proposition rather than the alternative …are much more probable given… proposition...than proposition...
10,000 - 1,000,000	...provide very strong support for the first proposition rather than the alternative …are far more probable given… proposition...than proposition...
1,000,000 and above	...provide extremely strong support for the first proposition rather than the alternative …are exceedingly more probable given… proposition...than proposition...

* Likelihood ratios corresponding to the inverse (1/X) of these values (X) will express the degree of support for the specified alternative compared to the first proposition.

**Forensic practitioners or their reports should avoid conveying the impression that a statement of the kind: "the forensic findings provide weak support for the first proposition compared to the alternative" is meaning that the findings provide (strong) support for the stated alternative. It just means that the findings are up to 10 times more probable if the first proposition is true than if the stated alternative is true. This is also the reason why the alternative should be explicitly stated. In cases where the reader could be mislead as described above, forensic practitioners shall add additional comments.

Figure 3.6: Verbal Representation of Likelihood Ratios [2]

4 Approach

Handwriting has been generally considered a representative of human behavioral characteristics. With the evolution of modern computing technologies, researchers have moved towards the automated analysis of handwriting. This shift has been reinforced by the interest various industries have in this field. One of the potentially most important applications of automated handwriting analysis systems is in forensic science.

Forensic handwriting/signature analysis however, has until now been carried out without actual application of automated systems. This is because there is a large gap between the needs of Forensic Handwriting Examiners (FHEs) and the existing methods of the Pattern Recognition (PR) community. The underlying issue is that most state-of-the-art handwriting/signature analysis systems cannot be directly applied to forensic cases. Current systems make a decision on whether a signature is either authentic or forged, and this chapter argues that this classification is not optimally informative and useful for forensic application. Meanwhile, the PR community has mainly moved by considering the cases which are either trivial for FHEs or not relevant for their daily casework. For instance, the difficult categories of disguised signatures and identifying the author of a forgery seem to be nearly forgotten in most of the current PR research. While these categories were at least mentioned in the last century [8, 9] they do not play a noticeable role in the recent works [12].

The main objective of this chapter is to bridge the gap between the two communities, i.e., the PR-researchers and FHEs. To achieve this objective, this chapter will first point out the major differences currently present in the terminology and its usage between the two communities, PR and FHEs. A common terminology useful for both the communities is suggested. Then, the need of data similar to what are available in forensic cases

[0] This chapter is an adapted version of work published in [20, 201] and submitted in a journal.

is highlighted. This chapter will then argue that the current output reporting schemes available in PR do not address the needs of forensic science. Finally, the use of a relatively simple adaptation of the current comparison methods to successfully address the needs of FHEs and PR simultaneously is proposed and a performance evaluation scheme adapted to such methods is described[1].

4.1 Common Terminology for PR and FHEs

As defined in Chapter 1, PR researchers generally consider signature verification as a 1 *to* 1 classification problem: either a signature is verified to belong to a given person or it is not. It is the task of determining whether or not a signature has been written by a specific author by comparing a questioned signature with a known genuine signature [9]. If a signature is verified to belong to the specified person, the person is called a *client* or *authentic/target/specimen/sample author/writer/signer*, otherwise the person is called an *impostor/forger* or *non-specimen writer* [24, 26]. In the PR community, the impostor's attempt or forgery is usually categorized into the following types.

- Type 1: produced by a forger after unrestricted practice but the level of expertise of the forger is not specified (the forger expertise can vary from a layperson to a professional impostor) [202].
- Type 2: produced by a forger who has no idea about and has not seen the genuine signatures of the authentic author. The forger has no knowledge about authentic author's name. In fact, genuine signatures of any person other than the specimen writer can be considered a forgery attempt of Type 2 against the specimen author [203].
- Type 3: produced by a forger which has knowledge about only the name of a specimen writer [204].
- Type 4: produced by an inexperienced forger after observing the specimen signature for some time without any particular knowledge about the spelling of the specimen author's name [27].
- Type 5: produced by an inexperienced forger, laymen after practicing the forgery for an unrestricted number of times [118] [119].

[1] Submitted

Table 4.1: Types of signature forgery as defined by different authors in the PR community

Authors	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
Justino et al. [203]	–	Random	Simple	–	–	–
Hanmandlu et al. [27]	–	–	Random	Unskilled	–	Skilled
Weiping et al. [204]	–	–	Casual	–	–	–
Ferrer et al. [118]	–	–	–	–	Simple	–
Huang et al. [119]	–	–	–	–	Targeted	–
Nguyen et al. [94]	–	Random	–	–	Targeted	–
Francisco et al. [178]	–	Random	–	–	Simple	–
Yeung et al. [179]	Skilled	Random	–	–	–	–
Bernadette et al. [183]	Skilled	Random	–	–	–	–
Liwicki et al. [137]	Simulation	–	–	–	Simulation	Simulation
Blumenstein et al. [205]	–	Random	–	–	Simulated	–
Bertolini et al. [206]	–	Random	Simple	–	Simulated	–
Liwicki et al. [202]	Skilled	Random	Random	–	Skilled	Skilled
Proposed	Skilled	[1]	Simple	Simple	Skilled	Skilled

[1] Removed in the proposed taxonomy.

- Type 6: produced by an experienced forger - usually a calligrapher - after practicing the forgery for an unrestricted number of times [27].

Note that in the PR community different names are often used for the same forgery type and sometimes the same name is used to refer to the different types of forgeries. Some recent examples are provided in Table 4.1. The discrepancies in the usage of various terms, as depicted in Table 4.1, point to the need for a common terminology for signature forgeries. The following section, therefore, proposes a common terminology that is supposed to be usable and informative simultaneously for both the communities, FHEs and PR researchers.

4.1.1 Recommendations

As depicted in Chapter 3: Figure 3.3, FHEs study various signing behaviors. These behaviors have varying importance in the daily work of an FHE [149]. The most important signing behaviors that FHEs would like to address with automatic systems, are the following.

- Genuine: signatures that belong to the specimen author.
- Forged/Simulated: signatures that belong to someone other than the specimen author.

- Disguised: signatures that belong to the specimen author but were written to look like a forgery.

Regarding the terminology used, agreeing on common definition would favor the application of automatic systems in real forensic casework. Therefore, this thesis proposes that the two communities use the following terms,

- Genuine: for authentic signatures.
- Forged/Simulated: for forged/simulated signatures, which can be simple or skilled.
- Simple forgery: a forgery where actual signatures are known but forgery is produced without any practice. Corresponds to Types 3 and 4 of Table 4.1
- Skilled forgery: same as simple forgery but produced after practice. Corresponds to Types 1, 5, and 6 of Table 4.1
- Disguised: written unnaturally by the genuine author with the purpose of denial at a later date.

This thesis proposes that the terms forgery and simulation may be used interchangeably as both are already in common use in PR, although it is stated that FHEs usually prefer using the term "simulation" compared to forgery, as a the term forgery is more specific to the verdicts of the courts. Note that for skilled forgeries the level of practice is not specified nor the type of forger, e.g., layman or calligrapher or professional forger. This is because some forensic handwriting analysis studies [39, 149] suggest that there are cases where some laypersons produce better forgeries than some calligraphers after practicing a forgery, and vice versa [207, 208]. In such a case, if a forgery is called a skilled forgery only when coming from a professional imposter or a calligrapher, a so-called skilled forgery may not be a good forgery attempt eventually. Furthermore, the performance of a professional impostor remains nearly the same after practicing a forgery by a varying number of practice trials. Therefore, categorizing on the basis of expertise may not be appropriate.

In many cases the PR research reveals results that are trivial/irrelevant at least with respect to forensic casework, e.g., a common practice of PR researchers is to report random forgeries but they are fictitious in the view of forensic experts. If random forgeries are included in a test set while evaluating the results of a system, a system having very low error rate may still not be suitable for forensic casework. On the other hand a system having a high error rate but considering skilled forgeries may yield better results in

Table 4.2: Results of some recent Signature Verification Systems. Here, F=Forgery, G=Genuine Signature, RF=Random forgery (not relevant for forensic casework), SF=Simple forgery, SK=Skilled forgery, SM=Simulated forgery, and T=Total number of signatures.

Study	Database	FAR(%)	FRR(%)
[33]	320(G) 320(F) 640(T)	0.11	0.02
[124]	300(G) 300(F) 600(T)	4.16	7.51
[209]	980(G) 980(F) 1960(T)	0.01(RF),4.29(SF),19.80(SK)	2.04
[127]	300(G) 600(F) 900(T)	4.41(RF),1.67(SF),15.67(SM)	10.33
[210]	2400(G) 0(F) 2400(T)	0.64	1,17
[31]	500(G) 0(F) 500(T)	9.81	3

forensic casework. Examples are reported in Table 4.2. Note that these examples are from PR literature and are presented here just to highlight the difficulty the FHEs face when viewing these results where different types of signature modalities are either differently defined or in some cases are combined with each other while reporting the overall system performance. As given in Table 4.2, Systems [209] and [127] are producing higher error rates with skilled and simulated forgeries than other systems which either do not specify the types of forgeries considered, e.g., Systems [33], [124], or do not use forgeries in evaluation like [210, 31]. By seeing the results like the ones reported by [210, 31], an FHE cannot say anything about which system will potentially suit real forensic casework. However Systems [209] and [127] are reporting their results on skilled forgeries separately which worths more for an FHE.

It is recommended here that the term *Random Forgery or Random Simulation, i.e., genuine signatures of other writers*, which corresponds to Type 2 in Table 4.1, should *not* be used. This is because random forgeries are just fictitious/newly developed signatures. They have usually no or minimal practical importance for the casework of FHEs. However, until recently they were regularly considered in the PR-literature [94, 178, 203, 27, 179, 183]. It is emphasized that if random forgeries are used, at all for some specific PR-research purposes, they must be reported separately from the other types and should not affect the overall performance evaluation. The other types of signing behaviors, e.g., disguised signatures, should be defined the way they are defined by FHEs. For PR researchers, this is a new way of looking at signatures and signing behaviors. However, the adoption of these definitions is important as they are consistent and in-line with the users' point of view, both in the PR community and FHEs. Please refer to Table 4.1 for the proposed classification of forgery types and their definitions.

4.2 Non-accessible datasets and non-representative data

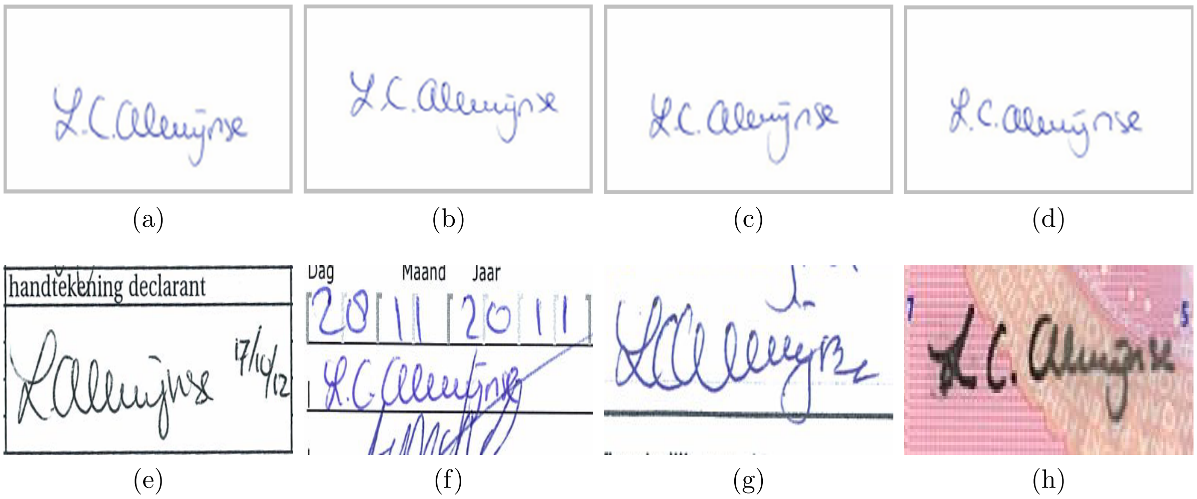
The data collection task is often underestimated in the process of developing signature verification or handwriting identification systems [3]. Like any other PR task, there is a continuous need of new signature data for training and testing the newly developed signature verification systems. Nearly all of the signature and handwriting datasets publicly available today have been collected under controlled conditions. More realistic data (even for other biometric modalities) are scarce. Systems trained on data collected in controlled environments are not much suited to forensic applications, and development of datasets particularly containing cases similar to forensic environments is required. Numerous automatic signature verification systems are reported in the PR literature over the last four decades. Most of them have been trained and tested for data that are collected in controlled environments and which cannot be a good representative of the data available from real world uncontrolled situations (see Figure 4.1 [3]). Any system trained on such data is not very useful for FHEs [137]. To further worsen the situation, many PR systems are not trained/tested on publicly available data and therefore the experiments are not repeatable/verifiable. Due to this, FHEs can never be sure of which systems can potentially be better applied to forensic casework.

4.2.1 Recommendations

To bridge this gap, the PR researchers should use data that are publicly available preferably collected by FHEs in forensic like situations. Through the course of this thesis, a large amount of such data are made publicly available, such as the data from various signature verification competitions jointly organized by PR researchers and FHEs. These include SigComp2011 [202], 4NSigComp2012 [13], SigWiComp2013 [95], and SigWiComp-2015 [211][2]. The author of this thesis has made all of these data publicly accessible from the Int. Association of Pattern Recognition (IAPR) TC-10/TC-11 databases portal[3]. Having different automated systems that report results on the same datasets may provide a comparative analysis of their performances. If application specific data are collected for special purposes they should be unbiased and have statistical significance. Moreover, at least the following most important information is required: Data collection procedure,

[2] available at http://tc11.cvc.uab.es/datasets/type/

[3] http://tc11.cvc.uab.es/datasets/



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

(e) (f) (g) (h)

Figure 4.1: Example signatures written by a writer in controlled environment (a-d), example signatures by the same specimen writer under uncontrolled conditions: e) signature on a declaration form, f) two overlapping signatures with restricted space for signing, g) signature on a receipt, written while standing, and h) signature on an ID-document, dating from 7 years back [3].

any specific restrictions applied while collection, any errors occurred and corrective measures taken, signers' education, sex, handedness, profession, origin, significant amount of specimen and questioned data, time span of data collection, and information about when and if the data will be publicly available.

4.3 Output Reporting

What should an automated signature verification system output in order to be successfully applicable in forensic casework? This is a substantial question for both PR researchers and FHEs.

The output produced usually by automated systems is not acceptable for presentation in the courts thereby making the use of automatic systems nearly impossible for FHEs [202]. Traditionally, automated signature verification systems report their decisions in a Boolean manner, i.e., if enough evidence of a forgery is present, a system reports a reject, otherwise an accept. Though, this is quite objective and may be significant in some fields like real time application, e.g., signature verification in banks, but a Boolean answer of *genuine* or *forged* is not adequate for FHEs. They are interested to exactly know how close is a questioned signature to a genuine signature when it is declared as forged and vice versa.

To bridge this gap, automated systems usually provide some sort of similarity score between 0 and 1, e.g., probability values. Here a value near 0 represents a forgery and a value near 1 represents genuine authorship. This again is inadequate for forensic casework. That is because mere scores/probability values in themselves raise many questions for FHEs and courts: How are these values related to the authorship (genuine or forged) and among themselves; how to compare different systems producing different values for the same questioned signature; how would an FHE establish that a value of 0.2 produced by one automated system is still more close of being genuine signature than a value of 0.4 produced by another system; how would these type of outputs be defended in courts?

Moreover, FHEs are interested to know the features contributing to the output. They would like to consider the features' uniqueness/rarity in a population, e.g., how rare the style of writing a special character in a population is? This information impacts the overall evaluation of an FHE while examining a signature sample. But, how would that relate to an automated system?

4.3.1 Recommendations

[4]

FHEs aim at providing the actual evidential value of the signature comparison to the court/jury. The jury then combines this with the other information and evidence available in the case. This evidential value is defined as the likelihood ratio (LR): the divided probabilities of the results of the analysis given either competing proposition, where an FHE weighs the likelihood of the observations given (at least) the following two hypotheses: [212, 213].

- H1: The questioned signature is an authentic signature of the reference writer;
- H2: The questioned signature is written by a writer other than the reference writer;

An important observation here is that an FHE reports a measure of evidential value, and not just an assignment of a class (genuine or forged). A signature verification system that only reports an assigned class is withholding information. Note that a remark about the general performance of the system can not tell us much about the system's performance in a particular case. A signature verification system can in fact give more

[4] submitted

information about classification rather than the pure classification itself. This is because it knows if the case was a close call (weak evidence), or a clear case (strong evidence).

Figure 4.2 shows two ways of obtaining LRs from the comparison scores reported by an automatic system. The upper scheme shows how the evidential value (LR) of a comparison score E in a particular case is now given by dividing the probabilities of observing $sc = E$ for same source comparisons and for different source comparisons.

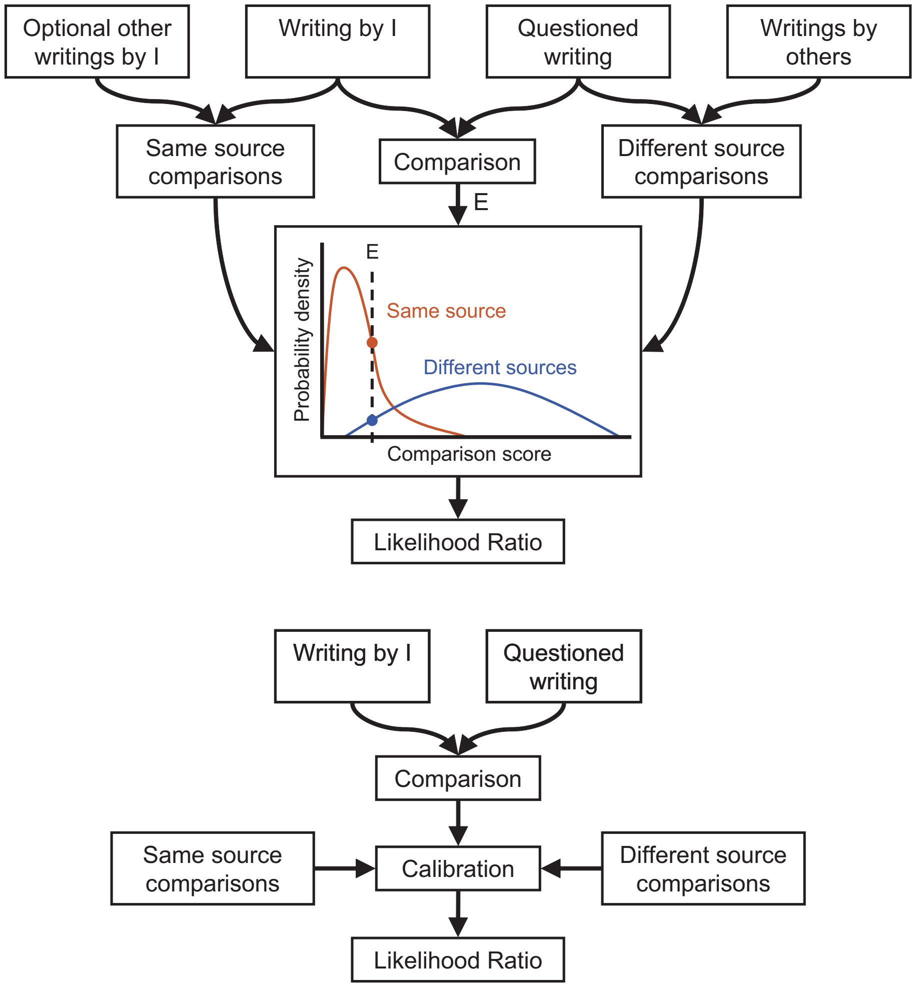


Figure 4.2: Comparison Schemes

The bottom scheme (in Figure 4.2) shows that such scores can also be converted into LRs by a so-called calibration procedure [214, 215], such as the one implemented in the FoCal toolkit[5]. Calibration procedures convert scores into LRs or LLRs (log-LRs) with a monotonic transformation, which therefore does not affect the discriminating power of the system. This has been realized by Brmmer et al. by using the Pool Adjacent

[5] http://focaltoolkit.googlepages.com

Violators (PAV) algorithm [216].

Note that so-far the computation of some kind of likelihood ratios has been realized in different ways only in some tools and frameworks, such as CEDAR-FOX [15][6] and WANDA project [17], but in general the PR community has not adopted the likelihood ratios at large. This makes the application of a majority of the state-of-the-art PR methods impossible in forensic casework.

Furthermore, PR researchers are usually not interested in continuous values like likelihood ratios. They usually demand an objective indicators of a system's performance. The next section deals with this issue.

4.4 Performance Evaluation

The $\widehat{C}_{llr}$ [216] is the cost due to both lack of discrimination and lack of calibration:

$$\widehat{C}_{llr} = \widehat{C}_{llr}^{min} + \widehat{C}_{llr}^{cal}$$

The $\widehat{C}_{llr}^{min}$ is a measure for the discrimination performance of the system, with lower values representing better discrimination. While there are other ways to study the performance of a system that produces LRs (such as the same source and different source LR distributions, or Tippett plots [19]), the $\widehat{C}_{llr}^{min}$ provides a one-dimensional performance metric that is required to directly determine which system performs better [216], e.g., when pointing out a winning system in a competition (as done with the EER or the AUC for classification systems in the past).

For reference, the LLR histograms for some extreme cases are provided in Figure 4.3. The LLR distributions for the true comparisons (genuine signatures) and false comparisons (forgeries) are depicted in green and red, respectively. Three distinct extreme cases are identified. The curves showing the uncalibrated evidential values are on the left and the calibrated LLR histograms obtained after calibrating with the Focal toolkit are on the right. Note that for optimal performance and thus minimal value of $\widehat{C}_{llr}^{min}$, the two curves must be optimally separated. The less overlap, the less cost is associated with the evidential values obtained.

In Case 1 the curves are perfectly separated and shifted towards the desired side. Therefore, the calibrated LLRs have the minimum cost, and without overlap $\widehat{C}_{llr}^{min} = 0$. In Case 2 the distributions of the evidential values completely overlap, i.e., the system

[6] the practicability of CEDAR-FOX is discussed in [14, 217]

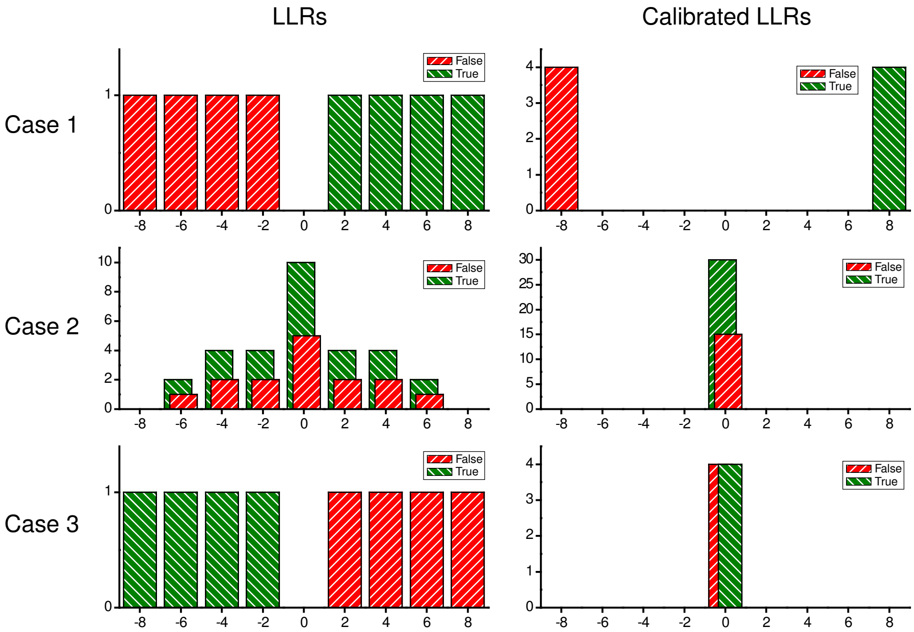


Figure 4.3: LLR histograms, extreme cases

cannot distinguish at all. Thus, for optimal $\widehat{C}_{llr}$ performance, it is best to always output LLRs equal to zero. In Case 3, which is a rather hypothetical case, the curves are perfectly separated but the outputs are always misleading. Again, it is best to map all values to zero. Note that in this case it would be better to invert the output of the uncalibrated system. The latter two cases (the perfectly ignorant system and the perfectly distinguishing but consistently lying system) result in a cost of $\widehat{C}_{llr}^{min} = 1$, which is the maximum value for a calibrated system. An even higher $\widehat{C}_{llr}$ would result if the curves were not calibrated.

In order to demonstrate the feasibility of the above presented evaluation scheme, through the course of this thesis, various signature verification competitions have been organized where this evaluation scheme is thoroughly tested. These competitions are an important part of this work where the author has not only focused on bridging the overall gaps between PR and forensic handwriting examination but also made many forensic like and multi-lingual data sets publicly available. The next chapter describes these competitions.

5 Validation

In order to effectively bridge the gap between PR and forensic examination of signatures and to validate the postulates of this thesis, the author organized various international signature verification competitions (co-joined with the leading PR researchers and forensic scientists). These competitions have not only validated the postulates reported in this thesis, but also provided various forensic like signature and handwriting datasets to the PR community in general. In particular, two major dimensions were considered in these competitions. The first dimension is taking more forensic like data to PR (which, to the best of author's knowledge, were never available to PR)–here the focus was to motivate PR researchers to develop systems that are capable of dealing with disguised signatures (previously largely forgotten in PR) along with the more traditional genuine and forged signatures. On account of this, two competitions, namely ICFHR 4NSigComp2010 (the author organized this competition before starting his PhD–however, the post competition experiments are mainly performed during this PhD), and ICFHR 4NSigComp2012 were organized. The second dimension was to motivate PR researchers to report the forensically relevant likelihood ratios, as discussed in the last chapter. Three competitions, namely ICDAR SigComp2011, ICDAR SigWiComp2013, and ICDAR SigWIcomp2015 are organized. The major findings and how the presented likelihood based evaluation scheme is validated in these competitions is discussed in the next sections.

[0] This chapter is an adapted version of work published in [137, 202, 13, 218, 211].

5.1 Signature Verification Competition 4NSigComp-2010

The ICFHR 4NSigComp2010 is the first ever signature verification competition that contained disguised signatures in the training and the evaluation sets. These data were collected by FHEs and the participating automatic systems were required to report;

- A Probability Value P between 0 and 1.
- A Decision Value D which could be either 1, 2, or 3.

The Probability Value P was compared to a predefined threshold t. A higher value $(P > t)$ indicated that the questioned signature was most likely a genuine one. A lower value $(P < t)$ indicated that the questioned signature was not genuine, meaning that it was not written by the reference author. A probability value of $(P = t)$ was considered as inconclusive. The Decision Value D represents the system's decision about the process by which the questioned signature was most likely generated. A Decision Value of 1 means that the underlying writing is natural: there was no or not enough evidence of any simulation or disguise attempt and the signature was written by the reference author. Decision Value 2 represented that the underlying writing process was unnatural: there was evidence of either a simulation or disguise attempt. Whereas a Decision Value 3 showed that the system was unable to decide if the underlying process was natural or unnatural: no decision could be made whether the signature was genuine, forged or disguised. The output reference table is provided in Table 5.1. It presents the various output possibilities. In this table, a value of P greater than t with output 1 means correct genuine authorship, with output 2, on the other hand, means that the author has made an attempt to disguise her/his identity. If the Decision Value is 3 then with any value of probability it is simply inconclusive. Any value of P less than t with decision value 2 indicates that the questioned signature was a result of a simulation or disguise process. The final assessment of the output values is given in Table 5.2. Two experiments were performed, one where the inconclusive ratings were ignored and another where they were counted as errors. There was no significant difference in the results and especially no change in the ranking of the systems since mostly none of the systems reported inconclusive decisions.

Table 5.1: Interpretation of the output

Decision	Probability		
Value	$P > t$	$P < t$	$P = t$
1	authentic	misleading	inconcl.
2	disguise	simulation	inconcl.
3	inconcl.	inconcl.	inconcl.

Table 5.2: Assessment of the output

True	Probability		
Answer	$P > t$	$P < t$	$P = t$
authentic	correct	incorr.	incorr./ignored
disguise	correct	incorr.	incorr./ignored
simulation	incorr.	correct	incorr./ignored

5.1.1 Data

The collection contained only off-line signature samples. The signatures were collected under supervision of Bryan Found and Doug Rogers. The images were scanned at 600 dpi resolution and cropped at the Netherlands Forensic Institute for the purpose of this competition.

The signature collection for training contained 209 images. These signatures comprised 9 reference signatures by the same writer, say A and 200 questioned signatures. The 200 questioned signatures comprised 76 genuine signatures written by the reference writer in his/her normal signature style; 104 skilled forgeries (written by 27 forgers freehand copying the signature characteristics of the reference writer by practicing for as long as they wished); and 20 disguised signatures written by the reference writer, here A.

The collection for evaluation contained 125 signatures. These comprised 25 reference signatures by the another writer, say B, and 100 questioned signatures. The 100 questioned signatures comprised 3 genuine signatures written by the reference writer in his/her normal signature style; 90 skilled forgeries (written by 34 forgers freehand copying the signature characteristics of the reference writer by practicing for as long as they wished); and 7 disguised signatures written by the reference writer, here B. All writings were made using the same make of ball-point pen and using the same make of paper, however, the color of ink was different in training (black) and evaluation (blue) set.

5.1.2 Submitted Systems

In total, 7 systems participated in the 4NSigComp2010. The following describes each of these systems briefly.

Biometric Recognition Group- ATVS -Univ. Autonoma de Madrid

This system is based on a fusion of two machine experts, one based on local analysis of the image [219] and a second approach based on allographic analysis [220]. The local matcher uses contour level features [219]. It is based on the features proposed for writer identification and verification using images of handwriting documents [220]. It computes the orientation of local contour fragments, as well as its curvature. The contour-direction distribution $f1$ is extracted by considering the orientation φ_1 of local contour fragments and computing its probability distribution. Curvature of the signature contour $f2$ is computed by considering two contour fragments attached at a common end pixel and computing the joint probability distribution of the directions φ_1 and φ_2 between that pixel and both fragments. As the algorithm runs over the contour, the two histograms of $f1$ and $f2$ are built, which are then normalized to a probability distribution. To compute the similarity between two signature images, the x^2 distance is used. This matcher outputs two distances, one for $f1$ and another one for $f2$. The matcher based on allographic analysis considers a signature as a stochastic pattern of handwritten shapes [220]. The probability distribution function (PDF) of these shapes in a given signature image is used to characterize the identity of the writer, which is computed using a common code-book of shapes obtained by means of clustering techniques. The code-book is generated using an external database of handwritten signatures [167]. This way, the code-book provides a common shape space and the PDF captures the individual shape usage preference of the signer. To compute the similarity between two signature images, the x^2 distance is used. Finally, fusion of the two machine experts is performed via linear combination of the individual scores [123]. Linear regression is used to compute the optimal fusion weights.

Université du Littoral Cote d'opale LISIC

To compare two signatures, this system computes a DTW similarity on their projections obtained by Mojette transform. Each image is first pretreated like the following

- Compute the (gray level) luminance matrix from the RGB image.
- Reduce the matrix in a bounding box.

- Compute the inverse video matrix.

Then, all projections of the image according to a 2-rank Farey series extended to $[0, \pi]$ are computed according to the Mojette transform algorithm. This encoding is derived from the Radon transform. Each projection can be interpreted as a spatial histogram of the luminance in a fixed axis. To answer the nature of the signature (naturally written or not), the similarity matrix on the given reference signatures is computed. From this matrix, a luminance threshold vector based on 1-NN algorithm is obtained. If there is a sufficient number of similarities between the questioned signature and the reference signature higher than this threshold, the questioned signature is considered to be naturally handwritten.

NifiSoft

This system computes several features based on the number of connected components, number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures and chain codes. Each feature F_i is computed for the questioned signature $F_i(q)$ and the N reference signatures $F_i(r)$, where $(r = 1, ..., N)$. The average absolute difference between the value of the feature F_i in the questioned signature and its values in the reference signatures is then computed. The obtained differences are combined via a logistic regression classifier trained either on only the 4NSigComp2010 database training set (partial training method) or on 4NSigComp2010 training set and complete SigComp2009 database (full training method).

Sabanci University

After preprocessing and size normalization steps, tessellate the image into a fixed number of zones using polar coordinate representation and extract gradient information in each zone. The extracted features of the query signature are classified using a user-dependent SVM that is trained with the reference signatures of the user and negative examples. A combination classifier is built which does score level combination of the user-dependent SVM classifier described above, with one based on normalized correlation and another similar to the first one, but using a user-independent SVM classifier.

Anonymous Systems I and II

Two anonymous systems participated in this study. These are commercial products and their internal details are unavailable. These systems are termed as “Anonym-I” and “Anonym-II” in the experiments.

Additional Systems

This section provides details about two additional systems which were not available in the 4NSigComp2010 but have been considered in this study. These systems were part of the 4NSigComp2012. They provided the author a possibility to analyze improvements in the results when more data were available (for further details about this, see Section 5.3).

Griffith University

This system employs the Gaussian Grid feature extraction technique [221] developed by the Blumenstein Lab at the School of (ICT) and the Institute for Integrated and Intelligent Systems (IIS), Griffith University in Australia. For completeness some details are included here, for further details refer to [221]. The Gaussian Grid feature extraction technique employs signature contours as its input. The following steps are performed. First, the input signature contour image is divided into $m * n$ zones. Then by tracing the contours in each block the 4-direction chain code histogram of each block is created. Every step from a pixel to its adjacent one of the four directions (horizontal, vertical, left-diagonal, and right-diagonal) are counted. There are four matrices of size $m * n$ for each direction, namely H, V, L, and R. After that Gaussian smoothing filter is applied and the value of each element of each matrix obtained is adjusted by dividing its value by the maximum value of the four matrices. Further from the two-matrix pairs horizontal (H) and vertical (V) matrices, left-diagonal (L) and right-diagonal (R) matrices, two new matrices are established. Eventually the feature vector is formed by merging the six matrices.

Anonymous System III

This system is primarily based on grid features introduced by Samuel et al. [124]. First, the signature image is spatially smoothed followed by binarization via combinations of local and global binarization techniques. After that the signature image is located and centralized via center of gravity and then divided into 64 cells. Then various features

Table 5.3: Summary of the systems performance on 2010 data with and without disguised signatures in the test set. EER: Equal Error Rate (with disguised signatures in the test set), EER*: Equal Error Rate (when disguised signatures removed from the test set)

System	Acc.	FAR	FRR	EER	EER*
AVTS	90.0	1.1	90	80	34
LISIC	54.0	41.1	90	58	41
NifiSoft (partial)	91.0	1.1	80	70	8
NifiSoft (full)	75.0	20.0	70	70	8
Sabanci University	**80.0**	**13.3**	**80**	**55**	28
Anonym-I	92.0	0.0	80	70	**0**
Anonym-II	20.0	87.0	10	60	21
Griffith University	30	70	70	70	4.44
Anonym-III	78	13	78	56	33

are extracted from each cell including, size of cell, it's center point, centroid, angle of inclinations each black pixel makes with the corners of the corresponding cell. After computing these features, thresholds are computed using means and variances. Following that, nearest neighbor approach is applied to decide on the result of each feature vector and finally a voting based classification based on different voting strategies is made. This system is termed as "Anonym-III" in experiments.

5.1.3 Performance Evaluation

The results are computed by generating Receiver Operating Characteristic (ROC) curves, containing the error rates based on varying the decision threshold t. The False Acceptance Rate (FAR–the percentage of wrongly accepted simulations/forgeries) and the False Rejection Rate (FRR–the percentage of signatures by the reference writer which have been wrongly interpreted as simulations/forgeries) are computed and the Equal Error Rates (EER–where FAR equals FRR) of different systems are given in Table 5.3. A crucial observation is that most of the systems could not handle disguised signatures (as given in Table 5.3). There was only one system that performed well on the disguised signatures, but this system showed a large error rate in detecting simulated signatures. A second set of experiments was made where the disguised signatures were excluded. The results to show that these are the disguised signatures which make the state-of-the-art systems fail badly. These results are given in Table 5.3 under EER*.

These results are quite interesting (see Figure 5.1 and Table 5.3) where one system has even reached an EER of 0% when disguised signatures were removed from the dataset.

Figure 5.1: Equal Error Rates on 2010 data (with and without disguised signatures)

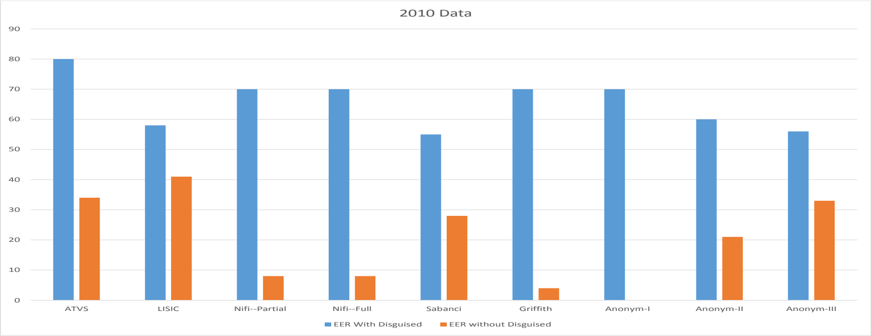


Table 5.4: Overview of the test data 4NsigComp2012

No. of signatures	Author "A1"	Author "A2"	Author "A3"
Reference	20	16	15
Questioned	250	100	100

This indicates the difficulty automatic systems face while classifying disguised signatures.

5.2 Signature Verification Competition 4NSigComp-2012

The ICFHR Competition on Automatic Forensic Signature Verification (4NsigComp2012) also contained previously unpublished skilled forgeries and disguised signatures. This time again the main focus is on automatic classification of disguised signatures along with genuine and forged signatures. A larger and more diverse dataset is used in this competition. Note that the participating systems this time had the possibility to train on the complete data from 4NSigComp2010 and this has helped the systems improve the results.

5.2.1 Data

The data contain only offline signature samples. The signatures were again collected under supervision of Bryan Found and Doug Rogers. The images were scanned at 300

Table 5.5: Breakup of Test Set Questioned Signatures

Type of signatures	Author “A1”	Author “A2”	Author “A3”
Disguised	47	08	09
Forged	160	42	71
Genuine	43	50	20

dpi resolution and cropped at the Netherlands Forensic Institute for the purpose of this competition.

The training set comprised training and test set of the 4NsigComp2010. For the test set signature samples were provided by the Forensic Expertise Profiling Laboratory (FEPL) of La Trobe University. It contained signature samples from three specimen writers/authors, A1, A2, and A3, respectively. The questioned samples were a mixture of genuine signatures, disguised signatures and skilled forgeries as given in tables 5.4 and 5.5 . All signatures were written using the same make of ball-point pen and the same make of paper. The questioned samples were numbered randomly, scanned and ink-jet or laser printed into a booklet.

For specimen author ’A1’, three normal signatures per day (written with a ball point pen) over a fifteen day period, six disguised signatures per day (written with a ball point pen) over a fifteen day period, and six normal signatures per day (written with a pencil) over a three day period were collected. From normal signatures pool the genuine and reference signatures for specimen author “A1” were drawn.

For forging the signatures of specimen author “A1”, two ’forgers’ were selected from the academic staff at La Trobe University. Each of the forgers was provided with six normal samples of the questioned signature written by ’A1’. Forgers were instructed that they could use any or all of the supplied specimen signatures as models for their forgeries. Forgers were also instructed that their forgeries must be unassisted (not tracings). Each forger was asked to complete the following task each day over a ten day period.

- 25 practice signatures (ball point pen)
- 5 forgeries (ball point pen)
- 5 forgeries (pencil)

The forgeries, other than the practice attempts, were used as a pool from which the questioned forged signatures were selected.

Table 5.6: Collective Results (EER: Equal Error Rate with disguised signatures, EER*: Equal Error Rate without disguised signatures)

System	Accuracy(%)	FRR	FAR	EER	EER*
Griffith	**85.11**	**15.82**	**14.29**	**15.82**	**14.16**
Nifi-Full	77.88	23.16	21.61	21.61	16.48
Sabanci	78.89	21.47	20.88	20.88	**13.19**
Anonym-III	71.11	28.81	28.94	28.81	20.35
Anonym-IV	30.67	62.71	73.63	62.71	68.14

Table 5.7: Author "A1" Results (EER: Equal Error Rate with disguised signatures, EER*: Equal Error Rate without disguised signatures)

System	Accuracy(%)	FRR	FAR	EER	EER*
Griffith	**93.60**	**6.67**	**6.25**	**6.25**	**4.38**
Nifi-Full	85.60	14.44	14.37	14.37	11.25
Sabanci	83.60	17.78	15.63	16.63	15.63
Anonym-III	72.00	27.78	28.13	27.78	16.25
Anonym-IV	19.60	80.00	80.63	80	81.4

For specimen authors "A2" and "A3", the normal and disguised signatures were written over a 10 and 15 days period respectively. From normal signatures pool the genuine and reference signatures were drawn for these specimen authors.

For forging the signatures of specimen author "A2", 31 adult forgers contributed. These individuals were volunteers drawn from a single private company. Each of the forgers was provided with three original normal samples of the signature written by the specimen writer. These forgers were instructed similar to the forgers of "A1", mentioned above.

For forging the signatures of "A3", six adult forgers contributed. These individuals were volunteers. Each of the forgers was provided with six original normal samples of the signature written by the specimen writer. These forgers were also instructed as above.

5.2.2 Submitted Systems

In total, five systems from five different institutions participated. Note that four of these systems (namely Griffith University, Nifi–Full, Sabanci University, and Anonym–III) are versions of the systems already discussed in Section 5.1 and those now have had the possibility to train on the 4NSigComp2010 data to handle disguised signatures in particular along with the genuine and forged signatures. Participants were allowed to be anonymous upon request. In the following details about the system "Anonym–IV" are provided. This system remained anonymous, however provided the following working details.

Anonym–IV

Given a scanned image as an input, first binarization, and then normalization with respect to skew, writing width and baseline location are performed. To extract the feature vectors from the normalized images, a sliding window approach is used. The width of the window is varied from one to three pixel and following geometrical features are computed at each window position, the mean pixel gray value, the centroid, vertical and horizontal second order moments, the locations of the uppermost, and lowermost black pixel and their positions and gradients with respect to the neighboring windows, the black to white transitions present within the entire window, the number of black-white transitions between the uppermost and lowermost pixel in an image column, and the proportion of black pixels to the number of pixels between uppermost and lowermost pixels are used. These features are already proposed in [222]. For classification various classifiers were employed and the best result were obtained by Gaussian Mixture Models.

5.2.3 Performance Evaluation

As stated earlier, the basic aim of the 4NsigComp2012 is to gauge the performance and applicability of some of the state-of-the-art signature verification systems in real forensic casework. For this purpose, disguised signatures were in the dataset. Experiments were performed on the complete test set as well as for each reference author in the test set individually.

The experiments were divided into two evaluation categories. The evaluation 1 for all the experiments considered genuine, forged and disguised signatures. The evaluation 2 for all the experiments considered only genuine and forged signatures. Table 5.6 shows the results when all the system were tested on the complete test set including disguised signatures. Here system 1 outperforms all the other systems both on accuracy and EER. When the disguised signatures were removed from the test set, evaluation 2, the system 3 performed better on the accuracy and FRR/FAR scales. Tables 5.7, 5.8, and 5.9 detail systems' results for individual authors. Note that different systems performed better for different authors. The System from Griffith University performed better for author

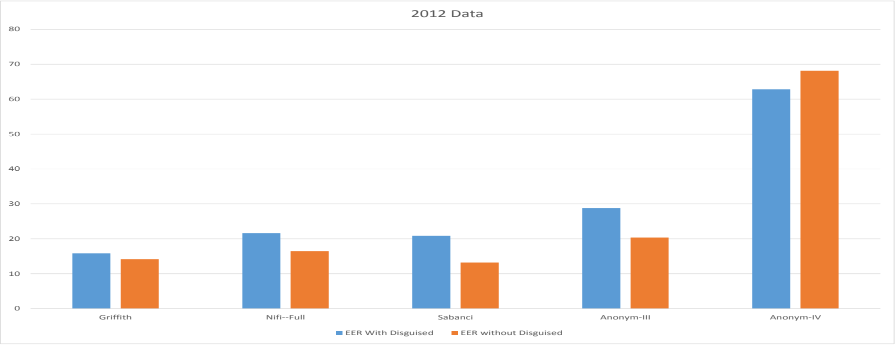


Figure 5.2: Equal Error Rates on 4NSigComp2012 data (with and without disguised signatures)

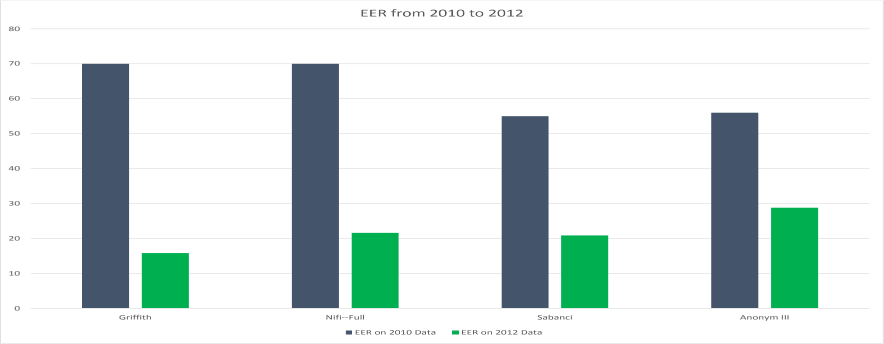


Figure 5.3: Reduction in Equal Error Rates from 2010 to 2012

$A1$, system from Sabanci university performed better for author $A2$, and system from Qatar university (Nifi–Full) performed better for author $A3$. Still on the complete dataset system from the Griffith university was the winner, this can be explained by difference in the sizes of the data from three authors. Author $A1$ has the largest data, which impact the overall results.

The results also show that the grid-based features which actually focus on many small regions of the signature seem to be the most efficient features for this dataset. However, we can also observe that the effectiveness of the features vary from one writer to another. Thus it is hard to draw some general conclusions of these results. Note that the system Anonym-IV also uses features from many small regions, however, instead of using a static grid it applies a more sophisticated method, which finally works worse on this data set.

Table 5.8: Author "A2" Results (EER: Equal Error Rate with disguised signatures, EER*: Equal Error Rate without disguised signatures)

System	Accuracy(%)	FRR	FAR	EER	EER*
Griffith	79.00	20.69	21.43	20.69	20
Nifi-Full	78.00	22.41	21.43	21.43	19.05
Sabanci	**87.00**	**13.79**	**11.90**	**11.90**	**9.52**
Anonym-III	74.00	25.86	26.19	25.86	23.81
Anonym-IV	34.00	67.24	64.29	65.80	64.29

Table 5.9: Author "A3" Results (EER: Equal Error Rate with disguised signatures, EER*: Equal Error Rate without disguised signatures)

System	Accuracy(%)	FRR	FAR	EER	EER*
Griffith	62	37.93	38.03	37.93	19.72
Nifi-Full	**68**	**31.03**	**32.39**	**31.03**	**5.63**
Sabanci	63	37.93	36.62	36.62	19.72
Anonym-III	65	34.48	35.21	34.48	19.72
Anonym-IV	41	58.62	59.15	58.62	74.65

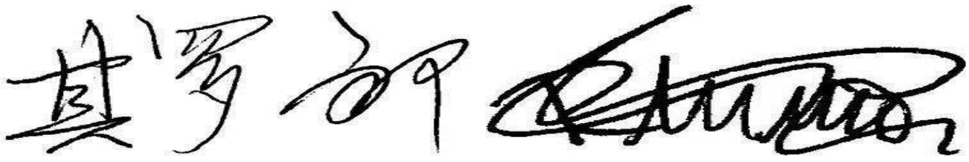
5.3 Discussion: Disguised Signatures

Disguised signatures arguably pose much difficulties for systems when an automatic classification of genuine, forged, and disguised signatures is needed. The above experiments have repeatedly proven this. As shown in Figure 5.1, the systems are producing quite encouraging results when there are no disguised signatures in the test set and in fact one system reaches an EER of 0%. However, in the presence of disguised signatures the same system reports an EER of 70%, and the best system could reach an EER of 55%. If we look the results closely, given in Table 5.3, the FAR – that is affected by the misclassified forgeries– is in fairly acceptable range for most of the systems (except the system Anonym-II which misclassified a lot of forgeries as genuine signatures). This shows that the systems are actually quite good in classifying forged and genuine signatures in the absence on disguised signatures. But the disguised signatures are a difficult genre for the current state-of-the-art systems. Notice that in Table 5.3 the FRR for nearly all the systems is equal or above 70%. This is because all of these systems completely failed in classifying the disguised signatures and every time misclassified the disguised signatures as forgeries. Note that, if an automatic system considers a disguised signature as a forgery, the aim of the person who disguised that signature is achieved (i.e., (s)he personally intended to make her/his signatures look like a forgery and now (s)he has succeeded). Please consider that Table 5.3 depicts the accuracy, FAR, FRR and EER when disguised signatures where in our test set and only the EER* which is achieved when disguised signatures were removed from the test set (actually for brevity the two tables are combined, the same is the case with all the next result tables as well).

The automatic systems on 2012 competition data have also shown mostly similar trends as on 2010 competition data. The system from Griffith university performed better compared to the other participants when disguised signatures were present in the test set and the system from Sabanci university was better than others when disguised signatures were removed. An interesting observation about the systems' behavior is that quite a large difference in systems performance was found for the two cases (disguised included vs. disguised removed) on 2010 data (see Figure 5.1), but comparatively very less performance variance on 2012 data (see Figure 5.2). This is because the participating systems for 2012 data have had the possibility to train on a dataset containing disguised signatures from the 4NSigComp2010. Furthermore, notice from tables 5.3, and 5.6, the EER* (case when disguised removed) has increased for some of the systems (e.g., Griffith, Sabanci, Nifi-Full) from 2010 to 2012. This is because they have now been exclusively trained for disguised signatures and by doing so although they are now able to classify many disguised signatures correctly but at the expense of some forgeries. So this a big area for improvement in the future. Please note that the system Anonym-IV is performing comparatively better in presence of disguised signatures than in their absence. This might have resulted due to over-tuning of this systems parameters to handle disguised signatures.

In addition to the collective results, the same experiments were performed on individual authors data from 2012. Here different systems performed better for different authors. As given in tables 5.4 and 5.5, there are three authors in the 2012 data with varying numbers of genuine, forged, and disguised signatures in the test set. The author "A1" has the maximum signatures available, i.e., 250, and for the other two authors there are 100 questioned signatures each.

Furthermore, as depicted in Figure 5.3, there has been a remarkable decrease in the EER from years 2010 to 2012 on the data containing disguised signatures. The best system on 2010 data was at an EER of almost 55%, which then dropped to almost 16% on 2012 data. This in fact is quite encouraging and gives hints for further improvements in the future.



(a) (b)

Figure 5.4: SigComp2011: Offline signature samples of Chinese (a) and Dutch (b).

5.4 Signature Verification Competition (SigComp2011)

This is the first ever competition where the signature verification community is motivated to enable their systems to compute the likelihood ratios instead of just computing the evidence. This is very important as it allows one to combine the FHE's evidence (from the results of an automated system) with other evidence presented in the courts. Participants were asked to produce a comparison score (e.g., a degree of similarity or difference), and the evidential value of that score, expressed as the ratio of the probabilities of finding that score when the questioned signature is a genuine signature and when it is a forgery (i.e., the likelihood ratio). Note that this competition has introduced a paradigm shift from the "decision" paradigm to an "evidential value" paradigm.

5.4.1 Data

This competition provided signature data in two languages, i.e., Dutch and Chinese. Figure 5.4 shows one sample signature of Chinese and Dutch each. Signatures were either genuine or forged/simulated. For training for both, Dutch and Chinese signatures, data from 10 reference writers were provided to the participants. Each writer contributed with 12 reference signatures and 12 questioned genuine signatures for both modes, offline and online. Furthermore, 36 forgeries were produced by 3 forgers for each reference writer. In addition to that, all the data (training and testing) of the SigComp2009 was also publicly available for training. The data for evaluation were similar. The only difference was that for Chinese just the data of 10 writers were used while for Dutch the data of 54 writers were used, leading to more representative results. More details are presented in Table 5.10.

Table 5.10: Training and Test set of the SigComp2011. Number of Authors (A), number of genuine signatures (G), and number of forged signatures (F).

	Training			Test		
	A	G	F	A	G	F
Chinese Offline	10	235	340	10	236	367
Chinese Online	10	230	430	10	245	461
Dutch Offline	10	240	123	54	1296	638
Dutch Online	10	330	119	54	1296	611

5.4.2 Submitted Systems

In total, thirteen systems were submitted for this competition. System 1 tessellates the signature into a fixed number of zones using polar coordinate representation and extracts gradient information in each zone. The extracted features of the query signature are classified using a user-dependent support vector machine (SVM) that is trained with the reference signatures of the user and negative examples. System 2 preferred remaining anonymous and did not provide details. System 3 includes two main phases: the evidence estimation phase and the calibration phase. For every two signatures, two types of descriptors, a local (at every sampled point based on the gradient in the gray scale image) and a global(in a skeleton image by using shape context) are computed. In the second phase, the FoCal framework [214] is used to calibrate the evidence score. Systems 4 and 5 are commercial products and from the same participant. Systems 6 and 7 use edge-based directional probability distribution features [223] and grapheme features [224]. System 8 is based on the methods introduced by Samuel [124]. System 9 remained anonymous and did not provide any description. Systems 1, 6, 7, and 9 were submitted for on-line and off-line signatures, systems 4 and 5 worked on on-line data only, and systems 3, 4, and 8, worked on off-line data only.

5.4.3 Performance Evaluation

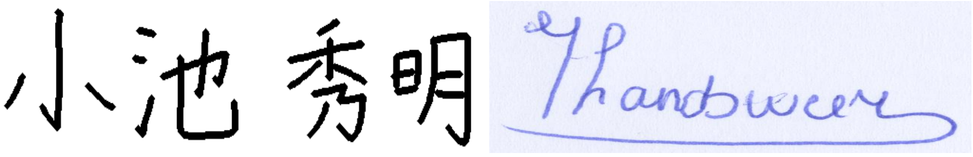
Four different signature verification tasks/scenarios were defined for the said competition, i.e., Chinese offline, Dutch offline, Chinese online, and Dutch online signature verification. The systems were evaluated according to several measurements. First, ROC-curves were generated to see at which point the equal error rate is reached. At this specific point the accuracy was also measured, i.e., the percentage of correct decisions with respect to all questioned signatures. The uniqueness of the said competition and associated evaluation was that the participants were required to report continuous scores that can be converted into LRs or Log LRs (LLRs) by various calibration procedures. The LLR and associated graphs (some examples given in Fig. 5.6) would help FHEs in presenting the results of automated systems in courts. One such conversion was applied using the FoCal toolkit [214] and first measured the cost of the log-likelihood ratios $\widehat{C}_{llr}$ and then the minimal possible value of $\widehat{C}_{llr}$, i.e., $\widehat{C}_{llr}^{min}$, as explained in Chapter 4. Note that a smaller value of $\widehat{C}_{llr}^{min}$ denotes a better performance of the method. The results of the competition are given in Table 5.11.

5.5 Signature Verification & Writer Identification Competitions (SigWiComp2013)

Again in this competition the PR community was motivated to enable their systems compute the likelihood ratios instead of just computing the evidence. The tasks this time were more challenging as more diverse forensic signature data (Japanese and Dutch) were made available to the participants. There were three tasks for signature modality, i.e., for Dutch offline, Japanese offline, and Japanese online signatures. An additional task was also defined in this competition for writer identification but it is out of the scope of the discussion at hand. Here the focus is only on the use of a novel evaluation scheme, based on likelihood ratios, and its application to signature verification where it produces more understandable, informative as well as objective results for both FHEs and PR-researchers.

5.5.1 Data

The Dutch signatures are collected by FHEs at the Netherlands Forensic Institute (NFI). The Japanese signatures are collected by PR-researchers at the Human Interface Laboratory, Mie University Japan. This combination of data collected by experts in the two communities (FHEs and PR) makes this competition quite unique and interesting. The Dutch offline signatures (both training and test) were collected under the supervision of Elisa van den Heuvel and Linda Alewijnse over a multiple year period. The signatures are available in offline format only. A preprinted paper was used with 12 numbered boxes for signing. The preprinted paper was placed underneath the blank writing paper. Four extra blank pages were added underneath the first two pages to obtain a soft writing surface. The signatures were scanned at 400 dpi, RGB color and saved as PNG images. Files were randomly numbered, so that file numbers do not link to the writer or signature



(a) (b)

Figure 5.5: SigComp 2011: offline signature samples of Japanese (a) and Dutch (b).

simulator. The training set comprised of signatures from training and test set of the SigComp 2009, and SigComp 2011. The evaluation set for this task was comprised of 27 authors where each author contributed 10 of her/his genuine signatures. There were 36 skilled forgeries on average available for each genuine author (varying from minimum 25 to maximum 60 forgeries per genuine author. The data breakup is provided in Table 5.12.

The Japanese signature data were collected with HP EliteBook 2730p tablet PC and self-made collection software built with Microsoft INK SDK. The online dataset consists of ascii files with the format: X, Y, Z. The sampling rate and resolution are 200Hz and 50 pixels/cm, respectively. The Z value on each line denoted only pen down (100) and pen up (0). The signature images in the offline dataset were generated from the online signatures. The training sets for both Japanese online and offline signature verification tasks were comprised of 11 authors with 42 genuine signatures of each author and 36 forgeries per author. The test sets for both Japanese online and offline verification tasks were comprised of 20 authors with 42 genuine signatures each and 36 corresponding forgeries per author. The data breakup is provided in Table 5.13.

5.5.2 Submitted Systems

10 systems for off-line and 3 system for on-line signature verification tasks were received. Note that the 10 systems submitted for off-line signature verification tasks were tested separately for Dutch and Japanese data.

System 1 uses signatures' baselines and loops and then generalizes the distribution of these features in all the genuine signatures. A similar process is repeated with forgeries and finally classification is performed using various distribution analysis methods. Details are available in [225]. Systems 2, System 3 and System 4 are from the same participant. The system, in general, uses two complementary features from each signature: Histogram of oriented gradients (HOG) and local binary patterns (LBP). HOG features are extracted in Cartesian and polar coordinates, LBP features are extracted only in

Cartesian coordinates; resulting in 3 kinds of features in total. For classification, the Support Vector Machines (SVMs) are used [226]. For the competition, the participants have further optimized the classifier weights for each dataset (Japanese and Dutch) using the provided data; these form the three submissions, i.e., the original version, and the two optimized versions. System 5 combines through a logistic regression classifier the geometrical features based on number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures, chain codes and edge based directional features. System 6 is based on edge directional based features. The method we starts with conventional edge detection that generates a binary image in which only the edge pixels are "on". Then each edge pixel is considered in the middle of a square neighborhood and checked in all directions emerging from the central pixel and ending on the periphery of the neighborhood for the presence of an entire edge fragment. All the verified instances are counted into a histogram that is finally normalized to a probability distribution which gives the probability of finding in the image an edge fragment oriented at the angle measured from the horizontal. For comparison purposes, the Manhattan Distance Metric is used. System 7 is based on the edge-hinge features which estimate the joint distribution of edge angles in a writer's handwriting. They are constructed by performing an edge detection using a Sobel kernel on the input images, and subsequently, measuring the angles of both edge segments that emanate from each edge pixel. To compare the signatures, the Manhattan Distance Metric is used. System 8 is based on multi-scale run length features [227] which are determined on the binary image taking into consideration both the black pixels corresponding to the ink trace and the white pixels corresponding to the background. The probability distribution of black and white run-lengths has been used. There are four scanning methods: horizontal, vertical, left-diagonal and right-diagonal. The runs lengths features are calculated using the gray level run length matrices and the histogram of run lengths is normalized and interpreted as a probability distribution. The method considers horizontal, vertical, left-diagonal and right-diagonal white run-lengths as well as horizontal, vertical, left-diagonal and right-diagonal black run-lengths extracted from the original image. To compare the signatures, the Manhattan Distance Metric is used. System 9 is based on the combination of both types of features used by the previous two methods: multi-scale edge-hinge features and multi-scale run-length features. Again for this method, the Manhattan Distance Metric is used. System 10 is based on the combination of three types of features used by the systems 6, 7, and 8: multi-scale edge-hinge features, multi-scale run-length features and the edge based directional features. Again for this method, the Manhattan Distance Metric is used to compare signatures. System 11, 12, and 13 perform online verification only and therefore tested for Japanese only. System 11 computes differences between the features of the questioned signature and the reference signatures at the signal level as well as the histogram level [228]. Those features include "x" and "y" coordinates, pressure, directions, angles, speed and angular speed. The System 12 is the Dynamic Time-Warping (DTW) based system described in [229], where the DTW score is slightly adjusted to obtain user-based thresholds and capture some new forgery clues. This adjustment is made by analyzing the reference signatures and the alignment of the query to the references. System 12 is the is similar to system 13, with the exception that pressure is also included in the DTW algorithm.

5.5.3 Performance Evaluation

The signature verification systems are evaluated for the following three tasks.

- Task 1: Dutch off-line signature verification
- Task 2: Japanese off-line signature verification
- Task 3: Japanese on-line signature verification

Again a smaller value of $\widehat{C}_{llr}^{min}$ denotes a better performance of the method. The results of the three defined tasks are given in tables 5.14, 5.15, and 5.16. As shown, different systems performed best on different tasks. The winner of the tasks 1, 2, and 3 are systems 2, 2, and 12 respectively.

5.6 Signature Verification & Writer Identification Competition (SigWIcomp2015)

This competition further strengthened the paradigm shift in automatic signature verification (emphasized by the author of the thesis, in the SigComp 2011, and tested heavily in SigWiComp2013) from the "decision paradigm" to an "evidential value paradigm".

5.6.1 Data

This time the signature data are available in Bengali, Italian, and German, and handwriting data in different writing styles in English only. Only signature tasks are under consideration in this chapter, so writer identification details are not provided here.

Italian Off-line Signatures

The Italian off-line signatures were collected at the University of Salerno from the University employees and students. Two unique aspects of these data are; these are actual signatures that were provided by individuals while filling various forms and applications at the University, and the signatures collection span over a period of time that is between 3 and 5 years depending on the subject. The forgeries are also made by students where they were allowed to practice the forgery as many times as they liked and then they produced skilled forgeries. For training, a set of 50 specimen genuine authors, with 5 reference signatures from each specimen author, was provided to the participants in binarized form. For evaluation, we used the same 50 specimen authors; with 10 corresponding questioned signatures each: containing genuine signatures and skilled forgeries, in binarized form. Detailed breakup is provided in Table 5.17.

Bengali (Bangla) Off-line Signatures

The signatures were collected from different parts of West Bengal, a state of India. The majority of the signatures were contributed by students. A total number of 240 genuine signatures were collected from 10 contributors (24 genuine signatures by each). For each contributor, all genuine specimens were collected in a single day's writing session. In order to produce the forgeries, the imitators were allowed to practice forgeries as long as they wished. A total number of 300 (30 signatures, 10 individuals) forged signatures were collected. The images were captured in 256 level Grey scale at 300 dpi and stored in TIFF format (Tagged Image File Format). A detailed breakup is provided in Table 5.18.

German On-line Signatures

The German on-line signatures were collected at the German Research center for Artificial Intelligence, Germany. A unique aspect of these data is that the data have been collected using a digitized pen rather than a tablet, i.e., by Anoto Pen [230]. This pen specializes in providing the look and feel of regular pens. It only demands to add Anoto dot pattern to any paper and data can be digitized seamlessly. The Anoto pattern makes it possible for the Anoto pen built-in camera to detect strokes and record signatures. The signature data were collected from employees of different financial institutions and students of University of Kaiserslautern who also participated in generation of skilled forgeries. The dataset consists of ASCII files with the format: X, Y, and Pressure with sampling rate of 75 Hz and resolution of 85 dpi. For training, 30 specimen genuine authors were selected, from whom each one provided 10 genuine reference signatures. For evaluation, data from the same 30 specimen genuine authors; with 15 corresponding questioned signatures each: containing genuine signatures and skilled forgeries were used. Detailed breakup is provided in Table 5.19.

Note that for all of the signature verification tasks, no forgeries were provided for training (see Tables 5.17, 5.18, and 5.19). This is important as; in the real world, one can never limit the forgery set, and also when forgeries are used for training–there is always a chance that an automatic system may learn to declare signatures as forgeries when they are coming from the forgers on whom the system is trained [39, 62].

5.6.2 Submitted Systems

40 systems initially registered for this competition, however, 30 systems from 13 different institutions eventually participated for the signature verification tasks. The details follow: 9 systems for off-line Italian signature verification, the same 9 systems optimized for off-line Bengali signature verification, and 12 system for the on-line German signature verification task. Table 5.21 provides details about affiliations of the participants. In the following we will describe each of these systems briefly by providing references so that interested readers may follow. Note that some of the participants preferred remaining anonymous after the results were declared.

Off-line Verification Tasks: Italian and Bengali

System 1 uses histogram of oriented gradients (HOG) and local binary patterns (LBP) extracted from local regions, together with the user-based and global SVM classifiers and a sophisticated classifier combination. The basic system description can be found in [226]. System 2 is based on edge directional based features. The method starts with conventional edge detection that generates a binary image in which only the edge pixels are “on”. Then each edge pixel is considered in the middle of a square neighborhood and checked in all directions emerging from the central pixel and ending on the periphery of the neighborhood for the presence of an entire edge fragment. For comparison purposes, the Manhattan Distance Metric is used. System 3 is based on the edge-hinge features which estimate the joint distribution of edge angles in a writer’s handwriting. They are constructed by performing an edge detection using a Sobel kernel on the input images, and subsequently, measuring the angles of both edge segments that emanate from each edge pixel. System 4 is based on multi-scale run length features which are determined on the binary image taking into consideration both the black pixels corresponding to the ink trace and the white pixels corresponding to the background. For details, refer to [227]. System 5 is based on the combination of both types of features used by the previous two methods: multi-scale edge-hinge features and multi-scale run-length features [231]. System 6 is based on the combination of three types of features used by the systems 2, 3, and 4: multi-scale edge-hinge features, multi-scale run-length features and the edge based directional features. System 7 and System 8 combine through a logistic regression classifier the geometrical features described in [228]. Those features are based on number of holes, moments, projections, distributions, position of barycenter, number of branches in the skeleton, Fourier descriptors, tortuosities, directions, curvatures, chain codes and edge based directional features. The two systems differ in weighting the features they are built on. System 9 a commercial product and the participating organization has requested not to mention their identity. The following basic details are provided about this system. The system has four main modules. The first one identify and extracts the signature from the image page. The second module applies filtering and other preprocessing on the extracted signature. The third module extracts various spatial, geometric, morphological, and statistical features. The last module allows the combination of multiple classifiers and uses DTW, Pearson Correlation, and Euclidean distance combined by an MLP neural network to make the final decision.

The same nine systems (Systems 1-9) have participated for both Italian and Bengali tasks (when optimized for the specific task), their details are therefore omitted.

On-line Verification Task: German

Systems 10 to 14 are from the same academic organization and they have requested not to mention their identity. All of these systems are based on DTW on various time functions derived from the position trajectory and the pressure information. These 5 systems differ in the specific functions considered and the type of score normalization applied. System 15 is based on the DTW algorithm and uses both global features (e.g., total writing time) and local features (e.g., pressure in a given point). If the total writing time of the signature to be verified is acceptable, for each local feature f (such as X-Y coordinates, pressure), the DTW distance between the time series related to the authentic signature and to the signature to be verified is computed. Then, DTW distances are combined with a weighted sum, by giving a weight to each of the different kinds of features. The similarity score is computed accordingly. System 16 computes multiple time series representation of signature including time series of X-coordinate, Y-Coordinate, Pressure, speed and centroid distance. Verification of genuine and forged signature is done using adaptation of multivariate m_Mediods based classification and anomaly detection approach as presented in [232]. System 17 and System 18 are commercial products designed by Cursor Insight. These systems specialize in calculating more than 70000 movement characteristics of each handwritten sample. For further details, refer to Cursor Insight[1]. System 19 compares the signatures using statistical data as well as a time based model. It extracts the changes of several dynamic features (direction, pressure and speed) over time. The statistical data contains static information like duration of the signature or duration and time of pen upliftings. The reference signatures are compared against each other to define the validity space of the model. The similarities of the features are combined into a final score using different weightings. System 20 computes differences between the features of the questioned signature and the reference signatures at the signal level as well as the histogram level [233]. Those features include x and y coordinates, pressure, directions, angles, speed and angular speed. System 21 uses the “x” and “y” coordinates only as features and Dynamic Time Warping (DTW) as the matching algorithm. Normalization of the query scores are done based on reference signature statistics. Details of this system can be found in [229].

5.6.3 Performance Evaluation

In the case of signature verification (both on-line and off-line), the task was to determine if a given questioned signature has been written by the author of the n reference specimen signatures or if it was forged by another writer. Again the evaluation is based on the cost of the log-likelihood ratios $\widehat{C}_{llr}$ using the FoCal toolkit [214], and finally the minimal possible value of $\widehat{C}_{llr}$, i.e., $\widehat{C}_{llr}^{min}$, as the final assessment value. A smaller value of $\widehat{C}_{llr}^{min}$ denotes a better performance of the method.

The results of the first three tasks are given in Tables 5.22, 5.23, and 5.24, respectively. As can be seen, different systems performed better on different tasks. The winner of the Italian off-line signature verification task, Task 1, is system 1 from Sabanci University, Turkey. The winner of the Bengali off-line signature verification task, Task 2, is system 4 from Tebessa University, Algeria. The winner of the German on-line signature verification task, Task 3, is system 17 from Cursor Insight (a company).

[1] www.cursorinsight.com

5.7 Discussion: Evaluation with Likelihood Ratios

The results of the last three international signature verification competitions, SigComp-2011, SigWiComp2013, and SigWIcomp2015, succinctly validate the importance of the presented evaluation scheme based on likelihood ratios. Note that the notion of likelihood and evaluation on the basis of Bayesian inference has been there since quite long [185, 234]. The FHEs already follow this somehow, though subjectively (please refer to [14]), and its usage in automatic applications has been studied by Brmmer et al. [216] and Gonzalez et al. [19]. Yet the PR, community in general was not very well aware of these issues. This dissertation makes the evaluation based on likelihood ratios (in accordance with the Bayesian inference) a current and more relevant topic by thoroughly validating its suitability for both PR (handwriting and signature analysis) and FHEs at the same time. The signature verification competitions organized through the course of this dissertation have not only created a general awareness in the signature verification community about the idea of reporting likelihood ratios, but have also quite successfully bridged the gaps between the two communities, PR and FHEs, at various levels as discussed in Chapter 4. Various interesting observations can be made when having a closer look at results presented in the last three sections. First, the presented evaluation metric not only considers the number of errors made by a system, but also looks into the severity of errors. The more severe an error is, the more cost will be associated with it. Therefore, a good EER does not always result in a good $\widehat{C}_{llr}^{min}$. For example, consider Table 5.11. System 9(b) performs quite well on the on-line Chinese data when looking at the EER, but has the worst $\widehat{C}_{llr}^{min}$. This might be explained by the fact that even a few large misleading answers spoil the overall performance with $\widehat{C}_{llr}^{min}$. Note that it reflects the practice, i.e., a system should not produce a high likelihood for a wrong decision, as this might result in wrong judgment with severe outcomes. Observations, similar to SigComp 2011, can be made when having a closer look at the evaluation results of SigWiComp2013. It is again obvious from the results that the system with the best FRR and FAR always turned out to have the best value of $\widehat{C}_{llr}^{min}$ but a good EER does not always result in a good $\widehat{C}_{llr}^{min}$, e.g., System 3 (see Table 5.14) performs better for the Task 1 when looking at the EER as compared to that of System 4 on the same data, but the $\widehat{C}_{llr}^{min}$ of System 4 is better than that of System 3. Again, a few large errors might spoil the overall performance with $\widehat{C}_{llr}^{min}$. Similar is the case with the Systems 8 and 9 on the same task. For further reference, the *llr* distributions of the genuine signatures (tar) and forgeries (nontar) for some systems on the Italian off-line dataset are provided in Figure 5.6 (in reference to

Table 5.22). Note that these curves are substantial for forensic experts. The overall performance of the systems can be judged by these curves. The more they overlay each other, the more is the cost of calibration attached to that system and eventually it is not a good system. In summary, these curves provide an overview of the system performance for FHEs. The farther the tar-curve goes to the left, the higher would be the cost of this misleading decision. Similarly, the farther the nontar-cuve goes to the right, the higher would be the cost. Note that both curves for the first system in Fig. 5.6 (system 9) are cluttered at the mean. It represents a high cost of llr. Ideally these curves should be well separated to achieve the minimum cost of llr, as the system with minimum value of $\widehat{C}_{llr}$ wins. The second system in Fig. 5.6 (system 4) shows a better performance in terms of $\widehat{C}_{llr}$. The best system on this database is system 1, with the curves shown at the bottom of Fig. 5.6. Here the curves for the two cases are well separated causing a small value of $\widehat{C}_{llr}$ for this system. This effect is most evident from the results of the most recent signature verification competition, SigWIcomp2015. Consider Table 5.23, there are at least three systems with the same EER of 1.67 each. Here if the evaluation was only considering the number of correct and incorrect classifications, based only on EER, there would be no clear winner. However, when looking into the likelihood ratios, one system was a clear winner. Figure 5.7 shows the complete CLLR curves for these systems for reference. Here it can be observed that though the said three systems made equal number of errors, but the distribution of errors for the three systems (on different desired sides of the graph, left or right) was different. This resulted in the different $\widehat{C}_{llr}^{min}$ values for the three systems. Since the $\widehat{C}_{llr}$ and $\widehat{C}_{llr}^{min}$ look both into the number of errors made by a system and also into the severity of errors by warping the scores of automatic systems. This makes the metric ($\widehat{C}_{llr}^{min}$) well suited for forensic applications where a severely mistaken system (although may be having a very low error rate) may lead a person to death, and thus considering the severity of errors is substantial for forensic casework.

Furthermore, the said evaluation metric is tested on many automatic systems, for a large amount of data available in multiple languages, e.g., Chinese, Dutch, English, Japanese, Italian, Bengali, and German. Every time the said evaluation metric reported consistent results. It shows the strength of this evaluation metric. In addition to this, the number of participants from 2011 to 2015 (through the course of this PhD) have increased greatly. In 2011, there were 13 participants, in 2013–there were 23, and most recently in 2015–there were 30 participants for the signature verification tasks (where the said evaluation metric was used), apart from others who are also using this evaluation metric after these competitions. This shows that the work done during the course of this thesis has gained acceptance in the community and more and more PR researchers are moving to enable there systems also report likelihood ratios as required by FHEs.

Some curves obtained by the evaluations performed on off-line signature verification systems from SigWIcomp2015 (Italian and Bengali) are provide in Appendix B for interested readers.

Table 5.11: SigComp2011: evaluation results of the four defined tasks.

Task 1: Chinese offline verification:

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1(a)	80.04	21.01	19.62	0.76	0.69
2	73.10	27.50	26.70	3.06	0.77
3	72.90	27.50	26.98	1.13	0.79
6(a)	56.06	45.00	43.60	1.26	0.89
7(a)	51.95	50.00	47.41	3.22	0.95
8	62.01	37.50	38.15	1.57	0.93
9(a)	61.81	38.33	38.15	6.23	0.92

Task 2: Dutch offline verification:

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1(a)	82.91	17.93	16.41	0.73	0.57
2	77.99	22.22	21.75	2.46	0.67
3	87.80	12.35	12.05	0.42	0.39
6(a)	95.57	4.48	4.38	0.71	0.13
7(a)	97.67	2.47	2.19	0.90	0.08
8	75.84	23.77	24.57	1.66	0.72
9(a)	71.02	29.17	28.79	4.13	0.79

Task 3: Chinese online verification:

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1(b)	84.81	12.00	16.05	0.57	0.35
4	93.17	6.40	6.94	0.41	0.22
5	93.17	6.40	6.94	0.42	0.22
6(b)	82.94	16.80	17.14	1.05	0.50
7(b)	85.32	13.60	14.97	0.91	0.46
9(b)	80.89	9.26	8.14	6.21	0.73

Task 4: Dutch online verification:

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1(b)	93.49	7.56	7.69	0.50	0.24
4	96.27	3.70	3.76	0.26	0.12
5	96.35	3.86	3.44	0.35	0.12
6(b)	91.82	8.33	8.02	0.53	0.29
7(b)	92.93	7.25	6.87	0.60	0.24
9(b)	88.56	11.11	11.27	6.43	0.40

Table 5.12: Dutch signatures. A: authors, G: genuine signatures, F: forged signatures, A*= SigComp2009 data, B*= SigComp2011 data

Mode	TrainingA*			TrainingB*			Test		
	A	G	F	A	G	F	A	G	F
Offline	12	60	1860	54	1296	648	27	270	972

Table 5.13: Japanese signatures. A: authors, G: genuine signatures, F: forged signatures

Mode	Training			Test		
	A	G	F	A	G	F
Offline	11	462	396	20	840	720
Online	11	462	396	20	840	720

Table 5.14: Results for Task 1: Dutch Offline Signature Verification

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1	67.90	31.85	32.14	2.005070	0.863449
2	**76.83**	**23.70**	**23.10**	**0.880048**	**0.642632**
3	75.56	24.44	24.44	1.086197	0.706733
4	74.93	25.19	25.05	0.979237	0.698044
5	73.76	26.67	26.18	3.941330	0.732123
6	72.14	27.41	27.93	0.966780	0.739129
7	70.87	29.63	29.06	1.103572	0.776319
8	69.16	31.11	30.80	1.021834	0.742239
9	70.15	29.63	29.88	1.053304	0.744848
10	72.95	27.41	27.00	1.023746	0.728198

Table 5.15: Results for Task 2: Japanese Offline Signature Verification

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1	72.70	27.23	27.36	4.692036	0.778976
2	**90.72**	**9.74**	**9.72**	**0.796040**	**0.339265**
3	89.82	10.23	10.14	0.814598	0.349146
4	86.95	13.04	13.06	0.831630	0.400977
5	66.67	33.33	33.33	0.940937	0.833182
6	76.70	23.60	23.06	0.980174	0.665241
7	73.98	26.07	25.97	1.021412	0.720444
8	68.33	31.35	31.94	1.006464	0.758588
9	72.10	27.89	27.92	1.013250	0.717454
10	74.59	25.41	25.42	1.002516	0.694036

Table 5.16: Results for Task 3: Japanese Online Signature Verification

ID	Accuracy(%)	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
11	70.55	29.56	30.22	1.176164	0.884223
12	**72.55**	**27.56**	**27.36**	**1.089183**	**0.744544**
13	72.47	27.56	27.50	1.114015	0.744793

Table 5.17: Italian Off-line signature data

Data	Authors	Genuine	Forged	Total
Training	50	250	0	250
Testing	50	229	249	478

Table 5.18: Bengali Off-line signature data

Data	Authors	Genuine	Forged	Total
Training	10	120	0	120
Testing	10	120	300	420

Table 5.19: German On-line signature data

Data	Authors	Genuine	Forged	Total
Training	30	300	0	300
Testing	30	150	300	450

Table 5.20: English Off-line handwritten text data

Data	Authors	Pages/Author	Total
Training	55	3	165
Testing	55	3	165

Table 5.21: Overview of the submitted systems

System	Modality	Participant	Mode
1	Signatures	Sabanci University, Turkey	Off-line
2-6	Signatures	Tebessa University, Algeria	Off-line
7,8	Signatures	Qatar University, Qatar	Off-line
9	Signatures	Commercial System	Off-line
10-14	Signatures	Anonymous	On-line
15	Signatures	Anonymous	On-line
16	Signatures	Bahria University, Pakistan	On-line
17-18	Signatures	Cursor Insight	On-line
19	Signatures	Commercial System	On-line
20	Signatures	Qatar University, Qatar	On-line
21	Signatures	Sabanci University, Turkey	On-line

Table 5.22: Results for Task 1: Italian Off-line Signature Verification

System	Participant	Accuracy	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1	**Sabanci University**	**99.16**	**0.87**	**0.80**	**0.655109**	**0.021358**
2	Tebessa University	63.60	36.68	36.14	0.993189	0.893270
3	Tebessa University	58.37	41.48	41.77	1.065696	0.952499
4	Tebessa University	65.06	34.93	34.94	1.074474	0.880930
5	Tebessa University	63.81	36.24	36.14	1.065475	0.901003
6	Tebessa University	65.48	34.50	34.54	1.041895	0.901003
7	Qatar University	56.69	43.23	43.37	8.901864	0.972708
8	Qatar University	53.77	46.29	46.18	13.111064	0.960163
9	Commercial System	46.44	53.28	53.82	1.003786	0.988845

Table 5.23: Results for Task 2: Bengali Off-line Signature Verification

System	Participant	Accuracy	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
1	Sabanci University	98.33	**1.67**	**1.67**	0.68895	0.052485
2	Tebessa University	95.00	5.00	5.00	0.933750	0.154390
3	Tebessa University	95.48	4.17	4.67	0.939850	0.116541
4	**Tebessa University**	**98.33**	**1.67**	**1.67**	**0.923886**	**0.039721**
5	Tebessa University	98.10	1.67	2.00	0.931253	0.060248
6	Tebessa University	98.33	**1.67**	**1.67**	0.929922	0.055556
7	Qatar University	55.70	44.23	44.37	1.161387	0.973043
8	Qatar University	91.43	8.33	8.67	2.839290	0.297842
9	Commercial System	45.71	51.67	55.33	0.998630	0.893893

Table 5.24: Results for Task 3: German On-line Signature Verification

System	Participant	Accuracy	FRR	FAR	$\widehat{C}_{llr}$	$\widehat{C}_{llr}^{min}$
10	Anonymous	78.00	22.00	22.00	0.948878	0.680014
11	Anonymous	76.67	23.33	23.33	0.944402	0.688105
12	Anonymous	78.89	21.33	21.00	0.855595	0.576762
13	Anonymous	79.33	20.67	20.67	0.874655	0.577098
14	Anonymous	78.00	22.00	22.00	1.262207	0.684602
15	Anonymous	58.89	41.33	41.00	0.996586	0.839468
16	Bahria University	70.00	30.00	30.00	1.025046	0.772073
17	**Cursor Insight**	**90.27**	**9.87**	**9.67**	**1.243891**	**0.290158**
18	Cursor Insight	83.41	16.45	16.67	1.496257	0.540563
19	Commercial System	85.33	14.67	14.67	0.807276	0.465681
20	Qatar University	56.67	43.33	43.33	3.345016	0.854712
21	Sabanci University	89.56	10.67	10.33	0.873383	0.304159

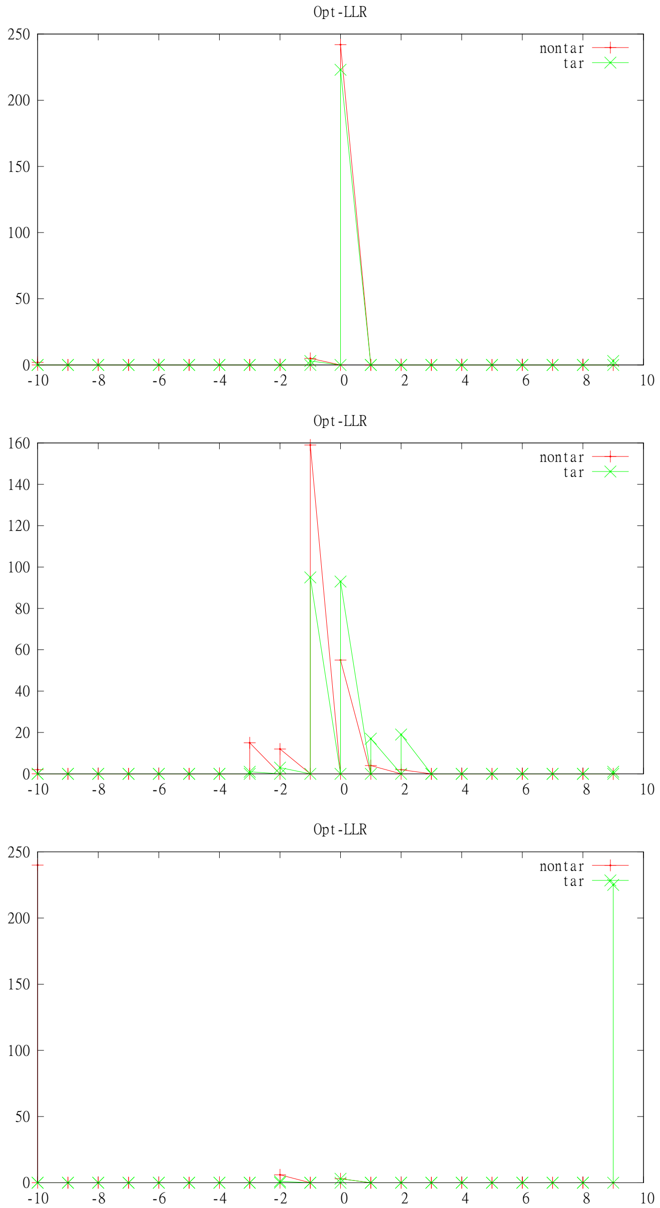
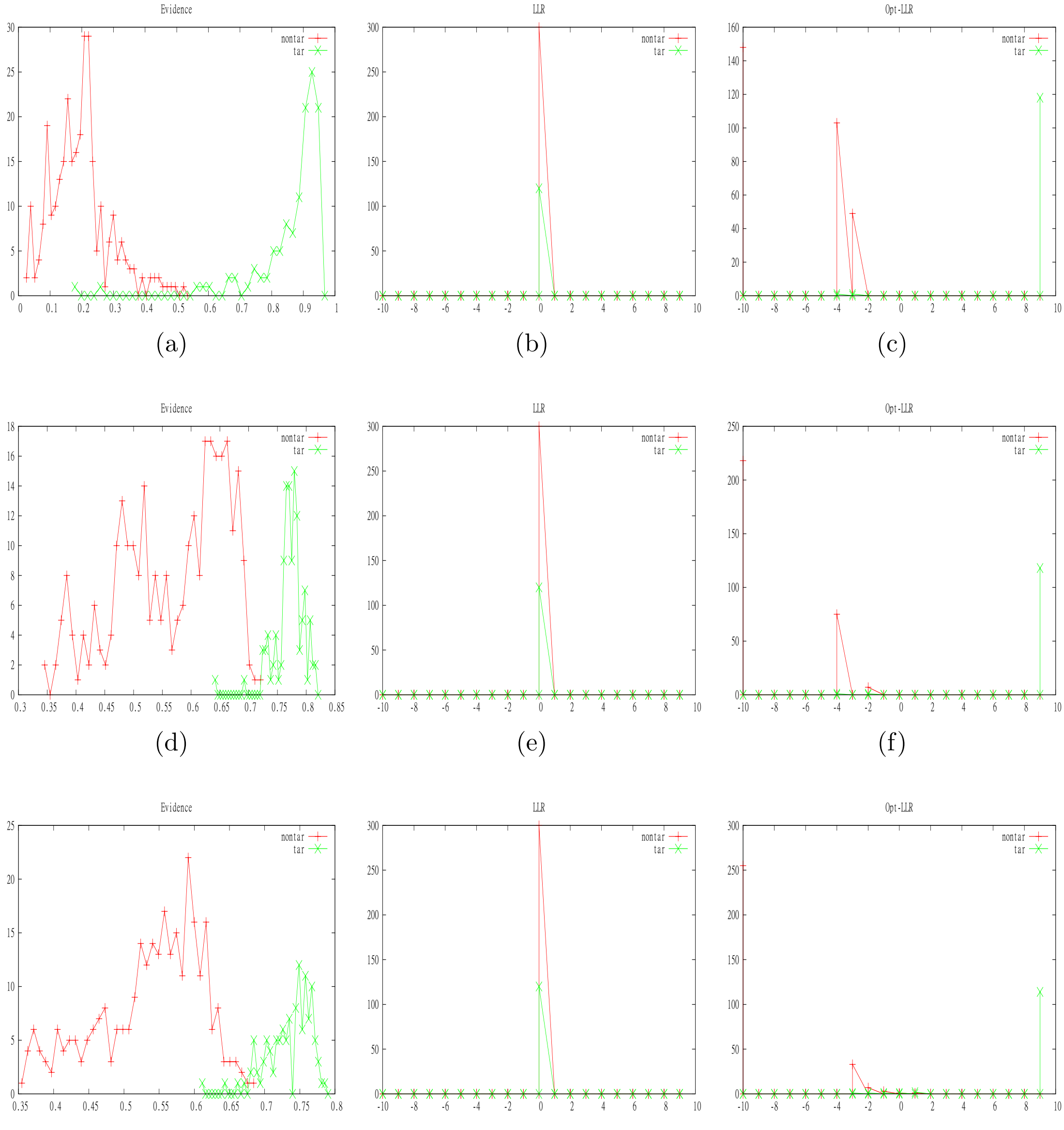


Figure 5.6: SigComp2015: Optimized/Calibrated LLR curves for a not so good system (top), a better system (middle), and the best system (bottom) on Italian off-line data.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 5.7: The evidence, likelihood, and optimized/calibrated likelihood curves of the three systems having same EER but different value for $\widehat{C}_{llr}^{min}$ on Bengali off-line data from SigWIcomp2015. (a-c) System ID 1, (d-f) System ID 4 (winner), (g-i) System ID 6.

6 Novel Local Features based Systems for Forensic Signature Verification

This chapter presents the local features/part-based systems proposed through the course of this thesis in order to perform signature verification particularly considering the implications of forensic casework. Previous literature, e.g., the works of Richiardi et al. [63] and Schlapbach et al. [24], instigated the initial idea of looking into local features and analyzing their suitability for signature verification with respect to forensic casework where a prime aspect is that the FHEs also have to look into the possibility of disguised behavior (please refer to Chapter 3). Furthermore, in-line with the contemporary research, some initial analysis carried out at the start of this thesis suggested the suitability of local features for forensic examination, e.g., [41, 62]. A key aspect is that usually in forensic cases, expert forgers carry out forgeries where the overall holistic appearance of forged signatures is extremely close to the respective genuine/specimen signatures. However, at minute details these forgeries differ from the specimen. Since local features are very good at capturing these finer details, they seem to possess good potential for solving such signature verification cases. Furthermore, various well-known local feature detectors and descriptors, e.g., SIFT [128], SURF [21], have over the years shown very good results for tasks, like character recognition, signature identification, and etc. Recently, a local feature based method, Scale Invariant Feature Transform (SIFT), has been introduced to the domain of handwriting [129] and is applied to writer retrieval and identification [60]. Some improvements in the basic SIFT descriptor have also been suggested for character recognition [130, 131]. Similarly another local feature based method, Speeded Up Robust Features (SURF), has been used for object and character recognition [56, 132, 21]. In

[0] This chapter is an adapted version of work published in [235, 57, 58, 59].

addition to this, some local keypoint detectors, e.g., Features from Accelerated Segment Test (FAST), have been used for problems like multiple object tracking [133], object recognition for smart phone platforms [134], and recognition of degraded handwritten characters [135].

Inspired from these, this thesis presents novel systems based on local feature approaches, SURF, FAST, and FREAK. Note that to the best of the author's knowledge, the systems presented here are the very first in this area that are reported in the literature, particularly for forensic signature verification cases involving disguised signature classification. Considering the importance of this work, the current chapter is divided into various sections where first the local feature approaches used in this thesis are summarized and then the novel systems based on these approaches are detailed.

6.1 Speeded Up Robust Features (SURF)

Speeded Up Robust Features (SURF) is a part-based/local feature approach which is primarily developed for object detection and recognition. Originally SURF was developed for gray scale images, however, now there are many variants of SURF [236, 237], which are also applicable to color images. The principle on which SURF is built is similar to another very popular local keypoint detector and descriptor, i.e., Scale Invariant Feature Transform (SIFT) [128], however, SURF is several times faster than SIFT [21]. In addition, it is more robust and provides invariance against many image transformations [21].

For extracting local/part-based features, it is first important to detect important areas (the so called "keypoints") in an image, which are significantly different from their neighbors. This difference can be based on change in gradients, edges, texture, etc. Once these locations are identified, the next step is to apply some measures/features on the area/patch to describe the information available in the area/patch. This step is referred to as keypoint description.

SURF provides a robust method for both detection and description of keypoints in an image. The speeded nature of SURF is due to the use of a special image representation called integral images [238]. SURF uses integral images to detect blob like structures in an image/rectangular grid/patch. An integral image is a sum of the values between the point under consideration and the origin. Integral images are already used by many researcher to improve the efficiency of different algorithms e.g., efficient adaptive thresholding [239], comparison of color images [240], etc. The use of integral images reduces the calculation of an area of rectangular region to only four operations, regardless of the size of area.

Therefore, all of the convolutions with different filters, sums, can be done very efficiently, and this is the main fuel which speeds up the SURF.

6.1.1 Keypoint detection

In principle, SURF keypoint detector is based on Hessian matrix, which is a matrix containing partial derivatives of a function. Considering a 2D image I as a function f of two variable (x, y). The Hessian matrix is given as

$$H(I(x,y)) = \begin{bmatrix} \frac{\partial^2 i}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{6.1}$$

Based on the above mentioned equation, the determinant of the Hessian, which is also called as discriminant, is given by

$$det(H) = \frac{\partial^2 I}{\partial x^2}\frac{\partial^2 I}{\partial y^2} - \left(\frac{\partial^2 I}{\partial x \partial y}\right)^2 \tag{6.2}$$

In addition to the use of integral images, another speed-up is achieved in SURF by using determinant of Hessian for both location and scale determination. The Hessian matrix with scale and space can be rewritten as follows.

$$H(X,\sigma) = \begin{bmatrix} Dxx(X,\sigma) & Dxy(X,\sigma) \\ Dxy(X,\sigma) & Dyy(X,\sigma) \end{bmatrix} \tag{6.3}$$

In equation 6.3 X is a point (x, y) in an image and σ is the scale. $Dxx(x, \sigma)$ is the partial derivative of image which is calculated using image convolution with scale normalized approximated Gaussian box filters. These box filters can be constructed for Gaussian derivatives in x, y and combined xy direction and can be computed very fast using the integral images. An approximated 9×9 weighted box filter in x, y and combined xy direction with the scale $\sigma = 1.2$ (which represents the highest spatial resolution and minimum scale) is shown in Figure 6.1. The determinant of the Hessian (equation 6.2) is considered as a blob response at location (x, y, σ). A keypoint can be found in different scales, as it is possible that the same image appears at different point of time in different resolutions. It is therefore important to have a keypoint detection mechanism which is scale invariant. This is achieved by using scale space representation, where the keypoints are looked for both in scale and space. The scale-space is divided into a number of response maps layers, called octaves, where each layer covers the double of scale. The

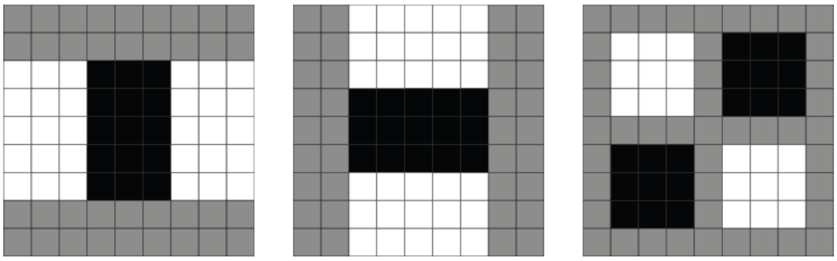


Figure 6.1: Weighted Box filter approximations in the x, y and xy-directions

9×9 box filter shown in Figure 6.1 corresponds to a real valued Gaussian filter of $\sigma = 1.2$. To obtain new layer the filter is up-scaled while maintaining the ratio. To filter out unnecessary keypoints, the responses on all of the detected layers are thresholded. This means that all those keypoints who have values less than some predefined threshold are removed. This threshold is also referred to as the Hessian threshold. On the remaining points, non-maximal suppression is preformed to find the set of probable candidate points. Finally, localization of keypoints in both scale and space is achieved by interpolation in the neighboring data by fitting 3D quadratic [241].

6.1.2 Descriptor

Once a keypoint is detected in an image, the next step is to perform measurements on that area and describe the content in the area using these measures. This step is also referred to as keypoint descriptor extraction phase. SURF uses location information and distribution of gradient in the region of keypoint to describe the content in the region.

While describing the content of a keypoint, it is important to find the reproducible orientation assessment, so that the descriptor extracted for the given patch is rotation invariant. Orientation assignment is done using the Haar wavelet response in x and y directions in a circular neighborhood of radius $6s$, where s is the scale at which the keypoint was detected. The responses are again computed using integral images, which results into a speeded computation. These responses are then represented as a vector, and the dominant orientation is estimated by summing up all the responses within a sliding orientation window covering an angle of $\pi/3$. A variant of SURF, called Up-right SURF (U-SURF), does not perform this step, which results into a more speeded but non-rotation invariant version of SURF. Once the orientation of the keypoint is determined, the next step is to extract the descriptor itself, which is done as follows:

1. Construct a square window of size $20s$ centered around the keypoint, oriented along

the estimated orientation.

2. Split the window in 4 × 4 sub regions.

3. Compute wavelet response in x and y direction for each sub region, referred to as dx and dy.

4. The sum of wavelet responses in each sub region is stored as first component of the descriptor ($\sum dx$, $\sum dy$). To store the information about the polarity of the intensity change, sum of absolute values of response is also stored, i.e., $\sum \mathsf{dx}$, $\sum \mathsf{dy}$.

This results into a 4 dimensional descriptor for each sub region, which in-turn results into a 64 dimensional descriptor for a keypoint.

6.2 Features from Accelerated Segment Test (FAST)

Keypoint detection is considered as an important step for many computer vision and image processing tasks, e.g., object tracking, image matching, object recognition, and so on. FAST [4] is an efficient method for identification of keypoints in an image. FAST keypoint detection uses machine learning to detect keypoints, which are stable, and computationally less expensive than machine of the keypoint detection methods like, Difference of Gaussian [128], SUSAN [242], Harris [243], etc. It is based on segment test criteria, which works by analyzing 16 pixels around the keypoint candidate pixel p. These points are labeled clockwise by a number from 1 to 16 (see Figure 6.2).

A following check is performed on the candidate pixel so that to decide whether or not to classify a point as a keypoint.

1. Set a threshold intensity value T.

2. If a set of N neighboring pixels in a circle have intensity greater than the candidate pixel intensity plus threshold intensity,
   OR If a set of N neighboring pixels in a circle have intensity less than the candidate pixel intensity minus threshold intensity, the candidate pixel is referred to as a keypoint.

The above process will return many keypoints adjacent to one another. To overcome this problem, non-maximal suppression is performed. However, to perform non-maximal suppression, it is necessary to first compute a score for each keypoint. The score is

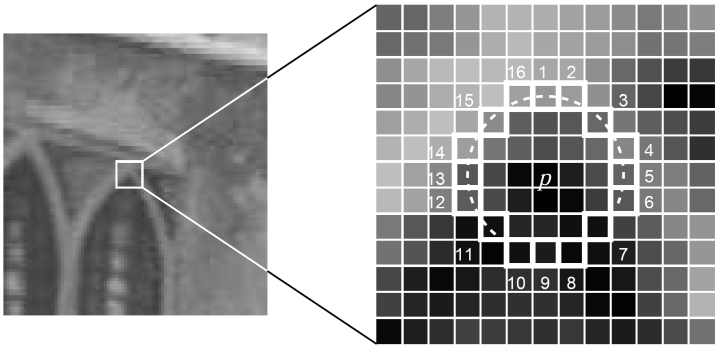


Figure 6.2: An example of segment test [4]

computed (using equation 6.4) by the sum of absolute difference between the pixels in the neighboring arcs and the center pixel. Adjacent keypoints are compared based on this score and the one with lower score is removed.

$$V = max \begin{cases} \sum(pixelvalues - p) & (value - p) > t \\ \sum(p - pixelvalues) & (p - value) > t \end{cases} \tag{6.4}$$

6.3 Fast Retina Keypoint (FREAK)

Another recently introduced part-based/local feature is Fast Retina Keypoint feature. Unlike SURF, which provides both keypoint detection and description, FREAK [5] focuses only on the keypoint descriptor extraction. FREAK [5] is binary local keypoint descriptor, which is inspired from the human visual perception. Particularly, it is inspired from retina, due to which it is referred to as Retina keypoint, whereas fast comes from the speed achieved by the binary nature of the descriptor. The main part, which is inspired form human visual system is the sampling grid used to compare pairs of pixel intensities. For instance we can use random pairs (as in BRIEF [244] and ORB [245]) or circular pattern with equally spaced circles concentric (as in BRISK [246]). FREAK uses retinal sampling grid (which is also circular) where density of points near the center are higher than the farther points [5]. The inspiration behind this sampling comes from retina where higher resolution is captured in the fovea (area near the center of retina), whereas a lower resolution is captured in the perifoveal (outermost area of retina). To realize this inspiration into FREAK, kernels of different sizes are used for every sample point [5]. The size of the Gaussian kernels is changed with respect to the log-polar retinal pattern. In addition, it includes redundancy that brings more discriminative power. Figure 6.3 shows the sampling pattern of FREAK.

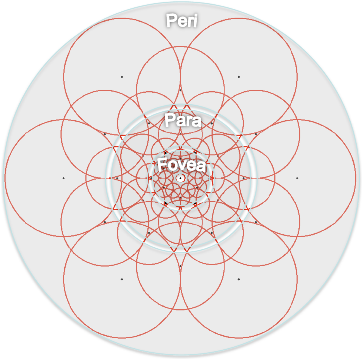


Figure 6.3: Sampling pattern used in FREAK [5]

A binary descriptor is constructed by thresholding the difference between selected pairs and corresponding Gaussian kernel(see equation 6.5).

$$F = \sum_{0 \le a < N} 2^a T(P_a) \tag{6.5}$$

Where F is a descriptor, P_a is a pair and N is the desired size of the descriptor.

$$T(P_a) = \begin{cases} 1 & if(I(P_a^{r1}) - I(P_a^{r2})) > 0 \\ 0 & otherwise \end{cases} \tag{6.6}$$

$I(P_a^{r1})$ is smoothed intensity of first field of the pair P_a and $I(P_a^{r2})$ is smoothed intensity of second field of the pair P_a.

An important step while creating a FREAK descriptor is the selection of pairs which are not highly correlated and discriminant. This is achieved by coarse-to-fine ordering of the difference of Gaussian. Furthermore, saccadic search is used to parse and construct the descriptor in several steps. It starts by searching with the first 16 bytes of the FREAK descriptor (which is actually the coarse information) and if the distance is smaller than a threshold, the comparison is continued to the next bytes to analyze finer information. This process even accelerates the final matching step. With this process, more than 90% of the candidates are discarded at the first parsing step. The orientation estimation is performed using the sum of local gradients over selected pairs. In total 45 pairs are used to estimate the orientation.

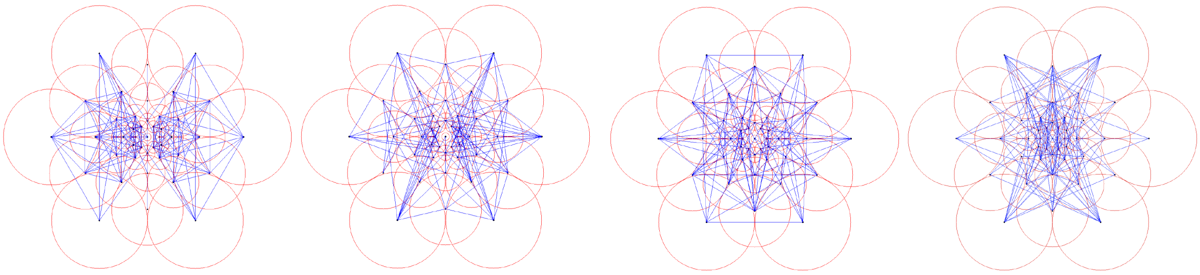


Figure 6.4: Coarse-to-fine sampling used in FREAK [5]

6.4 Signature Verification with Local Features: A Proof of Concept

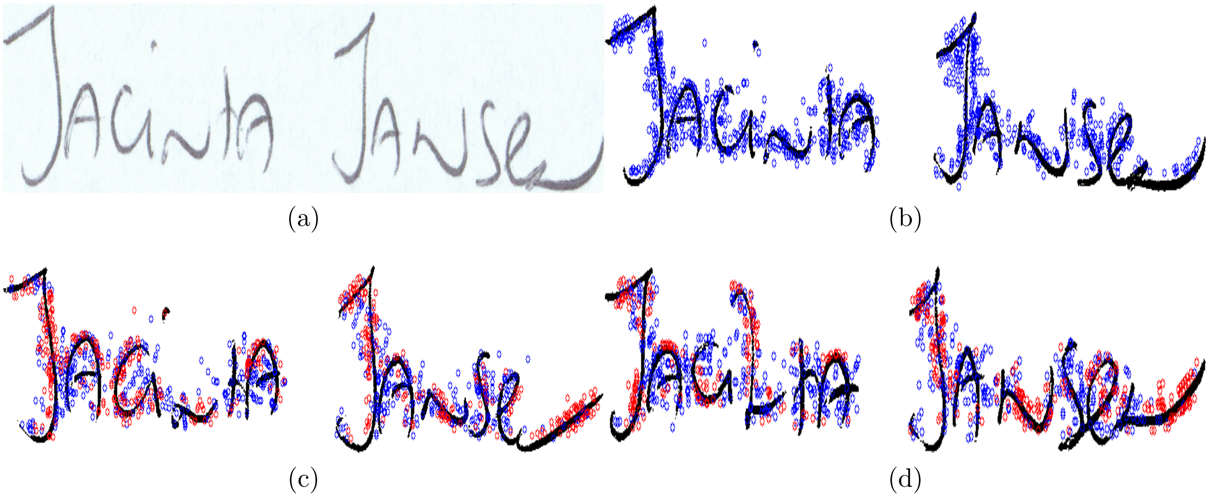
Local features, e.g., Speeded Up Robust Features (SURF), are already well known for performing various classification and recognition tasks. For details about SURF, please refer to Section 6.1. This section details the finding of initial experiments that were carried out in order to investigate the suitability of local features, in particular SURF, for signature verification. In this work, SURF were used along with the K-Nearest Neighbor approach for classifying disguise, forged, and genuine signature. For extraction of features, a Hessian threshold of 400 was used, i.e., all the keypoints having Hessian value less than 400 were neglected. This filtering is done to neglect unimportant features from the signatures and the Hessian threshold was empirically calculated. The SURF was applied to both the reference signatures and questioned signatures using the following methodology.

1. Pick the first key-point (having Hessian value greater than or equal to 400) from the reference signature R. Lets call this keypoint as R_1.

2. Compare R_1 with Q_1 (the first keypoint of the questioned signature Q). If the distance between R_1 and Q_1 is less than an empirically found threshold, θ, increase the positive count P_1 for R_1 by one score point.

3. Compare R_1 with Q_2 (the second keypoint of the questioned signature). Again, if the distance is less than θ, increase the positive count for R_1, i.e., P_1 by one further score point. Repeat this process until reaching Q_n, i.e., the last keypoint of the questioned signature Q.

4. Count the total number of reference signature keypoints, from R_1 to R_n, having positive count greater than 0. Calculate average by considering the total number of



(a) (b)

(c) (d)

Figure 6.5: Some example signatures. (a): A genuine reference signature, (b): Keypoints extracted from the genuine reference signature, (c): Keypoints extracted from a questioned forged signature, (d): Keypoints extracted from a questioned disguised signature.

reference keypoints (present in the R) and the reference keypoints with a non-zero positive score count. This represents the average number of reference signature keypoints that are present in the questioned signature. Based on this number and an empirically found threshold θ_t (which may vary for different authors), perform classification.

Figure 6.5(a) shows an example reference (genuine) signature and Figure 6.5(b) shows the corresponding SURF-keypoints extracted from this reference (genuine) signature. Similarly Figure 6.5(c) shows a questioned (forged) signature and Figure 6.5(d) shows a questioned (disguised) signature, respectively, with SURF-keypoints extracted. Here blue dots represent the original questioned keypoints and red dots represent the keypoints which are at a distance less than or equal to θ, when compared to reference signature keypoints.

The data from the 4NSigComp2010 signature verification competition data was used for experiments. This is the first ever publicly available dataset containing disguised signatures. For details about the dataset, please refer to Chapter 5, Section 5.1. The same experimental protocol was used as in the 4NSigComp2010. For training, 25 genuine reference signatures were available: however, only one of these 25 genuine reference signatures was used at a time to train the system. Note that, no forgery was used for training. This is a realistic forensic scenario [137] and is posed in the 4NSigComp2010 competition.

Table 6.1: Results of the experiments performed

Ref Signature	Experiment	Accuracy(%)	FAR(%)	FRR(%)
Best Case	1	93	7	0
	2	93	7.0	0
Worst Case	1	74	22.22	33.66
	2	72	26.66	40

6.4.1 Results

FHEs are generally interested in classifying genuine, disguised, and forged signatures at the same time. Accordingly, the following two sets of experiments are reported.

- Experiment 1: where the focus was on the classification of forged and genuine signatures only.
- Experiment 2: where the focus was on the complete classification of genuine, forged, and disguised signatures.

For comparing the reference signature's keypoints with the questioned signature keypoints, the Euclidean distance between the descriptors of the two was computed. In the first set of experiments, a single threshold θ_t was used for classification, where signatures with an average number of matching keypoints less than or equal to θ_t were considered as forged. While the signatures with average number of matching keypoints greater than θ_t were classified as genuine. For the second set of experiments, it was observed that the said three types of signatures were clearly separable by two decision boundaries. The proposed system classified all the disguised signatures at a large distance from genuine signatures while the forged signatures were in between the disguised and genuine signatures. Though this is quite contrary to the general thinking about disguised signatures according to which disguised signatures might be closer to genuine signatures as compared to forged signatures since they are written by the same authentic specimen author.

The findings of these experiment however refer towards another phenomenon, i.e., the *intention*, studied in detail by Michel [247]. Actually when a writer wants to disguise his/her signatures, the aim is to make the signatures as far as possible from his/her original signatures but without doing much to the subjective similarities. To do this, he/she usually writes some characters, parts, etc., differently while the other parts of the signature are kept more or less the similar. On the other hand the intention of a skilled forger is to go as close to the genuine signature as possible, therefore, skillfully forged signatures are much closer to the genuine signatures. In the second set of experiments the values of all the genuine signatures were clearly higher than all the forgeries and the disguised signatures showed even smaller values than the forgeries: yet the three were clearly separable with only a few outliers.

Note that in this proof of concept, only one authentic reference signature was used to train the local features based system at one time. This has both a high side and a low side. The high side being in forensic environments the number of available reference signatures is generally quite low. This makes this system quite suitable for application in the real world forensic casework. However, note that there are some issue related to the availability of very few training samples. First, a very low number of specimen/training samples is also not acceptable to FHEs. For FHEs it is nearly impossible and unacceptable to cast their observation on the basis of a single signature. A single signature can give no hint about intra-writer variability and very limited hint about writer specific characteristics along with a whole compendium of other missing attributes, due to which an FHE can not say anything about authorship even with a minimal probability. However, this proof of concept gave quite a hope about using local features for signature verification as it appeared that on some signatures the results were extremely good. Only when trained on some specific signatures the results turned worst. The best and the worst case results of these experiments are provided in Table 6.1. The experiments, following the above mentioned methodology, were repeated for each specimen signature separately and results were computed for all the questioned signatures. The large diversity between the best and worst case results (produced by two different specimen signatures) also reveal that local feature can also be used for initial scrutiny of specimen signatures so that to accept only the most consistent training samples while registering a user with the system. This is an important task that the author would pursue in the future.

After getting the above mentioned initial proof of concept, a generic methodology has been designed to use the various local feature based approaches for performing signature verification. The next section describes this methodology.

6.5 Signature Verification: Generic Local Features based Methodology

The proposed methodology for signature verification is based on part-based/local features. To perform part-based analysis, it is first required to extract keypoints from the signature images. The regions around these keypoints are then described using different descriptors. Note that different feature detectors and descriptors can be used individually or in a combined fashion to detect local keypoints and describe them. Hence, in the proposed methodology, FAST [4] keypoint detector is used to detect keypoints in signature images. FAST keypoint detector is computationally efficient in comparison to well known keypoint detection methods, e.g., SIFT [128], Harris [248], and SURF [21] (results are also provided when SURF keypoint detector is used for finding the potential local areas of interest). In addition, FAST gives a strong response on edges, which makes it suitable for the task of signature verification. For further details about FAST, please refer to Section 6.2. Once the keypoints are detected, descriptor for each of the keypoints is computed using a part-based descriptor, FREAK [5]. FREAK is a binary keypoint descriptor which is computationally very efficient. For further details about FREAK, please refer to Section 6.3. As the descriptors extracted using FREAK are binary, therefore Hamming distance is used for comparison of descriptors of query and reference signatures. The use of Hamming distance in-turn makes it computationally more efficient as it can be computed using a simple XOR operation on bit level.

To categories a signature as genuine, forged, or disguised, first it is binarized using the well known global binarization method OTSU [249]. The OTSU binarization is chosen since the data had fairly high resolution signature images and OTSU is also computationally efficient. After binarization, the following procedure is followed.

1. Apply the FAST or SURF keypoint detector on all the reference signatures, separately, to get the local areas of interest (keypoints) from these signatures.

2. Then, get the descriptors of all of these keypoints present in all reference images using the FREAK keypoint descriptor, which describes each keypoint with a 64 bit descriptor.

3. All of these keypoints and their associated descriptors describing local information are added into a database. The database thus contains features for all of the keypoints which are collected from all reference signature images.

4. Once the features database is created, keypoints and descriptors are extracted for the query/questioned signature.

5. Now a comparison is made between the query signature keypoints and the keypoints present in the features database for each corresponding author.

6. The same process of detecting local area of interest using FAST and then descriptors by FREAK is applied to the query image.

7. Take the first keypoint of the query Image and compare it with all the features present in the features database, one by one. If a query signature keypoint is at a distance less than θ, from any feature present in the features database, note the keypoint.

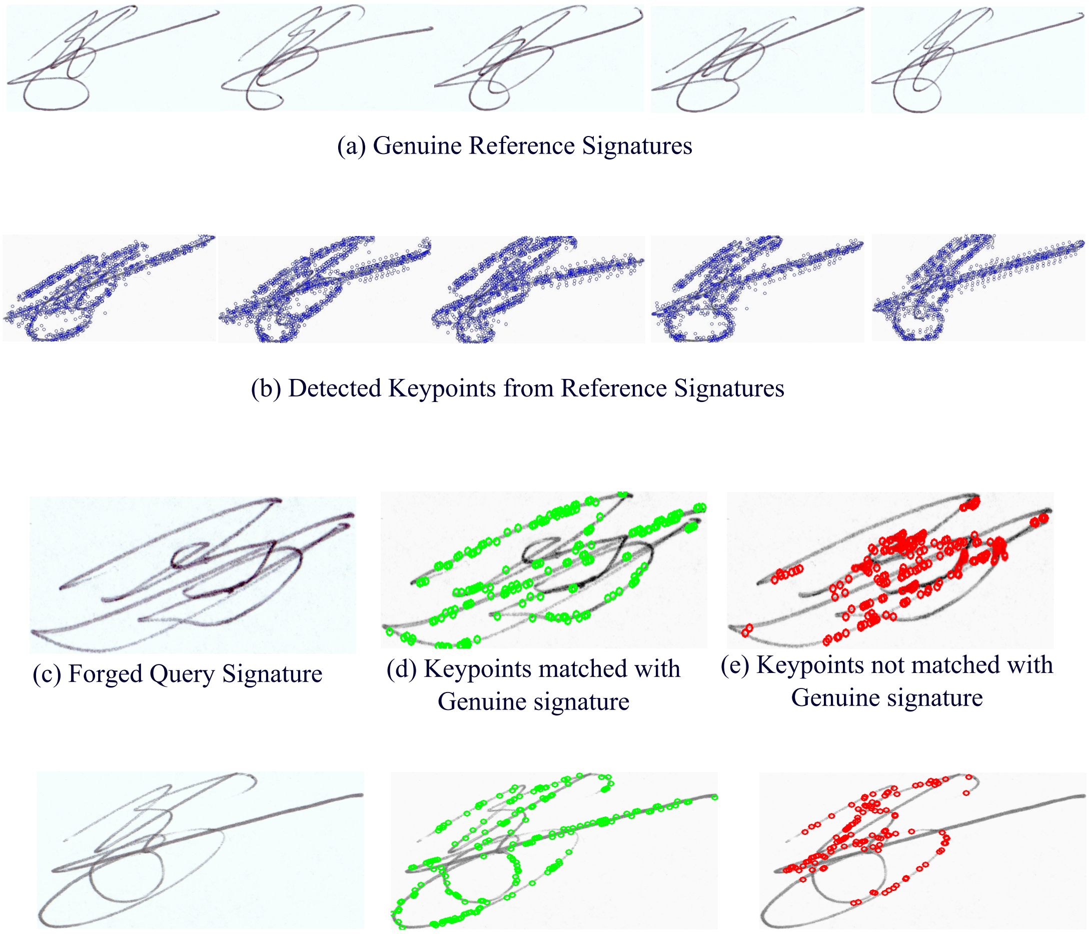
8. Keep this process going until all the query signature's keypoints are traversed.

9. Calculate the average by considering the total number of query keypoints and the query keypoints matched with the features database. This represents the average local features of the questioned signature that are present in the database of that author.

10. Now, if this average is greater than the threshold θ, (meaning, most of the questioned signature local features are matched with reference local features), the questioned signature is be classified as belonging to the authentic author, otherwise if this average is less than the threshold θ, (meaning, there are only a few query keypoints for whom any match is found), it does not belong to the authentic author.

The same procedure is followed when using the SURF or the FAST keypoint detector. Classification in both the cases is based on the local descriptions provided by FREAK features.

Figure 6.6 provides some example signatures where the above mentioned methodology is applied. Figure 6.6 (a) shows some sample genuine signatures and the related Figure 6.6 (b) shows the keypoints extracted from the reference signatures using FAST feature detector. Figure 6.6 (c) shows a questioned signature that in fact is a forgery attempt. Figure 6.6 (d) and (e) provide green and red keypoints on this forgery attempt. The green keypoints are those which belong to the query signature but they also matched with the keypoints present in the features database. The red keypoints are those which belong to the query signature but they did not match with the features database. Similarly,



(a) Genuine Reference Signatures

(b) Detected Keypoints from Reference Signatures

(c) Forged Query Signature

(d) Keypoints matched with Genuine signature

(e) Keypoints not matched with Genuine signature

(f) Disguised Query Signature

(g) Keypoints matched with Genuine signature

(h) Keypoints not matched with Genuine signature

Figure 6.6: Steps of Signature verification. Green dots: keypoints of the query/questioned signature that matched with the keypoints of the reference signatures. Red dots: keypoints of the query/questioned signature that did not match with the keypoints of the reference signatures.

Figure 6.6 (f), (g), and (h) are obtained when the query signature was a disguise attempt. From the Figure 6.6 (c through e), one can see that for lines which were more or less straight in nature there was a high match of query signature with the reference signature, even if the query signature was a forgery attempt. However, for curved lines/strokes (which are usually very specific of particular individuals) a large number of keypoints mismatched. For disguised query signature, Figure 6.6 (f through h), larger number of keypoints matched with the reference signatures. This is because the disguise signature belonged to the original/authentic author who left some traces of her/his identity while trying to disguise. This can be attributed to the brain motor activity which is always

Table 6.2: Summary of the comparisons performed between the proposed systems, SURF-FREAK, FAST-FREAK and the participants of 4NSigComp2010.

System (sec.)	FAR	FRR	EER	Time (sec.)
1	1.1	90	80	312
2	41.1	90	58	1944
3	20.0	70	70	85
4	0.0	80	70	19
5	13.3	80	55	45
6	87.0	10	60	730
7	1.1	80	70	65
(SURF-FREAK)	30	30	30	12
(**FAST-FREAK**)	**30**	**30**	**30**	**0.6**

utilized whether a person is trying to disguise or not.

6.6 Evaluation

Two experiments were performed for evaluating the efficiency and performance of the proposed systems.

- Experiment 1 that focused the classification of disguised, forged, and genuine signatures using FAST Kepypoint detector and FREAK features.
- Experiment 2 that focused the same classification using SURF Kepypoint detector and FREAK features.

The results of these experiments are provided in Table 6.2. As shown in the table, both of the newly proposed systems, i.e., SURF-FREAK and FAST-FREAK outperform all the participants of the 4NSigComp2010 signature verification competition. The best system from the competition could achieve an EER of 55% in the presence of disguised signatures in the test set. Both of the proposed systems, however, achieve an EER of 30% which is remarkable when compared to the other systems.

Furthermore, Table 6.2 also presents the performance comparison of the said systems on the basis of time. The time is given in seconds and is actually the average time taken by any algorithm to report its result on the authenticity of one questioned signature. For reporting this result, the system has to process the questioned as well as 25 reference signatures. Both of the proposed systems again outperformed all the participants. Specially the FAST-FREAK method is extremely time efficient. It succeeds from the other eight methods by a times of (520, 3240, 141, 31, 75, 1216, 108, and 20, respectively). All tests were performed at a machine with the following specifications.

- Processor: Intel Dual Core 1.73 GHz
- Memory: 1GB
- OS: WinXP Professional

The proposed methodology is generic in nature as different combinations of local feature detectors and descriptors can be used. Two systems are presented which follow the same underlying methodology, one based on SURF detector with FREAK descriptors, and the other based on FAST detector with FREAK descriptors. Other combinations of detectors and descriptors could also be applied likewise. The results indicated that the proposed methodology fairly suits local feature based approaches and beats the state-of-the-art by a large margin, both in term of time and error rate. Furthermore, a general drawback of most of the local features based approaches is the enormous amount of time they take to compute results. By using the proposed methodology, the presented systems were extremely efficient compared to other systems who were mostly relying on global features [137]. This shows that, if utilized properly, local feature approaches show the potential of improving both performance and efficiency of classification.

7
Stability Analysis for Signature Verification

Stability of signatures, both on- and off-line, has been studied in order to improve the results of signature verification [138, 139, 90]. Sacha et al. [140] emphasized that signature samples from some individuals result in significantly worse performance than other samples from the same individuals. Thus, selection of the most stable specimen samples is important. In the recent past, some static signature stability analysis techniques have been proposed that are later used for performing signature verification. For example, Impedovo et al. [250] use the regions on the upper and lower contours of the specimen signatures assuming that the upper and lower contours of a signature usually are crucial for performing verification. Similarly, Pirlo et al. [78] use an equi-mass segmentation approach to non-uniformly split signatures into a standard number of regions. Successively, a multiple matching technique is adopted to estimate stability of each region, based on cosine similarity. A crucial observation about the state-of-the-art signature analysis systems incorporating stability of signatures is that they do not consider disguised signatures (see Chapter 3) and the verification problem is usually solved by classifying signatures into two classes, i.e., either as genuine or forged. This classification is helpful in many fields, e.g., banking, but in some areas, e.g., forensic handwriting/signature analysis, disguised signatures (written by specimen authors but with the intention to disassociate the authorship) are important [137].

The purpose of writing this chapter is two-fold. First, it presents a novel analysis of stability of signatures on the basis of signature's local / part-based features. The Speeded Up Local features (SURF) are used for local analysis which give various clues about the potential areas from whom the features should be exclusively considered while performing signature verification. As such, this chapter is an initial attempt to take the

[0] This chapter is an adapted version of work published in [136, 40].

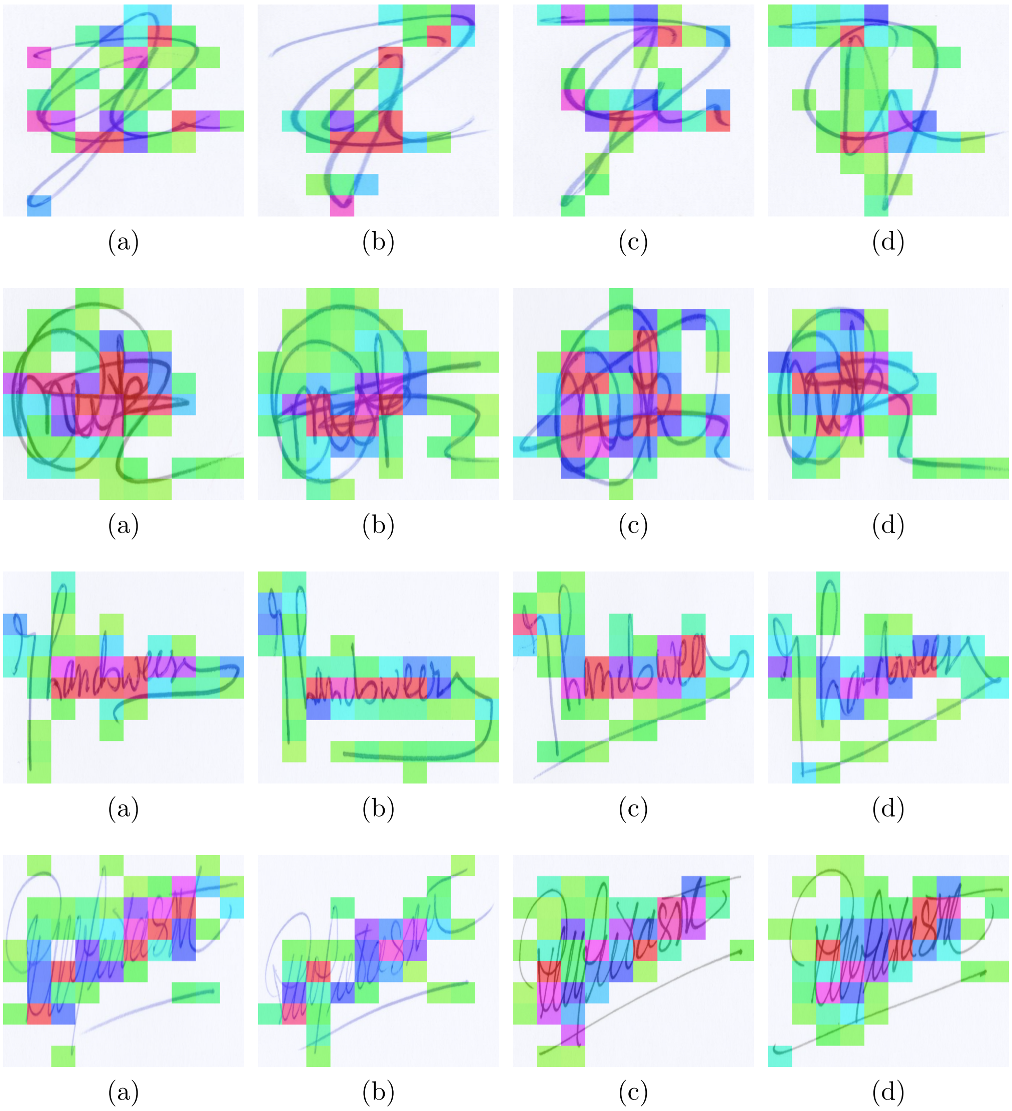
analysis of local stability of signatures to the much realistic domain of forensic signature verification. Second, based on the results of the local stability analysis, a novel signature verification system is proposed to classify genuine, forged, and disguised signatures. The proposed system achieved an equal error rate of 15%, which is considerably very low when compared against the state-of-the-art.

7.1 Local Stability Analysis

Humans generally show the so called intra-writer or within-writer variations when they write signatures. It is a common observation that if a person writes her/his signatures a few times, even by keeping the writing positions, paper, pen, postures, and etc., similar; the signatures vary to a lesser or a greater extent. The analysis of these variations, or conversely the stability of signatures, is very important and has been previously studied heavily [75, 250, 78, 83]. The analysis of stability of signatures can provide us with various insights into the actual signing processes. In general, the stability analysis is performed on global level (by looking at the overall shape changes in a signature) or on part-based/local level (by looking at the signatures' finer details specific to portions a signature).

This chapter details the signature stability analysis performed in this thesis, and then combines the findings to design a complete signature verification system. The first important decision about performing such an analysis is to whether perform it at global, or local, or both the levels simultaneously. The author opted to perform signature stability analysis at local level to explicitly consider disguised signatures where it is a common practice of signers, while disguising their signatures, to keep the entire signature similar to their original signatures except adding or removing a tiny portion or certain strokes. It was thus a natural choice as local features are least affected by such changes [41, 62]. In particular SURF is used for performing local analysis of signature stability. For details about SURF, please refer to Chapter 6, Section 6.1. For the extraction of SURF features a Hessian threshold of 1000 was used, i.e., all the keypoints having a Hessian value less than 1000 were neglected. The following hypotheses framed the foundation of this local features based signature stability analysis.

- H_1: The stability is not homogeneously distributed across the signature. In other words, keypoints from some areas will give more stable results than those from the other areas.



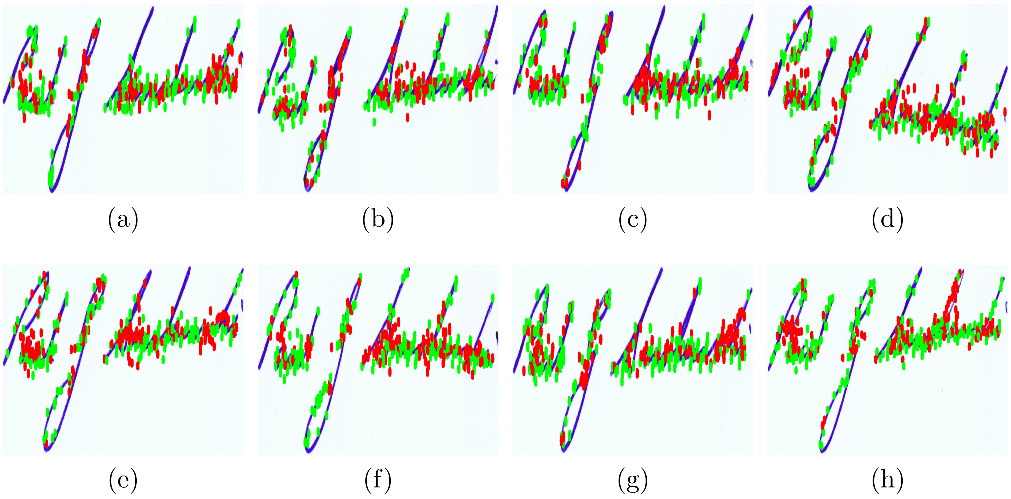
(a) (b) (c) (d)

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

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

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

Figure 7.1: Heat maps of some example specimen (genuine) signatures from four different authors (one genuine author in each row) showing the most stable (green) and the most unstable (red) parts along with the moderately stable parts (colors varying from green to red through blue).



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

(e) (f) (g) (h)

Figure 7.2: Example genuine reference signatures of one author. Green points are considered to be stable and are added to the reference keypoints database for performing verification. Red points are considered unstable and are not included in the reference keypoints database.

- H_2: The stability behavior is generalizable to other authors.

To verify these hypotheses, several experiments were performed on the genuine reference signatures of various authors w.r.t. their stability. The following procedure was followed.

First of all, for every author available in the dataset, one genuine signature was chosen at random (let it be “A”) from the specimen set (set of genuine reference signatures) of that author. SURF keypoints were extracted from the remaining genuine signatures available in the specimen set of that author. This made the reference signature database for comparing the distances of keypoints of signature “A” with that of its database. SURF keypoints were extracted from the signature “A” and these distances are compared with the concerned keypoints database. The Euclidean distance of every keypoint was calculated from the database and was assigned to a matrix. The process was repeated in an “n-1 cross validation” manner and eventually returned the distances of each keypoint of each genuine reference signature from the concerned reference keypoints dataset (when this signature was removed from the specimen data). After normalizing the values, the sum of these values was taken as the actual color of a bin in the histogram (each bin contained the sum of individual keypoint distances from the reference authors dataset for that author).

Figure 7.1 shows some example heat maps along with the superimposed original genuine signatures from four different specimen authors. It highlights the areas of signatures which are most stable and which are most unstable for each of the four genuine signatures of the specimen authors. Note that for each genuine signature, different areas can be stable or unstable. The goal is to identify regions, whose keypoints can confidently be used for classification. The stability of a region is defined by the proximity of the keypoints to a reference keypoints dataset. The white colored bins show the absence of any keypoints in that particular bin while the colors varying from green through blue to red indicate the bins having keypoints. The green color shows the stable regions, i.e., the regions having the keypoints which are at a minimum distance from the reference keypoints dataset. The red color shows the portions of the examined signature which are at the maximum distance from the reference keypoint database. If a part of a query signature has more similar parts in the reference (genuine) signature database, the heat map value at that part becomes closer to green. By inspecting the genuine specimen signatures of various authors, it is revealed that:

- The first hypothesis, H_1, can only be partially proved. The intersection parts for most of the genuine signatures are often unstable (so, shape around the intersection is unstable). However, this is not generalizable to all the intersection cases.
- The second hypothesis, H_2, is also verified partially where we found that the ascending and descending parts are often stable (even though they show a keen curve). Further analysis is required to solidify these findings.

Note that this heat map representation was directly realized because the presented analysis is based on a part-based method. If not, an elaborated non-linear registration method was needed to evaluate this local stability. The basic idea for utilizing the results of this analysis for the classification task is thus to use the keypoints in the regions which are known to have higher stability in general observation, e.g., a classification excluding most of the keypoints in the middle areas of reference genuine signatures.

7.2 Signature Verification

The following procedure is followed to perform signature verification.

1. Compute the keypoints from all the genuine specimen signatures of an author except only one genuine reference signature and make a temporary keypoints database.

2. Compute the keypoints from the remaining genuine reference signature and compare the distances of all of its keypoints from the temporary keypoints database.

3. Find the average distance and then mark all keypoints having distance less than or equal to the average distance as green and all other keypoints red. Figure 7.2 shows these examples for eight different genuine specimen signatures of an author. The final reference keypoints database will contain only the green (stable) keypoints.

4. Repeat this process for every genuine specimen signature in an "n-1 cross validation manner" and populate a final reference keypoints database using only the stable keypoints from different reference/specimen signatures.

5. Once the final reference keypoints database is created, keypoints and descriptors are extracted for the query/questioned signature. A comparison is made between the query signature keypoints and the keypoints present in our final reference keypoints database for that particular author.

6. Find the local keypoints from the query signature by using SURF. Then take the first keypoint of the query Image and compare it with all the features present in the final reference keypoints database, one by one. If a query signature keypoint is at a distance less than an empirically found threshold θ, note the keypoint. Keep this process going until all the query signature's keypoints are traversed.

7. Finally, calculate the probability of each query signature being genuine by considering the total number of query keypoints and the query keypoints matched with the final reference keypoints database. This represents the average local features of the questioned signature that are present in final reference keypoints database of that author.

Figure 7.2 shows that for curved lines/strokes (intersections) usually a large number of keypoints mismatched (depicted with many red keypoints in these areas). Hence, the initial findings point that genuine specimen authors themselves, in most of the cases, do not write very stable intersecting strokes and, therefore, these strokes can be neglected while performing verification.

Table 7.1: Summary of the comparisons performed.

System	Features	FAR	FRR	EER
1 [219]	Contour features	1.1	90	80
2 [137]	Different global statistics	41.1	90	58
3 [137]	Local and global combination	20.0	70	70
4 [137]	Gradient features	0.0	80	70
5	Unknown-commercial product	13.3	80	55
6	Unknown-commercial product	87.0	10	60
7 [137]	Local and global combination	1.1	80	70
8 [57]	SURF-FREAK	30	30	30
9 [57]	FAST-FREAK	30	30	30
10 [62]	Local sliding window	20	20	20
Proposed	**Locally stable SURF**	**15**	**10**	**15**

7.3 Evaluation

Data from 4NSigComp2010 signature verification competition were used for evaluation, please refer to Chapter 5: Section 5.1 for details about the data. Various tests were performed and a comparison of the proposed local stability analysis based signature verification system against ten state-of-the-art systems is provided. Seven systems are the participants of the 4NSigComp2010 signature verification competition while the remaining three are the later reported systems on the same data

As shown in Table 7.1, the proposed system outperforms all the participants of the 4NSigComp2010 signature verification competition as well as all the other systems. The best system from the competition could achieve an EER of 55% in presence of disguised signatures in the test set, while the best later reported system achieved an EER of 20%. The newly proposed system however achieves an EER of 15% which is remarkable when compared to the other systems.

A potential reason for the better performance of the proposed system can be that the proposed system exclusively considers the information available in the local areas of genuine reference signatures and while making a model for an author removes all the unstable features already. This sort of feature selection provides a good opportunity for the proposed system to base the final output strictly on the features which are very stable for an author.

8

Man vs. Machine: A Comparative Analysis for Forensic Signature Verification

Signatures are used as a seal of authenticity in our everyday life. There has always been a demand to authenticate this seal, particularly in cases where a signature became disputed. A large body of literature has developed which describes the theories, methods and techniques used to examine and evaluate the authenticity of questioned signatures (e.g., Osborn, 1929 [7]; Harrison, 1958 [252]; Conway, 1959 [253]; Hilton,1982 [254]; Ellen, 1989 [255]; Huber & Headrick, 1999 [256]). From the outset, the focus of forensic examinations was to both objectively and subjectively establish whether they were genuine (written by the specimen author) or simulated (written by an imposter/forger). In the contemporary forensic environment, signature authentication/verification is still universally carried out by humans. But outside of forensic science, computer based signature verification techniques continue to develop in response to commercial needs around quickly recording and authenticating an individualś mark [12]. Although the opinions of Forensic Handwriting Examiners (FHEs) remain the most popular method of signature authenticity determinations, computer based techniques are attracting increasing interest within the forensic community. The question here is; which is better: man or machine? To address this question, the current chapter empirically compares the performance of the two, on the same or similar material. The novelty of this work is; various state-of-the-art signature verification systems are applied to questioned signature problems which had already been worked by FHEs and then performed a comparative analysis of the two. The aim of this chapter is to make some inroads with respect to comparing the performance of the traditional human approach to that used by contemporary objective

[0] This chapter is an adapted version of work published in [58, 251].

techniques.

Note that the author does not envisage objective automated systems as replacements for FHEs; but foresee a great potential for automatic systems to assist human experts in signature analysis and interpretation in the future. It is, however, noted that there are significant limitations when using machines, since machines are generally trained on the case specific training data (containing the specimen signature samples alone from a said case). This training enables them learn the genuine signing behavior of the specimen author and by virtue of this training, they develop statistical and/or structural models for providing judgments about the questioned signatures. These models are highly influenced by the training data provided to them and therefore the representativeness of the training material is critical to ensure they model the relevant elements of the task that they are to carry out. Further, generally machines do not consider the lessons they learned while analyzing other cases (although there exist various techniques which can be applied to enable machines to utilize the writing behaviors they learned from various cases other than the specific case at hand (e.g., methods by Weber et al., 2009 [257]). Human experts, do rely on the case specific data, but also heavily rely on their previous knowledge of predictive features associated with genuine and forged behaviors and routinely apply this knowledge to the specific case at hand. The difference in philosophy between human and machine based examination strategies limits the machines to acting as a potentially good assistant, rather than a complete replacement of the FHE. Having said this, it is also recognized that commercially, outside forensic casework, machines are used in preference to humans (e.g., the banking industry) primarily due to issues associated with the volume of authentications that are required to be carried out, and the timeliness associated with task.

As stated in Chapter 2, today the PR community in general considers automatic signature verification to be a two-class pattern classification problem [12]. In some earlier PR studies it was defined differently where PR researchers at least mentioned other genres of signatures such as disguised signatures [9]. As a two class classifier, an automated system has to decide whether a given signature belongs to a referenced authentic author or not. If a system finds enough evidence of genuine authorship from the questioned signature, it considers the signature as genuine; otherwise it declares the signature forged/simulated. Clearly, this is not the case when FHEs approach signature comparison tasks. FHEs consider signature verification as a multi-class (at least three class) classification problem [20]. Along with genuine and forged signatures, they also consider the possibility that the observed combination of any similar and dissimilar features might result from disguise behavior, or might result from a myriad of other factors that could impact on the writing act and which may not be captured in the population of specimen material used in the comparison (for example illness, drug effects, writing surface effects, writer position etc.). For the purpose of this comparative study, automated systems that could look into the possibility of questioned signatures being genuine, forged, and disguised are considered. The PR systems, as well as human experts, were required to classify the given signatures in one of the three following classes, or to the class inconclusive (when they were unable to conclude anything about a signatures authenticity);

Genuine signatures: normal signatures written by the specimen writer. Forged signatures: written by some person other than the specimen writer where that person has tried to imitate the genuine signature of the specimen writer. Note that FHEs prefer using the term "simulated" rather than "forged" as the later term implies intent. However, in the thesis, the terms forged and simulated are used interchangeably since it is know which of the signatures were intentionally simulated, and also since in the PR community the term "forgery" is already in widespread usage. Disguised signatures: written by the specimen writer where there has been a deliberate attempt to change the features of the signature for the purpose of later denial. Typically the strategy associated with this behavior is either to introduce gross changes to the form that can easily be referred to, or to make the signature appear to be a forgery by executing the signature in a way that introduces feature changes that the genuine writer believes would be present if the signature was forged by another person (also referred to as auto-simulation behavior).

Table 8.1: Year-wise breakup of signature data

Year	Reference	Disguised	Forged	Genuine	Total
2001	20	47	160	43	270
2002	9	20	104	76	209
2004	16	8	42	50	116
2005	15	9	71	20	115
2006	25	7	90	3	125
Overall	85	91	467	192	835

8.1 Data

For the purpose of this study, blind test data collected by a La Trobe program run over the years 2001, 2002, 2004, 2005, and 2006, respectively, were used. Although the year by year data has not been published, the approach used and summary statistics have been presented [258, 259]. All the signatures were in the form of static images. The original signatures were scanned at 600 dpi resolution and cropped at the Netherlands Forensic Institute for the purpose of this study.-

For the year 2001, one specimen writer wrote three normal signatures per day (written with a ball point pen) over a fifteen day period, six disguised signatures per day (written with a ball point pen) over a fifteen day period, and six normal signatures per day (written with a pencil) over a three day period. From the normal signatures pool ,the genuine questioned signatures and the reference signatures (the set formed to which the questioned signatures would be compared) were constructed. Two forgers were selected from the academic staff at La Trobe University to forge the specimen writers signatures. Each of the forgers was provided with six normal samples of the specimen writers signature. The forgers were instructed that they could use any or all of the supplied specimen signatures as models for their forgeries. The forgers were also instructed that their forgeries must be unassisted (not tracings). Each forger was asked to complete the following task each day over a 10 day period: 25 practice signatures (ball point pen), 5 forgeries (ball point pen), and 5 forgeries (pencil). The forgeries, other than the practice attempts, were used as a pool from which the questioned forged signatures were selected. All the questioned samples were numbered randomly, scanned and ink-jet or laser printed into a booklet.

For the year 2001, the total selected corpus contained 270 signatures belonging to different signature categories as given in Table 8.1. For the year 2002, one specimen writer wrote fifteen normal and six disguised signatures per day over a seven day period. In addition to these signatures, the specimen writer provided an 81 genuine signature samples (27 pages containing three signatures per page). Signatures from this supplementary pool were provided to the forgers as examples of the signature they were required to forge. For forging the signatures of the specimen writer, 27 forgers were selected from volunteers drawn from groups such as secondary school teachers and professional organizations. Each of the forgers was provided with 3 normal samples of the signature written by the specimen writer. Forgers were instructed that they could use any or the entire supplied reference signatures as models for their forgeries. Forgers were also instructed that their forgeries must be unassisted (not tracings). Each forger was asked to complete the following tasks.

Inspect the genuine signature and, without practice, immediately attempt to forge it three times. Practice simulating the genuine signature fifteen times and then simulate the signature an additional three times.

The total selected corpus, for the year 2002, contained 209 signatures belonging to the different signature categories as given in Table 8.1. For the year 2004, one specimen writer wrote the normal and disguised signatures over a ten days period. From the normal signature pool the genuine and reference signatures were drawn for the specimen writer. Thirty one adult forgers were used to generate the forgery pool. These individuals were volunteers drawn from a single private company. Each of the forgers was provided with three genuine samples of the signatures written by the specimen writer. These forgers were instructed similarly to the forgers from years 2001 and 2002.

The total selected corpus, for the year 2004, contained 116 signatures belonging to the different signature categories as given in Table 8.1. For the year 2005, one specimen writer wrote the normal and disguised signatures over a ten days period. Six adult volunteer forgers were used to generate the forgery pool. Each of the forgers was provided with 3 original normal samples of the genuine signature written by the specimen writer. These forgers were instructed similarly to the forgers from previous years. The total selected corpus, for the year 2005, contained 115 signatures belonging to the different signature categories as given in Table 8.1.

For the year 2006, one specimen writer wrote the normal genuine and disguised signatures over a five day period. Seven disguised signatures and 25 normal genuine signatures were chosen from this subset. Thirty-four adult volunteer forgers contributed to the forgery set. The forgers were either lay persons or calligraphers. Similar instructions were given to the forgers as given in the previous years. The total selected corpus, for the year 2006, contained 125 signatures belonging to the different signature categories as given in Table 8.1. The results of the FHEs analysis of the data summarized in Table 8.1 were known.

8.2 Results of PR Systems

To determine how well PR approaches performed compared to FHEs, two off-line (static data only) signature verification competitions were organized at the 12th and 13th International Conferences on Frontiers in Handwriting Recognition (ICFHR). These were titled the “4NSigComp2010” and the “4NSigComp2012”.

In the 4NSigComp2010 competition, seven different systems were submitted and applied to the La Trobe signature test data. For details, refer to Chapter 5: Section 5.1.2. In addition to these seven systems, two further systems were used for these experiments; system 8 used Gaussian mixture models for classification while utilizing various local features extracted through a sliding window approach [62], and system 9 is from the author (for details, please see Chapter 7). In the 4NSigComp2012 competition, five systems were submitted. For details, refer to Chapter 5: Section 5.1.2.

The La Trobe signature images from the year 2002 were provided to the participants of the 4NSigComp2010 competition for use as training data. Signature images from the year 2006 La Trobe trial were used for evaluation. In the 4NSigComp2012 competition, the training and evaluation set (i.e., complete data from the 4NSigComp2010, i.e., the years 2002 and 2006 data) were provided as the training set to the participants, and the La Trobe data from the years, 2001, 2004, and 2005 were used for evaluation. The performance of the participating automated systems is reported here on a year-wise basis. The evaluation results are reported on the data from years 2001, 2004, 2005, and 2006 as the data from year 2002 were only used for training and were not included in any evaluation set.

All the participating systems had to classify signatures as genuine, forged, disguised, or whether they were unable to classify (this option was given to the systems as it is always present within the FHE paradigm). It is interesting to note, however, that the automated systems, despite of being provided with an option to report "inconclusive", generally provided a classification. Although the participants never explicitly stated the reasons for this, it is considered that a majority of automatic systems were initially developed for applications other than forensic. In such applications, e.g., in banking, automatic systems usually only accept or reject a signature. If rejected, authentication may further be carried out by bank staff where they can consider other proofs as well, e.g., passwords/ids, to allow the signer complete the transaction– in fact without the actual successful verification of signatures. Having said that, it is envisaged that in the future, as more automatic systems are developed for forensic applications, the option to report inconclusive opinions might also be considered.

Table 8.2: Results of automated systems on the years 2001, 2004, and 2005 data. FAR: False Acceptance Rate, FRR: False Rejection Rate, EER: Equal Error Rate. *: When disguised signatures were not included in the test set.

System	Accuracy(%)	FAR(%)	FRR(%)	EER(%)	EER*(%)
1	85.11	14.29	15.82	15.82	14.16
2	77.88	21.61	23.16	23.16	16.81
3	78.89	20.88	21.47	21.47	13.19
4	30.67	73.63	62.71	70.24	68.14
5	71.11	28.94	28.81	28.81	20.51

Table 8.3: Results of automated systems on the year 2006 data. FAR: False Acceptance Rate, FRR: False Rejection Rate, EER: Equal Error Rate. *: When disguised signatures were not included in the test set.

System	Accuracy(%)	FAR(%)	FRR(%)	EER(%)	EER*(%)
1	90	1.1	90	80	34
2	54	41.1	90	58	41
3	75	20	70	70	8
4	92	0	80	70	0
5	80	13.3	80	55	28
6	20	87	10	60	21
7	91	1.1	80	70	8
8	80	20	78	56	33
9	80	20	20	20	33

Table 8.2 shows the results when automated systems are evaluated for the years 2001, 2004, and 2005 data (in the 4NSigComp2012 competition) and Table 8.3 shows the results when the automated are evaluated systems for the year 2006 data. Both the Tables 8.2 and 8.3, show results when disguised signatures were removed from the evaluation set and the experiments were repeated. This is done to analyze the effect of the presence of disguised signatures on the performance of automatic systems. Note that, along with accuracy, the FRR and FAR are also reported in order to uncover the potential performance of the systems. In fact, accuracy is insufficient in representing the actual performance of a system (and to that effect also humans) when there are unequal number of different types of signatures, i.e., genuine, forged, and disguised, among the questioned data. The human performance is also plotted in the FRR/FAR space (see Figures 8.4 and 8.5). Furthermore, the EER (the point where FRR equals FAR) is also important as a single objective measure to rank systems performances when tested on the same data. This EER is not directly correlated with the accuracy and systems with varying accuracies can have the same EER, as shown in Table 8.3. One can also measure a system performance by putting different weight penalties when a system makes errors in identifying different types of signatures, i.e., genuine, forged, or disguised. This allows to mold the FRR/FAR metric with respect to application based preferences as whether to consider a misclassified forgery a greater error or a misclassified genuine signature a greater error and vice versa. However in these experiments, all the miss-classifications / errors are treated equally by giving them a penalty weight of 1. For further details and background issues about these metrics, please refer to Chapter 2: Section 2.6.

As given in Tables 8.2 and 8.3, the accuracy and error rates varied among systems. In general, the systems faced difficulties in classifying disguised signatures (considered most of the disguised signatures as forgeries) and nearly in all the cases (except for system 9 in Table 8.3, reasons for which in [62]) the systems performance increased when the disguised signature samples were removed from the questioned signatures. In Table 8.3, system 4 reached an error rate of 0% when disguised signatures were not considered in evaluation. In fact, this system was 100% correct in classifying forgeries and genuine signatures, but misclassified all the disguised signatures as forged. Note that one can evaluate the results of an automatic system by varying the numerical thresholds according to which an automatic system objectively classifies a signature as genuine, forged, or disguised. In contrast, there are no numerical thresholds for humans with respect to their opinions regarding the category of a signature. For example, a human expert cannot have a numerical objective threshold below which s(he) can consider a signature as forged and above which she can consider the same signature as genuine (and vice versa) and keep varying that objective threshold to give opinions about the signatures; thereby classifying certain signatures into one category at one threshold and into the other category at another threshold.

8.3 Results of FHEs

The evaluation of FHEs opinion data on the La Trobe trials had been carried out as part of the program offered over the trial years. For each of the yearly La Trobe trials, FHEs were provided with a hard copy image of each signature and an answer booklet. Examiners were informed that the date range over which the reference material was taken was around the time that the questioned samples were written. For one of the trials they were also informed that a calligrapher group was used in the production of some of the simulations (Dewhurst et al., 2008). FHEs were asked to express their opinion as to authenticity of each of the questioned signatures on a five-point scale. For simplicity, and a direct comparison, the levels of FHEs' opinions are not considered in this study.

Table 8.4: Results of the proficiency tests conducted with FHEs on the data from year 2001.

Type	**Genuine**	**Disguised**	**Forged**	**Total**
Correct Classified	128	571	2840	5039
Misleading (Errors)	30	461	265	756
Reported Inconclusive	105	895	3455	4455
Total	1763	1927	6560	10250

For the year 2001 La Trobe signature data, 51 answer booklets were submitted, comprising 10 peer reviewed responses (cross-checked by a second FHE), 31 individual responses (not peer-reviewed), and 10 experimental responses (from individuals and trainees). A total of 10250 authorship opinions were expressed by the group. Of these opinions 49.2% were correct, 7.4% were misleading and 43.5% were inconclusive. This translates into an error rate of 13.0% on the decisions (accuracy of 87.0%) when those opinions that were inconclusive were disregarded. The opinion data associated with these results is given in Table 8.4.

Table 8.5: Results of the proficiency tests conducted with FHEs on the data from year 2004.

Type	**Genuine**	**Disguised**	**Forged**	**Total**
Correct Classified	990	69	343	1402
Misleading (Errors)	1	13	9	23
Reported Inconclusive	9	78	488	575
Total	1000	160	840	2000

For the year 2004 La Trobe signature data, 21 answer booklets were submitted, comprising 7 peer reviewed responses (cross-checked by a second FHE), and 14 individual responses (not peer-reviewed). A total of 2000 authorship opinions were expressed by the group. Of these opinions 1402 (70.1%) were correct, 23 (1.2%) were misleading and 575 (28.8%) were inconclusive. This translates into an error rate of 1.6% on the decisions (accuracy of 98.4%) when those opinions that were inconclusive were disregarded. A detailed breakdown of these results is given in Table 8.5.

Table 8.6: Results of the proficiency tests conducted with FHEs on the data from year 2005.

Type	**Genuine**	**Disguised**	**Forged**	**Total**
Correct Classified	587	73	1263	1923
Misleading (Errors)	1	52	174	227
Reported Inconclusive	32	154	764	950
Total	620	279	2201	3100

For the year 2005 La Trobe signature data, in total, 31 answer booklets were submitted, comprising 5 peer reviewed responses (crosschecked by a second FHE), and 26 individual responses. A total of 3100 authorship opinions were expressed by the group. Of these opinions 1923 (62.0%) were correct, 227 (7.3%) were misleading, and 950 (30.6%)

were inconclusive. This translates into an error rate of 10.6% on the decisions (accuracy of 89.4%) when those opinions that were inconclusive were disregarded. A detailed breakdown of these results is given in Table 8.6.

For the La Trobe data collection of year 2006, in total, 33 answer booklets were submitted, comprising 11 peer reviewed responses (cross-checked by a second FHE) and 22 individual responses (not peer reviewed). A total of 3100 authorship opinions were expressed by the group. Of these opinions 40.5% were correct, 7.2% were misleading and 52.3% were inconclusive. This translates into an error rate of 15.2% on the decisions (accuracy of 84.8%) when those opinions that were inconclusive were disregarded. The opinion data associated with these results is given in Table 8.7. In addition to the collective results, various tests were performed to analyze the errors made by individual examiners.

Table 8.7: Results of the proficiency tests conducted with FHEs on the data from year 2006.

Type	**Genuine**	**Disguised**	**Forged**	**Total**
Correct Classified	93	10	1151	1254
Misleading (Errors)	0	111	113	224
Reported Inconclusive	0	96	1526	1622
Total	93	217	2790	3100

Figure 8.1 shows the examiner scores (inconclusive and misleading/incorrect opinions) by questioned signature category for the 2006 trial. The percentage inconclusive opinions are colored yellow and the percentage incorrect opinions are colored red. The x-axis depicts the examiners anonymous identification code. It can be seen that a large number of FHEs were either inconclusive or they misclassified the forgeries and disguised signatures. They, on the other hand, were quite good at identifying the genuine signatures individually. The results from the other years show a similar trend, however for brevity, they are not included here. Since one might predict that FHEs will exhibit a much wider range of performance success as compared to automatic systems, several other tests were performed to characterize the FHE data. The relationship between examiners experience and the total number of opinion errors (see Figure 8.2), and the relationship between the time examiners took to complete the trials and the total number of opinion errors (see Figure 8.3) is presented here for interest. For brevity, the results for the 2001 data are reported here. Both for Figure 8.2 and Figure 8.3, no simple correlation was found to exist between the two variables (at x and y axis). The experiments show that there is

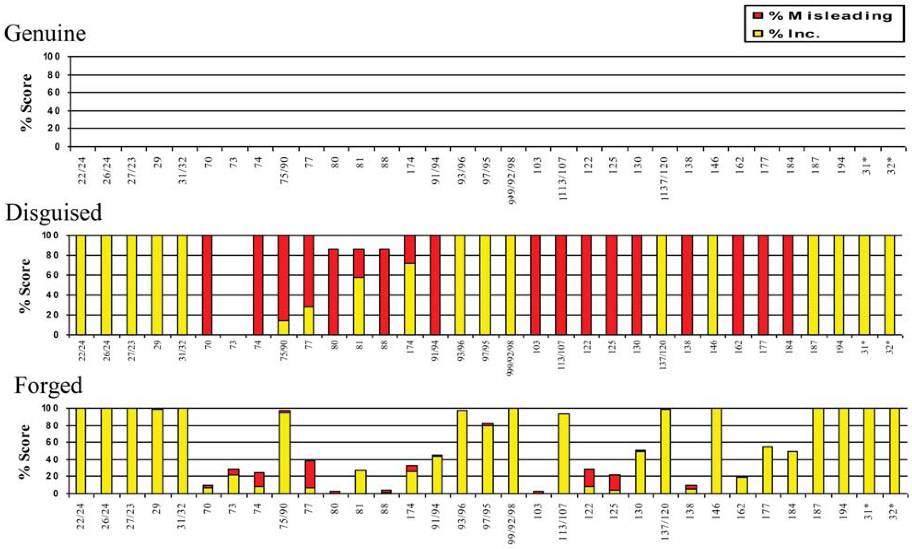


Figure 8.1: Results of individual FHEs on the La Trobe 2006 data. Total examiners: 33. X-axis: IDs of examiners, Y-axis: percentage score of each examiner.

no support for the notion that the validity of a trained examiners opinion can be referenced by the number of years the examiner has been practicing and also no support for the notion that the validity of a trained examiners opinion can be referenced by the amount of time the examiner spent performing the task. The time taken by automatic systems to complete the verification task for the data sets from different years was also computed. Most of the automatic systems were able to complete the task very efficiently, e.g., the most efficient automatic system was able to output results for the whole data of year 2006 in less than 100 seconds. The important point here is that the state-of-the-art automatic systems usually look only at specific information/evidence present in the signatures (and of course machines are quite efficient in processing specific information in this way). The humans in our study, on the other hand, use complex perceptual and cognitive processes to assess all of the features of the questioned signature trace and not surprisingly take vastly longer to perform the task. Taking this into consideration, a direct time comparison between man and machine is not reported in this chapter.

8.4 Comparison

The overall man vs. machine comparison was initially performed on the basis of collective accuracies of the two, man and machine. Table 8.8 provides the overall results of this performance comparison taking into account the complete signatures; i.e., along

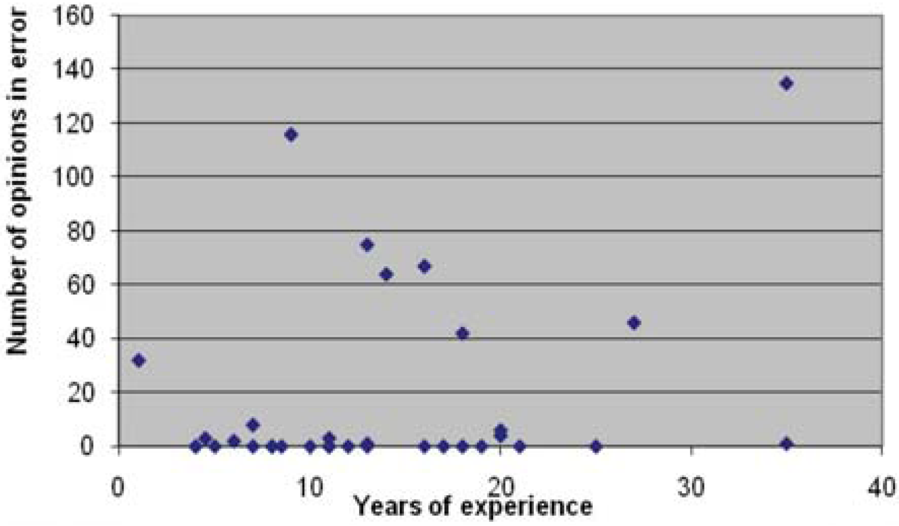


Figure 8.2: Relationship between examiners experience and the total number of opinion in errors. Points indicate the years of experience of FHEs and the corresponding number of opinions expressed in error.

with genuine and forged signatures, the disguised signatures were also considered while computing these results. The average as well as the best performances are reported to provide a clear comparison between man and machine. As shown in Table 8.8, there is much variation in human performance from trial to trial when compared to that of machines. For example, in the year 2001, the average human performance is at 44.8% although the best performance is at 100%. Similar trends can be seen for the results from other years. For humans, therefore, one can assume a large variance in performance in general; whereas for machines the average performance is at 70.8% and the best at 93.6% showing comparatively less variance. Here one may infer that most of the state-of-art automatic methods (applying different classification approaches) perform close to each other, and humans carry great performance diversity, both in terms of accuracy and speed. It is clear that automatic systems could provide a good supplementary objective tool for FHEs as they provide quite consistent results. Further, these systems can also be used to cut down a large population into a smaller population (due to their speed) when examining real world signature cases.

In order to measure the total performance capabilities of automatic systems, Receiver Operating Characteristic (ROC) curves [186]. These curves are given in Figure 8.4 (combined data from 2001, 2004, and 2005) and Figure 8.5 (data from 2006). Note that an automatic system has many possible points/thresholds on which it can operate to reach an opinion on classifying signatures as genuine, forged, or disguised. Therefore, it is

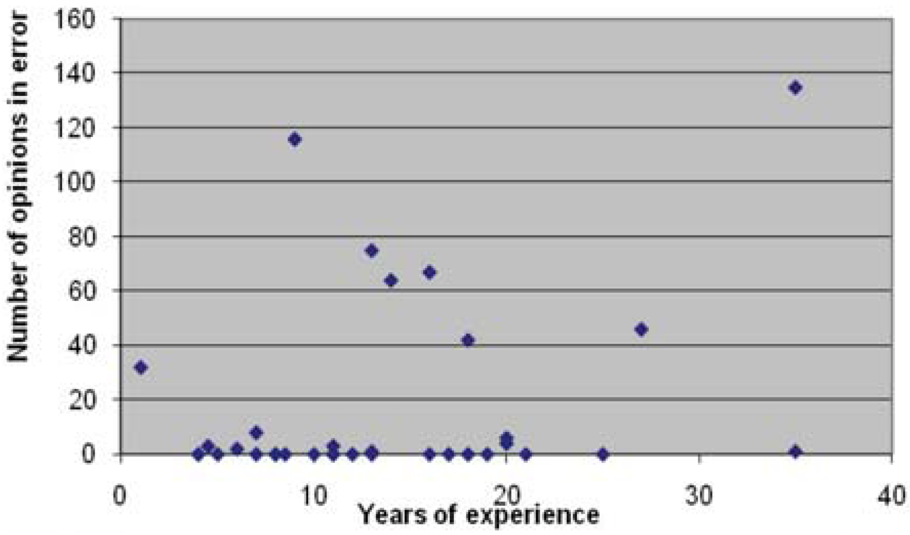


Figure 8.3: Relationship between the time taken by FHEs and the total number of opinions in error. Points indicate the time taken by FHEs to complete the trial and the corresponding number of opinions in error.

Table 8.8: Man vs. machine results for the data collections from various years.

Data from the year	**Accuracy**			
	Avg. Human	Avg. Machine	Best Human	Best Machine
2001	44.8	70.8	100	93.6
2004	66.2	70.4	97	87
2005	62	59.8	100	68
2006	38.8	71.7	91	92

preferred to represent the complete performance behavior in the form of so-called ROC curves. Generally, these curves are developed by considering the FRR on one axis and FAR on the other axis while varying the thresholds on the basis of which a system gives an opinion about signature type (genuine, forged, or disguised). This generates the complete behavior of a system on the given data in the form of a curve in the FRR/FAR space. In Figures 8.4 and 8.5, FHEs performance, unlike automated systems, is represented by single points. These points are calculated by looking into the overall experts performance. The false acceptance is calculated by taking the ratio of the forged questioned signatures which were misclassified as genuine by the examiners and the total forged questioned signatures. The inconclusive opinions were not considered. The false rejection was computed by taking the ratio of the sum of disguised and genuine questioned signatures which were misclassified as forged, and the total disguised and genuine questioned signatures.

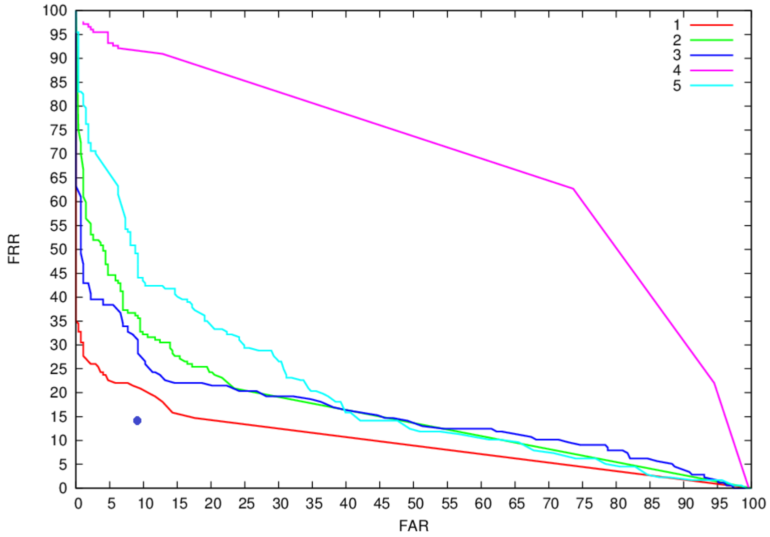


Figure 8.4: Man vs. machine comparison in the FAR/FRR space (on combined 2001, 2004, and 2005 data).

The inconclusive opinions were again neglected. These are plotted as single points in the FRR/FAR space (the same containing the ROC-curves for the automatic systems). Note that, unlike machines, a complete ROC curve of human performance is impossible since there are no objective numerical thresholds for humans who they can vary with respect to their opinions regarding signature classification [137].

As can be observed from Figures 8.4 and 8.5, humans outperformed most of the machines (except that proposed in this thesis outperformed humans at certain points). An important reason for these results is that humans used a possibility to note their opinion as inconclusive when they were unable to find enough evidence of genuine or forged authorship as per their analysis. The machines were also given this possibility but none of the machines used this, and nearly in all the cases came up with an opinion. Nonetheless, the humans also carried a great deal of previous knowledge in terms of their experiences in solving forensic cases, but machines relied on the case specific data alone. This might have also affected the performance of machines. Note that currently there are some limitations associated with automatic systems which are required to be overcome in order for these systems to be applicable in real world forensic cases. The main challenges are that these systems need to train on more forensic data captured from real casework to improve system learning [13], that some forensic environments require outputs in the form

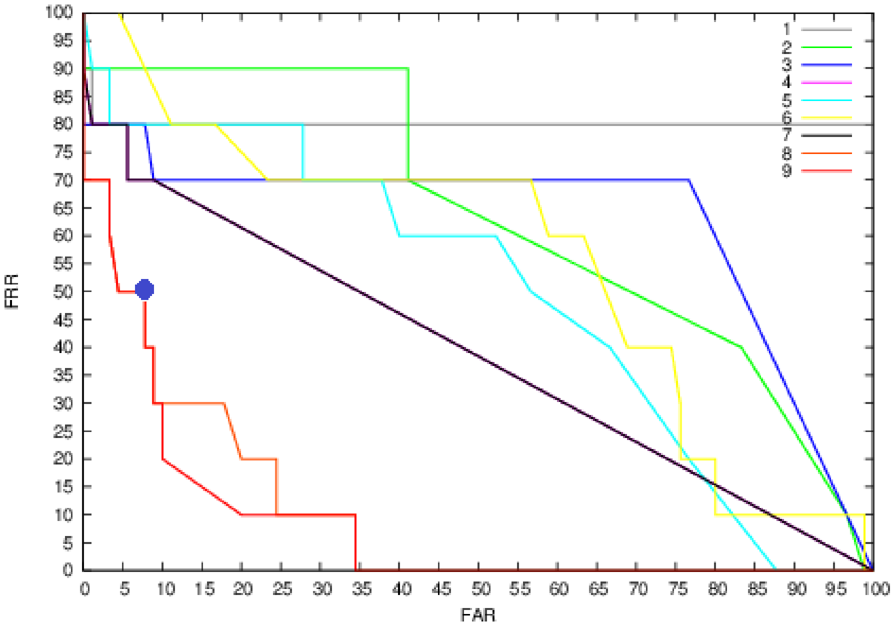


Figure 8.5: Man vs. machine comparison in the FAR/FRR space (on 2006 data). Systems 1-7 are participants of the 4NSigComp2010 competition while system 8 and 9 are added later.

of likelihood ratios (according to the Bayesian inference) to be acceptable as a laboratory output [19, 20], and that automatic systems should provide explanations of their outputs such that FHEs can weigh the probative value of their outputs accordingly. It is the case that no state-of-the-art automatic system is currently capable of completely fulfilling these desired requirements. Having said this, since automatic systems are extremely efficient when compared to humans, they have the potential to serve as assistants for human experts where they may potentially be guided by fast objective data. Many methods can be devised to enable machines to automatically incorporate knowledge from previous cases, e.g., using case based reasoning [257]. The challenge here is investigate whether it is really helpful, required, or even recommended to consider the incorporation of previous knowledge when automatically classifying signatures by machines.

8.5 Discussion

This chapter has provided the results of a detailed study performed in order to compare signature verification performance of FHEs against automated systems. As the technology around automated systems develops, the potential applications for these objective systems in forensic casework grow. This paper shows that the performance of automatic systems, although in many respects unlike that of FHEs, can result in characteristics approaching the average for FHEs participating in the La Trobe trials. This study suggests that different automatic systems, just like humans, were better on different data. However, there was not much variance in the performances of automatic systems which is unlike that of humans, as humans performance showed a great degree of variation from average to the best case performances.

The automated systems/machines encountered difficulties in correctly classifying disguised signatures. When disguised signatures were removed from the test data, some automatic systems could reach an EER of nearly 0% in one of the datasets. Similar to machines, FHEs also faced difficulties when they attempt to classify questioned signatures that are a product of disguise behavior. This is likely to result from the mixed signal that disguised signatures provide to both FHEs and the automated machines. Similarities may exist with the genuine signature, since it was written by the writer of the authentic signatures, and dissimilarities may exist due to the conscious changes to the signature made by the genuine writer in order to introduce features where denial can be claimed. The human experts faced problems in correctly classifying disguised signatures; however, they had freely used a possibility to declare their findings inconclusive on the basis of not being able to find enough evidence of genuine, forged, or disguised authorship from the signatures (the automatic systems were also provided this possibility but no participating system used this). In fact, a large number of human trials reported disguised and forged signatures as inconclusive. Furthermore, both humans and machines were in many cases accurate in identifying genuine signatures.

Performance comparisons of the type described here offer promise regarding the future of objective techniques in forensic casework. One must be careful, however, not to overestimate the potential of automated techniques since this study is based on data derived not from casework, but carefully constructed blind trials. The signature sets, both specimen and questioned, are therefore very clean with respect to controlling variables which are not normally able to be controlled in casework (e.g., controlling the representativeness of the population of specimen signatures, the type of writing instrument, the writing medium, the writing conditions etc.). In many cases, casework samples suffer from a lack of specimen signatures, or specimen signatures that may not be representative of the writerś normal behavior. However, still more research is needed to discover that to what extent these limitations impact the potential of automated systems so that to produce accurate and useful results for forensic casework.

9

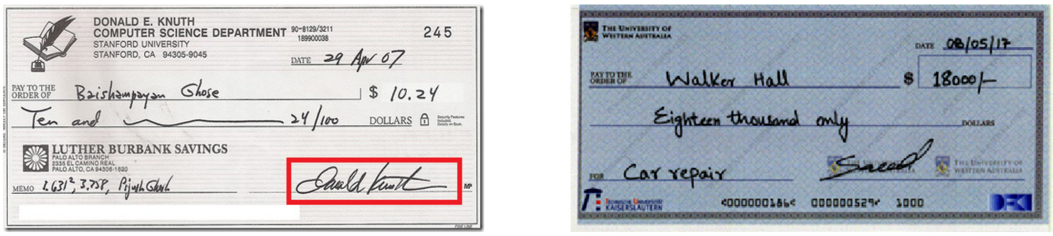
Signature Segmentation from Document Images

9.1 Introduction

Automatic systems are already being used in almost every field to facilitate different processes in daily life ranging from a simple vending machine to ATM machines and sophisticated systems for automatically diverting postal mails, etc. With the development in the field of document image analysis, there appeared different automatic systems which analyze documents and extract various types of information for different purposes, e.g., sorting of postal mails based on zip code [261], optical character recognition, automatic extraction of; names, addresses, numbers, and dates, etc. from documents [262]. Researchers are aiming to develop systems capable of verifying authenticity of documents on the basis of certain information extracted from these documents, e.g., signatures.

Signatures are considered as an important identity for authentication. Signatures are a widely used authentication mechanism in banks, contracts, credit card payments, etc. Researchers in the last 4 decades have already put a lot of effort for development of offline (using only spatial information, e.g., scanned signature images) and online (using both spatial and temporal/dynamic information) signature verification systems. In almost all of the existing systems it is assumed that signatures are already segmented. Training as well as testing of these systems is also performed on segmented signatures. In addition, existing publicly available datasets for development and evaluation of signature verification systems also contain only segmented signatures. However, in real world scenarios, signatures are a part of documents which also contain other information, e.g., bank checks (see Figure 9.1), invoices, contracts, credit card pay slip, wills, etc. In these scenarios, the existing signature verification and identification systems cannot be used as is, because of

[0] This chapter is an adapted version of work carried out and published collaboratively in [28, 260].



DONALD E. KNUTH
COMPUTER SCIENCE DEPARTMENT 90-8129/3211 189900038
STANFORD UNIVERSITY
STANFORD, CA 94305-9045

245

DATE 29 Apr 07

PAY TO THE ORDER OF Baishampayan Ghose $ 10.24

Ten and 24/100 DOLLARS

LUTHER BURBANK SAVINGS
PALO ALTO BRANCH
2335 EL CAMINO REAL
PALO ALTO, CA 94306-1620

MEMO 1.631², 3.758, Pijush Ghosh

Donald Knuth MP

(a)

THE UNIVERSITY OF WESTERN AUSTRALIA

DATE 08/05/17

PAY TO THE ORDER OF Walker Hall $ 18000/-

Eighteen thousand only DOLLARS

FOR Car repair

Saeed

THE UNIVERSITY OF WESTERN AUSTRALIA

Technische Universität KAISERSLAUTERN

⑆0000000186⑆ 000000529⑈ 1000

DFKI

(b)

Figure 9.1: Bank check images([6])

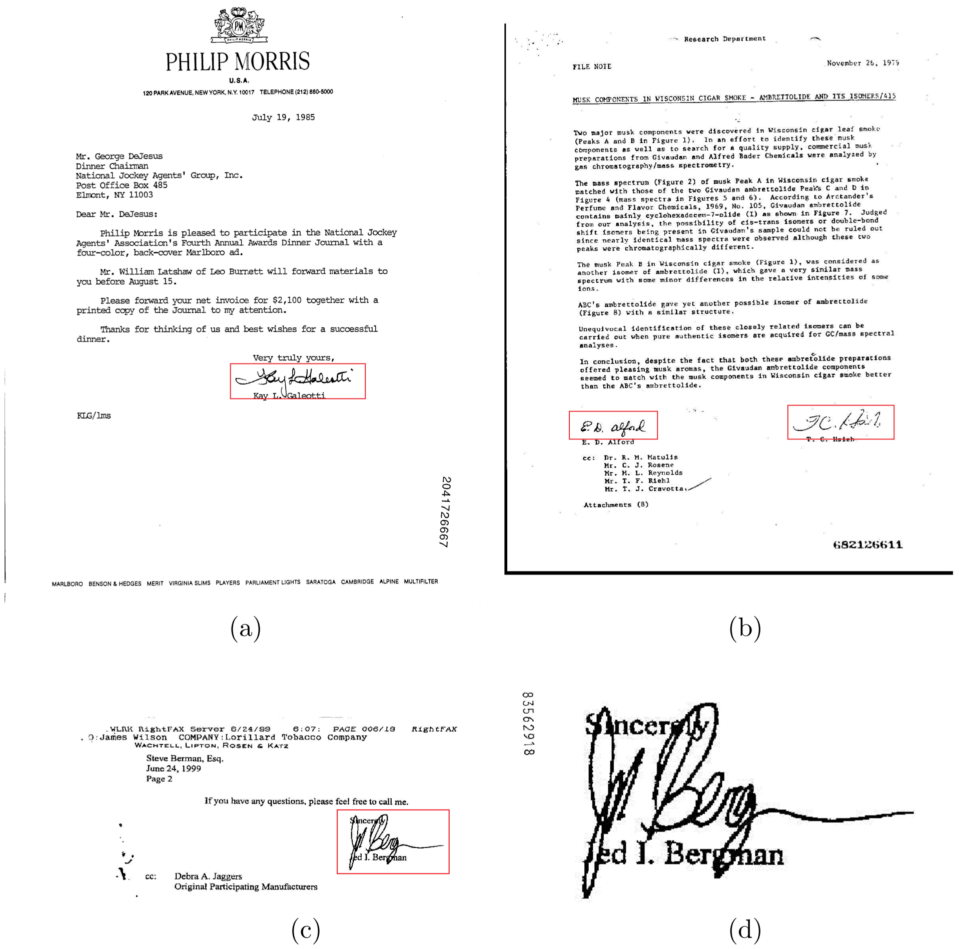
the lack of segmentation. Similarly in case of forensic documents, signatures are also often found in a non-segmented form. This chapter focuses on the challenges which must be tackled for the development of a complete automatic document analysis system capable of performing signature segmentation/extraction from documents and then performing signature verification. The remaining chapter is organized as follows. Section 9.2 provides an overview of the systems available for information (particularly, signature) segmentation. Section 9.3 provides details on the presented method for signature segmentation. Section 9.4 presents details of the dataset, the evaluation protocol, and the results of the proposed signature segmentation method. Finally, Section 9.5 provides details about the related open research questions raised by virtue of this chapter.

9.2 Existing Segmentation Systems

In the past, signature segmentation has not been considered by many researchers, especially in case of document images, whereas mostly the focus remained segmentation of handwritten from printed text. This section will provide an overview about the existing systems for segmentation of handwriting from machine printed text as well as segmentation of signatures in bank checks and documents.

9.2.1 Printed & Handwritten Text Segmentation

Imade et al. [263] proposed a method for segmentation and classification of printed character, handwritten character, photograph, and painted image regions. Feed-forward neural network is used for classification of different segments. Similarly, Kuhnke et al. [264] proposed a classification system which reads a raster image of a character and outputs confidence values for machine-written and hand-written character classes. Different features are extracted and passed to feed forward neural network for getting the confidence



PHILIP MORRIS
U.S.A.
120 PARK AVENUE, NEW YORK, N.Y. 10017 TELEPHONE (212) 880-5000

July 19, 1985

Mr. George DeJesus
Dinner Chairman
National Jockey Agents' Group, Inc.
Post Office Box 485
Elmont, NY 11003

Dear Mr. DeJesus:

Philip Morris is pleased to participate in the National Jockey Agents' Association's Fourth Annual Awards Dinner Journal with a four-color, back-cover Marlboro ad.

Mr. William Latshaw of Leo Burnett will forward materials to you before August 15.

Please forward your net invoice for $2,100 together with a printed copy of the Journal to my attention.

Thanks for thinking of us and best wishes for a successful dinner.

Very truly yours,

Kay L. Galeotti

Kay L. Galeotti

KLG/lms

2041726667

MARLBORO BENSON & HEDGES MERIT VIRGINIA SLIMS PLAYERS PARLIAMENT LIGHTS SARATOGA CAMBRIDGE ALPINE MULTIFILTER

(a)

Research Department

FILE NOTE

November 26, 1979

MUSK COMPONENTS IN WISCONSIN CIGAR SMOKE - AMBRETTOLIDE AND ITS ISOMERS/415

Two major musk components were discovered in Wisconsin cigar leaf smoke (Peaks A and B in Figure 1). In an effort to identify these musk components as well as to search for a quality supply, commercial musk preparations from Givaudan and Alfred Bader Chemicals were analyzed by gas chromatography/mass spectrometry.

The mass spectrum (Figure 2) of musk Peak A in Wisconsin cigar smoke matched with those of the two Givaudan ambrettolide Peaks C and D in Figure 4 (mass spectra in Figures 5 and 6). According to Arctander's Perfume and Flavor Chemicals, 1969, No. 105, Givaudan ambrettolide contains mainly cyclohexadecen-7-olide (I) as shown in Figure 7. Judged from our analysis, the possibility of cis-trans isomers or double-bond shift isomers being present in Givaudan's sample could not be ruled out since nearly identical mass spectra were observed although these two peaks were chromatographically different.

The musk Peak B in Wisconsin cigar smoke (Figure 1), was considered as another isomer of ambrettolide (I), which gave a very similar mass spectrum with some minor differences in the relative intensities of some ions.

ABC's ambrettolide gave yet another possible isomer of ambrettolide (Figure 8) with a similar structure.

Unequivocal identification of these closely related isomers can be carried out when pure authentic isomers are acquired for GC/mass spectral analyses.

In conclusion, despite the fact that both these ambretolide preparations offered pleasing musk aromas, the Givaudan ambrettolide components seemed to match with the musk components in Wisconsin cigar smoke better than the ABC's ambrettolide.

E. D. Alford

E. D. Alford

T. C. Hsieh

cc: Dr. R. M. Matulis
Mr. C. J. Rosene
Mr. M. L. Reynolds
Mr. T. F. Riehl
Mr. T. J. Cravotta

Attachments (8)

682126611

(b)

.WLRK RightFAX Server 6/24/99 6:07: PAGE 006/19 RightFAX
. O:James Wilson COMPANY:Lorillard Tobacco Company
WACHTELL, LIPTON, ROSEN & KATZ

Steve Berman, Esq.
June 24, 1999
Page 2

If you have any questions, please feel free to call me.

Sincerely,

Jed I. Bergman

cc: Debra A. Jaggers
Original Participating Manufacturers

83562918

(c)

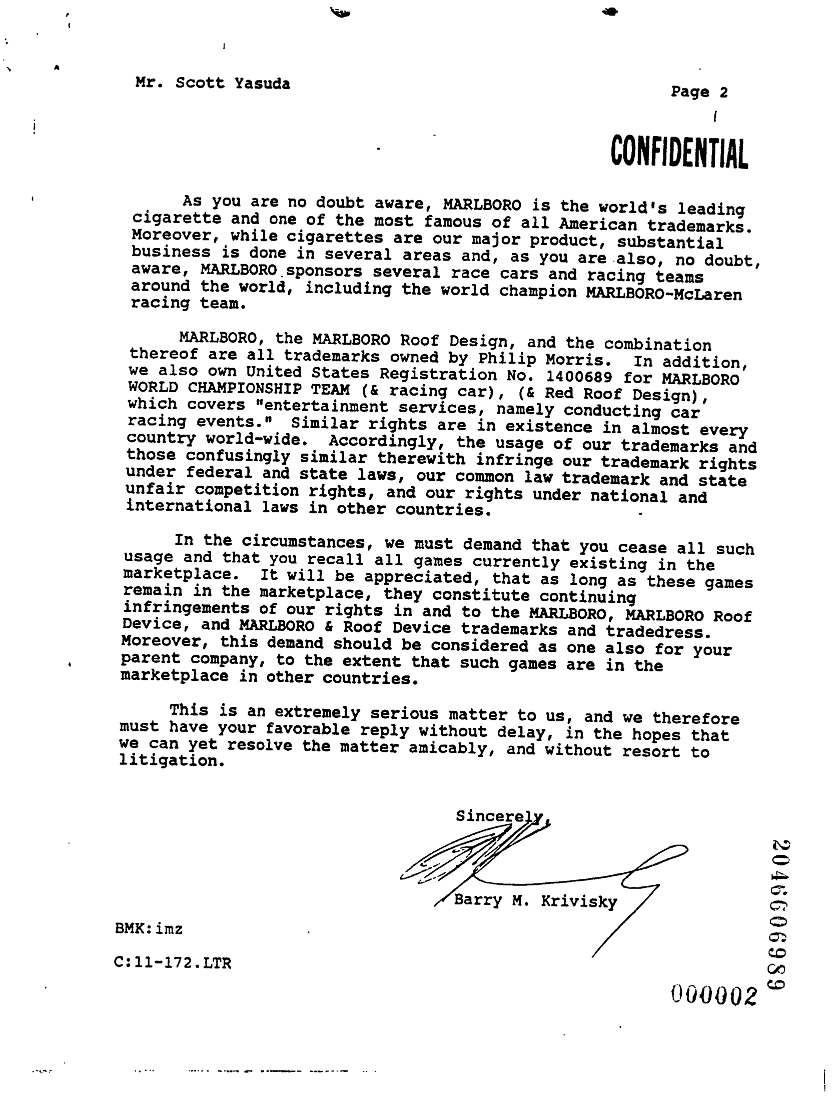
Sincerely,

Jed I. Bergman

(d)

Figure 9.2: (a), (b), (c) Signatures at different positions in document images, (d) Signature overlapping with text

score. Guo et al. [265] addresses the problem of separating handwritten annotations from machine-printed text within a document. Hidden Markov models (HMMs) are used to distinguish between machine-printed and handwritten materials. Zheng et al. [266] proposed a system for detection of machine printed and handwritten text in noisy documents. First, trained Fisher classifier is used for separation of machine printed and handwritten text from noise. After noise filtering, Markov Random Field is used to segment the machine printed text from handwritten. Similarly, Chanda et al. [267] used chain-code feature with Support Vector Machine (SVM) classifier for segmentation of machine printed text from handwritten. Recently, Saeed et al. [268] and Purnendu et al. [269] proposed systems for segmentation of handwriting from machine printed text in Farsi/Arabic and Bangla respectively.



Mr. Scott Yasuda

Page 2

CONFIDENTIAL

As you are no doubt aware, MARLBORO is the world's leading cigarette and one of the most famous of all American trademarks. Moreover, while cigarettes are our major product, substantial business is done in several areas and, as you are also, no doubt, aware, MARLBORO sponsors several race cars and racing teams around the world, including the world champion MARLBORO-McLaren racing team.

MARLBORO, the MARLBORO Roof Design, and the combination thereof are all trademarks owned by Philip Morris. In addition, we also own United States Registration No. 1400689 for MARLBORO WORLD CHAMPIONSHIP TEAM (& racing car), (& Red Roof Design), which covers "entertainment services, namely conducting car racing events." Similar rights are in existence in almost every country world-wide. Accordingly, the usage of our trademarks and those confusingly similar therewith infringe our trademark rights under federal and state laws, our common law trademark and state unfair competition rights, and our rights under national and international laws in other countries.

In the circumstances, we must demand that you cease all such usage and that you recall all games currently existing in the marketplace. It will be appreciated, that as long as these games remain in the marketplace, they constitute continuing infringements of our rights in and to the MARLBORO, MARLBORO Roof Device, and MARLBORO & Roof Device trademarks and tradedress. Moreover, this demand should be considered as one also for your parent company, to the extent that such games are in the marketplace in other countries.

This is an extremely serious matter to us, and we therefore must have your favorable reply without delay, in the hopes that we can yet resolve the matter amicably, and without resort to litigation.

Sincerely,

Barry M. Krivisky

BMK:imz

C:11-172.LTR

000002

2046069989

Figure 9.3: Original document

9.2.2 Signature Segmentation from Bank-checks

Jayadevan et al. [270] proposed a comprehensive survey on systems for processing of bank checks, but this survey does not contain information about signature segmentation methods proposed for bank checks. In case of bank checks, Djeziri et al. [42] and Madasu et al. [43] specifically proposed methods for extraction/segmentation of signatures. Djeziri et al. [42] proposed a method which is inspired from human visual perception and based on filiformity criteria. Using this criteria contour lines of objects are differentiated from handwritten lines. Madasu et al. [43] used sliding window to calculate the entropy and to fit the window to signature block. Sankari et al. [271] proposed an approach for segmen-

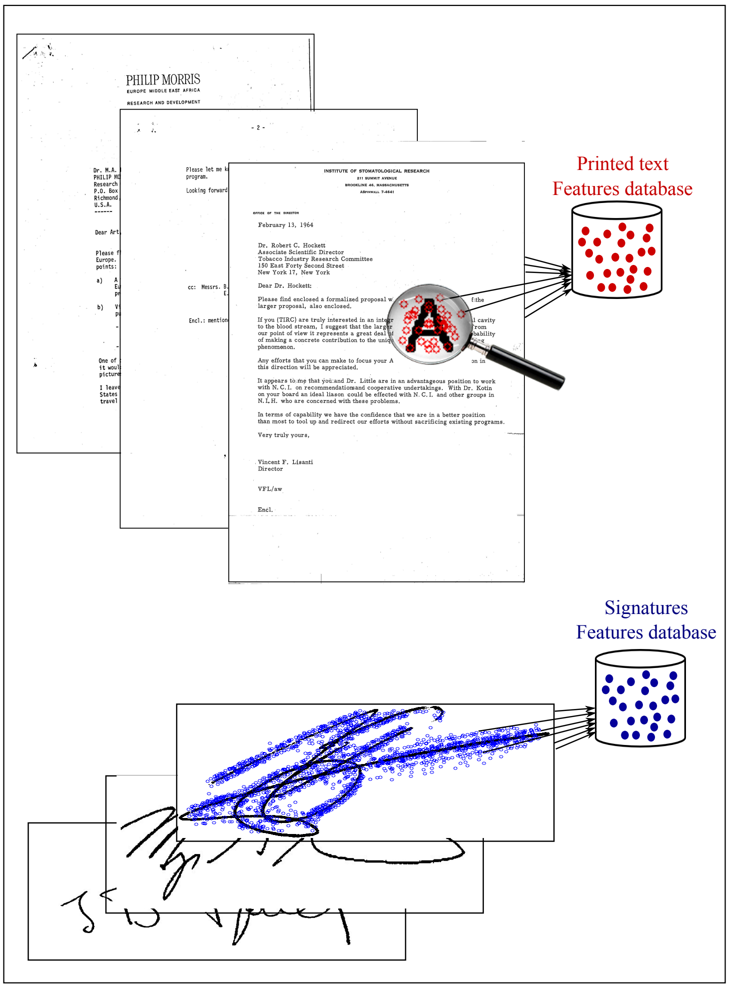
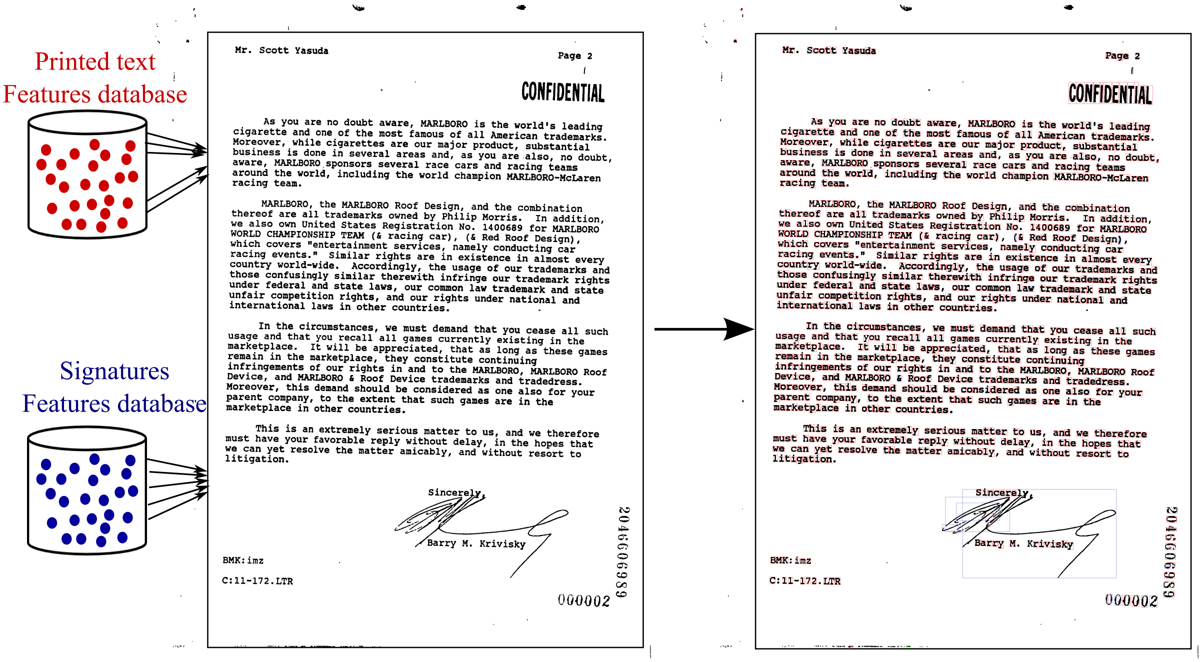


Figure 9.4: Training

tation of bank check account number and account holders signature from check images using prior knowledge of Cartesian coordinate space. These segmented regions are further used for training and verification using Hamming distance measures. In case of bank checks, a priori information about location of signatures is already available which makes the segmentation process comparatively easy. Therefore, most of the existing system for signature verification can be optimized and applied directly to bank checks if this prior information is available. In case where the prior information regarding particular location/position of signatures in a document is not known, application of current automatic signature verification systems is still a challenge.



Mr. Scott Yasuda

Page 2

CONFIDENTIAL

As you are no doubt aware, MARLBORO is the world's leading cigarette and one of the most famous of all American trademarks. Moreover, while cigarettes are our major product, substantial business is done in several areas and, as you are also, no doubt, aware, MARLBORO sponsors several race cars and racing teams around the world, including the world champion MARLBORO-McLaren racing team.

MARLBORO, the MARLBORO Roof Design, and the combination thereof are all trademarks owned by Philip Morris. In addition, we also own United States Registration No. 1400689 for MARLBORO WORLD CHAMPIONSHIP TEAM (& racing car), (& Red Roof Design), which covers "entertainment services, namely conducting car racing events." Similar rights are in existence in almost every country world-wide. Accordingly, the usage of our trademarks and those confusingly similar therewith infringe our trademark rights under federal and state laws, our common law trademark and state unfair competition rights, and our rights under national and international laws in other countries.

In the circumstances, we must demand that you cease all such usage and that you recall all games currently existing in the marketplace. It will be appreciated, that as long as these games remain in the marketplace, they constitute continuing infringements of our rights in and to the MARLBORO, MARLBORO Roof Device, and MARLBORO & Roof Device trademarks and tradedress. Moreover, this demand should be considered as one also for your parent company, to the extent that such games are in the marketplace in other countries.

This is an extremely serious matter to us, and we therefore must have your favorable reply without delay, in the hopes that we can yet resolve the matter amicably, and without resort to litigation.

Sincerely,

Barry M. Krivisky

BMK:imz

C:11-172.LTR

000002

2046606989

Mr. Scott Yasuda

Page 2

CONFIDENTIAL

As you are no doubt aware, MARLBORO is the world's leading cigarette and one of the most famous of all American trademarks. Moreover, while cigarettes are our major product, substantial business is done in several areas and, as you are also, no doubt, aware, MARLBORO sponsors several race cars and racing teams around the world, including the world champion MARLBORO-McLaren racing team.

MARLBORO, the MARLBORO Roof Design, and the combination thereof are all trademarks owned by Philip Morris. In addition, we also own United States Registration No. 1400689 for MARLBORO WORLD CHAMPIONSHIP TEAM (& racing car), (& Red Roof Design), which covers "entertainment services, namely conducting car racing events." Similar rights are in existence in almost every country world-wide. Accordingly, the usage of our trademarks and those confusingly similar therewith infringe our trademark rights under federal and state laws, our common law trademark and state unfair competition rights, and our rights under national and international laws in other countries.

In the circumstances, we must demand that you cease all such usage and that you recall all games currently existing in the marketplace. It will be appreciated, that as long as these games remain in the marketplace, they constitute continuing infringements of our rights in and to the MARLBORO, MARLBORO Roof Device, and MARLBORO & Roof Device trademarks and tradedress. Moreover, this demand should be considered as one also for your parent company, to the extent that such games are in the marketplace in other countries.

This is an extremely serious matter to us, and we therefore must have your favorable reply without delay, in the hopes that we can yet resolve the matter amicably, and without resort to litigation.

Sincerely,

Barry M. Krivisky

BMK:imz

C:11-172.LTR

000002

2046606989

Figure 9.5: Extracted and marked connected components from question document image

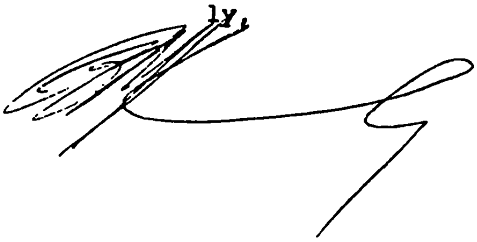


Figure 9.6: Extracted signatures from the document of Figure 9.3.

9.2.3 Signature Segmentation from Document Images

There are plenty of documents other than bank checks which also contain signatures, e.g., contracts, invoices, pay slips, wills, etc. Figure 9.2 shows that segmentation in these document images becomes more challenging as signatures can be at different parts of documents depending upon the content. To deal with signatures in complete documents, Zhu et al. [44] proposed method for segmenting signatures from complete document using saliency map. In addition to the method for signature segmentation, Zhu et al. [44] also introduced a publicly available dataset namely Tobacco-800 consisting of complex document images which contain information about signatures on printed text documents. Along with other information, Tobacco-800 contains patch level information for 900 signatures on complete documents. A complete document retrieval system is presented by

Zhu et al. [272] where signature matching is combined with detection framework from [44] to have a complete document retrieval system. Mandal et al. [273] proposed an approach for signature segmentation using conditional random fields. Results are reported on a subset of Tobacco-800 dataset i.e., 105 images out of 1290. One of the main problems of this approach is that it requires a large number of training samples. Mandal et al. [273] segment signatures on patch level, whereas in some cases some text is touching the signature components which may cause problems later in signature verification. To rectify the problem of touching characters, recently Partha et al. [274] proposed an approach for segmentation of signatures along with the segmentation of those characters which are touching signature strokes.

A major problem with all of the above mentioned approaches of signature segmentation, from document images, is that none of them is applied on the complete dataset (Tobacco-800). Subsets of Tobacco-800 dataset are used and it is not mentioned that which images are included in any particular subset. This makes it difficult to find the behavior of these methods in case of some other existing classes. Also none of the above mentioned systems reported efficiency of system in term of time and complexity.

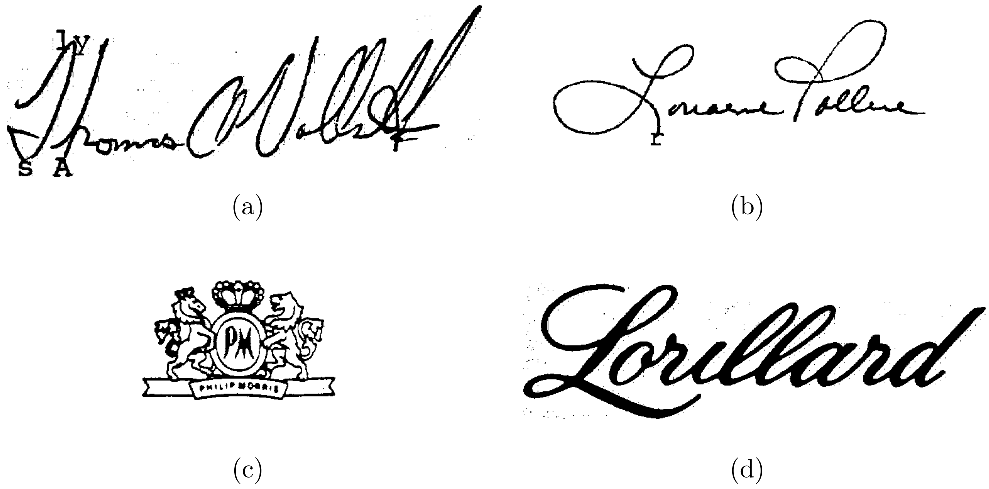
Furthermore, some commercial systems capable of finding one or two signatures in bank checks as well as IRD images and snippets and apply signature verification on these segmented signatures are available, e.g., SignatureXpert-2[1] by Parascript. The problem with commercial systems is that the data on which they are trained is never made available publicly. Also details about how these systems are working and the methods they are applying are also not known.

9.3 Proposed Method for Signature Segmentation

This Section provides an insight into the proposed method for signature segmentation. The proposed method is based on Local features, i.e., Speeded Up Robust Features (SURF). SURF is a part based approach that represents image as a set of keypoints. As part based approaches extract keypoints from the parts of image (which represent local features), it brings robustness against different variations in the image [132, 56].

For each of the SURF keypoints a 128 bit descriptor is extracted that represents the keypoint. This descriptor is used to find the similarity between different keypoints. For extraction of SURF features a Hessian threshold of 400 is used, i.e., all the keypoints

[1] http://www.parascript.com/recognition-products/forms-processing/signaturexpert-2



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

Figure 9.7: Examples of correctly segmented signatures (a, b) and false positives (c, d)

having Hessian threshold less than 400 were neglected. For details about the internal working of SURF, please refer to Chapter 6.

For training, 10 documents from the Tobaaco-800 dataset containing machine printed text and signatures are used. To train the system it is required that all the machine printed text is separated from the signatures (see Figure 9.4). As the Tobbaco-800 dataset does not include ground truth for machine printed text, two new images for each document were generated manually, i.e., printed text and signature images. These images were used for training. Connected components are extracted for each of the printed text as well as signature image of training set.

For all the connected components from the printed text image, the corresponding extracted keypoints and their respective descriptors are added to printed text features database. Similarly, for each connected component of signature image, extracted keypoints and their respective descriptors are added to signature features database. These two databases serve as reference for the matching of the features during testing.

To segment a signature from a document in the test set (as in Figure 9.3) which contains both signatures as well as printed text, connected components are extracted. For each connected component SURF features are extracted. The descriptor of every keypoint is compared with all the descriptors of printed text keypoints and signature keypoints from the reference databases. The Euclidean distance metric is used as a distance measure.

Finally, for classification of the connected components, a majority voting approach is applied. If a connected component's keypoint has less Euclidean distance to the signature keypoints as compared to the printed text keypoints, one vote is added to the signatures class and vice versa. The process is repeated until all of the connected components are assigned to one of the two classes (See Figure 9.5). Once all of the connected components are marked as printed text/signature, separate image for signature is generated. To segment the signature from the test document, the original image is cloned and bounding boxes of all connected components of printed text are filled with white color on that image, which in turn results in a segmented signature image.

As a post processing step horizontal run length smearing is performed on the segmented signature image. Applying smearing merges all of the neighboring components. Connected components are extracted from the smeared images and all of the small connected components are neglected. Remaining components are considered as signature patches. Figure 9.6 shows extracted signatures from the document of Figure 9.3. One of the main advantages of our approach is that it requires very limited number of training samples.

9.4 Evaluation

9.4.1 Dataset

Currently, to the best of the author's knowledge, there are two publicly available datasets that contain information about signature zones, i.e., the Tobbaco-800 dataset [44] and the Maryland Arabic dataset [275]. The Tobbaco-800 dataset contains 1290 images with machine printed and handwritten text in English as well as 900 labeled signatures. The Maryland Arabic dataset contains 169 handwritten images with both English and Arabic text along with 149 labeled signatures.

To generate results comparable to the other approaches, like the one by Zhu et al. [44], evaluation of the proposed method is performed on the Tobbaco-800 dataset. The Tobbaco-800 dataset contains mixed images with machine printed text, signatures, handwritten annotations and logos. However, the ground truth of this dataset only contains information about logos and signatures on the patch level. As mentioned earlier, the document analysis community has recently started considering the problem of signature segmentation and therefore, in the available datasets, currently only the patch level ground truth information about signatures is available and not on the stroke level.

To compare the current method with the recently proposed method by Mandal et al., [46], this work also uses a subset containing only machine printed text and signatures

Table 9.1: Signature Segmentation results on patch level

Method	Precision%	Recall%
Proposed	**56.52**	**100**
Mandal et al.(105 images) [46]	not reported by authors	98.56
Guangyu et al. [44, 45]	not reported by authors	92.8

from the Tobbaco-800 dataset.

9.4.2 Results

To evaluate the performance of the proposed method the precision and recall measures are used. As mentioned in Section 9.4.1, the ground truth contains only patch level information of the signatures. Therefore , precision and recall are also calculated on the patch level. A signature is considered detected if there is at least 40% of overlap between the ground truth and the detected signature patch.

The evaluation results of the proposed method are presented in Table 9.1. This method has recall of 100%, which means that all the signatures are extracted successfully. A minor drawback of this method, however, is that its precision is currently quite low. One reason is that the images containing logos were also considered in the test set (Figure 9.7 for some examples of false positives) and therefore logos are sometimes marked as signature patches. Adding the class “logo” might overcome this problem.

Figure 10.6 shows some of the segmentation results of the proposed method. Qualitatively, the correctly segmented signatures are comparable to manually cropped signatures. As can be seen, the proposed method preforms quite well on a difficult database. More than every second extracted patch is a signature and all signatures are found. This performance is already quite useful for practice as it is often very important to find all signatures. Using simple background knowledge, such as the probable position of a signature, the system might perform even better.

9.5 Open Research Areas and Ongoing Research

As given in Section 9.2, most of the existing systems are evaluated on subsets of Tobacco-800 dataset (to the best of author’s knowledge) as there is no publicly available dataset specifically designed for signature segmentation. Even the existing Tobacco-800 dataset only contains patch level information about signatures. This non-availability of datasets shows that still signature segmentation is not considered by many researchers. Since 2011, some researchers have considered this problem but still there is a lot more to be done.

In order to have good segmentation systems that are integrable with signature verification systems so that to be effectively usable in the real world, it is a must to first have some benchmark datasets on which signature segmentation systems can be evaluated in terms of their precision and recall as well as performance and quality. The author would start developing such a dataset in the future. Along with patch level information, this dataset will have also signature stroke information and would be usable for testing complete signature segmentation and verification frameworks for analysis of documents containing signatures.

In addition to the lack of dataset, another important area is to perform layout free segmentation of signatures, as signatures can be found at different locations in different documents (as shown in Figure 9.2). This means that automatic systems should be capable of finding signatures without using prior information about the layout of the document and the probable location of a signature. Some efforts have already been made in this direction [273, 28, 274].

Another important part is to tune signature verification systems in a way that they can distinguish between genuine and forged signatures even in the presence of some noise in the signature, e.g., touching characters or missing part of a signature. The existing signature verification systems mostly assume that questioned signature image contains no information other than the signature. Accordingly, different features are extracted from the whole questioned signature image and compared with the training samples for authorship attribution. However, Figure 9.2 (d) shows a very common scenario where most of the existing signature verification systems would falsely report the signature as a forgery. It is simply because of the presence of text in the signature image, which is considered as a part of signature during verification process. Therefore, it is required to tune existing signature verification systems so that to make them robust to noise and possible touching components. Use of local features by Malik et al. [62] has already shown promising results in signature verification where verification is performed on the basis of parts of signatures rather than considering the complete structure of signature.

10

Hyper-spectral Imaging for Signature Analysis

Forensic Handwriting Examiners (FHEs) use Multi-spectral ($4 - 20$ color channels) and Hyper-spectral (more than 20 color channels) imaging devices to discriminate different inks, same inks-with different aging, alterations to signatures, and for many other related applications. Inspired from the work of FHEs, this chapter presents a novel automatic method for signature segmentation from hyper-spectral document images (240 spectral bands between $400 - 900$ nm). The proposed method is based on a part-based key point detection technique, which does not use any structural information, but relies only on the hyper-spectral response of the document regardless of ink color. The proposed method is capable of segmenting (overlapping and non-overlapping) signatures from varying backgrounds like, printed text, tables, stamps, and logos, etc. Importantly, the proposed method can extract signature pixels and not just the bounding boxes. This is substantial when signatures are overlapping with text and/or other objects in image. Furthermore, during this work, a novel dataset comprising of 300 documents scanned using a high resolution hyper-spectral scanners has been developed. The chapter also presents the said dataset. Evaluation of the proposed method on this hyper-spectral dataset shows that the proposed method is able to extract signature pixels with the precision and recall of 100% and 79.31%, respectively. To the best of author's knowledge, this is the first time when HSI is used for signature segmentation on pixel level.

10.1 Introduction

Humans are trichromats. According to the trichromatic theory, humans possess three independent visual channels for perception of colors [277]. In accordance, RGB color

[0] This chapter is an adapted version of work carried out and published collaboratively in [276].

space is also defined by the combination of three different colors, i.e., Red, Green, and Blue (RGB) [278]. The reason for aligning the RGB space of cameras, scanners, displays and printers with the trichromatic human vision is that the colors appear realistic to us. Figure 10.1 shows the color spectrum which ranges from ultraviolet to infrared. The human eye can see objects that lie in the range of visible spectrum, everything outside this spectrum is not visible to the human eye. In addition, the range from 400-500 nm is perceived by the first cone and corresponds to the blue channel, 450-630 nm is perceived by the second cone and corresponds to the green channel, and 500-700 nm is perceived by the third cone and corresponds to the red channel in the RGB space [278, 279].

Machines, on the other hand, are not trichromats, i.e., they do not only rely on the 3 channel data. In machines, one can have a finer or coarser representation than the RGB model, depending on the requirements of the target application. Hyper-spectral imaging (HSI) divides the spectra into a large number of bands [280], which results in a very fine and detailed representation compared to the RGB model. In addition, HSI covers a wider range of spectra, starting from ultraviolet, including visible, and ranging to the infrared region. HSI is already being used in different fields like agriculture [281], medical [282], and mineralogy [283]. In the recent past, a few researchers from document image analysis community have started using HSI for different document analysis tasks. It has been mostly used for historical document analysis [284, 285, 286, 287, 288, 289, 290], ink separation [291, 292, 293], and document forgery detection [294, 295, 296] (See Section 10.2).

This chapter demonstrates an application of HSI to segment a very important biometric modality, i.e., handwritten signatures. Signatures are widely used for writer identification and verification. Note that a majority of signature verification and writer identification methods assume that signatures are available pre-segmented (taken out of the document – having no overlap whatsoever with other document contents like text or graphics) [12, 40]. However, in the real world, signatures are usually written on documents like bank checks, invoices, wills, letters, etc., where they overlap with other information/text present in the document. Therefore, to apply signature verification directly on document images, it is first required to segment the signatures from documents.

There exist many methods for signature segmentation [28, 297], which use color and/or structural information. These methods are mostly effective when either the approximate location of signatures is known or signatures are not overlapping with the machine-printed text. However, in many realistic scenarios, signatures are overlapping with text and it is very difficult to separate the pixels of signature from the text components. Figure 10.2 shows an example of no, partial, and complete overlapping signatures.

Existing methods work on RGB, gray scale, or sometime binary images for signature segmentation. This chapter presents a novel method for automatic signature segmentation from document images. This method uses the power of part-based keypoint detection and benefits from the fine and high dimensional representation of hyper-spectral imaging. In particular, the Speeded Up Robust Features (SURF) are used as part-based keypoint detector. A hyper-spectral camera having very high spectral resolution, i.e., 2.1 nm, and covering the wavelengths from visible (400 nm) to infrared (900 nm) regions is used in this work. In addition to the signature segmentation system, an HSI documents dataset consisting of 300 documents is also developed and presented here. These documents include non-overlapping as well as overlapping signatures with text and sometimes logos. Evaluation results show that the proposed system is capable of extracting almost all of the signatures with very high precision, even if the signatures are completely overlapping with the text components.

The remainder of this chapter is organized as follows. First, Section 10.2 provides an overview of the existing methods that use HSI for document image analysis. Section 10.3 introduces the new HSI dataset presented in this work. Section 10.4 provides insight into the novel automatic signature segmentation method. Finally, Section 10.5 describes evaluation protocol and experimental results.

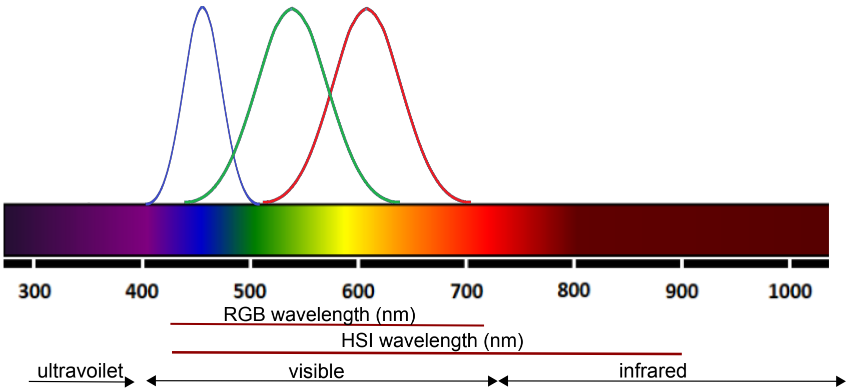
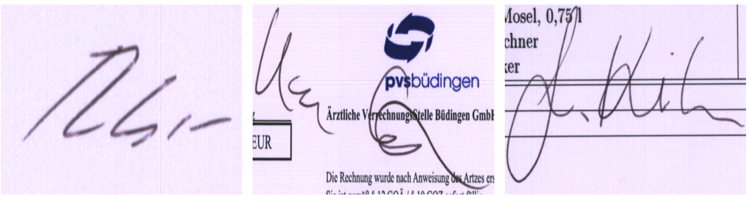


Figure 10.1: The color spectrum.


pvsbüdingen
Ärztliche Verrechnungsstelle Büdingen GmbI
EUR
Die Rechnung wurde nach Anweisung des Artzes ers
Mosel, 0,75 l
chner
ker

(a) (b) (c)

Figure 10.2: Signatures occurrences in documents. (a) no overlap, (b) partial overlap, (c) complete overlap

10.2 Related Work

This section provides an overview of some of the important works reported in reference to the use of HSI for document image analysis. In the document analysis community, HSI is mostly used for analysis of historical documents, ink mismatch detection, and forgery detection.

For historical documents, HSI has been used primarily for text recovery, character segmentation, and overall document enhancement. P. Shiel et al. [284] used HSI to perform quality text recovery, segmentation, and dating of historical documents. They performed segmentation on a 16th century paste-down cover and a multi-ink example typical of which is found in the late medieval administrative texts such as Gottingen's kundige book. Aalderink et al. [285] proposed a method for quantitative analysis of historical documents using hyperspectral imagining. They mapped the distribution of different types of ink and identified the corroded areas within a nineteenth century handwritten letter. In addition, they also proposed a method to enhance the visibility of hidden features like under-drawings on a seventeenth-century historical map. Lettner et al. [286] used spatial (stroke properties) and multi-spectral information in combination with Markov random field model for character segmentation in ancient documents. D. Goltz et al. [287] used HSI for enhancing the assessment of stains on the surface of historical documents. They use hyper-spectral imaging software (ENVI) for quantitatively assessing the extent of staining in two different documents (a treaty and a prayer book). Hollaus et al. [288] presented a method for enhancement of ancient and degraded writings using Fisher Linear Discriminate Analysis (LDA) on multispectral imaging. Hedjam et al. [289] used HSI for restoration of information from historical documents. The degraded information is restored by extracting, cleaning, and combing spectral response of visible and infrared wavelengths. Recently, Saleem et al. [290] also used LDA on multi-spectral imaging for enhancement of ancient and degraded pieces of handwriting which are barely visible by naked eye. Optical Character Recognition is used to evaluate the enhancement method.

For ink mismatch detection in documents, G. Reed et al. [293] have recently shown that HSI is a useful technique for examination of writing inks. They analyzed the spectral responses of red, blue, and black gel inks and achieved discriminative powers of 1.00, 0.90 and 0.40 for red, blue, and black gel inks respectively. Z. Khan et al. [291] considered handwritten notes drafted in various inks. They presented a publicly available dataset of HSI documents which can be used for ink mismatch detection and applied the k-means clustering for detecting different types of inks in the proposed dataset [292].

Furthermore, HSI is used to perform non destructive analysis of documents to detect document forgeries. E. B. Brauns et al. [294] proposed a method for non destructive analysis of potentially fraudulent documents using HSI. They used Fourier transform spectroscopy to achieve spectral discrimination for authentication of written and printed documents. In particular, fuzzy c-means clustering is used to analyze these images. A. Morales et al. [295] proposed a method to detect forgeries in handwritten documents. They used analysis of ink in documents and pen verification using HSI and Least Square SVM classification. The method works for automatic ink type identification, which is tested for 25 different types of pens and achieved an accuracy of 87.5%. C. S. Silva et al. [296] used hyperspectral imaging near infrared range (HSI-NIR) (from 928 - 2524 nm) to detect forgeries in documents. They analyzed three different types of forgeries i.e., obliterating text, adding text, and crossing lines. Principal component analysis (PCA) along with Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) is used for obliteration and adding text problems. To detect crossing lines MCR-ALS and Partial Least Squares Discriminant Analysis (PLS-DA) were used. The identification rate for obliterating text, adding text, and crossing lines is 43%, 82%, and 85% respectively.

10.3 Dataset

Currently, there are two publicly available datasets that contain information about signature zones, i.e., Tobacco-800 dataset [44] and Maryland Arabic dataset [275]. Both of these datasets contain binarized images, i.e., only one channel data are available for these images. A publicly available hyperspectral documents dataset is presented by Z. Khan et al. [292]. This dataset contains handwritten text and can be used only for ink separation. There is no publicly available dataset which contains hyperspectral document images and which can be used for signature segmentation.

This chapter presents a dataset which contains patches from 300 document images,

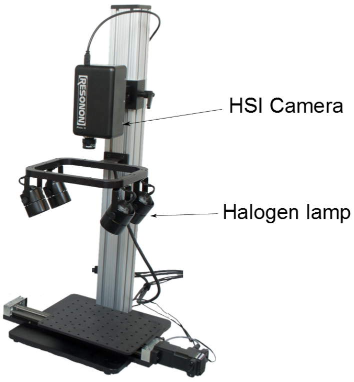


Figure 10.3: HSI Scanning Setup

Parameter	Values
Spectral Range (nm)	400 - 900
Spectral Resolution (nm)	2.1
Spectral Channels	240
Spatial Channels	640
Max Frame Rate (fps)	145
Bit Depth	12

Table 10.1: HSI Camera Specifications

scanned using hyperspectral camera with a very high spectral resolution of 2.1 nm. The documents used contain printed text mostly in black but include colored graphics and logos. Signatures are performed using different type of pens including oil and gel pens, having blue and black inks. The dataset of 300 documents is further split into training and test set. The training set contains 30 documents which are representative samples of the complete dataset. The test set includes the remaining 270 documents. Examples of none, partial, and complete overlapping signatures are available in both training and test sets. As ground-truth, bounding boxes of signatures are marked for each document.

The technical details of the camera used for capturing the HSI data are provided in Table 10.1. The setup of HSI system is shown in Figure 10.3. In addition to a high spectral resolution, this camera also covers complete visible region as well as infrared region till 900 nm. The image scanned using this hyperspectral camera has 240 bands. This means that each pixel has 240 values in contrast to the 3 values resulting in case of RGB scanning. Figure 10.5 shows an example image from the data.

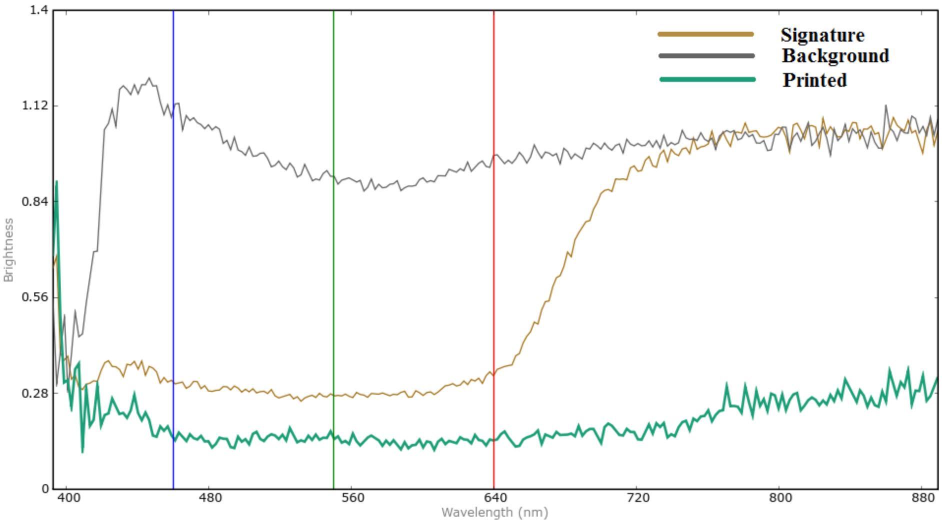


Figure 10.4: Spectral response of page background, machine-printed text, and signature pixels

10.4 Methodology

The proposed method for automatic signature segmentation is based on part-based keypoint detection. In addition, it uses the document's spectral response in order to locate the signatures in the document. As mentioned in Section 10.3, the documents are scanned using a hyperspectral camera having 240 bands ($\lambda_{1...240}$). This results in a representation where each pixel in the document has 240 values.

On inspection of the hyperspectral response of document images, it was noticed that printer ink has a consistent response across almost all of the 240 band. However, the pen inks had a significant variation in their response across the bands, especially in the infrared wavelengths. This can be observed in Figure 10.4 where spectral responses of page background (white), printed text, and signature pixels are shown. This observations serves as the building block for our methodology. Based on this observation, the first step is to locate the two most distinguishing bands out of the total 240 bands, i.e.,

1. The band where all of the objects on the document have non-significant response (everything is visible) including signatures. We refer this band as λ_{max}.

2. The band where all of the objects except signatures have non-significant response (everything except signatures is visible). We refer this band as λ_{min}

The λ_{max} and λ_{min} found using a part based keypoint detector. Part based keypoint detector locates most important points in the document which are referred as keypoints.

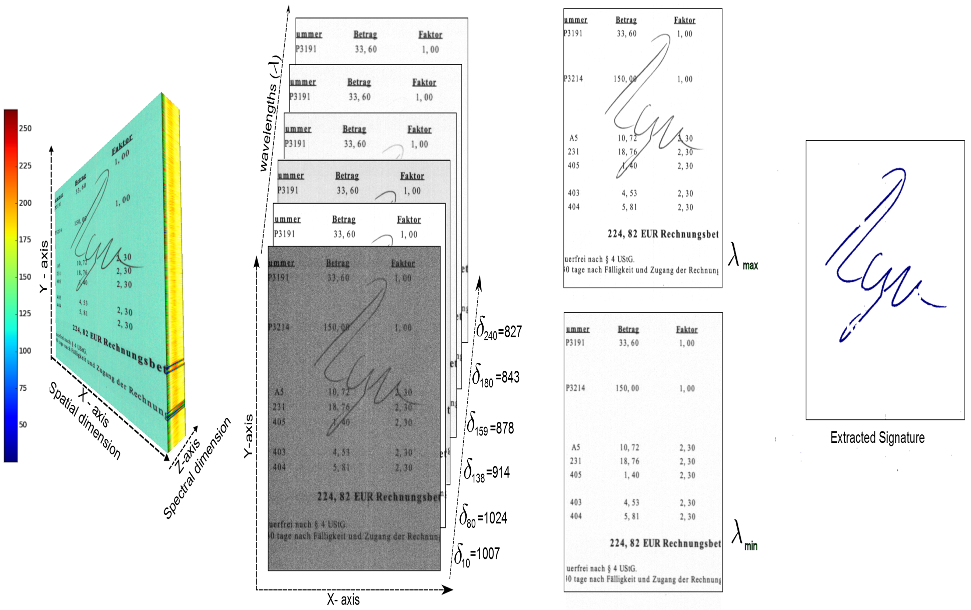


Figure 10.5: Signature Segmentation: An Example Case

The Speeded Up Robust Features (SURF) [298] is used for keypoint detection. However, the overall approach is not limited to SURF and can be used with other key-point detection techniques as well, e.g., SIFT [128], FAST [299], or BRISK [246].

For applying the part based keypoint detector, each HSI document image is treated as 240 grayscale images, each containing the spectral response of the document for the corresponding band. Before applying keypoint detector, noise removal is performed to remove small noise sparks which appear on most of the bands. This is done by applying the averaging filter. After noise removal, SURF keypoints detector is applied on each of the 240 grayscale images generated from the HSI document under consideration. The number of keypoints detected on each image are refereed as δ_n, where n corresponds to the band number. The band with the maximum number of keypoints is refereed as λ_{max} and the band with the minimum number of keypoints is refereed as λ_{min} (see Equation 10.1). Figure 10.5 shows images corresponding to HSI document and the number of keypoints (δ_n) detected on each image.

$$\lambda_{\text{max}\,/\,\text{min}}(n) = \begin{cases} \text{max} = \text{n}, & \text{if } \delta_n = \max(\delta_{1\ldots240}). \\ \text{min} = \text{n}, & \text{if } \delta_n = \min(\delta_{1\ldots240}). \end{cases} \tag{10.1}$$

Once these bands ($\lambda_{\text{max}}, \lambda_{\text{min}}$) are located, the next step is to separate the signature pixels from the remaining text/information. To do so, first, morphological opening is performed on the λ_{max} resulting in noise removal, which is then subtracted from λ_{min}. This results into some noise and signature pixels (λ_{sub}) (see Equation 10.2).

$$\lambda_{\text{sub}} = (\lambda_{\text{max}} \circ \text{mask}_{3\times3}) - \lambda_{\text{min}} \tag{10.2}$$

To remove the noise and get the signature pixels, morphological closing is performed on λ_{sub} and the resulting pixels are used as mask to extract the actual signature pixels from the document. An intersection of signature mask and λ_{max} results in the final signature image (see Equation 10.3). The enclosing rectangle containing the signature mask is refereed as the bounding box of the extracted signature patch. Figure 10.5 shows the complete work flow and the results produced by the proposed method at each step.

$$\text{signature pixels} = (\lambda_{\text{sub}} \bullet \text{mask}_{3\times3}) \cap \lambda_{\text{max}} \tag{10.3}$$

10.5 Evaluation

The standard precision and recall measures are used to report the performance of the system. Precision represented that how relevant the retrieved bounding boxes were, i.e. what percentage out of the retrieved bounding boxes are corresponding to signatures (see Equation 10.4). Recall indicated that out of all signatures bounding boxes which are present in the document how many are part of the retrieved bounding boxes (see Equation 10.5).

$$\text{Precision} = \frac{(\text{Signature BBox}) \cap (\text{Retrieved BBox})}{(\text{Retrieved BBox})} \tag{10.4}$$

$$\text{Recall} = \frac{(\text{Signature BBox}) \cap (\text{Retrieved BBox})}{(\text{Signature BBox})} \tag{10.5}$$

As mentioned in Section 10.3, only patch level ground truth is available. This means that bounding box corresponding to the signatures is provided in every case. The signature is considered detected if there is at least 50% of overlap between the ground truth

Table 10.2: Signature Segmentation results on patch level

Metric	Value%
Precision	100
Recall	79.31

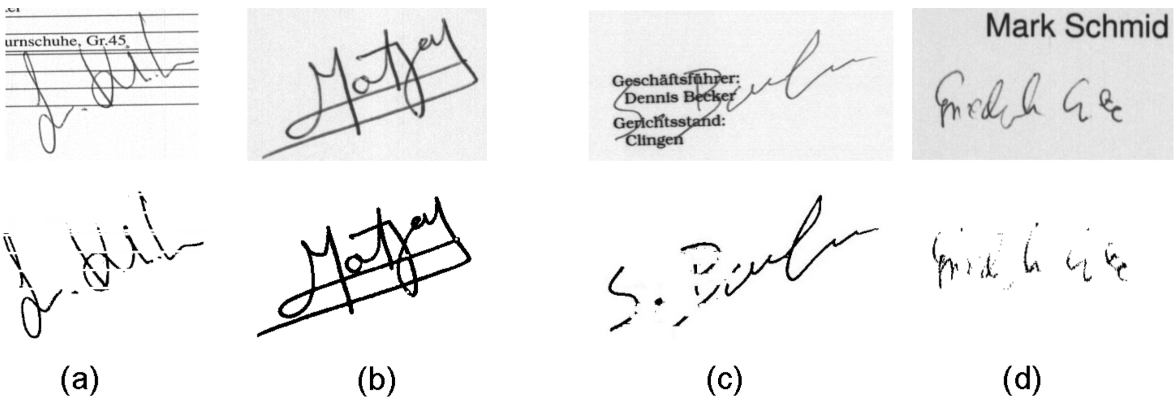


Figure 10.6: Signature Extraction Results: (a,b,c) Successfully Extracted (overlap $> 50\%$) (d) Failure (overlap $< 50\%$)

and the detected signature patch. Table 10.2 shows the evaluation results of the proposed automatic signature segmentation method.

The proposed method has a recall of 79.31% with the precision of 100%, which means that almost all the signatures are extracted successfully with high precision. Figure 10.6 shows some of the segmentation results. Qualitatively, the correctly segmented signatures are comparable to manually cropped signatures. An important highlight of the proposed method in comparison to the state-of-the-art signature segmentation methods is that, it not only provides the zones of signatures in documents but also extracts the signature pixels. This is very important in cases where signature is overlapping with any other information, because, if only signature zone is returned, it will include extra touching components as well. While for real applications, e.g., writer identification and verification, it is required to have only signature pixels and no other touching component. Figure 10.6 (a,c) shows the cases of overlapping signature and extraction results. It can be seen that the method has extracted all of the pixels which belong to signature only. Figure 10.6 (d) shows the failure case, which occurs due to wrong selection of λ_{min}. This can be improved by analyzing multiple bands around λ_{min} rather than using only a single band i.e., λ_{min}. This is an ongoing research and capturing a large dataset of hyper-spectral images and signatures and making this publicly available is planned for the future.

11 Associated Research

This chapter describes some tasks, associated with the core signature verification, that have been performed during the due course of this thesis. Though, this thesis primarily concerns with automatic signature verification and its application in forensic casework where mostly verification of off-line signatures is required. Nonetheless, on-line signatures are very often used in real world scenarios these days, and it is hoped that forensic experts will encounter cases of on-line signature verification in the future [95]. The following sections describe various experiments that have been performed in order to look into the various dimensions and implications of on-line data.

11.1 A Signature Verification Framework for Digital Pen Applications

This section presents a framework for real-time online signature verification scenarios. The main motivation of this work is to take signature verification to the most commonly occurring real world scenarios, particularly in industry, where signature verification is required. One of the important markets where signature verification is highly demanded is financial institutions. This section describes the application the proposed signature verification framework in different real world scenarios in connection with the Anoto digital pen.

Figure 11.1 illustrates the hardware layout of the Anoto digital pen. This pen specializes in providing the look and feel of regular pens. It only demands to add Anoto dot pattern to any paper and data can be digitized seamlessly. The Anoto pattern makes

[0] This chapter is an adapted version of results obtained from different collaborative research pursuits published in [230, 300, 301].

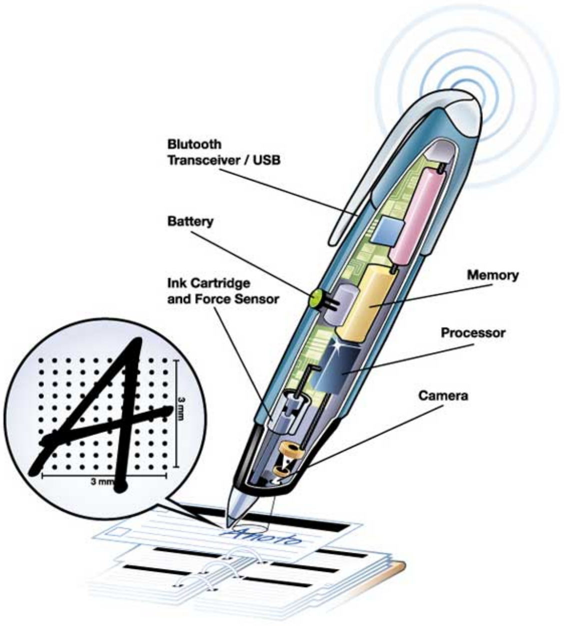


Figure 11.1: Anoto digital pen.

it possible for the Anoto pen's built-in camera to detect strokes and record signatures that then can be stored in an internal memory or sent via communication unit using Bluetooth/USB. Due to this ease of use, Anoto pens are finding applications in fields from health care to finance. The proposed signature verification framework is an attempt to take signature verification to every area where the Anoto pen finds an application. In particular, it has already been applied in test scenarios for financial institutions and product manufacturing companies.

11.1.1 Signature Verification Framework Overview

The general overview of the signature verification framework is illustrated in Figure 11.2. The online data are collected using the Anoto digital pen and saved in the pen's memory. The pen is then synchronized with some processing device like a computer or a mobile phone. Through this synchronization, data are sent to the Anoto Software Development Kit (SDK). Once the data are received at the SDK, the proposed framework picks the corresponding signatures data (questioned or referenced) and passes it to the signature verification module. The signature verification module then uses a Gaussian Mixture Model (GMM)-based approach to process the signature.

There can be two situations in the framework. In the first situation the user is

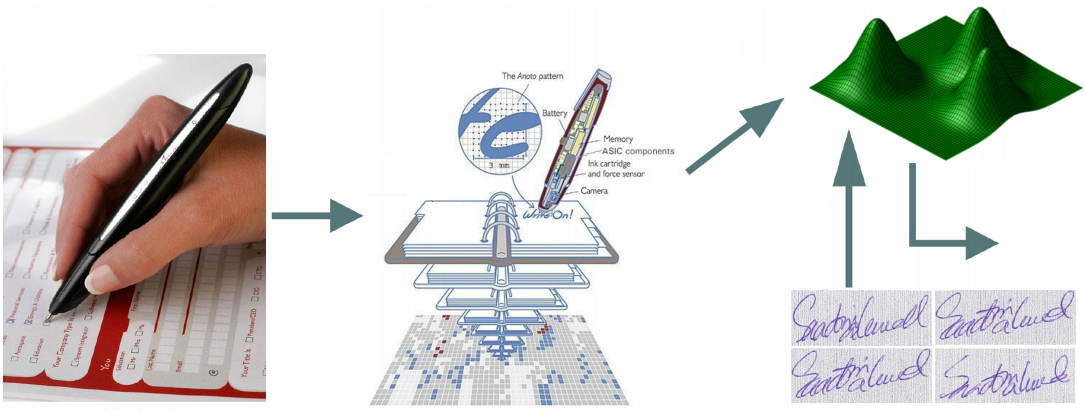


Figure 11.2: General overview of the proposed signature verification framework.

interacting with the framework for the first time. In this case (s)he has to provide her/his genuine signatures as the reference signatures and prove the identity by any other traditional secure way. Now, the framework generates reference GMMs for the user and stores them. In this way registration of a user is completed.

In the second situation a user interacts with the framework by providing her/his signatures and claiming to be some specific person. Now the framework takes the claimed person's GMMs (assuming that this person is already registered with the framework) and fits the questioned person/signature model. Currently, the system reports it's evaluation result in the form of probability values. Based on this value it can be decided whether the claiming person is the authentic writer or a forger.

11.1.2 Signature Verification Module

The signature verification system used for the framework is an adapted version of a previous system introduced by Liwicki et al. [302]. The basic details are given here only for completion, please refer to [302] for a complete description of the system. Given the online data as an input, the signatures are corrected with respect to their skew and then the following features are extracted; the pen-up/pen-down feature *(1)*; the pressure *(2)*; the speed *(3)*; the speed in x and y direction *(4,5)*; the acceleration *(6)*; the acceleration in x and y direction *(7,8)*; the log radius of curvature *(9)*; the normalized x- and y-coordinate *(10,11)*; the writing direction *(12,13)*; the curvature *(14,15)*; the vicinity aspect *(16)*; the vicinity slope *(17,18)*; the vicinity curliness *(19)*; the vicinity linearity *(20)*; the ascenders and descenders in the off-line vicinity of the considered point *(21,22)*; and the context map, where the two-dimensional vicinity of the point is transformed to a 3×3 map and the resulting nine values are taken as features *(23-31)*.

Gaussian Mixture Models (GMM) have been used to model the signatures of each person. More specifically, the distribution of feature vectors extracted from a person's handwriting is modeled by a Gaussian mixture density. For a D-dimensional feature vector denoted as x, the mixture density for a given writer (with the corresponding model A) is defined as:

$$p(x\|A) = \sum_{i=1}^{m} w_i p_i(x)$$

In other words, the density is a weighted linear combination of M uni-modal Gaussian densities, $p_i(x)$, each parametrized by a $D \times 1$ mean vector, and $D * D$ covariance matrix. For further details, please refer to [303].

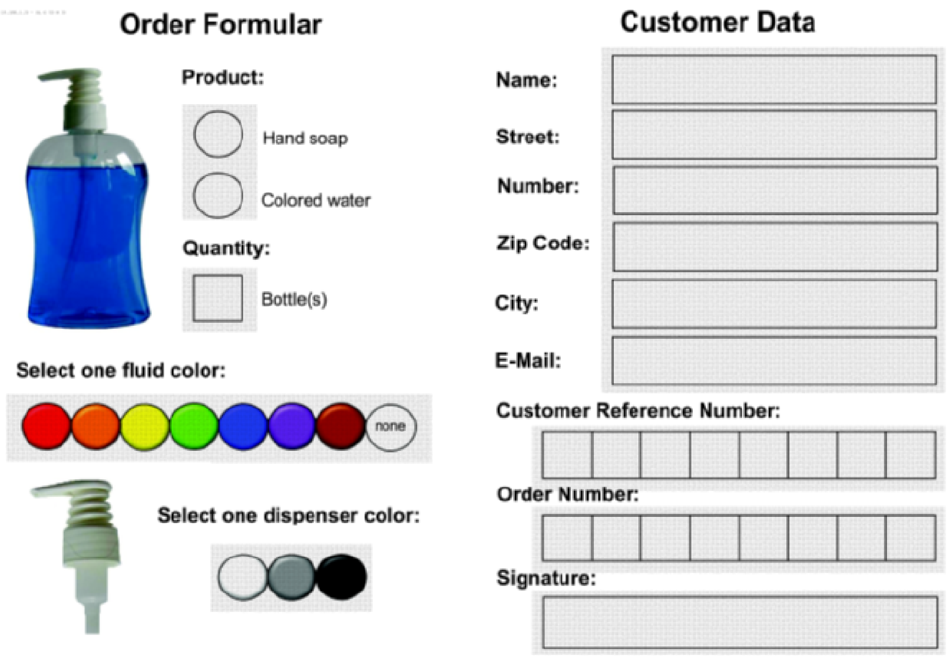
11.1.3 Application Scenarios

Automatic Order Processing

Highly customized products having shorter development life cycles is the demand of today's global market [304]. This makes efficient order processing an important area for any manufacturing company to improve. In traditional order processing, a client fills an order form and then posts/faxes it to the company. On receiving the order, the company follows its predefined procedure to establish authenticity of the order. One important modality in this process is using the signatures of the client and keeping them in the company's record. This is a time consuming process. Alternatively, web based forms can be used. However, web based order processing suffer from a compendium of difficulties for customer as explained by Doyle et al. [305].

To cope with this an approach for intelligent digital pen-based ordering has been recently introduced Koessling et al. [306]. Here, instead of traditional paper or Internet, a digital pen is used to take customer specific orders that are afterward used in production automation. The customer fills this form as a regular form and signs it. The pen is synchronized (attached to a computer via USB or Bluetooth) and the corresponding electronic form is mapped to the writing information. The final order is then sent to the *SmartFactory*^{KL} which allows for product automation (further details are provided in [306]).

A pen based interaction order form is shown in Figure 11.3. Here a user may select a product with different colors and enter her/his particulars using the Anoto pen. Note the signature field is also provided as it is on the traditional forms but now it is dealt with the proposed signature verification framework. The paper used here is exactly the



Order Formular

Product:

○ Hand soap

○ Colored water

Quantity:

□ Bottle(s)

Select one fluid color:

● ● ● ● ● ● ● ○ none

Select one dispenser color:

○ ● ●

Customer Data

Name: □

Street: □

Number: □

Zip Code: □

City: □

E-Mail: □

Customer Reference Number: □□□□□□□□

Order Number: □□□□□□□□

Signature: □

Figure 11.3: A pen based interaction order form.

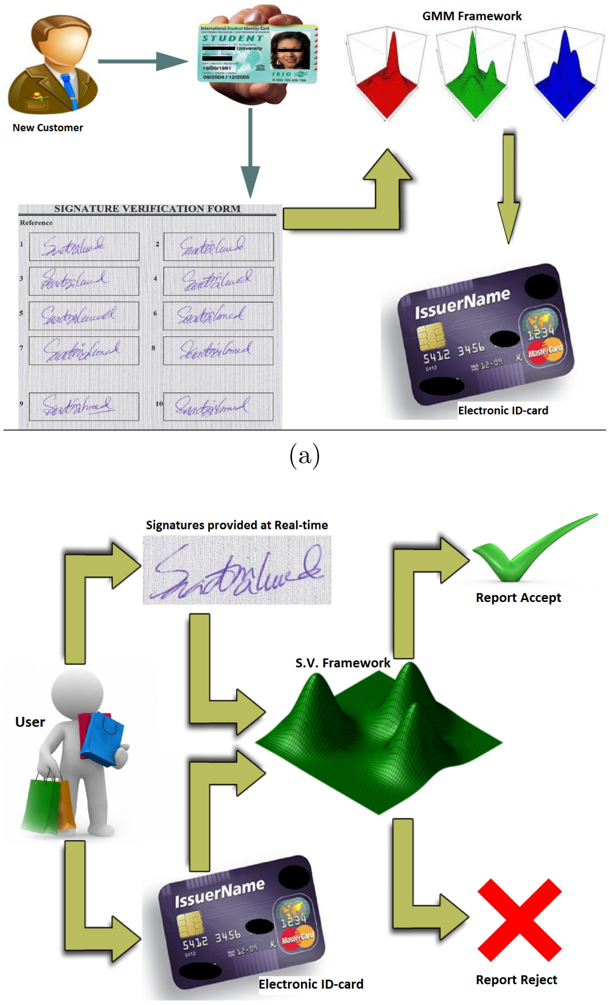
same as it was in the traditional approach except that now it also carries the Anoto dot pattern.

The final accept or reject of an order would now depend on the result of the signature verification module. If sufficient likelihood for authenticity can be established, the customized order is sent to the *SmartFactory*^{KL} and the customized production process is started. Otherwise, the order is rejected.

Signature Verification in Banks

Financial institutions bear substantial losses on account of insufficient signature verification mechanisms, where often, a visual comparison is performed to authenticate signatures. This section suggests the idea of integrating the GMM descriptions produced by the proposed framework into electronic ID-cards so that to help banks increase their immunity against fake credit or debit card users.

Two example application scenarios for the framework are illustrated in Figures 11.4(a) and 11.4(b), respectively. Figure 11.4(a) illustrates a scenario where a customer registers at a bank for the first time. At first the customer applies for opening an account by providing her/his identity. This may be done by any card containing picture and particulars of the applicant like a passport or personal ID-card. For opening an account



(a)

(b)

Figure 11.4: Application scenarios of the proposed framework: (a) registering a customer and generating an electronic ID-card; (b) establishing the authenticity of a customer.

the bank takes a certain number of online signatures from the customer and provides them to the framework proposed here. The framework is then applied to take GMM descriptions using various combinations of these genuine reference signatures. Note that already during this process a validation can be performed that all signatures are similar enough to produce a probability of being authentic. The obtained GMM descriptions are then stored on the electronic card of the customer. By doing this, the original signatures would not be available on the card (protecting the customer against fraud). Only the model used for comparison will be available. Furthermore, one may not generate the genuine signatures from the stored model thereby removing the danger of any security attack of such kind.

Now the customer, whenever using the card for monetary transaction may use his/her signatures instead of or additionally to pin codes/logins. Figure 11.4(b) illustrates this scenario. Here a customer has an electronic card (developed in the previous scenario) after purchasing goods at a supermarket tries to pay with this card. At the counter (s)he provides her/his electronic card along with signatures written with a digital pen. These signatures are transferred to a local computer having an instance of the framework where the authenticity of these signatures is judged. If the framework then reports an accept it refers that the customer is authentic. Otherwise, the customer might be a forger/imposter trying to use some other person's card (in which case, another authentication method can be applied). There can also be various other applications of these *behavioral information containing electronic cards* which are beyond the scope of this discussion.

11.1.4 Datasets and Evaluation

The evaluation of the framework was performed on two data sets. The first data set contained the data collected specifically using the Anoto digital pens on forms having Anoto dot pattern. For this collection, ten authors (male and female) from different countries aging between 18 to 40 provided their genuine signatures. Ten forgers (students, researchers, and a calligrapher) were asked to make skilled forgeries of each genuine author. Each genuine author contributed 9 of her/his signatures. Out of these nine, 7 signatures were used as reference signatures and remaining 2 were put in the test set for every genuine author. Each forger also produced 9 forgeries. As a whole, the test set for this data set contained 20 genuine signatures and 90 skilled forgeries.

The second data set was the NISDCC signature collection of the ICDAR 2009 online signature verification competition (SigComp)[1]. This data set consists of 60 authentic signatures written by 12 authors, 31 forgers produced skilled forgeries at the rate of 5 forgeries per genuine author.

Since the number of signatures in the first data set is quite low, the results are reported in terms of number of Falsely Accepted signatures (FA) and number of Falsely Rejected signatures (FR). The results are given in Table 11.1. The False Accept Rate (FAR) and False Reject Rate (FRR) with the Equal Error Rate (EER) for the second data set are however reported. These results are shown in Table 11.2. The evaluation results indicate initial success that is also triggering the interest of industry (financial institutions) in this area.

[1] publicly available for research purposes at http://sigcomp09.arsforensica.org

Table 11.1: Evaluation results of Anoto online data (data set 1)

Author-ID	FA	FR
1	0	0
2	0	0
3	2	1
4	0	1
5	0	1
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0

Table 11.2: Evaluation results of ICDAR 2009 online data (data set 2)

Overall (for all authors)	FAR	FRR	EER
GMM Framework	0.9	4.0	3.3

11.2 Online Signature Analysis

The motivation for this study is to compare the expedience of two easy to use and cheap devices for online signature verification and to use the state-of-the-art feature extraction. Both devices are ballpoint pens with a sensor attached. One of them measures pen acceleration with an accelerometer, while the other sensor is a gyroscope that measures pen angular momentum. This section shows how Legendre approximation and SVM classifier perform on the two databases collected with the above mentioned pen and compares their performances.

11.2.1 Hardware

Two ball point pens were used with different types of attached sensors to measure different dynamic features of the handwriting during the signing process. First a ballpoint pen with a tri-axial accelerometer attached close to the tip of the pen (see Figure 11.5) was used. The LIS352AX accelerometer chip was used because of its signal range, high accuracy, impressively low noise, and ease of use. The maximal possible sample rate is 1000 Hz. The properties of the equipment are detailed in [307]. Note that originally the equipment was developed for educational purposes as described in [308] and the idea here is to apply it for online signature verification.

For comparison, the accelerometer was replaced with a 2-axis gyroscope to measure the angular momentum of the pen during signing. A low-power and dual-axis LPR530AL gyroscope sensor was used which measures angular velocity along x and y axis. This sensor provides high resolution, has full scale of ±300°/s, and capable of detecting rates up to 140 Hz. Due to the size of this circuit board, the gyroscope could not be attached to the tip of the pen (see Figure 11.6).

11.2.2 Databases

Acceleration

The complete description of the AccSigDb2011 database can be found in [307]. It was recorded between January and March of 2011 and contains 600 signatures from 40 authors. Each of them contributed 10 genuine signature and 5 forged signature of a selected author in the following way; the contributors could choose which author/signature they are willing to forge and after allowing them practice, the forgeries were recorded. Afterward the database was extended, see the details in [309], and the extended database contains

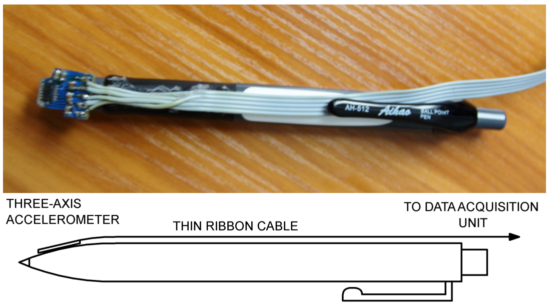


Figure 11.5: The accelerometer is mounted close to the tip of the pen

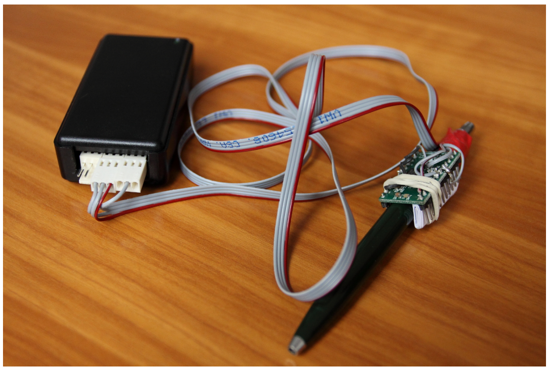


Figure 11.6: The gyroscope is mounted to the pen

300 additional signatures. 20 authors who contributed to the first version of the database were requested to repeat the same process (10 genuine, 5 forged signature per person) between April and May of 2011. This extension provided an opportunity to examine the similarities between signatures from the same author captured in two recording periods. Figure 11.7 shows the reduced signals of four signatures which belong to the same author.

Angular momentum

The similar recording process was repeated as to the previous one, however now gyroscope was attached to the ballpoint pen instead of the accelerometer. The sample rate during the signature recording was 100 Hz. This database is referred as GyroSigDb2012. 21 authors contributed to the GyroSigDb2012, each of them contributed ten signatures except one who gave 50 signature samples. Figure 11.8 shows two signatures from the same author. It shows the output voltage of the gyroscope directly. Each row belongs to one signature, the first column (left) shows the signal along the x-axis, the second (right) shows the signal along the y-axis.

Overlap

In this study only the overlapping part of the two databases is used for comparison as some authors contributed to both the databases. This overlap represents 10 authors and 300 signatures: 20 signatures per author from the AccSigDb and 10 signatures per authors from the GyroSigDb.

11.2.3 Comparison

Legendre approximation for feature extraction

Several feature extraction methods have been used in signature verification in the last few decades. Verification methods are based on local or global features. Local features are calculated for each point of the time sequence, global features characterize the whole signature. If global features are used, the length of the feature vectors are fixed, regardless of the length of the signature writing process. Since the same author does not sign in the same amount of time, thus if one can represent properly each signature globally with a fixed length of vector, it can reduce the differences between the signature of the same author. Moreover, the signatures with fixed length-representation are easier to compare and they are required for some biometric applications [310, 311].

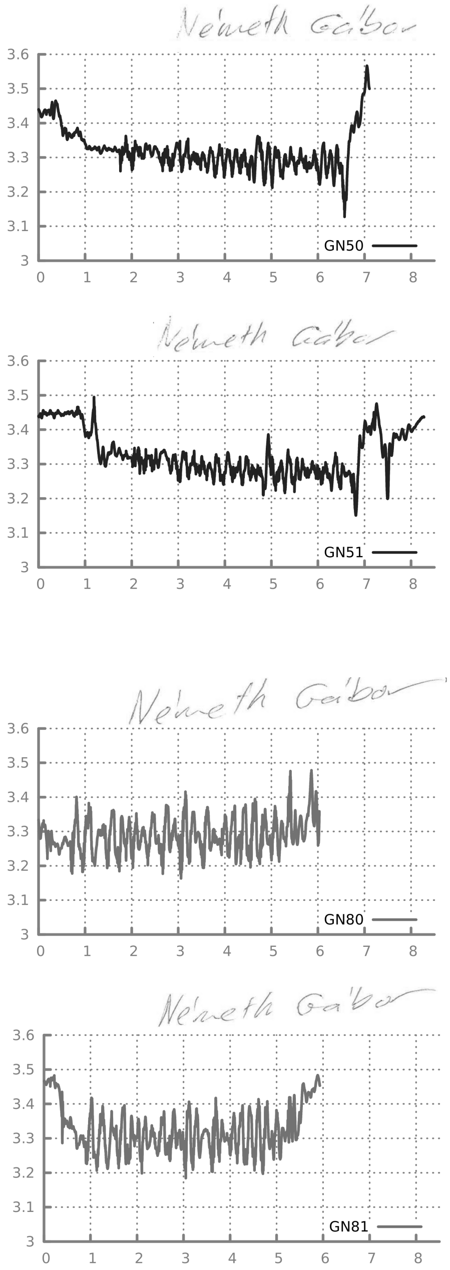


Figure 11.7: Four genuine signatures from NG (AccSig2011)

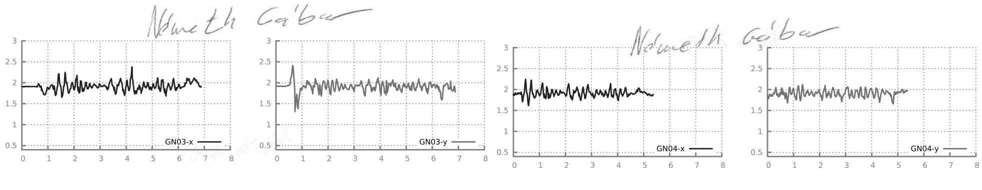


Figure 11.8: Two genuine signatures from NG (GyroSig2012, left/right: x/y axis)

N	1	3	5	10	11	13	15	17	18	19	20	23	25	30	35	40
AccSig	80.11	50.00	87.67	87.44	88.44	88.00	87.22	86.56	85.00	84.44	85.11	86.33	84.00	83.78	83.00	80.56
GyroSig	59.11	64.89	74.22	77.33	80.44	80.44	80.22	78.44	78.67	78.22	80.00	78.44	79.56	78.22	74.22	71.33

Table 11.3: Pairwise comparison of average accuracy on the two databases depending on the order (N) of Legendre series

N	1	3	5	10	11	13	15	17	18	19	20	23	25	30	35	40
AccSig	38.00	10.00	37.00	33.00	39.00	43.00	36.00	36.00	34.00	33.00	36.00	35.00	37.00	36.00	34.00	32.00
GyroSig	8.00	30.00	36.00	54.00	46.00	52.00	44.00	50.00	52.00	48.00	46.00	40.00	46.00	40.00	28.00	30.00

Table 11.4: Comparison of accuracy on the two databases depending on the order (N) of Legendre series

The order of the Legendre polynomial was varied from $N = 1$ to $N = 40$ in order to examine which order performs the best. Every signature in the AccSigDb2011 can be represented with a feature vector of length N. Signatures in the GyroSigDb2012 were represented similar way as in [312]. Thus for each signature the Legendre polynomial with order N for axes x and y, respectively was computed, and concatenated the coefficient vectors of the polynomials, so the feature vectors in these case have length of 2 N.

SVM classifier for classification

For classification the Support Vector Machine (SVM) classifier was used with radial basis function (RBF). Half of the signatures were used for training, and half for testing. The accuracy of the classification can be seen in Table 11.3. Beside this, the SVM classifier was used with every signature of the 10 authors mentioned above to see how SVM can separate the signatures of the 10 authors from each other into ten classes. The accuracy of the classification can be seen in Table 11.4 (i.e. the percent of correctly classified signatures).

11.2.4 Results

Table 11.3 and 11.4 show the values of the accuracies depending on the order of the Legendre polynomials. We varied the order N from 1 to 40.

Considering the binary classification of the signatures, the average accuracy is around 85%-89% in AccSigDb2011 and weakly dependent on the order of the polynomials. Contrarily, in GyroSigDb2012 the accuracy is always less than in AccSigDb2011 and it is decreasing if the order of the polynomials are greater than 15 (see Figure 11.9). If the order of the polynomials are greater than 30, the accuracy decreases significantly on the

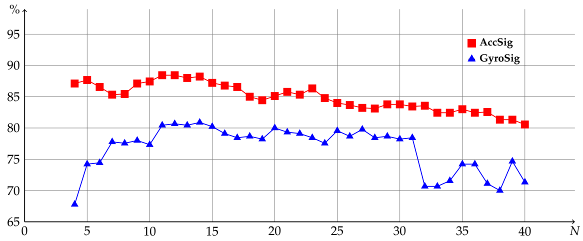


Figure 11.9: Accuracy of the pairwise binary classification. X-axis: Order of Polynomials, Y-axis: accuracy.

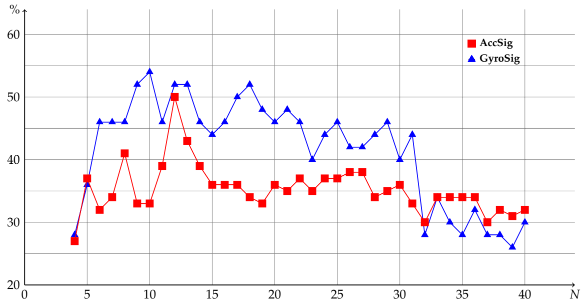


Figure 11.10: Accuracy of the multiclass classification. X-axis: Order of Polynomials, Y-axis: accuracy.

GyroSigDb2012. While using multi-class classification the accuracy is much lower than in previous case since we have more classes. In this case the proposed method results higher accuracy for GyroSigDb2012 compared to AccSigDb2011. It is around 50% if the order of the Legendre polynomials is less than 20, while it is just around $35 - 40\%$ on the AccSigDb2011 (see Figure 11.10).

These preliminary results show the potential of the two sensors for signature verification, however more experiments on larger data are planned to solidify the initial findings.

11.3 Online Verification with Kolmogorov-Smirnov Distribution Distance

This section describes an online signature verification approach based on the Kolmogorov-Smirnov distribution metric. Different feature distributions are compared in order to perform classification and distinguish between forged and genuine signatures.

11.3.1 Approach

The *SigComp2011* database (Dutch online data) [202] was used for evaluation of the proposed method. The signatures were collected by a WACOM Intuos3 A3 Wide USB Pen Tablet with sampling rate of 200Hz, resolution 2000 lines/cm and precision of 0.25mm. In the SigComp2011 dataset, the (x_t, y_t, p_t) values are captured as x and y coordinates of the signature, and pressure p at a given time t.

Preprocessing

The WACOM tablet is a large tablet with an active area of 488mm×305mm and it is possible for two signatures from the same person to be written on any part of the active area. Therefore the x, y coordinate values were shifted to the origin, so for each signatures S the minimal x and y values were subtracted from the original coordinates. Thus,

$$\begin{aligned} x'_t &= x_t - \min_k (x_k) \quad (t = 1, \ldots, T_S) \\ y'_t &= y_t - \min_k (y_k) \quad (t = 1, \ldots, T_S) \end{aligned}$$

where T_S is the number of sampled values of signature S.

Feature extraction

In addition to the above-mentioned features (i.e., shifted x, y, and p), the v velocity values based on the difference between the coordinates were also calculated. The difference between the coordinates is calculated simply using the difference formulas;

$$\begin{aligned} \Delta x_t &= x_{t+1} - x_t \quad (t = 1, \ldots, T_S - 1) \\ \Delta y_t &= y_{t+1} - y_t \quad (t = 1, \ldots, T_S - 1). \end{aligned}$$

and the velocity values are calculated in the following way:

$$v_t = \sqrt{\Delta x_t^2 + \Delta y_t^2} = \sqrt{\left(\frac{x_{t+1} - x_t}{\Delta t}\right)^2 + \left(\frac{y_{t+1} - y_t}{\Delta t}\right)^2}.$$

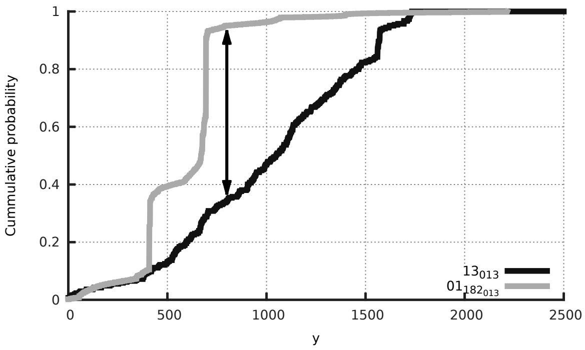


Figure 11.11: Plots of two empirical distribution functions, the black arrow denoting the two-sample KS distance

Classification

The classification is based in the distribution of different features, especially pressure and velocity values of the signatures.

In order to compare two distribution functions, the Kolmogorov-Smirnov test was used which calculates the maximal difference between the cumulative distribution functions. If $F_x = f(X < x)$ and $G_x = g(X < x)$ are the two distributions, then the Kolmogorov-Smirnov distance (KS-distance) between the two distribution functions is,

$$D(f,g) = \sup_{x} \textsf{F\_x-G\_x} \tag{11.1}$$

Figure 11.11 shows two empirical distribution functions and the arrow indicates the KS-distance of these distributions.

During the verification process the KS-distance was calculated pairwise between each reference signature for the same author and the questioned signature and the references for all the features. Dutch evaluation data set contains 12 reference and 24 questioned signatures for each author. Altogether there are 36 signatures per person. The KS-distance is always in the $[0, 1]$ interval so the distance matrices can be represented as a greyscale image. Examples are shown in Figure 11.12. The figure depicts the KS-distance values belonging to the author denoted by id 022. Each column represents different features (i.e., x, y, p, v) and contains three sub-figures. The first row shows the distance matrix which shows the distance between the reference signatures. The diagonal is black (whose color is assigned to zero) because there is no difference between identical distributions (the KS-distance between identical signatures is zero). The second row shows the distance values between the questioned forged signatures. Each row in the

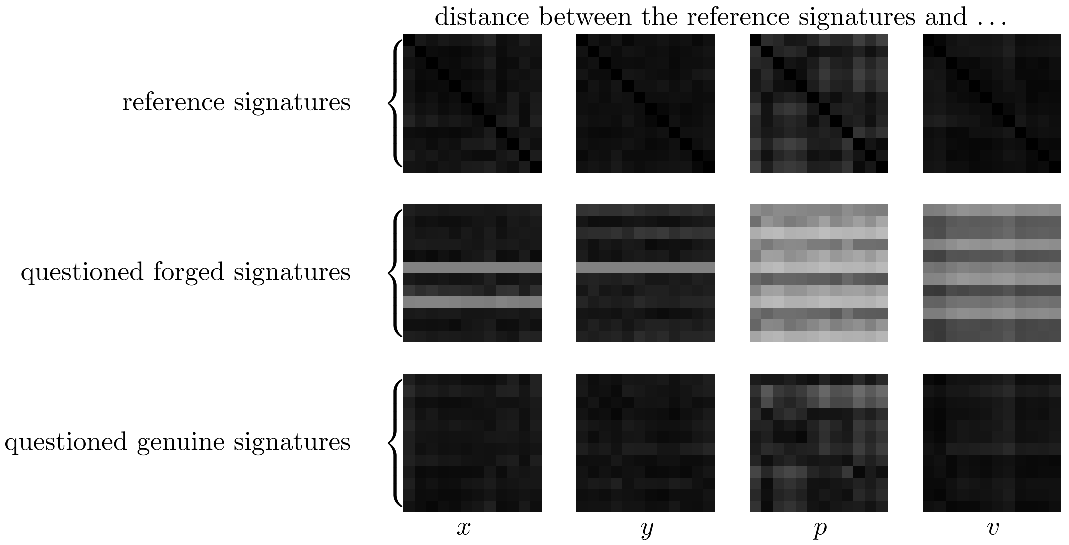


Figure 11.12: KS-distance matrix of the coordinates, pressure and velocity features of author 022 from the evaluation set of the SigComp2011 dataset

image contains 12 pixels which represent the 12 distances between the corresponding forged, questioned, and the reference signatures. The third row shows the distances between questioned genuine and reference signatures in the same way as the forged ones: Each row in the image represents the distances between the corresponding questioned genuine signature and the reference signatures.

Training

Based on the reference signatures of a particular author, the reference distance was calculated based on each author. First, the KS-distance between each pairs of reference signatures was calculated for an arbitrary feature and the reference distance was determined by taking the average, maximum, or minimum of these distances. During the testing phase, this reference distance was used for comparison purposes.

11.3.2 Algorithm

For each questioned signature the KS-distance from each corresponding reference signature was calculated and the KS-distance of the given questioned signature was defined as the average, maximum, or minimum of these KS-distances.

Algorithm 1 shows the verification steps. From the second to seventh line, the distances for the reference set are calculated, while lines 8-10 calculate the distances for the corresponding questioned signature. On line 11 the decision is based on the duration of the signature. On line 14, the main constraint for the (minimal, maximal, or average) distance of the questioned signature from the references is taken into account.

Another constraint impact was tested too, i.e., a signature duration constraint. The duration of the signature writing process being tested first. If the questioned signature duration was greater than 1.5 times the maximal duration of the reference signatures, and the same hold for the pen down duration (when $p_i > 0$), the questioned signature was rejected.

First the constraints of the different features were taken too into account separately (for x, y, p and v and for maximum, minimum, and average distance it means $4 \cdot 3 = 12$ different cases, if the time duration and maximal reference constraints are taken into account it makes $4 \cdot 12 = 48$ cases). For example, if the minimal KS-distance of the pressure values fulfilled the constraint in the line 14, hence if $\min / \text{average}(\text{dist}, p) < c \cdot \min(\text{distRef}, p)$ was true for the questioned signature, it was accepted as a genuine signature (certainly the duration constraint was take into account), otherwise it was rejected. The c threshold was varied from 0 to 7.

In the next step these constraints were combined in several ways: The constraints regarding to the x and y coordinates were combined with `AND` ($\wedge$) and `OR` ($\vee$), the p pressure and v velocity the same way. The pressure was combined with the x, y coordinates as well, as shown below.

11.3.3 Results

Table 11.6 lists the false acceptance and false rejection rates (FAR/FRR) on the Dutch dataset of *SigComp2011* based on the proposed method. The rows show which feature (x, y coordinates, pressure p or velocity v) and which distance was used (average, maximum, minimum). The columns show two scenarios. In scenario I the duration constraint was used (line 11 in Algorithm 1), while in scenario II this constraint was not applied. In addition the two scenarios (with and without time constraints) were also tested: in those the reference signature which differed the most from the other reference signatures were excluded, so they were not used during the decision process. These experiments negligibly worsened both the accuracy and equal error rates by approximately $0.5 - 1.0\%$ (not shown here).

For comparison purposes, the Dynamic Time Warping (DTW) distance for pressure and velocity features is included in Table 11.6. Moreover, Table 11.5 shows the results of the competing systems submitted to the SigComp2011 [202]. Values in bold indicate

Algorithm 1: Verification method

```
Input  : feature, references and questioned signature, threshold $c$
Output: genuine or forgery (decision)
k=1
for $i$=1 to NoReferences do
    for $j$=($i$+1) to NoReferences do
        distRef$_k$ = KS(reference$_i$, reference$_j$, feature)
        $k = k + 1$
    end
end
for $i$=1 to NoReferences do
    dist$_i$ = KS(reference$_i$, questioned, feature)
end
if  testDurationBool $\wedge$ duration(questioned) $> 1.5 \max$(duration(references))
then
    return forgery
else
    if max / min /avg(dist) $< c \cdot$ max / min /avg(distRef) then
        return genuine
    else
        return forgery
    end
end
```

the lowest FAR/FRR rates in that section of the table. Also, the different combination of constraints are compared when the same type of statistic is used for comparison (maximum, minimum, and average distances are compared both on the reference signatures and between the references and each questioned). Values in italics indicate the lowest FAR/FRR rates when only single features are considered.

It is evident that the classification scheme based on the KS-distance performed better than the others based on DTW have a performance comparable to the systems submitted to the *SigComp2011* competition. It is observed that if the time constraint is applied (column with label “with time constraint”), the equal error rates are roughly less than 13%. If the time constraint is not applied, the error rates increases by $10 - 20\%$.

For almost all the constraints, the best results were achieved with a minimum distance. In this case the reference distance was the minimum distance between the reference signatures and it was compared with the minimum distance of the questioned signature and the reference signatures. If we compare the performance of the different single features, we see the smallest error rates appear in the rows which belongs to the pressure feature .

ID	FAR	FRR
4	3.76	3.70
5	3.44	3.86
7	6.87	7.25
1	7.69	7.56
KS	**7.86**	**8.02**
6	8.02	8.33
9	11.27	11.11

Table 11.5: Results (FAR and FRR values) of SigComp2011 Dutch online competition

The minimum KS-distance gives good results when the coordinate constraints and pressure are used together (row $x \wedge y \wedge p$) otherwise (in the case of average and maximum) the constraint for pressure and velocity together (row $p \wedge v$) gives the lowest equal error rates. Besides the minimum distance, the time constraint improved the results significantly.

		With time constraint	**Without time constraint**
x	average	12.27%/12.19%	29.62%/29.01%
y		11.13%/11.27%	30.28%/30.56%
p		*9.66%/9.57%*	19.97%/19.91%
v		11.13%/11.11%	27.66%/27.78%
$x \wedge y$		10.47%/10.34%	26.19%/25.15%
$x \vee y$		12.44%/12.50%	29.79%/29.01%
$p \wedge v$		**7.86%/8.02%**	15.55%/15.74%
$p \vee v$		10.31%/10.19%	26.35%/26.54%
$(x \wedge y) \wedge p$		8.02%/7.72%	16.04%/16.05%
$(x \wedge y) \vee p$		10.80%/10.96%	21.93%/22.22%
x	max	11.95%/12.04%	31.75%/33.95%
y		11.29%/11.42%	32.90%/30.71%
p		*9.66%/9.88%*	23.73%/23.61%
v		10.97%/10.49%	26.35%/27.31%
$x \wedge y$		10.64%/10.34%	30.93%/31.02%
$x \vee y$		10.31%/11.73%	32.24%/33.64%
$p \wedge v$		**7.86%/8.02%**	16.69%/16.67%
$p \vee v$		11.46%/11.73%	27.00%/26.85%
$(x \wedge y) \wedge p$		8.84%/9.26%	19.97%/20.06%
$(x \wedge y) \vee p$		12.44%/11.73%	25.37%/26.23%
x	min	11.78%/11.88%	27.82%/27.62%
y		10.97%/10.96%	27.99%/28.55%
p		*9.00%/8.95%*	20.13%/20.22%
v		4.42%/50.46%	21.11%/50.46%
$x \wedge y$		9.82%/9.88%	22.91%/23.30%
$x \vee y$		11.29%/11.27%	26.35%/25.93%
$p \wedge v$		2.13%/52.01%	10.64%/52.01%
$p \vee v$		9.49%/9.41%	19.31%/19.44%
$(x \wedge y) \wedge p$		**8.67%/8.64%**	17.02%/17.13%
$(x \wedge y) \vee p$		10.47%/10.49%	20.13%/19.75%
DTW distance			
p	average	13.75%/13.89%	38.13%/38.89%
v		14.08%/14.04%	33.06%/33.18%
p	max	13.91%/14.04%	30.44%/30.86%
v		14.57%/14.66%	29.95%/30.09%
p	min	8.67%/62.19%	49.75%/62.04%
v		6.87%/74.54%	25.86%/74.54%

Table 11.6: FAR/FRR values

12

Conclusion and Future Work

This chapter summarizes the major conclusions that could be drawn from the issues considered in this thesis and the solutions presented to tackle them. Any limitations of these solutions are discussed along with the possible research that would overcome these limitations in the future.

12.1 Conclusion

This thesis is aimed at bridging the gap between the state-of-the-art automatic methods of signature verification and their application in forensic science. In the contemporary forensic environment, signature authentication/verification is still universally carried out by humans. However outside of forensic science, computer based signature verification techniques continue to develop in response to commercial needs around quickly recording and authenticating an individual's mark. Nonetheless, computer based techniques are attracting increasing interest within the forensic science community. However, forensic scientists – particularly Forensic Handwriting Examiners (FHEs) – see various problems with automatic systems which are hindering the actual application of such systems in forensic casework. These problems are primarily related to differences in the way signature verification is taken by pattern recognition (PR) researchers and FHEs. The author foresees the two communities, PR and FHEs, move apart with the day, and considers this a high time to bridge the gap between them. Furthermore, the author considers bridging these gaps essential as forensic science needs objective signature verification methods which could be incorporated in the routine work flow of FHEs and the state-of-the-art PR research has such methods. Still, due to not understanding the demands of forensic casework, the PR research seems moving away from forensic science and potentially losing this market. This thesis has surveyed the state-of-the-art of pattern recognition (PR) and forensic methods for automatic signature comparison and has identified four areas where the two communities, PR and FHEs, need to work together so that to benefit from each other. These are data, terminology, output reporting, and how evaluation of automatic systems is carried out today.

First, this thesis has argued that traditionally the signature data used in PR are not actual/close representative of the real world data (especially that available in forensic cases). The systems trained on such data, therefore, are not suitable for application in forensic environments. This particular gap can be bridged by providing more realistic data to PR researchers. To this end, various multilingual (e.g., Chinese, Dutch, English, Japanese, Italian, Bangla) signature and handwriting datasets are gathered in collaboration with FHEs and are made publicly available through the course of this thesis. A special attention has been given on providing data containing such cases which were largely forgotten by PR researchers, e.g., disguised signatures– where authentic authors purposefully make their signatures look like a forgery. By providing such data, this thesis has encouraged PR researchers to look into more realistic signature verification problems and develop solutions to their effect.

Second, this thesis has identified that the terminology used in the two communities differ greatly. In fact, even in PR, there is no standard terminology and people differ in the usage of various terms particularly relating to the various types of forged signatures/handwriting. This thesis has presented a new terminology that is equally useful for both forensic scientists and PR researchers. The proposed terminology is hoped to increase the general acceptability of automatic signature analysis systems in forensic science.

Third, this thesis has noted that the state-of-the-art signature verification systems generally report classification results in the form of either a binary decision (a hard yes/no) or a score (raw evidence) based on similarity/difference. This is not optimally informative and in fact not desired by FHEs and courts. An automatic system should not give verdicts/decisions as it has access to neither the priors of a judicial case nor to the associated non scientific evidence, like motives of the crime. The thesis has argued that automatic systems should rather report the probability of observing the evidence (e.g., a certain similarity/difference score) given the signature belongs to the acclaimed identity, and the probability of observing the same evidence given the signature does not belong to the acclaimed identity. By this, the thesis has urged the state-of-the-art automatic signature verification systems to move from making hard decisions to soft decisions and report likelihood ratios that actually represent the evidential value of the score rather than the raw score (evidence).

Fourth, this thesis has debated that the state-of-the-art evaluation methods like, equal error rate and area under curve do not address the needs of forensic science. These needs require an assessment of the evidential value of signature verification, rather than a hard/pure classification (accept/reject binary decision). This thesis demonstrates and validates the use of a relatively simple adaptation of the current verification methods based on the Bayesian inference dependent calibration of continuous scores rather than hard classifications (binary and/or score based classification). Note that the use of such likelihood based evaluation adaptation has made the signature verification results even more interpretable and usable for forensic experts as this evaluation scheme, unlike EER, takes into account the severity of errors as well to report on the final ranking of systems. For validating the use of such a scheme, through the course of this thesis various international signature verification competitions (ICDAR SigComp 2011, ICFHR 4Nsigcomp 2012, ICDAR SigWiComp 2013, ICDAR SigWIcomp 2015) have been organized. The ICFHR 4NSigComp 2010 for the first time ever included disguised signatures in the competition data. The SigComp 2011 laid the foundation of validating the evaluation adaptation discussed in this thesis where the participating systems were required to produce a comparison score, e.g., a degree of similarity or difference, and the evidential value of that score, for the first time ever. Reporting the results this way took the automated signature verification systems one step closer to the forensic casework. Then in 2102, 2013, and most recently in 2015 this shift from "decision" paradigm to "evidential value" paradigm has been further strengthened where large amount of forensic like data has been made available for developing systems tailored to forensic needs and reporting results in the form of likelihood ratios. Note that this thesis does not suggest PR researchers to stop reporting performance of their systems in terms of measures based on number of correct/incorrect decisions (like EER), rather it motivates them to also report results in the form of likelihood ratios. It will help the PR systems in finding a direct application in forensic casework. The outcome of these competitions not only validated the use of adaptation discussed in this thesis, but also encouraged PR researchers to start using such adaptations so that to enable their systems find application in the real world forensic casework.

After identifying the above mentioned major gaps between the state-of-the-art PR and forensic signature/handwriting analysis research, and providing their probable solutions, this thesis has presented some novel signature verification systems particularly designed to cater the needs of forensic casework. These systems are based on local or the so called part based features like, speeded up robust features, features from accelerated segment test, and fast retina keypoint features. The notion behind using these local feature based approaches is that the author is convinced that analysis of local features is of great worth with respect to forensic signature examination. This is because forensic casework has to consider very sophisticated signing/writing behaviors, e.g., skilled forgeries and disguised signatures, where the holistic view could be very close to other genres of signing behaviors, e.g., genuine, but still local information might contain important indicia for authorship attribution. The systems proposed in this thesis have clearly outperformed the state-of-the-art automatic systems on same data under the same evaluation protocols, thereby justifying the work presented in this thesis. Furthermore, this thesis has provided a novel concatenation of local features and signature stability and used them as a means of verifying signatures. The said local stability analysis has proven to be effective and has further improved the results achieved by using the local features based approaches reported in this thesis.

In order to further strengthen the outcomes of this thesis and to establish/validate the suitability of automatic signature verification systems for forensic casework, this thesis has provided a detailed man versus machine comparison by empirically comparing the performance of the two on the same / similar signature data. The aim is to make some inroads with respect to comparing the performance of the traditional human approach to that used by contemporary objective techniques. The results show that the state-of-the-art systems, and particularly the ones proposed in this thesis, have shown performance comparable to that of trained human forensic handwriting examiners and even outperformed them in some cases. This thesis, however, does not envisage objective automated systems as replacements for FHEs; but foresee a great potential for automatic systems to assist human experts in signature analysis and interpretation in the real world forensic casework. It is considered that there are significant limitations when using machines, since machines are generally trained on the case specific training data (containing the specimen signature samples alone from a said case). This training enables them learn the genuine signing behavior of the specimen author and by virtue of this training, they develop statistical and/or structural models for providing judgments about the questioned signatures. These models are highly influenced by the training data provided to them and therefore the representativeness of the training material is critical to ensure they model the relevant elements of the task that they are to carry out. Further, human experts rely on the case specific data but also heavily rely on their previous knowledge of predictive features associated with genuine and forged behaviors and routinely apply this knowledge to the specific case at hand which is somewhat different from machines. This difference in philosophy between human and machine based examination strategies limits the machines to acting as a potentially good assistant, rather than a complete replacement of the FHE. Having said this, it is also recognized that commercially, outside forensic casework, machines are used in preference to humans (e.g., in the banking industry) primarily due to issues associated with the volume of authentications that are required to be carried out, and the timeliness associated with task. The thesis further reveals that different automatic systems, just like humans, were better on different data. However, there was not much variance in the performances of automatic systems which is unlike that of humans, as humans performance showed a great degree of variations from average to the best case performance. This is potentially because machines are not influenced by psychological and physical changes as the humans do. With these findings, it is hoped that this thesis would motivate many FHEs around the world to use automatic signature analysis systems for their routine casework.

Along with the above mentioned prime concerns, the author has also looked into some areas related closely to forensic signature examination. While developing automatic signature verification systems, a general assumption is that the signatures would be readily available. However, it is a common observation that signature authentication is required on different types of documents like contracts, credit card payments, wills, and letters etc. To perform automatic verification of such signatures, they must be first extracted / segmented from these documents. This thesis underpins the challenges which must be tackled for the development of a complete automatic document analysis system capable of performing signature extraction / segmentation from documents and then performing signature verification. The thesis presents a novel local features based system for signature segmentation from a wide compendium of documents ranging from bank checks to contracts. Furthermore, the thesis also implied the use of Hyper-Spectral Imaging (HSI) for performing signature segmentation and introduced a novel HSI document datasets containing hundreds of HSI forensic like documents. A novel system, having a precision of 100%, is proposed for segmenting signatures from HSI documents. In addition to that, though currently forensic examiners at large perform analysis of signatures available off-line, they might face on-line signature verification cases in a few years or so. The thesis also looks into this dimension and performs an analysis of novel pen devices that could readily be used for signing and later for verification. The choice of analyzing these pens, instead of tablets, is quite conscious as the author believes that with time and advancement in technology, using these pen devices would be a preferred option for many writers as they allow signing on the paper (which is quite natural when compared to signing on tablets) and recording the on-line data at the same time.

12.2 Limitations

This section describes the limitations of both the scenarios faced in this thesis as well as of the solutions presented here. First, the thesis has made various forensic like signature verification datasets available to the PR community. With time these datasets are getting popular in the PR community and many systems have been reported for these data. Note that these data were collected by the author and other PR researchers in collaboration with forensic examiners, where a majority of data donors belonged to various reputed forensic science institutes of the world. The donors were mostly forensic experts or trainees. Further, the data have been collected in a systematic manner where additional information (e.g., age, sex, handedness, etc.) about donors has been recorded along with abiding by the general practices of good data collection. So, there is not much question in the quality of data. However, in PR a large number of systems require enormous amount of data for training, e.g., systems based on various variants of neural networks. This implies that to motivate the PR researchers at large, more forensic like data are required. Furthermore, it is a common practice in PR to train the systems on the so called negative samples. From the view point of FHEs, this is not optimal and in fact should not be done in PR as this is quite contrary to the real world scenarios where an FHE often finds some specimen signatures and a questioned signature without any negative samples. In addition to that, the behavior of any systems trained on negative samples is also affected by the choice of negative samples; which of course is not suitable for the courts.

Second, the terminology presented in this thesis is quite adequate in itself currently where it not only defines the three basic categories of signing behaviors, genuine, forged, and disguise, but also defines the different types of forgeries. However, there are even more natural and unnatural signing behaviors that are available in forensic science and yet have not been largely considered in PR. Nonetheless, the author advocates the usage of definitions of such writing/signing behaviors in the sense similar to what is used in forensic science. This would stop the terminology gap from widening in the future as well.

Third, the analysis of local stability of signatures has been performed in this thesis. Identifying local stable regions and then using them for verification is quite promising when viewed with respect to forensic like data (containing disguised behavior for example). However in other scenarios, like that faced in banks, global stability also seems of interest. Nonetheless, the two could be combined for optimizing systems for handling different real world situations at the same time.

Fourth, the use of HSI for signature segmentation though seems very promising, yet it has an inherent limitation, i.e., currently the HSI devices are too expensive and are not available freely in many areas of the world. Nonetheless, the author hopes that with time this technology will get cheaper and will benefit many areas of document analysis especially signature segmentation and verification.

12.3 Future Work

This thesis contributes both at a theoretical and application level, and accordingly the author sees various research dimensions to consider for the future.

First, to further solidify the findings of this thesis, collection of large signature datasets containing genuine, forged, and disguised signatures, and making them publicly available is planned. These datasets would be multilingual (similar to the ones currently presented in this thesis). These datasets would contain different signing behaviors (discussed in this thesis), and will have cases of varying difficulty. Both the forensic and general purpose aspect of signature verification would be considered. Furthermore, it would be desirable if FHEs could provide PR researchers with freely available realistic data collected from forensic situations. It may need some ethical considerations but freely available datasets with large number of realistic skilled forgeries and disguised signatures would help the PR researchers better optimize their systems for forensic scenarios. Note that the LLR analysis requires more data from more writers. Thus, it is planned to perform analysis on data that contain signatures from more reference writers and skilled forgers. Large and diverse test sets where signatures are produced by different authors under various different psychological and physical conditions may also yield interesting results. Noteworthy, a fundamental limitation with the current data sets is that forgeries are done by "genuine" people. Real "fraudsters" may have unique behavior like nervousness. Creating such data sets may be difficult, but should be addressed by FHEs so that to test the state-of-art automatic methods on "true" data.

Second, the terminology presented in this thesis would be further elaborated where other signing/writing behaviors which are studied by forensic experts and are not yet generally considered in PR research would be defined.

Third, the novel local feature based systems presented in this system would be further refined and special attention would be given to the analysis of stability in signatures. The analysis of signature stability is very important as it can not only largely influence the results of verification but during registration, based on analysis of stability, various specimen signatures could be rejected. This would result in a more stable, and compact feature set for each user and thus would bring better verification results. Furthermore, the effects of aging could be studied by analyzing stability of signatures collected from individuals over a period of time. This would enable development of signature verification systems that would be least affected by aging (sometimes the available specimen signatures and questioned signatures could have been written with a difference of a number of years).

Fourth, the use of digital pens, e.g., Anoto pen, for capturing the dynamics of signatures is increasing day by day. These pens are very similar to normal ball point pens but they have specialized mechanism for capturing signature dynamics with the help of a built-in camera and paper containing specialized dot pattern (the so called Anoto pattern). The other most common option for capturing signature dynamics in through tablets, like Wacom tablet. Both of these types of devices at least provide ASCII files with the format: X, Y, and Pressure. An important future research would be to capture the data from same individuals using the two devices simultaneously and analyzing the impact of capturing devices on verification performed thereafter. These results would help in finding the most suitable on-line signature collection devices for different field, like banking and forensic science.

In addition to that, another requirement in forensic casework is that FHEs often require knowledge about, e.g., important areas of signatures that support and explain the systems decision on genuine, forged, or disguised authorship. In other words, FHEs also require explanations from automated systems. This is sometimes substantial for them as they may correlate their findings with these explanations and present the results in courts. This feature, when included in automated systems, will increase their application in forensic scenarios to many folds and this would be an important task that the author has planned to pursue in the near future.

In the future, such cases should also be considered where automatic systems are applied to detect forgeries that are in fact “perfect copies” of original signatures. This would help in identifying the cases where fraudsters/forgers have taken the genuine signatures of an authentic author and pasted it somehow to pass automatic signature verification. For detecting a fraudster’s cut and paste attempt various image processing techniques can be applied, which in fact can be combined with an end to end signature segmentation and verification system where the system on detect perfect signature copies can consolidate their findings by applying such image processing techniques in the future.

Furthermore for the future, several related events (signature verification competitions and workshops) are planned along with the next Intl. Conf. on Document Analysis & Recognition (ICDAR) and Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR) conferences. The organization of these competitions does not exclude the organization of other scenarios, e.g., it is desired to join forces in the future, i.e., to use the same evaluation protocol for similar competitions. This makes it easier for the participants to adjust their systems and also allows for more possibilities of evaluating several scenarios.

Appendices

A

Abbreviations

A: Number of Authors

AER: Average Error Rate

AF: Artificial Forgeries DS: Disguised Signature

DSM: Dynamic Similarity Measure

EER: Equal Error Rate

FAR: False Acceptance Rate

FBQ: Forgeries of bad quality

FD: Full Dataset

FGQ: Forgeries of good quality

FRR: False Rejection Rate

FS: Forgeries/ Forged Signatures

GS: Genuine Signatures

N: Number of reference signatures

PA: Protected Approach

RF: Random Forgeries

S: Signatures

SF: Simple Forgeries

SK: Skilled Forgeries

T1, T2, ... : Task 1, Task 2, ...

TES: Test Set

TRS: Training Set

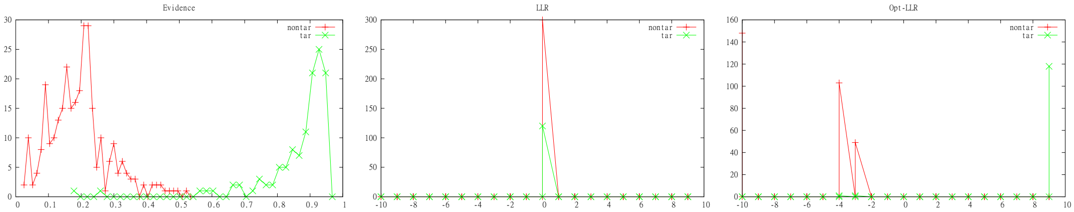
UPA: Un-Protected Approach

WDT: Writer Dependent Threshold

WIT: Writer Independent Threshold

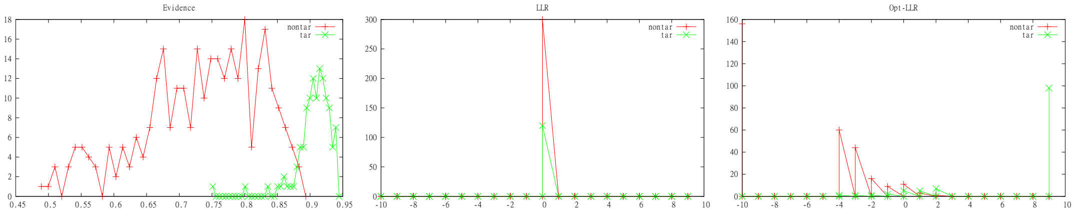
B

LLR Behavior of Off-line Signature Verification Systems from SigWIcomp2015



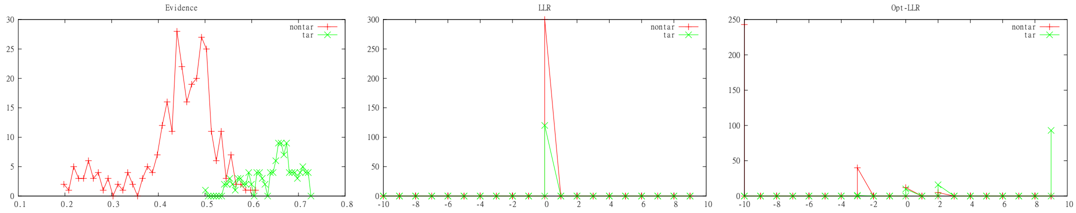
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.1: System 1 on Bengali data



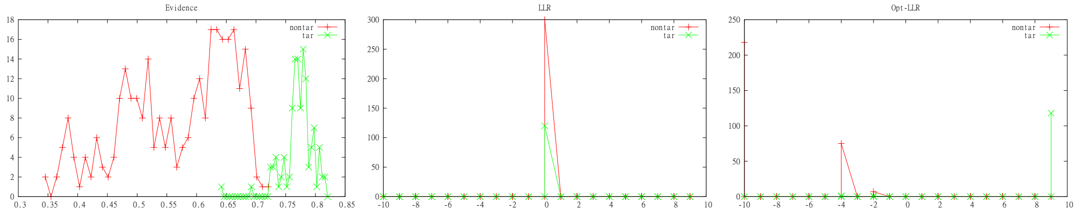
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.2: System 2 on Bengali data



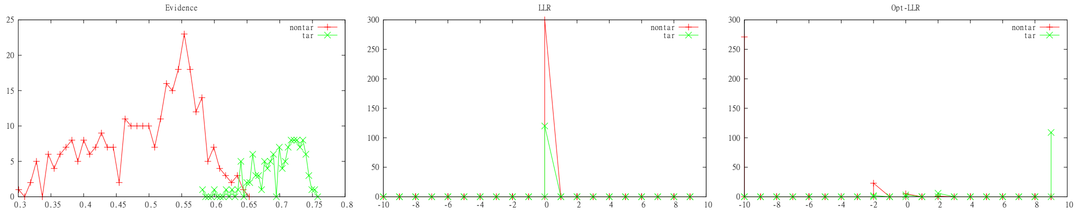
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.3: System 3 on Bengali data



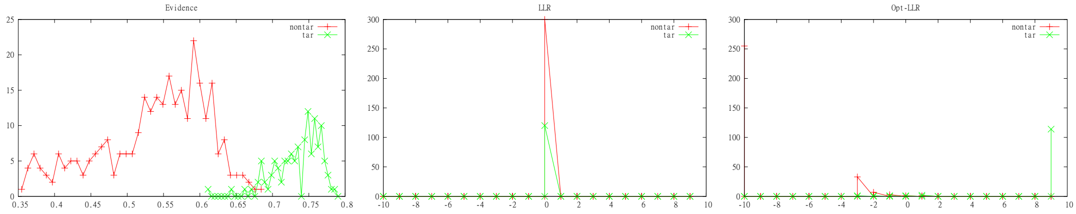
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.4: System 4 on Bengali data



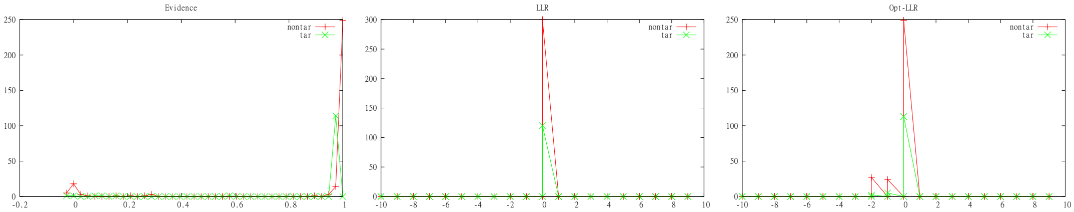
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.5: System 5 on Bengali data



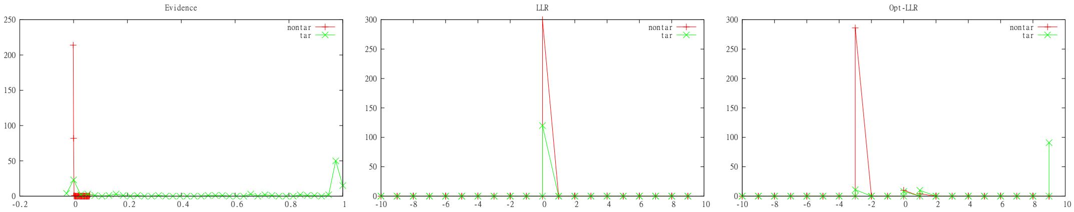
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.6: System 6 on Bengali data



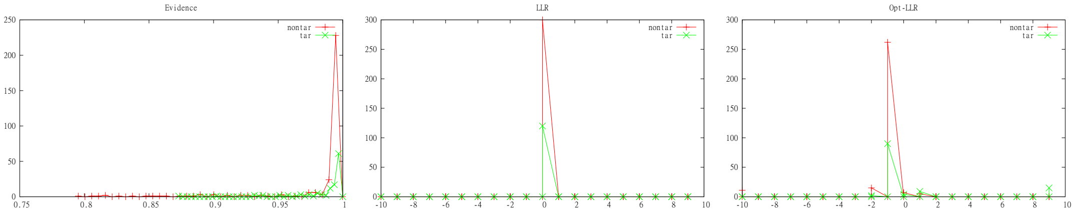
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.7: System 7 on Bengali data



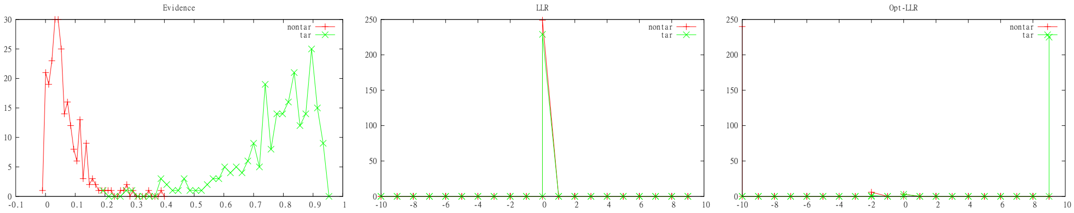
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.8: System 8 on Bengali data



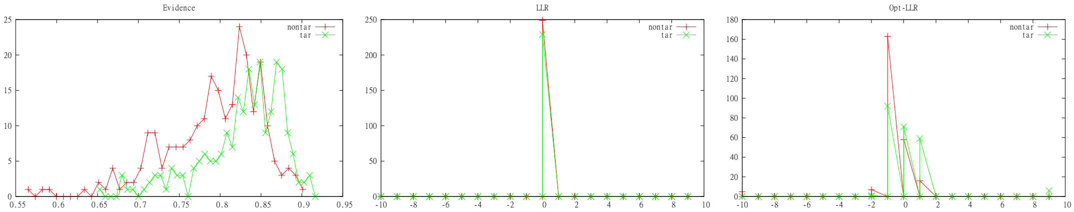
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.9: System 9 on Bengali data



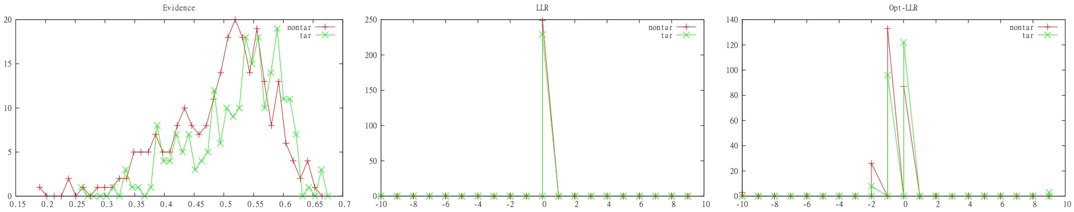
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.10: System 1 on Italian data



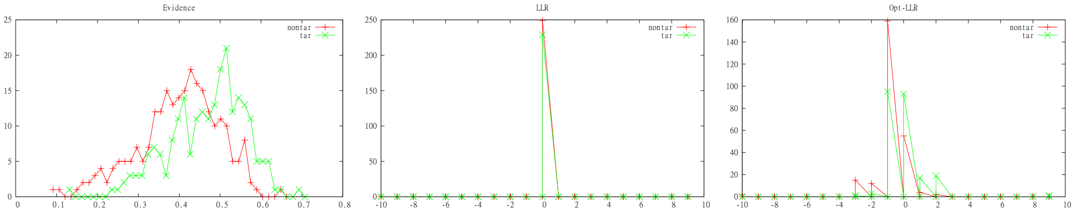
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.11: System 2 on Italian data



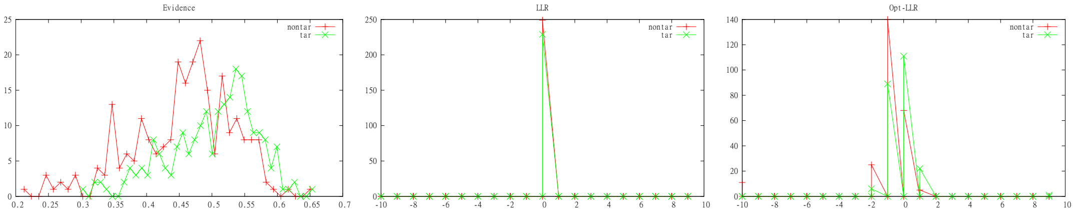
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.12: System 3 on Italian data



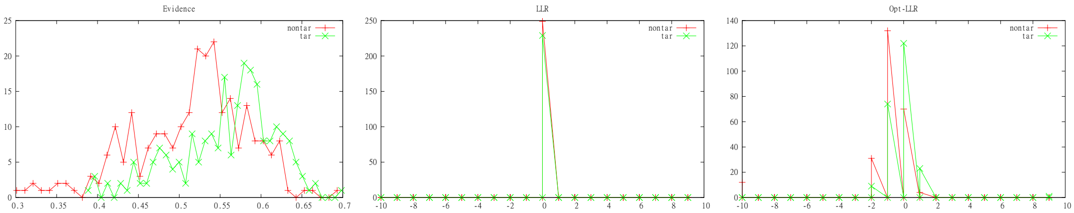
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.13: System 4 on Italian data



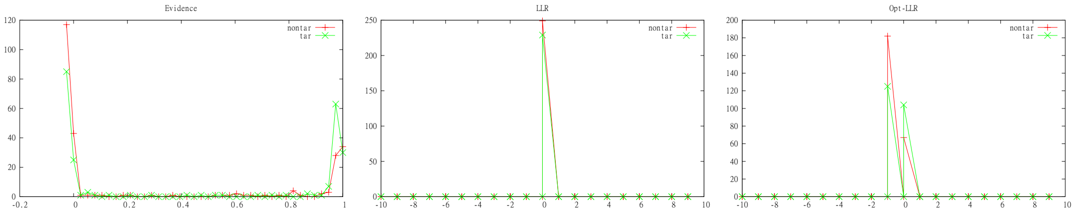
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.14: System 5 on Italian data



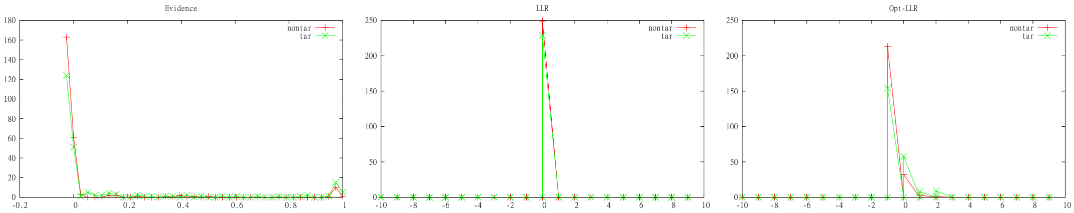
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.15: System 6 on Italian data



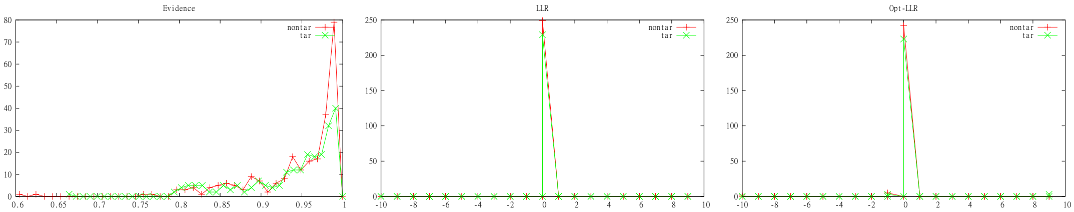
(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.16: System 7 on Italian data



(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.17: System 8 on Italian data



(a) Evidence (b) CLLR (c) Opt-CLLR

Figure B.18: System 9 on Italian data

Bibliography

[1] Vivian L. Blankers, C. Elisa van den Heuvel, Katrin Y. Franke, and Louis G. Vuurpijl. ICDAR 2009 signature verification competition. In *ICDAR*, pages 1403–1407, 2009.

[2] ENFSI Core group on strengthening the evaluation of forensic results across Europe (STEOFRAE). ENFSI guideline for evaluative reporting in forensic science. Technical report, The European Network of Forensic Science Institutes (ENFSI), 2015.

[3] Linda Alewijnse. Offline handwriting acquisition under controlled and uncontrolled conditions. In *Proceedings of the 2nd International Workshop on Automated Forensic Handwriting Analysis, AFHA 2013, Washington DC, USA*, pages 11–14, 2013.

[4] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005.*, volume 2, pages 1508 –1515 Vol. 2, Oct. 2005.

[5] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 510–517. IEEE, 2012.

[6] Historical and interesting bank checks. `http://www.smartmoneydaily.com/personalfinance/15-historical-interesting-bank-checks.aspx`. Accessed: 2015-06-28.

[7] A. S. Osborn. *Questioned documents, 2nd edition.* Nelson Hill, Chicago, 1929.

[8] Ordway Hilton. Can the forger be identified from his handwriting? *The Journal of Criminal Law, Criminology, and police Science*, 43:547–555, November-December 1952.

[9] R. Plamondon and G. Lorette. Automatic signature verification and writer identification – the state of the art. In *Pattern Recognition*, volume 22, pages 107–131, 1989.

[10] F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art 1989–1993. In R. Plamondon, editor, *Progress in Automatic Signature Verification*, pages 13–19. World Scientific Publ. Co., 1994.

[11] Rejean Plamondon and Sargur N. Srihari. On-line and off-line handwriting recognition: a comprehensive survey. *IEEE TPAMI*, 22(1):63–84, 2000.

[12] Donato Impedovo and Giuseppe Pirlo. Automatic signature verification: The state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(5):609–635, 2008.

[13] Marcus Liwicki, Muhammad Imran Malik, Linda Alewijnse, Elisa van den Heuvel, and Bryan Found. ICFHR2012 Competition on Automatic Forensic Signature Verification (4NsigComp 2012). In *13th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR 2012)*, page n.A., Bari, Italy, 2012.

[14] R. J. Verduijn, E. van den Heuvel, and R. Stoel. Forensic requirements for automated handwriting analysis systems. In *IGS*, pages 132–135, 2011.

[15] S. N. Srihari, B. Zhang, C. Tomai, S. Lee, Z. Shi, and Y. C. Shin. A system for handwriting matching and recognition. In *Symp. on Document Image Understanding Technology*, pages 67–75, 2003.

[16] M. Philipp. Fakten zu FISH, das forensische informations-system handschriften des bundeskriminalamtes — eine analyse nach ber 5 jahren wirkbetrieb. Technical report, Bundeskriminalamt, Germany, 1996. in German.

[17] K. Franke, L.R.B. Schomaker, C. Veenhuis, L.G. Vuurpijl, and I. Erp, M. van Guyon. WANDA: A common ground for forensic handwriting examination and writer identification. *ENFHEX News*, pages 23–47, 2004.

[18] B. Walch and D. Gantz. The forensic language-independent analysis system for handwriting identification (flash-id), 2013.

[19] Joaquin Gonzalez-Rodriguez, Julian Fierrez-Aguilar, Daniel Ramos-Castro, and Javier Ortega-Garcia. Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *Forensic Science International*, 155(2-3):126–140, 2005.

[20] Muhammad Imran Malik and Marcus Liwicki. From terminology to evaluation: Performance assessment of automatic signature verification systems. In *International Conference on Frontiers in Handwriting Recognition (ICFHR),*, pages 613–618. IEEE, 2012.

[21] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[22] Chandrshekar S. Patil and A. J. Patil. Article: A review paper on facial detection technique using pixel and color segmentation. *International Journal of Computer Applications*, 62(1):21–24, January 2013. Full text available.

[23] Cedric Neumann, Christophe Champod, Roberto Puch-Solis, Nicole Egli, Alexandre Anthonioz, and Andie Bromage-Griffiths. Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of forensic sciences*, 52(1):54–64, 2007.

[24] Andreas Schlapbach. *Writer identification and verification*, chapter 1, pages 2–4. Clearway Logistics Phase 2-3 (November 14, 2007), 2007.

[25] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology,*, 14(1):4–20, 2004.

[26] F. Bimbot and G. Chollet. *Assessment of speaker verification systems*, chapter 11, pages 408–480. Mouton de Gruyter, Berlin, Germany, 1997.

[27] Madasu Hanmandlu, Mohd. Hafizuddin Mohd. Yusof, and Vamsi Krishna Madasu. Off-line signature verification and forgery detection using fuzzy modeling. *Pattern Recogn.*, 38:341–356, March 2005.

[28] S. Ahmed, M.I. Malik, M. Liwicki, and A. Dengel. Signature segmentation from document images. In *ICFHR*, pages 425–429, Sept 2012.

[29] J. Llados and G. Sanchez. Indexing historical documents by word shape signatures. In *Ninth International Conference on Document Analysis and Recognition, 2007. ICDAR 2007.*, volume 1, pages 362–366, Sept 2007.

[30] Liwicki Marcus. *Recognition of Whiteboard Notes On-Line, Off-Line, and Combination.* PhD thesis, Institut fr Informatik und angewandte Mathematik, Universitt Bern, 2007.

[31] H. Baltzakis and N. Papamarkos. A new signature verification technique based on a two-stage neural network classifier. *Engineering Applications of Artificial Intelligence*, 14(1):95–103, February 2001.

[32] Reena Bajaj and Santanu Chaudhury. Signature verification using multiple neural classifiers. *Pattern Recognition*, 30(1):1–7, January 1997.

[33] Emre Ozgunduz, Tulin Senturk, and Elif KArsligil. Off-line signature verification and recognition by support vector machine. In *in Proc. European Signal Processing Conference*, 2005.

[34] Meenakshi K. Kalera, Sargur Srihari, and Aihua Xu. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(7):1339–1360, 2004.

[35] Julian Fierrez, Javier Ortega-Garcia, Daniel Ramos, and Joaquin Gonzalez-Rodriguez. Hmm-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recogn. Lett.*, 28(16):2325–2334, December 2007.

[36] Jane Bromley, Isabelle Guyon, Yann Lecun, Eduard Sckinger, and Roopak Shah. Signature verification using a "siamese" time delay neural network. In *In NIPS Proc*, 1994.

[37] George S Eskander, Robert Sabourin, and Eric Granger. Hybrid writer-independent–writer-dependent offline signature verification system. *IET Biometrics*, 2(4):169–181, 2013.

[38] Zhong hua Quan and Kun hong Liu. Online signature verification based on the hybrid hmm/ann model. *International Journal of Computer Science and Network Security*, 7(3), 2007.

[39] Muhammad Imran Malik, Linda Alewijnse, Marcus Liwicki, and Michael Blumenstein. Signature verification tutorial. *2nd International Workshop and Tutorial on Automated Forensic Handwriting Analysis (AFHA)*, 2013.

[40] M. I. Malik, Marcus Liwicki, Andreas Dengel, Seiichi Uchida, and Volkmar Frinken. Automatic signature stability analysis and verification using local features. In *14th International Conference on Frontiers in Handwriting Recognition*, pages 621–626. IEEE, 2014.

[41] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Evaluation of local and global features for offline signature verification. In *AFHA*, pages 26–30, 2011.

[42] S. Djeziri, F. Nouboud, and R. Plamondon. Extraction of signatures from check background based on a filiformity criterion. *TIP*, 7(10):1425–1438, October 1998.

[43] Vamsi Krishna Madasu, Mohd Hafizuddin, Mohd Yusof, M. Hanm, and Lu Ss. Automatic extraction of signatures from bank cheques and other documents. In *Proceedings of DICTA*, pages 591–600, 2003.

[44] Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Multi-scale structural saliency for signature detection. In *In Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2007)*, pages 1–8, 2007.

[45] Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Signature Detection and Matching for Document Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2015–2031, November 2009.

[46] Ranju Mandal, Partha Pratim Roy, and Umapada Pal. Signature segmentation from machine printed documents using conditional random field. In *ICDAR*, pages 1170–1174, 2011.

[47] Pavel Senin. Dynamic Time Warping Algorithm Review. Technical Report CSDL-08-04, Department of Information and Computer Sciences, University of Hawaii, Honolulu, Hawaii 96822, December 2008.

[48] J.-J. Brault and R. Plamondon. Segmenting handwritten signatures at their perceptually important points. *IEEE TPAMI*, 15(9):953–957, sep 1993.

[49] Jaeyeon Lee, Ho-Sub Yoon, Jung Soh, Byung Tae Chun, and Yun Koo Chung. Using geometric extrema for segment-to-segment characteristics comparison in online signature verification. *Pattern Recognition*, 37(1):93–103, 2004.

[50] T.H. Rhee, S.J. Cho, and J.H. Kim. On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries. In *ICDAR*, pages 645–649, 2001.

[51] Berrin Yanikoglu and Alisher Kholmatov. Online signature verification using fourier descriptors. *EURASIP J. Adv. Signal Process*, 2009:12:1–12:1, January 2009.

[52] F.A. Afsar, M. Arif, and U. Farrukh. Wavelet transform based global features for online signature recognition. In *9th IEEE International Multitopic Conference,*, pages 1–6, Dec 2005.

[53] Y. Liu, Z. Yang, and L. Yang. Online signature verification based on dct and sparse representation. *IEEE Transactions on Cybernetics*, PP(99):1–1, 2014.

[54] Helger Lipmaa. On differential properties of pseudo-hadamard transform and related mappings. In *Proceedings of the Third International Conference on Cryptology: Progress in Cryptology*, INDOCRYPT '02, pages 48–61, London, UK, UK, 2002. Springer-Verlag.

[55] Robert Sabourin, Ginette Genest, and Franoise Prteux. Pattern spectrum as a local shape factor for off-line signature verification. In *13th ICPR*, pages 43–48, 1996.

[56] W. Song, S. Uchida, and Marcus Liwicki. Comparative study of part-based handwritten character recognition methods. In *ICDAR*, pages 814–818, 2011.

[57] Muhammad Imran Malik, Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel. Freak for real time forensic signature verification. In *12th International Conference on Document Analysis and Recognition*, pages 971–975. IEEE, 2013.

[58] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. In *12th International Conference on Document Analysis and Recognition (ICDAR),*, pages 872–876. IEEE, 2013.

[59] M. I. Malik, Marcus Liwicki, and Andreas Dengel. Local features for off-line forensic signature verification. In *Advances in Digital Handwritten Signature Processing: A Human Artefact for e-Society*, pages 95–110. World Scientific, 2014.

[60] Stefan Fiel and Robert Sablatnig. Writer retrieval and writer identification using local features. In *DAS*, pages 145–149, 2012.

[61] S. Pal, S. Chanda, U. Pal, K. Franke, and M. Blumenstein. Off-line Signature Verification using G-SURF. In *ISDA*, pages 586–591, 2012.

[62] Marcus Liwicki and M. I. Malik. Surprising? power of local features for automated signature verification. In *15th Biennial International Graphonomics Society Conference*, pages 18–21. International Graphonomics Society, 2011.

[63] Jonas Richiardi, Hamed Ketabdar, and Andrzej Drygajlo. Local and global feature selection for on-line signature verification. In *In Proc. IAPR 8th International Conference on Document Analysis and Recognition*, pages 625–629, 2005.

[64] Katrin Franke. Analysis of authentic signatures and forgeries. In ZenoJ.M.H. Geradts, KatrinY. Franke, and CorJ. Veenman, editors, *Computational Forensics*, volume 5718 of *Lecture Notes in Computer Science*, pages 150–164. Springer Berlin Heidelberg, 2009.

[65] A.C. Ramachandra, J.S. Rao, K.B. Raja, K.R. Venugopla, and L.M. Patnaik. Robust offline signature verification based on global features. In *Advance Computing Conference, 2009. IACC 2009. IEEE International*, pages 1173–1178, March 2009.

[66] R. Sabourin and G. Genest. An extended-shadow-code based approach for off-line signature verification. I. evaluation of the bar mask definition. In *Conference B: 12th IAPR Int. Conf. on Computer Vision Image Processing and Pattern Recognition,*, volume 2, pages 450–453, Oct 1994.

[67] The-Anh Pham, Hong-Ha Le, and Nang-Toan Do. Offline handwritten signature verification using local and global features. *Annals of Mathematics and Artificial Intelligence*, pages 1–17, 2014.

[68] Vu Nguyen, M. Blumenstein, and G. Leedham. Global features for the off-line signature verification problem. In *10th International Conference on Document Analysis and Recognition,*, pages 1300–1304, July 2009.

[69] Fengxi Song, Zhongwei Guo, and Dayong Mei. Feature selection using principal component analysis. In *International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM),*, volume 1, pages 27–30, Nov 2010.

[70] Jukka Iivarinen, Kimmo Valkealahti, Ari Visa, and Olli Simula. Feature selection with self-organizing feature map. In Maria Marinaro and PietroG. Morasso, editors, *ICANN 94*, pages 334–337. Springer London, 1994.

[71] A. Marcano-Cede, J. Quintanilla-Dominguez, M.G. Cortina-Januchs, and D. Andina. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In *36th Annual Conference on IEEE Industrial Electronics Society*, pages 2845–2850, Nov 2010.

[72] Mark Michael and Wen-Chun Lin. Experimental study of information measure and inter-intra class distance ratios on feature selection and orderings. *IEEE Transactions on Systems, Man and Cybernetics,*, SMC-3(2):172–181, March 1973.

[73] J. Galbally, J. Fierrez, M.R. Freire, and J. Ortega-Garcia. Feature selection based on genetic algorithms for on-line signature verification. In *IEEE Workshop on Automatic Identification Advanced Technologies,*, pages 198–203, June 2007.

[74] A. Verikas and M. Bacauskiene. Feature selection with neural networks. *Pattern Recognition Letters*, 23(11):1323 – 1335, 2002.

[75] G. Pirlo and D. Impedovo. On the measurement of local stability of handwriting: An application to static signature verification. In *BIOMS*, pages 41–44, 2010.

[76] Donato Impedovo and Giuseppe Pirlo. Stability analysis of static signatures for automatic signature verification. In *ICIAP*, volume 6979 of *LNCS*, pages 241–247. Springer Berlin Heidelberg, 2011.

[77] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla, and C.A. Trullo. Analysis of stability in static signatures using cosine similarity. In *ICFHR*, pages 231–235, 2012.

[78] G. Pirlo and D. Impedovo. Cosine similarity for analysis and verification of static signatures. *IET Biometrics*, 2(4):151–158, 2013.

[79] Kai Huang and Hong Yan. Stability and style-variation modeling for on-line signature verification. *Pattern Recognition*, 36(10):2253 – 2270, 2003.

[80] Antonio Parziale, SalvatoreG. Fuschetto, and Angelo Marcelli. Exploiting stability regions for online signature verification. In Alfredo Petrosino, Lucia Maddalena, and Pietro Pala, editors, *New Trends in Image Analysis and Processing ICIAP*

2013, volume 8158 of *Lecture Notes in Computer Science*, pages 112–121. Springer Berlin Heidelberg, 2013.

[81] Antonio Parziale, Salvatore Gerardo Fuschetto, and Angelo Marcelli. Modeling stability in on-line signatures. *J. Forensic Document Examination*, 24:36–50, 2014.

[82] Giuseppe Pirlo, Donato Impedovo, Rejean Plamondon, Christian OReilly, A. Cozzolongo, R. Gravinese, and Andrea Rollo. Stability of dynamic signatures: From the representation to the generation domain. In Alfredo Petrosino, Lucia Maddalena, and Pietro Pala, editors, *New Trends in Image Analysis and Processing ICIAP 2013*, volume 8158 of *Lecture Notes in Computer Science*, pages 122–130. Springer Berlin Heidelberg, 2013.

[83] G Pirlo and D Impedovo. Stability analysis of dynamic signatures in multiple representation domains: application to automatic signature verification. *International Journal of Signal and Imaging Systems Engineering*, 7(3):180–188, 2014.

[84] Marianela Parodi, Juan Carlos Gómez, and Linda Alewijnse. Automatic online signature verification based only on FHE features: An oxymoron? In *14th International Conference on Frontiers in Handwriting Recognition, ICFHR*, pages 73–78, 2014.

[85] A. G. Dyer, B. Found, and D. Rogers. Visual attention and expertise for forensic signature analysis. *Journal of Forensic Sciences*, 51(6):1397–1404, 2006.

[86] J. Sita, B. Found, and D. Rogers. Forensic handwriting examiners' expertise for signature comparison. *Journal of Forensic Sciences*, 47:1117–1124, 2002.

[87] Hina Arora Sargur N. Srihari, Sung-Hyuk Cha and Sangjik Lee. Individuality of handwriting. *Journal of Forensic Sciences*, 47(4):1–17, 2002.

[88] Meinard Mueller. Dynamic time warping. In *Information Retrieval for Music and Motion*, pages 69–84. Springer-Verlag Berlin Heidelberg, 2007.

[89] L. Bovino, S. Impedovo, G. Pirlo, and L. Sarcinella. Multi-expert verification of hand-written signatures. In *ICDAR*, pages 932–936, Aug. 2003.

[90] S. Impedovo and G. Pirlo. Verification of handwritten signatures: an overview. In *ICIAP*, pages 191–196, Modena, Italy, 2007. IEEE Computer Society.

[91] C. Gruber, T. Gruber, S. Krinninger, and B. Sick. Online signature verification with support vector machines based on lcss kernel functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, 40(4):1088–1100, Aug 2010.

[92] J.-P. Drouhard, R. Sabourin, and M. Godbout. A neural network approach to off-line signature verification using directional {PDF}. *Pattern Recognition*, 29(3):415 – 424, 1996.

[93] Alan McCaben, Jarrod Trevathan, and Wayne Read. Neural network-based handwritten signature verification. *Journal of Computers*, 3(8):9–22, 2008.

[94] Vu Nguyen, M. Blumenstein, V. Muthukkumarasamy, and G. Leedham. Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines. In *ICDAR*, pages 734–738, sept. 2007.

[95] Muhammad Imran Malik, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found. Signature Verification and Writer Identification Competitions for On- and Offline Skilled Forgeries (SigWiComp2013). In *12th Int. Conf. on Document Analysis and Recognition*, pages 1477–1483, Washigton, DC, USA, 2013.

[96] Y. Guerbai, Y. Chibani, and N. Abbas. One-class versus bi-class svm classifier for off-line signature verification. In *International Conference on Multimedia Computing and Systems (ICMCS),*, pages 206–210, May 2012.

[97] Fatih Camci and Ratna B. Chinnam. General support vector representation machine for one-class classification of non-stationary classes. *Pattern Recognition*, 41(10):3021–3034, 2008.

[98] Cheila Bergamini, Luiz S. Oliveira, Alessandro L. Koerich, and Robert Sabourin. Combining different biometric traits with one-class classification. *Signal Processing*, 89(11):2117–2127, 2009.

[99] Kwang-Kyu Seo. An application of one-class support vector machines in content-based image retrieval. *Expert Systems with Applications*, 33:491–498, 2007.

[100] Belkacem Fergani, Manuel Davy, and Amrane Houacine. Speaker diarization using one-class support vector machines. *Speech Communication*, 50(5):355–365, 2008.

[101] Mennatallah Amer, Markus Goldstein, and Slim Abdennadher. Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description (ODD)*, pages 8–15. ACM, 8 2013.

[102] Larry M. Manevitz and Malik Yousef. One-class svms for document classification. *J. Mach. Learn. Res.*, 2:139–154, 2002.

[103] Larry Manevitz and Malik Yousef. One-class document classification via neural networks. *Neurocomputing*, 70(79):1466 – 1481, 2007.

[104] David M. J. Tax and Robert P. W. Duin. Combining one-class classifiers. In *Proceedings of Second International Workshop on Multiple Classifier SystemsMCS*, pages 299–308, 2001.

[105] Oleksiy Mazhelis. One-class classifiers : a review and analysis of suitability in the context of mobile-masquerader detection. *South African Computer Journal*, 36:29–48, 2006.

[106] J. Galbally, J. Fierrez, M. Martinez-Diaz, and J. Ortega-Garcia. Improving the enrollment in dynamic signature verfication with synthetic samples. In *ICDAR*, pages 1295–1299, july 2009.

[107] J. Galbally, J. Fierrez, M. Martinez-Diaz, J. Ortega-Garcia, R. Plamondon, and C. O'Reilly. Kinematical analysis of synthetic dynamic signatures using the sigma-lognormal model. In *ICFHR*, pages 113–118, nov. 2010.

[108] Miguel A Ferrer, Moisés Díaz-Cabrera, and Aythami Morales. Synthetic off-line signature image generation. In *International Conference on Biometrics (ICB),*, pages 1–7. IEEE, 2013.

[109] Miguel A Ferrer, Moises Diaz-Cabrera, and Aythami Morales. Static signature synthesis: A neuromotor inspired approach for biometrics. *Transaction on Pattern Analysis and Machine Intelligence*, 2014.

[110] Javier Galbally, Moises Diaz-Cabrera, Miguel A Ferrer, Marta Gomez-Barrero, Aythami Morales, and Julian Fierrez. On-line signature recognition through the combination of real dynamic data and synthetically generated static data. *Pattern Recognition*, 48(9):2921–2934, 2015.

[111] Luana Batista, Eric Granger, and Robert Sabourin. A multi-classifier system for off-line signature verification based on dissimilarity representation. In Neamat El Gayar, Josef Kittler, and Fabio Roli, editors, *Multiple Classifier Systems*, volume 5997 of *Lecture Notes in Computer Science*, pages 264–273. Springer Berlin Heidelberg, 2010.

[112] Bin Fang, C. H. Leung, Yuan Yan Tang, K. W. Tse, Paul C. K. Kwok, and Y. K. Wong. Off-line signature verification by the tracking of feature and stroke positions. *Pattern Recognition*, 36(1):91–101, 2003.

[113] B. Fang and Y. Y. Tang. Reduction of Feature Statistics Estimation Error for Small Training Sample Size in Off-Line Signature Verification. In *ICBA*, pages 526–532, 2004.

[114] Stephane Armand, Michael Blumenstein, and Vallipuram Muthukkumarasamy. Off-line Signature Verification using the Enhanced Modified Direction Feature and Neural-based Classification. In *International Joint Conference on Neural Networks*, pages 684–691, 2006.

[115] Peter S. Deng, Hong-Yuan M. Liao, Chin W. Ho, and Hsiao-Rong Tyan. Wavelet-Based Off-Line Handwritten Signature Verification. *Computer Vision and Image Understanding*, 76(3):173–190, December 1999.

[116] J. Drouhard, R. Sabourin, and M. Godbout. A neural network approach to off-line signature verification using directional PDF. *Pattern Recognition*, 29(3):415–424, March 1996.

[117] E. Fadhel and P. Bhattacharyya. Application of a Steerable Wavelet Transform using Neural Network for Signature Verification. *Pattern Analysis & Applications*, 2(2):184–195, June 1999.

[118] M.A. Ferrer, J.B. Alonso, and C.M. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE TPAMI*, 27(6):993–997, June 2005.

[119] Kai Huang and Hong Yan. Off-line signature verification using structural feature correspondence. *Pattern Recognition*, 35(11):2467–2477, November 2002.

[120] C.R. Prashanth, K. B. Raja, K. R. Venugopal, and L. M. Patnaik. Standard scores correlation based off-line signature verification system. In *ACT '09*, pages 49–53, Washington, DC, USA, 2009. IEEE Computer Society.

[121] V. Ramesh and M. Murty. Off-line signature verification using genetically optimized weighted features. *Pattern Recognition*, 32(2):217–233, February 1999.

[122] K. Ueda. Investigation of off-line japanese signature verification using a pattern matching. In *ICDAR*, pages 951–955, Aug. 2003.

[123] Jesus Francisco Vargas Bonilla, Miguel Angel Ferrer Ballester, Carlos Manuel Travieso Gonzalez, and Jesus Bernandino Alonso Hernandez. Offline signature verification based on pseudo-cepstral coefficients. In *ICDAR*, pages 126–130, USA, 2009. IEEE Computer Society.

[124] Ibiyemi Samuel Daramola Samuel. Novel feature extraction technique for off-line signature verification system. *International Journal of Engineering Science and Technology*, 2:3137–3143, 2010.

[125] N.A. Murshed, F. Bortolozzi, and R. Sabourin. Off-line signature verification using fuzzy artmap neural network. In *IEEE International Conference on Neural Networks*, volume 4, pages 2179–2184 vol.4, Nov 1995.

[126] Marianela Parodi, Juan C. Gomez, and Abdel Belaïd. A circular grid-based rotation invariant feature extraction approach for off-line signature verification. In *ICDAR*, pages 1289–1293, Washington, DC, USA, 2011. IEEE Computer Society.

[127] Cesar Santos, Edson J. R. Justino, Flavio Bortolozzi, and Robert Sabourin. An off-line signature verification method based on the questioned document expert's approach and a neural network classifier. In *IWFHR*, pages 498–502, Washington, DC, USA, 2004. IEEE Computer Society.

[128] D.G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, volume 2, pages 1150–1157 vol.2, 1999.

[129] Song Wang, Seiichi Uchida, Marcus Liwicki, and Yaokai Feng. Part-based methods for handwritten digit recognition. *Frontiers of Computer Science*, 7(4):514–525, 2013.

[130] Zhiyi Zhang, Lianwen Jin, Kai Ding, and Xue Gao. Character-sift: A novel feature for offline handwritten chinese character recognition. In *ICDAR*, pages 763–767. IEEE Computer Society, 2009.

[131] Zhen Jin, Kaiyue Qi, Yi Zhou, Kai Chen, Jianbo Chen, and Haibing Guan. Ssift: An improved sift descriptor for chinese character recognition in complex images. In *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on*, pages 1–5, 2009.

[132] Duy-Nguyen Ta, Wei-Chao Chen, Natasha Gelfand, and Kari Pulli. Surftrac: Efficient tracking and continuous object recognition using local feature descriptors. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR09)*, 2009.

[133] Kanghun Jeong and Hyeonjoon Moon. Object detection using fast corner detector based on smartphone platforms. *Computers, Networks, Systems and Industrial Engineering, ACIS/JNU International Conference on*, 0:111–115, 2011.

[134] Piotr Bilinski, Francois Bremond, and Mohamed Becha Kaaniche. Multiple object tracking with occlusions using hog descriptors and multi resolution images. In *ICDP*, pages 1–6, 2009.

[135] M. Diem and R. Sablatnig. Recognition of degraded handwritten characters using local features. In *ICDAR*, pages 221–225, 2009.

[136] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Part-based system for forensic signature verification involving disguised signatures. In *5th International Workshop on Computational Forensics*, pages 35–43, 2012.

[137] Marcus Liwicki, C. Elisa van den Heuvel, Bryan Found, and Muhammad Imran Malik. Forensic signature verification competition 4nsigcomp2010 - detection of simulated and disguised signatures. In *ICFHR*, pages 715–720. o.A., 2010.

[138] G. Congedo, G. Dimauro, A.M. Forte, S. Impedovo, and G. Pirlo. Selecting reference signatures for on-line signature verification. In Carlo Braccini, Leila DeFloriani, and Gianni Vernazza, editors, *Image Analysis and Processing*, volume 974 of *Lecture Notes in Computer Science*, pages 521–526. Springer Berlin Heidelberg, 1995.

[139] V. Di Lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo, A. Salzo, and L. Sarcinella. Selection of reference signatures for automatic signature verification.

In *Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International Conference on*, pages 597–600, 1999.

[140] Sascha Mller and Olaf Henniger. Evaluating the biometric sample quality of handwritten signatures. In Seong-Whan Lee and StanZ. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 407–414. Springer Berlin Heidelberg, 2007.

[141] Xu-Hong Xiao and Graham Leedham. Signature verification by neural networks with selective attention. *Applied Intelligence*, 11:213–223, September 1999.

[142] J. J. Brault and R. Plamondon. A complexity measure of handwritten curves: modeling of dynamic signature forgery. *IEEE Transactions on Systems, Man and Cybernetics,*, 23(2):400–413, 1993.

[143] M. C. Fairhurst, E. Kaplani, and R. M. Guest. Complexity measures in handwritten signature verification. In *in Proc. 1st Intl. Conf. on Universal Access in Human-Computer Interaction*, New Orleans, USA, 2001.

[144] T. Dewhurst, B. Found, and D. Rogers. The relationship between quantitatively modelled signature complexity levels and forensic document examiners' qualitative opinions on casework. *Journal of Forensic Document Examination*, 18:21–40, 2007.

[145] Bryan Found and Doug Rogers. Contemporary issues in forensic handwriting examination. a discussion of key issues in the wake of the starzecpyzel decision. *Journal of Forensic Document Examination*, 8:1–31, 1995.

[146] B. Found and D. Rogers. A consideration of the theoretical basis of forensic handwriting examination: The application of Complexity Theory? to understanding the basis of handwriting identification. *International Journal of Forensic Document Examiners*, 4:109–118, 1998.

[147] B. Found, D. Rogers, V. Rowe, and D. Dick. Statistical modelling of experts' perceptions of the ease of signature simulation. *Journal of Forensic Document Examination*, 11:73–99, 1998.

[148] Erwin Mattijssen. The relation between the ease of signature simulation and the quality of produced simulations. Master's thesis, University of Amsterdam, the Netherlands, 2009.

[149] Erwin Mattijssen, C. Elisa van den Heuvel, and Reinoud D. Stoel. The relation between signature complexity and the perceived quality of signature simulations. In *IGS*, pages 185–189, 2011.

[150] Avni PEPE, Douglas K. ROGERS, and Jodi C. SITA. A cognitive look into simulations of high and low complexity signatures. In *IGS*, pages 136–139, 2011.

[151] Avni PEPE, Douglas K. ROGERS, and Jodi C. SITA. A Consideration of signature complexity using simulators' gaze behaviour. *Journal of Forensic Document Examination*, 22:5–13, 2012.

[152] F. A. Fernandez, M. C. Fairhurst, J. Fierrez, and J O. Garcia. Impact of signature legibility and signature type in off-line signature verification. In *Biometrics Symposium*, pages 1–6, 2007.

[153] Linda C. ALEWIJNSE, C. Elisa VAN DEN HEUVEL, Reinoud D. STOEL, and Katrin FRANKE. Analysis of signature complexity. In *IGS*, pages 6–9, 2009.

[154] D.S. Guru and H.N. Prakash. Online signature verification and recognition: An approach based on symbolic representation. *IEEE TPAMI*, 31:1059–1073, 2009.

[155] Muhammad Talal Ibrahim, Matthew Kyan, M. Aurangzeb Khan, Khurram Saleem Alimgeer, and Ling Guan. On-line signature verification: Directional analysis of a signature using weighted relative angle partitions for exploitation of inter-feature dependencies. In *ICDAR*, pages 41–45, USA, 2009. IEEE Computer Society.

[156] Muhammad Talal Ibrahim, Matthew Kyan, M. Aurangzeb Khan, and Ling Guan. On-line signature verification using 1-d velocity-based directional analysis. In *ICPR*, pages 3830–3833, USA, 2010. IEEE Computer Society.

[157] Anil K. Jain, Friederike D. Griess, and Scott D. Connell. On-line signature verification. *Pattern Recognition*, 35(12):2963–2973, 2002.

[158] Emanuele Maiorana, Patrizio Campisi, Julian Fierrez, Javier Ortega-Garcia, and Alessandro Neri. Cancelable templates for sequence-based biometrics with application to on-line signature recognition. *Trans. Sys. Man Cyber. Part A*, 40(3):525–538, May 2010.

[159] B. Ly Van, S. Garcia Salicetti, and B. Dorizzi. On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 37(5), 2007.

[160] Mohammad M. Shafiei and Hamid R. Rabiee. A new on-line signature verification algorithm using variable length segmentation and Hidden Markov Models. In *ICDAR*, page 443, USA, 2003. IEEE Computer Society.

[161] L Yang, B K Widjaja, and R Prasad. Application of Hidden Markov Models for signature verification. *Pattern Recognition*, 28(2):161–170, 1995.

[162] Quen-Zong Wu, Suh-Yin Lee, and I-Chang Jou. On-line signature verification based on logarithmic spectrum. *Pattern Recognition*, pages 1865–1871, 1998.

[163] Dangxiao Wang, Yuru Zhang, Chong Yao, Jun Wu, Huimin Jiao, and Muli Liu. Toward force-based signature verification: A pen-type sensor and preliminary validation. *IEEE T. Instrumentation and Measurement*, 59(4):752–762, 2010.

[164] Ramanujan S. Kashi, Jianying Hu, W. L. Nelson, and W. Turin. On-line handwritten signature verification using Hidden Markov Model features. In *ICDAR*, pages 253–257, Washington, DC, USA, 1997. IEEE Computer Society.

[165] Edgard Nyssen, Hichen Sahli, and Kui Zhang. A multi-stage online signature verification system. *Pattern Analysis & Applications*, 5:288–295, 2002.

[166] C. E. Pippin. Dynamic signature verification using local and global features. Technical report, Georgia Inst. Inform. Technol., Atlanta, GA, July 2004.

[167] Julian Fierrez-Aguilar, Loris Nanni, Jaime Lopez-Pealba, Javier Ortega-Garcia, and Davide Maltoni. An on-line signature verification system based on fusion of local and global information. In *AVBPA'05*, pages 523–532, 2005.

[168] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.

[169] Christian Gruber, Thiemo Gruber, Sebastian Krinninger, and Bernhard Sick. Online signature verification with support vector machines based on LCSS kernel functions. *Trans. Sys. Man Cyber. Part B*, 40(4):1088–1100, August 2010.

[170] M.R. Hecker. *Forensische Handschriftenuntersuchung: eine systematische Darstellung von Forschung, Begutachtung und Beweiswert.* Internationale Kriminalistik. Kriminalistik Verlag, 1993.

[171] Sharifah Mumtazah Syed Ahmad, Loo Yim Ling, Rina Md Anwar, Masyura Ahmad Faudzi, and Asma Shakil. Analysis of the effects and relationship of perceived handwritten signature's size, graphical complexity, and legibility with dynamic parameters for forged and genuine samples. *Journal of Forensic Sciences*, 58(3):724–731, 2013.

[172] J. Galbally, J. Fierrez, , and J. Ortega-Garcia. Classification of handwritten signatures based on name legibility. In *Proc. SPIE Defense and Security Symposium, Biometric Technologies for Human Identification, BTHI*, volume 6539. Proc. SPIE, April 2007.

[173] Y. Qiao, J. Liu, and X. Tang. offline signature verification using online handwriting registration. In *CVPR*, pages 1–8, June 2007.

[174] Mayank Vatsa, Richa Singh, Pabitra Mitra, and Afzel Noore. Signature verification using static and dynamic features. In NikhilRanjan Pal, Nik Kasabov, RajaniK. Mudi, Srimanta Pal, and SwapanKumar Parui, editors, *Neural Information Processing*, volume 3316 of *Lecture Notes in Computer Science*, pages 350–355. Springer Berlin Heidelberg, 2004.

[175] Alan McCabe and Jarrod Trevathan. Handwritten signature verification using complementary statistical models. *Journal of Computers*, 4(7):670–680, 2009.

[176] Miguel A Ferrer, Aythami Morales, J Francisco Vargas, Ivan Lemos, and Mónica Quintero. Is it possible to automatically identify who has forged my signature? approaching to the identification of a static signature forger. In *10th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 175–179. IEEE, 2012.

[177] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. MCYT baseline corpus: a bimodal biometric database. *IEEE Proceedings - Vision, Image, and Signal Processing*, 150(6):395–401, 2003.

[178] F. Vargas, M. Ferrer, C. Travieso, and J. Alonso. Off-line Handwritten Signature GPDS-960 Corpus. In *ICDAR*, pages 764–768. IEEE, September 2007.

[179] Dit yan Yeung, Hong Chang, Yimin Xiong, Susan George, Ramanujan Kashi, Takashi Matsumoto, and Gerhard Rigoll. Svc2004: First international signature verification competition. In *ICBA*, pages 16–22. Springer, 2004.

[180] A. J. Mansfield and J. L. Waymen. Best practices in testing and reporting performance of biometric devices version 2.0. Technical report, National Physical Laboratory, San Jose State University, San Jose, California, United States, August 2002.

[181] Matthew Young. Final text of TR 19795-3, biometric performance testing and reporting, part 3: Modality specific testing. *Biometrics*, page 23, 2007.

[182] D. Petrovska. Biosecure multimodal evaluation compaign bmec 2007. In *IEEE Conference on Biometrics: Theory, Applications and Systems*, page o.A., Washington, DC, 2007.

[183] N. Houmani, A. Mayoue, and et al. BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures. *Pattern Recognition*, 45(3):993–1003, Aug 2011.

[184] Nesma Houmani, Sonia Garcia-Salicetti, Bernadette Dorizzi, Jugurta Montalvo, Jnio Coutinho Canuto, Mrio Vasconcelos Andrade, Yu Qiao, Xingxing Wang, Tobias Scheidat, Andrey Makrushin, Daigo Muramatsu, Joanna Putz-Leszczynska, Michal Kudelski, Marcos Faundez-Zanuy, Juan Pascual-Gaspar, Valentin Cardeoso-Payo, Carlos Vivaracho-Pascual, Enrique Argones Ra, Jos Luis Alba Castro, Alisher Kholmatov, and Berrin Yanikoglu. Biosecure signature evaluation campaign (esra2011): Evaluating systems on quality-based categories of skilled forgeries. In *Proceedings of the International Joint Conference on BIOMETRICS (IJCB)*, page o.A, Washington D.C., U.S.A., 2011.

[185] I. W. Evett. Bayesian inference and forensic science: Problems and perspectives. In *Journal of the Royal Statistical Society. Series D, The Statistician*, volume 36, pages 99–105. Blackwell Publishing for the Royal Statistical Society, 1987.

[186] Tom Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874, 2006.

[187] Alvin F. Martin, George R. Doddington, Terri Kamm, Mark Ordowski, and Mark A. Przybocki. The DET curve in assessment of detection task performance. In *EUROSPEECH*, pages 1895–1898, 1997.

[188] N. Houmani, S. Garcia Salicetti, and B. Dorizzi. A Novel Personal Entropy Measure confronted with Online Signature Verification Systems' Performance. In *2nd IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, September 2008.

[189] Sonia Garcia-Salicetti, Nesma Houmani, and Bernadette Dorizzi. A novel criterion for writer enrolment based on a time-normalized signature sample entropy measure. *EURASIP J. Adv. Signal Process*, 2009:9:1–9:12, January 2009.

[190] I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik. What size test set gives good error rate estimates? *IEEE TPAMI*, 20(1):52–64, Jan 1998.

[191] A. J. Mansfield, J. L. Wayman, Authorised Dr, Dave Rayner, and J. L. Wayman. Best practices in testing and reporting performance, 2002.

[192] B. Found. An introduction to the character of forensic investigations into questioned signatures. In *1st Int. Workshop on Automated Forensic Handwriting Analysis (AFHA), Beijing, China*, 2011.

[193] Linton A. Mohammed, Bryan Found, Michael Caligiuri, and Doug Rogers. The dynamic character of disguise behavior for text-based, mixed, and stylized signatures. *Journal of Forensic Sciences*, 56:S136–S141, 2011.

[194] Brian Found and D Rogers. Documentation of forensic handwriting comparison and identification method: a modular approach. *Journal of Forensic Document Examination*, 12:1–68, 1999.

[195] Bryan Found, Doug Rogers, and Carolyne Bird. Documentation of forensic handwriting method: A modular approach. Technical report, Victoria Police Forensic Services Department, Victoria, Australia, 2012.

[196] National Research Council Committee on Identifying the Needs of the Forensic Sciences Community. *Strengthening Forensic Science in the United States: A Path Forward.* The National Academies Press, 2009.

[197] Anders Nordgaard and Birgitta Rasmusson. The likelihood ratio as value of evidencemore than a question of numbers. *Law, Probability and Risk*, 11:303–315, 2012.

[198] Jane A. Lewis. *Forensic Document Examination, Fundamentals and Current Trends*. Elsevier Academic Press, 2014.

[199] sargur Srihari. iFOX. In *Presentation at Measurement Science and Standards in Forensic Handwriting Analysis Conference, Gaithersburg, MD, USA*, 2013.

[200] M. Schulte-Austum. D-Scribe. In *Presentation at Measurement Science and Standards in Forensic Handwriting Analysis Conference, Gaithersburg, MD, USA*, 2013.

[201] Marcus Liwicki, M. I. Malik, and Charles Berger. Towards a shared conceptualization for automatic signature verification. In *Advances in Digital Handwritten Signature Processing: A Human Artefact for e-Society*, pages 65–80. World Scientific, 2014.

[202] Marcus Liwicki, Muhammad Imran Malik, C. Elisa van den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. Signature verification competition for online and offline skilled forgeries (SigComp2011). In *ICDAR*, pages 1480–1484, 2011.

[203] Edson Justino, Flavio Bortolozzi, and Robert Sabourin. A comparison of SVM and HMM classifiers in the off-line signature verification. *Pattern Recognition Letters*, 26(9):1377–1385, July 2005.

[204] Weiping Hou, Xiufen Ye, and Kejun Wang. A survey of off-line signature verification. In *Int. Conf. on Intelligent Mechatronics and Automation*, pages 536–541, 26-31, 2004.

[205] M Blumenstein, Miguel A Ferrer, and J F Vargas. The 4nsigcomp2010 off-line signature verification competition: Scenario 2. In *ICFHR*, pages 721–726. IEEE, 2010.

[206] D. Bertolini, L.S. Oliveira, E. Justino, and R. Sabourin. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *Pattern Recognition*, 43:387–396, January 2010.

[207] Tahnee Dewhurst, Bryan Found, and Doug Rogers. Are expert penmen better than lay people at producing simulations of a model signature? *Forensic science international*, 180(1):50–53, 2008.

[208] B Found and D Rogers. The impact of forger practice on the validity of forensic document practitioners' opinions. In *Forensic Science International*, volume 136, pages 86–87. Elsevier, Ireland, 2003.

[209] L. Cordella, P. Foggia, C. Sansone, F. Tortorella, and M. Vento. A cascaded multiple expert system for verification. In *Multiple Classifier Systems*, volume 1857 of *Lecture Notes in Computer Science*, pages 330–339. Springer Berlin / Heidelberg, 2000.

[210] A. El-yacoubi, E. J. R. Justino, R. Sabourin, and F. Bortolozzi. Off-line signature verification using hmms and cross-validation. In *Proc. of the IEEE Workshop on Neural Networks for Signal Processing*, pages 859–868, 2000.

[211] Muhammad Imran Malik, Sheraz Ahmed, Angelo Marcelli, Umapada Pal, Michael Blumenstein, Linda Alewijns, and Marcus Liwicki. ICDAR2015 Competition on Signature Verification and Writer Identification for On- and Off-line Skilled Forgeries (SigWIcomp2015). In *13th Int. Conf. on Document Analysis and Recognition*, pages 1186–1190, Gammarth, Tunisia, 2015.

[212] K.A. Martire, R.I. Kemp, M. Sayle, and B.R. Newell. On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, 240:61 – 68, 2014.

[213] Henry Mornard. L'affaire Dreyfus: la revision du procés de Rennes, Ligue française pour la défense des droits de l'homme et du citoyen, 1907.

[214] Niko Brümmer and Johan du Preez. Application-independent evaluation of speaker detection. *Computer Speech & Language*, 20(2-3):230–275, 2006.

[215] Daniel Ramos Castro. *Forensic Evaluation of the Evidence using Automatic Speaker Recognition Systems.* PhD thesis, ATVS-Biometric Recognition Group, Universidad Autónoma de Madrid, Madrid, Spain, Dec. 2007.

[216] Niko Brümmer. *Measuring, refining and calibrating speaker and language information extracted from speech.* PhD thesis, University of Stellenbosch, South Africa, Oct. 2010.

[217] Reinier J. VERDUIJN, C. Elisa van den Heuvel, and Reinoud D. STOEL. Exploratory investigation on the performance of the cedar-fox system for forensic handwriting verification and identification. In *The 15th Int. Graphonomics Society Conf. (IGS2011)*, pages 177–180, Live Aqua Cancun, Mexico, June 2011.

[218] Muhammad Imran Malik, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found. Icdar 2013 competitions on signature verification and writer identification for on-and offline skilled forgeries (sigwicomp 2013). In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1477–1483. IEEE, 2013.

[219] A. Gilperez, F. Alonso-Fernandez, S. Pecharroman, J. Fierrez, and J. Ortega-Garcia. Off-line signature verification using contour features. In *ICFHR*, 2008.

[220] M. Bulacu and L. Schomaker. Text-independent writer identification and verification using textural and allographic features. *IEEE TPAMI*, 29(4):701–717, April 2007.

[221] Vu Nguyen and Michael Blumenstein. An application of the 2D gaussian filter for enhancing feature extraction in off-line signature verification. In *ICDAR*, pages 339–343, sept. 2011.

[222] U. V. Marti and H. Bunke. *Using a statistical language model to improve the performance of an HMM-based cursive handwriting recog- nition systems*, chapter 3, pages 65–90. World Scientific Publishing Co., Inc., 2002.

[223] S. Al-Ma'adeed, E. Mohammed, and D. Al Kassis. Writer identification using edge-based directional probability distribution features for arabic words. In *IEEE/ACS International Conference on Computer Systems and Applications*, pages 582–590, 2008.

[224] S. Al-Ma'adeed, A.-A. Al-Kurbi, A. Al-Muslih, R. Al-Qahtani, and H. Al Kubisi. Writer identification of arabic handwriting documents using grapheme features. In *IEEE/ACS International Conference on Computer Systems and Applications*, pages 923–924, 2008.

[225] Bence Kovari and Hassan Charaf. A study on the consistency and significance of local features in off-line signature verification. *Pattern Recogn. Lett.*, 34(3):247–255, February 2013.

[226] M. B. Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov. Offline signature verification using classifier combination of hog and lbp features. In *Int. Joint Conf. on Biometrics*, pages 1–7, 2011.

[227] Chawki Djeddi, Imran Siddiqi, Labiba Souici-Meslati, and Abdellatif Ennaji. Text-independent writer recognition using multi-script handwritten texts. *Pattern Recognition Letters*, In press, 2013.

[228] A. Hassaine, S. Al-Maadeed, and A. Bouridane. A set of geometrical features for writer identification. *Neural Information Processing. Springer Berlin Heidelberg*, 2012.

[229] Alisher Kholmatov and Berrin Yanikoglu. Identity authentication using improved online signature verification method. *Pattern Recognition Letters*, 26(15):2400–2408, 2005.

[230] Muhammad Imran Malik, Sheraz Ahmed, Andreas Dengel, and Marcus Liwicki. A signature verification framework for digital pen applications. In *DAS*, pages 419–423. IEEE, 2012.

[231] Chawki Djeddi, Labiba Souici-Meslati, and Abdellatif Ennaji. Writer recognition on arabic handwritten documents. In *ICISP*, pages 493–501, 2012.

[232] Shehzad Khalid and Shahid Razzaq. Frameworks for multivariate m-mediods based modeling and classification in euclidean and general feature spaces. *Pattern Recogn.*, 45(3):1092–1103, 2012.

[233] A. Hassaine and S. Al-Maadeed. An online signature verification system for forgery and disguise detection. *Neural Information Processing. Springer Berlin Heidelberg*, 2012.

[234] I.W. Evett. Towards a uniform framework for reporting opinions in forensic science casework. *Science & Justice*, 38(3):198–202, 1998.

[235] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Local features for forensic signature verification. In *New Trends in Image Analysis and Processing–ICIAP 2013*, pages 103–111. Springer, 2013.

[236] Peng Fan, Aidong Men, Mengyang Chen, and Bo Yang. Color-surf: A surf descriptor with local kernel color histograms. In *Network Infrastructure and Digital*

Content, 2009. IC-NIDC 2009. IEEE International Conference on, pages 726–730, Nov 2009.

[237] Jing Fu, Xiaojun Jing, Songlin Sun, Yueming Lu, and Ying Wang. C-surf: Colored speeded up robust features. In Yuyu Yuan, Xu Wu, and Yueming Lu, editors, *Trustworthy Computing and Services*, volume 320 of *Communications in Computer and Information Science*, pages 203–210. Springer Berlin Heidelberg, 2013.

[238] Paul Viola and Michael J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004.

[239] Faisal Shafait, Daniel Keysers, and Thomas Breuel. Efficient implementation of local adaptive thresholding techniques using integral images. In *Proceedings of the 15th Document Recognition and Retrieval Conference (DRR-2008)*, volume 6815. SPIE, 1 2008.

[240] Sekwon Yeom, Adrian Stern, and Bahram Javidi. Compression of 3d color integral images. *Opt. Express*, 12(8):1632–1642, Apr 2004.

[241] Matthew Brown and David Lowe. Invariant features from interest point groups. In *In British Machine Vision Conference*, pages 656–665, 2002.

[242] StephenM. Smith and J.Michael Brady. Susan a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[243] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[244] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, pages 778–792, Berlin, Heidelberg, 2010. Springer-Verlag.

[245] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the 2011 International Conference on Computer Vision*, ICCV '11, pages 2564–2571, Washington, DC, USA, 2011. IEEE Computer Society.

[246] Stefan Leutenegger, Margarita Chli, and Roland Yves Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. In *ICCV*, pages 2548–2555. IEEE, 2011.

[247] L. Michel. Disguised signatures. *Journal of the Forensic Science Society*, 18:25–29, 1978.

[248] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[249] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[250] Donato Impedovo and Giuseppe Pirlo. Stability analysis of static signatures for automatic signature verification. In *ICIAP*, volume 6979 of *LNCS*, pages 241–247. Springer, 2011.

[251] M. I. Malik, Marcus Liwicki, Andreas Dengel, and Bryan Found. Man vs. machine: A comparative analysis for signature verification. *J. Forensic Document Examination*, 24:21–35, 2014.

[252] W.R. Harrison. *Suspect documents: their scientific examination.* Praeger, New York, 1958.

[253] J.V.P. Conway. *Evidential documents.* Charles C Thomas,, Illinois, 1959.

[254] O. Hilton. *Scientific examination of questioned documents.* Elsevier Science Publishing Company, New York, 1982.

[255] D. Ellen. *The scientific examination of documents: methods and techniques.* Ellis Horwood Limited, West Sussex, 1989.

[256] A. M. Huber, R. A.and Headrick. *Handwriting identification: facts and fundamentals.* Boca Raton, CRC Press, 1999.

[257] Markus Weber, Thomas Roth-Berghofer, Volker Hudlet, Heiko Maus, and Andreas Dengel. Context-aware service discovery using case-based reasoning methods. In *KI*, volume 5803, pages 664–671. Springer-Verlag, Heidelberg, 9 2009.

[258] Bryan James Found, Douglas Kelman Rogers, et al. The initial profiling trial of a program to characterize forensic handwriting examiners' skill. *Journal of the American Society of Questioned Document Examiners*, 2003.

[259] Bryan Found and Doug Rogers. The probative character of forensic handwriting examiners identification and elimination opinions on questioned signatures. *Forensic science international*, 178(1):54–60, 2008.

[260] Sheraz Ahmed, , Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Extraction of signatures from document images for real world applications. In *Journal of the American Society of Questioned Document Examiners*, 2015.

[261] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, December 1989.

[262] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, Adam Funk, Angus Roberts, Danica Damljanovic, Thomas Heitz, Mark A. Greenwood, Horacio Saggion, Johann Petrak, Yaoyong Li, and Wim Peters. *Text Processing with GATE (Version 6)*. GATE, 2011.

[263] S. Imade, S. Tatsuta, and T. Wada. Segmentation and classification for mixed text/image documents using neural network. In *Proceedings. 2nd ICDAR*, pages 930 –934, oct 1993.

[264] K. Kuhnke, L. Simoncini, and Zs.M. Kovacs-V. A system for machine-written and hand-written character distinction. In *Proceedings. of 3rd ICDAR*, volume 2, pages 811 –814 vol.2, aug 1995.

[265] J.K. Guo and M.Y. Ma. Separating handwritten material from machine printed text using hidden markov models. In *Proceedings. 6th ICDAR*, pages 439 –443, 2001.

[266] Yefeng Zheng, Huiping Li, and D. Doermann. Machine printed text and handwriting identification in noisy document images. *TPAMI*, 26(3):337 –353, march 2004.

[267] Sukalpa Chanda, Katrin Franke, and Umapada Pal. Structural handwritten and machine print classification for sparse content and arbitrary oriented document fragments. In *Proceedings of SAC 10*, pages 18–22. ACM, 2010.

[268] Saeed Mozaffari and Parnia Bahar. Farsi/arabic handwritten from machine-printed words discrimination. In *Proceedings. ICFHR*. IEEE, 2012.

[269] Purnendu Banerjee and Bidyut Baran Chaudhuri. A system for hand-written and machine-printed text separation in bangla document images. In *Proceedings ICFHR*. IEEE, 2012.

[270] R. Jayadevan, S.R. Kolhe, P.M. Patil, and U. Pal. Automatic processing of handwritten bank cheque images: a survey. *IJDAR*, 15:267–296, 2012.

[271] M. Sankari, M. Benazir, and R. Bremananth. Verification of bank cheque images using hamming measures. In *ICARCV 10*, pages 2531 –2536, dec. 2010.

[272] Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Signature detection and matching for document image retrieval. *TPAMI*, 31(11):2015–2031, 2009.

[273] Ranju Mandal, Partha Pratim Roy, and Umapada Pal. Signature segmentation from machine printed documents using conditional random field. *ICDAR*, 0:1170–1174, 2011.

[274] Ranju Mandal, Partha Pratim Roy, and Umapada Pal. Signature segmentation from machine printed documents using contextual information. *IJPRAI*, 2012.

[275] Yi Li, Yefeng Zheng, David S. Doermann, and Stefan Jaeger. Script-independent text line segmentation in freestyle handwritten documents. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(8):1313–1329, 2008.

[276] Muhammad Imran Malik, Sheraz Ahmed, Faisal Shafait, Ajmal Saeed Mian, Christian Nansen, Andreas Dengel, and Marcus Liwicki. Hyper-spectral analysis for automatic signature extraction. In *17th Biennial Conference of the International Graphonomics Society*, 2015.

[277] T. Young. *The Bakerian Lecture: On the Theory of Light and Colours*. Bakerian lecture. Royal Society, 1802.

[278] Robert William Gainer Hunt. *The reproduction of colour*. John Wiley & Sons, 2005.

[279] G Wyszecki and WS Stiles. Color science: Concepts and methods, quantitative data and formulae. *John Wiley& Sons, New York*, 1982.

[280] Chein-I Chang. *Hyperspectral imaging: techniques for spectral detection and classification*, volume 1. Springer, 2003.

[281] D. Lorente, N. Aleixos, J. Gmez-Sanchis, S. Cubero, O.L. Garca-Navarrete, and J. Blasco. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, 5(4):1121–1142, 2012.

[282] Mihaela Antonina Calin, Sorin Viorel Parasca, Dan Savastru, and Dragos Manea. Hyperspectral imaging in the medical field: Present and future. *Applied Spectroscopy Reviews*, 49(6):435–447, 2014.

[283] Sven Schneider, Richard J. Murphy, and Arman Melkumyan. Evaluating the performance of a new classifier the gp-oad: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery. {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, 98(0):145 – 156, 2014.

[284] Patrick Shiel, Malte Rehbein, and John Keating. The ghost in the manuscript: Hyperspectral text recovery and segmentation. *Codicology and Palaeography in the Digital Age, M. Rehbein and PS und Torsten Schaßan, Eds. Norderstedt: Books on Demand*, pages 159–174, 2009.

[285] Bernard J Aalderink, Marvin E Klein, Roberto Padoan, Gerrit de Bruin, and Tedag Steemers. Clearing the image: A quantitative analysis of historical documents using hyperspectral measurements. In *Poster presented at the AIC 37th Annual Meeting*, 2009.

[286] M. Lettner and R. Sablatnig. Spatial and spectral based segmentation of text in multispectral images of ancient documents. In *10th ICDAR*, pages 813–817, July 2009.

[287] Douglas Goltz, Michael Attas, Gregory Young, Edward Cloutis, and Maria Bedynski. Assessing stains on historical documents using hyperspectral imaging. *Journal of Cultural Heritage*, 11(1):19–26, 2010.

[288] F. Hollaus, M. Gau, and R. Sablatnig. Enhancement of multispectral images of degraded documents by employing spatial information. In *12th ICDAR*, pages 145–149, Aug 2013.

[289] Rachid Hedjam and Mohamed Cheriet. Historical document image restoration using multispectral imaging system. *Pattern Recognition*, 46(8):2297 – 2312, 2013.

[290] F. Hollaus, M. Diem, and R. Sablatnig. Improving ocr accuracy by applying enhancement techniques on multispectral images. In *22nd ICPR*, pages 3080–3085, Aug 2014.

[291] Zohaib Khan, Faisal Shafait, and Ajmal S Mian. Towards automated hyperspectral document image analysis. In *AFHA*, pages 41–45, 2013.

[292] Z. Khan, F. Shafait, and A. Mian. Hyperspectral imaging for ink mismatch detection. In *12th ICDAR*, pages 877–881, Aug 2013.

[293] G Reed, K Savage, D Edwards, and N Nic Daeid. Hyperspectral imaging of gel pen inks: An emerging tool in document analysis. *Science & Justice*, 54(1):71–80, 2014.

[294] Eric B Brauns and R Brian Dyer. Fourier transform hyperspectral visible imaging and the nondestructive analysis of potentially fraudulent documents. *Applied spectroscopy*, 60(8):833–840, 2006.

[295] Aythami Morales, Miguel A Ferrer, Moises Diaz-Cabrera, Cristina Carmona, and Gordon L Thomas. The use of hyperspectral analysis for ink identification in handwritten documents. In *ICCST*, pages 1–5. IEEE, 2014.

[296] Carolina S Silva, Maria Fernanda Pimentel, Ricardo S Honorato, Celio Pasquini, José M Prats-Montalbán, and Alberto Ferrer. Near infrared hyperspectral imaging for forensic analysis of document forgery. *Analyst*, 139(20):5176–5184, 2014.

[297] Ranju Mandal, Partha Pratim Roy, and Umapada Pal. Signature segmentation from machine printed documents using conditional random field. In *11th ICDAR*, pages 1170–1174. IEEE, 2011.

[298] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417. Springer, 2006.

[299] Edward Rosten and Tom Drummond. Fusing points and lines for high performance tracking. In *ICCV*, volume 2, pages 1508–1515 Vol. 2, Oct 2005.

[300] Erika Griechisch, M.I. Malk, and Marcus Liwicki. Online signature analysis based on accelerometric and gyroscopic pens and legendre series. In *12th Int. Conf. on Document Analysis and Recognition (ICDAR)*, pages 374–378. IEEE, 2013.

[301] Erika Griechisch, M. I. Malik, and marcus Liwicki. Online signature verification based on kolmogorov-smirnov distribution distance. In *14th International Conference on Frontiers in Handwriting Recognition (ICFHR-2014)*, pages 738–742. IEEE, 2014.

[302] M. Liwicki. Evaluation of novel features and different models for online signature verification in a real-world scenario. In *Proc. 14th Conf. of the Int. Graphonomics Society*, pages 22–25, 2009.

[303] Andreas Schlapbach, Marcus Liwicki, and Horst Bunke. A writer identification system for on-line whiteboard data. *Pattern Recogn.*, 41:2381–2397, July 2008.

[304] A. Piccini, M.and Carpignano and P.C. Cacciabue. Supporting integrated design of control systems and interfaces: A human centred approach. In *Proceedings of the 8th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine-Systems*, pages 87–92, Kassel, Germany, 2001.

[305] Ronan McDonnell Julie Doyle, Zoran Skrba and Ben Arent. Designing a touch screen communication device to support social interaction amongst older adults. In *Proceedings of the 24th BCS International Conference on Human-Computer Interaction (HCI-2010)*, Dundee, Scotland, 2010.

[306] Holger Koessling, Dominic Gorecky Marcus Liwicki, Markus Weber Gerrit Meixner, and Andreas Dengel. Pen-based interaction forms for smarter product customization. In *Proceedings of the 18th International Federation of Automatic Control World Congress. World Congress of the International Federation of Automatic Control (IFAC-2011),*. IFAC, 4 2011.

[307] Horst Bunke, Jnos Csirik, Zoltan Gingl, and Erika Griechisch. Online signature verification method based on the acceleration signals of handwriting samples. In *CIARP*, volume 7042 of *Lecture Notes in Computer Science*, pages 499–506. Springer, 2011.

[308] Jnos Csirik, Zoltan Gingl, and Erika Griechisch. The effect of training data selection and sampling time intervals on signature verification. In *AFHA*, volume 768 of *Proceedings of CEUR Workshop*, pages 6–10. CEUR-WS.org, 2011.

[309] Katalin Kopasz, Péter Makra, and Zoltán Gingl. Edaq530: a transparent, open-end and open-source measurement solution in natural science education. *European Journal of Physics*, 32(2):491–504, March 2011.

[310] Pim Tuyls, AntonH.M. Akkermans, TomA.M. Kevenaar, Geert-Jan Schrijen, AskerM. Bazen, and RaimondN.J. Veldhuis. Practical biometric authentication with template protection. In Takeo Kanade, Anil Jain, and NaliniK. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 436–446. Springer Berlin Heidelberg, 2005.

[311] Haiyun Xu, Raymond N.J. Veldhuis, Tom A.M. Kevenaar, Anton H.M. Akkermans, and Asker M. Bazen. Spectral minutiae: A fixed-length representation of a minutiae set. In *Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, pages 1–6, Los Alamitos, 2008. IEEE Computer Society Press.

[312] Marianela Parodi, Juan C. Gomez, and Marcus Liwicki. Online Signature Verification Based on Legendre Series Representation: Robustness Assessment of Different Feature Combinations. In *2012 International Conference on Frontiers in Handwriting Recognition*, pages 379–384. IEEE, IEEE, sep 2012.

Muhammad Imran Malik

research highlights

Publications: 23
Citations: 181*
H-index: 8*

research interests

Machine Learning, Forensic Handwriting Analysis, Evaluation of Forensic Systems, Document Image and Signature Processing

personal information

Contact: Trippstadter Str. 121, 67663, Kaiserslautern Germany
+49 (176) 45755183
+49 (631) 20575-4818
mimalik@rhrk.uni-kl.de
iimranmalik@gmail.com

education

2011–2015 **PhD** Artificial Intelligence — **University of Kaiserslautern, Germany**
Bridging the gap b/w Pattern Recognition & Forensic Science
Focused forensic aspects of handwriting and signature analysis, development of generic signature verification methods for forensic science and evaluation of such methods so that to be acceptable for the courts/jury.

2008–2011 **Masters** of Computer Science — **University of Kaiserslautern, Germany**
Benchmarking Automatic Signature Verification Systems
Developed signature verification systems based on global features and various classifier combinations.

2000–2004 **Bachelors** of Computer Science — **Allama Iqbal Open University, Pakistan**
Real Time Object Tracking & Video Enhancement System
Developed for application in various Night Vision devices (NVDs).

fellowships and training

Sep.-Nov. 2014 **Visiting Researcher** — **University of Western Australia, Perth, Australia**
Devloped solutions for automatic segmentation and verification of information from documents using hyper-spectral imaging techniques, considering forensic casework and general purpose applications.

Nov. 2013-Jan. 2014 **Visiting Researcher** — **Kyushu University, Fukuoka, Japan**
Developed various systems for local/part-based analysis of stability of handwritten text and signatures with respect to forensic casework.

invited talks and tutorials

August 2014 **3rd Int. Workshop on Automatic Forensic Handwriting Analysis, USA**
American and Australasian Societies of Questioned Document Examiners, Hawaii, USA.
`http://www.dfki.de/afha/2014/index.html`

June 2013 **Faculty of Electrical Engineering, Kyushu University, Japan**
Institute of Electronics, Information, and Communication Engineers (IEICE), the largest ICT society in Japan. `http://www.ieice.org/~kyushu/koenkai/h25/koen_4.html`

March 2013 **Szeged University, Szeged, Hungary**
Department of Computer Algorithms & Artificial Intelligence
`http://www.inf.u-szeged.hu/~grerika/tutorial/Tutorial.pdf`

Sept. 2012 **13th Int. Conf. on Frontiers in Handwriting Recognition, Bari, Italy**
Tutorial Sessions: State-of-the-Art of Signature Processing

* Google Scholar

awards and recognitions

June 2013 **Best Student Paper Award**
16th Biennial International Graphonomics Society Conference (IGS-2013), Nara, Japan
Association of Forensic Document Examiners, USA. `http://afde.org`

June 2008 **HEC-DAAD - Scholarship**
Secured HEC /DAAD Open Merit Scholarship for Masters & PhD in Germany.

April 2005 **Topped in Bachelors Computer Science**
First position in BS (CS) in the Institute of Computer and Management Sciences, study center of Allama Iqbal Open University.

April 1997 **Topped in the Secondary School**
First position in Sadiq Public School Rawalpindi in Matriculation exam.

professional activities

- Member Leadership Board Int. Association of Pattern Recognition TC-6.
- Member Program Committee 13th Int. Conf. on Document Analysis & Recognition (ICDAR-2015).
- Organizer SigWiComp-competition at the Int. Conf. on Document Analysis & Recognition (ICDAR-2015).
- Co-chair Program Committee 4th Int. Workshop on Automatic Forensic Handwriting Analysis 2015, Nancy, France.
- Member Program Committee Int. Workshop on Image-based Smart City Applications, Genova, Italy.
- Co-chair Program Committee 3rd Int. Workshop on Automatic Forensic Handwriting Analysis 2014, Hawaii, USA.
- Chair Program Committee 2nd Int. Workshop on Automatic Forensic Handwriting Analysis 2013, Washington D.C., USA.
- Member Program Committee 6th Int. Wokshop on Computational Forensics 2014, Stockholm, Sweden.
- Member Program Committee Int. Workshop on Emerging Aspects on Handwritten Signature Processing 2013, Naples, Italy.
- Member Program Committee 1st Int. Workshop on Automatic Forensic Handwriting Analysis 2013, Beijing, China.
- Member International Graphonomics Society.
- Organizer SigWiComp-competition at the Int. Conf. on Document Analysis & Recognition 2013, Washington D.C., USA.
- Organizer 4NSigComp-competition at the Int. Conf. on Frontiers in Handwriting Recognition 2012, Bari, Italy.
- Organizer SigComp-competition at the Int. Conf. on Document Analysis & Recognition 2011, Beijing, China.
- Evaluator 4NSigComp-competition at the Int. Conf. on Frontiers in Handwriting Recognition 2010, Kolkata, India.
- Frequent Reviewer for journals; IEEE Trans. on Human Machine Systems, Computer Vision & Image Understanding, Pattern Recognition, Neurocomputing, Pattern Recognition Letters, IET Biometrics.

references

Prof. Andreas Dengel
Professor
Kaiserslautern
University of
Technology, Germany
andreas.dengel@dfki.de

Prof. Marcus Liwicki
apl.-Professor
Kaiserslautern
University of
Technology, Germany
liwicki@cs.uni-kl.de

Prof. Faisal Shafait
Assistant Professor
University of Western
Australia, Australia.
faisal.shafait@uwa.edu.au

- Frequent Reviewer for refereed int. conferences and workshops including; CVPR, ICDAR, ICPR, ICFHR, DAS, DRR, CBDAR, GCPR, KIS, AFHA, EAHSP.

publications

[1] Muhammad Imran Malik, Sheraz Ahmed, Faisal Shafait, Ajmal Saeed Mian, Christian Nansen, Andreas Dengel, and Marcus Liwicki. Hyper-spectral analysis for automatic signature extraction. In *17th Biennial Conference of the International Graphonomics Society*, 2015.

[2] Muhammad Imran Malik, Marcus Liwicki, Andreas Dengel, and Bryan Found. Man vs. Machine: A comparative analysis for offline signature verification. *Journal of Forensic Document Examination*, 24:21 –35, Oct 2014.

[3] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Local features for off-line forensic signature verification. In *Advances in Digital Handwritten Signature Processing: A Human Artefact for e-Society*, pages 95–110. World Scientific, 2014.

[4] Marcus Liwicki, Muhammad Imran Malik, and Charles Berger. Towards a shared conceptualization for automatic signature verification. In *Advances in Digital Handwritten Signature Processing: A Human Artefact for e-Society*, pages 65–80. World Scientific, 2014.

[5] Muhammad Imran Malik, Marcus Liwicki, Andreas Dengel, Seiichi Uchida, and Volkmar Frinken. Automatic signature stability analysis and verification using local features. In *14th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Crete, Greece*, pages 621–626. IEEE, 2014.

[6] Erika Griechisch, Muhammad Imran Malik, and Marcus Liwicki. Online signature verification based on kolmogorov-smirnov distribution distance. In *14th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Crete, Greece*, pages 738–742. IEEE, 2014.

[7] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. In *12th Int. Conf. on Document Analysis and Recognition (ICDAR), Washington D.C., USA*, pages 872–876. IEEE, 2013.

[8] Muhammad Imran Malik, Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel. FREAK for real time forensic signature verification. In *12th Int. Conf. on Document Analysis and Recognition (ICDAR), Washington D.C., USA*, pages 971–975. IEEE, 2013.

[9] Muhammad Imran Malik, Marcus Liwicki, Linda Alewijnse, Wataru Ohyama, Michael Blumenstein, and Bryan Found. ICDAR 2013 competitions on signature verification and writer identification for on-and offline skilled forgeries (sigwicomp 2013). In *12th Int. Conf. on Document Analysis and Recognition (ICDAR, Washington D.C., USA)*, pages 1477–1483. IEEE, 2013.

[10] Erika Griechisch, Muhammad Imran Malik, and Marcus Liwicki. Online signature analysis based on accelerometric and gyroscopic pens and legendre series. In *12th Int. Conf. on Document Analysis and Recognition (ICDAR, Washington D.C., USA)*, pages 374–378. IEEE, 2013.

[11] Muhammad Imran Malik, Marcus Liwicki, Andreas Dengel, and Bryan Found. Man vs. machine: A comparative analysis for forensic signature verification. In *16th Biennial Conf. of the Int. Graphonomics Society, Nara, Japan*, pages 9–13. Int. Graphonomics Society, 2013.

[12] Sheraz Ahmed, Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Towards signature segmentation & verification in real world applications. In *16th Biennial Conf. of the Int. Graphonomics Society, Nara, Japan*, pages 139–142. Int. Graphonomics Society, 2013.

[13] Erika Griechisch, Muhammad Imran Malik, and Marcus Liwicki. Online signature verification using accelerometer and gyroscope. In *16th Biennial Conf. of the Int. Graphonomics Society, Nara, Japan*, pages 143–146. Int.l Graphonomics Society, 2013.

[14] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Local features for forensic signature verification. In *New Trends in Image Analysis and Processing (ICIAP), Naples, Italy*, pages 103–111. Springer, 2013.

[15] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Part-based system for forensic signature verification involving disguised signatures. In *5th Int. Workshop on Computational Forensics (IWCF), Tsukuba, Japan*, pages 70–77. 21st Int. Conf. on Pattern Recognition (ICPR), 2012.

[16] Muhammad Imran Malik and Marcus Liwicki. From terminology to evaluation: Performance assessment of automatic signature verification systems. In *13th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Bari, Italy*, pages 613–618. IEEE, 2012.

[17] Marcus Liwicki, Muhammad Imran Malik, Linda Alewijnse, Elisa van den Heuvel, and Bryan Found. Icfhr 2012 competition on automatic forensic signature verification (4nsigcomp 2012). In *13th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Bari, Italy*, pages 823–828. IEEE, 2012.

[18] Sheraz Ahmed, Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Signature segmentation from document images. In *13th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Bari, Italy*, pages 425–429. IEEE, 2012.

[19] Muhammad Imran Malik, Sheraz Ahmed, Andreas Dengel, and Marcus Liwicki. A signature verification framework for digital pen applications. In *10th IAPR Int. Workshop on Document Analysis Systems (DAS, Gold Coast, Australia)*, pages 419–423. IEEE, 2012.

[20] Marcus Liwicki, Muhammad Imran Malik, C Elisa van den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. Signature verification competition for online and offline skilled forgeries (sigcomp2011). In *11th Int. Conf. on Document Analysis and Recognition (ICDAR), Beijing, China*, pages 1480–1484. IEEE, 2011.

[21] Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Evaluation of local and global features for offline signature verification. In *Int. Workshop on Automated Forensic Handwriting Analysis, ICDAR-2011, Beijing, China*, pages 26–30, 2011.

[22] Marcus Liwicki and Muhammad Imran Malik. Surprising? power of local features for automated signature verification. In *15th Biennial Int. Graphonomics Society Conference, Live Aqua Cancun, Mexico*, pages 18–21. Int. Graphonomics Society, 2011.

[23] Marcus Liwicki, C Elisa van den Heuvel, Bryan Found, and Muhammad Imran Malik. Forensic signature verification competition 4nsigcomp2010-detection of simulated and disguised signatures. In *12th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Kolkata, India*, pages 715–720. IEEE, 2010.