UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

INSTITUTO UNIVERSITARIO DE CIENCIAS Y TECNOLOGÍAS CIBERNÉTICAS

PROGRAMA DE DOCTORADO

DOCTORADO EN CIBERNÉTICA Y TELECOMUNICACIÓN

TESIS DOCTORAL

Writer identification using online handwritten passwords

AUTOR: D. Tobias H. Kutzner

DIRECTORES: Dr. D. Carlos Manuel Travieso González
Dra. Dña. Ingrid Bönninger

El Director **El Codirector** **El Doctorando,**

Las Palmas de Gran Canaria a

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Acknowledgements

I would like to thank my thesis advisors Prof. Dr. Carlos Manuel Travieso González from the Institute for Technological Development and Innovations (IDeTIC) Signals and Communications Department at ULPGC and Prof. Dr.-Ing. Ingrid Bönninger from the Institute of Medical Technology at BTU Cottbus – Senftenberg, for the opportunity to write this PhD thesis and for their excellent support all the time.

I would like to thank the professors of the Institute of Medical Technology at BTU for their support.

Furthermore, I would like to thank Prof. Dr. rer. nat. habil. Olga Wälder for support in special mathematical and statistical investigations.

I would like to thank Dr. Nina Wolfeil and Mr. Michael Mannel from International Relations Office at BTU Cottbus – Senftenberg for the renovation of the agreement between BTU and ULPGC and the possibility of ERASMUS stay.

Last but not least, I would like to thank my family for their support and understanding of the many leisure time I have sacrificed for this thesis.

Abstract

The aim of the thesis is to investigate the enhanced safety of authentication systems through the use of handwritten secure password. In state of the art we find some online and a lot of offline text and signature verification and recognition systems, but no system that use a handwritten secure password for authentication. Therefor in this thesis it will be investigated to increase safety through the use of a handwritten secure password for authentication systems. The improvement of results in writer recognition by increasing the safety of handwritten password and exploring the possibilities of writer recognition of short texts, such as passwords under practical conditions shall be examined. For this purpose, both available public databases as well as own databases with handwritten datasets will be used / produced as research basis. As input device a device with a touch screen display (smart phone, tablet) shall be used. The preprocessing of the data, extracting the features and the classification is investigated and applied. Research for a suitable segmentation as a function of the data format will be examined. The usefulness of new features will be tested in the databases considering different standard machine learning –based methods for feature selection and/or classification. Not only theoretical investigation with popular data mining tools will be applied. The system will be tested in a real-world application using a prototype that will be developed on Java for handwriting password verification and writer identification.

Contents

List of Figures

Figure **Page**

List of Tables

Table **Page**

Chapter 1 A Review of Biometric: Hypothesis

1.1 Introduction

In the last few years, the importance of IT security has increased significantly, and has been brought into the minds of people, not least through cyber-attacks on particularly sensitive facilities. Making existing systems even safer is an ongoing task. To safeguard IT networks and infrastructure, secure authentication has the top priority, while at the same time user-friendliness becomes more and more important. Chapter 1 of this thesis explains the motivation for the subject, and gives an overview of the current access systems divided into non-biometrical and biometrical systems and their evaluation. The previous related work of the author is shown. Finally, the hypothesis will be presented, which will be confirmed by this thesis, and the structure of the thesis will be described.

1.2 Motivation

In the digital century, digital authentication systems are becoming increasingly popular. A few years ago password or pin codes were the only way to get access to software systems or get money from the bank account. In the last years, biometric systems have been established more and more to increase the security in public and safety areas. Different kinds of biometric modalities have been implemented and many references can be found in this area, because the biometric market has been growing and the use of biometric applications is demanded in security area [1]. Now we are increasingly using cards with a chip or magnetic stripe, terminals or smartphones with corresponding apps to scan QR[1] codes and carry out NFC[2] communication for payment or access. In countless areas of everyday life an authentication is necessary for example: to get cash in a bank, for cashless payment, in the internet or supermarket, or simply to get access to buildings and rooms.

[1] Quick Response

[2] Near Field Communication based on RFC radio frequency identification

Nowadays it is usually necessary to enter a pin at a point of the process or to sign to complete the process which is state of the art, but cannot be regarded as very safe. Access using biometric information such as fingerprint, digital signature, and iris scan or speech recognition are currently the safest methods, which were applied in addition to many research projects [4].

1.3 Type of access

There are different ways to get access to buildings and rooms, for authentication to secure systems i.e. for payment and so on. The main kind of access are divided into two types: non-biometrical- and biometrical systems.

1.3.1 Non-biometrical systems

The following paragraph gives a short representation of the most important non-biometrical systems, PIN, QR- and barcode, NFC, and BLE.

a) PIN:

A personal identification number (PIN) or secret number is a number known only to one or a few persons, with which they can authenticate themselves against a machine. The redundant acronym PIN number or the PIN code are often used. In the narrower sense, PINs are numeric passwords [1].

In the case of the chip less card, this processing is carried out in a protected environment only after reading the data from the card; in the case of cards with a chip, this processing additionally provides a contribution protected by the connection to the reader.

A common application for PINs is authentication on a cash machine. This requires the input of a minimum four-digit number sequence to prevent or at least complicate account access by unauthorized persons. Also you can pay with the bank card and the associated PIN in many shops cashless.

Also for the Internet banking is usually a PIN necessary. With this PIN and the account data you can see his account, the money and the last bookings. With a TAN you can then make a transfer or other banking transactions.

PINs are also used to protect mobile phones from unauthorized use and in many other areas of application where a minimum of security is required. SIM cards for mobile phones are delivered with a PIN, PIN2, PUK and PUK2. All codes are stored on the SIM card. PINs are changeable, PUKs are not. The PUKs are used to unlock blocked PINs. The PIN2 is used to change special, often paid services [1].

© Bubbles / John Powell

Figure 1.1: Enter PIN Bank, Restaurant and Supermarket [2][3]

To introduce an overview of the current mobile payment procedures, here the actual most important: QR-Code, NFC and BLE[3]

b) QR- and barcode-based solutions:

The square QR codes and barcodes are versatile - both are used on-and off-line. A variety of procedures are currently under discussion at the stationary cash desk. There is on the one hand the possibility that the cash terminal displays a code that the customer scans. The payment is then executed with the pre-stored account information. Among other things, PayPal and the Otto Group with Yapital as well as some start-ups offer this QR code based option. The second variant, with the creation of a code on the customer's smartphone and the scan through the checkout, is offered by companies like PayCash (based on QR codes) and Paymey (based on barcodes)[4].

In [5] an approach to improving the insurance premium payment process is presented. QR code and the web-service technology are applied for developing the application on smart phones that can facilitate the customers for more convenient channel of premium payment. The approach will prevent the company from the loss of business opportunity caused by the disclosure of customer information to the business competitors.

In [6] a mobile payment system based on 2-Dimentional (2D) barcode for mobile users is presented and system architecture is discussed, design and implementation of the proposed mobile payment solution, as well as QR 2D barcode based security solutions. Unlike other existing mobile payment systems, the proposed payment solution in [6] provides distinct advantages to support buy-and-sale products and services based on 2D Barcodes. Unlike other

[3] Bluetooth Low Energy

mobile payment systems and solutions, the mobile payment system based on the QR code has the following distinctive features. a) It can be a two-dimensional bar code identification of the product which provides payment services to buy and sell, enable all goods and products with QR barcode identification to trade anywhere and at any time. b) The mobile payment system can be easily used in the case of not updating equipment. c) To improve the mobile user experience in the mobile payment, reduce user input, and it is easy to support after sale services, such as product delivery and pick-up and customer verification. d) Mobile phone payment transaction security can be improved.

c) NFC-based solutions:

NFC as a possibility of the contactless exchange of data by radio chip over a few centimeters also has several applications. For example, the chip may be located on a card, in a smartphone or as an intermediate solution on a sticker affixed to the telephone. This chip can be used to store data on payment processing - similar to a card payment. The advantage over the "normal" card is that the data is exchanged by radio. This eliminates the need to plug the card into a terminal. Several market participants offer their own solutions, including the credit card companies MasterCard with PayPass and Visa with payWave, Deutsche Telekom, Telefónica Germany and Vodafone with the cooperation project mpass (as well as their own solutions) and Google with its Wallet [4].

The money card of the German banks and “Sparkasse” banks was also made contactless as a new "Girogo" variant. However, this must be charged with a cash amount (prepaid solution) before use - similar to a "normal" wallet. This approach, which is rather impractical for customers, is to be supplemented by a contactless girocard payment. This non-contact debit card payment would resemble NFC-based solutions such as MasterCard's PayPass and PayWave credit cards as well as visa with PayWave from the expiration date: amounts up to 20 or 25 euros should be sent without contacting the card and without pre-charging -Solution) [4].

In [7] a new protocol destined to secure NFC mobile payment transactions between NFC smartphones and payment terminals is introduced. It allows to solve EMV security weaknesses by enhancing the classical EMV exchanged messages and adding a new security layer. It ensures: mutual authentication and non-repudiation, integrity and confidentiality of banking information, the validity of banking data that are not revoked. The protocol correctness is successfully analysed using the Scyther tool.

In [8] cloud-based NFC payment architecture for small traders is introduced: which allows them to benefit from their smartphones integrating NFC technology for use directly as NFC merchant payment devices, without needing to buy an external NFC payment terminal. In addition, the proposal introduces a new protocol aiming to secure NFC payment transactions.

d) BLE-based solutions:

The latest technology for payment in a stationary business is the communication of the customer with the trader via the wireless technology BLE. Compared to NFC, BLE has a higher range of about 10 meters. Smartphones within the radius of a BLE transmitter, also called "beacon", can be identified and used for payment. PayPal as well as Apple use this possibility in first field trials, to let the visitor settle the bill by half year. In the first half-year 2017, PayPal plans to cooperate with large and some smaller retailers to test the Beacon payment service on the German market. If the seller has the Bluetooth device in use and the customer installed the PayPal app, then searching the hand in or handbag after the smartphone or the credit card is unnecessary. The customer has to say at the checkout only that he wants to pay with PayPal and receives a receipt by mail: Because the mobile app connects to beacon automatically when entering the store, the process is handled without further action. The seller automatically gets specific information about the buyer, for example, user name and photo [4].

1.3.2 Biometrical systems

Let us take a step back to the origin of science dealing with biometrics as a science, which deals with measurements of living creatures and the necessary measuring and evaluation procedures.

> "Biometric or biometrics (also biometrics - of ancient Greek βίος bíos "life" and μέτρον métron "measure, scale") is the study of the measurement of living bodies based on biometrics, special security procedures have been developed in the area of communication and information technology that are aimed at the detection of biometric data. The biometric data of humans affect all body- physiological characteristics and behavioral structures." [9]

A definition:

Depending on the application area, there are different detailed definitions. In 1841, Christoph Bernoulli was one of the first scientists to use the term biometry in a very literal interpretation for the measurement and statistical evaluation of human life [9].

The concept of biometry has the two facets of biometric statistics and biometric recognition methods, which are also separated in practice.

Biometric statistics are concerned with the development and application of statistical methods for the evaluation of all kinds of living beings. It is used intensively by all life sciences. Karl Pearson (1857-1936) was a pioneer of scientific methods. In this context, biometrics is also used as a synonym for biostatistics[9].

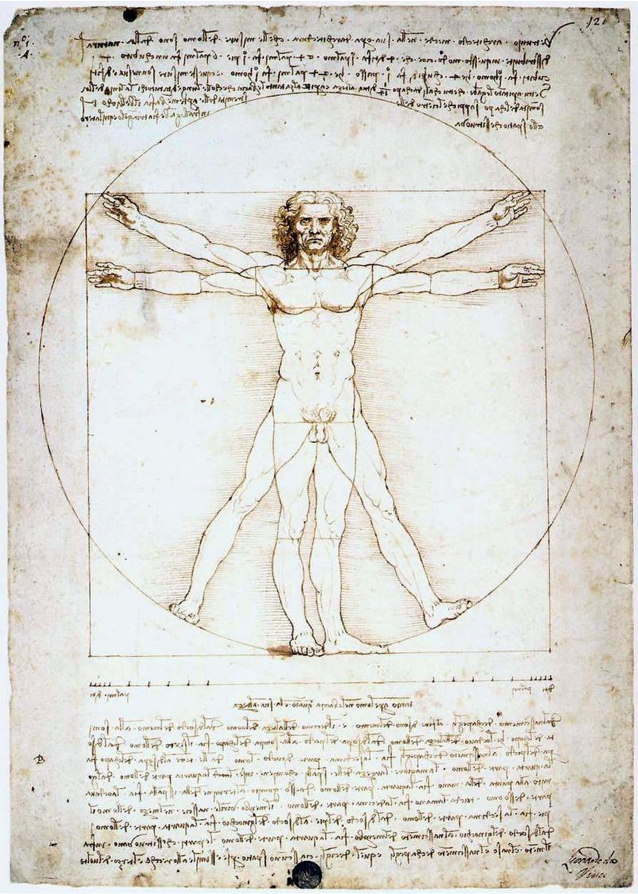


Figure 1.2: Leonardo da Vinci:The Vitruvian man [9]

Biometrics for the identification of persons were already used as identification methods. In 1879 Alphonse Bertillon developed a system for the identification of the identity, which was later called Bertillonage, based on 11 body lengths (anthropometry). In 1892 Francis Galton laid the scientific foundation for the use of the fingerprint (Daktyloskopie) [9].

Today, biometrics in the field of person recognition is also defined as the automated recognition of individuals, based on their behavioral and biological characteristics.
Further fields of application of biometrics are, for example, automated disease diagnosis methods. Biometrics lives from the interplay of the disciplines of life sciences, statistics, mathematics and computer science. Only today's information technology makes it possible to cope with the high computing power requirements of conventional biometric methods [9].

Physiological features include the iris and retina, the fingerprint, veins and facial recognition, hand geometry, ear shape, smell and blood count. In addition to the above-mentioned body-specific static features, there are also movement and behavioral features. Among these dynamic features are the signature dynamics, the tip behavior, the voice recognition, the movement of the lips and the human gait [13].
We divide biometric methods in physical traits such as face, fingerprint, and iris. They are very unique to every individual and are stable over an extended period of time. Hence, biometric systems, which are based on these traits, are usually accurate and reliable enough for identification purposes that involve one to many comparisons [10] -[11].

On the other hand, behavioral traits such as voice, gait, and signature may be susceptible to changes over time [10] and can be skillfully mimicked by impostor [12]. Thus designing an accurate behavioral based biometric system is a challenging task.

Let's take a closer look at today's most used verification methods:

a) Fingerprint:

Manufacturers of biometric systems also use the fingerprint, which is read mostly optically or electrically (eg capacitively), to identify legitimate unauthorized users. In order to prevent access to imitated fingerprints, temperature and pulse sensors can be integrated into the detection devices, which check whether a living finger has been placed on the device ("life detection"), which is only visibly effective. However, since capturing the fingerprint is reminiscent of a sovereign measure, this system is not popular with all users, which is why alternative biometric recognition systems are often used [14].

For extracting the minutiae, a special algorithm is used, by which the minutiae are put into a mathematical form. From the image provided by the fingerprint scanner, specific data is collected for each fingerprint, which is sufficient for learning-in or later comparison with existing fingerprint data. A concrete fingerprint can no longer be reconstructed from the minutia data [14].

The security of fingerprint systems is relatively small, since a fingerprint is easy to reproduce. The fingerprint sensors installed in mobile devices offer comfort in comparison to the password or pattern input, but only conditionally improve the security. The hardware hacker starbug was able to overcome the biometrics mechanism in 2014 just a few days after the appearance of the iPhone 5s, the first Apple device with touch ID [14].

For authentication, several minutiae are compared with existing reference data. Fingerprints are compared with the biometric fingerprinting method (dactyloscopy) in order to identify persons clearly. A biometric fingerprint can be used as an additional factor for two-factor authentication in computer networks, such as the open UAF standard of the FIDO alliance [14].

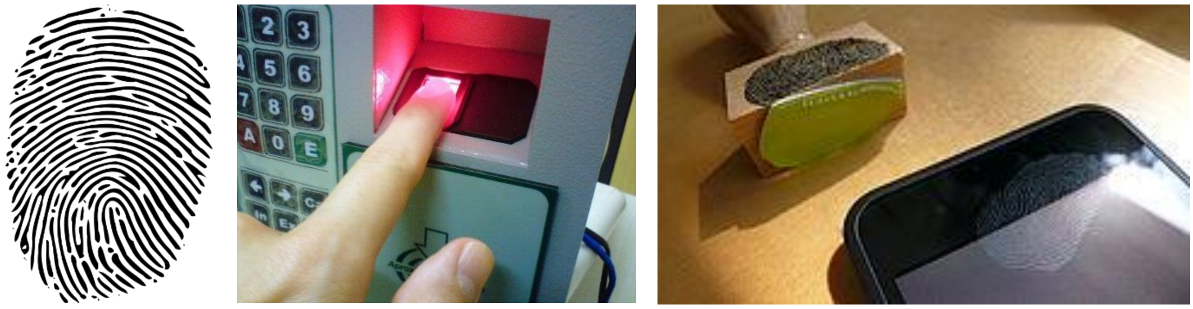


Figure 1.3: Fingerprint, Sensor, Forger Stamp iPhone [14]

b) Face:

Face recognition is the analysis of the appearance of visible features in the area of the frontal head given by the geometrical arrangement and texture properties of the surface.

It is necessary to distinguish between the localization of a face in the image and the assignment of the face to a particular person. In the first case, it is examined whether and where a face is to be seen, in the second, to whom it is concerned.

If we are concerned with recognizing the face in the sense of knowing what face it is, then one can distinguish two cases: If this is done by humans, the face of speech is spoken in the English-speaking space, while facial recognition is designated by machines as facial recognition [15].

2D method:

Simple facial recognition techniques use a two-dimensional (2D) geometric survey of particular features (e.g., eyes, nose, mouth). In this case, the position, distance and position thereof are determined. However, today's methods usually rely on complex calculations such as wavelet analysis (e.g. by means of Gabor transformation) or main component analysis. The National Institute of Standards and Technology (NIST) has repeatedly carried out comparative studies of various commercial and university procedures. The results show a clear increase in the detection performance within approximately 10 years. If the false rejection rate at a false acceptance rate of 0.1% in 1993 was still 79% in practice (i.e. almost four out of five people were not recognized at the time), this error rate is now (as of mid - 2006) from the most powerful to only 1 % (i.e. about one hundred people are not recognized). This rate is of the same magnitude as the current fingerprint or iris recognition methods and surpasses the faculties of human facial recognition [15].

In 2001, two computer scientists developed the Viola-Jones method for facial recognition named after them. The method is based on machine learning, also recognizes structures of a different kind, such as traffic signs for autonomous driving [15].

3D method:

In addition to two-dimensional biometric facial recognition, which is used to capture commercially available cameras, a new branch has been developed which is based on the three-dimensional (3D) detection (e.g. by strip projection) of the face. The additional information is intended to achieve higher recognition accuracy, better post-independence and over-reliance. TEST results from the NIST show that, as of mid-2006, the 2D methods are still superior to the 3D method with respect to the recognition performance [15].

In October 2016 it became known that 117 million Americans were in the facial recognition database of the FBI [15].

A recognized face can be used as a biometric factor for authentication.

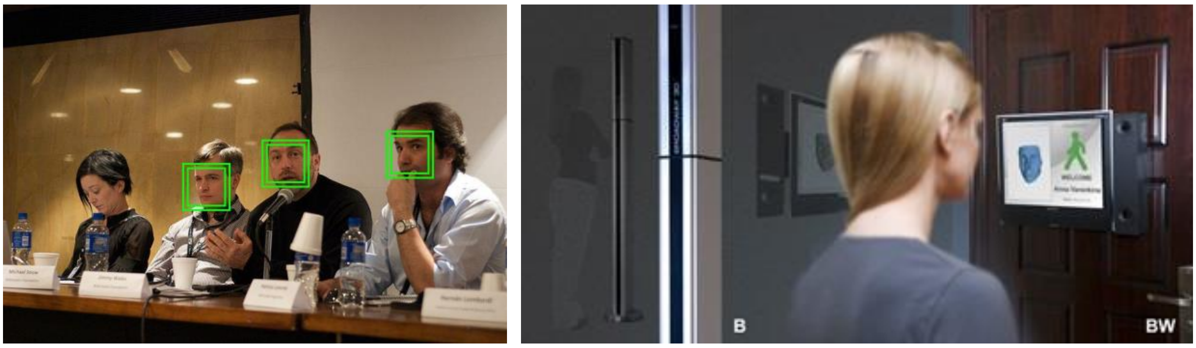


Figure 1.4: Biometric Face detection and Access Control [15][16]

c) Iris:

Iris recognition is a method of biometrics for the purpose of authenticating or identifying people. For this purpose, images of the iris (rainbow skin) of the eye are recorded with special cameras, the characteristic features of the respective iris are identified with algorithmic methods, converted into a set of numerical values (feature vector) and stored for the recognition or with a Or several templates already saved.

The original concept of using iris images for biometric recognition was developed and patented by Flom and Safir in 1987. The expiration of the patent in 2006 has led to increased research efforts worldwide [17].

The most widely used method and template (as of April 2007) in commercial application is the Iriscode based on algorithms of the mathematician John Daugman [17].

The US company EyeLock has been offering the first Iris scanner with USB port compatible with the U2F protocol of the FIDO alliance since January 2015 [17].

Commercial recognition methods capture about 260 individual optical features of the iris. These characteristics develop from a randomly controlled, morphogenetic process in the first months of life of a person and remain largely unchanged over the remaining lifetime. Even identical twins do not have an identical iris structure. The outstanding property of iris recognition in practical application is its extremely low number of false-positive comparison results compared to other biometrics methods, i.e., the likelihood of confusion of one iris with that of one eye of another person is almost zero. As a result, iris detection is a reliable identification method even in large databases with millions of personal data records, as well as for identification in access control situations without a primary recognition feature, i.e. without the use of identification cards or RFID tags [17].

False-negative (false-non-matched) results, i.e. instances of the non-recognition of a person actually detected, can occur, in particular in unfavorable conditions of the eye's eye, when the iris is caused by spectacle rims, reflections on eyeglass lenses or the typical ones Narrow eyelids is only inadequately visible [17].

A further characteristic is the low requirement for computational resources for the iris comparison. Therefore, the iris detection and recording is particularly suitable for mobile use in PDA-sized devices [17].

Together with facial and fingerprint recognition, iris recognition is one of the biometric forms provided by the ICAO for use in electronic passports (ePass). In order to ensure the worldwide, manufacturer-independent interoperability of the data, the standard ISO / IEC 19794-6 "exchange format based on iris images" specifies the requirements for iris image acquisition and storage that are applicable for this purpose [17].

For irrigation, the iris recognition is only suitable to a limited extent, since its iris structures disintegrate just a few minutes after the death of a person.

In mobile phones, iris recognition is used to unlock the phone instead of a PIN code or a fingerprint. This technology was first used in the Fujitsu ARROWS NX F-04G. Also the mobile phones of the Lumia 950 series from Microsoft use this technique of authentication [17].

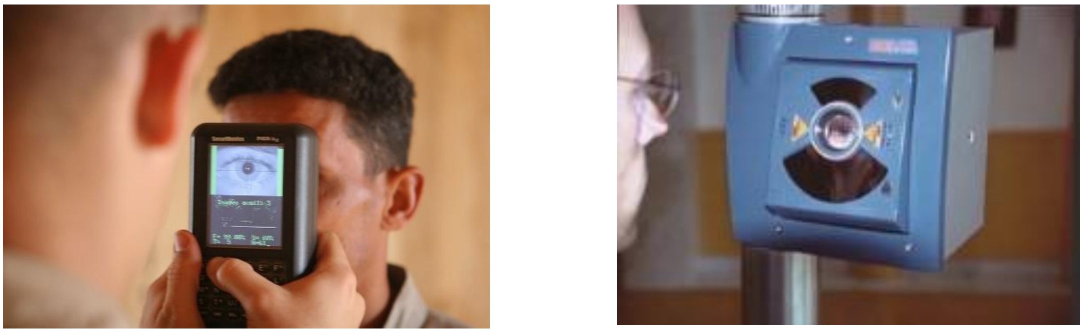


Figure 1.5: Iris detection and authentication [17][18]

d) Speech/Voice:

The speech recognition or also automatic speech recognition is a sub-area of applied computer science, engineering sciences and computer linguistics. It deals with the investigation and development of methods that make automata, in particular computers, the spoken language of the automatic data acquisition accessible. The speech recognition is to be distinguished from the voice or speech recognition, a biometric method for person identification. However, the realizations of these methods are similar [19].

At present, roughly two types of speech recognition can be distinguished:

- Speaker-independent speech recognition
- Speaker-dependent speech recognition

Characteristic for the speaker-independent speech recognition is the property that the user can immediately start the speech recognition without a previous training phase. The vocabulary, however, is limited to a few thousand words [19].

Speaker-dependent speech recognizers are trained by the user before their use (in newer systems: during use) on their own peculiarities of pronunciation. A central element is the individual interaction with the system in order to achieve an optimal speaker-dependent result (own terms, abbreviations, abbreviations, etc.). It is therefore not useful for applications with frequently changing users (e.g., call centers). In comparison, the vocabulary is much larger than that of the non-speaker recognizers [19].
Thus, current systems contain more than 300,000 word forms. It is also necessary to distinguish between:

Front-end systems and Back-end systems [19].

In front-end systems, the processing of the language and implementation into text takes place directly so that they can read the result practically without a significant time delay. The implementation can be done on the user's computer or cloud-based. Through the direct interaction between user and system, the highest recognition quality is achieved here. Likewise, the system can be controlled via commands and integration of other components such as real-time assistance systems. In back-end systems, however, the implementation is carried out in a time-delayed manner. This is usually done on a remote server. The text is only available with a delay. Such systems are still common in the medical field. Since there is no direct interaction between the speaker and the recognition result, an outstanding quality is only to be expected if the user already has experience with speech recognition [19].

Besides the size and flexibility of the dictionary, the quality of the acoustic recording plays a decisive role. Microphones that are placed directly in front of the mouth (for example, headsets or telephones) achieve a significantly higher detection accuracy than when the room microphone is more distant [19].

The development of speech recognition is proceeding very fast. Today (as of 2016), speech recognition systems are used, among other things, in smartphones such as Siri, Google Now, Cortana and Samsung's S Voice. Current speech recognition systems no longer have to be trained. The plasticity of the system is decisive for a high degree of accuracy beyond everyday language. In order to meet high demands, professional systems offer the user the possibility to influence the personal result by prescribing or auditioning [19].

The voice recognition is used today (as of 2016), e.g. in Siri, Google Now, Cortana, Amazon Echo / Alexa and Samsung's S Voice. With the high level of reliability in the everyday language (e.g. smartphones) or in the professional language (individualized professional systems), language to text can be converted, commands and controls can be carried out (command and control) or semantic analyzes (language understanding) [19].

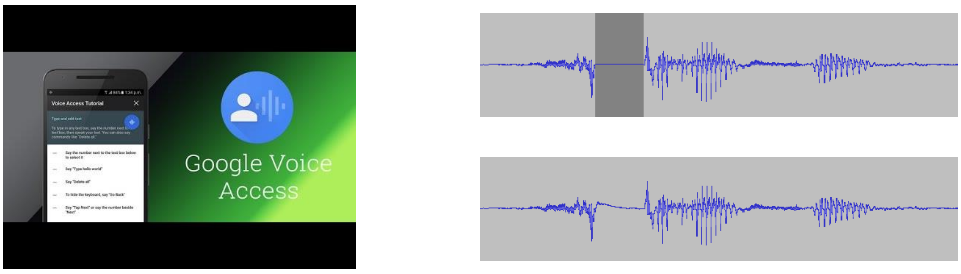


Figure 1.6: Speech/Voice detection access [19][21]

e) Signature:

The term "handwriting recognition" refers to all procedures which automatically recognize handwritten letters, digits, words or sentences and transform them into a file to be processed for the computer. Handprint Character Recognition (HCR) is an intelligent character recognition (ICR) and optical character recognition (OCR) has emerged [22].

Handwriting recognition is divided into off-line and online recognition.

For many purposes, like recognition of postal code and analysis of manuscripts, off-line recognition is an adequate method. Since manuscripts are characterized by many individual and national characteristics and are subject to the human psyche, the recognition methods analyze individual signs, but also the writing process and the writing speed. For these recognition methods online systems are necessary. The movement and character characteristics are stored and compared in an internal dictionary. With this recognition, detection rates of well over 90% are achieved [22].

Signatures are among the dynamic biometric features. The way in which people sign, how the print distribution runs, with which acceleration and speed the entire lettering and individual passage or letters are written, the steepness of the letters and the length of the interruptions within the signature are characteristic features which are not copied by other persons Can be. While the acceptance of systems for character recognition is very high, their invariability is not good, and the uniqueness is also relatively low at about 1: 10.000 [22].

Graphic tablets or touch screens are used for the recording of the signature or for the letter recognition. Since only the static characteristics of the signature are recorded, special pressure sensors with built-in sensors are used for the dynamic features [23].

f) Keyboard Dynamics / Movement:

Tip biometrics is based on the recognition that the typing behavior of each person on a keyboard is as individual as his manuscript. This means that every computer user can be recognized by his unique typing behavior [27].

Numerous characteristics, such as the duration, writing rhythm, or speed, are evaluated in the tip analysis. Thanks to a highly developed, sophisticated technology, reliable user recognition is ensured and the copying of a strange tip behavior is ruled out [27].
The application of tip biometry is conceivably simple in practice. Unlike the password login, the user does not need to remember anything: He types a short set and immediately gets access to his data. Any other user is rejected, because his profile does not match the one stored. The text to be typed appears open on the login screen. It does not have to be kept secret. What is decisive is not what the user types, but how he types - and even an attentive observer cannot imitate this [27].

Of course, the user must first train his / her profile, which usually takes two to three minutes. He/she types the login text a few times, and the recognition system generates the user's tip profile on this basis. He can then log in immediately using tip biometrics. The intelligent detection software updates this profile every time you log in successfully to adapt to long-term changes in the typing behavior. Short-term, everyday fluctuations in the typing behavior - for example, if the user types a little slower in the morning - tolerates the process without losing any separating sharpness [27].

Tip biometrics benefits from the fact that it can already look back on a long history. The discovery of the typing behavior as a sign of recognition already took place at the end of the 19th century in the time of telegraphing. In the First World War enemy movements of the enemy were identified in the Morse behavior of the radio operators. In the second world war British agents with their deposited Type Profile verified themselves. In the eighties of the last century, the idea emerged to secure computer accesses with tip biometrics. However, the advancement of technology should still prove to be a long way to go until the process achieved a high level of application [28].

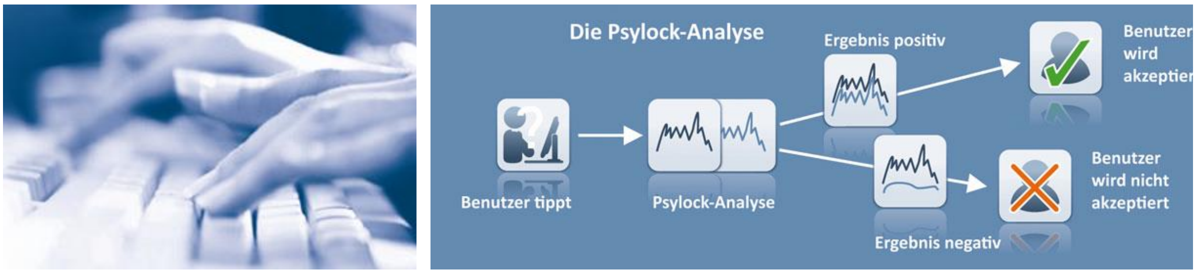


Figure 1.7: Keyboard Dynamics “Tipp Biometrie Psylock Analysis” [27][28]

A look at the table (Table 1.1) shows the various verification procedures with used sensors and the advantages and disadvantages of these methods:

Table 1.1: Biometric recognition features and their safety [13]

Biometric features	Sensors	Weaknesses
Fingerprint Hand Geometry	Sensor Chip Optical Scanners	Dirty or injured fingers Diseases
Sound/Voice Detection	Microphone	Background noise Disease-related changes
Face recognition Retina recognition Iris recognition	Camera Special camera Infrared laser	Clothing and weather dependent Diseases or injuries of the eye
Signature Keyboard dynamics Movement sequence	Signature Pad, Tablet, Smartphone, Terminal keyboard Camera	Mental and disease-related changes

Table 1.2 give a short evaluation of some of the mentioned verification methods [29]:

Table 1.2: Technical specifications for biometric systems / methods [29]

Characteristic	Template size (bytes)	Verification / registration time (sec)	False Acceptance Rate (FAR (%))	False Rejection Rate (FRR (%))
Finger image (minutiae)	900-1.200	0,5-20 / 10-30	0,01-0,0001	1,0-5,0
Hand geometry	10-20	2-5/-	0,1-5,0	0,2-5,0
Iris	< 512	0,5-10/-	0,01-1,0	0,1-2,0
Retina	40-96	> 1,5/< 30	0,0001	< 12
Face	< 1.300	1-5/< 30	0,5-2,0	1,0-3,0
Signature / Handwriting	400-1.500	5-15/30	1,6-20	2,8-25
Voice	1.500-3.000	> 1,5/-.	-.	-

Again and again, every year, there are a lot of card frauds, attacks on bank accounts or fraud where users are forced to visibly enter their data. This raises the question, how can we make these new procedures more secure, what has proven itself to perform a secure authentication? With a simple pin, when the deceiver get it is not difficult to get access to a system.

So what to do to solve this problem? The idea of this approach which is to use biometric online authentication with an own handwritten safe password and thus verifying whether the respective person is entitled to get access to the system or not can make attacks or frauds significantly more difficult and lead to an increase in the security of any authentication process.

1.3.3 Combination of non-biometrical and biometrical systems

This approach adopts the handwriting as the basis, which has been used as a secure authentication method since the early Middle Ages. Thus signatures are found among documents, such as the Ostarrîchi document of Emperor Otto III. of 996. While in Europe since the beginning of modern times the handwritten signing before witnesses is considered as legally binding, the stamped seal (Chinese seal 印, yìn, Japanese Hanko 判子) is still the binding legally valid signature in the East Asian cultural circle. Signature stamps are also common in other countries or institutions [26]. However, the signature alone is not 100% counterfeit proof; in order to achieve a plus of security is the idea to use a handwritten secure password.

1.4 Related work

In this paragraph related work of the author is described, that motivated him to deal with the analysis of handwritten passwords.

Before the beginning of this work in 2012 the author developed a client server application for a first simple user authentication with handwritten passwords on mobile devices [60]. (see Figure 1.8) The implemented prototype in [60] executes the segmentation on the mobile device. Feature extraction and classification run as server applications. The WEKA[4] tool was used to determine the most suitable algorithm for the recognition of the passphrase. The analyzes are carried out with 280 handwritten passwords. Of these, 176 are originals and 104 forgeries. As mobile devices, a HTC Desire with Android 2.2 and a Samsung Galaxy Ace with Android 2.3.3

[4] WEKA Datamining Framework from University of Waikato

were used. The proposed system recognizes the writer of the password, with a probability of 96.87%, the writer of the password, the false acceptance (FAR) rate was 12.5% when eight passwords of each writer are used for the classification model.

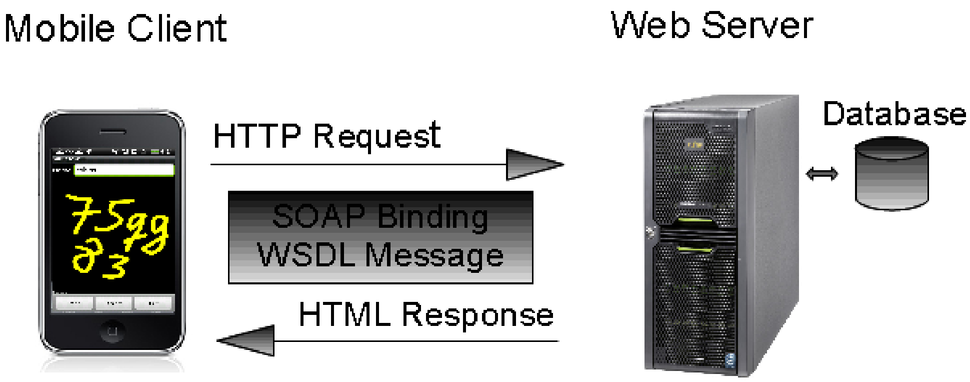


Figure 1.8: Handwriting recognition as a client-server solution [123]

The proposed systems have two modes: The Enrol Mode for collecting data and building the model. The Test Mode for Verification. (see Figure 1.9)

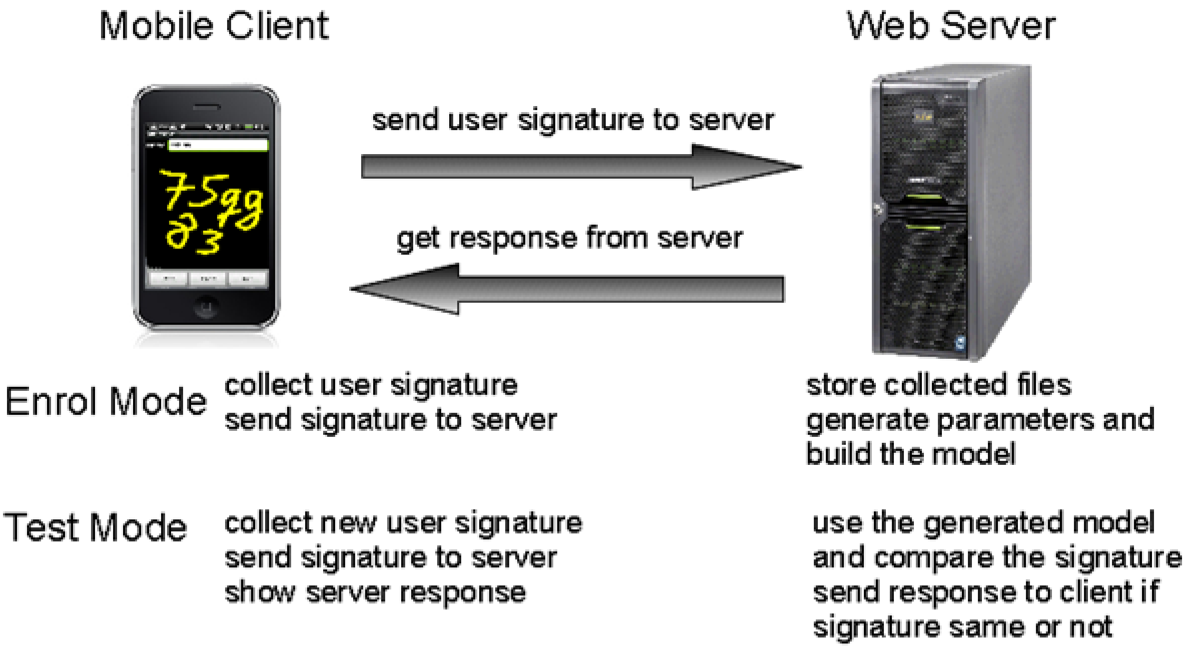


Figure 1.9: Enroll and Test Modes for Writer Recognition [123]

The experience in the implementation of smartphone Apps, client-server applications, and the idea of safe authentication systems the author used within the framework of his work as an administrator, manager and researcher in several eLearning projects. An overview with short description of the publications resulting from these activities is given below:

An App was developed for school mathematic in the study beginning phase and published as native app to Windows Phone and Android Store [31][33][38][40][43]. (see Figure 1.10 and 1.11)

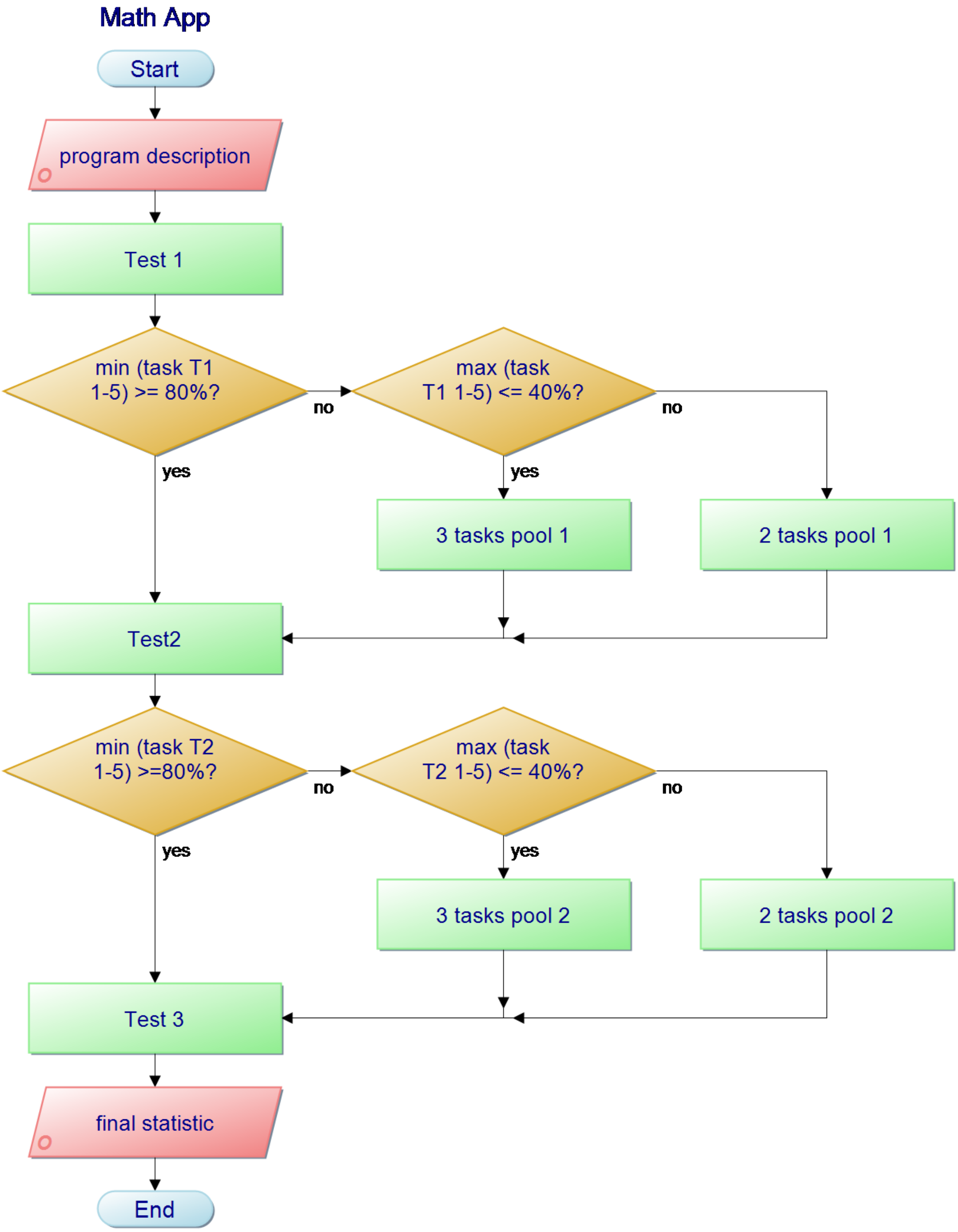
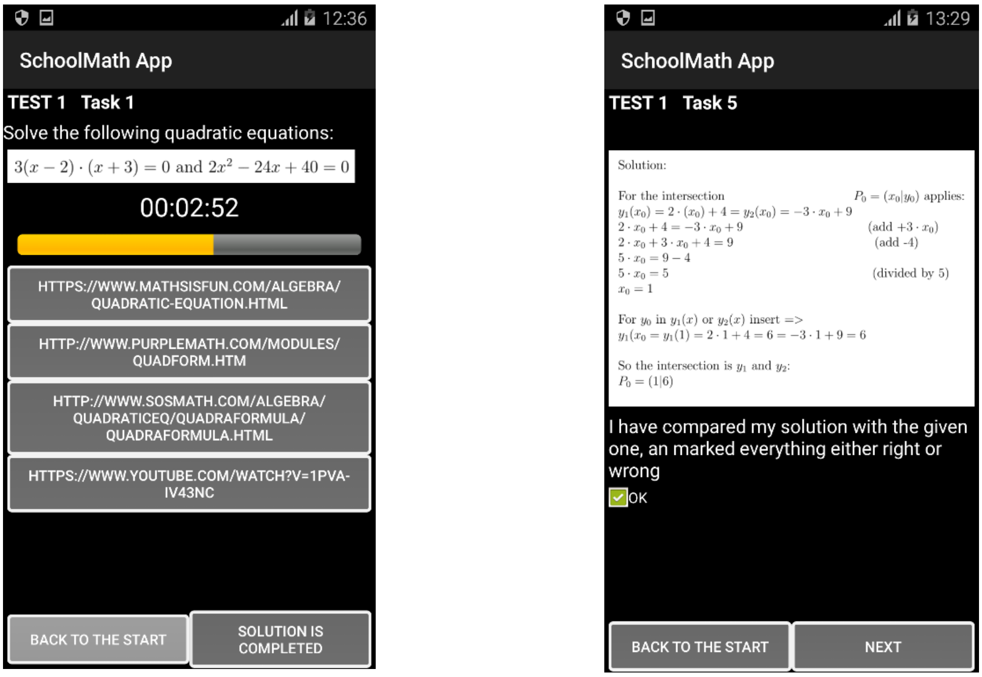


Figure 1.10: flowchart School Math App [43]



TEST 1 Task 1

Solve the following quadratic equations:

$3(x-2)\cdot(x+3)=0$ and $2x^2-24x+40=0$

00:02:52

HTTPS://WWW.MATHSISFUN.COM/ALGEBRA/QUADRATIC-EQUATION.HTML

HTTP://WWW.PURPLEMATH.COM/MODULES/QUADFORM.HTM

HTTP://WWW.SOSMATH.COM/ALGEBRA/QUADRATICEQ/QUADRAFORMULA/QUADRAFORMULA.HTML

HTTPS://WWW.YOUTUBE.COM/WATCH?V=1PVA-IV43NC

TEST 1 Task 5

Solution:

For the intersection $P_0=(x_0|y_0)$ applies:

$y_1(x_0)=2\cdot(x_0)+4=y_2(x_0)=-3\cdot x_0+9$

$2\cdot x_0+4=-3\cdot x_0+9$ (add $+3\cdot x_0$)

$2\cdot x_0+3\cdot x_0+4=9$ (add -4)

$5\cdot x_0=9-4$

$5\cdot x_0=5$ (divided by 5)

$x_0=1$

For y_0 in $y_1(x)$ or $y_2(x)$ insert =>

$y_1(x_0=y_1(1)=2\cdot 1+4=6=-3\cdot 1+9=6$

So the intersection is y_1 and y_2:

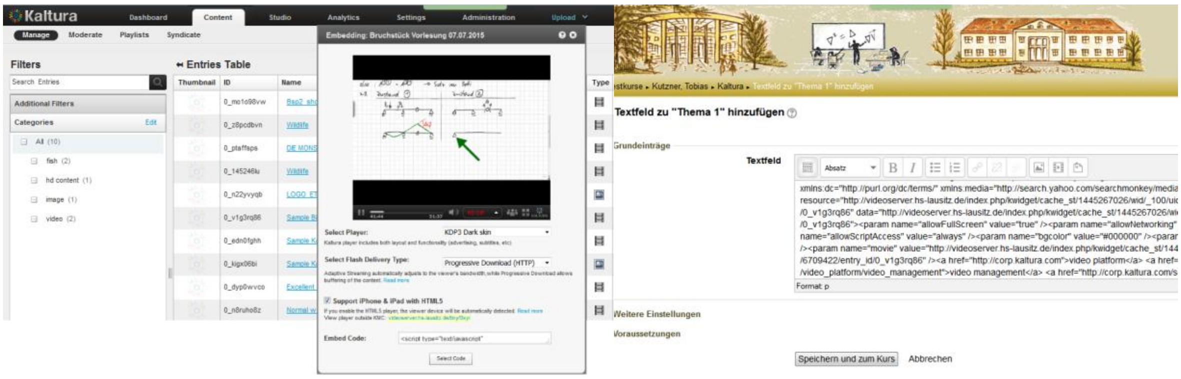
$P_0=(1|6)$

I have compared my solution with the given one, an marked everything either right or wrong

☑ OK

Figure 1.11: screenshot School Math App task with LaTeX text and solution [43]

As administrator eLearning and Video platforms have been set up, adapted and furtherly developed, server and modules were installed, implemented and updated [32][37].



Entries Table

Thumbnail	ID	Name
	0_mo1o98vw	Bap2_sho
	0_z8pcdbvn	Wildlife
	0_ptaffaps	DIE MONS
	0_145246lu	Wildlife
	0_n22yvyqb	LOGO_ET
	0_v1g3rq86	Sample B
	0_edn0fghh	Sample K
	0_kgx06bi	Sample K
	0_dyp0wvco	Excellent
	0_n8ruho8z	Normal w

Embedding: Bruchstück Vorlesung 07.07.2015

Select Player: KDP3 Dark skin

Kaltura player includes both layout and functionality (advertising, subtitles, etc)

Select Flash Delivery Type: Progressive Download (HTTP)

Adaptive Streaming automatically adjusts to the viewer's bandwidth, while Progressive Download allows buffering of the content. Read more

☑ Support iPhone & iPad with HTML5

If you enable the HTML5 player, the viewer device will be automatically detected. Read more

View player outside KMC: videoserver.hs-lausitz.de/tiny/5kyr

Embed Code: <script type="text/javascript"

stkurse ▸ Kutzner, Tobias ▸ Kaltura ▸ Textfeld zu "Thema 1" hinzufügen

Textfeld zu "Thema 1" hinzufügen

Textfeld

```
xmlns:dc="http://purl.org/dc/terms/" xmlns:media="http://search.yahoo.com/searchmonkey/media
resource="http://videoserver.hs-lausitz.de/index.php/kwidget/cache_st/1445267026/wid/_100/uic
/0_v1g3rq86" data="http://videoserver.hs-lausitz.de/index.php/kwidget/cache_st/1445267026/wi
/0_v1g3rq86"><param name="allowFullScreen" value="true" /><param name="allowNetworking"
name="allowScriptAccess" value="always" /><param name="bgcolor" value="#000000" /><parar
/><param name="movie" value="http://videoserver.hs-lausitz.de/index.php/kwidget/cache_st/144
/6709422/entry_id/0_v1g3rq86" /><a href="http://corp.kaltura.com">video platform</a> <a href=
/video_platform/video_management">video management</a> <a href="http://corp.kaltura.com/s
```

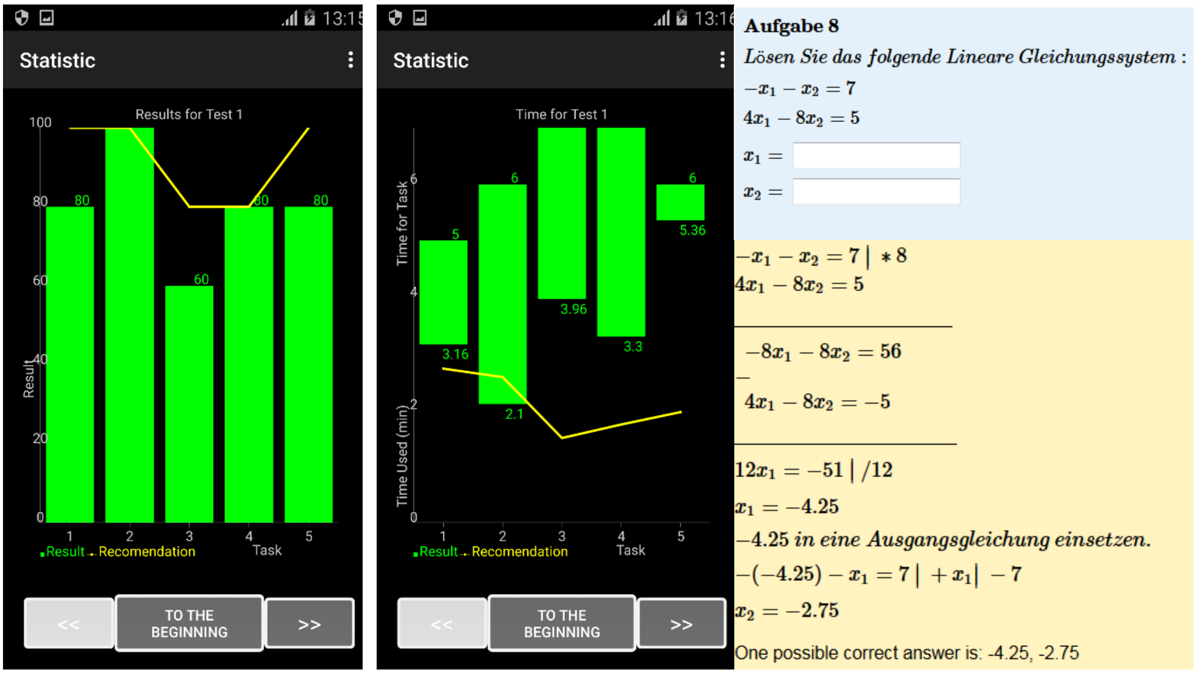
Format: p

Figure 1.12: Video content, embed code Kaltura[5] and eLearning Platform Moodle[6] [37]

Research where carried out on new methods to make the work easier for students, but also teachers, the algorithm of the math app was further developed and supplemented by statistics, new modules for the learning platform were tested, developed and installed especially for dynamic assessments and video assessments [34]-[36][39][41][42][44] (see Figure 1.13).

[5] Kaltura CE: Open Source Video Platform (Community Edition) www.kaltura.org

[6] Moodle: Open Source eLearning Platform https://moodle.org



Aufgabe 8

Lösen Sie das folgende Lineare Gleichungssystem :

$-x_1 - x_2 = 7$

$4x_1 - 8x_2 = 5$

$x_1 =$

$x_2 =$

$-x_1 - x_2 = 7 \mid *8$

$4x_1 - 8x_2 = 5$

$-8x_1 - 8x_2 = 56$

$-$

$4x_1 - 8x_2 = -5$

$12x_1 = -51 \mid /12$

$x_1 = -4.25$

-4.25 *in eine Ausgangsgleichung einsetzen.*

$-(-4.25) - x_1 = 7 \mid +x_1 \mid -7$

$x_2 = -2.75$

One possible correct answer is: -4.25, -2.75

Figure 1.13: Statistics with time and recommendation, Example of a dynamical task in Moodle [36][43]

1.5 Hypothesis

> ***The combination of handwriting and secure password leads in truly secure authentication.***

Whether this is really so, brings a plus of security, is scientifically proven and from this approach a system for the secure authentication will be developed and should be a research focal point of this work.

The following main points are investigated and referenced in this thesis:

1. Handwritten passwords are safer than keyboard-written passwords.

2. It’s possible to identify writers using the handwritten password.
3. Mobile devices are suitable for handwritten password input.
4. The features for the handwriting identification can be applied both to a handwritten password and a signature.
5. An impostor, with the knowledge of the password, is rejected on the basis of the biometric features of his handwriting.

1.6 Document structure

This thesis focuses on writer identification with a handwritten password. For this, the basic methods, terms and procedures are clarified and the State of the Art of the analysis of handwriting is described, common used public databases are analyzed, and the academic void is identified in Chapter 2. In Chapter 3 concept idea of the whole system is described. In Chapter 4 the methods for data collection, preprocessing, feature extraction, classification, and materials like used databases, the parameter reduction processes, and the used classification are shown. The realized experiments are explained. In Chapter 5 the results from the experiments are presented, evaluated and discussed. The last chapter, Chapter 6 summarizes the work with conclusions and future investigations.

Chapter 2 State of the art

2.1 Review Methodology

2.1.1 Biometrics: Conceptuality and Demarcation

The term biometrics is derived from the Greek terms "bios" for "life" as well as "metron" for "measure". Biometrics is therefore understood to mean methods which are based on a measurement of unique human characteristics and thus enable an unambiguous recognition of an individual [46].

Biometrics in the narrower sense refers to an automated recording and evaluation of individual features in which the recognition performance is taken over by computer technologies. Biological and / or anatomical properties (such as, for example, the structures of a face, an iris or a finger) and behavior-dependent properties (for example, the typing behavior, the voice, the gait or handwriting) could also be used or evaluated for biometric recognition. In principle, two classes of biometric methods can be distinguished [46]:

- Identification based on biometric characteristics;
- Verification of an identity claim based on biometric characteristics.

2.1.2 Biometric- Verification and Identification: Characteristics, Methods and Systems

The **biometric verification**, i.e. The confirmation of the claimed identity of the individual (1: 1 = the measured person is actually the one who claims it to be), and - the **biometric identification**, i.e. The recognition of an individual from a (defined) set of biometrically registered persons (1: n = the measured person is XY). (see Figure 2.1) In the case of verification, the current measurement data is compared with the existing data of the individual, e.g. On a chip card or a PDA (= Personal Digital Assistant), or can be stored centrally, with a predefined user identification, decentral (in the possession of the person). Within the scope of enrollment, e.g. In the case of a large-scale application for banking machines - it will often be necessary to temporarily store the biometric data at a further central location in order to load it onto the chip card for the corresponding user ("personalizing the chip cards"). The advantage from the point of view of data protection that the template is placed in the user's authority in the case of a decentralized verification is that safety disadvantages and associated possible problems of adhesion are counteracted by loss or damage [49]. In the case of identification, the biometric system compares the measured data with the - centrally stored - data of all previously registered and checks which template best matches that of the current user. This results in higher requirements with regard to the required database size and identification time. This type of biometric detection is currently used mainly in high-security areas with a small number of users or for police investigations [49]. (Biometric personal) authentication / authentication, which, however, has not yet been implemented in the German-speaking world [51], so that biometric personal identification is generally used, even if only a verification takes place . Authorization is the authorization (as the actual result of verifying the identity of the user), i.e. the authorization or authorization for access or action [52].

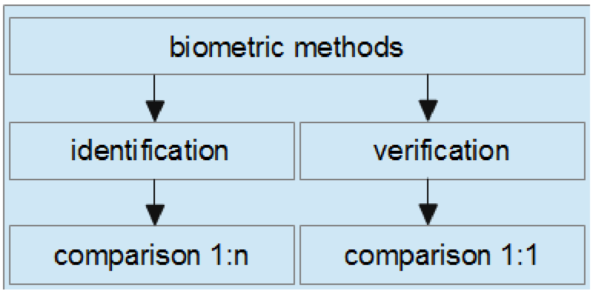


Figure 2.1: limitation identification / verification [46]

During **identification**, a currently collected measurement value is compared with many pre-stored reference values. Known examples of biometric identification are the use of fingerprints in the context of criminal investigations or the video surveillance of public spaces. For example, the faces of passers-by could be matched with those of known offenders, which are stored in central databases [46].

Within the second class of methods, the so-called biometric **verification**, however, only a comparison with a stored reference data set takes place. This corresponds to the use of a password or a PIN for authentication - in this case, a user simply identifies the identity he claims. A biometric verification thus represents a fundamental alternative to authentication on the basis of knowledge or possession (e.g. with a smartcard) [46].

The security of a password in the sense of erasability depends on the length and complexity of the selected string. However, complex passwords may not only exclude strangers, but the user-friendliness also decreases. The longer a string, the more difficult it is to remember it correctly and enter it without errors. Thus, security and usability are linked to each other in a fundamental conflict of objectives. An ideal system is characterized by the fact that both the safety of overcoming and the usability are particularly high. Traditional authentication methods are not (yet) able to accomplish this. The target conflict also applies to biometric methods, but the combination of different methods gives the possibility to approach the ideal method [46]. (see Figure 2.2)

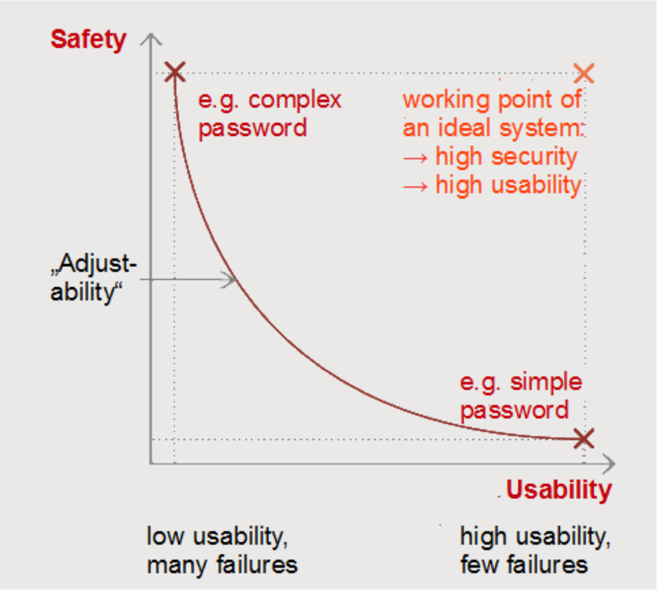


Figure 2.2: safety usability [46]

The comparison process, which takes place with biometric methods, is not absolutely exact, but a fuzzy comparison (see Figure 2.3), in contrast to knowledge-based methods. Just as different signatures of the same person do not completely agree, there are regularly differences in multiple measurements of a biometric feature. As well as human security personnel at the airport, an automated biometric procedure has a certain effect in the comparison of the current measured value and the stored reference value. While a person compares the current appearance with the passport and also estimates the body size and assesses the eye color, a technical comparison is based on a formalized pattern recognition process [46].

A common approach is to first extract individual features from biometric data. In the case of the use of fingerprints, these are, for example, branches, conspicuous loops or endings in the grooves of the fingerprint, the so-called minutiae. In a comparison for the purpose of the verification, a sufficient correspondence between the feature quantities is then examined. The required match or overlap corresponds to a required password length and thus indicates the overrunning safety. This means that the security of a biometric procedure per se is also dependent on the configuration of the system. Each biometric system thus has a so-called working line, which specifies the configurability or "adjustability". A working point is to be selected as a parameter on this line. Again, there is a balancing between usability and security, the more restrictive a system is configured (i.e. the higher the security level has been chosen), the more input attempts are necessary if necessary. In principle, biometric systems can offer a higher level of basic security compared to passwords [47]. However, especially in the case of everyday use, it must be ensured that necessary safety levels are not undershot. This could be done by, for example, specifying parameters from the manufacturer's side, which should prevent end users from making a mistake - and thus greatly reduce the security. However, further research and standardization efforts are necessary to ensure the stability and comparability of various biometric approaches with regard to the actual safety achieved [46].

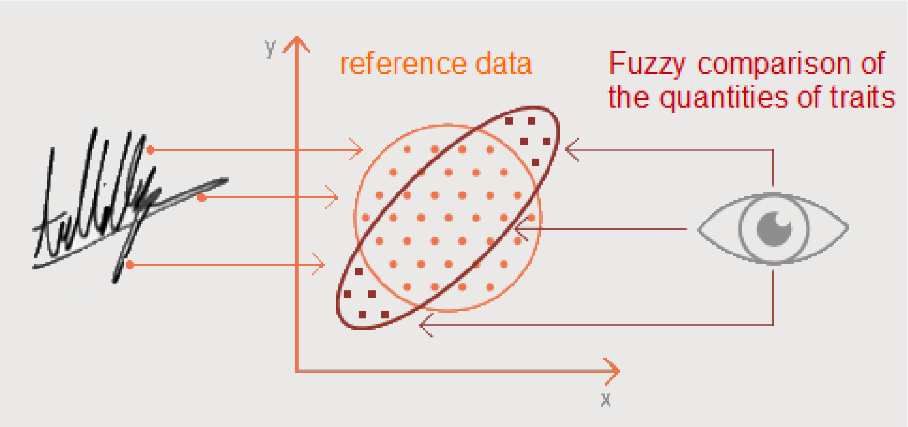


Figure 2.3: Fuzzy comparison, different modalities [46]

2.1.3 Enrolment, Template, Identification, and Verification

The basis of each biometric procedure is the so-called **enrolment**, irrespective of the feature used and the applied technique. It includes the first-time measurement and measurement of the biometric characteristic of the future users, the conversion of the "raw data" into a reference data record and the storage thereof, the so-called template. This represents the comparison value with which the new measurement data (at least to a high degree) must coincide in all subsequent biometric checks in order to be able to identify the user. This basic process of enrollment requires, therefore, both the highest technical requirements (with respect to sensitivity and accuracy, in order to produce individual, but also reproducible data sets) as well as the highest safety requirements. The main purpose of the application of biometric methods, i.e. The increase in the security of an overall system (for example, the cash dispensing at the machine) can only be achieved if the reference data record, the template, can be stored permanently protected. Particularly with regard to future large-scale deployments, many questions remain open (e.g. which and how many persons must be appropriately qualified to carry out the enrollment), the solution of which is likely to involve a high financial and organizational effort [49]. Two "operating modes" are distinguished in a biometric check of the users [50].

2.1.4 Measure or biometric quality

The Problems in practice are: Incorrect acceptance and false rejection.

Ideally, any biometric data set would be unique to a human individual and unambiguously assigned to it - originally collected reference data (template) and respectively measured data set would be identical. In practice, limitations of this ideal uniqueness, accuracy and reproducibility result for various reasons:

Each measurement process is a powerful information reduction. For reasons of principle (capacity), the collected data must be limited. In addition, there is the respective measuring limit (sensitivity) and accuracy of the sensor or of the overall system as well as the "noise" which cannot be avoided. The amount of data to be stored in the template should be minimized for technical reasons (memory size, transmission rates), but the accuracy is reduced. Behavioral characteristics always show a greater or lesser variance due to the nature of human motor skills. However, even physiological features are only limited in time. They can be temporarily or permanently altered by aging processes, diseases or injuries. Slight changes must therefore be tolerated by the system in "active" as well as "passive" methods. - In addition, disturbing environmental influences during the measurement, e.g. Different light conditions or temperature changes that can affect the performance of sensors [49].
The biometric system carries out a statistical comparison of the data sets of the template and the measurement. The result specifies a percentage value of the match. A hundred per cent agreement will practically never occur in the above-mentioned technical, physiological, or conditional limitations. Thus, for each system, a **threshold** (a value of the degree of matching of the reference and measured value) must be defined (e.g., 95%) from which the identification or verification is deemed to have taken place and the user is accepted as justified. This tolerance threshold has a large impact on how many users are either mistakenly accepted or are falsely rejected (or are forced to repeat the process several times). The rates of false rejection or false acceptance of a biometric system (**FRR** = False Rejection Rate);
FAR = (False Acceptance Rate) cannot be calculated theoretically, but must be determined empirically. FAR and FRR are affected in such a way that a decrease in false acceptance increases false rejection and vice versa. The absolute magnitude of the error rates, however, depends on the sensitivity and accuracy, i.e. the severity, of the overall system and is therefore

determined by the choice of the above-mentioned. Tolerance threshold. Less accurate systems such as voice recognition will either accept many users incorrectly (if the tolerance threshold is low) or falsely reject them. The iris scan, on the other hand, has a low FAR as well as a low FRR due to its high separation sharpness. Depending on the practical application, the rate of false rejection should be minimized (usually for reasons of comfort, i.e. avoiding frustrating errors), or the rate of false acceptance (especially with regard to a safety increase) by selecting and setting the tolerance threshold of the system. FAR and FRR are considered to be the most important parameters for the performance of a biometric system. If the values are the same, one speaks of "**EER**" = "Equal Error Rate". A further parameter which has been rarely treated in test scenarios so far, but which is very important in practice, is the rate of false enrollment and enrolment tests (**FER** = False Enrolment Rate), which can have a great impact on user acceptance [53].

For organizing classifiers and visualizing their performance Receiver Operating Characteristics (**ROC**) graphs are a useful technique. ROC graphs are commonly used in medical decision making, and in recent years have been increasingly adopted in the machine learning and data mining research communities [54].

The ROC curves describes the performance of a verification system on a test set by plotting the False Acceptance Rate (FAR) against the False Rejection Rate (FRR).
Figure 2.4 shows a confusion matrix and equations of several common metrics that can be calculated from it.

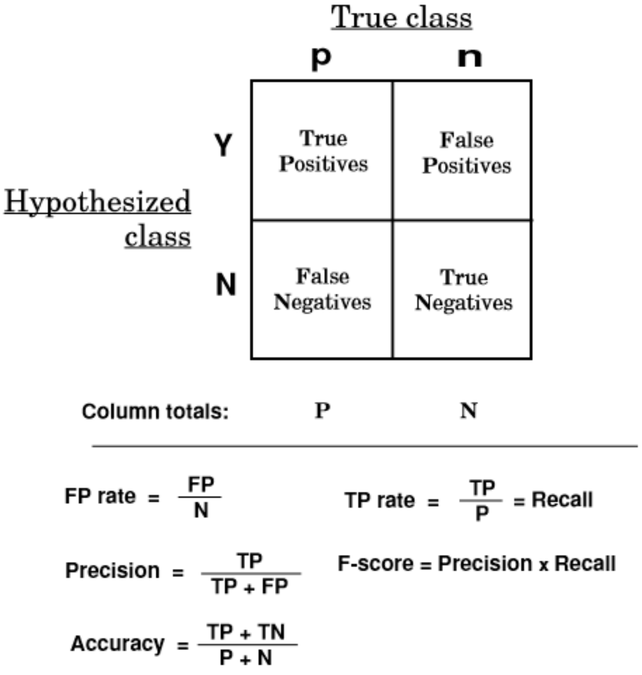


Figure 2.4: A confusion matrix and several common performance metrics that can be calculated from it [54].

The numbers along the major diagonal represent the correct decisions made, and the numbers off this diagonal represent the errors—the confusion—between the various classes. The **True Positive rate** (also called hit rate and recall) of a classifier is estimated as:

$$\text{TP rate} \approx \frac{\text{positives correctly}}{\text{classified total pos}}$$

The **False Positive rate** (also called false alarm rate) of the classifier is:

$$\text{FP rate} \approx \frac{\text{rate negatives incorrectly classified}}{\text{total negatives}}$$

Additional terms associated with ROC curves are:

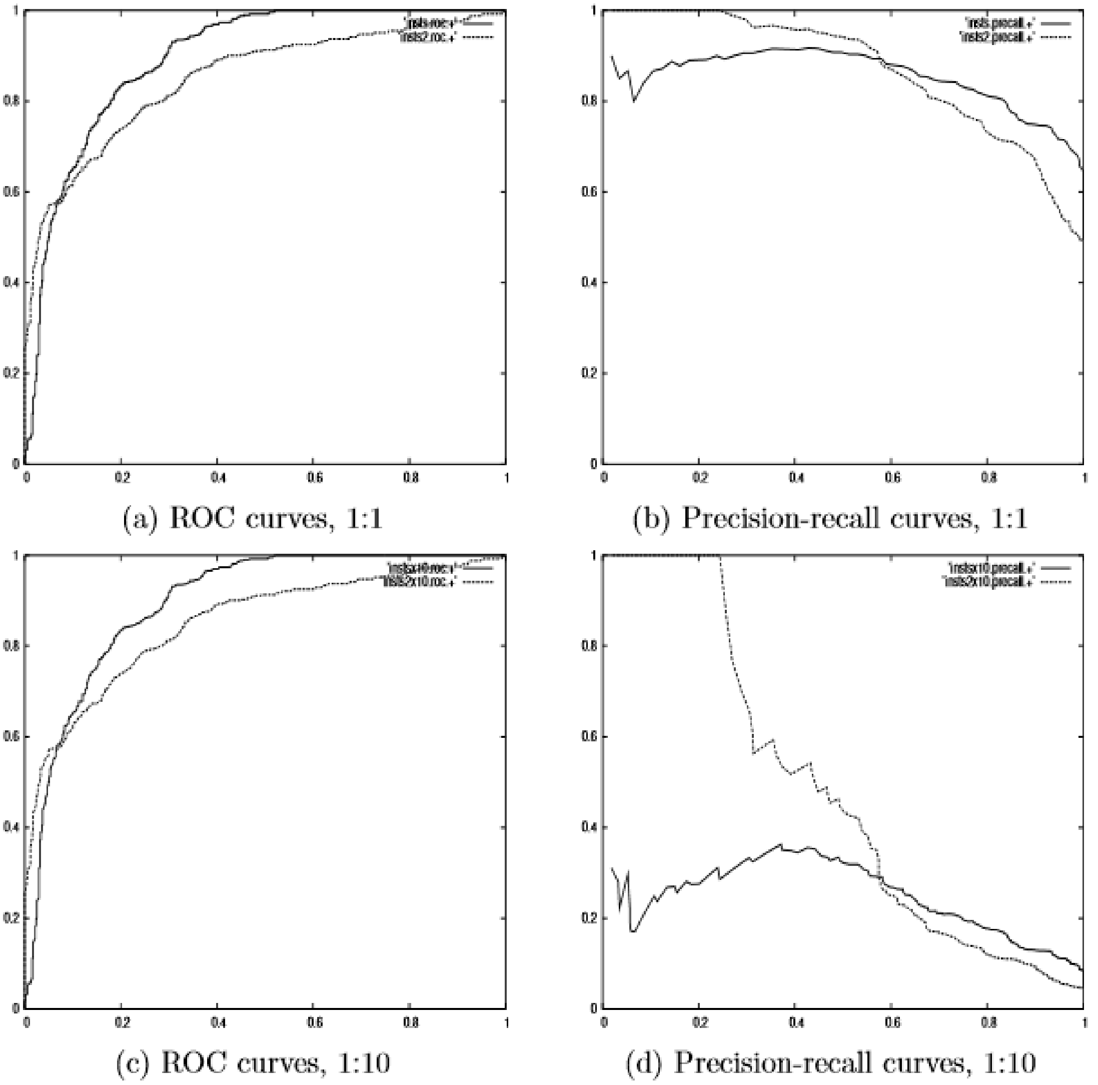
$$Sensitivity = Recall$$

$$Specitivity = \frac{\text{True negatives}}{\text{False positives } + \text{ True negatives}} = 1 - \text{FP rate}$$

$$\text{Positive predictive value} = \text{Precision}$$

To see the effect of class skew, consider the curves in figure 2.5, which show two classifiers evaluated using ROC curves and precision-recall curves. In 2.5a and b, the test set has a balanced 1:1 class distribution. Graphs 2.5c and d show the same two classifiers on the same domain, but the number of negative instances has been increased ten-fold. Note that the classifiers and the underlying concept has not changed; only the class distribution is different. Observe that the ROC graphs in 2.5a and 2.5c are identical, while the precision-recall graphs in 5b and 2.5d differ dramatically. In some cases, the conclusion of which classifier has superior performance can change with a shifted distribution [54].

Further information about ROC curves and the applications can be found in chapter 3 Methods and Materials.



(a) ROC curves, 1:1

(b) Precision-recall curves, 1:1

(c) ROC curves, 1:10

(d) Precision-recall curves, 1:10

Figure 2.5: ROC and precision-recall curves under class skew [54]

2.1.5 Signature-and Handwriting Recognition

In the case of signature recognition or handwriting recognition, not only the optical appearance of the signature (the lettering as an "off-line parameter") is decisive, but also features such as pressure, speed, acceleration, pick-up and drop-off points as well as pin- Parameter. The under / handwriting is today usually taken with a commercial tablet or PDA or touchscreen. As an alternative, special pins with sensors are also used, which record the parameters for the signature / manuscript's performance and transmit them for evaluation [50]. An extension of the signature analysis is provided by a handwriting system in which not only the signature, but

so-called "semantics" [53], are used for hand-written authentication. This can include predefined words, whole sentences or even small drawings. (see Figure 2.6)
An advantage of the use of "semantics" lies in the anonymity - even with centrally stored data sets this can be preserved. The methods can be set up in such a way that they can be controlled by the user themselves by changing the stored reference data record relatively quickly. This makes a clearer link to a declaration of the will possible. The "controllability" by the user should also be conducive to acceptance. Since the detection of the dynamic parameters is a life detection, the security against counterfeiting is quite high [55]. Due to the (still) high error rates, however, the systems have so far only been very limited [49].

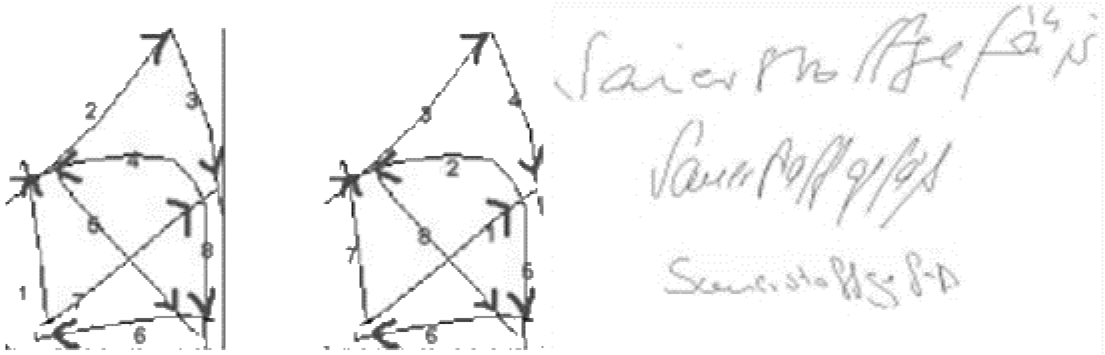


Figure 2.6: Online handwriting "Haus von Nikolaus" and given signature [53]

For the recognition of signatures, they act in the same manner as for normal handwriting recognition, a number of parameters are obtained from the signature and then passes their analysis by various methods such as ANN.

In the paper Personal authentication using signature recognition [56], various forms for signature recognition differentiating between dynamic (on-line) or static (off-line) are presented, on-line methods are based on analysis of the shape, speed, stroke, pen pressure and timing information, while offline recognition is based on technical analysis by recognizing the shape. But in the off-line recognition can also obtain other parameters as vertical and horizontal projection, center of gravity, the Hough Transform (HT) all of them being explained in Some Handwritten Signature Parameters in Biometric Recognition Process [57]. Also another interesting method for signature recognition is the one based on graphs [58], this work creating a graph from the sample of writing.

Additionally, several solutions for handwriting recognition have been proposed without the use of touch screens to collect writing. Besides the typical solutions that use scanners to take samples of writing on paper already there are other solutions like those based on Smart Pens, as for example the developed in Person Authentication by Handwriting in air using a Biometric Smart Pen Device [59], in this project for biometric identification they use a Biometric Smart Pen (BiSP) which reflects the dynamics of writing of a person writing with BISP in the air, because according to their study the difference between writing in the air and on paper is minimal.

The BiSP is a normal pen equipped with a series of sensors to measure (e.g. acceleration, the angle, grip pressure) characteristic of each person at the time of writing.

But are these really secure systems? according to a study called Evaluating the security of handwriting biometrics [60], these systems are not completely secure systems according as can be observed in the study were able to fool most of these systems. In the study they are able through a handwriting sample of the particular person to supplant with: Handwriting sample of the person without this is taken on a touch screen. Statistical studies of pen-stroke taken of population.

The present writer recognition systems are divided into off-line and on-line systems. On powerful machines working off-line systems [125][126][129] ,they use as parameters union of letters, correlation of the same words [126] or static and dynamic forms [129]. The recognition rate rises considerably, if not only characters are used for the identification [129]. The quality of the results depends on the number of used characters and the number of the repetition of the same word [129].

On-line handwriting recognition systems use the characters or words, which are written on a touch sensitive surface [127][128]. They interpret the movements and the stylus of writing. The problem of on-line systems are the limited resources of the mobile devices. To achieve better results an adaptive classifier is used, which is modified progressively every time the user writes a pattern [128]. But this method increases the computational time and the complexity of the system. The on-line recognition systems have good results of character recognition [127], but they cannot identify the writer because of the limited resources.

The basic idea of a distributed solution is also used in [130][131]. These systems recognize characters and text but not writers.

In the paper Combining Neural Networks as Context-Driven Search for On-Line, Printed Handwriting Recognition in the Newton [132] we find the description of Apple’s print recognizer (APR), the input are sequences of x- and y coordinates penning up and down information. The segmentation describes which stroke will be combined to produce segments groups of strokes that will be treated as possible letters are summarized. The classification stage evaluates each segment using the ANN classifier, and produces a vector of output activations that are used as letter class probabilities. The search stage uses these class probabilities together with models of lexical and geometric context to find the N most likely word or sentence hypotheses. (see Figure 2.5)

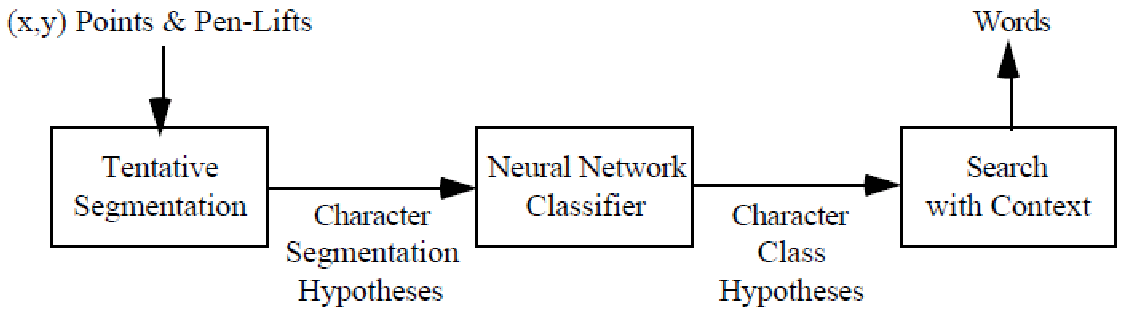

(x,y) Points & Pen-Lifts
Words
Tentative Segmentation
Character Segmentation Hypotheses
Neural Network Classifier
Character Class Hypotheses
Search with Context

Figure 2.7: Simple block diagram of hand print recognizer [132]

2.2 Most used databases in literature

In this paragraph the most used databases ATV-Signature Long Term Database, IAM online handwriting DB, SVC 2004 DB, and MCYT 100 DB are described shortly.

- **ATV-Signature Long Term Database (ATV- SLT DB)**

The dataset comprises the on-line signature data of the 29 common users to the BiosecurID and the Biosecure databases. These two signature subsets were acquired in a 15-month time span and present some unique features that make them especially suited for aging evaluation of on-line signature recognition systems. The general time distribution of the different sessions of the database is shown in Figure 2.8 [179].

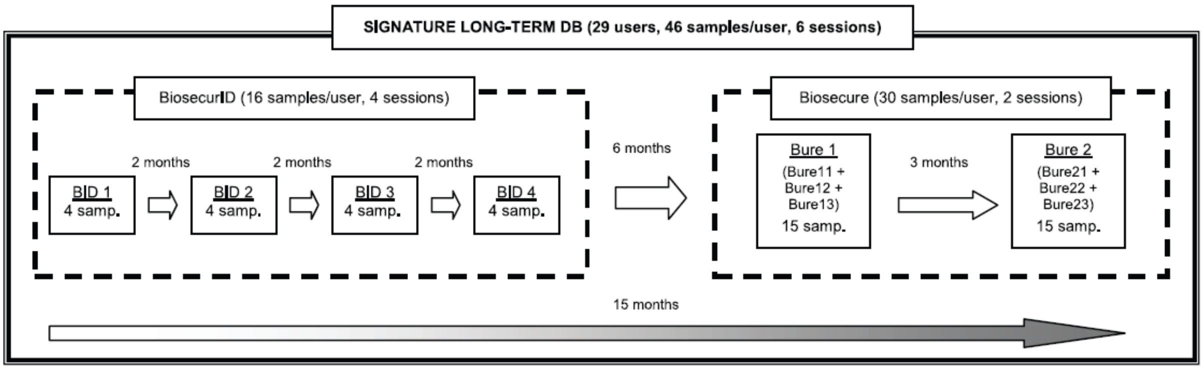

SIGNATURE LONG-TERM DB (29 users, 46 samples/user, 6 sessions)
BiosecurID (16 samples/user, 4 sessions)
2 months
2 months
2 months
BID 1 4 samp.
BID 2 4 samp.
BID 3 4 samp.
BID 4 4 samp.
6 months
Biosecure (30 samples/user, 2 sessions)
Bure 1 (Bure11 + Bure12 + Bure13) 15 samp.
3 months
Bure 2 (Bure21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

Figure 2.8: General time diagram of the different acquisition sessions that conform the Signature Long-Term Database [179]

The BiosecurID Signature Subset:

It comprises 16 original signatures per user (29 users). Samples were captured in 4 separate acquisition sessions (named BID1, BID2, BID3 and BID4 in Figure 6) The sessions were captured leaving a two-month interval between them, in a controlled and supervised office-like scenario. Users were asked to sign on a piece of paper, inside a grid that marked the valid signing space, using an inking pen. The paper was placed on the Wacom Intuos 3 pen tablet that captured the time signals of each signature at a 100Hz sampling rate (trajectory functions *x* and *y* with an accuracy of 0.25mm, pressure function p with a precision of 1024 pressure levels, and azimuth and altitude angles). All the dynamic information is stored in separate text files following the format used in the first Signature Verification Competition. ATV-SLT Signature Sample see Figure 2.9:

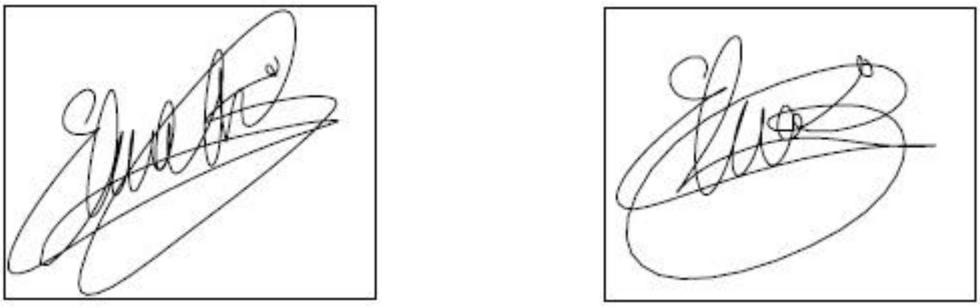


Figure 2.9: ATV-SLT signature sample [179]

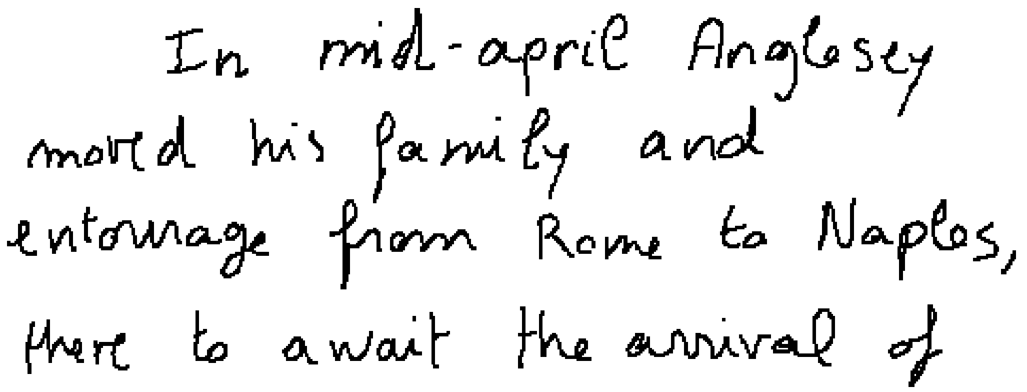
The Biosecure Signature Subset:

This dataset was captured 6 months after the BiosecurID acquisition campaign had finished. It comprises 30 original signatures per user (same 29 users as the BiosecurID subset) distributed in two acquisition sessions separated three months. The 15 original samples corresponding to each session were captured in three groups of 5 consecutive signatures with an interval of around 15 minutes between groups. The signature dataset was designed to be fully compatible with the BiosecurID one. The acquisition scenario and protocol are almost identical: as in the BiosecurID case, users had to sign using an inking pen on a piece of paper with a restricted space, placed over the Wacom Intuos 3 pen tablet. The dynamic information stored is the same as in BiosecurID and following also the SVC format.

- **IAM online handwriting DB**

The IAM On-Line Handwriting Database (IAM-OnDB) [185] contains forms of handwritten English text acquired on a whiteboard. It can be used to train and test handwritten text recognizers and to perform writer identification and verification experiments. The database was first published in [185] at the ICDAR 2005. The database contains forms of unconstrained handwritten text, acquired with the E-Beam System. The collected data is stored in xml-format, including the writer-id, the transcription and the setting of the recording. For each writer the gender, the native language and some other facts which could be useful for the analysis are stored in the database.

All XML-forms and also all extracted text lines are available for download as XML files. The extracted lines are also available for download as converted tif-images, which could be used for the recognition as proposed. All texts in the IAM On-line database are built using sentences provided by the LOB Corpus. Text sample see Figure 2.10:



In mid-april Anglesey
moved his family and
entourage from Rome to Naples,
there to await the arrival of

Figure 2.10: Text sample IAM Online Handwriting Database [185]

The IAM Online Handwriting Database is structured as follows:

- 221 writers contributed samples of their handwriting
- more than 1'700 acquired forms
- 13'049 isolated and labeled text lines in on-line and off-line format
- 86'272 word instances from a 11'059 word dictionary

- **SVC 2004 DB**

SVC2004 [186] consists of two separate signature verification tasks using two different signature databases. The signature data for the first task contain coordinate information only, but the signature data for the second task also contain additional information including pen orientation and pressure. The first task is suitable for SVC2004 participating teams on-line

signature verification on small pen-based input devices such as personal digital assistants (PDA) and the second task on digitizing tablets.
Each database has 100 sets of signature data. Each set contains 20 genuine signatures from one signature contributor and 20 skilled forgeries from at least four other contributors. Unlike physiological biometrics, the use of skilled forgeries for evaluation is very crucial to behavioral biometrics such as handwritten signature. Of the 100 sets of signature data, only the first 40 sets were released (on 25 October 2003) to participants for developing and evaluating their systems before submission (by 31 December 2003). While the first 40 sets for the two tasks are totally different, the other 60 sets (not released to participants) are the same except that the pen orientation and pressure attributes are missing in the signature data for Task 1. Although both genuine signatures and skilled forgeries were made available to participants, user enrollment during system evaluation accepted only five genuine signatures from each user, although multiple sets of the genuine signatures each were used in multiple runs. Skilled forgeries were not used during the enrollment process. They were only used in the matching process for system performance evaluation. Evaluation of signature verification performance for each user was only started after all users had been enrolled. Therefore, participants could make use of genuine signatures from other users to improve the verification accuracy for a user if they so wished [186].

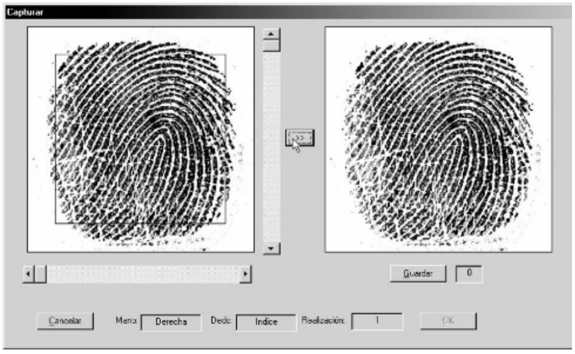
- **MCYT 100 DB**

The current need for large multimodal databases to evaluate automatic biometric recognition systems has motivated the development of the MCYT bimodal database [187]. The main purpose has been to consider a large scale population, with statistical significance, in a real multimodal procedure, and including several sources of variability that can be found in real environments. The acquisition process, contents and availability of the single-session baseline corpus are fully described. Some experiments showing consistency of data through the different acquisition sites and assessing data quality are also presented.

The basic idea behind the design of the MCYT database has been the optimization of the extent of its significance which, in a multimodal database, is related to (i) the number of individuals enrolled, (ii) the number of modalities per individual and (iii) the number of realizations (samples) for each modality (which should include those variability factors that can be found in the acquisition process). Because this significance increases with each one of the above-mentioned parameters (i) to (iii), the design of multimodal biometric databases (see, for example, BIOMET database [5]), usually seeks the maximization of each one of them (i.e. as many individuals as possible, as many modalities as possible and/or as many samples as possible).
In the design of the MCYT database where fingerprint and signature have been the selected biometric features, the maximization of the significance has been carried out in two different ways. On the one hand, the number of individuals has been maximized while maintaining the

number of realizations within a useful margin; on the other hand, the number of realizations has been maximized while maintaining the number of individuals within a useful quantity.

For each individual, the on-line signature capture session is achieved after the fingerprints are registered in the database. Since the acquisition of each on-line signature is accomplished dynamically, a graphics tablet is needed: the acquisition device used is a WACOM pen tablet, model INTUOS A6 USB. Each target user produces 25 genuine signatures, and 25 skilled forgeries are also captured for each user. These skilled forgeries are produced by the 5 subsequent target users by observing the static images of the signature to imitate, trying to copy them (at least 10 times), and then, producing the valid acquired forgeries in an easy way (i.e. each individual acting as a forger is requested to sign naturally, without artefacts, such as breaks or slowdowns). The signature ending is determined by setting a 3 s timer to the first zero pressure sample found (i.e. a pen up). If no samples with non-zero pressure value are detected in this interval, the capture process is stopped, and the complete signature is stored. Otherwise, the timer is reset until the next pen up is found. Finally, and as a reference, some example images from an MCYT_Fingerprint subcorpus are depicted see Figure 2.11. Example data from an MCYT_Signature subcorpus, azimuth and inclination angles of the pen. (see Figure 2.12)



Mano: Derecha Dedo: Indice Realización: 1

Figure 2.11: Fingerprint examples from MCYT_Fingerprint subcorpus [187]

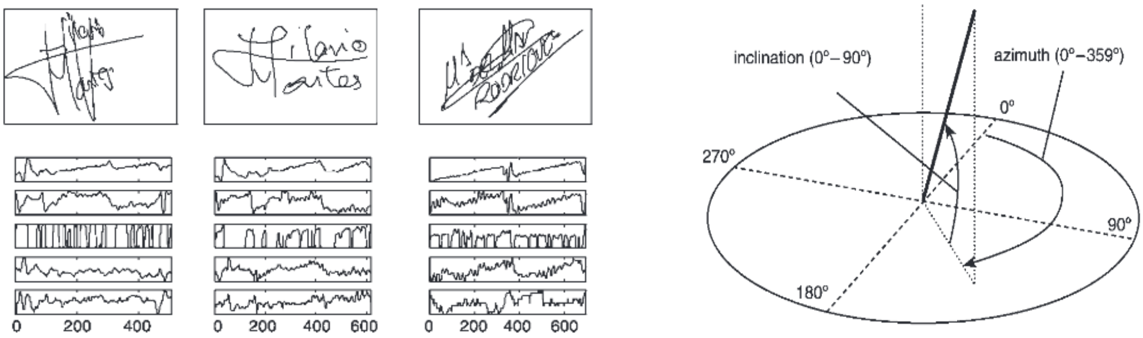


Figure 2.12: Signature Samples and angles of the pen with respect to the plane of the graphic card Intuos from Wacom [187]

A summary overview of other databases gives the Study Offline Handwritten Signature Verification Literature Review see Table 2.1 [71]:

Table 2.1: Commonly used offline signature datasets [71]

Dataset Name	Users	Genuine signatures	Forgeries
Brazilian (PUC-PR)	60 + 108	40	10 simple, 10 skilled
CEDAR	55	24	24
MCYT-75	75	15	15
MCYT-100	100	25	25
GPDS Signature 160	160	24	30
GPDS Signature 960	960	24	30
Grayscale	881	24	30

2.3 Experiments and results

This paragraph gives a summary of experiments and results of writer identification and verification systems.

2.3.1 Writer Identification and Verification

User authentication is the process by which the identity of a person is checked. For this review, two types of data are required, the reference data and the data whose authenticity is to be detected. Authentication is considered to be successful if both data are sufficiently matched. In authentication, two types of confirmation of identity can be distinguished. Identification is the comparison of all stored reference data with the authentication data. The person is considered to be identified if reference data can be found that are sufficiently matched with the currently displayed data. Verification is referred to when the system checks whether the authenticated authentication data matches the reference data of a claimed identity (e.g., user name) within a range of variations [45].

Figure 2.13 shows the general process of biometric verification. The basis for the verification is the stored reference data in the system or on the SmartCard or the pen. The process for generating this data is referred to in biometrics as reference data acquisition (Enrollment). The following processes of the biometric system are traversed: the biological data are collected by the sensor and the features are extracted (feature extraction). If the data are of sufficient quality and number (the majority of the systems requires several authentication data to generate the reference data), a reference data record can be generated by the system which is then stored. In the first step of verification, the detected biometric property (current authentication data) is then detected by the sensor and passed to the feature extractor after a possible preprocessing. It extracts the required features from the input data and creates a feature vector representing the feature within the system. This vector is compared in a comparison process to the feature vector of the reference data of the claimed identity. The similarity value represents the measure of the similarity of the two vectors to one another. This is the input value for the classification which decides on the result of the verification. If the system decides that the user is the one that he pretends, true is returned, false otherwise [45].

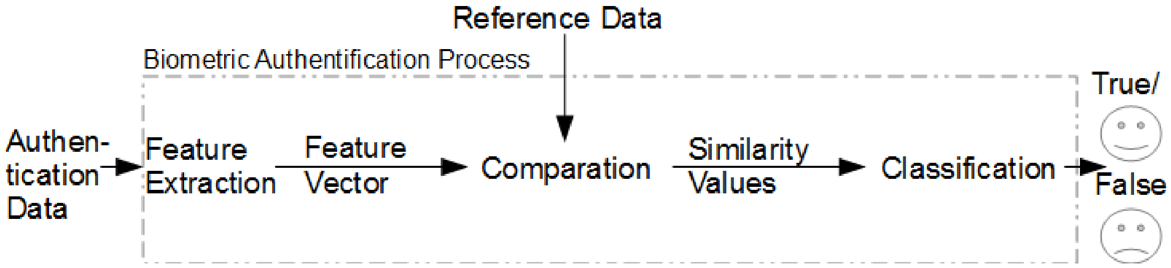


Figure 2.13: Schematic representation of a biometric verification process [45]

In science, a basic distinction is made between offline and online identification and verification.

- Offline: Handwriting text documents are mostly in scanned form of a text page or image that has been digitized by scanning.

So biometric offline and online writer identification is the basis for the construction of a classification model. It comprises the three steps of feature selection, measurement of characteristics and individualization.

- Online: Digital records of handwriting with additional information such as time, pressure, angle and more dynamic information are available in a dataset.

So biometric offline and online writer verification is characterized by the fact that it allows a certain person to use certain features that are individually assigned to them. It presupposes an action that is interpreted by an intelligent algorithm as an indication of its identity.

The following two chapters provide an overview of state-of-the-art on-line and offline identification and verification methods and results with handwritten signatures and passwords.

2.3.2 Offline Writer Identification and Verification

In [58] is suggested creating a graph from the sample of writing, from this graph are taken various data as a connection between segments, adjacent vertices, graph number by which proceeds to verify the signature obtaining a 94% identification accuracy.

On powerful machines working off-line systems [125][126][129] have a success rate of the writers by their handwriting of 92-99%. They use as parameters union of letters, correlation of the same words [126] or static and dynamic forms [129]. The recognition rate rises considerably, if not only characters are used for the identification (40%) but also words or handwritten texts (until 99%) [129].

In [60] the trace transform based affine invariant features are applied for signature verification. The trace and diametric functional are suitably chosen to derive a set of circus functions from each signature image. CEDAR (Center of Excellence for Document Analysis and Recognition) [60] database was used with a total of 24 signature samples collected from each of 55 writers. Some skilled forgers forged the signatures in the database of 55 writer's original signatures and 24 forgeries for each writer are collected. The FAR and FRR curves meet at a threshold of 0.069 with the EER as 24.4%. By setting the threshold value at 0.069, the error rates such as FAR and FRR are found to be 24.58% and 25.83% respectively.

In [61] a novel approach for offline signature recognition system is presented which is based on powerful global and local wavelet features (energy features). In classification stage; a simple Euclidean distance measure is used as decision tool. The average recognition accuracy obtained using this model ranges from 90% to100% with the training set of 10persons to 30 persons.

In [62] an off-line hand printed signature verification method is proposed to detect skilled forgeries of amateur type by using only two genuine signature templates of an individual. The experimentations were carried out on a database of 500 signatures collected from 125 individuals, which consists of three genuine signatures from each individual and one skilled forgery of each individual's signature. The matching is performed using a mixed method of correlation and distance calculation of the feature sets of genuine and skilled forgeries. Experimental results show an average verification rate of 81%.

In [64] a writer identification approach based on graphometrical and forensic features is proposed. An own database is used; composed of 100 users with 10 samples per each one, where each sample is a text in Spanish. An Artificial Neural Network is used as classifier and, according to the decision fusion module, the system reaches up to 94.6% of success rate.

In [65] a novel approach for verification of signatures based on curve matching using shape descriptor and Euclidian distance is introduced. The measurement of similarities is proceeded by finding correspondences between signatures shape descriptor (shape context) with Euclidian distance between the sample points of one signature and the sample point of another signature is attached. The GPDS960 signature corpus database is used: of 960 users, 95 were picked out at random, for random forgeries the first genuine signature of other 94 users was taken. The System reaches FRR 0.1011, FAR for Skilled forger 0.0421 and FAR Random (Untrained) 0.0867, the accuracy is between 89.89% and 95.78%.

In [66] an innovative design for signature verification which is able to extract features from an individual's signatures and uses those feature sets to discriminate genuine signatures from forgeries JAVA-PYTHON platform is proposed. A combination of genuine and fraud signature images is fed into the proposed model and ANN is used to classify these images. The learning rate is gradually increased from 0.01 and an optimized rate of 0.09 is observed with an accuracy of 95 % for 200 epochs. Experimental results suggest that it is able to deliver 95% of accuracy.

In [67] a low level stroke feature is used, which was originally proposed for recognition of printed Gujarati text, for off-line handwritten signature verification. Experiments were performed on the ICDAR 2009 Signature Verification Competition dataset which contains both genuine and forge signatures. Recognition was carried out using the Support Vector Machine (SVM) classifier with 3-fold cross validation. The Equal Error Rate (EER) achieved 15.59%, which is comparable to the ICDAR 2009 Signature Verification Competition Result.

In [68] a detail study on existing methods for handwritten character recognition is introduced. The Table 2.2 provides an overview of the applied methods, used records and results in the examined articles. Accuracy reaches up to 98.26%.

Table 2.2: overview applied methods, used records and results [68]

Paper	**Language**	**Online**	**Offline**	**Model / Features**	**Dataset Size**	**Accuracy Claimed**
[74]	English		√	HMM Gradient (4), Projection (6), Curvature	100 Writers X 5 sample of each = 13000 Samples	98,26%

				(6), Local Grad (4)	(26 Alphabet) Training Set : 2600 Char Testing Set: 10,400	
[82]	English		√	Multi scale Neural Network	10 sample X 26 Alphabets = 260 10 X 10 numeric = 100	85%
[83]	English			Single Layer NN. Binary features	26 Characters only	80%
[84]	Arabic		√	Modified Cascaded ANN 5 features from each block of character	Total 100 Characters from 10 persons	67%
[87]	Arabic		√	Beta Elliptical Model		98.00%

The study in [69] explores the effectiveness of two textural measurements on signature verification for offline skilled forgeries. Signature images corresponding to 521 writers from the GPDS960 database were used to evaluate the performance of these features. The database contains 24 genuine signatures and 30 forged signatures for a total of 881 different individuals. The error rates of different features vary significantly with run-length features outperforming all other features reporting an error rate of 14.70%. The proposed run-length features come out to be the most effective in terms of error rates.

In [70] two methods namely LDA and Neural Network for the offline signature from the scan signature image of signature detection and verification for security of an individual person provided. The research has focused the comparative analysis in terms of FRR, SSIM, MSE, and PSNR. These parameters are compared with the early work and the recent work. The method in [70] shows the best appropriate matching signature with FRR of 3,6%.

The report in [71] focuses on offline signature verification, characterized by the usage of static(scanned) images of signatures, where the objective is to discriminate if a given signature is genuine (produced by the claimed individual), or a forgery (produced by an impostor). An overview of how the problem has been handled by several researchers in the past few decades and the recent advancements in the field is presented. Inspired by these advancements, the experimental results still report somewhat large error rates for distinguishing genuine signatures from skilled forgeries, when large public datasets are used for testing, such as GPDS. Error rates are at least around 7-8% in the best reported results, even when the number of samples for training is around 10-15 (results are worse with 1-3 samples per user, which is a common scenario in banks and other institutions).

In the paper [72] finer intensity-based features and global geometry-based features are explained. Particularly, the finer features are computed for every sample point of a signature using a histogram of intensities, and the geometry based features are extracted using an adaptation of the shape context descriptor. The system was submitted to the signature verification competition (SignComp2011) and archived the following results. For the Dutch dataset the system is interpreted up to an accuracy rate of 87.8 % with FRR=12.35 % and FAR=12.05 %. For the Chinese dataset the system is interpreted up to an accuracy rate of 72.9 % with FRR=27.5 % and FAR=26.98 %.

In [73] a text and user generic model for writer verification that uses a combination of pen pressure information from ink intensity and writing indentations obtained by a multiband image scanner is proposed. A writer-specific dissimilarity representation to consider individual handwriting characteristics that affect model performance is proposed. Experimental results obtained using handwriting samples collected from 54 volunteers are reported. The results show a decrease in error rate compared with conventional methods from 10.0% to 4.0%.

Some robust machines working under offline systems reach a success rate in writer recognition based on their handwriting biometric from 26% to more than 99% [125][136]. An OCR-driven writer identification, developed by [125], achieved a maximum accuracy of >95% in experiments with 2000 words per writer of a database of 259 persons. When 125 words are used, the accuracy is 26%.
OCR-Recognition of handwritten passwords plus writing speed and pressure [170], handwritten graphical passwords [171].When examined critically we can realize the following disadvantages:

- Every person has only one personal signature.
- If a forger can falsify the signature he gets access to all of these systems.
- OCR-Recognition of a password eliminates the biometrical features of the writer.
- It is hard to draw a graphical password.

Biometric authentication systems are more reliable and harder to falsify. Many recent studies haves focused on offline systems [142] or only use offline parameters [161][163][164][168]**.**

2.3.3 Online Writer Identification and Verification

Online handwriting recognition systems use characters, which are written on a touch sensitive surface [138]-[140]. They interpret the movements and the stylus of writing. For writer identification, a point distribution model is used in [138]. Using a database of 50 handwritten words per 12 persons, the system reaches an identification rate of 97.3%. In [139], the character image is partitioned into 8x8 blocks with four directions (vertical, horizontal, left slant, and right slant) for each block (256 features). With 6000 Japanese Hiragana handwritten characters from 3 users, an average recognition rate of 99.92% is achieved. In [141], character prototypes were used, and an accuracy of 99.2% is reached with a database of 120 writers, 160 characters per writer. For the writer verification of 370 writers by four repetitions of the same handwritten word on an A4 tablet catalogues of pen-up and pen-down strokes are used [139]. The identification rate ranges from 80.6% to 95.6% when analysing a single word. The identification rate increases up to 99.7%, when two words are combined. Handwritten character recognition as an internet based services proposed in [141]. The characters of 190 users and 3.6 million samples are recognized by directional feature extraction and gradient feature extraction as feature extraction methods. The tests on different mobile OS achieved an average recognition rate of 96.17% in 29.5 milliseconds.

Compared to traditional authentication methods, such as input of password from keyboard, handwriting biometrics have been studied in the last years and used in this area because they are more reliable and harder to falsify. Previous research has been concentrated on writer identification using handwritten signature [173]-[174], handwritten signature plus keyboard written password or pin [155][156].

Online systems consider time, velocity [169][174], acceleration [175][177] and pressure features [178][179]. Computing time is studied mostly for online or offline parameter using systems [161][163][164][168]. But a high performance is more relevant for online authentication systems.

In [88] verification based on discrete wavelet transform (DWT) is carried out. Experiments are done with a signature database with 5 users, 20 genuine and 20 skilled forgery signatures. The recognition success rate for genuine signatures is 95% and the recognizing forgery signature as genuine is down of 9%.

In [89] new features called the forward and backward variances of signature for on-line signature verification proposed. In the proposed method, stable features of signature can be characterized by the forward and backward variances of signature. It is shown that signature can be verified by evaluating the difference of K-L coefficients of the forward and backward variances between the reference and signature to be verified. Experiments are performed on public signature database MCYT 100, consists of 5000 signatures from 100 people. The Equal Error Rate is 4.49%.

In [90] a set of rotation invariant descriptors is presented. The complexity of system because of high dimensionality of signature features is reduced by exploiting the statistical moments of the wavelet coefficients. To improve it further, suggested Fisher-Metric(FM) filtering develops a feature selection algorithm. In the end, taking advantage of artificial neural network on SVC2004 database blow up the performance of the proposed approach with False Acceptance Rate (FAR) of 3% and False Rejection Rate (FRR) of 2.5% with an Equal Error Rate (ERR) of less than 3%.

In [91] a non-parametric statistical test for a comparison of features and the verification of signatures is applied. The x, y coordinates, pressure, and velocity features both separately and combined where tested on the Dutch dataset of SigComp2011.For comparison purposes, the Dynamic Time Warping (DTW) distance for pressure and velocity features are included 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%.

The paper [92] describe the usage of the touch screen of smartphones or tablets in order to collect handwritten signature images and associated biometric markers derived from the motion direction of handwritten signatures that are written directly into the device touchscreen. These time base biometric markers can then be converted into signalling time waves, by using the dragging or lifting movements the user makes with a touch screen omnidirectional tip stylus, when he writes his signature at the device touchscreen per hand. Using the Euclidean Distance, it is possible to start verifying handwritten signatures by just using distances. It is also possible to obtain meaningful authentication results that tell us whether it is the legitimate user which is signing or if he is being impersonated by some other person. The results are improved using the time dimension for the verification process. In this case, the improvement was 10%.

In [93] a novel scheme, based on the Kinematic Theory of rapid human movements and its associated Sigma Log Normal model, to improve the performance of on-line signature verification systems is proposed. The approach combines the high performance of DTW-based systems in verification tasks, with the high potential for skilled forgery detection of the Kinematic Theory of rapid human movements. Experiments were carried out on the public available Biosecur ID multimodal database, comprising 400 subjects. Results show that the performance of the DTW-based system improves for both skilled and random forgeries. The database was divided into three independent sets so that results were not biased. Results showed improvements at all operating points, reaching around a 36% improvement at the EER.

The paper [94] describes a complete biometric algorithm for signature verification based on three stages. Signature is normalized by means of a pre-processing that removes irrelevant information. Subsequently, the captured signature is aligned with its template by applying a DTW algorithm. From this aligned signature the most salient features are extracted and used as input to a GMM model, whose output is used to confirm or deny the user's identity.
Recognition results: The accuracy of the proposed signature verification algorithm was tested on the MCyT database, which includes 100 users and contains 25 genuine signatures and 25 skilled forgeries for each one. Note that the best Equal Error Rate (ERR) (the parameter usually used for determining the quality of a biometric algorithm) is 2.74%, which is a good result for most signature verification systems. This accuracy level is comparable with those obtained by other biometric modalities. For instance, recent international competitions on fingerprint verification showed an average ERR ranged from 2% to 3% using different databases.

In [95] a signature is classified using a multidomain strategy. A signature is first split into different segments based on the stability model of a signer. Then, according to the stability model, for each segment, the most profitable domain of representation for verification purposes is detected. In the verification stage, the authenticity of each segment of the unknown signature is evaluated in the most profitable domain of representation. Signatures in the SUSIG database of handwritten signatures were used to test the system, according to the leave-one out strategy. The SUSIG database is composed of two sections: "Blind subcorpus" and "visual subcorpus." This paper used the table II signature verification results for the SUSIG database. When each stroke is verified using only the most stable domain of representation, are FRR = 2.15% and FAR =2.10%, when each stroke is verified using all the domains of representation, the verification results of the same system are FRR = 3.60% and FAR = 4.15%.

In [96] an online signature verification technique based on discrete cosine transform (DCT) and sparse representation is proposed. We find a new property of DCT, which can be used to obtain a compact representation of an online signature using a fixed number of coefficients, leading to simple matching procedures and providing an effective alternative to deal with time series of different lengths. Two databases are used SUSIG and SVC2004. SUSIG-Visual used in [96] contains a total of 1880 genuine signatures by 94 persons and 940 skilled (half are highly skilled) forgeries collected in two sessions from 94 persons. SVC2004 database contains signature data collected from a graphic tablet (WACOM Intuos) with 20 genuine signatures and 20 skilled forgeries per person. SUSIG-Visual (for “common threshold + skilled forgeries”) shows a result with EERs are 2.98%, 3.03%, and 2.46%, respectively.

In [97] the scarcely addressed case of only one available signature for training, the use of modified duplicates is proposed. The novel technique relies on a fully neuromuscular representation of the signatures based on the Kinematic Theory of rapid human movements and its Sigma-Lognormal model. These artificial duplicated signatures have demonstrated their utility in a reference set improving the performance of DTW. MCYT-330 Contribution I RI- EER I SI- EER Duplicated + HMM. [I] I 6.60 % I 15.60 % SUSIG Visual Sub-Corpus Contribution I RI- EER I SI- EER Duplicated + DTW (this work) I 3.61 % I 7.87 %

The proposal in [98] uses the Shannon Entropy, the Statistical Complexity, and the Fisher Information evaluated over the Bandt and Pompe symbolization of the horizontal and vertical coordinates of signatures. The proposal [98] is carried out on the freely available and widely used handwritten signatures database MCYT-100 subset of 100 persons. The results this proposal using five (ten, respectively) training samples, are ERR(%) = 0.19 (0.17, respectively).

In [99] a novel online finger-drawn PIN authentication technique is proposed that lets a user draw a PIN on a touch interface with her finger. The system provides some resilience to shoulder surfing without increasing authentication delay and complexity by using both the PIN as well as a behavioral biometric pattern in user verification. Our approach adopts the Dynamic Time Warping (DTW) algorithm to compute dissimilarity scores between PIN samples. To evaluate the system two shoulder surfing scenarios: 1) PIN attack where the attacker only knows the victim’s PIN but has no information about its drawing characteristic and 2) Imitation attack where an attacker has access to a dynamic drawing sequence of victim’s finger-drawn PIN in the form of multiple observations. Experimental results with a data set of 40 users and 2400 imitating samples from two attacks yield an Equal Error Rate (EER) of 6.7% and 9.9%.

The algorithm in [100] extracts the position coordinates of extreme points of reference signature and test signature in the signature curves, and then uses discrete Fréchet distance as the measure of the curve distance, respectively matching peak points and peak points, and valley points and valley points. By comparison of the same person’s signatures during different periods in signature sample base, the FRR(Fault Rejection Rate) is about 1.8%. By comparison of genuine signatures and random forged signatures in signature sample base, the FAR(Fault Acceptance Rate) is about 5%. By comparison of genuine signatures and skilled forged signatures, FAR is about 10.1%. According to the above experimental results, the used method based on discrete Fréchet distance can well judge the random forged signatures and skilled forged signatures, and it can also accurately verify genuine signatures.

In [101] writer verification of handwritten kanji characters on a digitizing tablet is investigated. A set of 1,230 samples is collected for this tablet from 41 subjects: 6 females and 35 males, including two left-handed subjects. In our experiment, the subjects were asked 5 times to write 4 kanji characters in a 24mm x 60mm rectangle printed on a sheet fixed on the tablet. After a certain time, interval of about a week or more, each subject inputted the same kanji samples. This data collection was done six times. In total, 5x6x41, or 1,230 samples are used for the following writer verification study. The experimental results have shown that this method is significant enough to verify a writer, since it has yielded an EER of about 1%.

In [102] a different and perhaps more ominous threat: the possibility that the attacker has access to a generative model for the behavior in question, along with information gleaned about the targeted user, and can employ this in a methodical search of the space of possible inputs to the system in an attempt to break the biometric is described. For the experiments, several small data sets were created. Two writers (the authors) wrote four different passphrases 20 or more times, which resulted in a total of 154 samples. The handwriting was collected using a Wacom Intuos digitizing tablet. By using the concatenative attack described in this paper, a 49% success rate could be achieved.

In [103] a general biometric hash generation scheme based on vector quantization of multiple feature subsets selected by genetic optimization is presented. Development and evaluation experiments are reported on the MCYT signature database, comprising 16,500 signatures from 330 subjects. It’s observed that the best string (feature subset) converges to an EER value of about 24% for skilled forgeries.

In [104] a new approach for searching within handwritten documents without textual recognition is proposed. The approximate string searching technique is utilized, known from the domain of bioinformatics, where it is typically used for finding pieces of gene sequences. Four different feature types for converting handwriting signals into strings are discussed, that are able to use the string searching algorithm. While evaluating the system with an own database of handwritten documents results of precision and recall rate of each of 81.5%. are achieved.

In [105] a DTW-based on-line signature verification system is presented and evaluated. The system is specially designed to operate under realistic conditions, it needs only a small number of genuine signatures to operate and it can be employed in almost any signature capable capture device. The system has been evaluated using four on-line signature databases (MCYT, SVC2004, BIOMET and MyIDEA) and its performance is among the best systems reported in the state of the art. Average EERs over these databases lay between 0.41% and 2.16% for random and skilled forgeries respectively.

2.4 Discussion

In this chapter is given a review of the methodology with biometric methods and their demarcation. The biometric methods for verification and identification were considered, important terms clarified and procedures for the evaluation of the results presented. The most widely used state-of-the-art databases were presented, and the results of the writer identification and verification were subdivided into off-line and online identification and verification.

The records were mostly created with a digitizing tablet, whiteboard with digital pen or signature pad over a long time period. Apart from offline data (scanned signatures or screenshots), the online data were also collected with the specific features such as coordinates, time, pressure and angle information.

We have taken a look at the public databases, which are mainly composed of records with signatures, single characters, words, and whole sentences.

The results in off-line and online publications for writer identification and verification are mainly based on normalization and segmentation with online features made up of coordinates, angle information, time, speed and pressure information, with off-line features such as primarily geometric features. For the classification, identification and verification, after filtering the features with filters such as Fisher Score, classifiers off-line like HMM and GMM based, online classifiers like NN, DTW, and other standard based classifiers are used often.

The results of the individual investigations vary strongly. This overview provides a summary of the results of the publications under investigation:

Accuracy	between	**67%**	and	**99%**
FAR	between	**2.5%**	and	**26.98%**
FRR	between	**1.5%**	and	**24.13%**
EER	between	**0.41%**	and	**27.7%**

Most prior research of writer identification and verification has been limited to the analyses of handwritten text or signature.

Previous research neglected to analyze handwritten passwords or pins for a safe system access. Furthermore, there exist no databases with handwritten safe passwords or pins. That is one point why new databases with safe passwords have to be created for the investigation of whether handwriting passwords are suitable on mobile devices. The handwritten secure passwords and signatures should be created using a standard tablet and smartphone.

2.5 Conclusion

In addition to the existing data bases, new databases with handwritten passwords for the investigations and the comparison must be created for the identification of the authors with handwritten passwords. The data must be captured with devices such as a tablet and a smartphone to show that the system to be developed is well-suited for authenticating on standard devices.

The offline and online features from the examined state of the art must be checked for the expediency specifically for handwriting passwords, filtered, if necessary new features are developed, investigated and tested in the system. The classifiers must also be tested for their suitability and tested with the data sets and features to find out which specific classifiers are best suited. For the real, to implement solution, the required time for feature extraction and classifying plays a decisive role.

Chapter 3 Concept of the system

Previous research neglected to analyze handwritten passwords for a safe system access and there exist no databases of handwritten safe passwords. Therefore, a concept for a system for following purposes:

1. Capturing data of handwritten passwords
2. Extraction of features
3. Classification
4. Model building
5. Writer identification and verification

has to be developed.

In this chapter a Client Server System of the four steps: Data Collection, Feature Extraction, Feature Selection and Classification with two modes (enrol and test mode) is proposed.

3.1 Password Verification – Client and Server

For online writer identification and verification, the mobile client server system as presented in [30][123] is used and further developed. (see Figure 3.1)

Short introduction of the concept from user view:

First the user enters their username using the keyboard of the mobile device. Then the handwritten password is entered via touch screen. The online algorithms on the mobile phone provide a first pre-processing of the handwritten password. The result is a file with the features of their handwritten password. The mobile phone sends the result file to the server. The server starts the feature extraction and uses the model from the classification system. The result, whether the user has access or not, is sent back to the mobile phone and user can get access to their virtual account, in case of a positive recognition. Moreover, in this mobile phone server solution, combined an online and offline writer recognition method, and verified the writer with the help of their handwritten password by the implementation of high performance verification methods.

This application can be used as an added value to the classical methods. The biometric information of a handwritten password for instance can be used for virtual access to a bank account via smartphone.

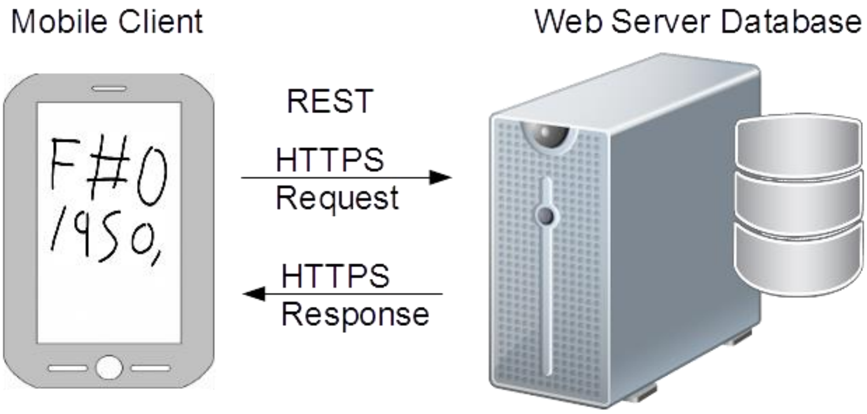


Figure 3.1: Online writer identification as a Mobile Client-Server Solution

For the proposed concept, a prototype with an Android smart phone as mobile client and a Linux system with a database at server site is developed. For a more secure, state of the art well-structured service the REST (Representation State Transfer) program paradigm is used with HTTPS as transfer protocol. The purpose of REST is to focus on machine-to-machine communication. REST is a simple alternative to similar procedures used in the previous implementation with SOAP, WSDL, and the related RPC method. The tight guidelines of REST help to build well-structured services and support the use of clean URLs. This is one of the main advantages of the proposed REST-Server architecture.

The proposed concept for writer identification and verification works with two modes:

1. **Enrol Mode** for training the system with the writer data (see Figure 3.2, grey part):

The system is divided into ten steps in the Enrol Mode for collecting data and building the model. On the mobile client there is a program to create a new user account. The user enters at first their username (see Figure 4.2, step 1) by keyboard and then their handwritten password (step2) by touch screen. The client pre-processing program (step 3) detects the points and time coordinates in the written characters and recognizes the segments of the handwriting. Users have to write a defined number of their handwritten password for building the model.
In case of a writing error the user will have the possibility to correct the mistake (step 4). After every successful input of a handwritten password, it will be stored in a file, whose identification is the username. After preprocessing where the segmentation is carried out (step 3) the file with point coordinates and time values contained will be sent to the server (step 5). The system saves minimum three files with offline parameters for each user. The number of files the system

sends depend on the number of handwritten passwords the user has to write this is depending on the experiments adjustable in the System. The next step is the feature extraction taking place on the server (step 6). The classification follows (step 8) and finally, the model is built (step 9) and stored in the database (step 10).

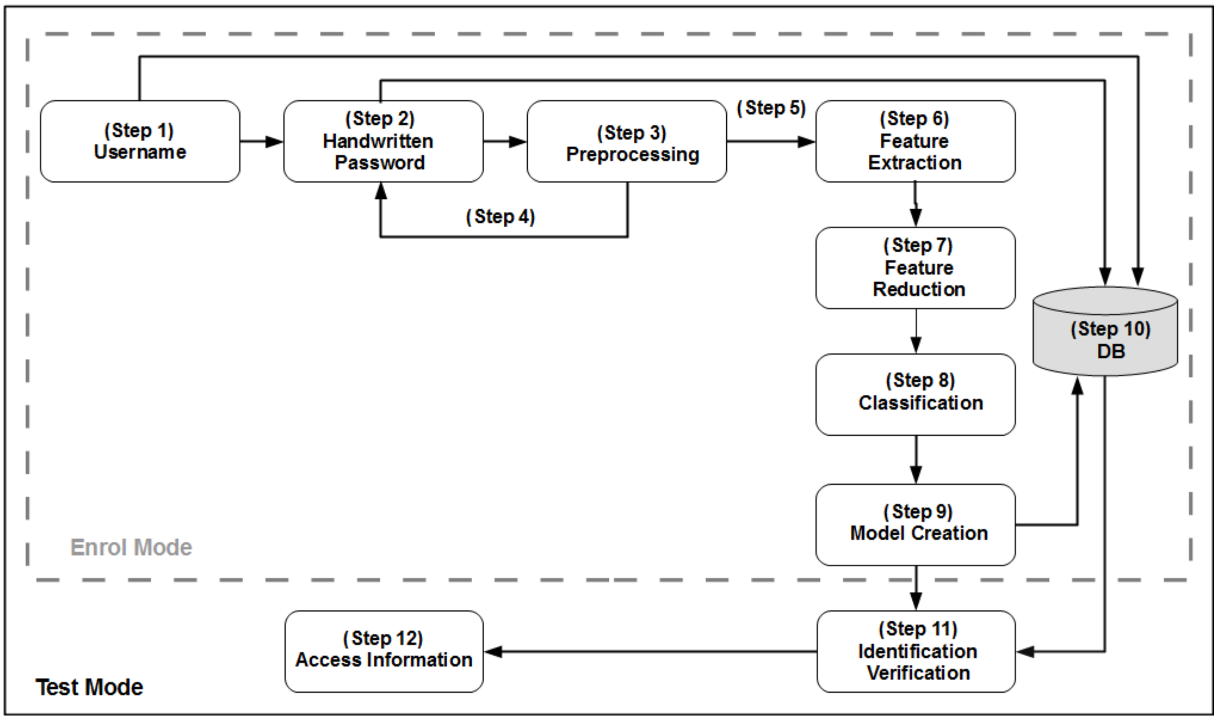


Figure 3.2: Enrol and Test Mode for writer identification and verification [133]

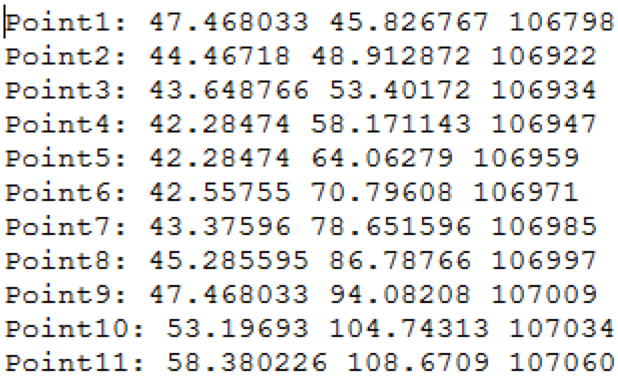
2. **Test Mode** Test Mode for the access control (see Figure 3.2, black part):

The 12 steps of the Test Mode are shown in the black part of Figure 4.2. At first, the user enters their username by keyboard (step 1) and then their handwritten password by touch screen (step 2). The offline pre-processing algorithm for the Test Mode collects the character point coordinates and time and generate segment data from the handwritten password. The user can cancel and repeat their writing (step 4). The touch screen phone sends the file containing the online information to the server (step 5). On the server, the feature extraction is applied to offline and online data (step 6). Feature reduction carried out (step 7) and classification (step 8) is done with the selected features. In the next step, the identification or verification starts with the model built in the Enrol Mode (step 11) and stored in the database (step 10). Finally, the result is sent back to the mobile phone of the client and the user can get access to the system or not (step 12).

3.1.1 Client

For the mobile client the pre-processing program was developed with parameter sampling like point coordinates, time values and touch down and touch up information for the identification of the segments. The features *segments, time*, and *points* were extracted at mobile client.

In order to read data from the mobile client, an application was written in JAVA using Android SDK. During the recording process, automatic selected pixels will be saved in a plain text format, every single pixel of which is represented by a pair of XY value in Cartesian coordinate system and saved in a separate line. A third value indicates the timestamp, which is exactly the moment the pixel is captured (See Figure 3.3).



```
Point1: 47.468033 45.826767 106798
Point2: 44.46718 48.912872 106922
Point3: 43.648766 53.40172 106934
Point4: 42.28474 58.171143 106947
Point5: 42.28474 64.06279 106959
Point6: 42.55755 70.79608 106971
Point7: 43.37596 78.651596 106985
Point8: 45.285595 86.78766 106997
Point9: 47.468033 94.08208 107009
Point10: 53.19693 104.74313 107034
Point11: 58.380226 108.6709 107060
```

Figure 3.3: Sample data arranged in an index-x-y-time sequence

In addition to the online data, the screenshot of the signature is stored as PNG for the generation and analysis of the offline features.

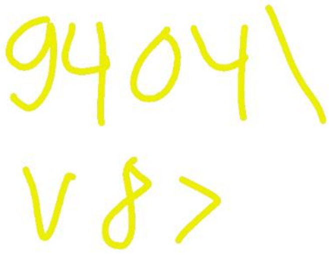
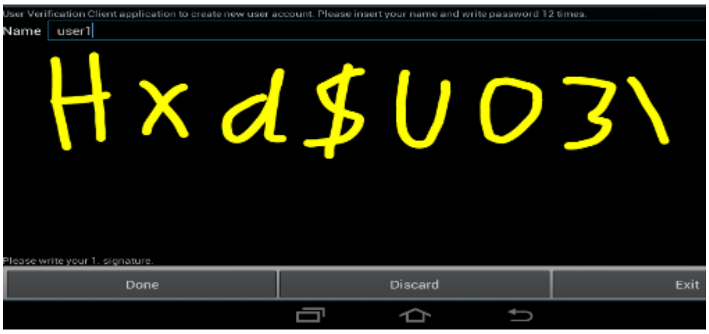


Figure 3.4: Image of password in PNG format

The user enters their username into a text field and his handwritten password into the drawing area. Figure 3.5 shows the client application (Enrol Mode) and Figure 3.6 (Test Mode) on the Android mobile device.

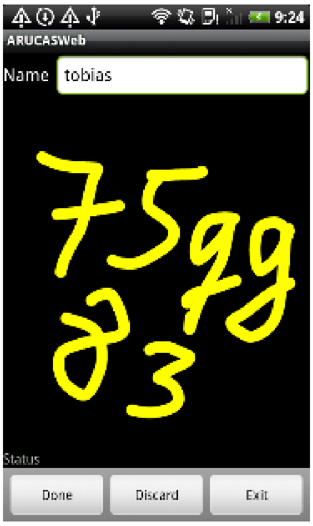


User Verification Client application to create new user account. Please insert your name and write password 12 times.

Name user1

Please write your 1. signature.

Figure 3.5: Mobile Client Enrol Mode



ARUCASWeb

Name tobias

Status

Figure 3.6: Mobile Client Test Mode

Depending on the experiment different devices are used, so for example the test user has to write the signature with finger or capacitive pen at android smartphone or tablet or digitizer pad (see Figure 3.7).

In the test version the written password is visible for evaluation purposes, in the real application it is invisible.

Figure 3.7: Bamboo Pad with pen

The collected username and features of the client pre-processing is transferred to the server for further processing and storage of the data.

3.1.2 Server

The data transferred by the client with HTTPS is stored on the server in a PostgreSQL database and processed using the algorithms of the data mining tool Weka [188], which are implemented in the REST Web service. The features are extracted and the model is generated and stored in the database using parameter reduction algorithms and the classifier in the Enrolment Mode.

In the Test Mode the features are extracted from the data transferred by the client and an identification / verification is carried out with the help of the generated model. The result is sent back to the client.
The original and preprocessed handwritten passwords have to be stored in a database (see Figure 3.8).

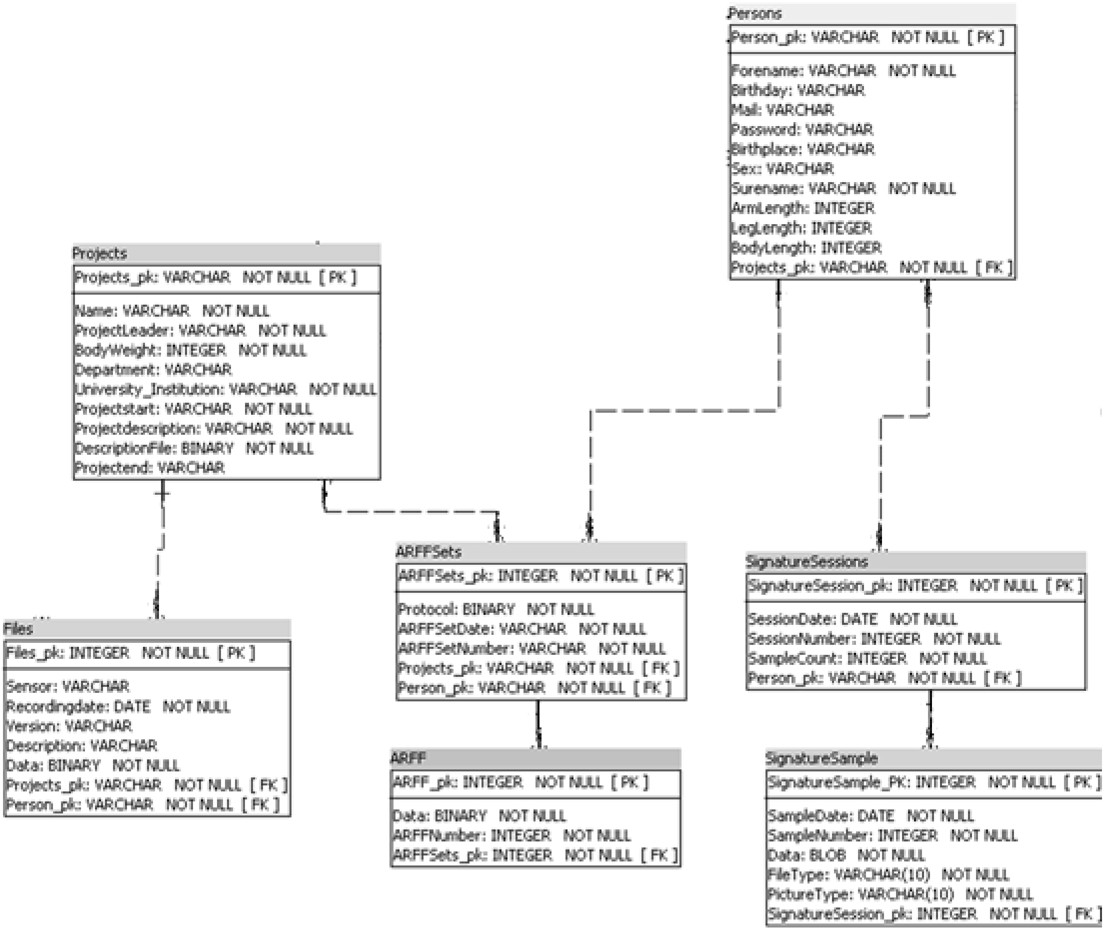

Persons
Person_pk: VARCHAR NOT NULL [PK]
Forename: VARCHAR NOT NULL
Birthday: VARCHAR
Mail: VARCHAR
Password: VARCHAR
Birthplace: VARCHAR
Sex: VARCHAR
Surename: VARCHAR NOT NULL
ArmLength: INTEGER
LegLength: INTEGER
BodyLength: INTEGER
Projects_pk: VARCHAR NOT NULL [FK]
Projects
Projects_pk: VARCHAR NOT NULL [PK]
Name: VARCHAR NOT NULL
ProjectLeader: VARCHAR NOT NULL
BodyWeight: INTEGER NOT NULL
Department: VARCHAR
University_Institution: VARCHAR NOT NULL
Projectstart: VARCHAR NOT NULL
Projectdescription: VARCHAR NOT NULL
DescriptionFile: BINARY NOT NULL
Projectend: VARCHAR
ARFFSets
ARFFSets_pk: INTEGER NOT NULL [PK]
Protocol: BINARY NOT NULL
ARFFSetDate: VARCHAR NOT NULL
ARFFSetNumber: VARCHAR NOT NULL
Projects_pk: VARCHAR NOT NULL [FK]
Person_pk: VARCHAR NOT NULL [FK]
SignatureSessions
SignatureSession_pk: INTEGER NOT NULL [PK]
SessionDate: DATE NOT NULL
SessionNumber: INTEGER NOT NULL
SampleCount: INTEGER NOT NULL
Person_pk: VARCHAR NOT NULL [FK]
Files
Files_pk: INTEGER NOT NULL [PK]
Sensor: VARCHAR
Recordingdate: DATE NOT NULL
Version: VARCHAR
Description: VARCHAR
Data: BINARY NOT NULL
Projects_pk: VARCHAR NOT NULL [FK]
Person_pk: VARCHAR NOT NULL [FK]
ARFF
ARFF_pk: INTEGER NOT NULL [PK]
Data: BINARY NOT NULL
ARFFNumber: INTEGER NOT NULL
ARFFSets_pk: INTEGER NOT NULL [FK]
SignatureSample
SignatureSample_PK: INTEGER NOT NULL [PK]
SampleDate: DATE NOT NULL
SampleNumber: INTEGER NOT NULL
Data: BLOB NOT NULL
FileType: VARCHAR(10) NOT NULL
PictureType: VARCHAR(10) NOT NULL
SignatureSession_pk: INTEGER NOT NULL [FK]

Figure 3.8: Part Data Model DBMS server

The classes for storing the data of the REST Service transfer objects are shown in figure 3.9.

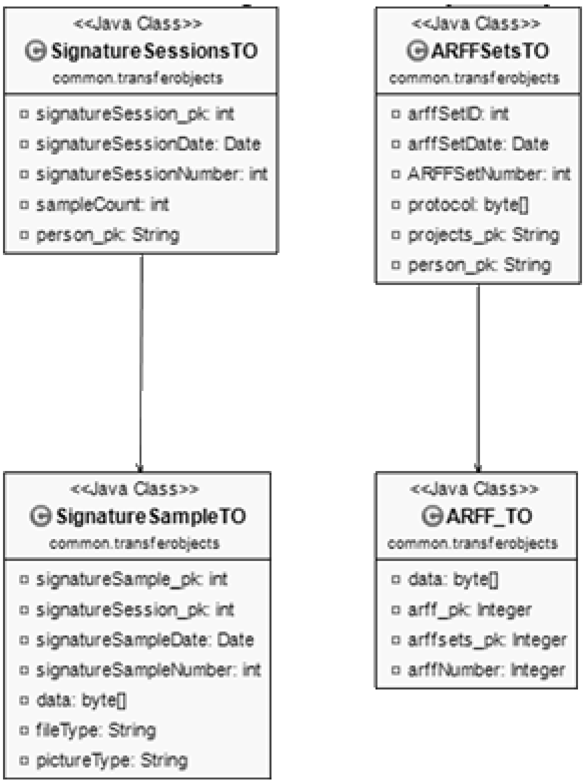


Figure 3.9: Part Class Transfer Objects REST Web service

After the description of the whole structure of client server system an overview of the feature extraction, classification methods and the used databases follows.

Chapter 4 Methods and Materials

This chapter provides an overview of the methods used in data collection, preprocessing, feature extraction, classification and materials like datasets and databases used for the experiments.

The process of building an identification and verification system consists of a training phase and a classification phase [106]. The training process essentially consists of 3 components: preprocessing, feature extraction and classification, regardless of the type of the input. This applies not only to the handwriting, but also to all other identification and verification procedures or general pattern recognition tasks. (see Figure 4.1)

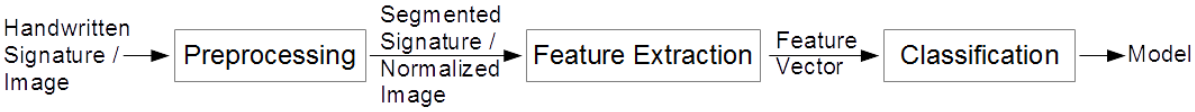

Handwritten Signature / Image
Preprocessing
Segmented Signature / Normalized Image
Feature Extraction
Feature Vector
Classification
Model

Figure 4.1: training phase handwriting identification and verification

In the training phase a model is built and trained with handwritten signatures. In the classification phase, the system has to classify a handwritten signature from an unknown writer (identification) see Figure 4.2 or from a known writer (verification) see Figure 4.3. In the test phase, the trained model is used to identify or verify the writer.

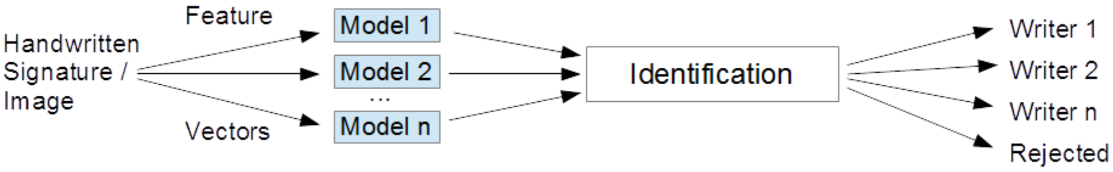


Figure 4.2: writer identification

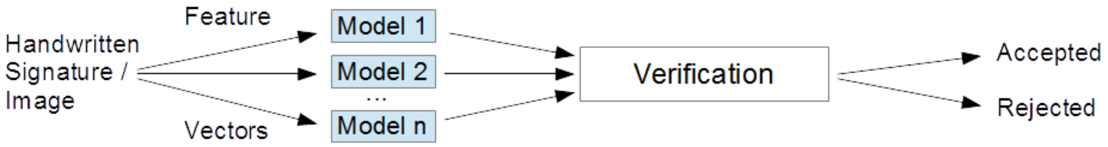


Figure 4.3: writer verification

The input media for the process of writer identification / verification consist either of paper or stored (scanned) images (off-line) or of a digital input medium such as smartphone, tablet, digital pen or pad with time and pressure information (on-line). Let's start with the databases, which are presented in further detail in the following chapter.

4.1 Databases

In [60] the writer identification was carried out with a self-developed database of 280 handwritten block letter words, 176 originals and 104 forgeries. To validate the extracted features and classification algorithms for a secure authentication system, public databases were search.
As already mentioned in chapter 2.2, there are a large number of databases available for handwriting recognition, but only a few are available special for the task of handwriting verification. This paragraph provides an overview and detailed description of the used and adapted public databases as well as the self-developed databases.

4.1.1 Public Databases

- **ATV-Signature Long Term Database (ATV- SLT DB)**

The On-Line Signature Long-Term Database contains 6 sessions generated by 27 users over 15 months. The quality change of the verification results over a 15 months-period of is examined. For the impostor test, a 7th session was added 3 years after developing the first datasets, the impostor session, with 6 signatures for each user [178]. (see Figure 4.4)

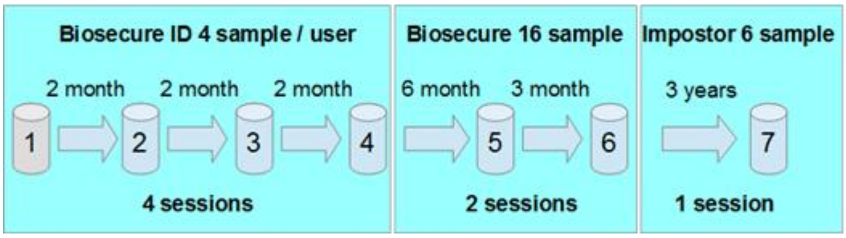


Figure 4.4: Signature Long Term DB and Impostor Session 0

For the generation of the impostor signatures it was necessary to generate the images of the original signatures. These images were generated as PNG graphic from the SVC files with a Java program. The counterfeiter in the experiment used this as a template to falsify the signatures. (see Figure. 4.5).

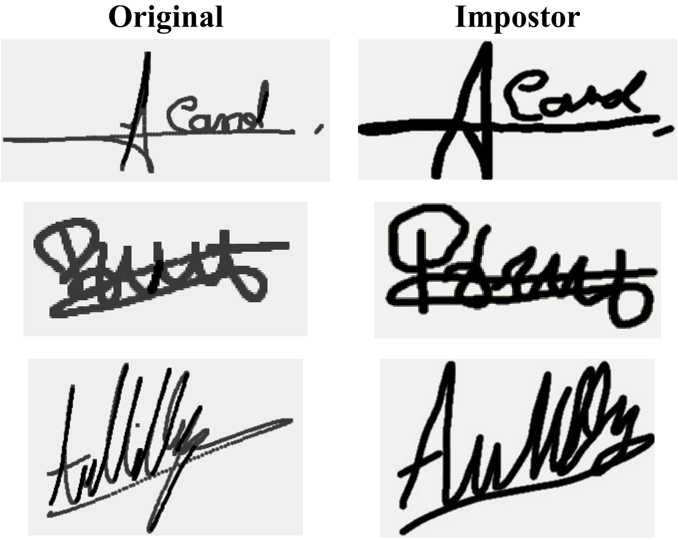
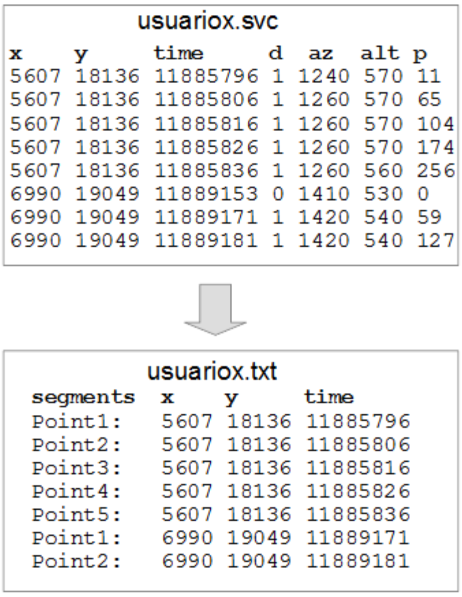


Figure 4.5: Samples: PNG from SVC and Impostor PNG [178]

To validate the features of [60] with the data from ATV-SLT DB the given SVC Files were transformed to the required data format. With a Java program the following transformation was done (see Figure 4.6).



```
usuariox.svc
x    y     time     d  az   alt p
5607 18136 11885796 1  1240 570 11
5607 18136 11885806 1  1260 570 65
5607 18136 11885816 1  1260 570 104
5607 18136 11885826 1  1260 570 174
5607 18136 11885836 1  1260 560 256
6990 19049 11889153 0  1410 530 0
6990 19049 11889171 1  1420 540 59
6990 19049 11889181 1  1420 540 127
```

```
usuariox.txt
segments  x    y     time
Point1:   5607 18136 11885796
Point2:   5607 18136 11885806
Point3:   5607 18136 11885816
Point4:   5607 18136 11885826
Point5:   5607 18136 11885836
Point1:   6990 19049 11889171
Point2:   6990 19049 11889181
```

Figure 4.6: SVC to TXT transformation [178]

- **IAM online handwriting DB**

To determine the possibility of using the same features like in [60] for writer verification with handwritten single character words, the experiments are working with the IAM online handwriting database with cursive texts for verification. The database contains 1760 handwritten text examples of 220 persons. The text contains between 40 and 60 words on one and more than one line. The IAM online handwriting database considers 8 samples of each user. (see Figure 4.7)

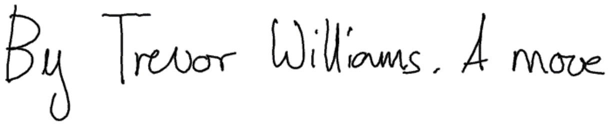
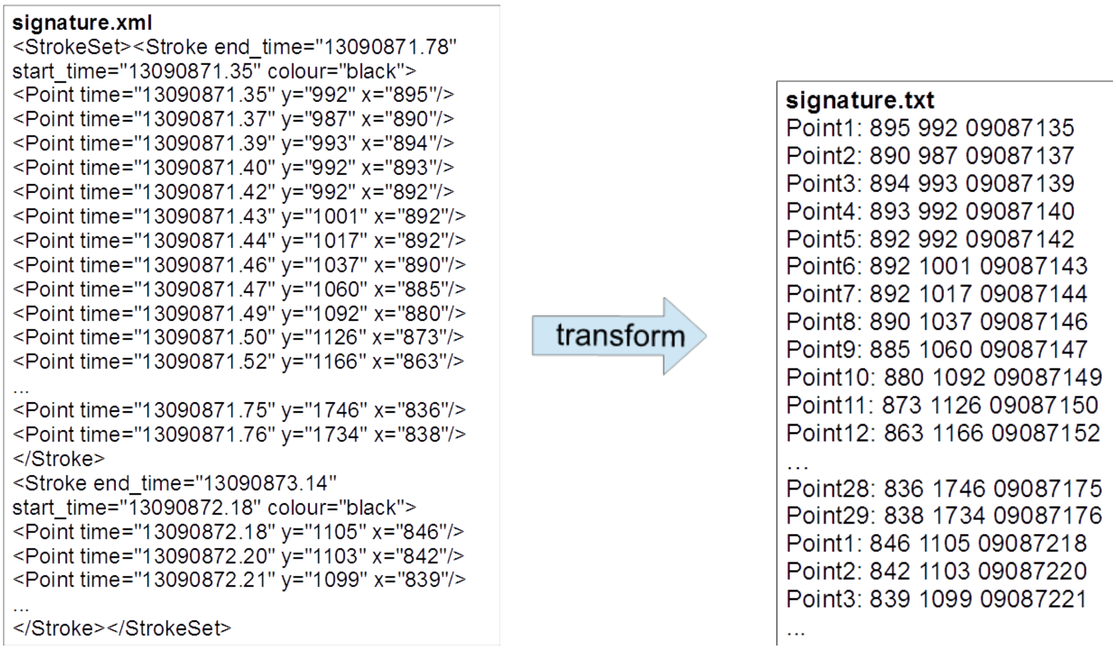


Figure 4.7: Sample of handwritten text of IAM online handwriting database

In order to use the proposed system (detailed information see chapter 4), the IAM data have to be transformed into a defined format. A XML parser program is used to transform the IAM data to our format. (see Figure 4.8)



signature.xml

```
<StrokeSet><Stroke end_time="13090871.78"
start_time="13090871.35" colour="black">
<Point time="13090871.35" y="992" x="895"/>
<Point time="13090871.37" y="987" x="890"/>
<Point time="13090871.39" y="993" x="894"/>
<Point time="13090871.40" y="992" x="893"/>
<Point time="13090871.42" y="992" x="892"/>
<Point time="13090871.43" y="1001" x="892"/>
<Point time="13090871.44" y="1017" x="892"/>
<Point time="13090871.46" y="1037" x="890"/>
<Point time="13090871.47" y="1060" x="885"/>
<Point time="13090871.49" y="1092" x="880"/>
<Point time="13090871.50" y="1126" x="873"/>
<Point time="13090871.52" y="1166" x="863"/>
...
<Point time="13090871.75" y="1746" x="836"/>
<Point time="13090871.76" y="1734" x="838"/>
</Stroke>
<Stroke end_time="13090873.14"
start_time="13090872.18" colour="black">
<Point time="13090872.18" y="1105" x="846"/>
<Point time="13090872.20" y="1103" x="842"/>
<Point time="13090872.21" y="1099" x="839"/>
...
</Stroke></StrokeSet>
```

signature.txt

```
Point1: 895 992 09087135
Point2: 890 987 09087137
Point3: 894 993 09087139
Point4: 893 992 09087140
Point5: 892 992 09087142
Point6: 892 1001 09087143
Point7: 892 1017 09087144
Point8: 890 1037 09087146
Point9: 885 1060 09087147
Point10: 880 1092 09087149
Point11: 873 1126 09087150
Point12: 863 1166 09087152
...
Point28: 836 1746 09087175
Point29: 838 1734 09087176
Point1: 846 1105 09087218
Point2: 842 1103 09087220
Point3: 839 1099 09087221
...
```

Figure 4.8: Transformation process

- **SVC 2004 DB**

As described before in paragraph 2.2 the SVC2004 DB consists of two separate signature verification tasks using two different signature databases. The signature data for the first task contain coordinate information only, but the signature data for the second task also contain additional information including pen orientation and pressure. The first task is suitable for SVC2004 participating teams on-line signature verification on small pen-based input devices such as personal digital assistants (PDA) and the second task on digitizing tablets.

The data was transformed for our experiments from the original format description into our format description in order to carry out the investigations:

In each text file, the signature is simply represented as a sequence of points. The first line stores a single integer which is the total number of points in the signature. Each of the subsequent lines corresponds to one point characterized by seven features listed in the following order:

X-coordinate - scaled cursor position along the x-axis
Y-coordinate - scaled cursor position along the y-axis
Time stamp - system time at which the event was posted
Button status - current button status (0 for pen-up and 1 for pen-down)
Azimuth - clockwise rotation of cursor around the z-axis
Altitude - angle upward toward the positive z-axis
Pressure - adjusted state of the normal pressure

```
2933 5678 31275775 0 1550 710 439
2933 5678 31275785 1 1480 770 420
3001 5851 31275795 1 1350 830 433
                 …
3794 3960 31275945 1 1320 770 465
```

Figure 4.9: SVC2004 user example.txt

Transformation process into our format was done by the Java program. Specification of the transformed data:

Segmentation begins new (count each position change) when pen down for writing
X-coordinate - scaled cursor position along the x-axis
Y-coordinate scaled cursor position along the y-axis
Time system time at which the event was posted

Basic data set of the output data after transformation:

```
Point1: 2933 5678 31275785
Point2: 3001 5851 31275795
Point3: 3114 6139 31275805
…
Point17: 3794 3960 31275945
```

Figure 4.10: SVC2004 transformed user example.txt

The described public databases are useful for the verification of quality of the extracted features, but for the writer identification in an authentication system a database of secure passwords is necessary.

4.1.2 Databases of BTU and ULGC

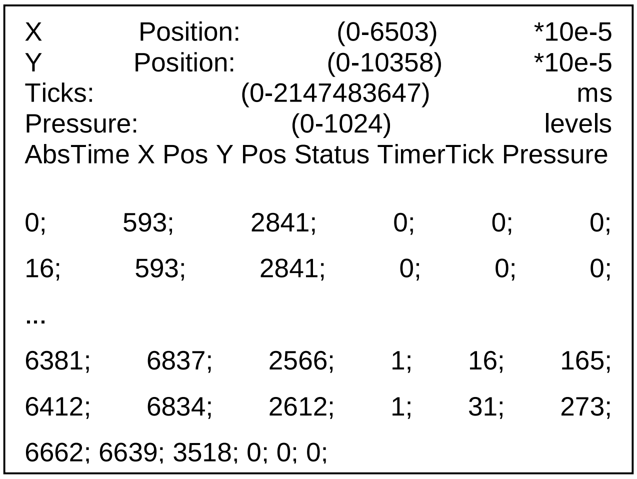
Databases with online handwritten secure password are not available. In order to meet the requirements of handwritten password identification with a mobile device and a secure password, to test and compare, different databases were developed at the BTU and the ULPGC.

a) BD (Base de Datos) online handwriting database

In order to carry out the examinations with handwriting, a self-developed large database with handwritten passwords from 150 users was available at the beginning of the work. However, further investigations of the database showed that this database is only conditionally suitable for the tests with handwritten secure passwords. A description of this database and why it is only conditionally suitable as follows in this paragraph. This is one of the reasons why further own databases have been developed during this thesis.

The database was created at the University of Las Palmas de Grand Canaria (ULPGC) with Tablet PC and digital pen. The database contains of 150 users stored in 150 folders with between 6 and 8 sample files per user with username and password. The data files are structured as follows:

It is captured and stored the maximum of x and y positions, how long the writer writes, and the maximum of pressure. In every row are given from left to right the absolute time, x position, y position, status, timer tick and pressure, and finally at the end of the file the password. (see Figure 4.11)



```
X          Position:         (0-6503)        *10e-5
Y          Position:         (0-10358)       *10e-5
Ticks:              (0-2147483647)               ms
Pressure:                (0-1024)            levels
AbsTime X Pos Y Pos Status TimerTick Pressure

0;        593;       2841;      0;       0;      0;

16;        593;       2841;      0;       0;      0;

...

6381;      6837;      2566;     1;      16;    165;

6412;      6834;      2612;     1;      31;    273;

6662; 6639; 3518; 0; 0; 0;
```

Figure 4.11: Biometric user file sample BD

Since there was no more detailed description of the records of the whole database, in a first step the data were analyzed randomly optically using a Java program, transformed into PNG graphics. It was visible which name and which signature the users actually wrote. And whether if the structure of each file is correct and if there are enough usable datasets for the tests.

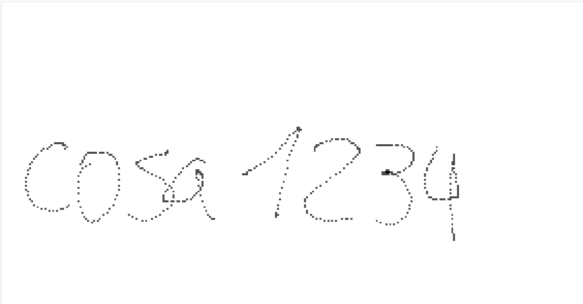


Figure 4.12: sample username and password

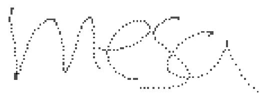


Figure 4.13: sample username

This first simple optical analysis led to the conclusion that many of the database records were not usable because empty, errors, gaps, different passwords, or usernames were written for the same user. This led to a more detailed investigation carried out in [124]. So we can see e.g. in the diagram figure 4.14 (number of points (y-axis) and height of the signature(x-axis)) there is a significant increase in the number of points for the data set and some clear outliers in the right half suggesting that most raw data were most likely manipulated by a small group of persons who have simulated different signatures. This pattern occurs in many variable pairs. These and the previous studies led to the decision that only a small part of the data set is usable for the tests. A detailed description of the investigations and results can be found in Chapter 5 Evaluation.

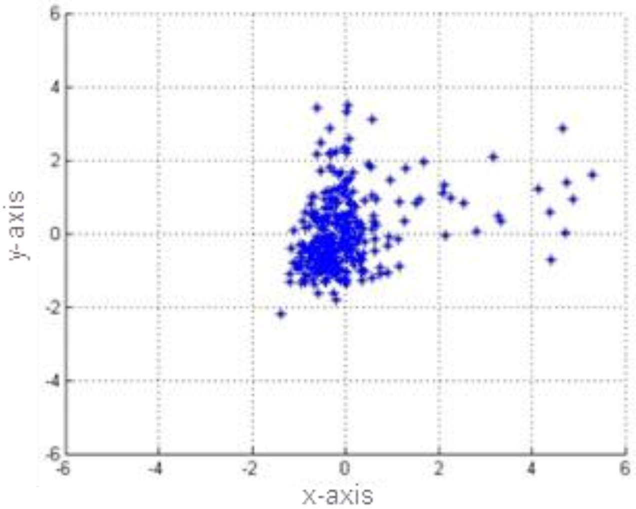


Figure 4.14: Displays the result of points for two variables (number of points (y-axis) and height of the signature (x-axis)) after PCA [124].

Further self-developed databases specially developed and adapted to the requirements of the thesis:

b) Password DB-9

The database contains of 108 handwritten genuine (12 samples came from nine randomly selected users) and 36 impostors (four false samples from nine randomly selected users) written on a HTC Desire mobile phone with Android 2.2. The signature samples are stored in folders sorted by user name, in the folders are the respective signatures as TXT and PNG. (see Figure 4.15 and 4.16) The impostors were collected with the same Android device and are stored in an extra folder in the same structure. The alleged counterfeiters knew the password.

```
Point1: 63.8 62.55833 8726438
Point2: 63.8 66.54999 8726555
Point3: 63.8 71.53958 8726573
Point4: 63.8 76.52916 8726596
…
Point40: 74.76 169.33 8728165
```

Figure 4.15: handwritten genuine sample TXT

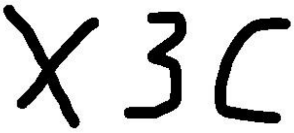


Figure 4.16: handwritten genuine sample PNG

c) Secure Password DB-32

For the secure password DB 32 the 32 test persons wrote generated safe passwords (single characters) with a length of 8 characters. The passwords were automatically generated from the page to create secure passwords [143] with a developed macro and stored into a list. A total of 384 training samples on a supervised system were created with different secure password character sequences like “Vty|b-sR” and written by 32 subjects subsequently [151]. The password character sequence was written repeatedly by each identical subject 12 times in one session. The handwritten secure passwords are collected with an Android Galaxy Tablet. The structure of the database is the same like Password DB 9 with TXT and PNG files (see figure 4.17).

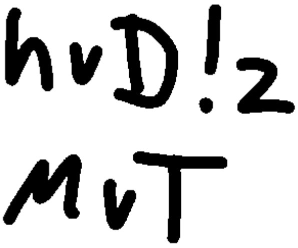


Figure 4.17: handwritten secure password sample PNG

d) Secure Password DB-150 with sessions

The Secure Password-DB-150 contains of 2700 handwritten secure passwords with 8 single characters of 150 users with 6 samples of each user collected in three sessions. The passwords where collected at ULPGC from 2015 to 2016.

Detailed structure of table for collecting safe handwritten passwords:

Table 4.1: Table for data collection secure password

# Users	Password	Nick	Sex	Age	ate of 1º Sessi	er of sa	te of 2º Sess	ber of sa	te of 3º Sessi	er of sa	ber o
50	9404\V8>	María	F	23	28.05.2015	6	02.06.2015	6	04.06.2015	6	18
51	L~k'wR_<	Patricia	F	36	28.05.2015	6	04.06.2015	6	16.06.2015	6	18
52	I,`vj'16	Antonio	M	34	28.05.2015	6	16.06.2015	6	16.06.2015	6	18
53	2bu\vR;`	Sandra	F	42	27.07.2015	6	29.07.2015	6	10.09.2015	6	18

The handwritten passwords are single characters, written in one or more lines (see Figure 4.18).

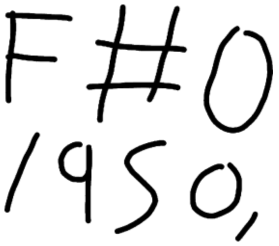
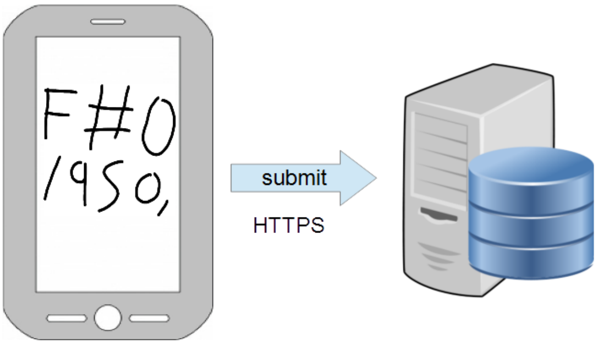


Figure 4.18: Sample of handwritten password in secure password-DB-150

The passwords are written on a display of an Android Tablet, preprocessed by a Java program, and transferred to a database at server (see Figure 4.19).



password.txt file generation
with user name, point,
segment and time information

store user password information
into database (Secure Password 100 DB),
generate features,store into database

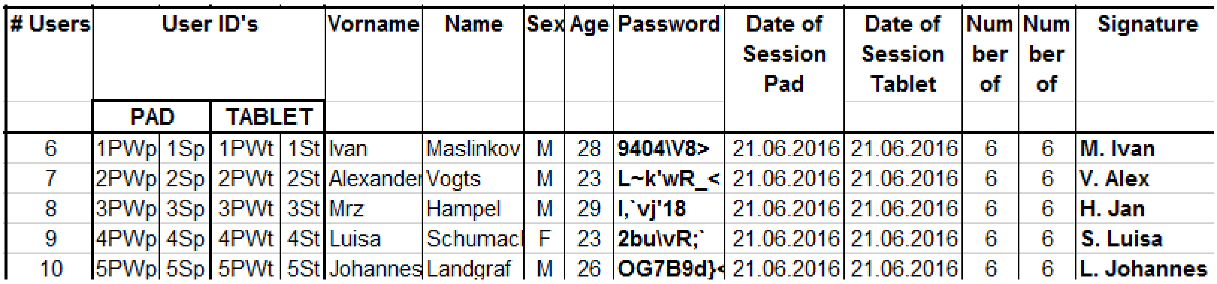
Figure 4.19: Structure of processing the Secure Password DB-150

- **Secure password and signature DB-30 (Wacom Pad and Android Tablet**

The database contains 720 handwritten secure passwords with 8 single characters and signatures of 30 users collected in one session at BTU in 2016. The secure password and signatures each of a user collected with Bamboo Pad (digital pen with pressure information) and Android Tablet (without pressure information).

Detailed structure of table for collecting safe handwritten passwords and signature with Pad and Tablet:

Table 4.2: Table for data collection secure password and signature pad and tablet



# Users	User ID's				Vorname	Name	Sex	Age	Password	Date of Session Pad	Date of Session Tablet	Num ber of	Num ber of	Signature
	PAD		TABLET											
6	1PWp	1Sp	1PWt	1St	Ivan	Maslinkov	M	28	**9404\V8>**	21.06.2016	21.06.2016	6	6	**M. Ivan**
7	2PWp	2Sp	2PWt	2St	Alexander	Vogts	M	23	**L~k'wR_<**	21.06.2016	21.06.2016	6	6	**V. Alex**
8	3PWp	3Sp	3PWt	3St	Mrz	Hampel	M	29	**I,`vj'18**	21.06.2016	21.06.2016	6	6	**H. Jan**
9	4PWp	4Sp	4PWt	4St	Luisa	Schumacl	F	23	**2bu\vR;`**	21.06.2016	21.06.2016	6	6	**S. Luisa**
10	5PWp	5Sp	5PWt	5St	Johannes	Landgraf	M	26	**OG7B9d}<**	21.06.2016	21.06.2016	6	6	**L. Johannes**

The handwritten passwords are single characters, written in one or more lines. The handwritten signatures are single characters and together written characters (signs) written in one line. (see Figure 4.20 and Figure 4.21)

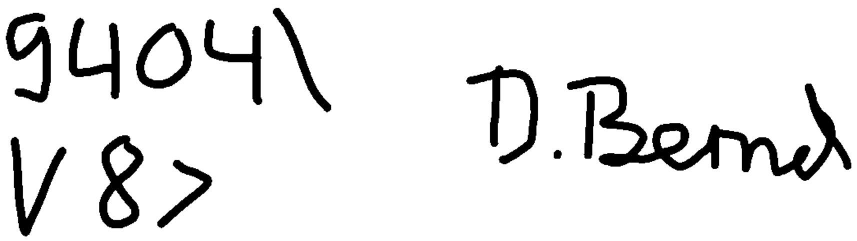


Figure 4.20: Sample of handwritten secure password and signature tablet

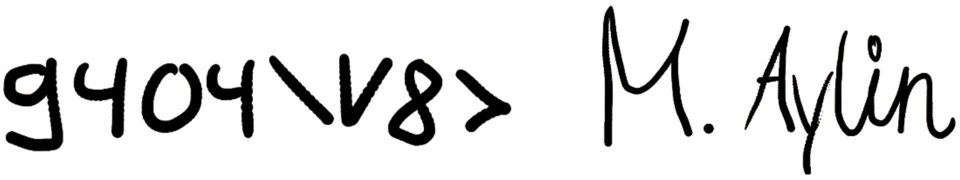


Figure 4.21: Sample of handwritten secure password and signature pad

The passwords and signatures are written on a display of an Android Tablet and Signature Pad with digital pen, preprocessed by a Java program.

4.2 Preprocessing

In preprocessing also needs to distinguish between online handwriting data and offline handwriting data. Off-line data consist of binary or gray-scale images, where in the online data are features like point information (X and Y coordinates), time and pressure information are included. The methods for preprocessing of the data are strongly dependent on both the input media and the writing style. Different variants for scanning, image enhancement and normalization, as well as different types of features, are briefly described here according to offline procedure and online procedure.

The following methods essentially depend on the type of the input data:

Online:

- Scanning the pen trajectory
- Processing time-delayed strokes (e.g., i-dots or t-strokes, which may be inserted only at the end of the word)

Offline:

- Eliminate interference (noise, document quality, contamination) caused by the scanning or already caused by the original
- Skeleton or contour
- Determination of the lower and upper base line (basic line and core height)
- Normalization of size, skew and slant

On-line data has different effects than off-line data. Thus, e.g. The scanning of the pen trajectory is necessary to compensate for different writing speeds which have no influence on the word class. Also the time when stepped marks (t-dashes, i-, u-points, etc.) are made is an important decision-making feature for the process of signature verification. On-line data are often skeleted already.

In preprocessing area, signature segmentation is a complex task too since different signatures produced by the same writer can differ from each other due to local stretching, compression, omission or additional parts [107]. Technique such as segmentation by pen-down and pen-up signals was introduced by G.Dimauro et al.[108]. Furthermore, a technique used in

segmentation called dynamic time warping was introduced by L. Bovino et al. [109]. Both of them belong to online method.
On the other hand, malfunctions which affect the typeface (scan, copy, fax machine) occur only with off-line data. Also a skeleton (or dilution) or contour determination of the thickness of the text (depending on the pen) is a sensible reduction of the writing variables only with off-line data.

4.2.1 Normalization

The normalization procedures concern both on-line and off-line data. In this case the type of application, i.e. the scoring or independent of the scribe, is more decisive. For a writing-dependent system, the writing characteristics such as size and pitch are generally constant and therefore need not necessarily be normalized. In a non-scripted system, the differences between the scribes are too big to avoid standardization. The normalization steps are generally dependent on the position of the baseline lines. The upper and lower base line often cannot be defined so clearly. (see Figure 4.22)

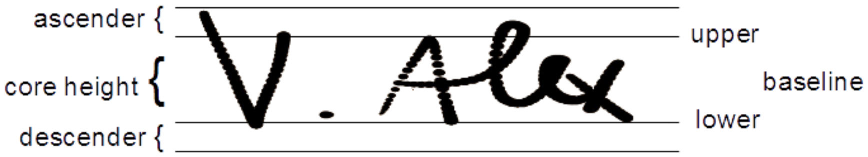


Figure 4.22: Definition upper lower baseline

For example, the tendency to draw differs strongly between different writers, in particular between right-handers and left-handers, and also in italics.

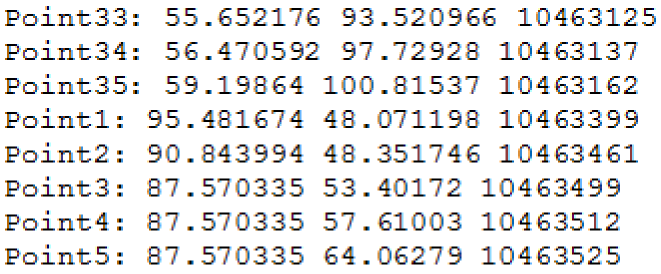
It is important to consider the extent to which standardization results in an improvement in the identification of the characteristics and in the identification / verification, since the characteristics of the individual handwriting of the writer will be negatively affected by excessive normalization.

As the concept of the thesis is based on the use of online handwriting, the segmentation of the data could be restricted to preprocessing. The segmentation was performed on the smartphone client as preprocessing. Since not only online data but also offline data were generated and examined, some special off-line parameters were also generated for testing and evaluation, a special preprocessing had to be carried out for these parameters. A detailed description of the results can be found in Chapter 5 Evaluation.

4.2.2 Segmentation

To generate the online parameters, segmentation is carried out on the mobile client in the preprocessing. The preprocessing begins with parameter sampling like point coordinates, time values and touch down and touch up information for the identification of the segments.
A new segment begins whenever the pen or finger is placed down on the display for writing and ends when the pen or finger is lifted up again.

In Figure 4.23 sample raw data are shown the whole data are ordered in time sequence line by line and indexed based on each single segment. In this sample, Point25 is the last collected point of the former segment and Point1 is the first collected point of the latter segment.



```
Point33: 55.652176 93.520966 10463125
Point34: 56.470592 97.72928 10463137
Point35: 59.19864 100.81537 10463162
Point1: 95.481674 48.071198 10463399
Point2: 90.843994 48.351746 10463461
Point3: 87.570335 53.40172 10463499
Point4: 87.570335 57.61003 10463512
Point5: 87.570335 64.06279 10463525
```

Figure 4.23: signature sample raw data segmentation

4.3 Feature Extraction

In feature extraction area, features classified into more detailed subset, such as function-features and parameter-features [107]. Function features are acquired by using a relevant time function such as position, velocity and acceleration. Meanwhile parameter features are elements, which are characterized as a vector of the signature. Such as total signature time duration, pen-down time ratio, number of pen-downs and pen-ups.

Also in the case of the feature extraction, a distinction is made between on-line and off-line, as in preprocessing. Possible features in the on-line handwriting recognition are the following:

- Change of writing direction, curvature
- Pen printing time, speed and pressure
- Write direction of successive sample points
- Local high above the base line, width and orientation of the signature

For the off-line handwriting recognition, there the following possible selection of features:

- Local altitude above and below the base line
- Determination of curvatures, loops, orientations and vertices
- DCT (discrete cosine transform) or Fourier coefficients and moments
- Transformation or compression of features using neural networks (NN)

These features (on- and off-line) are grouped into feature vectors whose sequence (temporal or spatial vector sequence) is used for the training of the classifiers for identification and verification. Alternatively, the feature vectors may be subjected to additional transformations to reduce the features (e.g., LDA: Linear Discriminant Analysis, formation of difference vectors, combination of adjacent feature vectors) before they are used for classification.

In the revised concept of this Thesis three groups of features were extracted: Geometric features, Static parameters and Temporal features.
The concept for feature extraction, the number of 67 features are extracted for the system and the tests in the Thesis will be described more detailed in the Chapter 4 Concept.

The basis for good identification and verification are good features, in the course of this thesis, a variety of different features have been newly developed and a number of features mentioned in State of the Art have been implemented. The features are again divided into the online features and the offline features. Here, all the features extracted from the online information are now presented as online features and the features extracted from the PNG graphics as offline features. In addition, we distinguish between geometrical, statistical and temporal features. In addition to the temporal features, there are still features, since the current commercially available tablets and smartphones do not react to pressure, the difficulty in this work lies in the lack of pressure features and in achieving the same results as with pressure features. In the data collection, however, the offline data were also recorded and used for comparative purposes for some experiments.

Let us take a look more closely at all features developed for the concept and the experiments for this thesis:

4.3.1 Online Features

A total of 67 online features were extracted, these are presented in more detail here divided into geometrical features, statistical features and temporal features.

a) Geometrical Features

In particular, the feature extraction consists of the following 27 geometrical features:

1. The feature *POINTS* is the number of the pixels occupied for the signature on the display.

2. The feature *SEGMENTS* consider the number of segments of the signature. A new segment begins when the display is touched by pencil, to begin writing process, and ends when pencil is removed from the display.

3. The feature *EUCLID* consider the Euclidean distance between the single points of the segment.

4. The feature *POINTANGLE* determines the angle between terminal points in the beginning of a segment in relation to the lower display border (see Equation 4.4):

$$\alpha = \arctan\left(\frac{|y_2 - y_1|}{|x_2 - x_1|}\right) \tag{4.4}$$

where (x_1,y_1) is first point of the segment and (x_2,y_2) is last point of the segment.

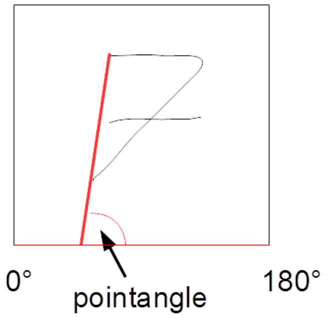


Figure 4.24: pointangle

5. The feature *HYPANGLE* determines the angle between first and last point of the segment in relation to the hypotenuse of the segment (see Equation 4.5):

$$\varphi = \arccos\left(\frac{\vec{a_1}\cdot\vec{a_2}}{\left|\vec{a_1}\right|\cdot\left|\vec{a_2}\right|}\right) \quad \text{with} \quad \vec{a_1} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \text{ and } \vec{a_2} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \tag{4.5}$$

and

$$\left|\vec{a_1}\right| = \sqrt{x_1^2 + y_1^2}$$
$$\left|\vec{a_2}\right| = \sqrt{x_2^2 + y_2^2}$$

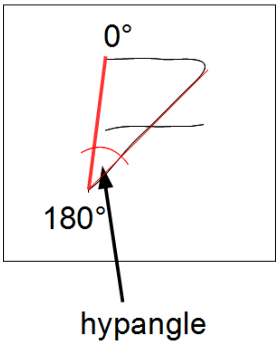


Figure 4.25: hypangle

6. The features *WIDTH* and *HEIGHT* refer to the extreme points of the segment. *SURFACE* means the calculated area. (see Figure 4.26)

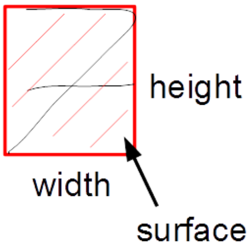


Figure 4.26: width height surface

7. The feature *REGANGLE* determines the angle in relation to the regression straight of the segment,

$$\alpha = \arctan\left(\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right), \tag{4.6}$$

where n is the number of points in the segment and (x_i, y_i) are the coordinates of point i.

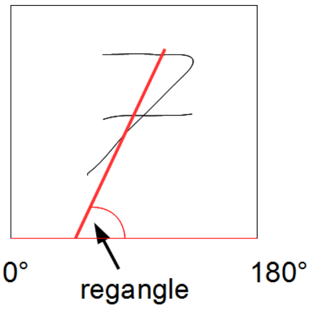


Figure 4.27: regangle

8. The feature *HORIZONTAL POINTANGLE* is calculated from the horizontal corner between the terminal points of the segments of whole signature. To calculate this angle a point displaced eight units to the left from the center of rectangle has to be selected in this implementation. (see Figure 4.28, Equation 4.7)

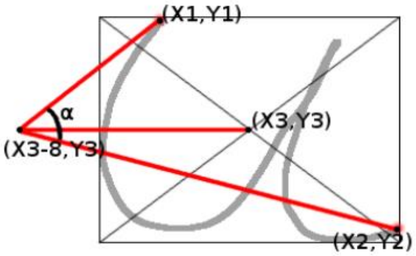


Figure 4.28: Horizontal pointangle

$$a = \arctan\frac{\left|\frac{y_3 - y_1}{x_3 - x_1 - 8}\right| - \left|\frac{y_3 - y_2}{x_3 - x_2 - 8}\right|}{1 + \left|\frac{y_3 - y_1}{x_3 - x_1 - 8}\right| - \left|\frac{y_3 - y_2}{x_3 - x_2 - 8}\right|} \tag{4.7}$$

9. The feature *SPHI* (Segment length distance relation) is the length of the segment (Euclidean distance of all points) divided by the distance between starting point and endpoint of the segment (Equation 4.8):

$$SPHI = \frac{\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{\sqrt{(y_1 - x_1)^2 + (y_n - x_n)^2}} \tag{4.8}$$

10. The number of pen-ups *NUM_STROKES*, actually equivalent to the number of strokes, hence it can be directly calculated according to the index in raw data.

11. The Relative Duration of Writing *RELATIVE_WRITING_DURATION* defined as below:

$$\text{Pen - Down Duration} \big/ \text{Total Duration}$$

It can be calculated according to the timestamps and index in raw data.

12. The Width and Height of a Character Sequence *WORD_WIDTH* and *WORD_HEIGHT* (see Figure 4.29)

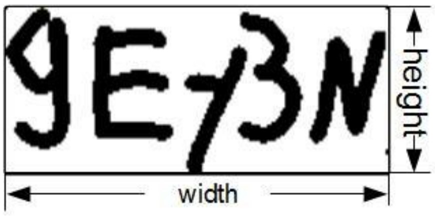


Figure 4.29: Width and Height of character sequence

13. The Horizontal Midpoint *WORD_MID_X* feature, which is always half the value of Width of a Character Sequence. (see Figure 4.29)

14. The Vertical Midpoint *WORD_MID_Y* feature, which is always half the value of Height of a Character Sequence. (see Figure. 4.29)
15. The Average Slant average tilt angle *AvgSLT* among all down-strokes is shown in Equation. (4.9). A definition of down-stroke and why down-strokes are better than general strokes can be found in [144].

$$AvgSLT = \frac{1}{N}\sum_{i=1}^{N} SLT_i \tag{4.9}$$

Every single tilt angle (SLT) is calculated according to the line between the first and the last point in a down-stroke shown in Equation (4.10). Angle is increasing anti-clock wisely. Hence, a vertical down-stroke has a 90-degree tilt angle.

In order to split meaningful down-stroke from an entire stroke, the following three threshold values are used in related algorithm:

$$SLT\begin{cases} \arctan\left(\frac{y_{last}-y_{first}}{x_{first}-x_{last}}\right) & x_{last} < x_{first} \\ 0.5\pi & x_{last} = x_{first} \\ \pi - \arctan\left(\frac{y_{last}-y_{first}}{x_{first}-x_{last}}\right) & x_{last} > x_{first} \end{cases} \tag{4.10}$$

16. The *MAX_ANGLE* (40 degree in this implementation) is defined as the maximum absolute angle allowed in a down-stroke, of which the angle is calculated according to the connection line between each pair of adjacent points based on the horizontal plane. As a result of this constraint, the sub-stroke such as the horizontal part in the bottom of the character "t" (See Figure. 4.30) will be excluded.
17. The *DELTA_ANGLE* (20 degrees in the implementation) is defined as the maximum angle difference between each pair of adjacent connection lines. As a result of this constraint, the sub-stroke, which has some sudden turning point smaller than *MAX_ANGLE* but greater than *DELTA_ANGLE*, will also be excluded. Consequently, the split down-strokes will be more vertical-like.
18. The *RELATIVE_HEIGHT* (one-third the height of the entry word) is defined as the minimum relative height, which should be fulfilled by the sub-stroke.

    As a result of this constraint, some quite vertical, however, very short sub-stokes will be excluded. An example can be seen in Figure 4.30 at the middle part of the character "z".

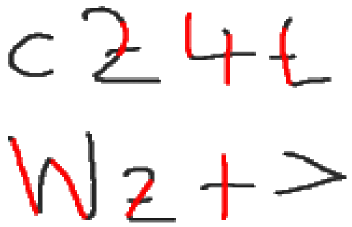


Figure 4.30: Down-strokes are rendered as red in this sample

19. The Amplitude of Slant *SLANT_AMPLITUDE* is difference between the minimum and the maximum tilt angle *AmpSLT* shown in Equation (4.10).

$$\begin{aligned} AmpSLT &= SLT_{\max} - SLT_{\min} \\ SLT_{\max} &= MAX\,(\text{Set of SLT value}) \\ SLT_{\min} &= MIN\,(\text{Set of SLT value}) \end{aligned} \tag{4.10}$$

The next eight features are geometrical moments, which indicate certain characteristics of points distribution among the entire character sequence. At this point, some basic equations of geometrical moments will be given [144].

The geometrical moments of *(p+q)th* order of handwriting samples of *N* sample points is calculated as shown in Equation (4.11).

$$m_{pq} = \sum_{i=1}^{N} x_i^p y_i^q \tag{4.11}$$

The central moments *(μ)* of *(p-q)th* order are calculated as Equation (4.12):

$$\mu_{pq} = \sum_{i=1}^{N} (x_i - \bar{x})^p (y_i - \bar{y})^q \tag{4.12}$$

Inertial moments defined as λ can be calculated from second order moments using Equation (4.12) as shown in Equation (4.13), (4.14):

$$\lambda_1 = \frac{(\mu_{20} + \mu_{02}) + \sqrt{(\mu_{20} - \mu_{02})^2 + 4(\mu_{11})^2}}{2} \tag{4.13}$$

$$\lambda_2 = \frac{(\mu_{20} + \mu_{02}) - \sqrt{(\mu_{20} - \mu_{02})^2 + 4(\mu_{11})^2}}{2} \tag{4.14}$$

With the help of Eq. (4.11) to Eq. (4.14), the next eight features are calculated.

20. The *ORIENTATION* describes the object orientation and is given as θ, which represents the angle between the object and the horizontal axis.

$$\theta = \frac{2}{\pi} \arctan\left(\frac{\mu_{02} - \mu_{20} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4(\mu_{11})^2}}{2\mu_{11}} \right) \tag{4.15}$$

21. The *INERTIAL RATIO* the moment of inertia I is defined as the ratio of the angular momentum of a system to its angular velocity around a principal axis. (Equation 4.16)

$$I = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \tag{4.16}$$

22. The *ASPECT RATIO equals* the value of *A* and, as such, it is a useful feature to represent the aspect ratio of the character it is then normalized by: (Equation 4.17)

$$A = \frac{1}{2}\left(\frac{\mu_{20} - \mu_{02}}{\mu_{20} + \mu_{02}} + 1 \right) \tag{4.17}$$

23. The *SPREADNESS* is the value of µ20 + µ02 and, as such, it reflects the spread of object shape. To normalize the feature, the character size is shown in Equation (4.18) and the normalized feature is calculated as shown in Equation (4.19).

$$size = \sqrt{\left(\left(x_{\max} - x_{\min}\right)\left(y_{\max} - y_{\min}\right)\right)} \tag{4.18}$$

$$S = \frac{2\sqrt{\left(\mu_{20} + \mu_{02}\right)/m_{00}}}{\sqrt{\left(x_{\max} - x_{\min}\right)\left(y_{\max} - y_{\min}\right)}} \tag{4.19}$$

Next three features calculated from 3rd order moments. To obtain normalized features from 3rd order moments, *µ30, µ03, µ21* and *µ12* are split into positive part and negative part.

24. The *HORIZONTAL SKEWNESS*, the *3rd* order moment *µ30* represents the skewness of object in horizontal direction with *µ30 > 0* corresponding to the situation that the object stresses on left case *µ30 < 0* corresponding to that the object stresses on right. Equation (4.20)

$$\theta_h = \frac{1}{2}\left(\frac{\mu_{30}^{+} - \mu_{30}^{-}}{\mu_{30}^{+} + \mu_{30}^{-}} + 1\right) \tag{4.20}$$

25. The VERTICAL SKEWNESS is similar to *µ30*, *µ03* represents the skewness in vertical direction with *µ03 > 0* corresponding to that the object stresses on the upper part and *µ03 < 0* to that the object stresses on the lower part. Equation (4.21)

$$\theta_v = \frac{1}{2}\left(\frac{\mu_{03}^{+} - \mu_{03}^{-}}{\mu_{03}^{+} + \mu_{03}^{-}} + 1\right) \tag{4.21}$$

26. The *BALANCE OF HORIZONTAL EXTENSION*, describes the horizontal extension of points in the characteristic function of *µ21*. Therefore, *µ21* represents the balance of horizontal extension of points. When *µ21 > 0,* the object has larger horizontal extension in lower part than in upper part. Equation (4.22)

$$E_h = \frac{1}{2}\left(\frac{\mu_{21}^{+} - \mu_{21}^{-}}{\mu_{21}^{+} + \mu_{21}^{-}} + 1 \right) \tag{4.22}$$

27. The *BALANCE OF VERTICAL EXTENSION* is similar to *µ21*, *µ12* represents the balance of vertical extension, and *µ12 > 0* shows that the object has larger vertical extension in right than in left part.

$$E_v = \frac{1}{2}\left(\frac{\mu_{12}^{+} - \mu_{12}^{-}}{\mu_{12}^{+} + \mu_{12}^{-}} + 1 \right) \tag{4.23}$$

b) Statistical Features

For the concept 27 different statistical features are extracted like:

28. The *STANDARD_DERIVATION_Y*, Standard derivation in y direction with number of all n points. (see Equation 4.24)

$$sdY = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{4.24}$$

29. The *MEAN_NUMBER_POINTS/SEGMENT* is the average of all points per segment. (see Equation 4.25)

$$NPS = \frac{\sum_{i=1}^{n} x_i}{\sum no\_segments} \tag{4.25}$$

30. The $DTW_{x,y,t}$ (Dynamic Time Warping Distance Feature)

    The Dynamic Time Warping Distance is calculated by two m,n Dimensional vectors $s[0 \ldots m]$, $t[0 \ldots n]$. The DTW Matrix has n rows, m columns and the first column and row is initialized by infinity. For every pair $s(i)$ and $t(i)$ the Euclidean distance is calculated, the result is called cost. The minimum of $DTW_{i-1,\,j}$, $DTW_{i,\,j-1}$ and $DTW_{i-1,\,j-1}$ is add to the cost. This is done for every index in the DTW Matrix. The Result of the DTW is stored at $DTW_{n,m}$. (see Equation 4.26, 4.27)

$$\begin{gathered} DTW_{i,j} = f(s_i, t_j) \\ +minimum(DTW_{i-1,j}, DTW_{i,j-1}, DTW_{i-1,j-1}) \end{gathered} \tag{4.26}$$

$$f(x,y) = |x - y| \tag{4.27}$$

    For each point the x-, y- and time stamp values were collected separately. After that the vectors are split in two vectors. The DTW is calculated for every vector pair. (see Equation 4.28)

$$\begin{gathered} DTW_X = DTW(x_{1,}x_2); \\ DTW_y = DTW(y_{1,}y_2); \\ DTW_t = DTW(t_{1,}t_2); \end{gathered} \tag{4.28}$$

c) Temporal (Speed and Relation) Features

For the Concept 14 different temporal, speed and relation features are extracted like:

31. The *TIME*, the whole time user need to write one secure password sample or text or signature sample.
32. The *TIME_MAX_X*, the time T, when the maximum x point of the signature is reached. (see Equation 4.29)

$$T_{\max\ X} = \max(x_i, x_{i+1}) = \frac{1}{2}x_i + \frac{1}{2}x_{i+1} + \frac{1}{2}\left|x_i - x_{i+1}\right| \tag{4.29}$$

33. The *TIME_V_MAX,* the time when the maximum speed is reached. (see Equation 4.30)

$$\max Vx = \frac{\max(x_{i+1}, x_i)}{\max(tx_{i+1} - tx_i\ )} \tag{4.30}$$

34. The *TIME_X_NEG,* the whole time of negative x movements:

$$txneg = \sum_{i=1}^{n} Tx, x < 0 \tag{4.31}$$

35. The *TIME_X_POS* or *TIME_Y_POS*, the whole time of positive x or y movements:

$$txpos = \sum_{i=1}^{n} Tx, x >= 0\ ,\ typos = \sum_{i=1}^{n} Ty, y >= 0 \tag{4.32}$$

36. The *RELATION NVxz* or *NVyz*, the number of the points horizontally *NSx* or vertically *NSy* in dependence to the whole speed:

$$NVxz = \frac{NSx}{v}\ ,\ NVyz = \frac{NSy}{v} \tag{4.33}$$

37. *RELATION VX MAX*

$$\mathrm{Re}\,lV\ \max = \frac{Tv\max}{v} \tag{4.34}$$

38. *RELATION VX* or *VY POSITIVE*:

$$Vxpos = \frac{txpos}{v},\quad Vypos = \frac{typos}{v} \tag{4.35}$$

39. *RELATION VX* or *VY NEGATIVE*:

$$Vxneg = \frac{txneg}{v},\quad Vyneg = \frac{tyneg}{v}, \tag{4.36}$$

40. *MAX VX* or *VY*:

$$\max Vx = \frac{\max(x_{i+1}, x_i)}{\max(tx_{i+1} - tx_i\ )},\quad \max Vy = \frac{\max(y_{i+1}, y_i)}{\max(ty_{i+1}, ty_i\ )} \tag{4.37}$$

41. *SPEED ALL*, the speed *v* about the whole signature time *t*:

$$v = \frac{\sqrt{\sum (x_i - y_i)^2}}{\sum t_i} \tag{4.38}$$

42. *RELATION POINTS SPEED*, the relation between number of all points *n* and the speed *v*:

$$NV = \frac{\sum_{i=1}^{n} x_i}{v} \tag{4.39}$$

All these extracted features are used for the proposed concept and the experiments detailed introduced see Chapter 5 Evaluation.

4.3.2 Offline Features

Beside the online features generated from online point information and time information, offline features generated from PNG graphics of the signatures are implemented and tested. This paragraph describes the features generated from offline information.

1. **Zernike moments:**

Zernike moments were suggested by Teague [190] as a form of orthogonal moments as early as 1980 to solve both the problem of redundancy (cf. Geometrical Moments) and the problem of rotational invariance. The Zernike moments are complex-valued moments and use the Zernike polynomials (*Vpq*) as radial polynomials:

$$Z_{pq} = \frac{p+q}{\pi} \int_0^{2\pi} \int_0^1 [V_{pq}(r,\theta)] * f(r\ cos\theta, r\ sin\theta) r\ dr\ d\theta \tag{4.40}$$

Where $p = 0, \ldots, \infty, \mid q \mid \leq p$ and $p\text{-} \mid q \mid$ just. The Zernike polynomials are the only polynomials in x and y which define an orthogonal basis for the set of complex-valued functions over the unit circle $x2 + y2 \leq 1$, e.g.

$$\langle V_{nm}, V_{qp} \rangle = \iint_{x^2+y^2 \leq 1}^{1} [V_{nm}(x,y)] * V_{pq}(x,y) dx\ dy = \delta_{np}\ \delta_{np} \tag{4.41}$$

The function must be a polynomial in *x* and *y* in order to achieve rotational and scaling invariance. (Translation is achieved by replacing the regular moments with central moments) In general, however, the Zernike moments are defined by polar coordinates. The basic functions of order *p* and repetition *q* are given by:

$$R_{pq} = \sum_{s=0}^{(p-|q|)/2} (-1)^s \frac{(p-a)!}{s!\left(\frac{p+|q|}{2}-s\right)!\left(\frac{p+|q|}{2}-s\right)!} r^{p-2s} \qquad (4.42)$$

And the radial polynomials *Rpq (r)* are defined as follows:

$$Z_{pq}^{f} = \frac{\langle f, V_{pq}\rangle}{V_{pq}, V_{pq}} \qquad (4.43)$$

We obtain a linearly independent polynomials of *degree* ≤ *n* on the basis of the constraint *p*- | *q* | in total $\frac{1}{2}$ *(n + 1) (n + 2)* [190]. The Zernike moment of an image function f with order p and repetition *q* is the length of the orthogonal projection of *f* on the Zernike basis function *Vpq* with:

$$\langle f, g\rangle = \iint_{r\leq 1}^{1} f(x,y) g * (x,y) dx\, dy \qquad (4.44)$$

Zernike moments have already been used extensively for character recognition of solid, binary symbols. However, they can also be used for gray-scale images. It is thus possible to extract both rotation-invariable and rotation-variant features, in the latter case the imaginary part is ignored. Zernike moments are projections of the input image onto the space, which is spanned by the orthogonal vectors *V* (as defined above). The Zernike Moments are calculated until the 15th order for the experiments in Chapter 4. For a discrete digital image, the Zernike moments of order p and repetition q can be calculated as follows:

2. **Fourier coefficients:**

A periodic function can be represented by a Fourier series as a series of sine and cosine functions whose frequencies are integral multiples of the fundamental frequency ω = 2π / T.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cdot \cos(n\omega t) + b_n \cdot \sin(n\omega t)) \qquad (4.45)$$

The Fourier coefficients *an* and *bn* can be calculated using the following Euler's formulas:

$$a_n = \frac{2}{T}\int_{c}^{c+T} f(t)\cos(n\omega t)\,dt \quad b_n = \frac{2}{T}\int_{c}^{c+T} f(t)\cos(n\omega t)\,dt \quad n = 1,2,\ldots \tag{4.46}$$

C represents a shift of the interval and can be arbitrarily selected. In this method, a maximum number of Fourier coefficients can be calculated, that, *f (t)* is approximated by a finite trigonometric polynomial *fn (t).*

$$f_n(t) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cdot \cos(k\omega t) + b_k \cdot \sin(k\omega t)) \tag{4.47}$$

From the read-in coordinates of the characters, the Fourier Coefficients are computed and then compared with one another. The coefficients can be used to transform the curves of the characters. In Figure 4.29, *C* is represented as an original and as an approximation by the number of coefficients. The Fourier Coefficients are calculated until the 100th order in the experiments chapter 5.

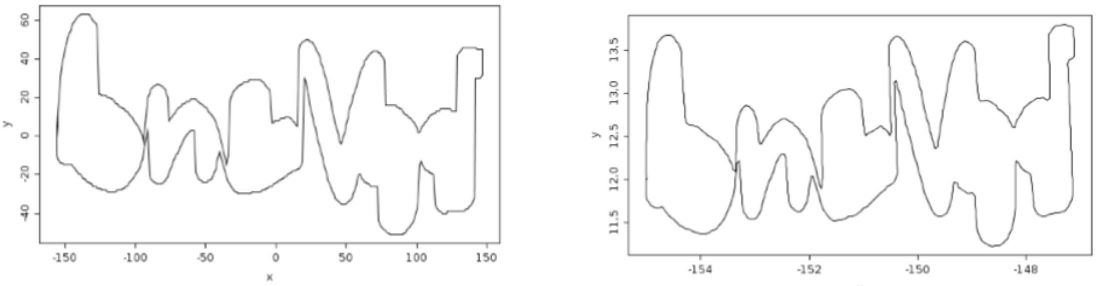


Figure 4.29.: Sample original (left) and representation Fourier Transformation (right)

4.4 Feature Reduction

One possibility to improve the performance of the identification and verification system is the reduction of the number of features. It is emphasized that we should look characteristics of a handwriting sample in different ways and at different directions. However, it does not mean that more features being extracted from a handwriting sample, more advantages we can take from them. On the contrary, overfit problem will occur if there are too many features being analyzed simultaneously. That’s the reason why the evaluation and selection procedure of features is necessary and important. There have already bunch of different evaluation methods been studied and researched. For this thesis for example, an evaluation method based on fisher score criteria were implemented and will be introduced.

In summary, it can be said that the parameter reduction contributes [112]:

- simplification of models to make them easier to interpret by researchers/users
- shorter training times
- to avoid the curse of dimensionality
- enhanced generalization by reducing overfitting (formally, reduction of variance)

There are several methods of feature reduction (see Figure 4.31) with some advantages and disadvantages for feature selection and feature extraction methods.

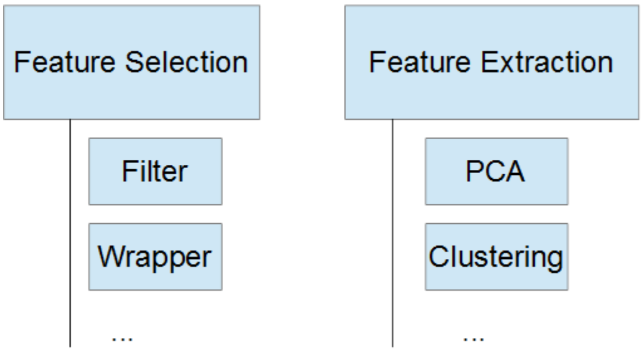


Figure 4.31: Methods of Feature Reduction

For example:

- Filters rank each feature and select the highest ranked features. They are efficient and fast to compute.

- Wrappers search for the highest subset of features with heuristic methods like forward selection and backward elimination. The disadvantage of Wrappers is the interaction with the classifier, and they are computationally intensive.

- PCA is based on a similarity measure like Euclidean distance or Mahalanobis distance. The advantages are: a lack of redundancy of given data, the orthogonal components and a reduced complexity in images grouping with the use of PCA [114]. The disadvantages are: The covariance matrix is difficult to be evaluated in an accurate manner [113]. Even the slightest invariance could not be captured by the PCA unless the training data explicitly provide this information [115].

- Clustering is a process of dividing a set of data (or objects) into a set of meaningful sub-classes, called clusters. There are two main methods for clustering K-Means and Hierarchical clustering. The advantages are: K-means is computationally fast, it is a simple and understandable unsupervised learning algorithm. The disadvantages are: it is difficult to identify the initial clusters. It is difficult to predict the value of K, because the number of clusters is fixed at the beginning. The final cluster pattern is dependent on the initial patterns [116].

In addition to PCA for the investigation of feature reduction, the filters Fisher Score and Info Gain Attributes Evaluation were used in this thesis, as these two filters facilitate very fast results on the one hand and convince with their good results on the other. More specifically, details about feature reduction methods used and results in this thesis will be found in Chapter 5 Evaluation.

Let’s take a closer look at the two filter methods Fisher Score and Info Gain Attribute Evaluation used for this thesis:

a) Fisher Score

The Generalized Fisher Score [117] is a joint feature selection criterion, which aims at finding a subset of features, which maximize the lower bound of traditional Fisher Score. It also resolves redundant problems in feature selection process.

The mathematical description of Fisher Score is shown as below in Equation (4.40).

$$F\left(X^{j}\right)=\frac{\sum_{i=1}^{N} n_i \cdot \left(\mu_i^j-\mu^j\right)^2}{\left(\sigma^j\right)^2} \tag{4.40}$$

$$\sigma^j=\sum_{i=1}^{N} n_i(\sigma_i^j)^2$$

In which,
j - the j-th feature.
i - the i-th class, which couble be interpreted as i-th subject in our test.
n - the size of the instances for certain class
μ - the mean value for certain class
σ - the standard deviation for certain class

b) Info Gain Attribute Evaluation

At second Information Gain Attribute Evaluation (IG) [151] is used for ranking. This ranker evaluates the worth of an attribute by measuring the information gain with respect to the class. The mathematical description of Information Gain Attribute Evaluation is shown as below in Equation (4.41).

$$\begin{aligned}
&IG = H(Y) - H(Y/X) = H(X) - H(X/Y)\\
&H(Y) = -\sum_{y\in Y} p(y)\log_2(p(y))\\
&H(Y/X) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y/x)\log_2(p(y/x))
\end{aligned} \tag{4.41}$$

In which,
H(Y) is the entropy of Y.
H(Y/X) is the entropy of Y after observing X.
p(y) is the marginal probability density function for the random variable Y.
p(y/x) is the conditional probability of y given x.

4.5 Classification

The classification is based on the assumption that the model which has been trained with data from writer *A* produces a higher score than any other models when confronted with the signature of writer *A*. In case of writer identification, the handwriting is either assigned to one of the n writers with the highest confidence score or it is rejected if the confidence measure is below a given threshold. In case of writer verification, the handwriting is accepted if the confidence measure is above a given threshold, otherwise the the handwriting is rejected.

In classification area, a comparison technique using Euclidean distance was introduced by R. S. A. Araujo et al. [110]. Meanwhile a Bayesian approach was introduced by D. Muramatsu Et al. [111]. Furthermore, there are still bunch of different comparison techniques such as correlation, Support Vector Machine (SVM), Neural Network (NN).

Hidden Markov Models (HMMs) [120], are often used for character recognition, speech recognition and offline handwriting recognition. A major advantage of classification over HMMs is that the language or writer sequence to be recognized need not be explicitly segmented.

The classification for online identification and verification is often done with NN and a variety of other classifiers, in the literature none of the classifiers stands out as outstanding.

The classification is carried out on the server with the algorithms of the Weka datamining framework.

The Weka workbench (WEKA [188]) is an organized collection of state-of-the-art machine learning algorithms and data preprocessing tools. The basic way of interacting with these methods is by invoking them from the command line. However, convenient interactive graphical user interfaces are provided for data exploration, for setting up large-scale experiments on distributed computing platforms, and for designing configurations for streamed data processing. These interfaces constitute an advanced environment for experimental data mining. The system is written in Java and distributed under the terms of the GNU General Public License [189].

Because we tested several algorithms in Weka and reached good result, in addition to the functions for parameter reduction, which are partly implemented themselves and partly also from the Weka framework, the following classifiers are used for the system and the experiments:

- Bayes Net
- Naïve Bayes
- k-Nearest Neighbour
- K-Star

Implementation in the REST Webservice on the server allows all classifiers to be implemented in Weka, and further new classifiers can be implemented. For the tests in Chapter 5, the above-mentioned classifiers were used, as well. Implementation in the REST Web service on the server allows all classifiers implemented in Weka and further new classifiers can be implemented. For the tests in Chapter 5, the above-mentioned classifiers were also used. The results and the discussion will show why these classifiers were restricted.

At this point we introduce and clarify some basic concepts about the Bayes Network, the Naive Bayes Classifier and the Nearest Neighbour Classifier. In the following two chapters, with realizing some of these basic concepts and understanding the related mathematical descriptions, who will help us to analyze and interpret the results, which are collected using the classification techniques mentioned above.

a) A Brief Introduction into the Bayes Network (BN)

Bayes network plays an important part in the area of data mining and machine learning. It reduces the storage space compared to the complete joint probability distribution based on the same number of variables using the relationship called conditionally independence between different variables. Two variables *A* and *B* are called conditionally independent, given variable *C* if:

$$P(A, B|C) = P(A|C) \cdot P(B|C) \qquad (4.42)$$

In order to using Bayes Network, first thing is to build a network structure based on the problem which is ready to be solved. The structure of a Bayes Network is mathematically called a directed acyclic graph (DAG). Each node in graph represents a variable and related variable will be connected by an edge. After the structure being built, for each variable, a conditional probability table (CPT) could be calculated and estimated based on training data.

As we should know that the traditional probability could only be calculated based on finite discretized values, such as binary values 'true' and 'false'. However, in this thesis, the biometric features are mostly more like real number values. As a result, before create CPTs based on training data, a procedure called discretization is necessary. Consequently, the real number values will be discretized and categorized into different intervals. So that each interval could be treat as a whole for the calculation of probabilities.

b) Naïve Bayes Classifier (NB)

The Bayes network plays an important role in the area of data mining and machine learning. It reduces the storage space compared to the distribution of the complete joint probability based on the same number of variables using the relationship called conditionally independence between different variables. Two variables *A* and *B* are called conditionally independent, given variable *C* if:

$$P(A,B|C) = P(A|C) \cdot P(B|C) \tag{4.42}$$

In order to use Bayes Network, the first thing is to build a network structure based on the problem which is ready to be solved. The structure of a Bayes Network is mathematically called a directed acyclic graph (DAG). Each node in graph represents a variable and the related variable will be connected by an edge. After the structure is being built, for each variable, a conditional probability table (CPT) could be calculated and estimated based on training data. As we should know, the traditional probability could only be calculated based on finite discretized values, such as the binary values 'true' and 'false'. However, mostly in this thesis, the biometric features are more like real number values x. As a result, before creating CPTs based on training data, a procedure called discretization is necessary. AS a consequence, the real number values will be distinguished and categorized into different intervals, so that each interval could be treated as a whole for the calculation of probabilities.

a) Naïve Bayes Classifier (NB)

The Naïve Bayes classifier is similar to the Bayes Network, which, is, however, furtherly simplified using conditionally independence in Equation (4.42). Firstly, let us have a look at the Bayes Formula,

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{4.43}$$

Apply Equation (4.43) to biometric features, assuming there are three features F1, F2 and F3, altogether. So that the probability of a handwriting sample belonging to Person *A* is,

$$P(A|F_1, F_{2,}F_3) = \frac{P(F_1, F_{2,}F_3|A) \cdot P(A)}{P(F_1, F_{2,}F_3)} \tag{4.44}$$

Assuming that the features F1, F2 and F3 are conditionally independent, based on Equation (4.42), Equation (4.44) can be further simplified into Equation (4.45),

$$P(A|F_1, F_{2,}F_3) = \frac{P(F_1|A) \cdot P(F_2|A) \cdot P(F_3|A) \cdot P(A)}{P(F_1, F_{2,}F_3)} \tag{4.45}$$

As the denominator in Equation (4.44) is constant, assume X is the set of all writers, a naive Bayes Formula can be defined as:

$$Writer_{Naive-Bayes} = argmax\ P(Writer = x \in X) \prod_{i=1}^{n} P(F_i\ | Writer = x \in X). \qquad (4.46)$$

b) k Nearest Neighbour Classifier (*k-NN*)

The k Nearest Neighbour Classifier is a linear classifier based on the distance from the test data to other training data, a data value can be represented by a point in an n-dimensional space (considered that we have n features). Distance can be calculated as a normal Euclidian distance or another distance such as hamming distance. For a One-Nearest-Neighbour classifier, these test data will be classified as the class, that the nearest data point belongs to.

In pattern recognition, the *k-NN* algorithm is a non-parametric method used for classification and regression [121]. In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression [122]:

- In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If *k* =1, then the object is simply assigned to the class of that single nearest neighbor.
- In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms [121].
Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/*d*, where *d* is the distance to the neighbour [121].

The neighbors are taken from a set of objects for which the class (for *k*-NN classification) or the object property value (for *k*-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required [121].

Chapter 5 Evaluation of Experiments and Results

This chapter is divided into two parts. In the first part two different own datasets with handwritten passwords are experimentally evaluated. In the second part own and public datasets with handwritten secure passwords, signatures and text samples are experimentally evaluated. The robustness of the datasets for the identification and verification of writers, the developed features and the classifiers is evaluated in various experiments. At the end of the chapter will be shown a summary of the best results with the used datasets, features and classifiers.

5.1 Evaluation Experiments with Handwritten Passwords

For the experiments with handwritten passwords, BD-150 which has been introduced in chapter I paragraph 1.1.2 section a) and password DB-9 which has been introduced in the same chapter paragraph 1.1.2 section b), were used. From the BD-150 only a part of the datasets (100 users) were selected for the experiments because many datasets are incomplete.

5.1.1 Experiments with Handwritten Password DB-9 and 10 Features

5.1.1.1 Experimental Setup

For the first experiment the Handwritten Password DB-9 is used with 144 samples (108 genuine and 36 impostors) of handwritten passwords by nine users. For the classification the dataset is split into a training set with 72 genuine samples (e.g. deposited passwords as security) and two test sets, one with 36 handwritten genuine samples and one with 36 impostor samples (e.g. for login) [123].

5.1.1.2 Feature Extraction / Classification

There are two modes for data collection, feature extraction and classification like described in Chapter 3:

In the Enrol Mode and Test Mode, three features (segments, time, points) on the mobile device, and at server, the feature extraction algorithm extracts seven geometric features (pointangle, hypangle, regangle, wide, height, surface, euclid). Besides, it builds and stores the model file from the classification in the model database.

For classifying, the three classifiers Naïve Bayes, k-Nearest Neighbour and Bayes Net are used.

5.1.1.3 Results

The achieved accuracy rate ranges from 90.62% to 96.87% for percentage split train/test with 66%. The best identification results, up to 96.87%, are reached by Naive Bayes Classifier. The FAR when using Naive Bayes Classifier reached 11.11% (see Table 5.1). This first experiment is working in an acceptable time range of 0.5 to 1.5 seconds depending on the WIFI connection. To get the real computational time, there are carried out some time tests on server and mobile client. The computational time for the whole process is subject to the constraints of internet connection, being from 0.5 to 1.5 seconds for testing (see Table 5.2).

TABLE 5.1: CLASSIFICATION AND IDENTIFICATION

Classifier	FAR (in %)	Success Rate (in %)
k-Nearest Neighbour	11.11	90.62
Bayesian Network	22.22	93.75
Naïve Bayes	11.11	96.87

TABLE 5.2: TEST OF COMPUTATIONAL TIME

Test	Computational Time (in ms)
Feature extraction on server	58
Build model with WEKA (Naïve Bayes)	20
Verify user and show result on mobile client	from 500 to 1500

The proposed solution is based on a touch screen phone and a server. It shows that our approach can reach a high security for the biometric user verification for an access control system. The system reaches a verification rate of 96.87% for distinguishing among different writers using a combination of online and offline identification and the features *segments, time, points, pointangle, hypangle, regangle, wide, height, surface,* and *euclid*.

Normally, a user who knows the password of another user always gains illegal access to a system. With this combination of a handwritten password on a mobile device and server-side classification algorithms the illegal access is reduced to FAR of 11.11%.

5.1.2 Experiments with DB-9 and BD-150 and 10 Features

5.1.2.1 Experimental Setup

In this experiment the two data sets DB-9 (a) and BD-150 (b) were used as described in paragraph 4. The records were performed once without transformation of the data and with transformation of the data. The cluster analysis and the principal component analysis (PCA) were used to transform the data [124]. For the classification, the classifiers Naïve Bayes, KNN and K Star were used. For features extraction the same 10 predominantly geometric features as used in paragraph 5.1.1 were used.

Both data sets are analyzed separately with the results of the classification with and without transformation and the second dataset is specifically tested related to the indication of possible manipulations and thus only a part of the data can be used for further experiments.

5.1.2.2 Feature Selection / Transformation / Classification

With the aid of cluster analysis, the 10 features extracted from the raw data were analyzed for their "spatial proximity". If one proceeds from the Euclidean metric in an N-dimensional space, the 10 features are first considered as 10 independent clusters. N denotes the number of collected signatures. If we want to imagine these clusters as circles in which the radii begin to grow uniformly, overlapping circle groups are gradually formed, which are now regarded as new intermediate clusters. After the analysis of both data sets, one could proceed from four equally spaced clusters after a certain step. The single-linkage method is used for hierarchical cluster analysis.

With the help of principal component analysis (PCA), four main components were also determined from 10 features. This meant that the information contained in 10-dimensional vectors was reduced to a four-dimensional space. The adapted model was then examined for its accuracy. In order to avoid the usual criticism of PCA, each of the 10 features are subjected to a simple transformation (see Equation 5.1) which corresponds to the usual standardization of the random variables in stochastics and is called a z-transformation Has made "dimensionless".

$$z(i) = \frac{x(i) - \bar{x}(i)}{s(i)}, \quad i = 1,...,10 \qquad (5.1)$$

On the basis of the investigation of the correlation matrix for 10 features, it became clear that the variable with the number 2 ("Number of pixels occupied by the signature on the touch screen") and the variable with the number 10 Correlate strongly (the correlation coefficient according to Pearson is over 0.97), see Table 5.3. After transformation (1) an arithmetic mean value was formed from these variables. The resulting new variable (which in the next paragraph is schematically designated by (2 + 10) / 2) thus describes the "workload of the user, which includes his signature".

TABLE 5.3: CORRELATION MATRIX FOR 10 FEATURES: (A) 1ST DATA SET, (B) 2ND DATA SET

1,0	-0,35	-0,36	-0,2	-0,6	-0,61	-0,56	0,33	0,16	-0,28
	1,0	-0,01	-0,11	0,19	0,17	0,19	-0,23	-0,03	0,97
		1,0	-0,22	-0,08	0,1	-0,14	0,09	0,33	0,01
			1,0	0,63	0,38	0,6	-0,38	-0,22	-0,13
				1,0	0,75	0,93	-0,69	-0,08	0,08
					1,0	0,9	-0,54	0,08	-0,01
						1,0	-0,7	-0,03	0,04
							1,0	-0,32	-0,16
								1,0	-0,03
									1,0

(A)

1,0	-0,34	0,15	0,06	-0,34	-0,25	-0,32	0,28	-0,10	-0,34
	1,0	-0,18	-0,06	0,66	0,30	0,62	-0,40	0,2	0,99
		1,0	-0,08	-0,35	0,17	-0,21	0,37	-0,22	-0,18
			1,0	0,14	0,25	0,17	-0,22	0,27	-0,06
				1,0	0,47	0,94	-0,64	0,31	0,66
					1,0	0,62	-0,38	0,12	0,3
						1,0	-0,54	0,28	0,62
							1,0	-0,44	-0,4
								1,0	0,2
									1,0

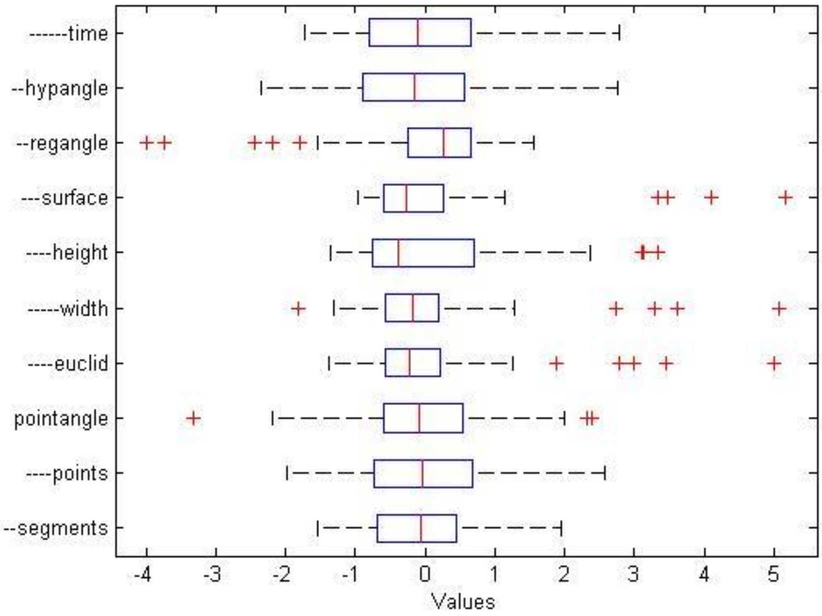
(B)

According to an analogous principle, the following four features [(2 + 10) / 2, 3, 6, 9] were selected for a partial analysis which correlated weakest linearly with one another. The word "sub-analysis" refers to the statistical analysis of these 4 variables instead of the original 10. The variable (2 + 10) / 2 is the above-mentioned "workload of the user that includes his signature". The features with the numbers 3 and 9 are special angles and the variable 6 was the maximum height of the signature.

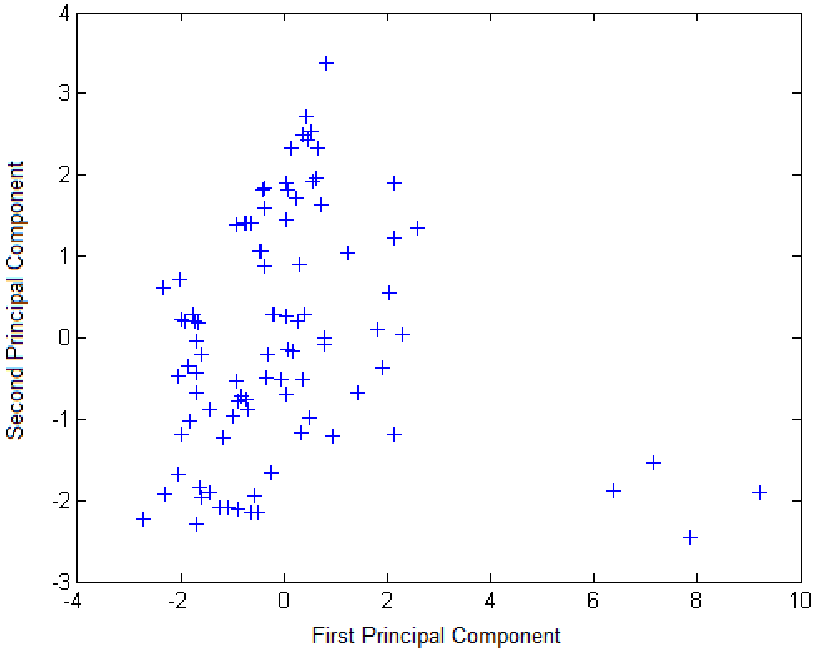
5.1.2.3 Results

1. First dataset DB-9

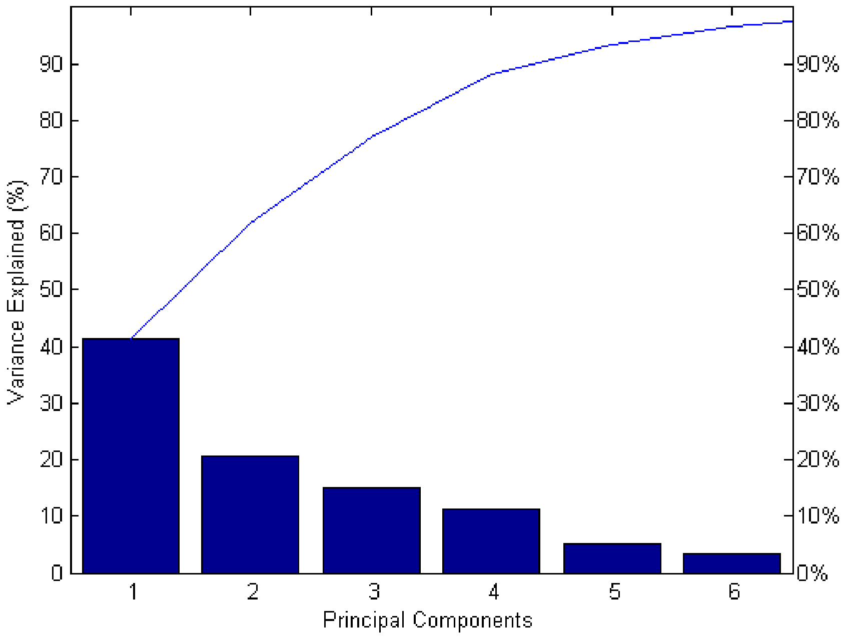
Figure 1 shows the box plot, the results of the PCA for the data transformed according to (1). The 1 th and 2 th (latent) main components can be interpreted with a little imagination as "work" and "geometry" of the signature.



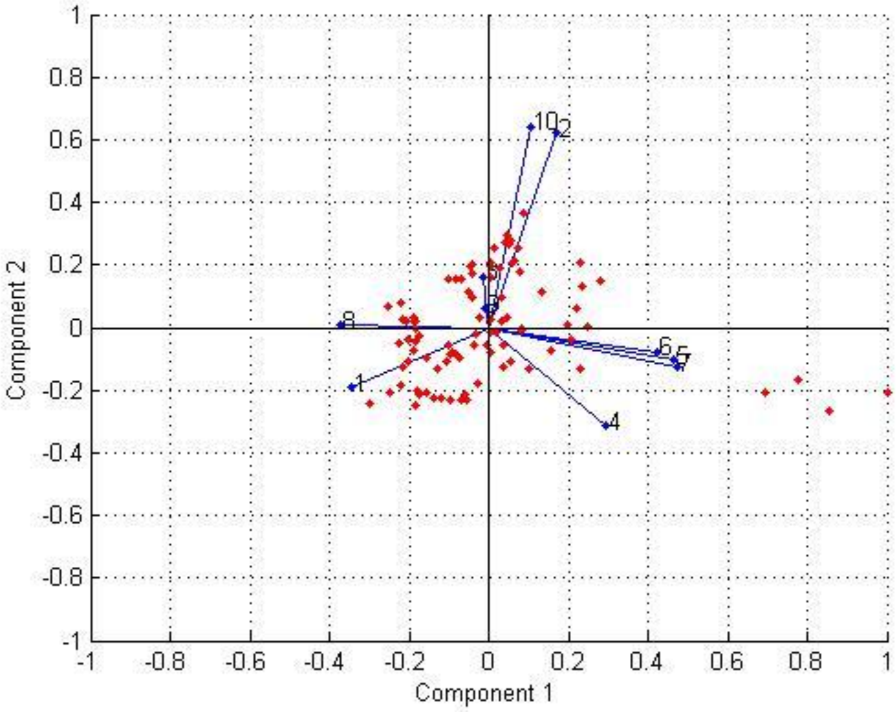
(A)



(B)


Variance Explained (%)
0
10
20
30
40
50
60
70
80
90
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
1
2
3
4
5
6
Principal Components

(C)


Component 2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Component 1
10
2
3
9
8
1
6
5
7
4

(D)

Figure 5.1: (A) box plot; (B) 1-th vs 2-th main component; (C) The variances of the corresponding main components in % to the total variance; (D) Biplot for two main components. Here the complete variable set of 10 variables was used for the first data set.

Figure 5.2 demonstrates the results of the cluster analysis for the first data set, which corresponds well to the results of the PCA in Figure 5.1 (D) (spatial proximity).

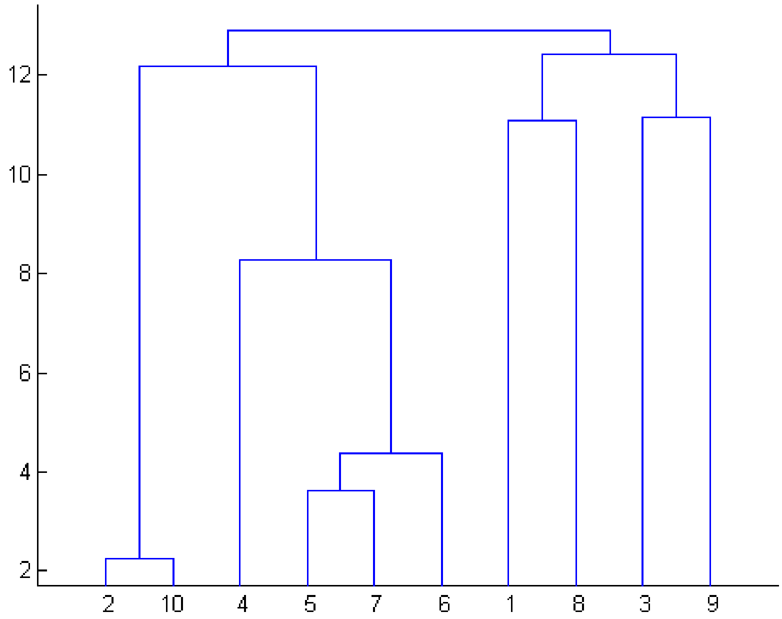
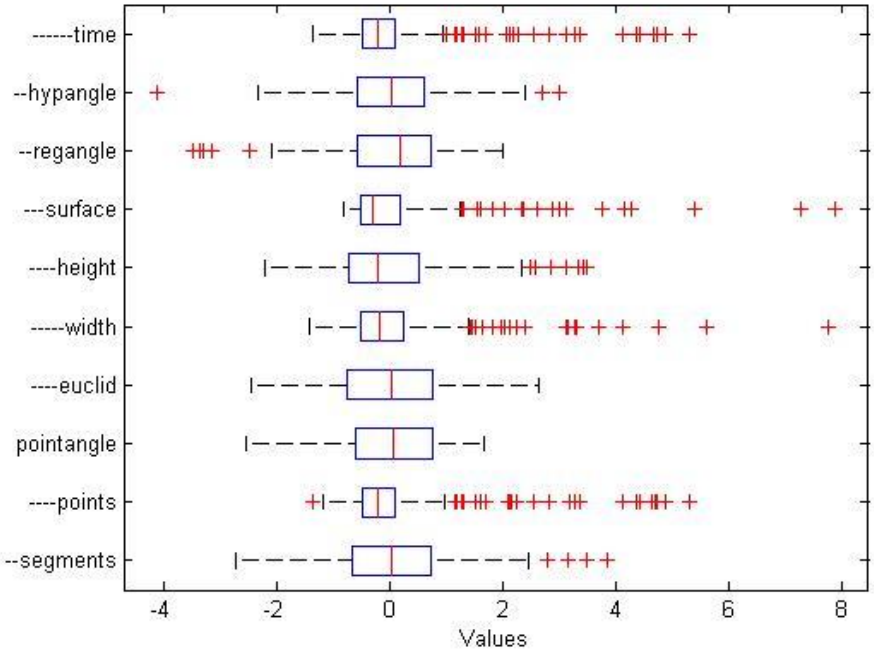


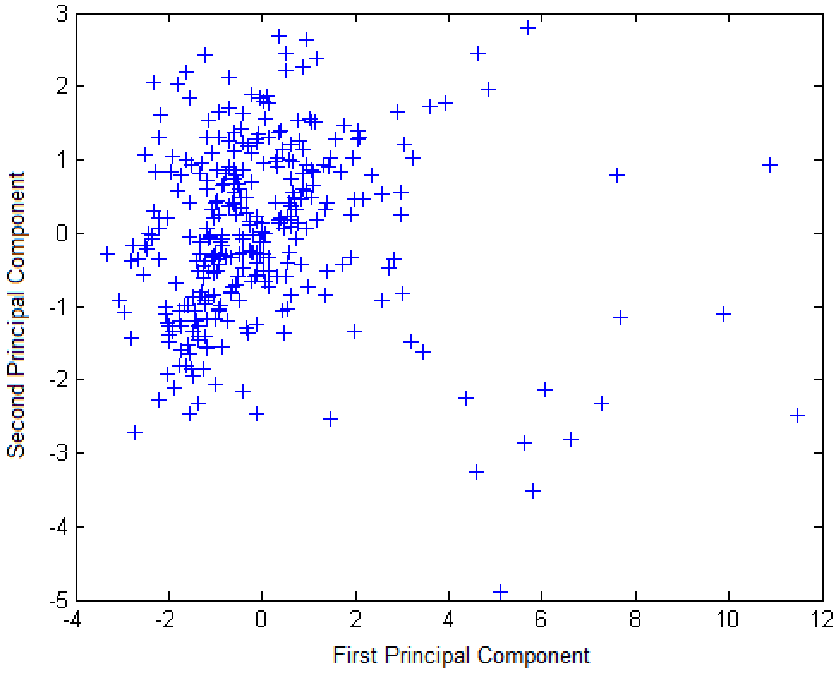
Figure 5.2: Results of the cluster analysis for the first data set after transformation.

2. Second Dataset BD-150

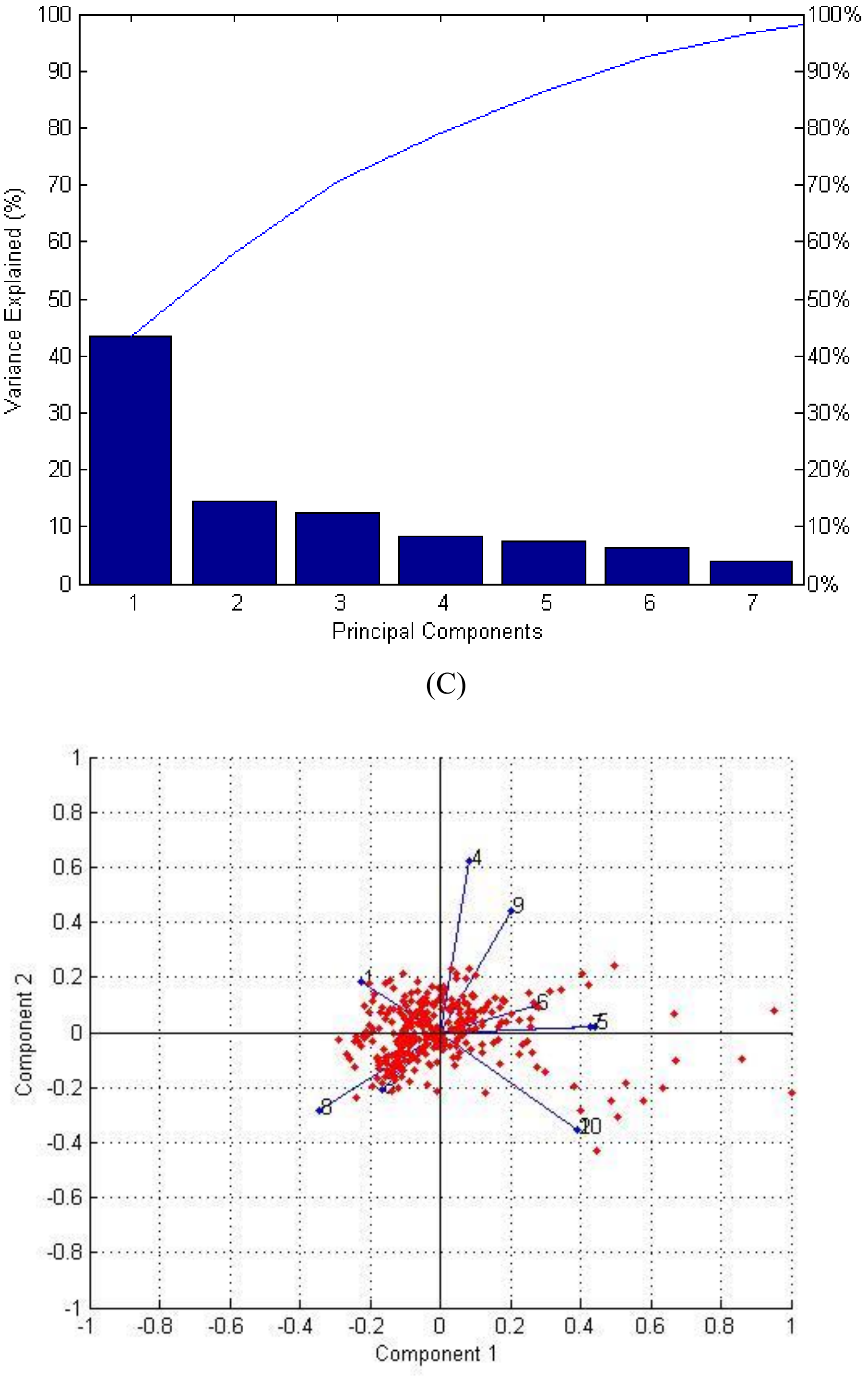
Analogous to the first data set, the transformation of the raw data was carried out first. The results of the subsequent principle component analysis are shown in Figure 5.3.


------time
--hypangle
--regangle
---surface
----height
-----width
----euclid
pointangle
----points
--segments
-4
-2
0
2
4
6
8
Values

(A)


3
2
1
0
-1
-2
-3
-4
-5
-4
-2
0
2
4
6
8
10
12
First Principal Component
Second Principal Component

(B)

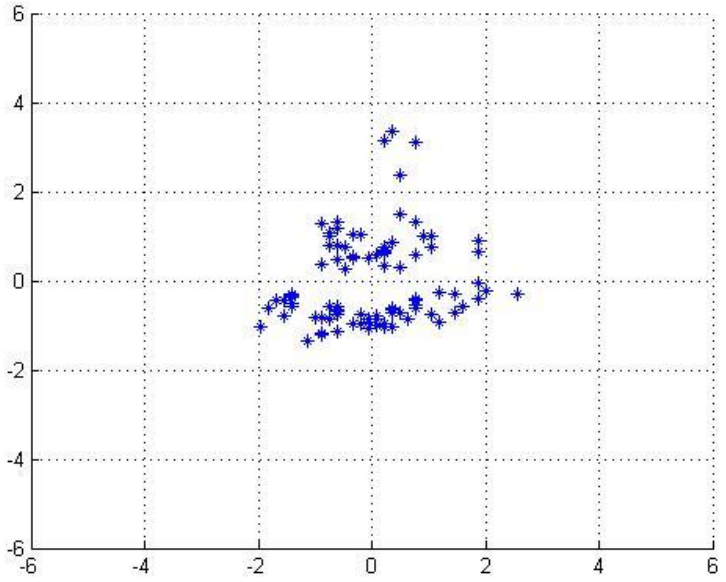


(C)

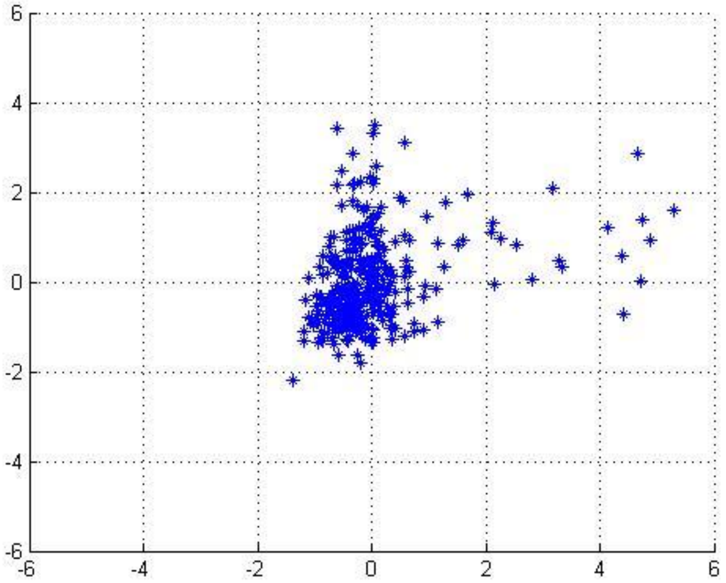
(D)

Figure 5.3: (A) box plot; (B) 1-th vs 2-th main component; (C) The variances of the corresponding main components in % to the total variance; (D) Biplot for two main components. Here the complete variable set of 10 variables was used for the second data set.

Figure 5.4 shows the point cloud for two variables (number of points vs. height of the signature) after the transformation (1) for the 1st and 2nd data record. 4 (A) and 4 (B), it can be seen that the two distribution patterns of the points can not be regarded as similar. 5 (B), on the one hand, a significant accumulation of the points and, on the other hand, some clear outliers in the right half, suggesting that most raw data were most likely manipulated by a small group of persons who simulated the different signatures. This pattern occurs in many variable pairs.



(A)



(B)

Figure 5.4: The point clouds after transformation (1): (A) for the first data set and (B) for the second data set. The following variables were used: number of points (y-axis) vs height of the signature (x-axis).

Table 5.4 summarizes the most important characteristic values for the experiment for both datasets. The reduction to 4 variables improved the accuracy of the projection in the main component analysis, but the detection rate deteriorated. This was primarily due to the loss of information. In addition, the outliers in Figures 2 (A) and 4 (A) had an effect on the results of the main component analysis. For the second data set, there was a marked improvement in the recognition rate after transformation (highlighted in Table 5.4).

Table 5.4: Comparison of the results: (a) recognition rate without and with transformation (1) for both data sets; (b) Influence of the variable reduction on the accuracy of the projection and on the recognition rate for the first data set with transformation

	Success rate of correct classification without transformation (1)			**Success rate of correct classification without transformation (1)**		
Classifier	**Naïve-Bayes**	**KNN**	**KStar (K*)**	**Naïve-Bayes**	**KNN**	**KStar (K*)**
1. Dataset N = 96	96,87%	90,62%	90,63%	97,85%	98,2%	100%
2. Dataset N = 326	19,82%	47,75%	**4,5%**	20,72%	47,75%	**43,24%**

(A)

1. Dataset according to The transformation (1), Number of variables	**Accuracy of the Projection, MQA**	**Success rate with:**		
		Naïve-Bayes	**KNN**	**KStar (K*)**
N=10	3,15	97,85%	98,2%	100%
N=4 [(2+10)/2, 3, 6, 9]	1,99	94,62%	92,47%	81,25%

(B)

The best results of the combination of online and offline identification with rounded 99% were achieved with the KNN classification as well as with eight passwords. The described transformation (1) of the variables leads, among other things, to a marked increase in the accuracy of the projection in the main component analysis. In addition, the graphical

presentation of the point cloud of pairs of variables is sharpened so far that the possible manipulations during the production of raw data (2nd data record) can be easily detected.
The results of the cluster analysis were consistent with the results of the PCA. However, the use of both methods is also somewhat cautious to enjoy: the variables used have outliers and are not stochastically independent. The use of cross-validation methods for variable reduction is not recommended because the extracted ten variables are heterogeneous and can not be subdivided into "content-like" groups. The biometric information contained in the various signatures is also heterogeneous. As seen in the second data set, the manipulation of most of the data material by a small group of users degrades the rate of detection. This assumption could be confirmed even after a visual comparison with the original biometric data. The KStar (K *) - classifier reacted very sensitively to this manipulation: the recognition rate nearly increased tenfold after the transformation (1) for the second data set.

5.2 Evaluation Experiments with Handwritten Secure Passwords, Signatures and Texts

In the previous chapter the results of writer identification of handwritten passwords with two different private datasets were presented. Now the robustness of the extracted features and classification methods should be proved with handwritten secure passwords, signatures and text samples from private and public datasets.

In this chapter different experiments with:

- Private databases: Secure Password DB-32, Secure Password DB-150 with sessions and Secure Password and Signature DB-30,
- Public databases: ATV-SLT-DB, IAM online handwriting DB and SVC 2004 DB

are evaluated. Detailed Information about the databases are given in Chapter 2 Paragraphs 1.1 a-c and 1.2 c-e.

5.2.1 Experiments with Secure Password DB-32 and 25 Features

5.2.1.1 Experimental Setup

In this experiment are 25 static and dynamic biometric features (see Paragraph 5.2.1.2) from a handwritten character password sequence on an android touchscreen device selected. For the writer verification the classification algorithms of WEKA framework used [151].
As database the Secure Password DB-32 as described in chapter 2 with safe passwords with a length of 8 characters written by each person 12 times is used. For the last test 96 password samples written three times every safe password character sequence by different subjects with knowledge of the password are collected. According to Table 5.1, all the test samples were randomly split into training- and test-sets pairs. After that, there are altogether 384 sets pairs to run the first test. At last, Bayes-Nets, KStar and K-Nearest Neighbor classifiers were used to classify each set pair simultaneously. In order to acquire a trusted and reliable result, the structure of the test is shown in the following table. (see Table 5.5)

TABLE 5.5: TEST STRUCTURE USING 6 DIFFERENT LEVEL OF PERCENTAGE-SPLIT

N% split \ classifiers	Bayes-Nets	KStar	KNN
50%			
60%	*N% indicates the entire test samples are split into*		
70%	*a pair of training set (N%) and test set (1 —N%)*		
80%	*based on each writer individually.*		
90%	*Each level has 384 randomly generated pairs.*		
Cross F 20			

5.2.1.2 Feature Extraction and Selection

The following 25 features are extracted for the experiments (see Table 5.6) (detailed description see Chapter 4):

TABLE 5.6: 25 GEOMETRICAL, STATISTICAL AND TEMPRAL FEATURES

Type of Feature	Feature
Geometrical:	WORD_WIDTH and WORD_HEIGHT
	SLANT_AMPLITUDE
	HORIZONTAL SKEWNESS
	VERTICAL SKEWNESS
	POINTS

	POINTAGNEL
	SEGMENTS
	REGANGLE
	EUCLID
	WIDTH, HEIGHT and SURFACE
	HYPANGEL
Statistical:	NUM_STROKES
	RELATIVE_WRITING_DURATION
	WORD_MID_X
	WORD_MID_Y
	AvgSLT
	RELATIVE_HEIGHT
	ORIENTATION
	INERTIAL RATIO
	ASPECT RATIO
	SPREADNESS
	BALANCE of HORIZONTAL EXTENSION
Temporal:	TIME

The feature selection (see Figure 5.5) is done by using the Generalized Fisher Score (detailed description see Chapter 4).

5.2.1.3 Results

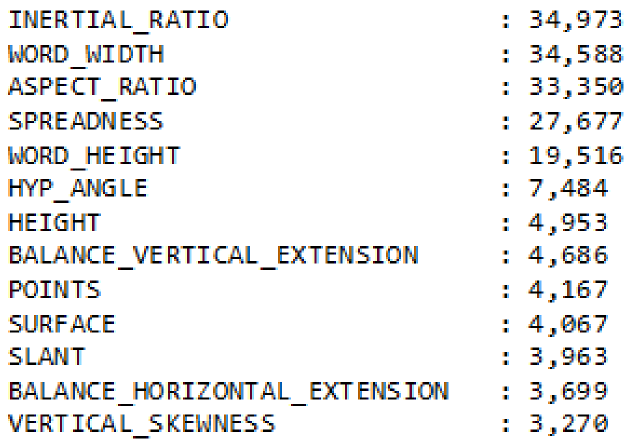
The result of the first test is shown in Table 5.7:

TABLE 5.7: SUCCESS RATES OF A CORRECT CLASSIFICATION FIRST TEST

classifiers / N%split	Bayes-Nets	KStar	KNN
50%	86.15%	91.79%	93.33%
60%	90.38%	94.87%	95.51%
70%	91.45%	95.73%	97.44%
80%	93.59%	96.15%	97.44%
90%	97.44%	97.44%	97.44%
Cross F 20	92.58%	94.63%	95.40%

By using the Fisher Score (detailed description see Chapter 4) of each feature can be independently calculated based on all test samples acquired from 32 subjects. The result is shown in Figure 5.5 and Figure 5.6.

According to the results, which were given above, some assumptions and conclusions can be made. Firstly, it is undoubtedly that the features Orientation and Slant Amplitude got the lowest score. This means, that the angle between the whole signature and the horizontal axis as well as the difference between maximum and minimum of amplitude for all single strokes are inexpedient for feature generation with this dataset. Secondly the highest score is won by the features Inertial Ratio, Aspect Ratio and Spreadness. These are generated features from second order moments, this means, they depend on the squares of the divergence from character coordinate according to their likelihood weights what seems to be a very good differentiation criterion for this dataset.



```
INERTIAL_RATIO                 : 34,973
WORD_WIDTH                     : 34,588
ASPECT_RATIO                   : 33,350
SPREADNESS                     : 27,677
WORD_HEIGHT                    : 19,516
HYP_ANGLE                      : 7,484
HEIGHT                         : 4,953
BALANCE_VERTICAL_EXTENSION     : 4,686
POINTS                         : 4,167
SURFACE                        : 4,067
SLANT                          : 3,963
BALANCE_HORIZONTAL_EXTENSION   : 3,699
VERTICAL_SKEWNESS              : 3,270
```

Figure 5.5: First part result Fisher Score of different features

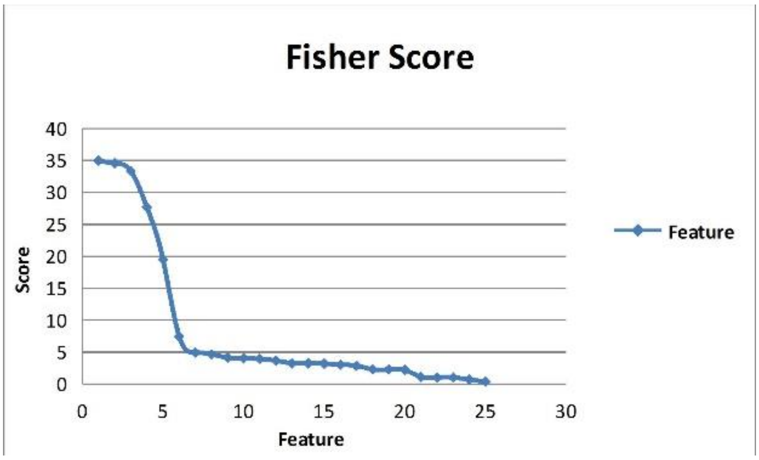


Figure 5.6: The line chart of different Fisher Score

Word- width and height got a high score too which means width and height of the same character sequences had a significant difference among different writers.

For second test the top 23 features are selected according to their Fisher Score (see Fig. 9, Fig. 10). The Bayes-Net classifier gave for the mean little bit les result than before. The KStar and KNN classifiers gave better performance compared to the selection of full feature set. As a result, in the second test can be determined with Fisher Score the best results for two classifiers and split of training- and test set achieved in the second test. The results of the second test using 23 features are shown in Table 5.8.

TABLE 5.8: SUCCESS RATES OF A CORRECT CLASSIFICATION SECOND TEST

classifiers / N% split	Bayes-Nets	KStar	KNN
50%	83.59%	93.33%	94.36%
60%	89.10%	97.44%	96.15%
70%	89.74%	97.44%	97.44%
80%	92.31%	98.72%	98.72%
90%	97.44%	97.44%	97.44%
Cross F 20	92.07%	95.14%	96.16%

The results of the impostor test (false accepted) with 32 subjects, 12 training samples each subject and test with three forgeries, two forgeries and one forgery are shown in Table 5.9:

TABLE 5.9: FALSE ACCAPTANCE RATES OF A CORRECT CLASSIFICATION SECOND TEST

classifiers / forgeries	Bayes-Nets	KStar	KNN
3	13.54%	10.42%	12,50%
2	18.46%	10.77%	15.38%
1	18.75%	12.50%	15.63%

It has been demonstrated, that the best results will be achieved with a high number of training-sets, the relation 9/10 to 1/10 and 8/10 to 2/10 delivers best result. After fisher score feature ranking and selection better result for two best classifiers are achieved. It can be ascertained that with Fisher Score and a sensible reduction of the features the results of the classification can be easily improved. The final Imposter test has shown that with this method using the best classifiers the access to an account with knowledge of the password can be lowered up to FAR of 10.42% if the forger tried three times to enter the handwritten password.

5.2.2 Experiments with Secure Password DB-32 and 39 Features

5.2.2.1 Experimental Setup

To achieve the best possible results 39 features used for this experiment, beside some statistical features are primarily time, speed and relation features used [160]. Specifically, to the 25 features from the experiment paragraph 5.2.1 the features are expanded with the following features described in Chapter 4 (see Table 5.10):

TABLE 5.10: 39 GEOMETRICAL, RELATIONS, TIME AND SPEED FEATURES

Type of Feature	Feature
Geometrical:	SPHI
	HORIZONTAL_POINT_ANGLE
Relations:	MEDIAN X or Y
	POINTS LOWER HORLINE or UPPER HORELNE
	POINTS ON HORLINE
	STANDARD DERIVATION X or Y
	MEAN NUMBER POINTS/SEGMENTS
	X or Y NUMBER POINTS SEGMENTS
	RELATION POINTS SPEED
	RELATION NVxz or Nvyz
	RELATION VX MAX
	RELATION VX or VY POSITIVE
	RELATION VX or VY NEGATIVE
Time and Speed:	TIME MAX X or Y
	TIME MIN X or Y
	SPEED ALL
	TIME_X_NEG or TIME_Y_NEG
	TIME_X_POS or TIME_Y_POS
	MAX VX or VY
	MIN VX or VY

For the FAR test 3 forgeries per user are collected from user with knowledge of the password.

According to Table 5.11, all the test samples were randomly split into training- and test-sets pairs. After that, we had altogether 384 samples to run the first experiment. At last, Bayes-Nets, Naive Bayes, k-Nearest Neighbor (KNN) and Multi-Layer Perceptron (MLP) classifiers were used to classify each set.

TABLE 5.11: TEST STRUCTURE USING 6 DIFFERENT LEVEL OF PERCENTAGE-SPLIT

classifiers N% split	Bayes-Nets	KStar	KNN	MLP
50%	*N% indicates the entire test samples are split into a pair of training set (N%) and test set (1 —N%) based on each writer individually.* *Each level has 384 randomly generated pairs.*			
60%				
70%				
80%				
90%				
Cross F 20				
AVG time in s	*Average time in seconds each classifier need for classification process.*			

With different ranking algorithms we want to reduce the features to achieve the same or better result after ranking and reduce the processing time for classification. In the tables we will show the results after classification with different classifiers. The number of features, result after classification and time for classification process will be shown and discussed.

5.2.2.2 Feature Selection

For feature selection the Generalized Fisher Score is used again. As second ranker the Information Gain Attribute Evaluation (IG) is used (detailed description see chapter 4). This ranker evaluates the worth of an attribute by measuring the information gain with respect to the class.

5.2.2.3 Classification

For the first experiment Table 5.12 we work with all 39 features, different classifier and percentage split in training- and test set. The average time (AVG time) will be extracted for every classifier. For second experiment we select parameter with Generalized Fisher Score and run same experiments like Table 5.11: Test structure using 6 different level of percentage-split. For third experiment we select features with Information Gain Attribute Evaluation ranker and run same test like Table 5.11 and for the last experiment (Test with Imposter) we use parameter from best result Table 5.12 and run same experiments like Table 5.11 to get best result for false acceptance rate.

5.2.2.4 Results

The result of the first test is shown in Table 5.12. In this test we work with all 39 features the fastest classification is done with Naive Bayes and KNN, best result for classification we reach with Bayes Net classifier. Cross validation for MLP delivered after five minutes no result therefore we left it out.

TABLE 5.12: STRUCTURE USING 6 DIFFERENT LEVEL OF PERCENTAGE-SPLIT (39 FEATURES)

classifiers N%split	Bayes-Nets	Naïve Bayes	KNN	MLP
50%	95.41%	89.29%	89.29%	94.90
60%	95.54%	91.08%	92.36%	96.82
70%	95.76%	90.68%	93.22%	96.61
80%	98.72%	97.44%	93.56%	96.15
90%	100%	97.44%	94.87%	97.44
Cross F 20	98.98%	98.47%	95.41%	-
AVG time in s	0.06	$<10^{-2}$	$<10^{-2}$	12.10

The result of the second test is shown in Table 5.13. The second test is carried out with 31 features after using Generalized Fisher Score for feature selection. The AVG time is reduced because less parameters for classification, the result is bit worse than before. Best results are reached with Naïve Bayes and KNN classifier.

TABLE 5.13: SUCCESS RATES OF A CORRECT CLASSIFICATION (31 FEATURES AFTER USING GENERALIZED FISHER SCORE)

classifiers N%split	Bayes-Nets	Naïve Bayes	KNN	MLP
50%	91.33%	86.22%	93.88%	96.43
60%	91.72%	89.81%	94.27%	96.18
70%	91.53%	90.68%	94.10%	94.10
80%	94.87%	98.72%	98.72%	97.44

90%	94,87%	97.44%	97.44%	97.44
Cross F 20	97.45%	98.47%	96.94%	-
AVG time in s	0.03	<10^{-2}	<10^{-2}	7.25

The result of the third test is shown in Table 5.14. In this test we work with 25 features after ranking with Information Gain Attribute Evaluation and reach best results of 100% for correct classification with Bayes Net and Naive Bayes classifiers. The classification process is very fast for both classifiers what the AVG time shows. This could affect positively in a real-time application system.

TABLE 5.14: SUCCESS RATES OF A CORRECT CLASSIFICATION (25 FEATURES AFTER RANKING WITH INFORMATION GAIN ATTRIBUTE EVALUATION)

classifiers N%split	Bayes-Nets	Naïve Bayes	KNN	MLP
50%	95.41%	88.27%	93.37%	93.88
60%	95.54%	92.36%	93.63%	93.63
70%	95.76%	88.98%	94.07%	93.22
80%	97.44%	100.00%	94.87%	96.15
90%	100,00%	97.44%	94.87%	97.44
Cross F 20	98.46%	98.47%	95.92%	-
AVG time in s	0.04	<10^{-2}	<10^{-2}	7.70

The results of the impostor test (false accepted) with 32 users, 12 training samples each user and test with one forgery, two forgeries and three forgeries after ranking with Information Gain Attribute Evaluation are shown in Table 5.15.

TABLE 5.15: FOURTH TEST: IMPOSTOR TEST FALSE ACCEPTED

classifiers forgeries	Bayes-Nets	Naïve Bayes	KNN	MLP
1	3.13%	12.50%	9,38%	15.63%
2	3.13%	12.50%	12.50%	12,50%
3	4.17%	10.42%	11,46%	10.42%
AVG time in s	0.16	0.03	<10^{-2}	15.52

The classification with all selected 39 features with priority time, speed and relations produces with 100% correct classification for Bayes Net classifier delivers already good results. After the ranking with Fischer Score the calculation time could be shortened by the parameter reduction for the classification. The third test delivered the best result with only 25 features and the impostor test with Bayes Net classifier the best result with FAR of 3.13% and 1 or 2 forgeries.

5.2.3 Experiments with ATV-SLT DB – with 7 Sessions and 64 Features

5.2.3.1 Experimental Setup

For this experiment the ATV-Signature Long Term Database (ATV- SLT DB) is used (see chapter IV for detailed description) collected in 6 sessions (6 signatures for each user collected in a time period of 15 month). 64 static and dynamic biometric features are extracted from the ATV-SLT DB sessions. For the impostor test a 7th session is added. The three known classifiers Naïve Bayes, Bayes Net, and KNN are used for the experiments 0.

The main objectives of the experiments are:

- whether the forgery resistance is changing over a long time period.
- whether the use of displays without pressure sensors can be recommended in authentication systems
- whether the authentication system is forgery resistant, if skilled the impostor sees all original signatures during the forgery.

5.2.3.2 Feature Selection

After transformation the feature data files with the segmentation, x, y coordinates and times were generated. From this data files the parameters were generated. Beside some statistical parameters primarily time, speed and relation parameter were used. The detailed description of the extracted parameters from the online signatures are in chapter 4.

5.2.3.3 Results

Four experiments were performed: first, the first four sessions were used separately and one session for training and one for testing. Each session contains four samples. In the second experiment two sessions' pairs were combined for training and one for testing. In the third experiment all 6 sessions were used and split percentage into train and test data for the classification. In the final experiment the impostor test was performed with all six sessions for training and the seventh session for testing. The following part figure out the results of the experiments:

1. **First Experiment**

The results of the first experiment training and testing all first four sessions separately and classify with three classifiers Naïve Bayes, Bayes Net, and KNN are shown in Table 5.16, Table 5.17 and Table 5.18:

TABLE 5.16: TEST NAÏVE BAYES CLASSIFIER

test session train	**first**	**second**	**third**	**fourth**
first	100%	83.33%	63.89%	40.74%
second	77.78%	100%	66.67	37.04%
third	51.85%	53.70%	100%	75.00%
fourth	47.22%	45.37%	86.81%	100%

TABLE 5.17: TEST BAYES NET CLASSIFIER

test session train	**first**	**second**	**third**	**fourth**
first	100%	96.30%	89.81%	78.70%
second	89.15%	100%	89.81%	77.78%
third	94.44%	87.04%	100%	83.33%
fourth	82.41%	74.07%	89.81%	100%

TABLE 5.18: TEST KNN CLASSIFIER

test session train	**first**	**second**	**third**	**fourth**
first	100%	94.44%	67.50%	56.48%
second	92.59%	100%	68.52%	51.85%
third	66.67%	66.67%	100%	82.41%
fourth	58.33%	53.70%	87.04%	100%

The first and second sessions are relatively similar and achieve good results, the third and fourth session achieves worse results compared with the first and second session. The best result of 96.30% correctly classified instances delivers the Bayes Net classifier with the first session for training and the second session for testing.

2. Second experiment

The results of the second experiment that combines session pairs as training sets and use one session as test set are shown in Table 5.19, Table 5.20 and Table 5.21:

TABLE 5.19: TEST NAÏVE BAYES CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	53.70%	35.19%
first_third	100%	99.07%	100%	78.70%
first_fourth	100%	94.44%	88.89%	100%
second_third	92.60%	100%	100%	75.00%
second_fourth	96.30%	100%	93.52%	100%
third_fourth	67.69%	67.59%	100%	100%

TABLE 5.20: TEST BAYES NET CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	88.89%	78.70%
first_third	100%	97.22%	100%	89.81%
first_fourth	100%	95.37%	96.30%	100%
second_third	95.37%	100%	100%	92.60%
second_fourth	97.22%	100%	98.15	100%
third_fourth	89.81%	89.81%	100%	100%

TABLE 5.21: TEST KNN CLASSIFIER

test session train	first	second	third	fourth
first_second	100%	100%	65.74%	60.19%
first_third	100%	95.37%	100%	83.33%
first_fourth	100%	93.52%	97.22%	100%
second_third	97.22%	100%	100%	83.33%
second_fourth	98.15%	100%	96.30%	100%
third_fourth	69.44%	66.67%	100%	100%

In this experiment, two sessions with eight training samples and four test samples per user were summarized. The best result of 98.15% correctly classified deliver the Bayes Net and the k-Nearest Neighbor classifier with the second_fourth session for training, Bayes Net with the third session for testing and k-Nearest Neighbor with the first session for testing.

3. Third Experiment

For the third experiment all samples were split beginning from first up to sixth session randomly into training- and test pairs. After that, there are altogether 1296 sets ((4 samples x 4 sessions + 16 samples x 2 sessions) x 27 writer) to run the third experiment. At last, Bayes-Nets, Naïve Bayes and k-Nearest Neighbor classifiers were used to classify each pair simultaneously. The results of the third experiment are shown in Table 5.22 and Figure 5.7:

TABLE 5.22: TEST SPLIT ALL CLASSIFIER

classifiers N%split	Bayes-Nets	Naive Bayes	KNN
50%	98.02%	97.36%	96.54%
60%	98.35%	97.33%	97.53%
70%	98.46%	96.98%	97.80%
80%	98.35%	97.12%	97.12%
90%	99.17%	97.52%	98.35%
Cross F 20	98.85%	97.86%	97.86%
AVG Time in s	0.35	0.06	0.01

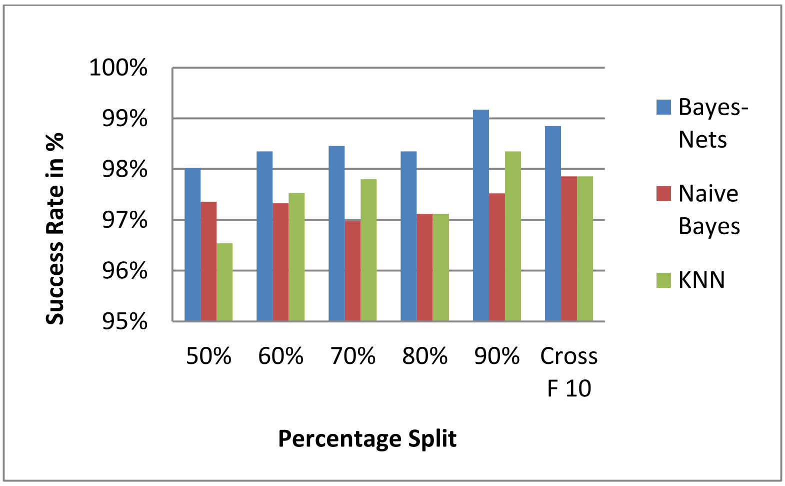


Figure 5.7: Bar chart of the third experiment

4. Fourth Experiment

For the fourth experiment all six sessions were used for training and the seventh session with six skilled forgeries per user for testing. There are all together 162 forgeries for the experiment. First all sessions were trained separately and secondly all sessions together. Finally, the average time for the classification process was calculated. The results of the fourth experiment are shown in Table 5.23 and Figure 5.8:

TABLE 5.23: IMPOSTOR TEST

classifiers session	Bayes-Nets	Naive Bayes	KNN
first	10.49%	7.41%	10.49%
second	11.11%	8.02%	11.73%
third	8.64%	4.32%	4.93%
fourth	8.02%	7.41%	4.32%
fifth	9.64%	9.88%	9.88%
sixth	7.41%	10.49%	10.49%
all sessions	11.73%	2.47%	11.11%
AVG Time in s	0.39	0.02	0.05

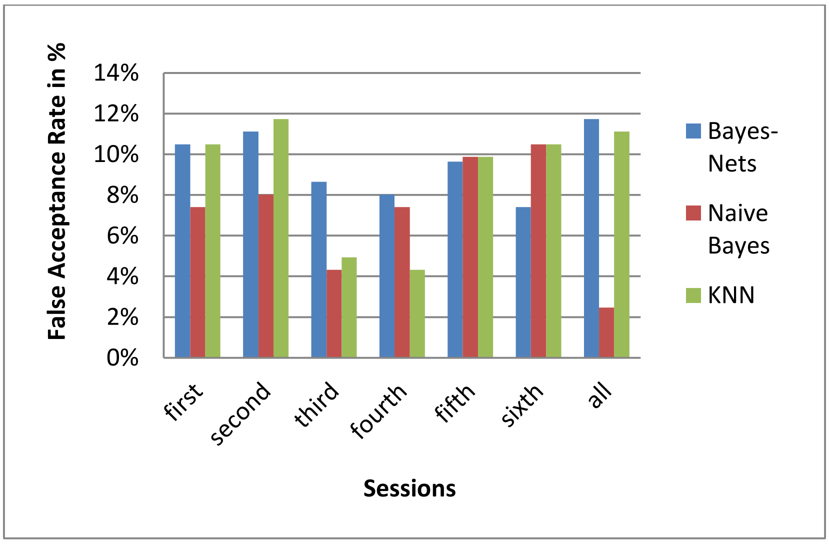


Figure 5.8: Bar chart of the fourth experiment

In this series of experiments, it can be noted that we can recommend the use of displays without pressure sensors, because the writer verification results with and without pressure seem not be different. The results of using all 64 features can be confirmed. Aging of handwriting over a time period of 15 month has a negative influence to the correct writer verification, if only one session is used as training set. The results can be improved by using more sessions of signatures. The use of many signatures over a long time period is recommended.

The recommended system is forgery resistant for skilled forger who can see the original signature with the FAR of 2.47%, if Naïve Bayes for the classification and all 48 signatures per writer of the time period of 15 months are used.

5.2.4 Evaluation of Secure-Block Letters Passwords and Cursive Texts with 67 Features

5.2.4.1 Experimental Setup

As described in Chapter 4.1 and 4.2 the IAM handwriting public database with text samples of 220 persons and the private secure password DB-150 with sessions of secure passwords of 150 Users is used for this experiments

The main objective of this experiment is to test the verification system by using cursive texts as well as block letter words. For the efficiency tests, the extracted 67 features are classified into three parameter types geometrical, statistical, and temporal parameters, and two feature reduction methods Fisher Score, and Info Gain Attribute Evaluation are applied. At first the classification of the handwritten passwords by using the different parameter types geometrical, statistical and temporal parameters separately are carried out (see Table 5.24).

TABLE 5.24: SEPARATE PARAMETER TESTS

Type Features	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10 Split 66%	Result correct classified in %		
AVG Time in s		AVG Time in s		

Afterwards all 67 features described in chapter 4 used for the writer identification.

5.2.4.2 Feature Selection

From all data records of the two databases the identical 67 features are extracted, divided in:

1. Geometrical Features: 26 different geometrical features are extracted.
2. Statistical Features: 27 different statistical features are extracted.
3. Temporal Features:14 different temporal features are extracted.

5.2.4.3 Feature Reduction

The already used before two filter methods are used also for this experiment:

- Fisher Score
- Info Gain Attribute Evaluation

5.2.4.4 Results

I. SECURE-PASSWORD-DB-150 (BLOCK LETTERS)

1. **Geometrical Features**

Table 5.25 shows, that the geometrical features for the secure passwords deliver not the best results. K-Nearest Neighbor classifier archives the best result for cross validation. It seems the geometrical parameters are not suitable well for this dataset. The time for classification is less than one second.

Table 5.25: Geometrical Features Secure Passwort DB 150

Features	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
Geom etrical	Cross 10	55.33%	69,67%	76%
	Split 66%	44,12%	41,18%	63%
AVG Time in s		0.11	$<10^{-2}$	$<10^{-2}$

2. **Statistical Features**

In Table 5.26 is demonstrated, that the statistical features deliver better results than geometrical parameters. The k-Nearest Neighbor classifier delivers best result for classification.

TABLE 5.26: STATISTICAL FEATURES SECURE PASSWORD DB 150

Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
Statis-tical	Cross 10	72,33%	72,83%	84,50%
	Split 66%	59,31%	49,02%	72,55%
AVG Time in s		0.11	<10^{-2}	<10^{-2}

3. Temporal Features

In Table 5.27 is demonstrated, that the temporal features deliver best result, the best classifier is Bayes Net classifier closely followed by Naïve Bayes classifier for cross validation. The time for the classification is the same like before.

TABLE 5.27: TEMPORAL PARAMETERS SECURE PASSWORD DB 150

Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
Temp oral	Cross 10	97,83%	96,50%	87,17%
	Split 66%	97,06%	85,20%	78,43%
AVG Time in s		0.11	<10^{-2}	<10^{-2}

4. Using All Features

In Table 5.28 is shown that using all features of parameter contributes once again leads to an improvement of the classification results. Although geometrical and statistical parameters individually achieved worse result, by the use of all parameters the classification rate grows nearly about 2 percent for the best result. Bayes Net protrudes again with the best result for the cross validation split 10%. Using all parameters affects a little bit negatively the average time is for the classification. It's also not surprising, however, with less number of the parameters time the system need less time for classification. However, for the practical application this time should be reduced under retention of the good classification results. In the next Table 5.29 we will have additionally a look for the experiments for the feature reduction starting with Fisher Score.

TABLE 5.28: ALL FEATURES SECURE PASSWORD DB 150

		Classifiers		
	Classification	***Bayes Net***	***Naïve Bayes***	***K Nearest Neighbor***
All Features	Cross 10	99,00%	94.83%	94,17%
	Split 66%	97,06%	78,92%	87,75%
AVG Time in s		0.56	0.03	$<10^{-2}$

TABLE 5.29: ALL FEATURES FISHER SCORE SECURE PASSWORD DB 150

Score	**Ranked attributes**
2028141,973	TIME_V_MAX
1984766,469	TIME_MAX_Y
1897897,559	TIME_VX_MIN
1887958,830	TIME_VY_MAX
1869689,600	TIME_VY_MIN
1868679,810	TIME_VX_MAX
1867024,784	TIME_MIN_Y
1841432,169	TIME_MAX_X
1841432,169	TIME_MIN_X
13,080	STANDARD_DERIVATION_Y
11,425	INERTIAL_RATIO
10,814	ASPECT_RATIO
10,172	SPHI
9,618	WORD_HEIGHT
8,080	DTW_Y
7,492	MEDIAN_Y
7,342	SPREADNESS
7,240	TIME_Y_NEG
6,642	TIME_X_POS
6,636	TIME_X_NEG
6,563	TIME_Y_POS
6,544	REGRESSION_LOWER_HORANGLE

6,480	REGRESSION_ON_HORANGLE
6,403	HEIGHT
6,203	CENTRAL_POINT
4,697	VERTICAL_SKEWNESS
4,689	MAX_VY
4,608	TIME
4,577	HYP_ANGLE

After reducing parameter with Fisher Score the results for Naïve Bayes and K-Nearest Neighbor classifier are a bit better. The AVG time for classification is nearly the same, because high number of parameters for classification. (see Table 5.30)

TABLE 5.30: ALL FEATURES RESULTS FISHER SCORE SECURE PASSWORD DB 150

All Parameters	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	99,00%	95,50%	96,17%
	Split 66%	97,06%	79,90%	87,75%
AVG Time in s		0.51	0.03	$<10^{-2}$

The ranked result of Info Gain Attribute Select, the ranker selects 45 features for classification. (see Table 5.31) Nearly all temporal parameters have the highest score with distance followed by the statistical features and the geometrical features.

TABLE 5.31: RANKED RESULT INFO GAIN ATTRIBUTE SELECT SECURE PASSWORD DB 150

Score	**Ranked attributes**
6.5849	TIME_V_MAX
6.578	TIME_VY_MIN
6.578	TIME_MAX_X
6.578	TIME_MIN_X
6.578	TIME_VX_MIN
6.578	TIME_MIN_Y
6.578	TIME_MAX_Y

6.5773	TIME_VX_MAX
6.5711	TIME_VY_MAX
2.2422	ASPECT_RATIO
2.1864	STANDARD_DERIVATION_Y
2.0582	INERTIAL_RATIO
1.7612	REGRESSION_ON_HORANGLE
1.7574	REGRESSION_LOWER_HORANGLE
1.6921	CENTRAL_POINT
1.6054	TIME_Y_POS
1.5582	MEDIAN_Y
1.5236	VERTICAL_SKEWNESS
1.5193	SPHI
1.4362	WORD_HEIGHT
1.3829	SEGMENTS
1.3829	NUM_STROKES
1.371	TIME_Y_NEG
…	…

In Table 5.32 is demonstrated, that after ranking with Info Gain Attribute Select the results for Naïve Bayes and K-Nearest Neighbor classifier are growing once again up a bit in spite of considerably reducing of the number of parameters. By the parameter reduction the classification time considerably decreases to AVG of 0.24 s for Bayes Net classifier.

TABLE 5.32: ALL FEATURES INFO GAIN ATTRIBUTE SELECT SECURE PASSWORT DB 150

All Parameters	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	99,00%	96,33%	96,50%
	Split 66%	97,06%	83,33%	93.14%
AVG Time in s		0.24	$<10^{-2}$	$<10^{-2}$

Now these results are compared with the results of the analysis of the handwritten signatures from the IAM online handwriting DB.

II. IAM ONLINE HANDWRITING DATABASE (CURSIVE TEXT)

1. Geometrical Features

In Table 5.33is demonstrated, the best result of 54.62% correct classification we achieved with Naïve Bayes classifier. Summarized we reach only averagely good results for the geometrical parameter.

TABLE 5.33: GEOMETRICAL FEATURES IAM ONLINE HANDWRITING DB

	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
Geometrical Features	Cross 10	17,18%	54,62%	34,10%
	Split 66%	9,81%	34,53%	28,11%
AVG Time in s		0.45	0.01	$<10^{-2}$

2. Statistical Features

In Table 5.34 is shown, that statistical parameters results are a little bit worse than the geometrical parameters best result with 35.32% correct classification reached the Naïve Bayes classifier for cross validation. The time for classification is nearly the same like for the geometrical parameters.

TABLE 5.34: STATISTICAL FEATURES IAM ONLINE HANDWRITING DATABASE

	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
Statistical Features	Cross 10	16,54%	35,32%	10,06%
	Split 66%	10,75%	20,38%	10,38%
AVG Time in s		0,44	0,02	$<10^{-2}$

3. Temporal Features

In Table 5.35 is demonstrated, the temporal achieved best result of 96.15% correct classified for Naïve Bayes classifier with cross validation. The maximum time for classification is with 0.5s only bit longer than the time for the other features.

TABLE 5.35: TEMPORAL FEATURES IAM ONLINE HANDWRITING DATABASE

Temporal Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
	Cross 10	95.84%	96.15%	58.21%
	Split 66%	89.43%	90.94%	48.87%
AVG Time in s		0.5	$<10^{-2}$	$<10^{-2}$

4. **Using All Features**

In Table 5.36 all features summarized where shown. A small increase compared to the temporal parameters can be noted. The best classifier is Bayes Net with 98.65% correct classification. The time for the classification lies with 1.41s still within the scope of.

TABLE 5.36: ALL FEATURES IAM ONLINE HANDWRITING DATABASE

All Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
	Cross 10	98,65%	93,53%	36,03%
	Split 66%	94,53%	80,57%	31,70%
AVG Time in s		1,41	0.05	$<10^{-2}$

5. **Feature Reduction using Fisher Score**

In Table 5.37 are demonstrated, the ranked results of Fisher Score Reduction for IAM online handwriting DB, the ranker select 59 parameters for classification. By far the best score achieves the time parameters followed by geometrical and statistical parameters.

TABLE 5.37: FISHER SCORE ALL FEATURES IAM ONLINE HANDRWITING DB

Score	Ranked attributes
344702,213	TIME_MAX_Y
344309,419	TIME_V_MAX
340892,335	TIME_VX_MAX
337972,869	TIME_VY_MAX
329873,423	TIME_VY_MIN
328062,080	TIME_MAX_X

328062,080	TIME_MIN_X
327360,471	TIME_MIN_Y
323060,762	TIME_VX_MIN
18,411	SLANT
9,381	TIME
9,348	POINTS
9,328	MEAN_NUMBER_POINTS/SEGMENT
8,913	RELATION_VX_MAX
6,362	WIDTH
5,661	REG_ANGLE
3,451	X_NUMBER_POINTS/SEGMENT
3,047	RELATION_NVyz
2,832	RELATIVE_WRITING_DURATION
2,726	RELATION_NVxz
2,645	HEIGHT
2,597	Y_NUMBER_POINTS/SEGMENT
2,540	POINT_ANGLE
…	…

After Fisher Score Reduction in Table 5.38 where demonstrated, that the results for Naïve Bayes and K-Nearest Neighbor are a bit better than before. By the reduction of the parameters the AVG Time decreases to 1.3s for Bayes Net classifier.

Table 5.38: Results All Features IAM Online Handrwiting DB

All Features	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	98,65%	96,92%	64,42%
	Split 66%	94,53%	85,85%	59,10%
AVG Time in s		1,3	0.04	$<10^{-2}$

6. Feature reduction using Info Gain Attribute Select

In Table 5.39 is demonstrated the ranked result of Info Gain Attribute Select, the ranker selected 22 features for the classification.

TABLE 5.39: INFO GAIN ATTRIBUTE SELECT ALL FEATURES IAM ONLINE HANDRWITING DB

Score	Ranked attributes
7.4759	TIME_MIN_Y
7.4754	TIME_MIN_X
7.4754	TIME_MAX_X
7.4744	TIME_VX_MAX
7.4737	TIME_MAX_Y
7.4737	TIME_V_MAX
7.4732	TIME_VY_MIN
7.4727	TIME_VY_MAX
7.4719	TIME_VX_MIN
1.9984	SLANT
1.9317	RELATION_VX_MAX
1.5304	WIDTH
1.526	TIME
1.4895	POINTS
1.0784	MEAN_NUMBER_POINTS/SEGMENT
0.7132	REG_ANGLE
0.6634	HEIGHT
0.6405	EUCLID
0.6214	SURFACE
0.62	RELATION_TIME_GAP_ALL
0.5813	X_NUMBER_POINTS/SEGMENT
0.570	Y_NUMBER_POINTS/SEGMENT
…	…

In Table 5.40: Results Info Gain Attribute Select All Features IAM Online Handwriting DB is demonstrated that the result improved for Naïve Bayes and K-Nearest Neighbor classifier again in spite of the reduction on only 22 parameters. The time becomes considerably shorter with only 0.62s for Bayes Net classifier for classification after ranking with Info gain Attribute Select.

TABLE 5.40: RESULTS INFO GAIN ATTRIBUTE SELECT ALL FEATURES IAM ONLINE HANDWRITING DB

All Features	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	98.65%	98.34%	87.44%
	Split 66%	94.53%	93.40%	82.26%
AVG Time in s		0.62	0.01	$<10^{-2}$

The findings from this Experiments may be summarized as follows:

1. The 67 features those where developed for writer identification and verification of handwritten passwords are readily applicable for single character words as well as handwritten cursive texts. The best results delivered 99% correct classification for handwritten single character words and 98.34% for cursive texts.
2. The highest effects for correct classification results have the temporal features. But the result can be improved by additional using geometrical and statistical features.
3. Nevertheless, a feature reduction by Fisher Score or Info Gain Attribute Evaluation improves the results and the program performance.
4. The best classification results are reached using by Bayes Net and Naïve Bayes Classifier.

5.2.5 Experiments with offline features and secure password DB-150

5.2.5.1 Experimental Setup

In this experiment the offline features from the PNG which were created next to the online data are analyzed. The features are generated as described in Chapter 4 from the Zernike Moments up to order 15 and the Fourier Descriptors up to order 100. Data basis is the already used DB-32 database (described in Chapter 4) with 32 users who have written their secure password on a mobile device. Finally, an examination with all features is carried out. The classifiers Naive Bayes, Bayes Net and k-Nearest Neighbour are used again.

5.2.5.2 Feature Selection

From the dataset, 18 features for the Zernike Moments and 32 features for the Fourier Descriptors were extracted. For the final test all 117 online and offline features are extracted. For the final evaluation of the quality of the features, a ranking with Info Gain Attribute Evaluation and all 117 features is carried out.

5.2.5.3 Results

The results in Table 5.41 show that the classification result for Zernike Moments Generally give poor results, the best result with 57.90% correctly classified instances delivers the KNN classifier for the cross-validation with 10 folds. The Fourier Descriptors see Table 5.42 are similarly bad, the Naïve Bayes classifier delivers with 43.87% the best result for cross-validation with 10 folds.

In Table 5.43 the results for all features (117 offline and online features) are presented The results are nearly as good as in the previous experiment (see Table 5.28) with all online features (67), the best result is Bayes Net Classifier with 98.09%

Table 5.41: Zernike Moments Secure Password DB-150

Zernike Moments	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	11.44%	56.46%	57.90%
	Split 66%	3.91%	52.12%	52.77%
AVG Time in s		0.49	$<10^{-2}$	$<10^{-2}$

Table 5.42: Fourier Descriptors Secure Password DB-150

Fourier Descriptors	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	34.32%	43.87%	30.07%
	Split 66%	17.05%	39.09%	25.19%
AVG Time in s		0.94	0.02	$<10^{-2}$

Table 5.43: All Features (online and offline) Secure Password DB-150

All Features	**Classification**	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
	Cross 10	98.09%	91.24%	95.75%
	Split 66%	95.61%	89.47%	94.19%
AVG Time in s		3.94	0.05	0.01

The ranking with Info Gain Attribute Select Table 5.44 shows that, despite the shortening of the Fourier descriptors in the experiment, some of the descriptors get a high score in the ranking. This allows the assumption that special Fourier Descriptors are capable of improving the results of classification. The Zernike Moments, on the other hand, are found at the end of the lowest-score raking list and therefore do not appear to improve the classification. Since the current offline features are generally bad, they also require a higher time for feature extraction and classification, so they were not considered in further experiments.

Table 5.44: Info Gain Attribute Select All Features (online and offline) Secure Password DB-150

Score	**Ranked attributes**
6.8939	TIME_V_MAX
6.8897	TIME_MAX_Y
6.8896	TIME_VX_MIN
6.8838	TIME_MIN_Y
6.8813	TIME_VY_MAX
6.8808	TIME_VX_MAX
6.8793	TIME_MIN_X
6.8793	TIME_MAX_X
6.8721	TIME_VY_MIN
2.0707	STANDARD_DERIVATION_Y
2.0232	ASPECT_RATIO
1.8856	WORD_HEIGHT
1.849	INERTIAL_RATIO
1.7143	SPHI
1.5939	HEIGHT

1.5128	KomplexFourierKoeff_Betrag_C_44
1.506	KomplexFourierKoeff_Betrag_C_666
1.4544	KomplexFourierKoeff_Betrag_C_333
1.4533	SEGMENTS
…	…
0	ZERNIKE_MOMENT_12_4_BETRAG
0	KomplexFourierKoeff_Betrag_C_-225
0	KomplexFourierKoeff_Betrag_C_-108
0	KomplexFourierKoeff_Betrag_C_97
0	KomplexFourierKoeff_Betrag_C_-321
0	ZERNIKE_MOMENT_3_1_BETRAG
0	KomplexFourierKoeff_Betrag_C_-168
0	HORIZONTAL_POINT_ANGLE
0	ZERNIKE_MOMENT_5_3_BETRAG

5.2.6 Experiments with secure password and signature DB-30 and 67 Features

5.2.6.1 Experimental Setup

The password and signature DB-30 described in chapter 4 is used for this experiment. This dataset was generated using secure passwords and signatures that were created by one and the same 30 users on Tablet and Signature Pad. In the experiment, it will be examined how the use of the password and the signature affect the classification results with the known 3 classifiers. On the other hand, it will be evaluated how different devices such as the tablet and the signature pad (higher resolution) affect the results of the classification with all 67 developed features. In the end, a graphical evaluation of the tests with tablet, pad with signature and password is shown and evaluated

5.2.6.2 Results

Table 5.45 shows the result of classification of all 67 online features generated form Signature with Tablet. Bayes Net classifier delivers the best results with 100%. In Table 5.46, a similar result with the best result of 100% for correctly classified with Bayes Net classifier for the results from the feature extraction and classification of the data from the secure password.

Table 5.45: All 67 Features Secure Password and Signature DB-30 Signature using Tablet

	Classification	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
All Features	Cross 10	100%	94.19%	96.51%
	Split 66%	98.26%	91.39%	96.55%

Table 5.46: All 67 Features Secure Password and Signature DB-30 Secure Password using Tablet

	Classification	**Classifiers**		
		Bayes Net	***Naïve Bayes***	***K Nearest Neighbor***
All Features	Cross 10	100%	96.51%	96.51%
	Split 66%	98.26%	91.38%	96.55%

In Table 5.47, the features from the DB-30 signatures written on the pad are extracted and classified. Here too, very good results of 100% correctly classified for Bayes Net and k-Nearest Neighbor. The classification with the features for the password created on the pad (see Table 5.48) provides equally good results for Bayes Net and k-Nearest Neighbor classifier.

Table 5.47: All 67 Features Secure Password and Signature DB-30 Signature using Pad

All Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
	Cross 10	100%	90.74%	100%
	Split 66%	100%	88.89%	100%

TABLE 5.48: ALL 67 FEATURES SECURE PASSWORD AND SIGNATURE DB-30 SECURE PASSWORD USING PAD

All Features	Classification	Classifiers		
		Bayes Net	*Naïve Bayes*	*K Nearest Neighbor*
	Cross 10	100%	96.67%	100%
	Split 66%	100%	90%	100%

The graphical evaluation with ROC analysis shows the results of the training and test with password and signature on pad. In Figure 5.9, all records of the password from the pad were used for the training and all signatures from the pad for the test with KNN classifier. A very good result is obtained near 1, which suggests that the features used are independently of what the writer writes identifies the writer with a high probability.

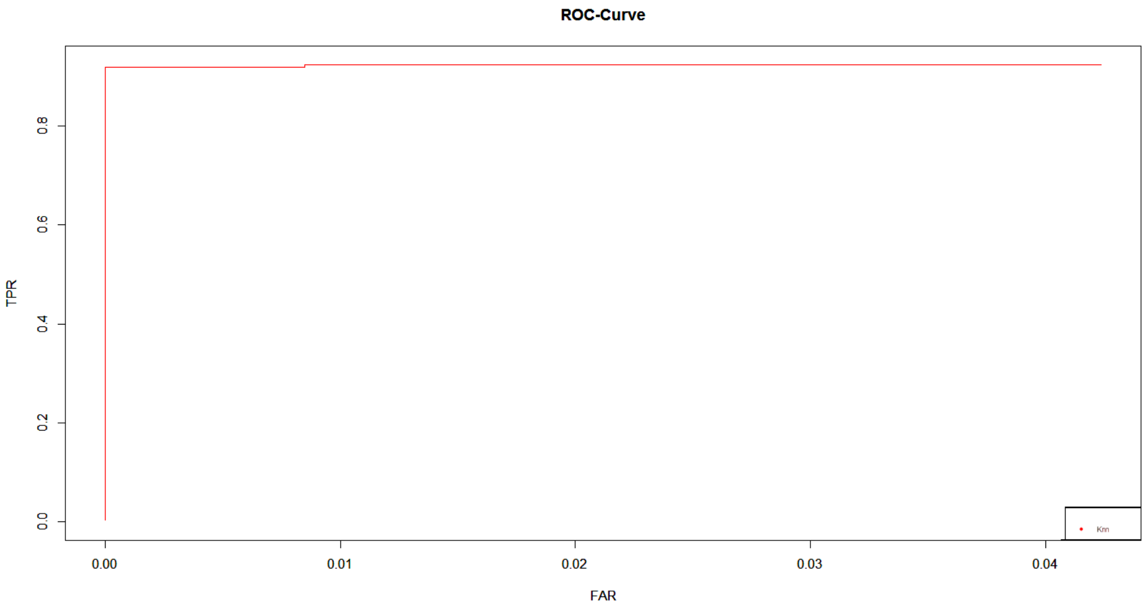


Figure 5.9: ROC analysis password (training) and signature (test) using pad with KNN

In figure 5.10, all records of the password from the tablet were used for the training and all signatures from the tablet for the test with KNN classifier. A good result is obtained near 0.9, which suggests that the features used independently also for the tablet of what the writer writes identifies the writer with a high probability.

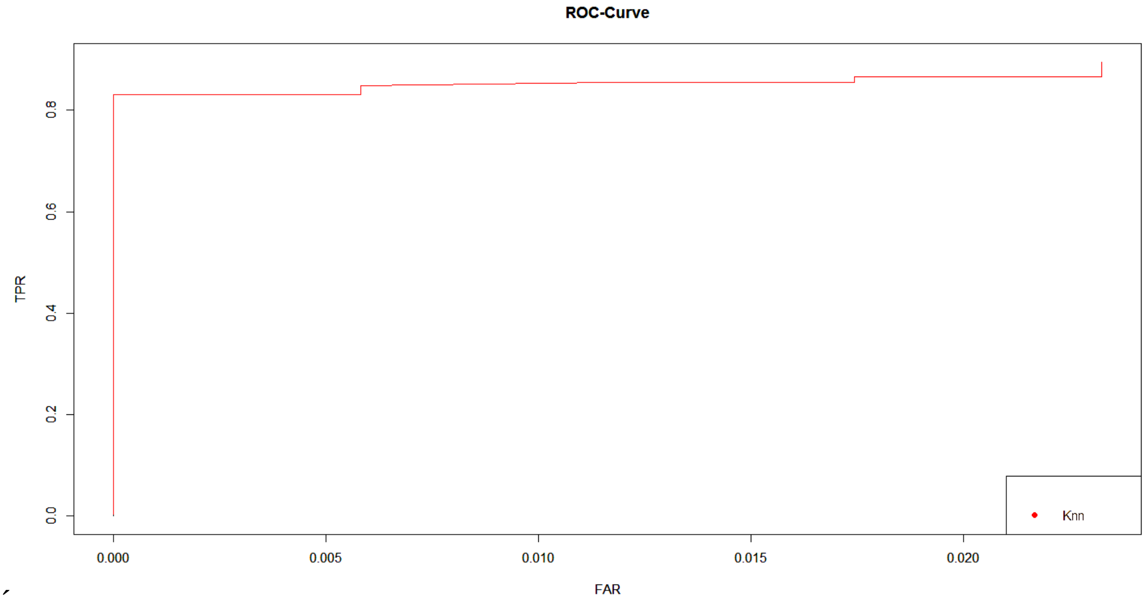


Figure 5.10: ROC analysis password (training) and signature (test) using tablet with KNN

In Figure 5.11, all records of the signatures from the tablet were used for the training and all passwords from the pad for the test with KNN classifier. A good result is obtained near 0.9, which suggests that the features used independently also for the device (pad and tablet) of what the writer writes identifies the writer with a high probability.

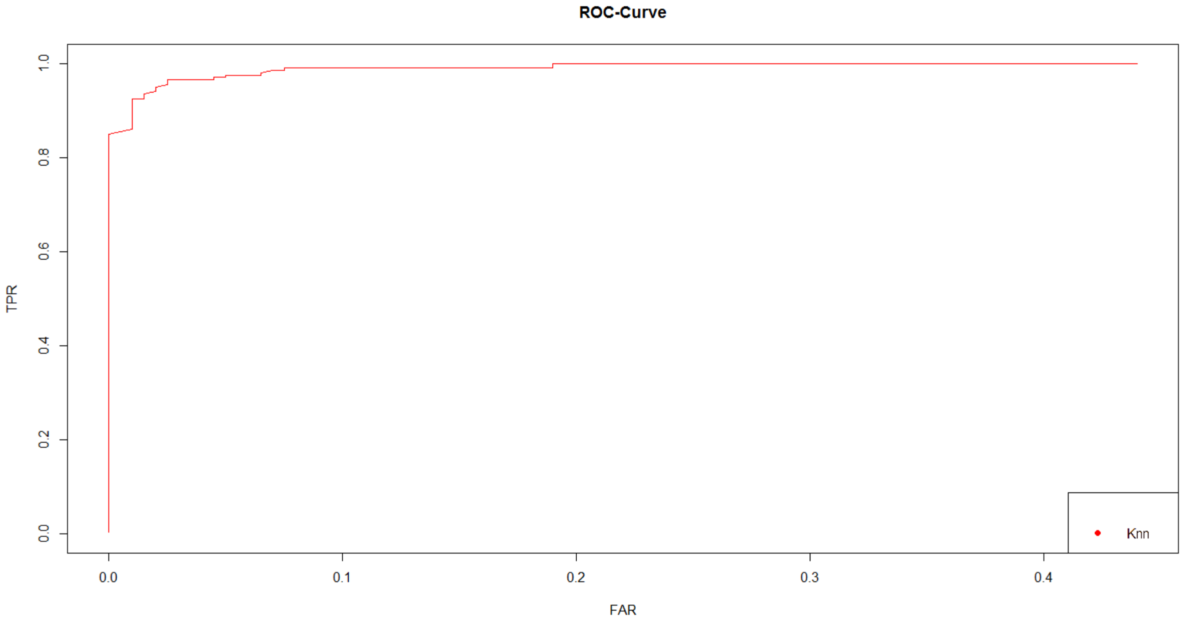


Figure 5.11: ROC analysis password (training) using tablet and signature (test) using pad with KNN

In Figure 5.12, all records of the signature from the tablet were used for the training and all signatures from the pad for the test with KNN classifier. A very good result is obtained near 1, which suggests that the features used independently also for signature with pad and tablet of what the writer writes identifies the writer with a high probability.

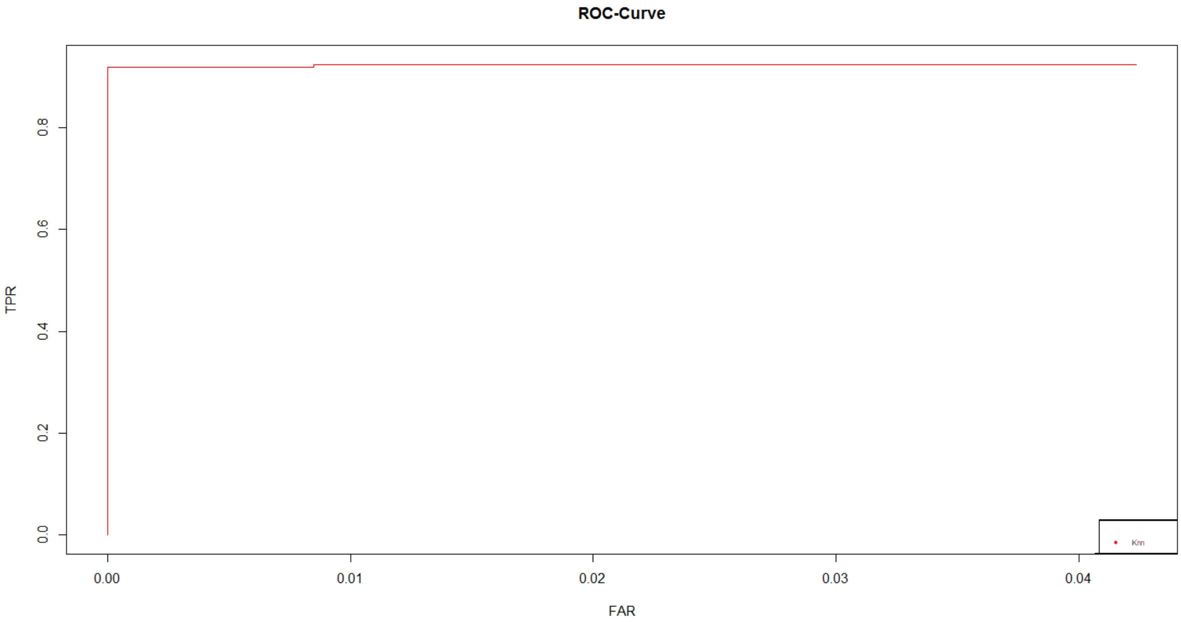


Figure 5.12: ROC analysis signature tablet and pad with KNN

5.3 Summary Results

In Table 5.49 the results of the tests are summarized. It can be ascertained that the results could be improved by stinging the number of features. An exception to this are the offline features, which did not lead to any improvement in the interaction with the online features. The results are strongly dependent on the data sets both the self-developed databases as well as the public databases can be achieved good results. Prerequisites are however always as error-free correctly collected records is not the case, the results deteriorate strongly. Both passwords, as well as signatures as well as cursive texts provide good results with the features and procedures used, independent of the device.

TABLE 5.49: SUMMARY BEST RESULTS ALL EXPERIMENTS

Experiment	*Database | Features*	*Accuracy (best result)*	*FAR*
1 (5.1.1)	DB-9 | 10	96.87%	11.11% - 22.22%
2a (5.1.2)	DB-9 | 10	100%	
2b (5.1.2)	BD-150 | 10	47.75%	
3 (5.2.1)	DB-32 | 25	98.72%	10.42% - 18.75%
4 (5.2.2)	DB-32 | 39	100%	3.13% - 12.50%
5 (5.2.3)	ATV-SLT DB | 64	100%	2.47% – 11.73%
6a (5.2.4)	DB-150 | 67	99%	
6b (5.2.4)	IAM HW DB | 67	98.65%	
7 (5.2.5)	DB-150 | 117	98.09%	
8 (5.2.6)	DB-30 | 67	100%	2% - 10%

Chapter 6 Conclusion and Future Work

6.1 Summary

First, it can be said, the hypothesis presented in this thesis is validated according to the results of the experiments of Chapter 5, and the underlying publications based on the developed databases and extracted features for writer identification.
The research cooperation between BTU and ULPGC and realization of this thesis were only possible thanks the renewed cooperation agreement between BTU and ULPGC.

In this thesis a system for writer identification with handwritten passwords is presented. Features for identification and verification have been developed and tested. For the tests, besides public databases, new databases were developed specially for secure handwritten passwords and signatures. The methods for preprocessing, feature extraction, classification to identify the writer were presented and applied. In several experiments, the developed features were examined for their robustness and suitability to identify writers specifically with smartphone and tablet with secure passwords and signatures for the identification of writers.

After an overview of this work and the motivation for this work was clarified in chapter 1, the various biometric and non-biometric systems for access control were presented, evaluated and the hypothesis established.

In Chapter 2 the methods for verification, identification and measurement of biometric quality are explained. The most important public databases are presented and the structure of self-developed databases is presented. A detailed overview of the most important state of the art publications of recent years on offline and online writer verification and identification is given and discussed. The results will be used for the next two chapters, Concept and Methods and Materials.

The concept presented in Chapter 3 is based on a client server solution with the two modes enrol and test. The handwritten data of the passwords are collected with an app on smartphone or tablet, segmented and sent to the server. Not only online data are collected but also offline data, which is important for the tests with different parameters in Chapter 5. On the server, the parameters extraction and classifying takes place. The data are stored in a database on the server and the result of the identification is sent back to the client.

In Chapter 4 methods and materials the basic methods for writer identification and verification are presented. In addition to the files with coordinates, segment and time, the images of the signatures are also saved as graphical files, especially for the background noise of online and offline data. The databases are presented divided into publicly available and privately created or already existing and extended databases. The secure password DB-150, which was created in 2015 at the ULPGC for this thesis, and the secure password and signature DB-30 in 2016 was created with smartphone and tablet for this thesis at the BTU Cottbus - Senftenberg. Both databases formed the basis for important investigations with secure passwords as well as signatures and delivered as in chapter 5 evaluated good results. It should also be mentioned at this point the investigation of already existing BD-150 online handwriting database which as the investigation shows from a small group of people was created and thus only conditionally usable and so for the thesis must be developed new databases. Furthermore, in Chapter 4 are explained the most important development of this thesis, the features divided by online and offline features. The online features again subdivided according to geometric, statistical and temporal features. The special pressure features don't exist on tablet and smartphones that's why they are not considered at this thesis. Consciously, no pressure features were developed, since normal smartphones and tablets do not have pressure sensors so in this thesis, a possibility ought and found good results for the writer identification even without pressure. In total, 67 online features were developed and with the features Zernike Moments (20 up to order 15) and Fourier coefficients (30 up to order 100) additionally 50 offline features, which had not so good results in the investigations and therefore not used for the implemented system. Furthermore, the feature selection with the most important procedures for feature extraction we explained PCA and clustering and feature selection like Fisher Score and Info Gain Attribute Evaluation. Finally, the classifiers K-Nearest Neighbor, KStar, Naïve Bayes and Bayes Net are presented.

In Chapter 5 the experiments are presented and evaluated. A total of 6 experiments were carried out with different databases, different number of features and classifiers. In the first part of the chapter we worked with handwritten passwords and 10 mainly geometric features, and in the second part with secure passwords, signatures and texts from the developed databases and public databases. In the second part, the features were extended from 25 to 67 online features and for the test with the offline features to additional 50 features, where calculated by the Zernike Moments are the worst and the temporal features are the best. There are still development potentials for the temporal features as well as for the Fourier Coefficients, some of which were ranked according to the ranking with Fisher Score in the upper part of the ranking list. The k-Nearest Neighbor and Bayes Net classifiers performed best on average across all experiments, and results of up to 100% accuracy were obtained for the classification and the FAR can be reduced up to 3.13% for safe passwords and 2.47% for handwritten texts. The ranking with Fisher Score and Info Gain Attribute Select helped to improve the time for the classification due to the reduction of the features and, at the same time, to improve the results, and thus remained mostly in the hundredth of a millisecond range for the developed prototypes and later practical use of the system a great advantage for best results.

The hypothesis presented in this thesis, which leads to "The combination of handwriting and secure password leads in truly secure authentication", can be confirmed according to the results of the experiments presented in Chapter 5, supervised bachelor- and master thesis and publications [123][124][133][151][160]:

1. **Handwritten passwords are safer than keyboard-written passwords.**
In addition, the secure passwords, as well as the texts with single characters and the signatures lead with the developed features to high secure identification of the writer. Compared to the passwords and PIN written with the keyboard, this method offers more security by incorporating the biometric features such as generated from writing style and writing time. With the now developed features, the security increases even if the writer knows the password and the writer has observed when writing. This makes it nearly impossible to forge a password and get access to a such secured system.

2. **It's possible to identify writers using the handwritten password.**
The use of the secure password results in up to 100% accuracy and the FAR rate could be lowered up to 3.13% (secure password) or 2.47 (handwritten texts) for best result in the experiments.

3. **Mobile devices are suitable for handwritten password input.**
The secure passwords and signatures, signed by smartphone and tablet, prove to be safe despite the missing pressure features and can keep up with the state of the art results generated with pressure features. So it is possible to install this system on standard devices such as smartphones and tablets without sacrificing security. The test with tablet and pad has shown that the identification with both tablet and pad with secure password as well as signature supplies up to 100% accuracy supplies.

4. **The features for the handwriting identification can be applied both to a handwritten password and a signature.**
In addition, the secure passwords, as well as the texts with single characters and the signatures lead with the developed features to very secure identification of the writer. The last test in chapter 4 paragraph 4.2.6 with datasets with secure password and signature collected from 30 users with tablet and pad has shown that the developed features can be used for passwords as well as signatures with different devices and deliver equally good results.

5. **An impostor, with the knowledge of the password, is rejected on the basis of the biometric features of his handwriting.**

Compared to the passwords and PIN written with the keyboard, the applied methods in this thesis offers more security by incorporating the biometric features such as generated from writing style and writing time. With the now developed features, the security increases even if the writer knows the password and the writer has observed when writing. This makes it nearly impossible to forge a password and get access to a such secured system.

6.2 Future Work

This draft provides the basis for the identification and verification of handwritten passwords and can be further developed in the future. The use of the system for the protection of bank accounts, for payment systems, access control and for the protection of confidential data on mobile devices is conceivable.

A practical security system must be further improved, e.g. so, that the handwritten password is not visible on the display and the data should be encrypted as well as during the transmission to the server and on the smartphone during the preprocessing and on the server database.

In addition to the use in the technical area like security for bank account, a use in medicine is also thinkable, for example, to support medical diagnoses of diseases, which are accompanied by a change in the typeface of the patients (e.g. Parkinson’s disease). For a prototype for Parkinson’s disease analysis the feature extraction and classification algorithms of this thesis already were used.

References

[1] Persönliche Identifikationsnummer – Wikipedia. (n.d.). Retrieved April 07, 2017, from https://de.wikipedia.org/wiki/Pers%C3%B6nliche_Identifikationsnummer

[2] Personal Identification Number (PIN) Definition. (n.d.). Retrieved April 07, 2017, from https://www.thebalance.com/pin-number-definition-and-explanation-315344

[3] Hackers can steal your details from chip and PIN machines used in shops and restaurants Daily Mail Online. (n.d.). Retrieved April 07, 2017, from http://www.dailymail.co.uk/news/article-2180849/Hackers-steal-details-chip-PIN-machines-used-shops-restaurants.html

[4] Mobile Payment: QR-Code, NFC oder BLE? | E-Commerce Magazin. (n.d.). Retrieved April 05, 2017, from http://www.e-commerce-magazin.de/mobile-payment-qr-code-nfc-oder-ble

[5] Wengnoon, R., & Limpiyakorn, Y. (2014). Extension of Insurance Premium Payment to Mobile Application with QR Code. *2014 International Conference on Information Science & Applications (ICISA)*. doi:10.1109/icisa.2014.6847397

[6] Ma, T., Zhang, H., Qian, J., Hu, X., & Tian, Y. (2015). The Design and Implementation of an Innovative Mobile Payment System Based on QR Bar Code. *2015 International Conference on Network and Information Systems for Computers*. doi:10.1109/icnisc.2015.35

[7] Madhoun, N. E., & Pujolle, G. (2016). Security Enhancements in EMV Protocol for NFC Mobile Payment. *2016 IEEE Trustcom/BigDataSE/ISPA*. doi:10.1109/trustcom.2016.0289

[8] El Madhoun, N., & Pujolle, G. (2016). A secure cloud-based NFC payment architecture for small traders. *2016 3rd Smart Cloud Networks & Systems (SCNS)*. doi:10.1109/scns.2016.7870562

[9] Biometrie – Wikipedia. (n.d.). Retrieved April 06, 2017, from https://de.wikipedia.org/wiki/Biometrie

[10] Jain, A., Ross, A., & Pankanti, S. (2006). Biometrics: A Tool for Information Security. *IEEE Transactions on Information Forensics and Security*, *1*(2), 125-143. doi:10.1109/tifs.2006.873653

[11] Wijesoma, W. S., Yue, K. W., Chien, K. L., & Chow, T. K. (2001). Online Handwritten Signature Verification for Electronic Commerce over the Internet. *Web Intelligence: Research and Development*, 227-236. doi:10.1007/3-540-45490-x_27

[12] Yanushkevich, S. (2006). Synthetic Biometrics: A Survey. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. doi:10.1109/ijcnn.2006.246749

[13] Biometrie: biometrics: ITWissen.info. (n.d.). Retrieved April 04, 2017, from http://www.itwissen.info/Biometrie-biometrics.html

[14] Fingerabdruck – Wikipedia. (n.d.). Retrieved April 05, 2017, from https://de.wikipedia.org/wiki/Fingerabdruck

[15] Gesichtserkennung – Wikipedia. (n.d.). Retrieved April 06, 2017, from https://de.wikipedia.org/wiki/Gesichtserkennung

[16] Artec ID präsentiert biometrische Gesichtserkennung und Zugangskontrolle | Telematik.tv. (n.d.). Retrieved April 06, 2017, from http://telematik-markt.de/telematik/artec-id-pr%C3%A4sentiert-biometrische-gesichtserkennung-und-zugangskontrolle

[17] Iris-Erkennung – Wikipedia. (n.d.). Retrieved April 04, 2017, from https://de.wikipedia.org/wiki/Iris-Erkennung

[18] Iriserkennung als Zugangssystem bei den Olymischen Spielen. (n.d.). Retrieved April 04, 2017, from https://www.pressetext.com/news/20000914005

[19] Spracherkennung – Wikipedia. (n.d.). Retrieved April 04, 2017, from https://de.wikipedia.org/wiki/Spracherkennung

[20] Android 'Voice Access' Lets You Control a Phone by Speaking to it - Mobile App Reviews. (n.d.). Retrieved April 04 from http://uk.pcmag.com/apps/76670/news/android-voice-access-lets-you-control-a-phone-by-speaking-to

[21] (HINDI) Control Your Phone with Voice Google Voice Access. (2017, July 6). Retrieved from https://www.youtube.com/watch?v=duDaJrdvdKc

[22] Handschrifterkennung: HCR (handprint character recognition): ITWissen.info. (2017, April 6). Retrieved from http://www.itwissen.info/Handschrifterkennung-handprint-character-recognition-HCR.html

[23] Unterschriftserkennung: signature detection: ITWissen.info. (n.d.). Retrieved April 6, 2017, from http://www.itwissen.info/Unterschriftserkennung-signature-detection.html

[24] Die 25+ besten Ideen zu Digital Signature auf Pinterest | kleiner Business-Plan. (n.d.). Retrieved April 7, 2017, from https://de.pinterest.com/explore/digital-signature/

[25] DeitY to Launch Aadhaar-Based eSignature Service in One Month. (n.d.). Retrieved April 7, 2017, from http://www.iamwire.com/2015/09/deity-launch-aadhaar-based-esinature-service-month/122471

[26] Unterschrift – Wikipedia. (n.d.). Retrieved November 2, 2016, from https://de.wikipedia.org/wiki/Unterschrift

[27] Digitale Handschrift: Tipp-Biometrie verhindert Password-Sharing | sicherheit.info. (n.d.). Retrieved November 2, 2016, from http://www.sicherheit.info/artikel/2103819

[28] Per Tippverhalten anmelden ▪ LANline. (n.d.). Retrieved April 7, 2017, from http://www.lanline.de/tippverhalten-anmelden-html

[29] Behrens, M., & Roth, R. (2001). Grundlagen und Perspektiven der biometrischen Identifikation. *Biometrische Identifikation,* 8-26. doi:10.1007/978-3-322-90843-8_2

[30] Nutzer-Authentifizierung mittels handschriftlichen Passworten auf Mobil-Geräten, Kutzner Tobias, Bönninger Ingrid, Travieso Carlos, 12. Wissenschaftstage der Hochschule Lausitz (FH) – University of Applied Sciences, 2012

[31] School Mathematics App for Study Beginners – DEM12, Tobias Kutzner, Christian Steinert, Olga Wälder, Online Educa Berlin, 2013

[32] Best Practice Beispiele: Aus dem Logbuch eines Moodle-Administrators an Hochschulen und Universitäten., Tobias Kutzner, Heiko Lehmann, Vortrag: Moodle Mahara Moot Konferenz Universität Leipzig, 2014

[33] Studienvorbereitung mit eLearning Tools, Kutzner Tobias, Steinert Christian, Moderne Lehre Gestalten, Tagung 2014 Martin-Luther-Universität Halle-Wittenberg, 2014

[34] Unterstützung der Studienvorbereitung durch den Einsatz ausgewählter eLearning Tools, Kutzner Tobias, Steinert Christian, E-Learning Symposium 2014 Potsdam, Mobil und vernetzt - studieren im digitalen Zeitalter, 2014

[35] Dynamische E-Tests als Methode der Studienbegleitung im ersten Semester, Steinert Christian, Kutzner Tobias, E-Didaktik-Tagung Göttingen "Lehre auf neuen Wegen", 2015

[36] Study accompaniment with EAssessments, Christian Steinert, Tobias Kutzner, InnoEducaTIC 2015, Jornadas Iberoamaricanas de Innovación Educativa en el ámbito de las TIC Las Palmas de Gran Canaria, 12-13 de noviembre de 2015, 2015

[37] Best Practice: Streaming Server for educational videos at Universities, Kutzner Tobias, Steinert Christian, InnoEducaTIC 2015, Jornadas Iberoamaricanas de Innovación Educativa en el ámbito de las TIC Las Palmas de Gran Canaria, 12-13 de noviembre de 2015, 2015

[38] School mathematics app for study beginners, Kutzner Tobias, Steinert Christian, Book of abstracts, 21st global, cross-sector conference on technology supported learning and training Berlin, December 2 – 4, 2015

[39] Dynamische Self Assessments mit Moodle- Gegenüberstellung des Nutzens und Aufwands bei der Erstellung, Steinert Christian, Kutzner Tobias, Erfolgsfaktor (en im) Selbststudium: Diskursive Fachtagung TH Wildau, 2016

[40] Bilinguale mobile Anwendung (App) zur Auffrischung der Mathematikkenntnisse in der Studieneingangsphase, Tobias Kutzner, Christian Steinert, Olga Wälder, Grundfragen Multimedialen Lehrens und Lernens, Die offene Hochschule: Vernetztes Lehren und Lernen Tagungsband GML^2, 2016

[41] Intelligentes einbinden von Ergebnissen interaktiver Lernvideos in Lernmanagementsysteme, Steinert Christian, Kutzner Tobias, Die vermessen(d)e Bildung: Möglichkeiten und Risiken digital vernetzter Technologien, Junges Forum für Medien und Hochschulentwicklung Darmstadt, 10.- 11. Juni 2016, 2016

[42] Best Practice: Massive Online Courses at the BTU Cottbus-Senftenberg, Steinert Christian, Kutzner Tobias, Palekhov Dmitry, Leptien Eva, nnoEducaTIC 2016 - III Jornadas Iberoamericanas de Innovación Educativa en el ámbito de las TIC, Las Palmas de Gran Canaria, 17-18 de noviembre de 2016, 2016

[43] Mathematics App as Mobile Assessment beside LMS Assessment, Kutzner Tobias, Steinert Christian, InnoEducaTIC 2016 - III Jornadas Iberoamericanas de Innovación Educativa en el ámbito de las TIC, Las Palmas de Gran Canaria, 17-18 de noviembre de 2016, 2016

[44] Kriterienorientierte Untersuchung videobasierter E-Assessments, Steinert Christian, Kutzner Tobias, Falke Tobias, Teaching is Touching the Future & ePS 2016 - Kompetenzorientiertes Lehren, Lernen und Prüfen Universitätsverlag Webler Bielefeld, 2017

[45] Untersuchung der Möglichkeit eines biometrischen On-Pen Matching Tobias Scheidat · Claus Vielhauer Otto-von-Guericke Universität Magdeburg {tobias.scheidat claus.vielhauer}@iti.cs.uni-magdeburg.de, D A C H Security 2006·syssec (2006) 392-404

[46] Working Paper of the Institute for Innovation and Technology | Nr. 21, Institut für Innovation und Technik (iit) Steinplatz 1, 10623 Berlin, iit perspektive Nr. 21 September 2014

[47] Jakobsson, M., Taveau, S. (2012): The Case for Replacing Passwords with Biometrics. In: In Mobile Security Technologies (MoST), 2012

[48] Art /_Roth Thomas Petermann Arnold Sauter, Studie_Biometrische Identifikationssysteme Sachstandsbericht Arbeitsbericht Nr. 76 Büro für Technikfolgeabschätzung im Deutschen Bundestag (TAB), 2002

[49] Dittmann, J., Mayerhöfer A., Vielhauser C. (2001), Platanista GmbH: Biometrische Systeme FuE Diffusionstendenzen und Anwendungen. Im Auftrage des deutschen Bundestages, 2001

[50] Behrens, M., & Roth, R. (2001). Biometrische Identifikation aus Nutzerperspektive — empirische Befunde. *Biometrische Identifikation,* 195-220. doi:10.1007/978-3-322-90843-8_11

[51] Nolde V., Leger, L. (2002). Biometrische Verfahren, Körpermerkmale als Passwort, ISBN-13: 978-3871564642, 2002

[52] Maier, K. (1994). TeleTrusT-Security Application Programming Interface. *Sicherheitsschnittstellen — Konzepte, Anwendungen und Einsatzbeispiele,* 167-178. doi:10.1007/978-3-663-06728-3_13

[53] Vielhauer, C. (2000). Handschriftliche Authentifikation für digitale Wasserzeichenverfahren. *Sicherheit in Netzen und Medienströmen,* 134-148. doi:10.1007/978-3-642-58346-9_12

[54] ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, Tom Fawcett, Hewlett-Packard Company, 2003

[55] Kästner, G., Breitenstein, O., Scholz, R., & Reiche, M. (2002). *Journal of Materials Science: Materials in Electronics, 13*(10), 593-595. doi:10.1023/a:1020152215266

[56] Kalenova, D. (2003). Personal authentication using signature recognition. Department of Information Technology, Laboratory of Information Processing, Lappeenranta University of Technology, 2003

[57] Porwik, P., & Para, T. (2007). Some Handwritten Signature Parameters in Biometric Recognition Process. *2007 29th International Conference on Information Technology Interfaces*. doi:10.1109/iti.2007.4283767

[58] Fotak, T., Bača, M., & Koruga, P. (2011). Handwritten signature identification using basic concepts of graph theory. Wseas Transactions on Signal Processing, 7(4), 117-129., 2011

[59] Scharffenberger, G., Kempf, K. (2011) Person Authentication by Handwriting in air using a Biometric Smart Pen Device. In: Proc. of the international Conference of the Biometrics Special Interest Group(BIOSIG), Darmstadt, Germany.219-226, 2011

[60] Ballard, L., Lopresti, D., & Monrose, F. (2006). Evaluating the security of handwriting biometrics. In Tenth International Workshop on Frontiers in Handwriting Recognition., 2006

[61] Kumar, M. M., & Puhan, N. B. (2014). Offline signature verification using the trace transform. *2014 IEEE International Advance Computing Conference (IACC)*. doi:10.1109/iadcc.2014.6779473

[62] Angadi, S., & Gour, S. (2014). Euclidean Distance Based Offline Signature Recognition System Using Global and Local Wavelet Features. *2014 Fifth International Conference on Signal and Image Processing*. doi:10.1109/icsip.2014.19

[63] Jayadevan, R., Subbaraman, S., & Patil, P. M. (2007). Verification of hand printed signature images using discrete dyadic wavelet transform. *2007 International Conference on Industrial and Information Systems*. doi:10.1109/iciinfs.2007.4579199

[64] Vasquez, J. L., Travieso, C. M., & Alonso, J. B. (2013). Off line writer identification based on graphometric parameters. *2013 IEEE 17th International Conference on Intelligent Engineering Systems (INES)*. doi:10.1109/ines.2013.6632815

[65] Narayan, N. P., Bonde, S. V., & Doye, D. (2015). Offline signature verification using shape dissimilarities. *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*. doi:10.1109/iccict.2015.7045677

[66] Choudhary, S., Kaul, S., Mishra, S., & Arun J.B. (2015). Signature verification using Java - Python for small computational devices. *2015 IEEE International Advance Computing Conference (IACC)*. doi:10.1109/iadcc.2015.7154788

[67] Joshi, M. A., Goswami, M. M., & Adesara, H. H. (2015). Offline handwritten Signature Verification using low level stroke features. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. doi:10.1109/icacci.2015.7275778

[68] A Literature Review on Hand Written Character Recognition, Mansi Shah and Gordhan, B. Jethava, Indian Streams Research Journal Vol -3 , ISSUE –2, March.2013 ISSN:-2230-785, 2013

[69] Djeddi, C., Siddiqi, I., Al-Maadeed, S., Souici-Meslati, L., Gattal, A., & Ennaji, A. (2015). Signature Verification for Offline Skilled Forgeries Using Textural Features. *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. doi:10.1109/sitis.2015.40

[70] Sharma, V., & Singh, N. (2016). Valid Signature Detection and Verification for Security of Individual Person. *Circulation in Computer Science, 1*(1), 50-53. doi:10.22632/ccs-2016-251-26

[71] Offline Handwritten Signature Verification Literature Review, Luiz G.Hafemann, Robert Sabourin, Luiz S.Oliveira, arXiv:1507.07909v2 [cs.CV] 19 Aug 2015, 2015

[72] Pham, T., Le, H., & Do, N. (2014). Offline handwritten signature verification using local and global features. *Annals of Mathematics and Artificial Intelligence*, *75*(1-2), 231-247. doi:10.1007/s10472-014-9427-5

[73] Okawa, M., & Yoshida, K. (2015). Text and User Generic Model for Writer Verification Using Combined Pen Pressure Information From Ink Intensity and Indented Writing on Paper. *IEEE Transactions on Human-Machine Systems*, *45*(3), 339-349. doi:10.1109/thms.2014.2380828

[74] Rajib Lochan Das, Binod Kumar Prasad, Goutam Sanyal, "HMM based Offline Handwritten Writer Independent English Character Recognition using Global and Local Feature Extraction", International Journal of Computer Applications (0975 – 8887), Volume 46– No.10, pp. 45-50, May 2012

[75] Bhatia, N. and Vandana, "Survey of Nearest Neighbor Techniques," International Journal of Computer Science and Information Security, Vol. 8, No. 2, (2001), 302-305.

[76] Rajbala Tokas,Aruna Bhadu, "A comparative analysis of feature extraction techniques for handwritten character recognition", International Journal of Advanced Technology & Engineering Research, Volume 2, Issue 4, pp. 215-219, July 2012

[77] Amritha Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using backpropagation neural networks", International journal on Advanced computing, Vol. 3, No. 4, pp. 51 - 58, July 2012

[78] Tirtharaj Dash, "Time efficient approach to offline hand written character recognition using associative memory net.", International Journal of Computing and Business Research, Volume 3 Issue 3 September 2012

[79] Imaran Khan Pathan,Abdulbari Ahmed Bari Ahmed Ali, Ramteke R.J., "Recognition of offline handwritten isolated Urdu character ", International Journal on Advances in Computational Research, Vol. 4, Issue 1, pp. 117-121, 2012

[80] Ashutosh Aggarwal, Rajneesh Rani, RenuDhir, "Handwritten Devanagari Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277-128X), Vol. 2, Issue 5, pp. 8590, May 2012

[81] Velappa Ganapathy, Kok Leong Liew, "Handwritten Character Recognition Using Multi scale Neural Network Training Technique", World Academy of Science, Engineering and Technology, pp. 32-37, 2008

[82] T.Som, Sumit Saha, "Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering,Vol.5, No.2, pp. 11-15, Dec 2011

[83] Rakesh Kumar Mandal, N R Manna, "Hand Written English Character Recognition using Row- wise Segmentation Technique", International Symposium on Devices MEMS, Intelligent Systems & Communication, pp. 5-9, 2011.

[84] Farah Hanna Zawaideh, "Arabic Hand Written Character Recognition Using Modified Multi-Neural Network", Journal of Emerging Trends in Computing and Information Sciences (ISSN 2079-8407), Vol. 3, No. 7, pp. 1021-1026, July 2012, 2012

[85] J Pradeep, E Shrinivasan and S.Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), vol. 3, No 1, Feb 2011., 2011

[86] Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering (ISSN: 2231 - 2307), Vol. 2, Issue 2, pp. 504-508, May 2012, 2012

[87] Moncef Charfi, Monji Kherallah, Abdelkarim El Baati, Adel M. Alimi, "A New Approach for Arabic Handwritten Postal Addresses Recognition", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 3, pp. 1-7, 2012

[88] Fahmy, M. M. (2010). Online handwritten signature verification system based on DWT features extraction and neural network classification. *Ain Shams Engineering Journal, 1*(1)*,* 59-70. doi:10.1016/j.asej.2010.09.007

[89] Wibowo, C. P., Thumwarin, P., & Matsuura, T. (2014). On-line signature verification based on forward and backward variances of signature. *The 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE).* doi:10.1109/jictee.2014.6804069

[90] Statistical on-line signature verification using rotation-invariant dynamic descriptors, M. R. Nilchiyan, R. Bte Yusof, E. Alavi, 978-1-4799-7862-5 © 2015 IEEE, 2015

[91] Griechisch, E., Malik, M. I., & Liwicki, M. (2014). Online Signature Verification Based on Kolmogorov-Smirnov Distribution Distance. *2014 14th International Conference on Frontiers in Handwriting Recognition*. doi:10.1109/icfhr.2014.129

[92] Goncalves, R. P., Augusto, A. B., & Correia, M. E. (2015). Time/space based biometric handwritten signature verification. *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*. doi:10.1109/cisti.2015.7170483

[93] Gomez-Barrero, M., Galbally, J., Fierrez, J., Ortega-Garcia, J., & Plamondon, R. (2015). Enhanced on-line signature verification based on skilled forgery detection using Sigma-LogNormal Features. *2015 International Conference on Biometrics (ICB)*. doi:10.1109/icb.2015.7139065

[94] Lopez-Garcia, M., Ramos-Lara, R., Miguel-Hurtado, O., & Canto-Navarro, E. (2014). Embedded System for Biometric Online Signature Verification. *IEEE Transactions on Industrial Informatics*, *10*(1), 491-501. doi:10.1109/tii.2013.2269031

[95] Pirlo, G., Cuccovillo, V., Diaz-Cabrera, M., Impedovo, D., & Mignone, P. (2015). Multidomain Verification of Dynamic Signatures Using Local Stability Analysis. *IEEE Transactions on Human-Machine Systems*, *45*(6), 805-810. doi:10.1109/thms.2015.2443050

[96] Liu, Y., Yang, Z., & Yang, L. (2015). Online Signature Verification Based on DCT and Sparse Representation. *IEEE Transactions on Cybernetics*, *45*(11), 2498-2511. doi:10.1109/tcyb.2014.2375959

[97] Diaz, M., Fischer, A., Plamondon, R., & Ferrer, M. A. (2015). Towards an automatic on-line signature verifier using only one reference per signer. *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. doi:10.1109/icdar.2015.7333838

[98] Rosso, O. A., Ospina, R., & Frery, A. C. (2016). Classification and Verification of Handwritten Signatures with Time Causal Information Theory Quantifiers. *PLOS ONE, 11*(12), e0166868. doi: 10.1371/journal.pone.0166868

[99] Van Nguyen, T., Sae-Bae, N., & Memon, N. (2014). Finger-drawn pin authentication on touch devices. *2014 IEEE International Conference on Image Processing (ICIP)*. doi:10.1109/icip.2014.7026013

[100] Zheng, J., Gao, X., Zhan, E., & Huang, Z. (2008). Algorithm of On-Line Handwriting Signature Verification Based on Discrete Fréchet Distance. *Advances in Computation and Intelligence*, 461-469. doi:10.1007/978-3-540-92137-0_51

[101] Nakamura, Y., & Kidode, M. (2006). Online Writer Verification Using Kanji Handwriting. *Multimedia Content Representation, Classification and Security*, 207-214. doi:10.1007/11848035_29

[102] The Effectiveness of Generative Attacks on an Online Handwriting Biometric, Daniel P. Lopresti and Jarret D. Raim, T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 1090–1099, 2005. c _ Springer-Verlag Berlin Heidelberg 2005

[103] Freire, M. R., Fierrez, J., Galbally, J., & Ortega-Garcia, J. (n.d.). Biometric Hashing Based on Genetic Selection and Its Application to On-Line Signatures. *Advances in Biometrics*, 1134-1143. doi:10.1007/978-3-540-74549-5_118

[104] Schimke, S., & Vielhauer, C. (2007). Similarity searching for on-line handwritten documents. *Journal on Multimodal User Interfaces*, *1*(2), 49-54. doi:10.1007/bf02910058

[105] Pascual-Gaspar, J. M., Cardeñoso-Payo, V., & Vivaracho-Pascual, C. E. (2009). Practical On-Line Signature Verification. *Advances in Biometrics*, 1180-1189. doi:10.1007/978-3-642-01793-3_119

[106] A. K. Jain, R. P. W. Duin and J. Mao. Statistical pattern recognition: A review. IEEE Transaction on Pattern Analyses and Machine Intelligence, 22(1):4-37, 2000

[107] Impedovo, D., & Pirlo, G. (2008). Automatic Signature Verification: The State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(5), 609-635. doi:10.1109/tsmcc.2008.923866

[108] Dimauro,G.; Impedovo, S.;Pirlo,G.:“A stroke-oriented approach to signature verification”in From Pixels to Features III–Frontiers in Handwriting Recognition, S.Impedovo and J.C. Simon, Eds. Amsterdam,TheNetherlands:Elsevier,1992,pp.371-384

[109] Bovino, L., Impedovo, S., Pirlo, G., & Sarcinella, L. (n.d.). Multi-expert verification of hand-written signatures. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings*. doi:10.1109/icdar.2003.1227796

[110] Araujo,A. S. R.; Cavalcanti, C. D. G.; Filho, C. B. D. C. E.: “On-line verification for signatures of different sizes” presentedatthe10th Int. Workshop Front. Handwriting Recognition. (IWFHR10),La Baule, France,Oct.2006

[111] Muramatsu, D., Kondo, M., Sasaki, M., Tachibana, S., & Matsumoto, T. (2006). A Markov Chain Monte Carlo Algorithm for Bayesian Dynamic Signature Verification. *IEEE Transactions on Information Forensics and Security*, *1*(1), 22-34. doi:10.1109/tifs.2005.863507

[112] Feature selection - Wikipedia. (n.d.). Retrieved May 15, 2017, from https://en.wikipedia.org/wiki/Feature_selection

[113] Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., … Worek, W. (n.d.). Overview of the Face Recognition Grand Challenge. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. doi:10.1109/cvpr.2005.268

[114] Asadi, S., Rao, C., & Saikrishna, V. (2010). A Comparative study of Face Recognition with Principal Component Analysis and Cross-Correlation Technique. *International Journal of Computer Applications, 10*(8), 17-21. doi:10.5120/1502-2019

[115] Li, C., Diao, Y., Ma, H., & Li, Y. (2008). A Statistical PCA Method for Face Recognition. *2008 Second International Symposium on Intelligent Information Technology Application*. doi:10.1109/iita.2008.71

[116] M, N. (2012). A Comprehensive Overview of Clustering Algorithms in Pattern Recognition. *IOSR Journal of Computer Engineering*, *4*(6), 23-30. doi:10.9790/0661-0462330

[117] Raje, S., Mehrotra, K., & Belhe, S. (2015). Writer adaptation of online handwritten recognition using Adaptive RBF Network. *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. doi:10.1109/icdar.2015.7333850

[118] L. Rabiner und B. Juang. “An Introduction to Hidden Markov Models.” IEEE ASSP Magazine, Seiten 4–16, 1986.

[119] E. Schukat-Talamazzini. Automatische Spracherkennung - Grundlagen, statistische Modelle und effiziente Algorithmen. Vieweg, Braunschweig, 1995.

[120] S. Young, J. Jansen, J. Odell, D. Ollason und P. Woodland. The HTK Book. University of Cambridge, 2000.

[121] Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician, 46*(3)*,* 175. doi:10.2307/2685209

[122] k-nearest neighbors algorithm - Wikipedia. (n.d.). Retrieved May 16, 2017, from https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[123] Kutzner, T., Biometric Handwriting Password Recognition for Mobile Devices as Client-Server solution, Master's Thesis, Hochschule Lausitz [FH] - University of Applied Sciences (2012), 2012

[124] Wälder O., Kutzner T., Eine empirische Studie zur Verifikation von Unterschriften und zur Indikation von Fälschungen Olga Wälder und Tobias Kutzner Projektgruppe zu Blended Learning, Hochschule Lausitz (FH) AUSTRIAN JOURNAL OF STATISTICS Volume 42 (2013), Number 2, 101–116, 2013

[125] Huaigu Cao ; Prasad, R. ; Natarajan, P. ; OCR-Driven Writer Identification and Adaptation in an HMM Handwriting Recognition System, Document Analysis and Recognition (ICDAR), 2011 International Conference on, pp. 739-743, 2011.

[126] Santana, O.; Travieso, C.M.; Alonso, J.B.; Ferrer, M.A.; Writer identification based on graphology techniques, Aerospace and Electronic Systems Magazine, IEEE, pp. 35-42, 2010.

[127] Ming-Yen Tsai ; Leu-Shing Lan ; Online Writer Identification Using The Point Distribution Model, Systems, Man and Cybernetics, 2005 IEEE International Conference on, pp. 1264 - 1268, 2005.

[128] Miyao, H.; Maruyama, M.; Writer Adaptation for Online Handwriting Recognition System Using Virtual Examples, Document Analysis and Recognition, 2009. ICDAR '09. 10th International, 2009

[129] Dzido, Robert ; Gehrke, Martin; Steinke, Karl-Heinz: Erkennung von Schreibern mittels handgeschriebener Buchstaben, Writer recognition by handwritten characters, Tagungsband der Jahrestagung der deutschen Gesellschaft für Informatik, Lübeck September 2009., 2009

[130] Wang, D.; Zhu, B.; Nakagawa,M.: A Digital Ink Recogntion Server for Handwritten Japanese Text , Document Analysis and Recognition (ICDAR), 2011 International Conference on , pp 146-150, 2011.

[131] Yan Gao; Lanwen Jin; Cong He; Guibin Zhou: Handwriting Character Recognition as a Service: A New Handwriting Recognition System Based on Cloud Computing, Document Analysis and Recognition (ICDAR), 2011 International Conference on
pp. 885 - 889, 2011.

[132] Larry S.; Yaeger, Brandyn J.; Webb, Richard F. Lyon: Combining Neural Networks as Context-Driven Search for On-Line, Printed Handwriting Recognition in the Newton, Association for the Advancement of Artificial Intelligence (www.aaai.org), 2010

[133] Kutzner, T., Travieso, C. M., Bonninger, I., Alonso, J. B., & Vasquez, J. L. (2013). Writer identification on mobile device based on handwritten. *2013 47th International Carnahan Conference on Security Technology (ICCST).* doi:10.1109/ccst.2013.6922063

[134] Huaigu, C., Prasad, R., Natarajan, P.: OCR-Driven Writer Identification and Adaptation in an HMM Handwriting Recognition System, International Conference on Document Analysis and Recognition (ICDAR), pp. 739-743 (2011)

[135] Santana, O., Travieso, C.M., Alonso, J.B., Ferrer, M.A.: Writer identification based on graphology techniques, IEEE Aerospace and Electronic Systems Magazine, pp. 35-42 (2010)

[136] Gehrke, M., Steinke, K., Dzido, R.: Writer Recognition by Characters, Words and Sentences, IEEE International Carnahan Conference on Security Technology,pp. 281-288 (2009)

[137] Hassaine, A., Al-Maadeed, S., Alja'am, J.M., Jaoua, A., Bouridane, A.: The ICDAR2011, 2011

[138] Arabic Writer Identification Contest, International Conference on Document Analysis and Recognition (ICDAR), pp. 1470-1474 (2011), 2011

[139] Ming-Yen, T., Leu-Shing, L.: Online Writer Identification Using The Point Distribution Model, 2005 IEEE International Conference on Systems, Man and Cybernetics, pp. 1264 – 1268 (2005), 2005

[140] Miyao, H., Maruyama, M.: Writer Adaptation for Online Handwriting Recognition System Using Virtual Examples, 10th International Conference on Document Analysis and Recognition, pp. 1156 – 1160 (2009), 2009

[141] Sesa-Nogueras, E.: Discriminative power of online handwritten words for writer recognition, 2011 IEEE International Carnahan Conference on Security Technology (ICCST), pp. 1-8 (2011), 2011

[142] Tan, G.X., Viard-Gaudin, C., Kot, A.C.: Automatic Writer Identification Framework for Online Handwritten Documents Using Character Prototypes, Pattern Recognition, vol. 42, pp. 3313-3323 (2009), 2009

[143] Test Your Password - Free Secure Password Tester. (n.d.). Retrieved June 8, 2017, from http://testyourpassword.com

[144] CHAPRAN, J. (2006). BIOMETRIC WRITER IDENTIFICATION: FEATURE ANALYSIS AND CLASSIFICATION. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(04), 483-503. doi:10.1142/s0218001406004831

[145] Maged M.M. Fahmy, Online handwritten signature verification system based on DWT features extraction and neural network classification, doi.org/10.1016/j.asej.2010.09.007, 2010

[146] Arabic Writer Identification Contest, International Conference on Document Analysis and Recognition (ICDAR), pp. 1470-1474 (2011), 2011

[147] Ming-Yen, T., Leu-Shing, L.: Online Writer Identification Using The Point Distribution Model, 2005 IEEE International Conference on Systems, Man and Cybernetics, pp. 1264 – 1268 (2005), 2005

[148] Sesa-Nogueras, E.: Discriminative power of online handwritten words for writer recognition, 2011 IEEE International Carnahan Conference on Security Technology (ICCST), pp. 1-8 (2011), 2011

[149] Miyao, H., Maruyama, M.: Writer Adaptation for Online Handwriting Recognition System Using Virtual Examples, 10th International Conference on Document Analysis and Recognition, pp. 1156 – 1160 (2009), 2009

[150] Tan, G.X., Viard-Gaudin, C., Kot, A.C.: Automatic Writer Identification Framework for Online Handwritten Documents Using Character Prototypes, Pattern Recognition, vol. 42, pp. 3313-3323 (2009), 2009

[151] Kutzner, T., Fanyu Ye, Bonninger, I., Travieso, C., Malay Kishore Dutta, & Singh, A. (2015). User verification using safe handwritten passwords on smartphones. *2015 Eighth International Conference on Contemporary Computing (IC3).* doi:10.1109/ic3.2015.7346651

[152] Singh, A. ; Choubey, A. ; Bandaru, S. ; Mohan, L. ; Dhiman, M. , The ultimate signature identifier?, 2011 3rd International Conference on Electronics Computer Technology (ICECT), Vol. 2, pp.52-56, 2011.

[153] Fouad, M. ; Alsulaiman, F. ; El Saddik, A. ; Petriu, E., Revocable handwritten signatures with haptic information, IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE), pp. 108-11, 2011.

[154] Vajpai, J. ; Arun, J.B. ; Vajpai, I. , Dynamic signature verification for secure retrieval of classified information, Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1-4, 2013.

[155] Piotr Porwik, Tomasz Para. Some Handwritten Signature Parameters in Biometric Recognition Process, Information Technology Interfaces, 2007. ITI 2007. 29th International Conference, pp. 185-190, 2007.

[156] Yanyan Li ; Mengjun Xie ; Jiang Bian, USign – A security enhanced electronic consent model, 36th Annual International Conference of Engineering in Medicine and Biology, pp. 4487-4490, 2014.

[157] Eskander, G.S. ; Sabourin, R. ; Granger, E. , Signature based Fuzzy Vaults with Boosted Feature Selection, IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), pp. 131-138, 2011.

[158] Omri, F. ; Foufou, S. ; Hamila, R. ; Jarraya, M. , Cloud-based mobile system for biometrics

authentication, 13th International Conference on Telecommunications (ITST), pp. 325-330, 2013.

［159］ Jebriel, S. ; Poet, R. , Automatic registration of user drawn graphical passwords, 6th International Conference on Computer Science and Information Technology (CSIT), pp. 172-177, 2014.

［160］ Kutzner, T., Dietze, M., Bonninger, I., Travieso, C. M., Dutta, M. K., & Singh, A. (2016). Online handwriting verification with safe password and increasing number of features. *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*. doi:10.1109/spin.2016.7566777

［161］ Porwik, P.; Para, T., Some Handwritten Signature Parameters in Biometric Recognition Process, Information Technology Interfaces, 2007. ITI 2007. 29th International Conference, pp. 185-190, 2007.

［162］ Jain, R.; Doermann, D., Writer identification Using an Alphabet of Contour Gradient Descriptors, 12th International Confrerence on Document Analysos and Recognition, pp. 550-554, 2013.

［163］ Eskander, G.S. ; Sabourin, R. ; Granger, E. , Signature based Fuzzy Vaults with Boosted Feature Selection, IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), pp. 131-138, 2011.

［164］ Guo Xian Tan; Viard-Gaudin, C.; Kot, A.C., A stocastic nearest Neighbor character prototype appraoch for online writer identification, 19th International Conference on Pattern Recognition, pp. 1-4, 2008

［165］ Namboodiri, A.M.; Jain, A.K., Online handwitten script recognition, IEEE Transaction on Pattern Analysis and Machine Intelligance, pp. 124-130, 2004.

［166］ Jebriel, S. ; Poet, R. , Automatic registration of user drawn graphical passwords, 6th International Conference on Computer Science and Information Technology (CSIT), pp. 172-177, 2014.

[167] Shivram, A.; Ramaiah, C.; Govindaraju, V. , Data Sufficiency for Online Writer Identification: A Comparative Study of Writer Style Space vs. Feature Space Models, 22nd Internatonal Conference on Pattern Recognition, pp. 3122-3125, 2014.

[168] Shivram, A.; Ramaiah, C.; Govindaraju, V., A hierachical Bayesian approach to online writer identification, IET Biometrics, pp. 191-198, 2013.

[169] Graves, A.; Liwicki, M.; Fernandez, S.; Bertolami, R.; Bunke, H.; Schmidhuber, J., A Novel Connectionist System for Unconstrained Handwriting Recognition, IEEE Transaction on Pattern Analysis and Machine Intelligance, pp. 855-868, 2009.

[170] Martinez-Diaz, M.; Fierrez, J.; Galbally, J., The DooDB Graphical Password Database: Data Analysis and Benchmark Results, IEEE Jounals & Magazines, pp. 596-605, 2013.

[171] Omri, F. ; Foufou, S. ; Hamila, R. ; Jarraya, M. , Cloud-based mobile system for biometrics authentication, 13th International Conference on Telecommunications (ITST), pp. 325-330, 2013.

[172] Fouad, M. ; Alsulaiman, F. ; El Saddik, A. ; Petriu, E., Revocable handwritten signatures with haptic information, IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE), pp. 108-11, 2011.

[173] Singh, A. ; Choubey, A. ; Bandaru, S. ; Mohan, L. ; Dhiman, M. , The ultimate signature identifier?, 2011 3rd International Conference on Electronics Computer Technology (ICECT), Vol. 2, pp.52-56, 2011.

[174] Vajpai, J. ; Arun, J.B. ; Vajpai, I. , Dynamic signature verification for secure retrieval of classified information, Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1-4, 2013.

[175] Renuka, R.; Suganya, V.; Arun Kumar, B., Online Hand Written Character Recognition using Digital Pen for Static Authentication, 2014 International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 – 05, 2014, Coimbatore, INDIA, pp. 1-5, 2014.

[176] Yang, J.; Chang, W.; Bang W.C.; Choi E.S.; Kang, K.H.; Cho, S.J.; Kim, D.Y., Analysis and Compensation of errors in input device based inertial sensors, IEEE International Conference on Information Technology: Coding and Computing, pp. 790-796, 2004.

[177] Vahab Iranmanesh, Sharifah Mumtazah Syed Ahmad, Wan Azizun Wan Adnan, Salman Yussof, Olasombo Ayodeji Arigbabu, Fahad Layth Malallah, Online Handwritten Signature Verification Using Neural Network Classifier Based on Principal Component Analysis, The Scientific World Journal, 2014.

[178] Kutzner, T., Bonninger, I., Travieso, C. M., Dutta, M. K., & Singh, A. (2016). Study of long-term quality of online signature verification systems. *2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*. doi:10.1109/ccintels.2016.7878206

[179] Aging in Biometrics: An Experimental Analysis on On- Line Signature Javier Galbally*, Marcos Martinez-Diaz, Julian Fierrez Biometric Recognition Group-ATVS, Universidad Autonoma de Madrid, Madrid, Spain 2013.

[180] N. Bouadjenek, h. Nemmour, Y. Chibani, Age, Gender and Handedness Prediction from Handwriting using Gratient Features, 13th International Conference on Docment Analysis and Recognition (ICDAR), pp.116-1120, 2105.

[181] M. Erbilek, M. Fairhurst, M. Da Costa-Abreu, Improved age prediction from biometric data using multimodal configurations, Intern. Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-7, 2014.

[182] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobé, On the relevance of age in handwritten biometric recognition, IEEE International Carnahan Conference on Security Technology (ICCST), pp. 105-109, 2012.

[183] N. Sae-Bae, N. Memon, Online Signature Verification on Mobile Devices, *Fellow, IEEE 2014, pp. 933-974.,2014.*

[184] Aging in Biometrics: An Experimental Analysis on On-

Line Signature Javier Galbally*, Marcos Martinez-Diaz, Julian Fierrez Biometric Recognition Group-ATVS, Universidad Autonoma de Madrid, Madrid, Spain 2013.

[185] Liwicki, M. and Bunke, H.: IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard. 8th Intl. Conf. on Document Analysis and Recognition, 2005, Volume 2, pp. 956 – 961

[186] SVC2004: First International Signature Veri_cation Competition Dit-Yan Yeung1, Hong Chang1, Yimin Xiong1, Susan George2, Ramanujan Kashi3, Takashi Matsumoto4, and Gerhard Rigoll5 1 Hong Kong University of Science and Technology, Hong Kong
2 University of South Australia, Australia 3 Avaya Labs Research, USA 4 Waseda University, Japan 5 Munich University of Technology, Germany

[187] BIOMETRICS ON THE INTERNET
MCYT baseline corpus: a bimodal biometric database 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 IEE Proc.-Vis. Image Signal Process., Vol. 150, No. 6, December 2003

[188] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (n.d.). Retrieved June 7, 2017, from http://www.cs.waikato.ac.nz/ml/weka/

[189] The WEKA workbench. (2017). *Data Mining,* 553-571. doi:10.1016/b978-0-12-804291-5.00024-6

[190] Teague, M. R. (1980). Image analysis via the general theory of moments*. *Journal of the Optical Society of America, 70*(8), 920. doi:10.1364/josa.70.000920

Appendix A

A.1 Resumen

Capítulo 1:

En los últimos años, la importancia de la seguridad en las TI ha aumentado significativamente y se ha puesto en la mente de las personas las precauciones que se deben tomar para evitar un ataque cibernético. Hacer que los sistemas existentes sean más seguros es una tarea que continua en constante progreso. Para proteger las redes informáticas y las infraestructuras, la autenticación segura tiene la máxima prioridad, mientras que al mismo tiempo la facilidad de uso es cada vez más importante. El Capítulo 1 de esta tesis explica la motivación para el autor, y da una visión general de los sistemas de acceso actuales divididos en sistemas no biométricos y biométricos, así como su evaluación. Además, se muestra el trabajo realizado por el autor anteriormente. Por último, se presentará la hipótesis, que se confirmará por esta tesis, y la estructura de la tesis.

1.2 Tipos de acceso

1.2.1 Sistemas no biométricos

Hay diferentes maneras de conseguir acceso a los edificios y habitaciones. El principal tipo de acceso se puede dividir en dos tipos: los sistemas no biométricos y biométricos. A continuación, se da una explicación corta de los sistemas no biométricos más importantes, PIN, QR-, código de barras, NFC, y BLE.

a) PIN:

Un número de identificación personal (PIN) o número secreto, es un número conocido generalmente por una persona. Con esta identificación pueden autenticarse en una máquina y las contraseñas son del tipo numérico [1].

Figura 1.1: introducir el PIN del Banco

b) QR- y soluciones basadas en códigos de barras:

Los códigos QR y códigos de barras son versátiles y pueden ser usados tanto online como offline. En [6] se muestra un sistema de pago móvil basado en dos dimensiones (2D) usando códigos de barras y se analiza la arquitectura del sistema, diseño e implementación, así como soluciones basadas en códigos de barras QR 2D.

c) Soluciones Basadas en NFC:

NFC hace posible el intercambio de información sin necesidad de tener contacto, basta con unos pocos centímetros y tiene varias aplicaciones. Por ejemplo, el chip puede estar ubicado en una tarjeta, dentro de un teléfono inteligente o mediante una etiqueta fijada al teléfono. La ventaja sobre la tarjeta “normal” es que los datos se intercambian por radio. Esto elimina la necesidad de conectar la tarjeta en un terminal.

c) Soluciones basadas en BLE:

La última tecnología para realizar el pago en un negocio es la comunicación del cliente con el operador a través de la tecnología inalámbrica BLE. En comparación con NFC, BLE tiene un mayor alcance, de unos 10 metros. Los teléfonos inteligentes dentro del radio de un transmisor BLE, también llamados "Balizas", pueden ser identificados y utilizados para el pago. En la primera mitad del año 2017, Paypal tiene previsto cooperar con los minoristas grandes y algunos más pequeños para probar el servicio de pago de Beacon en el marcado alemán.

1.2.2 Sistemas biométricos

Dependiendo del área de aplicación, existen diferentes definiciones detalladas. En 1841, Christoph Bernoulli fue uno de los primeros científicos que utilizó el término biometría en una interpretación muy literal para la medición y la evaluación estadística de la vida humana [9].
Dividimos los métodos biométricos en rasgos físicos como la cara, huella digital, y el iris. Son únicos para cada individuo y son estables durante un período prolongado de tiempo. Por lo tanto, los sistemas biométricos, que se basan en estos rasgos, son por lo general precisos y lo suficientemente confiables para fines de identificación que implican una o más comparaciones [10] - [11].
A continuación, se muestran los métodos de verificación más utilizados en la actualidad:

a) Huella dactilar:

Los fabricantes de sistemas biométricos también utilizan la huella digital, que se lee en su mayoría ópticamente o eléctricamente (por ejemplo, capacitivamente), para identificar a los usuarios no autorizados. Con el fin de evitar el acceso no deseado con huellas dactilares que no son del usuario, sensores de temperatura y del pulso se pueden integrar en los dispositivos de detección para comprobar si un dedo vivo ha sido colocado en el dispositivo.

La seguridad de los sistemas de huellas digitales es relativamente pequeña, ya que una huella digital es fácil de reproducir. Los sensores de huellas digitales instalados en los dispositivos móviles ofrecen comodidad en comparación a la introducción de la contraseña o patrón de desbloqueo. El pirata informático de hardware Starbug fue capaz de superar el mecanismo de biométrico en 2014 tan sólo unos días después de la aparición del iPhone 5s, el primer dispositivo de Apple con touch ID [14].

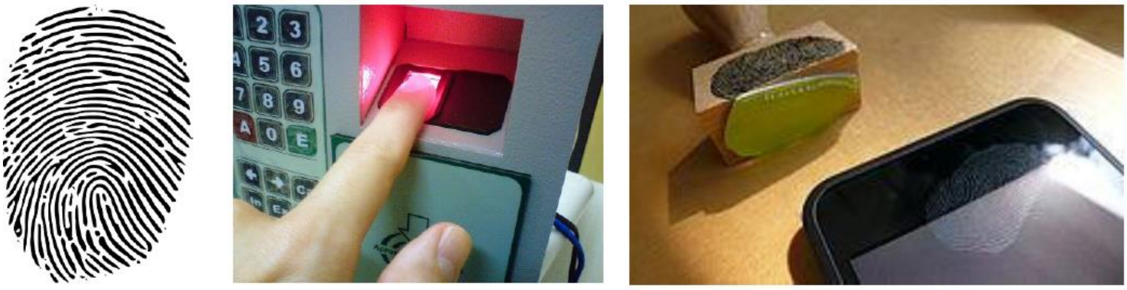


Figura 1.3: Huella digital, Sensor, Sello falsificador de huella para iphone [14]

b) Cara:

El reconocimiento facial es el análisis de la aparición de características visibles en la zona de la cabeza frontal dada por las propiedades geométricas y la textura de la superficie. Para esto se pueden utilizar dos métodos, 2D y 3D.

c) Iris:

El reconocimiento del iris es un método biométrico con el propósito de autenticación o identificación de personas. Para este fin, las imágenes del iris del ojo se graban con cámaras especiales, los rasgos característicos de los respectivos iris se identifican con métodos algorítmicos, convertidas en un conjunto de valores numéricos y se almacenan para el reconocimiento con una o varias plantillas que se han guardado.

El concepto original de la utilización de imágenes del iris para el reconocimiento biométrico fue desarrollado y patentado por Flom y Safir en 1987. La expiración de la patente en el año 2006 ha dado lugar a un mayor esfuerzo de investigación en todo el mundo [17].

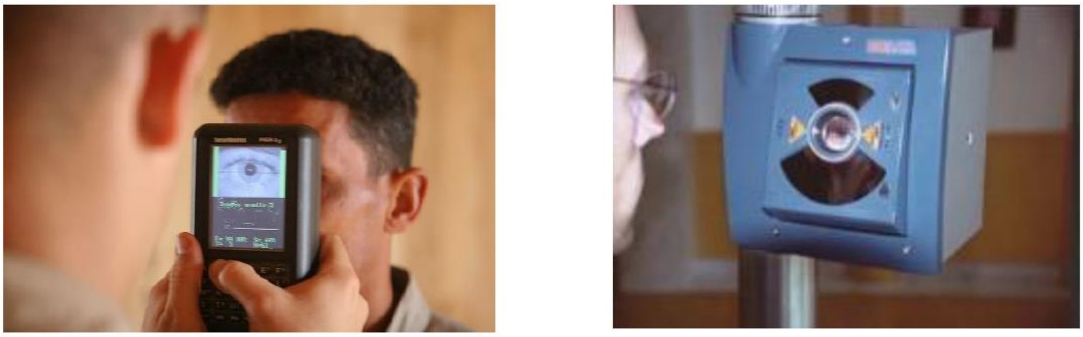


Figura 1.5: detección de Iris y autenticación [17] [18]

d) Voz:

El reconocimiento de voz o también reconocimiento automático del habla es una subárea de la informática aplicada, ciencias de la ingeniería y lingüística informática.
En la actualidad, aproximadamente se pueden distinguir dos tipos de reconocimiento de voz:

- reconocimiento de voz independiente del hablante
- reconocimiento de voz dependiente del locutor

Una característica para el reconocimiento de voz independiente del hablante es la propiedad que el usuario tiene para iniciar de inmediato el reconocimiento de voz sin una fase de formación previa. El vocabulario, sin embargo, se limita a unos pocos miles de palabras [19]. En cambio, los reconocimientos de voz dependientes del hablante son entrenados por el usuario antes de su uso (en sistemas más nuevos: durante el uso) en sus propias peculiaridades de pronunciación. Un elemento central es la interacción individual con el sistema a fin de lograr un resultado óptimo dependiente del hablante (términos propios, abreviaturas, etc.). Por tanto, no es útil para aplicaciones con los usuarios que cambian con frecuencia (por ejemplo, centros de llamadas).

e) Firma:

El término "reconocimiento de escritura" se refiere a todos los procedimientos que reconocen automáticamente las letras escritas a mano, números, palabras o frases y las transforman en un archivo para ser procesado por el ordenador. El reconocimiento de escritura se divide en off-line y online.

La tabla (Tabla 1.1) muestra los diversos procedimientos de verificación con los sensores usados y las ventajas y desventajas de estos métodos:

Tabla 1.1: Tipos de reconocimiento biométricos y su seguridad [13]

características biométricas	**Sensores**	**Debilidades**
Huella dactilar Geometría de la mano	Chip sensor Los escáneres ópticos	Los dedos sucios o heridos Enfermedades
Sonido / detección de voz	Micrófono	Ruido de fondo y cambios en la voz relacionados a enfermedades
Reconocimiento facial reconocimiento Retina reconocimiento del iris	Cámara Cámara especial láser infrarrojo	Vestimenta y clima Enfermedades o problemas en el ojo
Firma la dinámica del teclado	Tablet, teléfono inteligente	cambios mentales y las enfermedades relacionadas
secuencia de movimientos	Cámara	

1.2.3 Combinación de sistemas biométricos y no biométricos.

Este enfoque adopta la letra como base, esta se ha utilizado como un método de autenticación segura desde principios de la Edad Media. Estas firmas han sido encontradas en numerosos documentos, tales como el documento Ostarrîchi del emperador Otto III. de 996. Mientras que en Europa desde el comienzo de los tiempos modernos la firma manuscrita en presencia de testigos se considera como legalmente vinculante, el sello estampado (sello chino 印, Yin, Hanko japonesa 判 子) sigue siendo la firma jurídicamente válida en el círculo cultural de Asia Oriental. Sellos de firma son también común en otros países o instituciones [26]. Sin embargo, la firma por sí sola no es 100% a prueba de falsificaciones; con el fin de lograr un plus de seguridad surge la idea de utilizar una contraseña segura escrita a mano.

1.3 Trabajo relacionado.

En este párrafo se describe el trabajo relacionado del autor que lo motivó a tratar con el análisis de las contraseñas escritas a mano.

Antes del comienzo de esta obra en el año 2012 el autor desarrolló una aplicación cliente-servidor para una primera autentificación de usuario sencilla con contraseñas escritas a mano en dispositivos móviles [60]. (Véase la Figura 1.8) El prototipo implementado en [60] ejecuta la segmentación en el dispositivo móvil. La extracción de características y clasificación se ejecutan como aplicaciones del servidor. La herramienta WEKA 4 se utilizó para determinar el algoritmo más adecuado para el reconocimiento de la frase de la contraseña. Los análisis se llevaron a cabo con 280 contraseñas escritas a mano. De éstos, 176 son originales y 104 falsificaciones. En cuanto a los dispositivos móviles, se utilizaron un HTC con Android 2.2 y un Samsung Galaxy Ace con Android 2.3.3. El sistema propuesto reconoce al escritor de la contraseña, con una probabilidad del 96,87%, la falsa aceptación (FAR) fue de 12,5% cuando se utilizan ocho letras para la contraseña de cada escritor para el modelo de clasificación.

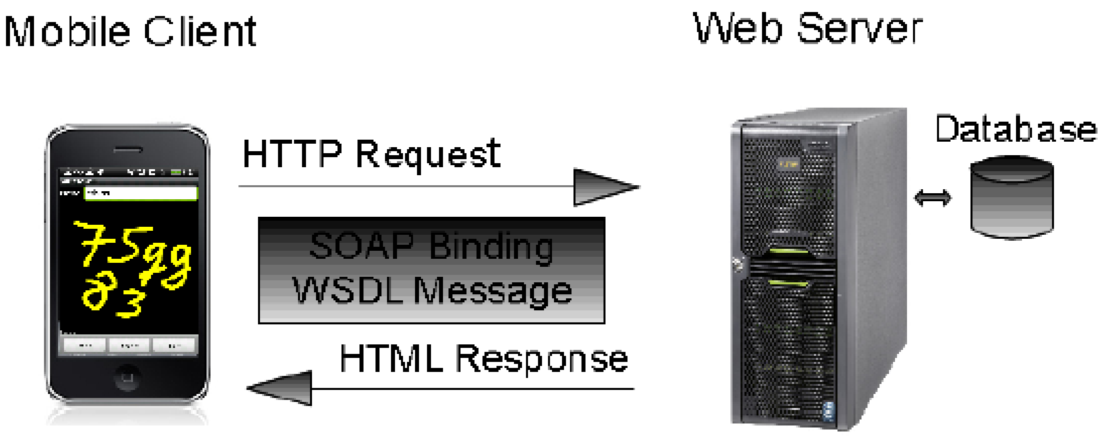


Figura 1.8: Reconocimiento de escritura como una solución cliente-servidor [127]

Los sistemas propuestos tienen dos modos: el modo Enrol para la recogida de datos y la construcción del modelo. El modelo de prueba para la verificación (mirar Figura 1.9)

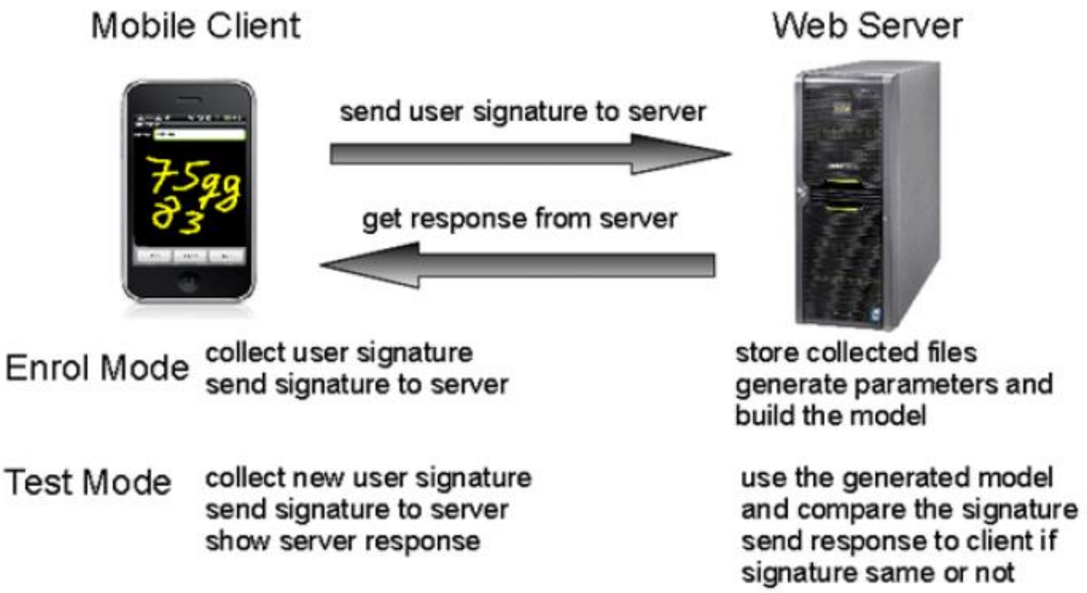


Figura 1.9: Modo Enrol y el modo de prueba para el reconocimiento de escritura [127]

1.4 Hipótesis

La combinación de escritura a mano junto con una contraseña es una opción realmente segura, aporta un plus de seguridad, se ha demostrado científicamente y desde este enfoque se desarrolló un sistema para la autenticación segura.

Los siguientes puntos se investigan y se referencian en esta tesis:

1. Contraseñas escritas a mano son más seguras que las contraseñas de teclado.
2. Es posible identificar los escritores que utilizan la contraseña escrita a mano.
3. Los dispositivos móviles son adecuados para la introducción de la contraseña escrita a mano.
4. Las características para la identificación de escritura a mano se pueden aplicar tanto a una contraseña escrita y una firma.
5. Un impostor, con el conocimiento de la contraseña, es rechazado sobre la base de las características biométricas de su puño y letra.

1.5 Estructura del documento

Esta tesis se centra en la identificación del escritor con una contraseña escrita a mano. Para ello, los métodos básicos, términos y procedimientos se aclaran y se describen en el estado del arte del análisis de escritura a mano, además, se analizan las bases de datos públicas usadas en el capítulo 2. En el capítulo 3, se describe el concepto de la idea de todo el sistema. En el capítulo 4 se muestran los métodos para la recolección de datos, procesamiento previo, extracción de características, la clasificación, y materiales como bases de datos utilizadas, los procesos de reducción de los parámetros, y la clasificación utilizada. Los experimentos realizados se explican. En el capítulo 5 se presentan los resultados de los experimentos, evaluados y discutidos. El último capítulo, el capítulo 6 resume el trabajo con las conclusiones y futuras investigaciones.

Capítulo 2: Estado del arte

2.1 metodología de revisión

2.1.1 Biometría: conceptualidad y demarcación

El término biometría se deriva de los términos griegos "bios" para "la vida", así como "Metron" de "medida". por lo tanto, Biometría hace referencia a los métodos que se basan en una medición de características humanas únicas y un reconocimiento inequívoco de un individuo [46].

Los datos biométricos en el sentido más estricto se refieren a unos registros y evaluaciones de las características individuales en los que el rendimiento del reconocimiento se toma con las tecnologías de un ordenador. Propiedades biológicas y / o anatómicas (tales como, por ejemplo, las estructuras de una cara, un iris o un dedo) y propiedades de comportamiento dependiente (por ejemplo, el comportamiento de escribir, la voz, el modo de andar o de escritura a mano) podrían también ser utilizados o evaluados por el reconocimiento biométrico.

2.1.2 Biometría- Verificación e identificación: Características, Métodos y Sistemas

La **verificación biométrica,** es decir, La confirmación de la identidad de la persona (1:1 = la persona medida es en realidad el que afirma que es), y - la **identificación biométrica,** es decir, el reconocimiento de un individuo a partir de un conjunto (definido) de las personas biométricamente registradas (1:n = la persona medida es XY). (Véase la Figura 2.1) En el caso de la verificación, los datos de medición actual se comparan con los datos existentes de la persona, por ejemplo, en una tarjeta de chip o una (Personal Digital Assistant) PDA, o también se pueden almacenar centralmente, con la identificación de un usuario predefinido de forma descentralizada (en posesión de la persona). Dentro del alcance de la inscripción, por ejemplo, en el caso de una aplicación a gran escala de máquinas bancarias - a menudo será necesario almacenar temporalmente los datos biométricos en una ubicación central adicional con el fin de cargarla en la tarjeta de chip para el usuario correspondiente ("personalización de las tarjetas del chip"). La ventaja desde el punto de vista de la protección de datos es que la plantilla se coloca en la autoridad del usuario en el caso de una verificación descentralizada es que las desventajas de seguridad y los posibles problemas asociados de adhesión son contrarrestadas por pérdidas o daños [49]. En el caso de la identificación, el sistema biométrico compara los datos medidos con los datos previamente almacenados centralmente y comprueba qué mejor plantilla coincide con el del usuario actual. Esto resulta en requisitos más altos en relación con el tamaño de la base de datos y en tiempo de identificación. Este tipo de detección biométrica se utiliza actualmente en áreas de alta seguridad con un pequeño número de usuarios o para investigaciones de la policía [49].

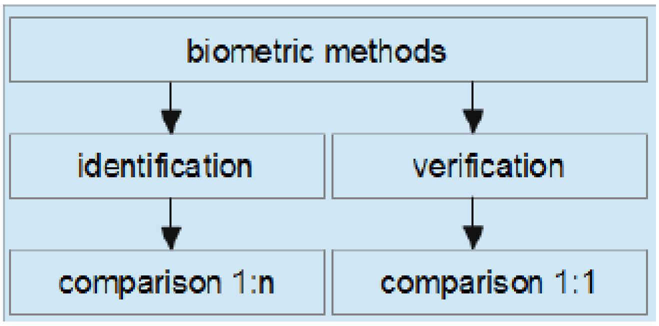


Figura 2.1: Identificación / verificación [46]

Durante la **identificación,** un valor de medición actualmente recogido se compara con muchos valores de referencia previamente almacenados. Los ejemplos conocidos de identificación biométrica son el uso de huellas digitales en el contexto de las investigaciones penales o de video vigilancia de los espacios públicos. Por ejemplo, las caras de los transeúntes podrían coincidir con los de delincuentes conocidos, que se almacenan en bases de datos centrales [46]. El proceso de comparación, que tiene lugar con los métodos biométricos, no es absolutamente exacto, pero en una comparación fuzzy (véase la figura 2.3), en contraste con los métodos conocidos. Así como diferentes firmas de la misma persona no está de acuerdo completamente, regularmente hay diferencias en múltiples mediciones de una característica biométrica. Así como al personal de seguridad humana en el aeropuerto, un procedimiento biométrico automatizado tiene un cierto efecto en la comparación del valor medido actual y el valor de referencia almacenado. Mientras que a una persona se compara el aspecto actual con el pasaporte y también estima el tamaño del cuerpo y evalúa el color de los ojos, una comparación técnica se basa en un proceso de reconocimiento de patrones formalizado [46].

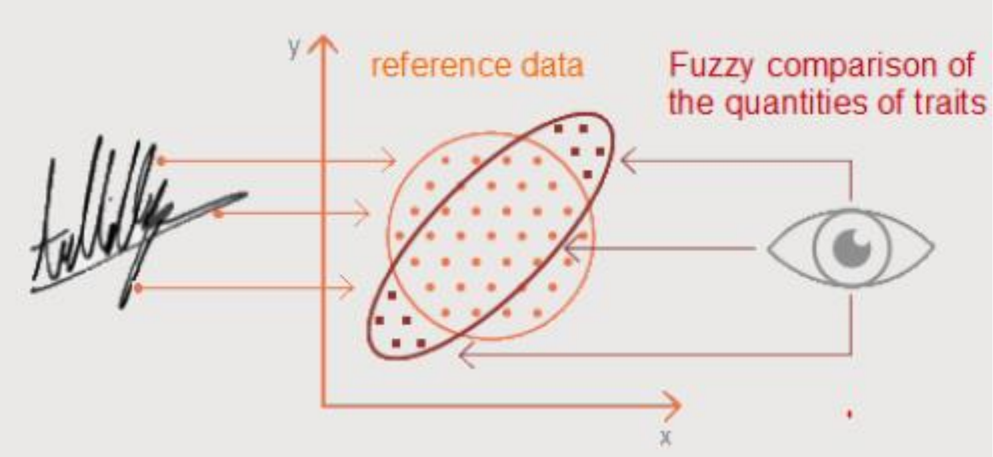


Figura 2.3: Comparación Fuzzy, diferentes modalidades [46]

2.1.3 Inscripción, plantilla, identificación y verificación

La base de cada procedimiento biométrico es la llamada **inscripción,** con independencia de la función que se utiliza y la técnica aplicada. Incluye la medición por primera vez y la medición de la característica biométrica de los futuros usuarios, la conversión de los "datos en bruto" en un registro de datos de referencia y el almacenamiento del mismo, la llamada plantilla. Esto representa el valor de comparación con las que los nuevos datos de medición (por lo menos a un alto grado) deben coincidir en todos los controles biométricos posteriores con el fin de ser capaz de identificar al usuario. Este proceso básico de inscripción requiere, por lo tanto, los más altos requisitos técnicos (con respecto a la sensibilidad y precisión, a fin de producir, sino también los conjuntos de datos reproducibles individuales), así como los requisitos de seguridad más altos. El objetivo principal de la aplicación de métodos biométricos, es decir, el aumento de la seguridad de un sistema general (por ejemplo, la dispensación de efectivo en una máquina) sólo puede lograrse si el registro de datos de referencia, la plantilla, se puede almacenar protegida de forma permanente. En particular, con respecto a las futuras implementaciones a gran escala, muchas preguntas siguen abiertas (por ejemplo, cuáles y cuántas personas deben estar debidamente calificadas para llevar a cabo la inscripción), la solución de los cuales es probable que implique un alto esfuerzo económico y de organización [49].

2.1.4 Medida o calidad biométrica

Los problemas en la práctica son: la aceptación incorrecta y falso rechazo.

Idealmente, cualquier conjunto de datos biométricos sería único para un individuo humano y sin ambigüedades asignado a él - los datos de referencia recogidos originalmente (plantilla) y conjunto de datos medidos respectivamente serían idénticos. En la práctica, existen limitaciones del caso ideal por varias razones:

Cada proceso de medición es una potente reducción de la información. Por razones de capacidad, los datos recogidos deben ser limitados. Además, tenemos el límite respectivo de medición (sensibilidad) y la precisión del sensor o del sistema en general, así como el "ruido" que no puede ser evitado. La cantidad de datos a ser almacenados en la plantilla debe minimizarse por razones técnicas (tamaño de la memoria, las tasas de transmisión), pero la precisión se reduce. Las características de comportamiento siempre muestran una variación mayor o menor debido a la naturaleza de las habilidades motoras humanas. Sin embargo, incluso las características fisiológicas sólo están limitadas en el tiempo. Pueden ser temporal o permanentemente alterados por procesos de envejecimiento, enfermedades o lesiones. por lo tanto, ligeros cambios deben ser toleradas por el sistema en "activo", así como los métodos "pasivos". Además, también tenemos perturbaciones ambientales durante la medición, por ejemplo, diferentes condiciones de luz o cambios de temperatura que pueden afectar el rendimiento de los sensores [49]. El sistema biométrico lleva a cabo una comparación estadística de los conjuntos de datos de la plantilla y la medición. El resultado especifica un valor porcentual del emparejamiento. Por esto, nunca ocurrirá una coincidencia del cien por cien para el sistema. Así, para cada sistema, existirá un **límite (**un valor del grado de coincidencia de la referencia y el valor medido) debe ser definido (por ejemplo, 95%) a partir del cual se considera la identificación o verificación y el usuario se acepta. Este umbral de tolerancia tiene un gran impacto en cuántos usuarios están erróneamente aceptados o bien cuántos son erróneamente rechazados (o se ven obligados a repetir el proceso varias veces).

Las tasas de rechazo falso o falsa aceptación de un sistema biométrico (**FRR** = Tasa de falso rechazo);

FAR = (Tasa de aceptación falsa) no se puede calcular teóricamente, sino que debe ser determinada empíricamente. FAR y FRR se ven afectados de tal manera que una disminución de falsa aceptación aumenta falso rechazo y viceversa. La magnitud absoluta de las tasas de error, sin embargo, depende de la sensibilidad y la precisión, es decir, la dificultad de todo el sistema y por lo tanto se determina por lo mencionado anteriormente (umbral de tolerancia). Sistemas menos precisos, como el reconocimiento de voz o bien se aceptan muchos usuarios de forma incorrecta (si el umbral de tolerancia es bajo) o falsamente los rechazan. El escaneo del iris, por el contrario, tiene un bajo FAR, así como una baja FRR debido a su alta nitidez de separación. Dependiendo de la aplicación práctica, la tasa de falso rechazo debe reducirse al mínimo (por lo general por razones de comodidad), o la tasa de falsa aceptación (especialmente con respecto a un aumento de la seguridad) mediante la selección del umbral de tolerancia del sistema. FAR y FRR se consideran que son los parámetros más importantes para la realización de un sistema biométrico. Si los valores son los mismos, se habla de " **EER "="** Equal Error Rate". Otro parámetro que ha sido raramente tratado en escenarios de prueba hasta ahora, pero que es muy importante en la práctica, es la tasa de falsos registros (**FER** = False Enrolment Rate), que puede tener un gran impacto en la aceptación de los usuarios [53].

2.1.5 Firma y reconocimiento de escritura

En el caso del reconocimiento de firma o del reconocimiento de escritura, no sólo el aspecto óptico de la firma (las letras como un "parámetro off-line") es decisiva, pero también cuentan la presión, la velocidad, la aceleración, etc. La escritura a mano hoy en día, por lo general, se toma con una tableta comercial o PDA o pantalla táctil. Una extensión del análisis de firmas es proporcionada por un sistema de escritura a mano en el que no sólo la firma, la llamada "semántica" [53], se utiliza para la autenticación de la escritura a mano. Esto puede incluir palabras predefinidas, frases enteras o incluso pequeños dibujos. (Véase la Figura 2.6) Una ventaja de la utilización de la "semántica" se encuentra en el anonimato, incluso con conjuntos de datos almacenados centralmente estos pueden ser preservados. Los métodos se pueden configurar de una manera tal que pueden ser controlados por el usuario, cambiando el registro de datos de referencia almacenado de manera relativamente rápida. La "capacidad de control" por el usuario también debe ser conducente a la aceptación. Puesto que la detección de los parámetros dinámicos es una detección de vida, la seguridad contra la falsificación es bastante alta [55]. Debido a las las altas tasas de error, sin embargo, los sistemas hasta ahora han sido muy limitados [49].

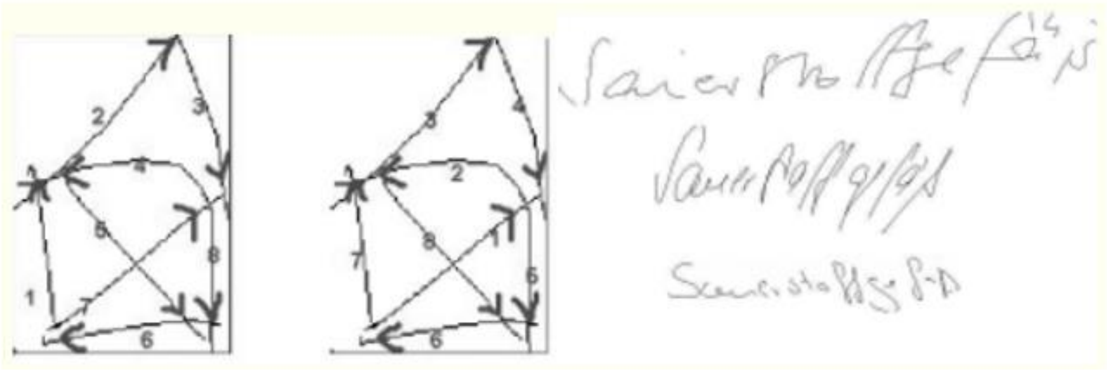


Figura 2.6: Escritura manual online “Haus von Nikolaus” y la firma dada [53]

2.1.6 Identificación del escritor y verificación

La autenticación de usuarios es el proceso mediante el cual se comprueba la identidad de una persona. Para esta revisión, se requieren dos tipos de datos, los datos de referencia y los datos cuya autenticidad va a ser detectada. La autenticación se considera que es exitosa si ambos datos son lo suficientemente parejos. En la autenticación, dos tipos de confirmación de la identidad pueden ser distinguidos. La identificación es la comparación de los datos de referencia almacenados con los datos de autenticación. La persona se considera identificado si los datos de referencia se pueden encontrar y estos corresponden con los datos mostrados actualmente. La verificación hace referencia a cuando el sistema verifica si los datos de autenticación autenticada coinciden con los datos de referencia de una identidad declarada (por ejemplo, nombre de usuario) dentro de una gama de variaciones de [45].

Figura 2.13 muestra el proceso general de verificación biométrica. La base para la verificación son datos de referencia almacenados en el sistema o en la tarjeta inteligente o la pluma. El proceso para la generación de estos datos se hace referencia en la biometría como adquisición de datos de referencia (inscripción). Los datos biológicos son recogidos por el sensor y se extraen las características. Si los datos son de suficiente calidad y cantidad (la mayoría de los sistemas requieren varios datos de autenticación para generar los datos de referencia), un registro de datos de referencia puede ser generado por el sistema que luego se almacena. En la primera etapa de la verificación, la propiedad biométrica detectada (los datos de autenticación actuales) mediante el sensor y se pasa al extractor de características después de un posible procesamiento previo (no incluido en la figura 5, ya que no forma parte de este sistema). Extrae las características requeridas de los datos de entrada y crea un vector de características que representa la función dentro del sistema. Este vector se compara en un proceso de comparación con el vector de características de los datos de referencia de la identidad que se reclama. El valor de la similitud representa la medida de la similitud de los dos vectores respecto al otro. Este es el valor de entrada para la clasificación que decide sobre el resultado de la verificación. Si el sistema determina que el usuario es el que se pretende, se devuelve true, false en caso contrario [45].

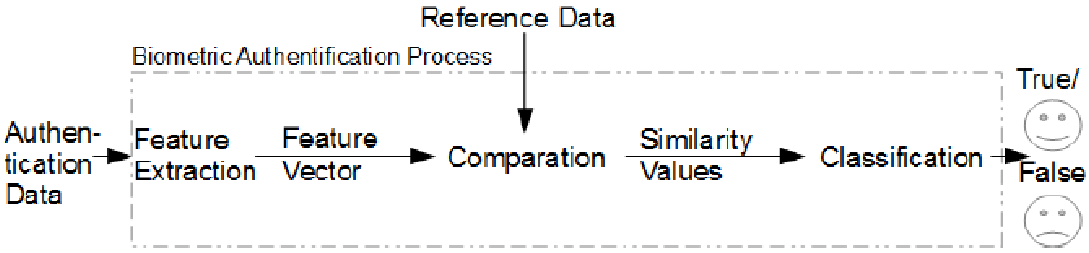


Figura 2.13: Representación esquemática de un proceso de verificación biométrica [45]

2.1.7 Identificación de escritura offline y verificación

En [58] se sugirió la creación de un gráfico a partir de las muestras de la escritura, de este gráfico se tomaron diversos datos como una conexión entre los segmentos, los vértices adyacentes, etc. Con esto se obtuvo una precisión de identificación del 94%.

En máquinas potentes que trabajan con sistemas offline [129], [130] [133] tienen una tasa de éxito en cuanto a la identificación de los escritores de su puño y letra de 92-99%. Estos utilizan como parámetros, la correlación de las mismas palabras [130] o formas estáticas y dinámicas [133]. La tasa de reconocimiento se eleva considerablemente, si se utilizan no sólo caracteres para la identificación (40%), sino también palabras o textos escritos a mano (hasta 99%) [133].

2.1.8 Identificación de escritura online y verificación

Los sistemas de reconocimiento de escritura en línea utilizan caracteres, que se escriben sobre una superficie sensible al tacto [142] - [144]. Ellos interpretan los movimientos y el estilo de la escritura. Para la identificación de los escritores, un modelo de distribución de puntos se utiliza en [142]. Usando para ello una base de datos de 50 palabras escritas a mano por 12 personas, el sistema alcanza una tasa de identificación de 97,3%. En [143], la imagen del personaje se divide en bloques de 8x8 con cuatro direcciones (vertical, horizontal, inclinación a la izquierda, y la inclinación derecha) para cada bloque (256 características). Con 6000 caracteres japoneses (hiragana) escritos a mano a partir de 3 usuarios, se logra una tasa de reconocimiento promedio de 99,92%. En [145], se utilizaron prototipos de caracteres, y una precisión de 99,2% se alcanza con una base de datos de 120 escritores, 160 caracteres por escritor. El reconocimiento de caracteres escritos como un servicio basado en Internet se propone en [146]. Los caracteres de 190 usuarios y 3,6 millones de muestras son reconocidas por la extracción de características direccionales y la extracción de características del gradiente. Las pruebas en diferentes OS móviles alcanzaron una tasa de reconocimiento media de 96,17% en 29,5 milisegundos. En comparación con los métodos tradicionales de autenticación, como la entrada de la contraseña desde el teclado, la biometría de escritura a mano se ha estudiado en los últimos años y se utiliza en esta área, ya que son más fiables y más difíciles de falsificar. La investigación anterior se ha concentrado en la identificación del escritor usando firma manuscrita [155] - [157], firma escrita a mano junto con una contraseña escrita desde teclado o PIN [158] - [159] [160].

2.2 Discusión

En este capítulo se presenta una revisión de la metodología con los métodos biométricos y su demarcación. Se consideraron los métodos biométricos para la verificación e identificación y los términos importantes y procedimientos para la evaluación de los resultados presentados. Se presentaron las bases de datos más utilizadas en el estado del arte, y los resultados de la identificación y verificación de los escritores se subdividieron en offline y online. Los registros fueron creados en su mayoría con una tablet digitalizadora y una pizarra con lápiz digital. Aparte de los datos offline (firmas o imágenes escaneadas), también se recogieron los datos online con los parámetros específicos, tales como coordenadas, el tiempo, la presión y la información del ángulo. Hemos echado un vistazo a las bases de datos públicas, que están compuestos principalmente de registros de firmas, caracteres individuales, palabras y frases enteras.

Los resultados de las investigaciones individuales varían fuertemente. El presente documento ofrece un resumen de los resultados de las publicaciones que se encuentran bajo investigación:

Precisión entre **67%** y **99%**
FAR entre **2.5%** y **26.98%**
FRR entre **1.5%** y **24.13%**
EER entre **0.41%** y **27.7%**

La mayoría de la investigación previa para la identificación del escritor ha sido limitada a los análisis de texto escritos a mano o firmas.

2.3 Conclusión

Además de las bases de datos existentes, las nuevas bases de datos con contraseñas escritas a mano para las investigaciones y la comparación, deben crearse para la identificación de los autores con las contraseñas escritas a mano. Los datos deben ser capturados con dispositivos tales como una tableta y un teléfono inteligente para mostrar que el sistema desarrollado es muy viable para la autenticación en los dispositivos estándar. Los parámetros online y offline del estado del arte que se han examinado, deben ser revisados específicamente para las contraseñas de escritura a mano, filtrándolos, o si es necesario, desarrollar nuevos parámetros para la investigación y probarlos en el sistema. Los clasificadores también deben ser probados por su conveniencia y probados con los conjuntos de datos y parámetros para saber qué clasificadores específicos son los más adecuados. Para el caso real, para implementar la solución, el tiempo requerido para la extracción de características y clasificación juega un papel decisivo.

Capitulo III: Concepto del sistema

Anteriores investigaciones abandonaron el análisis de las contraseñas escritas a mano para un sistema de acceso seguro y por ello no existe ninguna base de datos de contraseñas seguras escritas a mano. Por lo tanto, aquí tenemos un concepto para un sistema con estas características:

1. La captura de datos de contraseñas escritas a mano
2. La extracción de características
3. Clasificación
4. Construcción del modelo
5. identificación del escritor y verificación

Tiene que ser desarrollado.

En este capítulo, encontraremos un sistema cliente-servidor de cuatro pasos: Recolección de datos, extracción de características, Selección de características y clasificación con dos modos (registrarse y modo de prueba).

3.1 Introducción del concepto desde la vista del usuario

Primero el usuario introduce su nombre de usuario utilizando el teclado del dispositivo móvil. A continuación, se introduce la contraseña escrita a mano a través de la pantalla táctil. Los algoritmos online en el teléfono móvil proporcionan una primer pre-procesamiento de la contraseña escrita a mano. El resultado es un archivo con las características de las contraseñas escritas a mano. El teléfono móvil envía el archivo resultante al servidor. En el servidor se inicia la extracción de características y utiliza el modelo del sistema de clasificación. El resultado, si el usuario tiene acceso o no, es enviado de nuevo al teléfono móvil y puede iniciar un acceso a su cuenta virtual en el caso de un reconocimiento positivo.

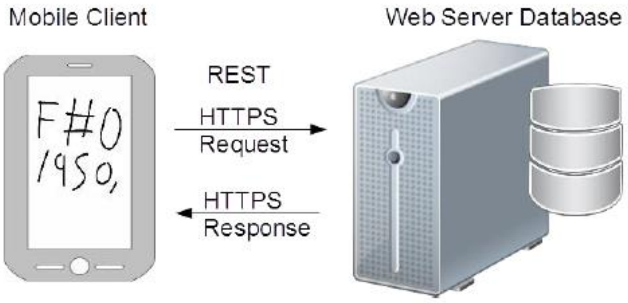


Figura 3.1: Identificación de un escritor en línea como una solución cliente-servidor móvil

Con el concepto propuesto, se ha desarrollado un prototipo con un teléfono inteligente Android como cliente móvil y un sistema Linux con la base de datos en el lugar del servidor. Para un servicio más seguro, en el estado del arte existe como paradigma el servicio REST (Representation State Transfer), el cual se usa con HTTPS como protocolo de transferencia. El propósito de REST es centrarse en la comunicación de máquina a máquina. REST es una alternativa simple a procedimientos similares utilizados en la realización anterior con SOAP, WSDL, y el método RPC relacionada. Las estrictas directrices de ayuda REST para construir servicios bien estructurados apoyan el uso de URLs limpias y esta es una de las principales ventajas de la arquitectura propuesta REST-servidor.

3.2 Servidor

Los datos transferidos por el cliente con HTTPS se almacenan en el servidor en una base de datos PostgreSQL y son procesados utilizando los algoritmos de la herramienta de minería de datos Weka [239], que se implementa en el servicio web REST. Las características se extraen y se genera el modelo, se almacenan en la base de datos utilizando algoritmos de reducción del parámetro y el clasificador en el modo de registro.

En el modo de prueba, las características se extraen de los datos transferidos por el cliente y una identificación/verificación se lleva a cabo con la ayuda del modelo generado. El resultado se envía de vuelta al cliente.

Las contraseñas escritas a mano originales y preprocesadas tienen que ser almacenadas en una base de datos (véase la figura 3.8).

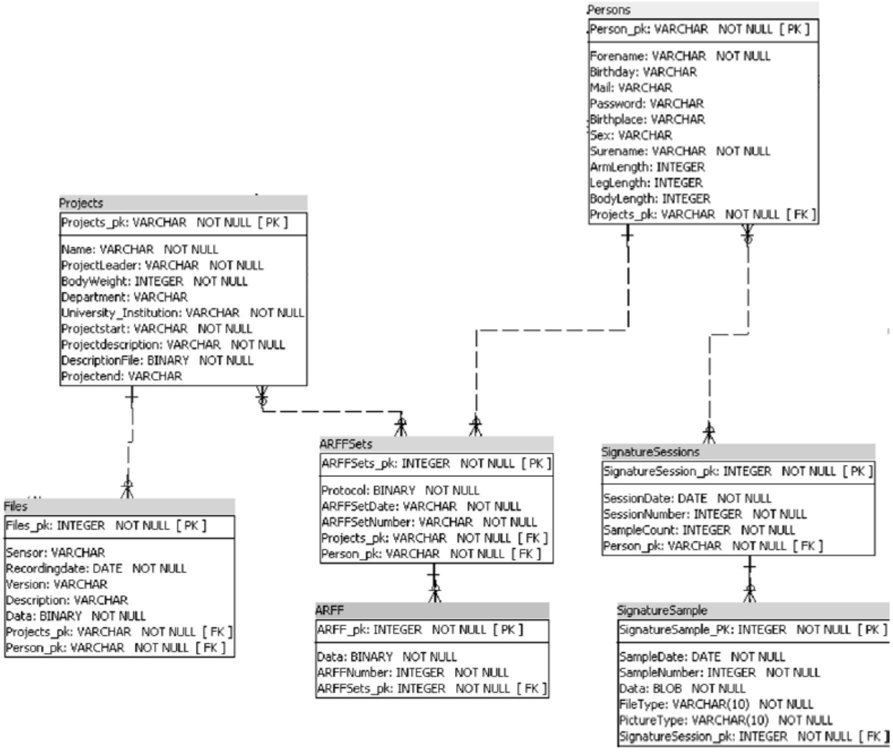

Persons
Person_pk: VARCHAR NOT NULL [PK]
Forename: VARCHAR NOT NULL
Birthday: VARCHAR
Mail: VARCHAR
Password: VARCHAR
Birthplace: VARCHAR
Sex: VARCHAR
Surename: VARCHAR NOT NULL
ArmLength: INTEGER
LegLength: INTEGER
BodyLength: INTEGER
Projects_pk: VARCHAR NOT NULL [FK]
Projects
Projects_pk: VARCHAR NOT NULL [PK]
Name: VARCHAR NOT NULL
ProjectLeader: VARCHAR NOT NULL
BodyWeight: INTEGER NOT NULL
Department: VARCHAR
University_Institution: VARCHAR NOT NULL
Projectstart: VARCHAR NOT NULL
Projectdescription: VARCHAR NOT NULL
DescriptionFile: BINARY NOT NULL
Projectend: VARCHAR
ARFFSets
ARFFSets_pk: INTEGER NOT NULL [PK]
Protocol: BINARY NOT NULL
ARFFSetDate: VARCHAR NOT NULL
ARFFSetNumber: VARCHAR NOT NULL
Projects_pk: VARCHAR NOT NULL [FK]
Person_pk: VARCHAR NOT NULL [FK]
SignatureSessions
SignatureSession_pk: INTEGER NOT NULL [PK]
SessionDate: DATE NOT NULL
SessionNumber: INTEGER NOT NULL
SampleCount: INTEGER NOT NULL
Person_pk: VARCHAR NOT NULL [FK]
Files
Files_pk: INTEGER NOT NULL [PK]
Sensor: VARCHAR
Recordingdate: DATE NOT NULL
Version: VARCHAR
Description: VARCHAR
Data: BINARY NOT NULL
Projects_pk: VARCHAR NOT NULL [FK]
Person_pk: VARCHAR NOT NULL [FK]
ARFF
ARFF_pk: INTEGER NOT NULL [PK]
Data: BINARY NOT NULL
ARFFNumber: INTEGER NOT NULL
ARFFSets_pk: INTEGER NOT NULL [FK]
SignatureSample
SignatureSample_PK: INTEGER NOT NULL [PK]
SampleDate: DATE NOT NULL
SampleNumber: INTEGER NOT NULL
Data: BLOB NOT NULL
FileType: VARCHAR(10) NOT NULL
PictureType: VARCHAR(10) NOT NULL
SignatureSession_pk: INTEGER NOT NULL [FK]

Figura 3.8: Parte del modelo de datos del servidor DBMS

Las clases para almacenar los datos de los objetos de transferencia del resto del servicio se muestran en la figura 3.9.

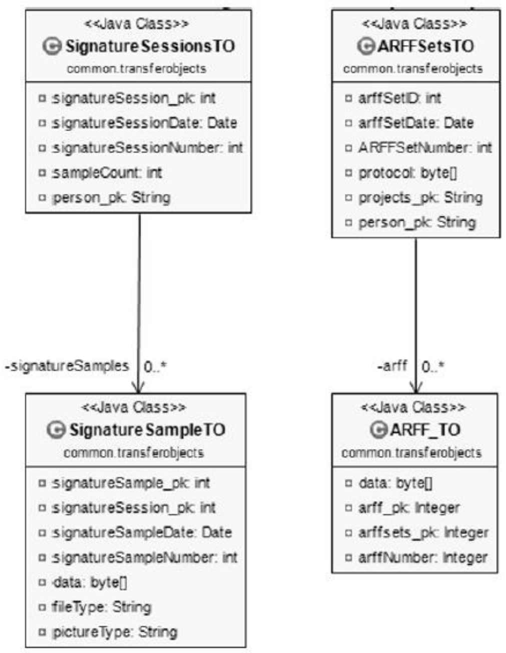


Figura 3.9: Parte de las clases de la transferencia de objetos en REST Web service

Después de la descripción de toda la estructura del sistema del servidor-cliente, continuamos con una visión general de la extracción de características, los métodos de clasificación y las bases de datos utilizadas.

Capitulo IV. Métodos y materiales

Este capítulo proporciona una visión general de los métodos utilizados en la recogida de datos, procesamiento previo, extracción de características, clasificación y materiales como conjuntos de datos y bases de datos utilizadas para los experimentos. Para construir un sistema de identificación y verificación, se puede apreciar una fase de entrenamiento y una fase de clasificación [110]. El proceso de entrenamiento consta esencialmente de 3 componentes, el preprocesamiento, la extracción de características y la clasificación, independientemente del tipo de la entrada. Esto se aplica no sólo a la escritura, sino también a todos los demás procedimientos de identificación y verificación o tareas de reconocimiento de patrones en general. (Véase la Figura 3.1)

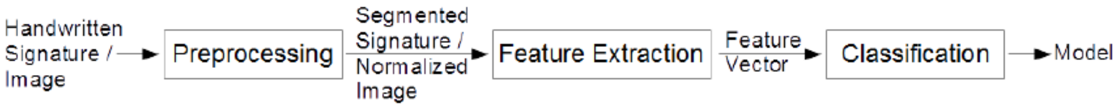

Handwritten Signature / Image
Preprocessing
Segmented Signature / Normalized Image
Feature Extraction
Feature Vector
Classification
Model

Figura 4.1: Fase de entrenamiento para la identificación y verificación de los escritores.

En la fase de formación, el modelo se construye y se entrena con la firma manuscrita. En la fase de clasificación, el sistema tiene que clasificar una firma escrita a mano por un escritor desconocido (identificación) véase la figura 3.2 o de un escritor conocido (verificación) véase la Figura 3.3. En la fase de prueba, el modelo entrenado se utiliza para identificar o verificar el escritor.

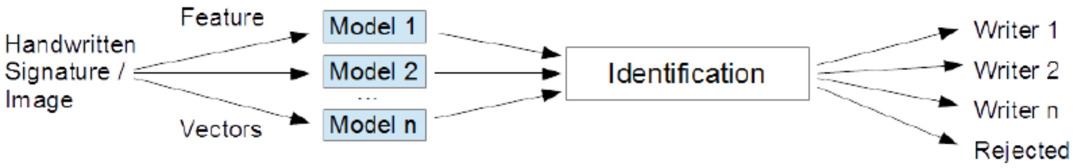


Figure 3.2: writer identification

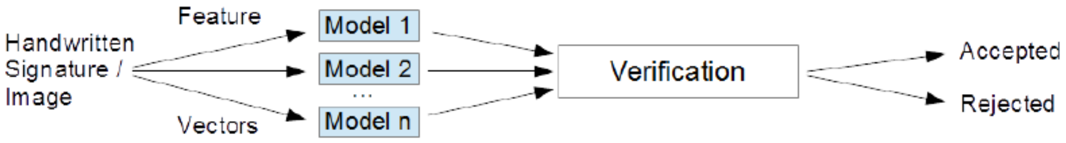


Figure 4.3: writer verification

Los medios de comunicación de entrada para el escritor de identificación / verificación consiste en cualquiera de las imágenes almacenadas (escaneados) (ofline) de papel o de un medio de entrada digital, tales como teléfonos inteligentes, tablets, un lápiz digital y la información de presión (online). Vamos a empezar con las bases de datos, que se presentan con más detalle en el capítulo siguiente.

4.1 Bases de datos

En [60] la identificación del escritor se llevó a cabo con una base de datos de desarrollo propio de 280 bloques de palabras escritas a mano, 176 originales y 104 falsificaciones. Para validar las características extraídas y algoritmos de clasificación para un sistema de autenticación segura, se buscaron bases de datos públicas. Como ya se ha mencionado en el capítulo 2.2, hay un gran número de bases de datos disponibles para el reconocimiento de escritura a mano, pero sólo unos pocos están disponibles en especial para la tarea de verificación de escritura a mano. El presente apartado ofrece un resumen y descripción detallada de las bases de datos públicas utilizadas y adaptadas, así como las bases de datos de desarrollo propio.

4.1.1 Bases de datos públicas

- **ATV-Signature Long Term Database (ATV- SLT DB)**

La base de datos ATV-SLT DB consta de 6 sesiones generadas por 27 usuarios a lo largo de 15 meses. Se examinó el cambio de la calidad de los resultados de la verificación durante un período de 15 meses. Para la prueba de un impostor se añadió una séptima sesión 3 años después de desarrollar los primeros conjuntos de datos [183]. (Véase la Figura 4.2)

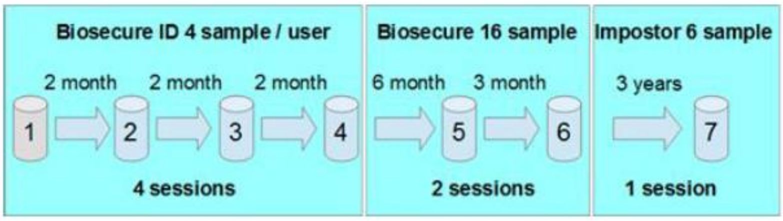


Figura 4.2: SLT DB y la sesión del impostor

Para la generación de las firmas impostoras era necesario generar las imágenes de las firmas originales. Estas imágenes se generaron como PNG gráfico de los archivos de SVC con un programa Java. El falsificador en el experimento usó esto como una plantilla para falsificar las firmas. (Véase la Figura. 4.4).

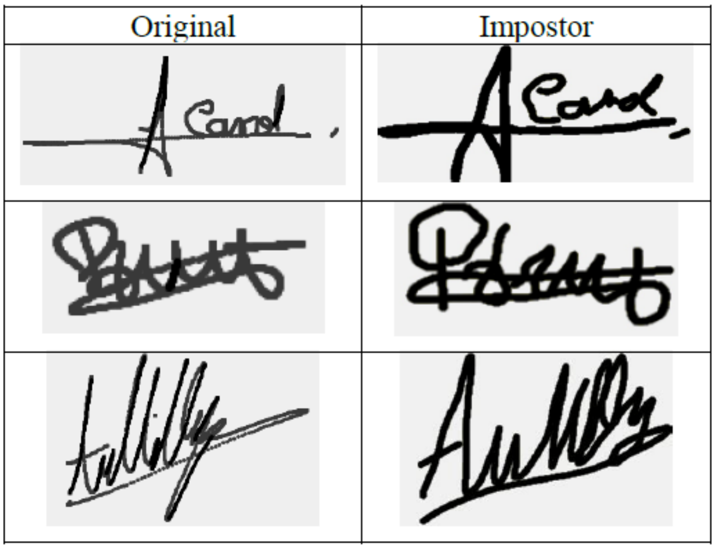


Figura 4.4: Muestras PNG de SVC y la firma del impostor en PNG [183]

- IAM online handwriting DB

Para determinar la posibilidad de utilizar las mismas características como en [60] para la verificación del escritor con palabras escritas a mano de un solo carácter, los experimentos en [194] están trabajando con la base de datos de escritura en línea IAM con textos en cursiva para su verificación. La base de datos contiene 1760 ejemplos de textos escritos a mano de 220 personas. El texto contiene entre 40 y 60 palabras en una y más de una línea. La base de datos de escritura en línea IAM considera 8 muestras de cada usuario [194]. (Véase la Figura 4.4)

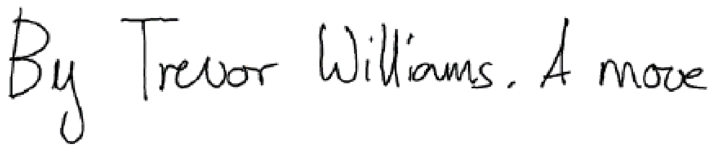


Figura 4.4: Ejemplo de texto escrito a mano de IAM [190]

4.2 Preprocesamiento

En pre-procesamiento también es necesario distinguir entre los datos de escritura online y offline de los datos de escritura a mano. los datos offline consisten en imágenes binarias o en escala de grises, donde en los datos online son características como la información del punto (coordenadas X e Y), el tiempo y la información de presión están incluidos. Los métodos para el procesamiento previo de los datos dependen en gran medida de los medios de comunicación de entrada y el estilo de escritura. Diferentes variantes para la exploración, mejora de la imagen y la normalización, así como diferentes tipos de características, se describen brevemente a continuación según el procedimiento offline y procedimiento online.

Los siguientes métodos esencialmente dependen del tipo de datos de entrada:

Online:

- El escaneo de la trayectoria de la pluma
- El procesamiento de trazos que sufren un delay (por ejemplo, puntos en las íes, que pueden ser insertados sólo al final de la palabra)

Offline:

- Eliminar la interferencia (ruido, la calidad del documento, la contaminación) causada por el escaneo o ya causada por la original.
- Esqueleto o contorno
- Determinación de la línea de base inferior y superior
- La normalización de tamaño, inclinación y la inclinación

4.2.1 Normalización

Las normalizaciones de los procedimientos conciernen tanto a los datos online como offline. Para un sistema de escritura dependiente, las características de escritura, tales como el tamaño y el nivel son generalmente constantes y por lo tanto necesariamente no necesita ser normalizado. En un sistema sin guion, las diferencias entre los escribas son demasiado grandes para evitar la estandarización. Los pasos de normalización son dependientes generalmente de la posición de las líneas de referencia. La línea de base superior e inferior a menudo no puede ser definida con tanta claridad. (Véase la figura 4.20).

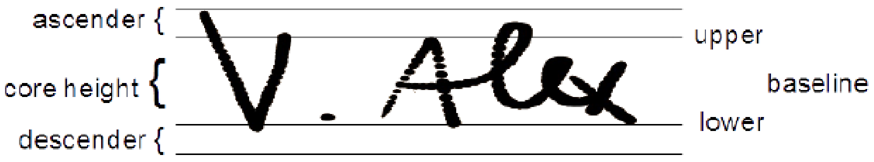


Figura 4.20: Definición de línea base superior e inferior

Por ejemplo, la tendencia a dibujar difiere mucho entre diferentes autores, en particular entre los diestros y zurdos, y también en cursiva.

4.2.2 Segmentación

Para generar los parámetros en línea, la segmentación se lleva a cabo en el teléfono del cliente en el preprocesado. El preprocesamiento comienza con el muestreo de parámetros como coordenadas de puntos, los valores de tiempo y el tacto para la identificación de los segmentos. Un nuevo segmento comienza cuando el lápiz o el dedo se coloca en la pantalla para la escritura y termina cuando el lápiz o el dedo se levanta de nuevo.

4.3 Extracción de características

En la zona de extracción de características, las características se clasifican en subconjuntos más detallados. Las características de la función se adquieren mediante el uso de una función en el tiempo, como la posición, velocidad y aceleración. Mientras tanto, las características de los parámetros son elementos, que se caracterizan como un vector de la firma. Tales como la duración total de la firma, las veces que se levanta al escribir, etc.

También en el caso de la extracción de características, se hace una distinción entre online y offline, como en el preprocesamiento. Las posibles características en el reconocimiento de la escritura online son las siguientes:

- Cambio en la dirección de escritura, la curvatura
- Tiempo de escritura, velocidad y presión
- Máximos locales a la hora de escribir sobre la línea, la anchura y la orientación de la firma

Para el reconocimiento de escritura offline, se pueden encontrar las siguientes características:

- Máximos y mínimos en la escritura sobre la línea base
- Determinación de las curvaturas, bucles, orientaciones y vértices
- DCT (transformada discreta del coseno) o coeficientes de Fourier y momentos
- Transformación o compresión de características utilizando redes neurales (NN)

La base para una buena identificación y verificación son buenas características, a lo largo de esta tesis, diversas características se han desarrollado y se han implementado una serie de funciones que se mencionan en el estado del arte. Las características se dividen a su vez en las funciones online and offline. Aquí, todas las características extraídas de la información online se presentan ahora como funciones online y las características extraídas de los gráficos PNG como características offline. Además, se distingue entre características geométricas, estadísticas y temporales. Además de las características temporales, hay más características importantes, ya que las actuales tabletas y teléfonos inteligentes disponibles en el mercado no detectan ningún tipo de presión, la dificultad de este trabajo radica en la falta de características de presión y en el logro de los mismos resultados que si usáramos características de presión. En la recolección de datos, sin embargo, los datos offline fueron también grabados y utilizados con fines comparativos para algunos experimentos.

4.4 Reducción de características

Una posibilidad para mejorar el rendimiento del sistema de identificación y verificación es la reducción del número de características. Se hace hincapié en que debemos buscar características de una muestra de escritura a mano de diferentes maneras y en diferentes direcciones. Sin embargo, esto no significa que a mayor número de características que se extraen de una muestra de escritura a mano, más ventajas podremos tomar de ellos. Por el contrario, será necesario realizar un sobreajuste si hay demasiadas características que se analizan simultáneamente. Esta es la razón por la cual es necesario e importante la evaluación y los procedimientos de la selección de características. Para esta tesis, por ejemplo, se ha utilizado un método basado en las puntuaciones de Fisher.

En resumen, se puede decir que la reducción de parámetros contribuye a[116]:

- Simplificación de modelos para que sean más fáciles de interpretar por los investigadores / usuarios
- tiempos más cortos de entrenamiento
- mejora de la generalización mediante la reducción de overfitting (formalmente, la reducción de la varianza) Hay varios métodos de reducción de la función (véase la figura 4.30) con algunas ventajas y desventajas para la selección de características y métodos de extracción de características.

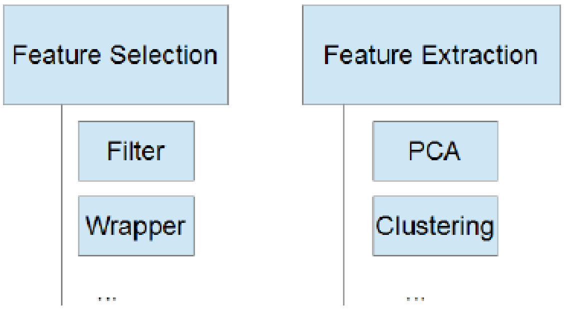


Figura 4.30: Métodos para la reducción de funciones

Capítulo V: Evaluación de los experimentos y resultados

Este capítulo está dividido en dos partes. En la primera parte se evaluaron experimentalmente dos conjuntos de datos privados con diferentes contraseñas escritas a mano. En la segunda parte, se evaluaron experimentalmente los conjuntos de datos públicos y privados con contraseñas seguras de firmas y textos de prueba. Se evalúa también la robustez de los conjuntos de datos para la identificación y verificación de los escritores, las características desarrolladas y los clasificadores. Al final del capítulo se muestra un resumen de los mejores resultados con las bases de datos, características y clasificadores utilizados.

5.1.1 Resultados usando contraseña DB-9

Para el primer experimento, se utilizaron contraseñas escritas a mano DB-9. Se utilizaron con 144 muestras (108 originales y 36 impostores) de contraseñas escritas a mano por nueve usuarios. Para la clasificación del conjunto de datos se dividió un conjunto de entrenamiento con 72 muestras originales y 2 conjuntos de prueba, una con 36 muestras auténticas escritas a mano y una con 36 muestras falsas (por ejemplo, para el inicio de sesión).

El índice de exactitud logrado varía de 90,62% a 96,87%. Los mejores resultados de la identificación fueron de hasta 96,87% y se alcanzaron por el clasificador Naive Bayes. El FAR utilizando Naive Bayes alcanzó 11,11%. Este primer experimento está trabajando en un intervalo de tiempo de 0,5 a 1,5 segundos dependiendo de la conexión WIFI. Para obtener el tiempo computacional real, no se llevan a cabo algunas pruebas de tiempo en el servidor y en el cliente móvil. El tiempo de cálculo para todo el proceso está sujeto a las limitaciones de la conexión a Internet, siendo de 0,5 a 1,5 segundos para las pruebas.

La solución propuesta se basa en un teléfono de pantalla táctil y un servidor. Esto demuestra que nuestro enfoque puede alcanzar una alta seguridad para la verificación biométrica de un usuario para un sistema de control de acceso. El sistema alcanza una tasa de verificación de 96,87% para distinguir entre los diferentes escritores que utilizan una combinación online y offline y la identificación de las características como segmentos, tiempo, puntos, ancho, altura, superficie, etc.

Normalmente, un usuario que conoce la contraseña de otro usuario siempre obtiene acceso ilegal a un sistema. Con esta combinación de una contraseña escrita a mano en un móvil y algoritmos de clasificación en el lado del servidor, el acceso ilegal se reduce con un FAR del 11,11%.

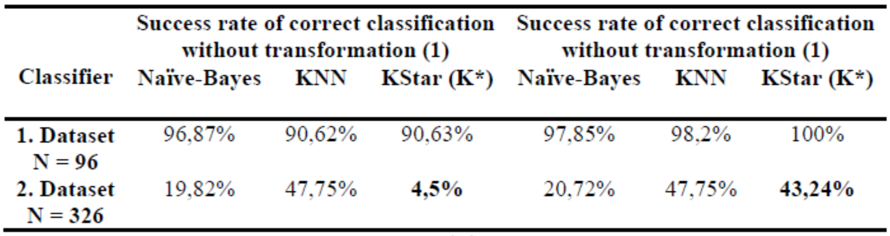
5.1.2 Experimentos con DB-9, BD-150 y 10 características

En este experimento los datos de dos conjuntos DB-9 (a) y BD-150 (b) se utilizaron como se describe en el apartado 4. Los registros se realizaron una vez sin la transformación de los datos y con la transformación de los datos. Un análisis de agrupamiento y el análisis de componentes principales (PCA) se utilizaron para transformar los datos. Para la clasificación, se utilizaron los clasificadores Naïve Bayes, KNN y K estrella. Para la extracción de las 10 características predominantemente geométricas se hizo de la misma forma que en el párrafo 5.1.1.

Ambos conjuntos de datos se analizaron por separado con los resultados de la clasificación con y sin transformación.

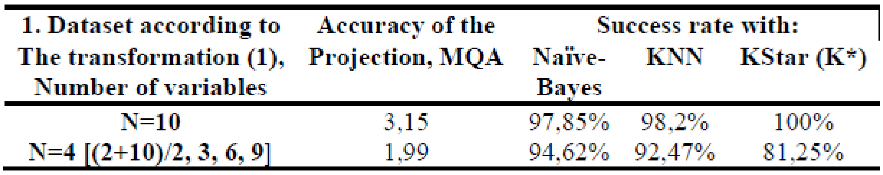
La Tabla IV resume los valores característicos más importantes en el experimento para ambos conjuntos de datos. La reducción de 4 variables mejoró la exactitud de la proyección en el análisis de componentes principales, pero la tasa de detección se deterioró. Esto se debió principalmente a la pérdida de información. Además, los valores atípicos en las figuras 2 (A) y 4 (A) tenían un efecto sobre los resultados del análisis de componentes principales. Para el segundo conjunto de datos, hubo una marcada mejoría en la tasa de reconocimiento después de la transformación (resaltado en la Tabla IV.).

Tabla IV. Comparación de los resultados: (A) tasa de reconocimiento sin y con transformación (1) para los dos conjuntos de datos; (B) Influencia de la reducción de variables en la tasa de reconocimiento para el primer data set con transformación



	Success rate of correct classification without transformation (1)			Success rate of correct classification without transformation (1)		
Classifier	Naïve-Bayes	KNN	KStar (K*)	Naïve-Bayes	KNN	KStar (K*)
1. Dataset N = 96	96,87%	90,62%	90,63%	97,85%	98,2%	100%
2. Dataset N = 326	19,82%	47,75%	**4,5%**	20,72%	47,75%	**43,24%**

(A)



1. Dataset according to The transformation (1), Number of variables	Accuracy of the Projection, MQA	Success rate with:		
		Naïve-Bayes	KNN	KStar (K*)
N=10	3,15	97,85%	98,2%	100%
N=4 [(2+10)/2, 3, 6, 9]	1,99	94,62%	92,47%	81,25%

(B)

Los mejores resultados de la combinación entre la identificación online y offline se puede aproximar a 99%. La transformación descrita en (1) de las variables conduce, entre otras cosas, a un marcado incremento en la exactitud de la proyección en el análisis de componentes principales. Los resultados del análisis de los grupos fueron consistentes con los resultados de la PCA. Sin embargo, el uso de ambos métodos es también un tanto peculiar: las variables utilizadas tienen valores atípicos y no son estocásticamente independientes. El uso de métodos de validación cruzada para la reducción de variables no se recomienda debido a que las variables extraídas son heterogéneas y no pueden subdividirse en grupos "Content-como". La información biométrica contenida en las diversas firmas también es heterogénea. Como se ve en el segundo conjunto de datos, la manipulación de la mayor parte del material de datos por un pequeño grupo de usuarios degrada la tasa de detección. Esta suposición se pudo confirmar, incluso después de una comparación visual con los datos biométricos originales. El clasificador KStar (K *) - reaccionó muy sensible a esta manipulación: la tasa de reconocimiento casi aumentó diez veces después de la transformación (1) para el segundo conjunto de datos.

5.2 Resultados y Experimentos de evaluación con contraseñas seguras escritas a mano, firmas y textos

En el capítulo anterior se presentaron los resultados para la identificación de escritores usando contraseñas escritas a mano con dos conjuntos de datos privadas diferentes. Ahora la robustez de las características extraídas y métodos de clasificación será probada con contraseñas seguras a mano, firmas y muestras de texto a partir de bases de datos públicas y privadas.

En este capítulo se encontrarán diferentes experimentos con:

- bases de datos privadas: contraseña segura DB-32, contraseña segura DB-150 con sesiones y contraseña segura y firma DB-30.
- bases de datos públicas: ATV-SLT-DB, DB IAM escritura online y SVC 2004 DB son evaluados. Información detallada acerca de las bases de datos están en el Capítulo 2 párrafos 1.1 ac y 1,2 ce.

También se van a mostrar los resultados en función del número de características que se han ido sumando, hasta llegar las 67 características propuestas, y a la aplicación de la reducción de características sobre las mismas, con la finalidad de tratar de mejorar el porcentaje de éxito, sobre las anteriores bases de datos.

5.3 Resultados resumidos

En la tabla VI se resumen los resultados de las pruebas. Se puede comprobar que los resultados podrían ser mejorados por picar el número de características. Una excepción a esto son las características fuera de línea, que no condujeron a ninguna mejora en la interacción con las características en línea. Los resultados son fuertemente dependientes de los conjuntos de datos tanto las bases de datos auto-desarrolladas como las bases de datos públicas se pueden lograr buenos resultados. Sin embargo, los requisitos previos son siempre que los registros correctamente recolectados sin error no son el caso, los resultados se deterioran fuertemente. Tanto las contraseñas, como las firmas, así como los textos cursivos proporcionan buenos resultados con las características y procedimientos utilizados, independientemente del dispositivo.

TABLA VI: Resumen de los mejores resultados de todos los experimentos acorde a las secciones de la versión en lengua inglesa.

Experiment	***Database | Features***	***Accuracy (best result)***	***FAR***
1 (5.1.1)	DB-9 | 10	96.87%	11.11% - 22.22%
2a (5.1.2)	DB-9 | 10	100%	
2b (5.1.2)	BD-150 | 10	47.75%	
3 (5.2.1)	DB-32 | 25	98.72%	10.42% - 18.75%
4 (5.2.2)	DB-32 | 39	100%	3.13% - 12.50%
5 (5.2.3)	ATV-SLT DB | 64	100%	2.47% – 11.73%
6a (5.2.4)	DB-150 | 67	99%	
6b (5.2.4)	IAM HW DB | 67	98.65%	
7 (5.2.5)	DB-150 | 117	98.09%	
8 (5.2.6)	DB-30 | 67	100%	2% - 10%

Capítulo 6: Conclusión y trabajo futuro

6.1 Resumen

En primer lugar, puede decirse que la hipótesis presentada en esta tesis se valida de acuerdo con los resultados de los experimentos del Capítulo 5 y las publicaciones subyacentes basadas en las bases de datos desarrolladas y las características extraídas para la identificación del escritor.

La cooperación de investigación entre BTU y ULPGC para la realización de esta tesis sólo fueron posibles gracias al renovado acuerdo de cooperación entre BTU y ULPGC.

En esta tesis se presenta un sistema para la identificación del escritor con contraseñas manuscritas. Se han desarrollado y probado características para la identificación y verificación. Para las pruebas, además de bases de datos públicas, se desarrollaron nuevas bases de datos especialmente para contraseñas y firmas manuscritas seguras. Se presentaron y aplicaron los métodos de preprocesamiento, extracción de características, clasificación para identificar al autor. En varios experimentos, se examinaron las características desarrolladas para su robustez y conveniencia para identificar a escritores específicamente con smartphone y tableta con contraseñas seguras y firmas para la identificación de escritores.

Después de una visión general de este trabajo y la motivación para este trabajo se aclaró en el capítulo 1, los diferentes sistemas biométricos y no biométricos para el control de acceso se presentaron, evaluaron y se estableció la hipótesis.

En el capítulo 2 se explican los métodos de verificación, identificación y medición de la calidad biométrica. Se presentan las bases de datos públicas más importantes y se presenta la estructura de las bases de datos auto-desarrolladas. Se da y se discute una visión detallada de las publicaciones más importantes del estado de la técnica de los últimos años sobre la verificación e identificación de escritores en línea y sin conexión. Los resultados se utilizarán para los próximos dos capítulos, Concepto y Métodos y Materiales.

El concepto presentado en el capítulo 3 se basa en una solución de servidor cliente con los dos modos de inscripción y prueba. Los datos manuscritos de las contraseñas se recogen con una aplicación en smartphone o tablet, segmentada y enviada al servidor. No sólo se recopilan datos en línea sino también datos fuera de línea, lo cual es importante para las pruebas con diferentes parámetros en el Capítulo 5. En el servidor, se lleva a cabo la extracción de parámetros y la clasificación. Los datos se almacenan en una base de datos en el servidor y el resultado de la identificación se devuelve al cliente.

En los métodos y materiales del Capítulo 4 se presentan los métodos básicos para la identificación y verificación del escritor. Además de los archivos con coordenadas, segmento y tiempo, las imágenes de las firmas también se guardan como archivos gráficos, especialmente para el ruido de fondo de datos en línea y fuera de línea. Las bases de datos se presentan divididas en bases de datos públicamente disponibles y creadas privadamente o ya existentes y extendidas. La contraseña segura DB-150, creada en 2015 en la ULPGC para esta tesis, y la contraseña segura y firma DB-30 en 2016 fue creada con smartphone y tableta para esta tesis en el BTU Cottbus - Senftenberg. Ambas bases de datos formaron la base para investigaciones importantes con contraseñas seguras, así como firmas y entregadas como en el capítulo 5 evaluaron buenos resultados. También debe mencionarse en este punto la investigación de BD-150 ya existente base de datos de escritura a mano que como la investigación muestra de un pequeño grupo de personas fue creado y por lo tanto sólo condicional utilizable y por lo que para la tesis deben ser desarrolladas nuevas bases de datos. Además, en el capítulo 4 se explica el desarrollo más importante de esta tesis, las características divididas por las características en línea y fuera de línea. Las funciones en línea de nuevo subdividido de acuerdo a las características geométricas, estadísticas y temporales. Las características especiales de presión no existen en la tableta y los teléfonos inteligentes por eso no se consideran en esta tesis. Conscientemente, no se desarrollaron características de presión, ya que los teléfonos inteligentes normales y las tabletas no tienen sensores de presión, así que, en esta tesis, una posibilidad debe y encontró buenos resultados para la identificación del escritor incluso sin presión. En total, se desarrollaron 67 características en línea y con las características Zernike Moments (20 hasta el pedido 15) y coeficientes de Fourier (30 hasta el orden de 100) además de 50 características fuera de línea, que no había resultados tan buenos en las investigaciones y por lo tanto no se utiliza para El sistema implementado. Además, la selección de características con los procedimientos más importantes para la extracción de características explicamos PCA y agrupación y selección de funciones como Fisher Score y Info Gain Attribute Evaluation. Finalmente, se presentan los clasificadores K-Nearest Neighbor, KStar, Naïve Bayes y Bayes Net.

La hipótesis presentada en esta tesis, que lleva a "La combinación de escritura a mano y contraseña segura conduce a una autenticación verdaderamente segura", puede confirmarse de acuerdo con los resultados de los experimentos presentados en el capítulo 5, tesis y publicaciones supervisadas de licenciatura y maestría. [124] [133] [151] [160]:

1. Las contraseñas manuscritas son más seguras que las contraseñas escritas con el teclado. Además, las contraseñas seguras, así como los textos con caracteres individuales y las firmas conducen con las características desarrolladas a una alta identificación segura del escritor. En comparación con las contraseñas y PIN escritos con el teclado, este método ofrece más seguridad mediante la incorporación de las características biométricas, tales como generados a partir de estilo de escritura y tiempo de escritura. Con las funciones ahora desarrolladas, la seguridad aumenta incluso si el escritor conoce la contraseña y el escritor ha observado al escribir. Esto hace casi imposible forjar una contraseña y obtener acceso a un sistema seguro.

2. Es posible identificar a los escritores utilizando la contraseña manuscrita.

El uso de la contraseña segura da como resultado una precisión de hasta el 100% y la tasa FAR podría bajarse hasta 3.13% (contraseña segura) o 2.47 (textos manuscritos) para obtener el mejor resultado en los experimentos.

3. Los dispositivos móviles son adecuados para la entrada de contraseña manuscrita.

Las contraseñas y firmas seguras, firmadas por smartphone y tableta, resultan seguras a pesar de las características de presión que faltan y pueden mantenerse al día con el estado de los resultados de arte generados con características de presión. Por lo tanto, es posible instalar este sistema en dispositivos estándar como smartphones y tabletas sin sacrificar la seguridad. La prueba con la tableta y el cojín ha demostrado que la identificación con la tableta y el cojín con la contraseña segura así como la firma suministra hasta 100% de exactitud suministra.

4. Las características de la identificación de escritura a mano pueden aplicarse tanto a una contraseña manuscrita como a una firma.

Además, las contraseñas seguras, así como los textos con caracteres únicos y las firmas conducen con las características desarrolladas a la identificación muy segura del escritor. La última prueba en el capítulo 4, párrafo 4.2.6, con conjuntos de datos con contraseña segura y firma recopilada de 30 usuarios con tableta y pad, ha demostrado que las características desarrolladas pueden usarse tanto para contraseñas como para firmas con diferentes dispositivos y resultados igualmente buenos.

5. Un impostor, con el conocimiento de la contraseña, es rechazado sobre la base de las características biométricas de su escritura.

En comparación con las contraseñas y el PIN escritos con el teclado, los métodos aplicados en esta tesis ofrecen más seguridad al incorporar las características biométricas, como las generadas por el estilo de escritura y el tiempo de escritura. Con las funciones ahora desarrolladas, la seguridad aumenta incluso si el escritor conoce la contraseña y el escritor ha observado al escribir. Esto hace casi imposible forjar una contraseña y obtener acceso a un sistema seguro.

6.2 Trabajo futuro

Este borrador proporciona la base para la identificación y verificación de contraseñas manuscritas y se puede desarrollar en el futuro. Es concebible el uso del sistema para la protección de cuentas bancarias, sistemas de pago, control de acceso y protección de datos confidenciales en dispositivos móviles.

Un sistema de seguridad práctico debe ser mejorado aún más, p. Por lo que la contraseña manuscrita no es visible en la pantalla y los datos deben ser cifrados, así como durante la transmisión al servidor y al teléfono inteligente durante el preproceso y en la base de datos del servidor.

Además del uso en el área técnica como la seguridad para la cuenta bancaria, un uso en medicina también es pensable, por ejemplo, para apoyar el diagnóstico médico de enfermedades, que van acompañadas de un cambio en la tipografía de los pacientes (por ejemplo, la enfermedad de Parkinson). Para un prototipo de análisis de la enfermedad de Parkinson se utilizaron los algoritmos de extracción y clasificación de características de esta tesis.