
BIOMETRICS

Edited by **Jucheng Yang**

INTECHWEB.ORG

Biometrics

Edited by Jucheng Yang

Published by InTech

Janeza Trdine 9, 51000 Rijeka, Croatia

Copyright © 2011 InTech

All chapters are Open Access articles distributed under the Creative Commons Non Commercial Share Alike Attribution 3.0 license, which permits to copy, distribute, transmit, and adapt the work in any medium, so long as the original work is properly cited. After this work has been published by InTech, authors have the right to republish it, in whole or part, in any publication of which they are the author, and to make other personal use of the work. Any republication, referencing or personal use of the work must explicitly identify the original source.

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Mirna Cvijic

Technical Editor Teodora Smiljanic

Cover Designer Jan Hyrat

Image Copyright Andy Piatt, 2010. Used under license from Shutterstock.com

First published July, 2011

Printed in Croatia

A free online edition of this book is available at www.intechopen.com

Additional hard copies can be obtained from orders@intechweb.org

Biometrics, Edited by Jucheng Yang

p. cm.

ISBN 978-953-307-618-8

INTECH OPEN ACCESS
PUBLISHER

INTECH open

free online editions of InTech
Books and Journals can be found at
www.intechopen.com

Contents

Preface IX

Part 1 Physical Biometrics 1

- Chapter 1 **Speaker Recognition 3**
Homayoon Beigi
- Chapter 2 **Finger Vein Recognition 29**
Kejun Wang, Hui Ma,
Oluwatoyin P. Popoola and Jingyu Li
- Chapter 3 **Minutiae-based Fingerprint
Extraction and Recognition 55**
Naser Zaeri
- Chapter 4 **Non-minutiae Based Fingerprint Descriptor 79**
Jucheng Yang
- Chapter 5 **Retinal Identification 99**
Mikael Agopov
- Chapter 6 **Retinal Vessel Tree as Biometric Pattern 115**
Marcos Ortega and Manuel G. Penedo
- Chapter 7 **DNA Biometrics 139**
Masaki Hashiyada
- ### **Part 2 Behavioral Biometrics 155**
- Chapter 8 **Keystroke Dynamics Authentication 157**
Romain Giot, Mohamad El-Abed
and Christophe Rosenberger
- Chapter 9 **DWT Domain On-Line Signature Verification 183**
Isao Nakanishi, Shouta Koike, Yoshio Itoh and Shigang Li

Part 3 Medical Biometrics 197

- Chapter 10 **Heart Biometrics: Theory, Methods and Applications 199**
Foteini Agrafioti, Jiexin Gao and Dimitrios Hatzinakos
- Chapter 11 **Human Identity Verification Based on Heart Sounds:
Recent Advances and Future Directions 217**
Francesco Beritelli and Andrea Spadaccini
- Chapter 12 **Investigation of Temporal Change in Heartbeat
in Transition of Sound and Music Stimuli 235**
Makoto Fukumoto and Hiroki Hasegawa
- Chapter 13 **The Use of Saliva Protein Profiling as a
Biometric Tool to Determine the Presence
of Carcinoma among Women 249**
Charles F. Streckfus and Cynthia Guajardo-Edwards

Preface

Biometrics uses methods for unique recognition of humans based upon one or more intrinsic physical or behavioral traits. In computer science, particularly, biometrics is used as a form of identity access management and access control. It is also used to identify individuals in groups that are under surveillance.

The key objective of the book is to provide comprehensive reference and text on human authentication and people identity verification from physiological, behavioural and other points of view (medical biometrics). It aims to publish new insights into current innovations in computer systems and technology for biometrics development and its applications.

The book consists of 13 chapters, each focusing on a certain aspect of the problem. The book chapters are divided into three sections: physical biometrics, behavioral biometrics and medical biometrics. In the first physical biometrics section, there are seven chapters. Chapter 1 provides an in-depth look at speaker recognition and address many practical and algorithmic issues related to the design and utilization of speaker recognition. In chapter 2 the author proposes some new algorithms for finger vein recognition such as using oriented filtering, template matching with relative distance and angle and wavelet moment fusing with PCA and LDA transform. In chapter 3 the author gives the recent advancements in the field of minutia-based fingerprint extraction and recognition. Chapter 4 provides a comprehensive idea about some of the well-known non-minutiae based descriptors during the last two decades and also proposes a novel non-minutiae based fingerprint descriptor with tessellated invariant moment features and Support Vector Machine (SVM). In chapter 5 the retina scanning technique is considered in detail throughout its historical evolution and try to use the birefringence of the retinal nerve fiber layer (RNFL) as a basis for successful identification. Chapter 6 proposes a fully-automatic authentication system using the retinal vessel tree pattern as biometric characteristic. As the most reliable personal identification, DNA is intrinsically digital and does not change during a person's life, and even after death. In chapter 7 the author proposes a method for generating a personal ID comprising short tandem repeat (STR) and single nucleotide polymorphism (SNP) information which are used in personal identification in forensic application.

In Section 2, two kinds of behavioral biometrics: keystroke dynamics and DWT domain on-line signature verification are introduced in chapter 8 and chapter 9 respectively. In section 3, medical biometrics constitutes another category of new biometric recognition modalities that encompasses signals which are typically used in clinical diagnostics, so chapter 10 gives a survey on heart biometrics with its theory, methods and applications. Chapter 11 proposes the usage of heart sounds for biometric recognition, describes the strengths and the weaknesses of the novel trait and analyzes in detail the methods developed so far and their performance. Chapter 12 investigates the temporal change in heartbeat intervals in a transition between different sound stimuli, since observing temporal change in heartbeat is important and contributes to improvement of exposure method of music and sound. In chapter 13 the author proposes the use of saliva protein profiling as a biometric tool to authenticate the presence of carcinoma among women.

The book was reviewed by editor Dr. Jucheng Yang, and some guest editors, such as Dr. Girija Chetty, Dr. Norman Poh, Dr. Loris Nanni, Dr. Jianjiang Feng, Dr. Dongsun Park, Dr. Sook Yoon and other.

Dr. Jucheng Yang
Professor
School of Information Technology,
Jiangxi University of Finance and Economics,
Nanchang, Jiangxi province,
China

Part 1

Physical Biometrics

Speaker Recognition

Homayoon Beigi
Recognition Technologies, Inc.
U.S.A.

1. Introduction

Speaker Recognition is a multi-disciplinary technology which uses the vocal characteristics of speakers to deduce information about their identities. It is a branch of biometrics that may be used for *identification*, *verification*, and *classification* of individual speakers, with the capability of *tracking*, *detection*, and *segmentation* by extension.

A speaker recognition system first tries to model the vocal tract characteristics of a person. This may be a mathematical model of the physiological system producing the human speech or simply a statistical model with similar output characteristics as the human vocal tract. Once a model is established and has been associated with an individual, new instances of speech may be assessed to determine the likelihood of them having been generated by the model of interest in contrast with other observed models. This is the underlying methodology for all speaker recognition applications. The earliest known papers on speaker recognition were published in the 1950s (Pollack et al., 1954; Shearme & Holmes, 1959).

Initial speaker recognition techniques relied on a human expert examining representations of the speech of an individual and making a decision on the person's identity by comparing the characteristics in this representation with others. The most popular representation was the *formant* representation. In the recent decades, fully automated speaker recognition systems have been developed and are in use (Beigi, 2011).

There have been a number of tutorials, surveys, and review papers published in the recent years (Bimbot et al., 2004; Campbell, 1997; Furui, 2005). In a somewhat different approach, we have tried to present the material, more in the form of a comprehensive summary of the field with an ample number of references for the avid reader to follow. A coverage of most of the aspects is presented, not just in the form of a list of different algorithms and techniques used for handling part of the problem, as it has been done before.

As for the importance of speaker recognition, it is noteworthy that *speaker identity* is the only biometric which may be easily tested (identified or verified) remotely through the existing infrastructure, namely the telephone network. This makes speaker recognition quite valuable and unrivaled in many real-world applications. It needs not be mentioned that with the growing number of cellular (mobile) telephones and their ever-growing complexity, speaker recognition will become more popular in the future.

There are countless number of applications for the different branches of speaker recognition. If audio is involved, one or more of the speaker recognition branches may be used. However, in terms of deployment, speaker recognition is in its early stages of infancy. This is partly due to unfamiliarity of the general public with the subject and its existence, partly because of the limited development in the field. These include, but are certainly not limited to, *financial*,

forensic and legal (Nolan, 1983; Tosi, 1979), *access control and security*, *audio/video indexing and diarization*, *surveillance*, *teleconferencing*, and *proctorless distance learning* Beigi (2009).

Speaker recognition encompasses many different areas of science. It requires the knowledge of *phonetics*, *linguistics* and *phonology*. *Signal processing* which by itself is a vast subject is also an important component. *Information theory* is at its basis and *optimization theory* is used in solving problems related to the training and matching algorithms which appear in *support vector machines (SVMs)*, *hidden Markov models (HMMs)*, and *neural networks (NNs)*. Then there is *statistical learning theory* which is used in the form of *maximum likelihood estimation*, *likelihood linear regression*, *maximum a-posteriori probability*, and other techniques. In addition, *Parameter estimation* and *learning techniques* are used in *HMM*, *SVM*, *NN*, and other underlying methods, at the core of the subject. *Artificial intelligence* techniques appear in the form of *sub-optimal searches* and *decision trees*. Also *applied math*, in general, is used in the form of *complex variables theory*, *integral transforms*, *probability theory*, *statistics*, and many other mathematical domains such as *wavelet analysis*, etc.

The vast domain of the field does not allow for a thorough coverage of the subject in a venue such as this chapter. All that can be done here is to scratch the surface and to speak about the inter-relations among these topics to create a complete speaker recognition system. The avid reader is recommended to refer to (Beigi, 2011) for a comprehensive treatment of the subject, including the details of the underlying theory.

To start, let us briefly review different biometrics in contrast with speaker recognition. Then, it is important to clarify the terminology and to describe the problems of interest by reviewing the different manifestations and modalities of this biometric. Afterwards, some of the challenges faced in achieving a practical system are listed. Once the problems are clearly posed and the challenges are understood, a quick review of the production and the processing of speech by humans is presented. Then, the state of the art in addressing the problems at hand is briefly surveyed in a section on theory. Finally, concluding remarks are made about the current state of research on the subject and its future trend.

2. Comparison with other biometrics

There have been a number of biometrics used in the past few decades for the recognition of individuals. Some of these markers have been discussed in other chapters of this book. A comparison of voice with some other popular biometrics will clarify the scope of its practical usage. Some of the most popular biometrics are *Deoxyribonucleic Acid (DNA)*, *image-based* and *acoustic ear recognition*, *face recognition*, *fingerprint* and *palm recognition*, *hand* and *finger geometry*, *iris* and *retinal recognition*, *thermography*, *vein recognition*, *gait*, *handwriting*, and *keystroke recognition*.

Fingerprints, as popular as they are, have the problem of not being able to identify people with damaged fingers. These are, for example, construction workers, people who work with their hands, or maybe people without limbs, such as those who have either lost their hands or their fingers in an accident or those who congenitally lack fingers or limbs. According to the National Institute of Standards and Technology (NIST), this is about 2% of the population! Also, latex prints of finger patterns may be used to spoof some sensors.

People, with damaged irides, such as some who are blind, either congenitally or due to an illness like glaucoma, may not be recognized through iris recognition. It is very hard to tell the size of this population, but they certainly exist. Additionally, one would need a high quality image of the iris to perform recognition. Acquiring these images is quite problematic. Although there are long distance iris imaging cameras, their field of vision may easily be

blocked by uncooperative users through the turning of the head, blinking, rolling of the eyes, wearing of hats, glasses, etc. The image may also not be acceptable due to lighting and focus conditions. Also, irides tend to change due to changes in lighting conditions as the pupils dilate or contract. It is also possible to spoof some iris recognition systems, either by wearing contact lenses or by simply using an image of the target individual's irides.

Of course, there is also a percentage of the population who are unable to speak, therefore they will not be able to use speaker recognition systems. The latest figures for the population of deaf and mute people in the United States reflected by the US Census Bureau set this percentage at 0.4% for deaf and mute individuals (USC, 2005). Spoofing, using recordings is also a concern in practical speaker recognition systems.

In terms of public acceptance, fingerprint recognition has long been associated with criminology. Due to these legacy associations, many individuals are wary of producing a fingerprint for fear of its malicious usage or simply due to the criminal connotation it carries. As an example, a few years ago, the United States government required capturing the image and fingerprint of all tourists entering the nation's airports. This action offended many tourists to the point that some countries such as Brazil placed a reciprocal system in place only for U.S. citizens entering their country. Many people entering the U.S. felt like they were being treated as criminals, only based on the act of fingerprinting. Of course, since many other countries have been adopting the fingerprint capture requirement, it is being tolerated by travelers much better, around the world.

Because facial, iris, retinal images, and fingerprints have a sole purpose of being used in recognition, they are somewhat harder to capture. In general, the public is more wary of providing such information which may be archived and misused. On the other hand, speech has been established for communication and people are far less likely to be concerned about parting with their speech. Even in the technological arena, the use of speech for telephone communication makes it much more socially acceptable.

Speaker recognition can also utilize the widely available infrastructure that has been around for so long, namely the telephone network. Speech may be used for doing remote recognition of the individual using the existing telephone network and without the need for any extra hardware or other apparatus. Also, speaker recognition, in the form of tracking and detection may be used to do much more than simple identification and verification of individuals, such as a full diarization of large media databases. Another attractive point is that cellular telephone and PDA-type data security needs no extra hardware, since cellular telephones already have speech capture devices, namely microphones. Most PDAs also contain built-in microphones. On the other hand, for fingerprint and image recognition, a fingerprint scanner and a camera would have to be present.

Multimodal biometrics entail systems which combine any two or more of these or other biometrics. These combinations increase the accuracy of the identification or verification of the individual based on the fact that the information is obtained through different, mostly independent sources. Most practical implementations of biometric system will need to utilize some kind of multimodal approach; since any one technique may be bypassed by the eager impostor. It would be much more difficult to fool several independent biometric systems simultaneously. Many of the above biometrics may be successfully combined with speaker recognition to produce viable multimodal systems with much higher accuracies. (Viswanathan et al., 2000) shows an example of such a multimodal approach using speaker and image recognition.

3. Terminology and manifestations

In addressing the act of *speaker recognition* many different terms have been coined, some of which have caused great confusion. *Speech recognition* research has been around for a long time and, naturally, there is some confusion in the public between *speech* and *speaker recognition*. One term that has added to this confusion is *voice recognition*.

The term *voice recognition* has been used in some circles to double for *speaker recognition*. Although it is conceptually a correct name for the subject, it is recommended that the use of this term is avoided. *Voice recognition*, in the past, has been mistakenly applied to *speech recognition* and these terms have become synonymous for a long time. In a speech recognition application, it is not the voice of the individual which is being recognized, but the contents of his/her speech. Alas, the term has been around and has had the wrong association for too long.

Other than the aforementioned, a myriad of different terminologies have been used to refer to this subject. They include, *voice biometrics*, *speech biometrics*, *biometric speaker identification*, *talker identification*, *talker clustering*, *voice identification*, *voiceprint identification*, and so on. With the exception of the term *speech biometrics* which also introduces the addition of a speech knowledge-base to speaker recognition, the rest do not present any additional information.

3.1 Speaker enrollment

The first step required in most manifestations of speaker recognition is to enroll the users of interest. This is usually done by building a mathematical model of a sample speech from the user and storing it in association with an identifier. This model is usually designed to capture statistical information about the nature of the audio sample and is mostly irreversible – namely, the enrollment sample may not be reconstructed from the model.

3.2 Speaker identification

There are two different types of speaker identification, *closed-set* and *open-set*. Closed-set identification is the simpler of the two problems. In close-set identification, the audio of the test speaker is compared against all the available speaker models and the speaker ID of the model with the closest match is returned. In practice, usually, the top best matching candidates are returned in a ranked list, with corresponding confidence or likelihood scores. In closed-set identification, the ID of one of the speakers in the database will always be closest to the audio of the test speaker; there is no rejection scheme.

One may imagine a case where the test speaker is a 5-year old child where all the speakers in the database are adult males. In closed-set Identification, still, the child will match against one of the adult male speakers in the database. Therefore, closed-set identification is not very practical. Of course, like anything else, closed-set identification also has its own applications. An example would be a software program which would identify the audio of a speaker so that the interaction environment may be customized for that individual. In this case, there is no great loss by making a mistake. In fact, some match needs to be returned just to be able to pick a customization profile. If the speaker does not exist in the database, then there is generally no difference in what profile is used, unless profiles hold personal information, in which case rejection will become necessary.

Open-set identification may be seen as a combination of closed-set identification and speaker verification. For example, a closed-set identification may be conducted and the resulting ID may be used to run a speaker verification session. If the test speaker matches the target speaker based on the ID, returned from the closed-set identification, then the ID is accepted

and passed back as the true ID of the test speaker. On the other hand, if the verification fails, the speaker may be rejected all-together with no valid identification result. An open-set identification problem is therefore at least as complex as a speaker verification task (the limiting case being when there is only one speaker in the database) and most of the time it is more complex. In fact, another way of looking at verification is as a special case of open-set identification in which there is only one speaker in the list. Also, the complexity generally increases linearly with the number of speakers enrolled in the database since theoretically, the test speaker should be compared against all speaker models in the database – in practice this may be avoided by tolerating some accuracy degradation (Beigi et al., 1999).

3.3 Speaker verification (authentication)

In a generic speaker verification application, the person being verified (known as the test speaker), identifies himself/herself, usually by non-speech methods (e.g., a username, an identification number, et cetera). The provided ID is used to retrieve the enrolled model for that person which has been stored according to the enrollment process, described earlier, in a database. This enrolled model is called the *target speaker model* or the *reference model*. The speech signal of the test speaker is compared against the target speaker model to verify the test speaker.

Of course, comparison against the target speaker's model is not enough. There is always a need for contrast when making a comparison. Therefore, one or more competing models should also be evaluated to come to a verification decision. The competing model may be a so-called (universal) background model or one or more cohort models. The final decision is made by assessing whether the speech sample given at the time of verification is closer to the target model or to the competing model(s). If it is closer to the target model, then the user is verified and otherwise rejected.

The speaker verification problem is known as a one-to-one comparison since it does not necessarily need to match against every single person in the database. Therefore, the complexity of the matching does not increase as the number of enrolled subjects increases. Of course in reality, there is more than one comparison for speaker verification, as stated – comparison against the target model and the competing model(s).

3.3.1 Speaker verification modalities

There are two major ways in which speaker verification may be conducted. These two are called the *modalities* of speaker verification and they are *text-dependent* and *text-independent*. There are also variations of these two modalities such as *text-prompted*, *language-independent text-independent* and *language-dependent text-independent*.

In a purely *text-dependent* modality, the speaker is required to utter a predetermined text at enrollment and the same text again at the time of verification. Text-dependence does not really make sense in an identification scenario. It is only valid for verification. In practice, using such text-dependent modality will be open to *spoofing* attacks; namely, the audio may be intercepted and recorded to be used by an impostor at the time of the verification. Practical applications that use the text-dependent modality, do so in the text-prompted flavor. This means that the enrollment may be done for several different textual contents and at the time of verification, one of those texts is requested to be uttered by the test speaker. The chosen text is the prompt and the modality is called *text-prompted*.

A more flexible modality is the *text-independent* modality in which case the texts of the speech at the time of enrollment and verification are completely random. The difficulty with this

method is that because the texts are presumably different, longer enrollment and test samples are needed. The long samples increase the probability of better coverage of the idiosyncrasies of the person's vocal characteristics.

The general tendency is to believe that in the text-dependent and text-prompted cases, since the enrollment and verification texts are identical, they can be designed to be much shorter. One must be careful, since the shorter segments will only examine part of the dynamics of the vocal tract. Therefore, the text for text-prompted and text-dependent engines must still be designed to cover enough variation to allow for a meaningful comparison.

The problem of spoofing is still present with text-independent speaker verification. In fact, any recording of the person's voice should now get an impostor through. For this reason, text-independent systems would generally be used with another source of information in a multi-factor authentication scenario.

In most cases, *text-independent* speaker verification algorithms are also *language-independent*, since they are concerned with the vocal tract characteristics of the individual, mostly governed by the shape of the speaker's vocal tract. However, because of the coverage issue discussed earlier, some researchers have developed text-independent systems which have some internal models associated with phonemes in the language of their scope. These techniques produce a text-independent, but somewhat language-dependent speaker verification system. The language limitations reduce the space and, hence, may reduce the error rates.

3.4 Speaker and event classification

The goal of classification is a bit more vague. It is the general label for any technique that pools similar audio signals into individual bins. Some examples of the many classification scenarios are gender classification, age classification, and event classification. Gender classification, as is apparent from its name, tries to separate male speakers and female speakers. More advanced versions also distinguish children and place them into a separate bin; classifying male and female is not so simple in children since their vocal characteristics are quite similar before the onset of puberty. Classification may use slightly different sets of features from those used in verification and identification, depending on the problem at hand. Also, either there may be no enrollment or enrollment may be done differently. Some examples of special enrollment procedures are, *pooling enrollment data* from like classes together, using *extra features in supplemental codebooks* related to specific natural or logical specifics of the classes of interest, etc. (Beigi, 2011).

Although these methods are called speaker classification, sometimes, the technique are used for doing event classification such as classifying speech, music, blasts, gun shots, screams, whistles, horns, etc. The feature selection and processing methods for classification are mostly dependent on the scope and could be different from mainstream speaker recognition.

3.5 Speaker segmentation, diarization, detection and tracking

Automatic segmentation of an audio stream into parts containing the speech of distinct speakers, music, noise, and different background conditions has many applications. This type of segmentation is elementary to the practical considerations of speaker recognition as well as speech and other audio-related recognition systems. Different specialized recognizers may be used for recognition of distinct categories of audio in a stream.

An example is the ever-growing tele-conferencing application. In a tele-conference, usually, a host makes an appointment for a conference call and notifies attendees to call a telephone number and to join the conference using a special access code. There is an increasing

interest from the involved parties to obtain transcripts (minutes) of these conversations. In order to fully transcribe the conversations, it is necessary to know the speaker of each statement. If an enrolled model exists for each speaker, then prior to identifying the active speaker (*speaker detection*), the audio of that speaker should be segmented and separated from adjoining speakers. When speaker segmentation is combined with speaker identification and the resulting index information is extracted, the process is called *speaker diarization*. In case one is only interested in a specific speaker and where that speaker has spoken within the conversation (the timestamps), the process is called *speaker tracking*.

3.6 Knowledge-based speaker recognition (speech biometrics)

A knowledge-based speaker recognition system is usually a combination of a speaker recognition system and a speech recognizer and sometimes a natural language understanding engine or more. It is somewhat related to the *text-prompted* modality with the difference that there is another abstraction layer in the design. This layer uses knowledge from the speaker to test for liveness or act as an additional authentication factor. As an example, at the enrollment time, specific information such as a Personal Identification Number (PIN) or other private data may be stored about the speakers. At the verification time, randomized questions may be used to capture the test speaker's audio and the content of interest. The content is parsed by doing a transcription of the audio and using a natural language understanding (Manning, 1999) system to parse for the information of interest. This will increase the factors in the authentication and is usually a good idea for reducing the chance of successful impostor attacks – see Figure 1.

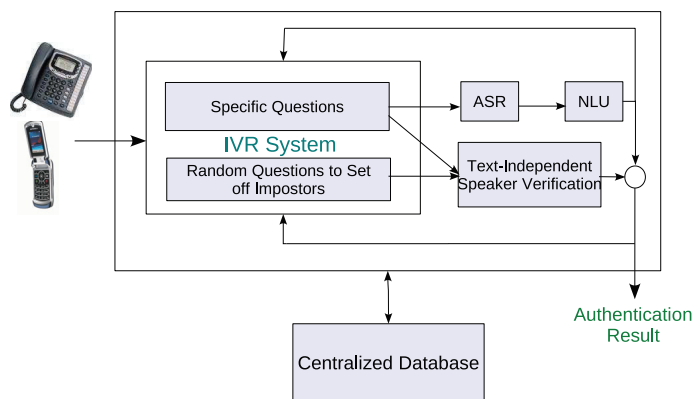


Fig. 1. A practical speaker recognition system utilizing speech recognition and natural language understanding

4. Challenges of speaker recognition

Aside from its positive outlook such as the established infrastructure and simplicity of adoption, speaker recognition, too, is filled with difficult challenges for the research community. Channel mismatch is the most serious difficulty faced in this technology. As an example, assume using a specific microphone over a channel such as a cellular communication channel with all the associated band-limitations and noise conditions in one session of using

a speaker recognition system. For instance, this session can be the enrollment session for instance.

Therefore, all that the system would learn about the identity of the individual is tainted by the channel characteristics through which the audio had to pass. On the hand, at the time of performing the identification or verification, a completely different channel could be used. For example, this time, the person being identified or verified may call from his/her home number or an office phone. These may either be digital phones going through voice T1 services or may be analog telephony devices going through analog switches and being transferred to digital telephone company switches, on the way.

They would have specific characteristics in terms of dynamics, cut-off frequencies, color, timber, etc. These channel characteristics are basically modulated with the characteristics of the person's vocal tract. Channel mismatch is the source of most errors in speaker recognition. Another problem is signal variability. This is by no means specific to speaker recognition. It is a problem that haunts almost all biometrics. In general, an abundance of data is needed to be able to cover all the variations within an individual's voice. But even then, a person in two different sessions, would possibly have more variation within his/her own voice than if the signal is compared to that of someone else's voice, who possesses similar vocal traits.

The existence of wide intra-class variations compared with inter-class variations makes it difficult to be able to identify a person accurately. *Inter-class* variations denote the difference between two different individuals while *intra-class* variations represent the variation within the same person's voice in two different sessions.

The signal variation problem, as stated earlier, is common to most biometrics. Some of these variations may be due to aging and time-lapse effects. Time-lapse could be characterized in many different ways (Beigi, 2009). One is the aging of the individual. As we grow older, our vocal characteristics change. That is a part of aging in itself. But there are also subtle changes that are not that much related to aging and may be habitual or may also be dependent on the environment, creating variations from one session to another. These short-term variations could happen within a matter of days, weeks, or sometimes months. Of course, larger variations happen with aging, which take effect in the course of many years.

Another group of problems is associated with background conditions such as ambient noise and different types of acoustics. Examples would be audio generated in a room with echos or in a street while walking and talking on a mobile (cellular) phone, possibly with fire trucks, sirens, automobile engines, sledge hammers, and similar noise sources being heard in the background. These conditions affect the recognition rate considerably. These types of problems are quite specific to speaker recognition. Of course, similar problems may show up in different forms in other biometrics.

For example, analogous conditions in image recognition would show up in the form of noise in the lighting conditions. In fingerprint recognition they appear in the way the fingerprint is captured and related noisy conditions associated with the sensors. However, for biometrics such as fingerprint recognition, the technology may more readily dictate the type of sensors which are used. Therefore, in an official implementation, a vendor or an agency may require the use of the same sensor all around. If one considers the variations across sensors, different results may be obtained even in fingerprint recognition, although they would probably not be as pronounced as the variations in microphone conditions.

The original purpose of using speech has been to be able to convey a message. Therefore, we are used to deploying different microphones and channels for this purpose. One person, in general uses many different speech apparatuses such as a home phone, cellphone, office

phone, and even a microphone headset attached to a computer. We still expect to be able to perform reasonable speaker recognition using this varied set of sensors and channels. Although, as mentioned earlier, this becomes an advantage in terms of ease of adoptability of speaker recognition in existing arenas, it also makes the speaker recognition problem much more challenging.

Another problem is the presence of vocal variations due to illness. Catching a cold causes changes to our voice and its characteristics which could create difficulties in performing accurate speaker recognition. Bulk of the work in speaker recognition research is to be able to alleviate these problems, although not every problem is easily handled with the current technology.

5. Human speech generation and processing

A human child develops an inherent ability to identify the voice of his/her parents before even learning to understand the content of their speech. In humans, speaker recognition is performed in the right (less dominant) hemisphere of the brain, in conjunction with the functions for processing pitch, tempo, and other musical discourse. This is in contrast with most of the language functions (production and perception) in the brain which are processed by the *Broca* and *Wernicke* areas in the left (dominant) hemisphere of the *cerebral cortex* (Beigi, 2011).

Speech generation starts with the speech content being developed in the brain and processed through the nervous system. It includes the intended message which is created in the brain. The abstraction of this message is encoded into a code that will then produce the language (language coding step). The brain will then induce neuro-muscular activity to start the vocal tract in vocalizing the message. This message is transmitted over a channel starting with the air surrounding the mouth and continuing with electronic devices and networks such as a telephone system to transmit the coded message.

The resulting signal is therefore transmitted to the air surrounding the ear, where vibrations travel through different sections of the outer and the middle ear. The *cochlear* vibrations excite the *cilia* in the inner ear, generating neural signals which travel through the *Thalamus* to the brain. These signals are then decoded by different parts of the brain and are decoded into linguistic concepts which are understood by the receiving individual.

The intended message is embedded in the abstraction which is deduced by the brain from the signal being presented to it. This is a very complex system where the intended message generally contains a very low bit-rate content. However, the way this content undergoes transformation into a language code, neuro-muscular excitation, and finally audio, increase the bit-rate of the signal substantially, generating great redundancy.

Therefore a low information content is encoded to travel through a high-capacity channel. This small amount of information may easily be tainted by noise throughout this process.

Figure 2 depicts a control system representation of speech production proposed by (Beigi, 2011). Earlier, we considered the transformation of a message being formed in the brain into a high-capacity audio signal. In reality, the creation of the audio signal from the fundamental message formed in the brain may be better represented using a control system paradigm.

Let us consider the *Laplace transform* of the original message being generated in the brain as $U(s)$. We may lump together the different portions of the nervous system at work in generating the control signals which move the vocal tract to generate speech, into a controller block, $G_c(s)$. This block is made up of $G_b(s)$ which makes up those parts of the nervous system

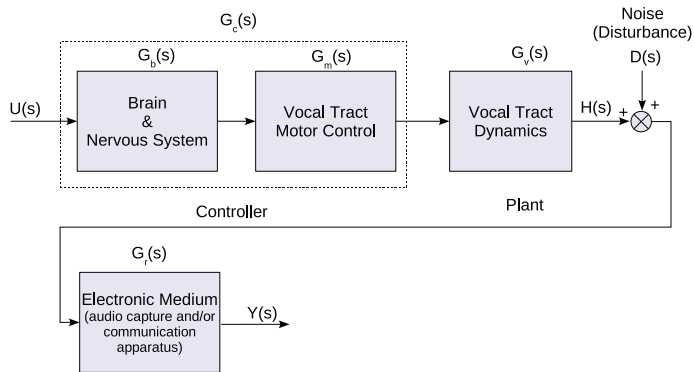


Fig. 2. Control system representation from (Beigi, 2011)

in the brain associated with generating the motor control signals and $G_m(s)$ which is the part of the nervous system associated with delivering the signal to the muscles in the vocal tract. The output of $G_c(s)$ is delivered to the vocal tract which is seen here as the plant. It is called $G_v(s)$ and it includes the moving parts of the vocal tract which are responsible for creating speech. The output, $H(s)$, is the *Laplace transform* of the speech wave, exciting the transmission medium, namely air. At this point we may model all the noise components and disturbances which may be present in the air surrounding the generated speech. The resulting signal is then transformed by passing through some type of electronic medium through audio capture and communication. The resulting signal, $Y(s)$ is the signal which is used to recognize the speaker.

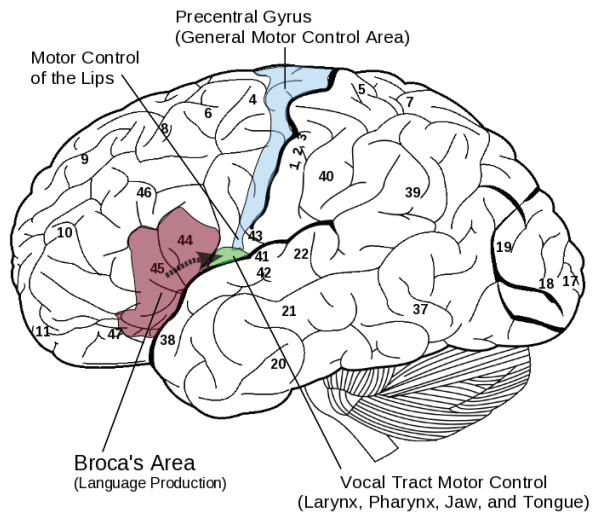


Fig. 3. Speech production in the Cerebral Cortex – from (Beigi, 2011)

Figure 3, borrowed from (Beigi, 2011), shows the superimposition of the interesting parts of the brain associated with producing speech. *Broca's area* which is part of the frontal lobe is

associated with producing the *language code* necessary for *speech production*. It may be seen as a part of the $G_b(s)$ in the control system representation of speech production. The *Precentral Gyrus* shown in the blue color is a long strip in the frontal lobe which is responsible for our *motor control*. The lower part of this area which is adjacent to Broca's area is further split into two parts. The lower part of the blue section is responsible for lip movement. The green part is associated with control of our *larynx, Pharynx, jaw, and tongue*. Together, these parts make up part of the $G_m(s)$ which is the second box in the controller. Note the proximity of these control regions to Broca's area, which is the coding section. Due to the slow transmission of chemical signals, the brain has evolved to allow for messages to travel quickly from $G_b(s)$ to $G_m(s)$ by utilizing proximity.

Note that *Broca's area* is also connected to the *language perception* section known as *Wernicke's area*. This will allow the feedback and refinement of the outgoing message. The vocal tract produces a carrier signal, based on its inherent dynamics, which is modified by the signal being generated by the $G_c(s)$. This is the actual plant which was called $G_v(s)$ in the control paradigm (Figure 2).

In *text-independent speaker recognition*, we are only concerned with learning the characteristics of the carrier signal in $G_v(s)$. *Speech recognition*, on the other hand, is concerned with decoding the intended message produced by Broca's area. This is why the signal processing is quite similar between the two disciplines, but in essence each discipline is concerned with a different part of the signal. The total time-signal is therefore a convolution of these two signals. The separation of these convolved signals is quite challenging and the results are therefore tainted in both disciplines causing a major part of the recognition error. Other sources are due to many complex disturbances along the way.

Figure 4 shows the major portion of the vocal tract which begins with the trachea and ends at the mouth and at the nose. It has a very plastic shape in which many of the cavities can change their shapes to be able to adjust the plant dynamics of Figure 2.

6. Theory and current approaches

The plasticity of the shape of the vocal tract makes the speech signal a non-stationary signal. This means that any segment of it, when compared to an adjacent segment in the time domain, has substantially different characteristics, indicating that the dynamics of the system producing these sections varies with time.

As mentioned in the *Introduction*, the first step is to store the vocal characteristics of the speakers in the form of speaker models in a database, for future reference. To build these models, certain features should be defined such that they would best represent the vocal characteristics of the speaker of interest. The most prevalent features used in the field happen to be identical to those used for speech recognition, namely, Mel Frequency Cepstral Coefficients (MFCCs) – see (Beigi, 2011).

6.1 Sampling

A Discrete representation of the signal is used for Automatic Speaker Recognition. Therefore we need to utilize the sampling theorem to help us determine the appropriate sampling frequency to be used for converting the continuous speech signal into its discrete signal representation.

One must therefore ensure that the sampling rate is picked in accordance with the guidelines set by the *Whittaker-Kotelnikoff-Shannon (WKS)* sampling theorem (Beigi, 2011). The WKS sampling theorem requires that the sampling frequency be at least two times the *Nyquist*

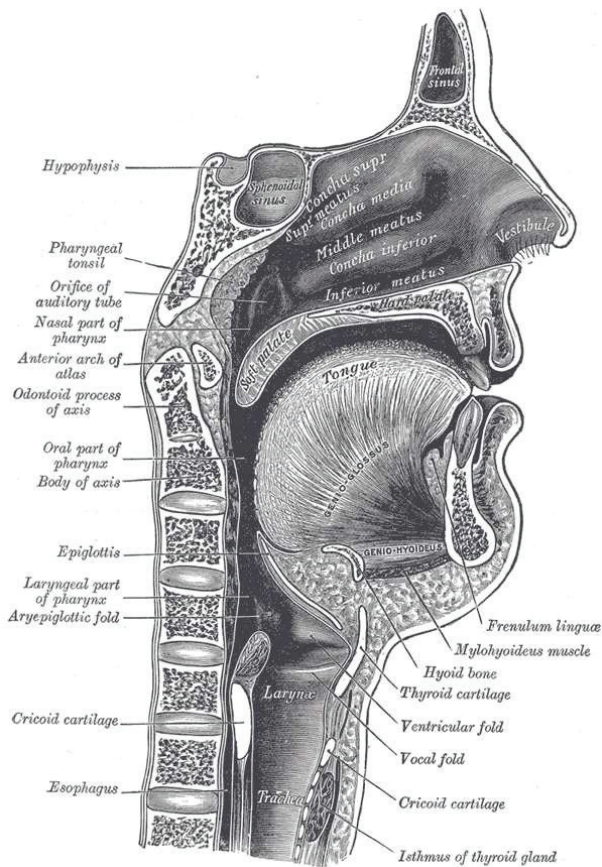


Fig. 4. Sagittal section of Nose, Mouth, Pharynx, and Larynx; Source: *Gray's Anatomy* (Gray, 1918)

Critical frequency. The *Nyquist critical frequency* is really the highest frequency content of the analog signal. For simplicity, normally an *ideal sampler* is used, which acts like the multiplication of an impulse train with the analog signal, where the impulses happen at the chosen sampling frequency.

In this representation, each sample has a zero width and lasts for an instant. The sampling theorem may be stated in words by requiring that the sampling frequency be greater than or equal to the *Nyquist rate*. The *Nyquist rate*, is defined as two times the *Nyquist critical frequency*.

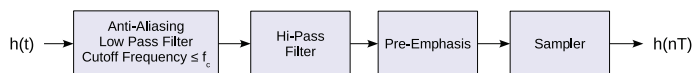


Fig. 5. Block diagram of a typical sampling process

Figure 5 shows a typical sampling process which starts with an analog signal and produces a series of discrete samples at a fixed frequency, representing the speech signal. The discrete samples are usually stored using a Codec (Coder/Decoder) format such as *linear PCM*, *μ -Law*,

a-Law, etc. Standardization is quite important for interoperability with different recognition engines (Beigi & Markowitz, 2010).

There are different forms for representing the speech signal. The simplest one is the *speech waveform* which is basically the plot of the sampled points versus time. In general, the amplitude is normalized to dwell between -1 and 1 . In its quantized form, the data is stored in the range associated with the quantization representation. For example, for a *16-bit signed linear PCM*, it would go from -32768 to 32767 .

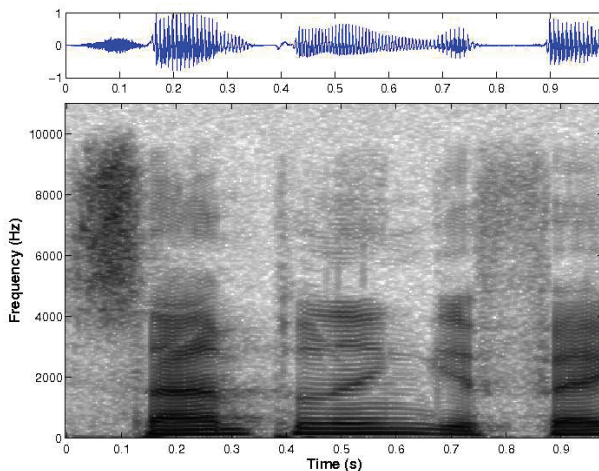


Fig. 6. Narrowband spectrogram of a speech signal

Another representation is, so-called, the *spectrogram* of the signal. Figure 6 shows the *narrowband spectrogram* of a signal. A sliding window of 23 ms was used for to generate this figure. The spectrogram shows the frequency content of the speech signal as a function of time. It is really a three-dimensional representation where the z-axis is depicted by the darkness of the points on the figure. The darker the pixel, the stronger the signal strength in that frequency for the time slice of choice. An artifact of the narrowband spectrogram is the existence of the horizontal curved lines across time. A *speech waveform* representation has also been plotted on top of the spectrogram of Figure 6 to show the relation between different points in the waveform with their corresponding frequency content in the spectrogram.

The system of Figure 5 should be designed so that it reduces *aliasing*, *truncation*, *band-limitation*, and *jitter* by choosing the right parameters, such as the sampling rate and volume normalization. Figure 7 shows how most of the fricative information is lost going from a 22 kHz sampling rate to 8 kHz. Normal telephone sampling rates are at best 8 kHz. Mostly everyone is familiar with having to qualify fricatives on the telephone by using statements such as “S” as in “Sam” and “F” as in “Frank”.

6.2 Feature extraction

Cepstral coefficients have fallen out of studies in exploring the arrival of echos in nature (Bogert et al., 1963). They are related to the spectrum of the log of spectrum of a speech signal. The frequency domain of the signal in computing the *MFCCs* is warped to the *Melody (Mel)* scale. It is based on the premise that human perception of pitch is linear up to 1000 Hz and then becomes nonlinear for higher frequencies (somewhat logarithmic). There are models of the

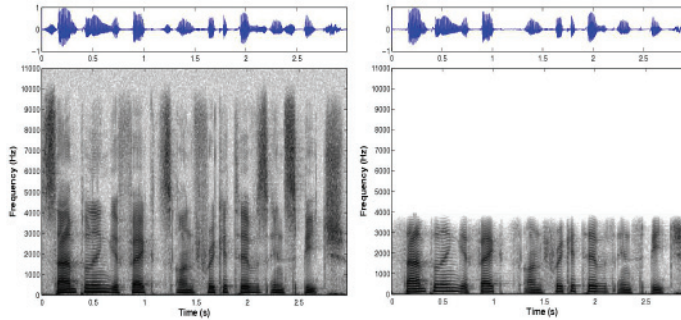


Fig. 7. Utterance: “Sampling Effects on Fricatives in Speech”, sampled at 22kHz (left) and 8kHz (right)

human perception based on other warped scales such as the *Bark scale*. There are several ways of computing Cepstral Coefficients. They may be computed using the *Direct Method*, also known as *Moving Average (MA)* which utilizes the *Fast Fourier Transform (FFT)* for the first pass and the *Discrete Cosine Transform (DCT)* for the second pass to ensure real coefficients. This method usually entails the following steps:

1. Framing – Selecting a sliding section of the signal with a fixed width in time which is then moved with some overlap. The sliding window is generally about *30ms* with an overlap of about *20ms* (*10ms* shift).
2. Windowing – A window such as a Hamming, Hann, Welch, etc. is used to smooth the signal for the computation of the *Discrete Fourier Transform (DFT)*.
3. FFT – The *Fast Fourier Transform (FFT)* is generally used for approximating the DFT of the windowed signal.
4. Frequency Warping – The FFT results are warped in the frequency domain in accordance with the Melody (Mel) or Bark scale.
5. MFCC – The Mel Frequency Cepstral Coefficients (MFCC) are computed.
6. Mel Cepstral Dynamics – Delta and Delta-Delta Cepstra are computed based on adjacent MFCC values.

Some use the *Linear Predictive*, also known as *AutoRegressive (AR)* features by themselves: *Linear Predictive Coefficients (LPC)*, *Partial Correlation (PARCOR)* – also known as *reflection coefficients*, or *log area ratios*. However, mostly the LPCs are converted to cepstral coefficients using autocorrelation techniques (Beigi, 2011). These are called Linear Predictive Cepstral Coefficients (LPCCs). There are also the Perceptual Linear Predictive (PLP) (Hermansky, 1990) features, shown in Figure 9. PLP works by warping the frequency and spectral magnitudes of the speech signal based on auditory perception tests. The domain is changed from magnitudes and frequencies to loudness and pitch (Beigi, 2011).

There have been an array of other features used such as *wavelet filterbanks* (Burrus et al., 1997), for example in the form of *Mel-Frequency Discrete Wavelet Coefficients* and *Wavelet Octave Coefficients of Residues (WOCOR)*. There are also *Instantaneous Amplitudes and Frequencies* which are in the form of *Amplitude Modulation (AM)* and *Frequency Modulation (FM)*. These features come in different flavors such as *Empirical Mode Decomposition (EMD)*, *FEPSTRUM*, *Mel Cepstrum Modulation Spectrum (MCMS)*, and so on (Beigi, 2011).

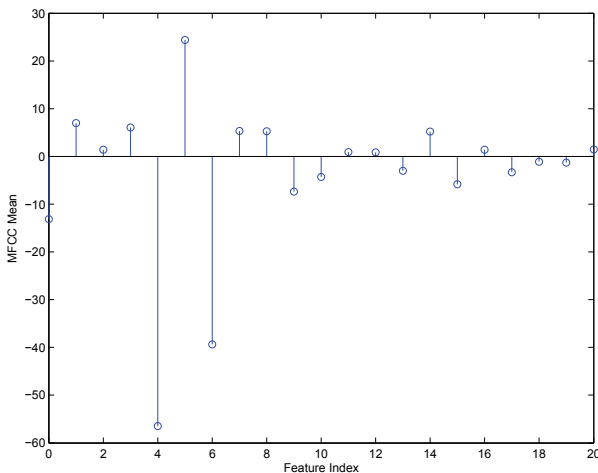


Fig. 8. A sample MFCC vector – from (Beigi, 2011)

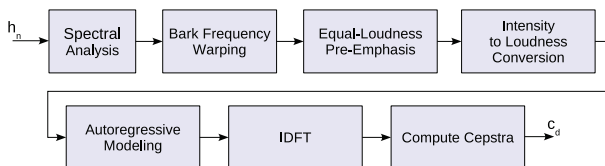


Fig. 9. A typical Perceptual Linear Predictive (PLP) system

It is important to note that most audio segments include a good deal of silence. Addition of features extracted from silent areas in the speech will increase the similarity of models, since silence does not carry any information about the speaker's vocal characteristics. Therefore, *Silence Detection (SD)* or *Voice Activity Detection (VAD)* (Beigi, 2011) is quite important for better results. Only segments with vocal signals should be considered for recognition. Other preprocessing such as *Audio Volume Estimation* and normalization and *Echo Cancellation* may also be necessary for obtaining desirable result (Beigi, 2011).

6.3 Speaker models

Once the features of interest are chosen, models are built based on these features to represent the speakers' vocal characteristics. At this point, depending on whether the system is text-dependent (including text-prompted) or text-independent, different methods may be used. Models are usually based on HMMs, GMMs, SVMs, and NNs.

6.3.1 Gaussian Mixture Models (GMM)

In general, there are many different modeling scenarios for speaker recognition. Most of these techniques are similar to those used for speech recognition modeling. For example, a *multi-state ergodic Hidden Markov Models* is usually used for text-dependent speaker recognition since there is textual context. As a special case of Hidden Markov Models, *Gaussian Mixture Models (GMM)* are used for doing text-independent speaker recognition. This is probably the most popular technique which is used in this field. GMMs are basically single-state degenerate HMMs.

The models are tied to the type of learning that is done. A popular technique is the use of a Gaussian Mixture Model (GMM) (Duda & Hart, 1973) to represent the speaker. This is mostly relevant to the text-independent case which encompasses speaker identification and text-independent verification. Even text-dependent techniques can use GMMs, but, they usually use a GMM to initialize Hidden Markov Models (HMMs) (Poritz, 1988) built to have an inherent model of the content of the speech as well. Many speaker diarization (segmentation and ID) systems use GMMs. To build a Gaussian Mixture Model of a speaker's speech, one should make a few assumptions and decisions. The first assumption is the number of Gaussians to use. This is dependent on the amount of data that is available and the dimensionality of the feature vectors.

Standard clustering techniques are usually used for the initial determination of the Gaussians. Once the number of Gaussians is determined, some large pool of features is used to train these Gaussians (learn the parameters). This step is called training. The models generated by training are called by many different names such as *background models*, *universal background models (UBM)*, *speaker independent models*, *Base models*, etc.

In a GMM, the models are parameters for collections of multi-variate normal density functions which describe the distribution of the Mel-Cepstral features (Beigi, 2011) for speakers' enrollment data. This distribution is represented by Equation 1.

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

where $\begin{cases} \mathbf{x}, \boldsymbol{\mu} \in \mathcal{R}^d \\ \boldsymbol{\Sigma} : \mathcal{R}^d \mapsto \mathcal{R}^d \end{cases}$

In Equation 1, $\boldsymbol{\mu}$ is the mean vector where,

$$\boldsymbol{\mu} \triangleq \mathcal{E} \{ \mathbf{x} \} \triangleq \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x}) d\mathbf{x} \quad (2)$$

The so-called "Sample Mean" approximation for Equation 2 is,

$$\boldsymbol{\mu} \approx \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{x}_i \quad (3)$$

where N is the number of samples and \mathbf{x}_i are the Mel-Cepstral feature vectors (Beigi, 2011).

The Variance-Covariance matrix of a multi-dimensional random variable is defined as,

$$\boldsymbol{\Sigma} \triangleq \mathcal{E} \left\{ (\mathbf{x} - \mathcal{E} \{ \mathbf{x} \}) (\mathbf{x} - \mathcal{E} \{ \mathbf{x} \})^T \right\} \quad (4)$$

$$= \mathcal{E} \left\{ \mathbf{x} \mathbf{x}^T \right\} - \boldsymbol{\mu} \boldsymbol{\mu}^T \quad (5)$$

This matrix is called the *Variance-Covariance* since the diagonal elements are the variances of the individual dimensions of the multi-dimensional vector, \mathbf{x} . The off-diagonal elements are the covariances across the different dimensions. Some have called this matrix the *Variance* matrix. Mostly in the field of Pattern Recognition it has been referred to, simply, as the *Covariance* matrix which is the name we will adopt here.

The Unbiased estimate of $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\Sigma}}$ is given by the following expression,

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{i=0}^{N-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \quad (6)$$

$$= \frac{1}{N-1} [\mathbf{S}_{xx} - N(\boldsymbol{\mu}\boldsymbol{\mu}^T)] \quad (7)$$

where the sample mean $\boldsymbol{\mu}$ is given by Equation 3 and the second order sum matrix, \mathbf{S}_{xx} is given by,

$$\mathbf{S}_{xx} = \sum_{i=0}^{N-1} \mathbf{x}_i \mathbf{x}_i^T \quad (8)$$

After the training is done, generally, the basis for a speaker independent model is built and stored in the form of the above statistics. At this stage, depending on whether a *universal background model* (UBM) (Reynolds et al., 2000) or *cohort models* are desired, different processing is done. For a UBM, a pool of speakers is used to optimize the parameters of the Gaussians as well as the mixture coefficients, using standard techniques such as *maximum likelihood estimation* (MLE), *Maximum a-Posteriori* (MAP) adaptation and *Maximum Likelihood Linear Regression* (MLLR). There may be one or more *Background models*. For example, some create a single background model called the UBM, others may build one for each gender, by using separate male and female databases for the training. *Cohort models* (Beigi et al., 1999) are built in a similar fashion. A cohort is a set of speakers that have similar vocal characteristics to the target speaker. This information may be used as a basis to either train a Hidden Markov Model including textual context, or to do an expectation maximization in order to come up with the statistics for the underlying model.

At this point, the system is ready for performing the enrollment. The enrollment may be done by taking a sample audio of the target speaker and adapting it to be optimal for fitting this sample. This ensures that the likelihoods returned by matching the same sample with the modified model would be maximal.

6.3.2 Support vector machines

Support vector machines (SVMs) have been recently used quite often in research papers regarding speaker recognition. Although they show very promising results, most of the implementations suffer from huge optimization problems with large dimensionality which have to be solved at the training stage. Results are not substantially different from GMM techniques and in general it may not be warranted to use such costly optimization calculations.

The claim-to-fame of *support vector machines* (SVMs) is that they determine the boundaries of classes, based on the training data, and they have the capability of *maximizing the margin* of class separability in the feature space. (Boser et al., 1992) states that the number of parameters used in a support vector machine is automatically computed (see *Vapnik-Chervonenkis* (VC) *dimension* (Borges, 1998; Vapnik, 1998)) to present a solution in terms of a linear combination of a subset of observed (training) vectors, which are located closest to the decision boundary. These vectors are called *support vectors* and the model is known as a *support vector machine*.

Vapnik (Vapnik, 1979) pioneered the statistical learning theory of SVMs, which is based on minimizing the classification error of both the training data and some unknown (held-out) data. Of course, the core of support vector machines and other *kernel* techniques stems from

much earlier work on setting up and solving *integral equations*. Hilbert (Hilbert, 1912) was one of the main developers of the formulation of *integral equations* and *kernel transformations*.

One of the major problems with SVMs is their intensive need for memory and computation power at the training stage. Training of SVMs for speaker recognition also suffers from these limitations. To address this issue, new techniques have been developed to split the problem into smaller subproblems which would then be solved in parallel as a network of problems. One such technique is known as *cascade SVM* (Tveit & Engum, 2003) for which certain improvements have also been proposed in the literature (Zhang et al., 2005).

Some of the shortcomings of SVMs have been addressed by combining them with other learning techniques such as *fuzzy logic* and *decision trees*. Also, to speed up the training process, several techniques based on the decomposition of the problem and selective use of the training data have been proposed.

In application to speaker recognition, experimental results have shown that SVM implementations of speaker recognition are slightly inferior to GMM approaches. However, it has also been noted that systems which combine GMM and SVM approaches often enjoy a higher accuracy, suggesting that part of the information revealed by the two approaches may be complementary (Solomonoff et al., 2004). For a detailed coverage, see (Beigi, 2011).

In general SVMs are two-class classifiers. That's why they are suitable for the speaker verification problem which is a two-class problem of comparing the voice of an individual to his/her model versus a background population model. N-class classification problems such as speaker identification have to be reduced to N two-class classification problems where the i^{th} two-class problem compares the i^{th} class with the rest of the classes combined (Vapnik, 1998). This can become quite computationally intensive for large-scale speaker identification problems. Another problem is that the Kernel function being used by SVMs is almost magically chosen.

6.3.3 Neural networks

Another modeling paradigm is the neural network perspective. There are quite a number of different neural networks and related architectures such as feed forward networks, TDNNs, probabilistic random access memory or pRAM models, Hierarchical Mixtures of Experts or HMEs, etc. It would take an enormous amount of time to go through all these and other possibilities. See (Beigi, 2011) for details.

6.3.4 Model adaptation (enrollment)

For a new person being enrolled in the system, the base speaker-independent models are modified to match the *a-posteriori statistics* of the enrolled person or target speaker's sample enrollment speech. This is done by any technique such as *maximum a-posteriori probability estimation (MAP)*, for example, using *expectation maximization (EM)*, or *maximum likelihood linear regression* for text-independent systems or simply by modifying the counts of the transitions on a *hidden Markov model (HMM)* for text-dependent systems.

7. Speaker recognition

At the identification and verification stage, a new sample is obtained for the test speaker. In the identification process, the sample is used to compute the likelihood of this sample being generated by the different models in the database. The identity of the model that returns the highest likelihood is returned as the identity of the test speaker. In identification, the

results are usually ranked by these likelihoods. To ensure a good dynamic range and better discrimination capability, log of the likelihood is computed.

At the verification stage, the process becomes very similar to the identification process described earlier, with the exception that instead of computing the log likelihood for all the models in the database, the sample is only compared to the model of the target speaker and the background or cohort models. If the target speaker model provides a better log likelihood, the test speaker is verified and otherwise rejected. The comparison is done using the *Log Likelihood Ratio (LLR)* test.

An extension of speaker recognition is diarization which includes segmentation followed by speaker identification and sometimes verification. The segmentation finds abrupt changes in the audio stream. *Bayesian Information Criterion (BIC)* (Chen & Gopalakrishnan, 1998) and *Generalized Likelihood Ratio (GLR)* techniques and their combination (Ajmera & McCowan, 2004) as well as other techniques (Beigi & Maes, 1998) have been used for the initial segmentation of the audio. Once the initial segmentation is done, a limited speaker identification procedure allows for tagging of the different parts with different labels. Figure 10 shows such a results for a two-speaker segmentation.

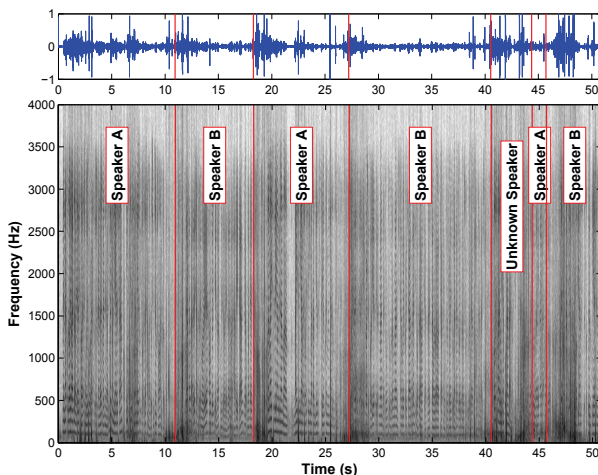


Fig. 10. Segmentation and labeling of two speakers in a conversation using turn detection followed by identification

7.1 Representation of results

Speaker identification results are usually presented in terms of the error rate. They may also be presented as the error rate based on the true result being present in the top N matches. This case is usually more prevalent in the cases where identification is used to prune a large set of speakers to only a handful of possible matches so that another expert system (human or machine) would finalize the decision process.

In the case of speaker verification, the method of presenting the results is somewhat more controversial. In the early days in the field, a *Receiver Operating Characteristic (ROC)* curve was used (Beigi, 2011). For the past decade, the *Detection Error Trade-Off (DET)* curve (Martin et al., 1997; Martin & Przybocki, 2000) has been more prevalent, with a measurement of the cost of producing the results, called the *Detection Cost Function (DCF)* (Martin & Przybocki, 2000).

Figures 11 and 12 show sample DET curves for two sets of data underscoring the difference in performances. Recognition results are usually quite data-dependent. The next section will speak about some open problems which degrade results.

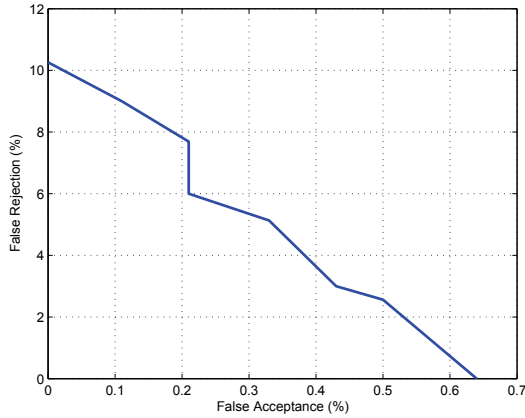


Fig. 11. DET Curve for quality data

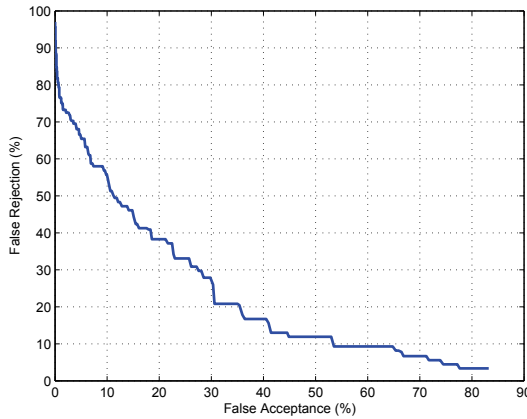


Fig. 12. DET Curve for highly mismatched and noisy data

There is a controversial operating point on the DET curve which is usually marked as the point of comparison between different results. This point is called the *Equal Error Rate (EER)* and signifies the operating point where the false rejection rate and the false acceptance rate are equal. This point does not carry any real preferential information about the “correct” or “desired” operating point. It is mostly a point of convenience which is easy to denote on the curve.

8. State of the art

In designing a practical speaker recognition system, one should try to affect the interaction between the speaker and the engine to be able to capture as many vowels as possible. Vowels are periodic signals which carry much more information about the resonance subtleties of the vocal tract. In the text-dependent and text-prompted cases, this may be done by actively designing prompts that include more vowels. For text-independent cases, the simplest way is to require more audio in hopes that many vowels would be present. Also, when speech recognition and natural language understanding modules are included (Figure 1), the conversation may be designed to allow for higher vowel production by the speaker.

As mentioned earlier, the greatest challenge in speaker recognition is *channel-mismatch*. Considering the general communication system given by Figure 13, it is apparent that the channel and noise characteristics at the time of communication are modulated with the original signal. Removing these channel effects is the most important problem in information theory. This is of course a problem where the goal is to recognize the message being sent. It is, however, a much bigger problem when the quest is the estimation of the model that generated the message – as it is with the speaker recognition problem. In that case, the channel characteristics have mixed in with the model characteristics and their separation is nearly impossible. Once the same source is transmitted over an entirely different channel with its own noise characteristics, the problem of learning the source model becomes even harder.

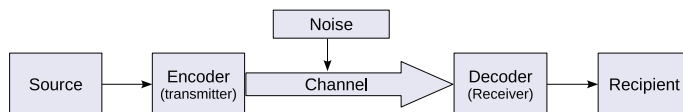


Fig. 13. One-way communication

Many techniques are used for resolving this problem, but it is still the most important source of errors in speaker recognition. It is the reason why most systems that have been trained on a predetermined set of channels, such as landline telephone, could fail miserably when cellular (mobile) telephones are used. The techniques that are being used in the industry are listed here, but there are more techniques being introduced every day:

- **Spectral Filtering and Cepstral Liftering**
 - Cepstral Mean Subtraction (CMS) or Cepstral Mean Normalization (CMN) (Benesty et al., 2008)
 - Cepstral Mean and Variance Normalization (CMVN) (Benesty et al., 2008)
 - Histogram Equalization (HEQ) (de la Torre et al., 2005) and Cepstral Histogram Normalization (CHN) (Benesty et al., 2008)
 - AutoRegressive Moving Average (ARMA) (Benesty et al., 2008)
 - RelATive SpecTrAl (RASTA) Filtering (Hermansky, 1991; van Vuuren, 1996)
 - J-RASTA (Hardt & Fellbaum, 1997)
 - Kalman Filtering (Kim, 2002)
- **Other Techniques**
 - Vocal Tract Length Normalization (VTLN) – first introduced for speech recognition: (Chau et al., 2001) and later for speaker recognition (Grashey & Geibler, 2006)
 - Feature Warping (Pelecanos & Sridharan, 2001)

- Feature Mapping (Reynolds, 2003)
- Speaker Model Synthesis (SMS) (R. et al., 2000)
- Speaker Model Normalization (Beigi, 2011)
- H-Norm (Handset Normalization) (Dunn et al., 2000)
- Z-Norm and T-Norm (Auckenthaler et al., 2000)
- Joint Factor Analysis (JFA) (Kenny, 2005)
- Nuisance Attribute Projection (NAP) (Solomonoff et al., 2004)
- Total Variability (i-vector) (Dehak et al., 2009)

Recently, depending on whether GMMs are used or SVMs, the two techniques of *joint factor analysis (JFA)* and *nuisance attribute projection (NAP)* have been used respectively, in most research reports.

Joint factor analysis (JFA) (Kenny, 2005) is based on factor analysis (FA) (Jolliffe, 2002). FA is a linear transformation which makes the assumption of having an *explicit model* which differentiates it from principal component analysis (PCA) and linear discriminant analysis (LDA). In fact in some perspective, it may be seen as a more general version of PCA. FA assumes that the underlying random variable is composed of two different components.

The first component is a random variable, called the *common factors*, which has a lower dimensionality compared to the combined random state, X , and the observation, Y . It is called the vector of *common factors* since the same vector, $\Theta : \mathcal{R}^1 \mapsto \mathcal{R}^M, M \leq D$, is a component of all the samples of \mathbf{y}_n .

The second component is the, so called, vector of *specific factors*, or sometimes called the *error* or the *residual* vector. It is denoted by $E : (\mathbf{e})_{D1}$. Therefore, this linear FA model for a specific random variable, $\tilde{Y} : \tilde{\mathbf{y}} : \mathcal{R}^q \mapsto \mathcal{R}^D$, related to the observed random variable Y may be written as follows,

$$\tilde{\mathbf{y}}_n = \mathbf{V}\boldsymbol{\theta}_n + \mathbf{e}_n \quad (9)$$

where $\mathbf{V} : \mathcal{R}^M \mapsto \mathcal{R}^D$ is known as the *factor loading* matrix and its elements, $(\mathbf{V})_{dm'}$, are known as the *factor loadings*. Samples of random variable $\Theta : (\boldsymbol{\theta}_n)_{M1}, n \in \{1, 2, \dots, N\}$ are known as the vectors of *common factors*, since due to the linear combination nature of the factor loading matrix, each element, $(\boldsymbol{\theta}_n)_{m'}$, has a hand in shaping the value of (generally) all $(\tilde{\mathbf{y}}_n)_d, d \in \{1, 2, \dots, D\}$. Samples of random variable $E : \mathbf{e}_n, n \in \{1, 2, \dots, N\}$ are known as vectors of *specific factors*, since each element, $(\mathbf{e}_n)_d$ is specifically related to a corresponding, $(\tilde{\mathbf{y}}_n)_d$.

JFA uses the concept of FA to split the space of the model parameters into speaker model parameters and channel parameters. It makes the assumption that the channel parameters are normally distributed, have a smaller dimensionality, and are common to all training samples. The model parameters, on the other hand, are common for each speaker. This separation allows for learning the channel characteristics in the form of separate model parameters, hence producing pure and somewhat channel-independent speaker models.

Nuisance attribute projection (NAP) (Solomonoff et al., 2004) is a method of modifying the original *kernel*, being used for the *support vector machine (SVM)* formulation, to one with the capability of telling specific channel information apart. The premise behind this approach is that by doing so, in both training and recognition stages, the system will not have the ability to distinguish channel specific information. This channel specific information is what is dubbed nuisance by (Solomonoff et al., 2004). NAP is a projection technique which assumes that most of the information related to the channel is stored in specific low-dimensional subspaces

of the higher dimensional space to which the original features are mapped. Furthermore, these regions are assumed to be somewhat distinct from the regions which carry speaker information.

Some even more recent developments have been made in speaker modeling. The *identity vector* or *i-vector* is a new representation of a speaker in a space of speakers called the *total variability space*. This model came from an observation by (Dehak et al., 2009) that the *channel space* in JFA still contained some information which may be used to distinguish speakers. This triggered the following representation of the GMM supervector of means ($\boldsymbol{\mu}$) which contains both speaker- and channel-dependent information.

$$\boldsymbol{\mu} = \boldsymbol{\mu}_I + \mathbf{T}\boldsymbol{\omega} \quad (10)$$

In Equation 10, $\boldsymbol{\mu}$ is assumed to be *normally distributed* with $\mathcal{E}\{\boldsymbol{\mu}\} = \boldsymbol{\mu}_I$, where $\boldsymbol{\mu}_I$ is the GMM supervector computed over the speaker- and channel-independent model which may be chosen to be the *universal background model*. The covariance for $\boldsymbol{\mu}$ is assumed to be $\text{Cov}(\boldsymbol{\mu}) = \mathbf{T}\mathbf{T}^T$, where \mathbf{T} is a low-rank matrix, and $\boldsymbol{\omega}$ is the *i-vector* which is a standard normally distributed vector ($p(\boldsymbol{\omega}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$). The *i-vector* represents the coordinates of the speaker in the, so-called, *total variability space*.

9. Future of the research

There are many challenges that have not been fully addressed in different branches of speaker recognition. For example, the large-scale speaker identification problem is one that is quite hard to handle. In most cases when researchers speak of large-scale in the identification arena, they speak of a few thousands of enrolled speakers. As the number of speakers increases to millions or even billions, the problem becomes quite challenging. As the number of speakers increases, doing an exhaustive match through the whole population becomes almost computationally implausible. Hierarchical techniques (Beigi et al., 1999) would have to be utilized to handle such cases. In addition, the speaker space is really a continuum. This means that if one considers a space where speakers who are closer in their vocal characteristics would be placed near each other in that space, then as the number of enrolled speakers increases, there will always be a new person that would fill in the space between any two neighboring speakers. Since there are intra-speaker variabilities (differences between different samples taken from the same speaker), the intra-speaker variability will be at some point more than inter-speaker variabilities, causing confusion and eventually identification errors. Since there are presently no large databases (in the order of millions and higher), there is no indication of the results, both in terms of the speed of processing and accuracy.

Another challenge is the fact that over time, the voice of speakers may change due to many different reasons such as illness, stress, aging, etc. One way to handle this problem is to have models which constantly adapt to changes (Beigi, 2009).

Yet another problem plagues speaker verification. Neither background models nor cohort models are error-free. Background models generally smooth out many models and unless the speaker is considerably different from the norm, they may score better than the speaker's own model. This is especially true if one considers the fact that nature is usually Gaussian and that there is a high chance that the speaker's characteristics are close to the smooth background model. If one were to only test the target sample on the target model, this would not be a problem. But since a test sample which is different from the target sample (used for creating the model) is used, the intra-speaker variability might be larger than the inter-speaker variability between the test speech and the smooth background model.

There are, of course, many other open problems. Some of these problems have to do with acceptable noise levels until break-down occurs. Using a cellular telephone with its inherently bandlimited characteristics in a very noisy venue such as a subway (metro) station is one such challenge.

Given the number of different operating conditions in invoking speaker recognition, it is quite difficult for technology vendors to provide objective performance results. Results are usually quite data-dependent and different data sets may pronounce particular merits and downfalls of each provider's algorithms and implementation. A good speaker verification system may easily achieve an 0% EER for clean data with good inter-speaker variability in contrast with intra-speaker variability. It is quite normal for the same "good" system to show very high equal error rates under severe conditions such as high noise levels, bandwidth limitation, and small relative inter-speaker variability compared to intra-speaker variability. However, under most controlled conditions, equal error rates below 5% are readily achieved. Similar variability in performance exists in other branches of speaker recognition, such as identification, etc.

10. References

- Ajmera, J. & McCowan, I. and Bourlard, H. (2004). Robust speaker change detection, *IEEE Signal Processing Letters* 11(8): 649–651.
- Auckenthaler, R., Carey, M. & Lloyd-Thomas, H. (2000). Score normalization for text-independent speaker verification systems, *Digital Signal Processing* 10(1–3): 42–54.
- Beigi, H. (2009). Effects of time lapse on speaker recognition results, *16th International Conference on Digital Signal Processing*, pp. 1–6.
- Beigi, H. (2011). *Fundamentals of Speaker Recognition*, Springer, New York. ISBN: 978-0-387-77591-3.
- Beigi, H. & Markowitz, J. (2010). Standard audio format encapsulation (safe), *Telecommunication Systems* pp. 1–8. 10.1007/s11235-010-9315-1.
URL: <http://dx.doi.org/10.1007/s11235-010-9315-1>
- Beigi, H. S., Maes, S. H., Chaudhari, U. V. & Sorensen, J. S. (1999). A hierarchical approach to large-scale speaker recognition, *EuroSpeech 1999*, Vol. 5, pp. 2203–2206.
- Beigi, H. S. & Maes, S. S. (1998). Speaker, channel and environment change detection, *Proceedings of the World Congress on Automation (WAC1998)*.
- Benesty, J., Sondhi, M. M. & Huang, Y. (2008). *Handbook of Speech Processing*, Springer, New York. ISBN: 978-3-540-49125-5.
- Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacrétaz, D. & Reynolds, D. (2004). A tutorial on text-independent speaker verification, *EURASIP Journal on Applied Signal Processing* 2004(4): 430–451.
- Bogert, B. P., Healy, M. J. R. & Tukey, J. W. (1963). The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking, in M. Rosenblatt (ed.), *Time Series Analysis*, pp. 209–243. Ch. 15.
- Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers, *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* 2: 121–167.

- Burrus, C. S., Gopinath, R. A. & Guo, H. (1997). *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, New York. ISBN: 0-13-89600-9.
- Campbell, J.P., J. a. (1997). Speaker recognition: a tutorial, *Proceedings of the IEEE* 85(9): 1437–1462.
- Chau, C. K., Lai, C. S. & Shi, B. E. (2001). Feature vs. model based vocal tract length normalization for a speech recognition-based interactive toy, *Active Media Technology, Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 134–143. ISBN: 978-3-540-43035-3.
- Chen, S. S. & Gopalakrishnan, P. S. (1998). Speaker, environment and channel change detection and clustering via the bayesian information criterion, *IBM Technical Report, T.J. Watson Research Center*.
- Dehak, N., Dehak, R., Kenny, P., Brummer, N., Ouellet, P & Dumouchel, P. (2009). Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification, *Interspeech*, pp. 1559–1562.
- de la Torre, A., Peinado, A. M., Segura, J. C., Perez-Cordoba, J. L., Benitez, M. C. & Rubio, A. J. (2005). Histogram equalization of speech representation for robust speech recognition, *IEEE Transaction of Speech and Audio Processing* 13(3): 355–366.
- Duda, R. O. & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York. ISBN: 0-471-22361-1.
- Dunn, R. B., Reynolds, D. A. & Quatieri, T. F. (2000). Approaches to speaker detection and tracking in conversational speech, *Digital Signal Processing* 10: 92–112.
- Furui, S. (2005). 50 years of progress in speech and speaker recognition, *Proc. SPECOM*, pp. 1–9.
- Grashey, S. & Geibler, C. (2006). Using a vocal tract length related parameter for speaker recognition, *Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006: The*, pp. 1–5.
- Gray, H. (1918). *Anatomy of the Human Body*, 20th edn, LEA and FEBIGER, Philadelphia. Online version, New York (2000).
URL: <http://www.Bartleby.com>
- Hardt, D. & Fellbaum, K. (1997). Spectral subtraction and rasta-filtering in text-dependent hmm-based speaker verification, *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, Vol. 2, pp. 867–870.
- Hermansky, H. (1990). Perceptual linear predictive (plp) analysis of speech, 87(4): 1738–1752.
- Hermansky, H. (1991). Compensation for the effect of the communication channel in the auditory-like analysis of speech (rasta-plp), *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-91)*, pp. 1367–1370.
- Hilbert, D. (1912). *Grundzüge Einer Allgemeinen Theorie der Linearen Integralgleichungen (Outlines of a General Theory of Linear Integral Equations)*, Fortschritte der Mathematischen Wissenschaften, heft 3 (Progress in Mathematical Sciences, issue 3), B.G. Teubner, Leipzig and Berlin. In German. Originally published in 1904.
- Jolliffe, I. (2002). *Principal Component Analysis*, 2nd edn, Springer, New York.
- Kenny, P. (2005). Joint factor analysis of speaker and session variability: Theory and algorithms, *Technical report*, CRIM.
URL: <http://www.crim.ca/perso/patrick.kenny/FAtheory.pdf>
- Kim, N. S. (2002). Feature domain compensation of nonstationary noise for robust speech recognition, *Speech Communication* 37(3–4): 59–73.

- Manning, C. D. (1999). *Foundations of Statistical Natural Language Processing*, The MIT Press, Boston. ISBN: 0-26-213360-1.
- Martin, A., Doddington, G., Kamm, T., Ordowski, M. & Przybocki, M. (1997). The det curve in assessment of detection task performance, *Eurospeech 1997*, pp. 1–8.
- Martin, A. & Przybocki, M. (2000). The nist 1999 speaker recognition evaluation – an overview, *Digital Signal Processing* 10: 1–18.
- Nolan, F. (1983). *The phonetic bases of speaker recognition*, Cambridge University Press, New York. ISBN: 0-521-24486-2.
- Pelecanos, J. & Sridharan, S. (2001). Feature warping for robust speaker verification, *A Speaker Odyssey - The Speaker Recognition Workshop*, pp. 213–218.
- Pollack, I., Pickett, J. M. & Sumbly, W. (1954). On the identification of speakers by voice, *Journal of the Acoustical Society of America* 26: 403–406.
- Poritz, A. B. (1988). Hidden markov models: a guided tour, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP-1988)*, Vol. 1, pp. 7–13.
- R., T., B., S. & L., H. (2000). A model-based transformational approach to robust speaker recognition, *International Conference on Spoken Language Processing*, Vol. 2, pp. 495–498.
- Reynolds, D. A. (2003). Channel robust speaker verification via feature mapping, *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, Vol. 2, pp. II-53–6.
- Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models, *Digital Signal Processing* 10: 19–41.
- Shearme, J. N. & Holmes, J. N. (1959). An experiment concerning the recognition of voices, *Language and Speech* 2: 123–131.
- Solomonoff, A., Campbell, W. & Quillen, C. (2004). Channel compensation for svm speaker recognition, *The Speaker and Language Recognition Workshop Odyssey 2004*, Vol. 1, pp. 57–62.
- Tosi, O. I. (1979). *Voice Identification: Theory and Legal Applications*, University Park Press, Baltimore. ISBN: 978-0-839-11294-5.
- Tveit, A. & Engum, H. (2003). Parallelization of the incremental proximal support vector machine classifier using a heap-based tree topology, *Workshop on Parallel Distributed Computing for Machine Learning*.
- USC (2005). Disability census results for 2005, World Wide Web.
URL: <http://www.census.gov>
- van Vuuren, S. (1996). Comparison of text-independent speaker recognition methods on telephone speech with acoustic mismatch, *International Conference on Spoken Language Processing (ICSLP)*, pp. 784–787.
- Vapnik, V. N. (1979). *Estimation of Dependences Based on Empirical Data*, russian edn, Nauka, Moscow. English Translation: Springer-Verlag, New York, 1982.
- Vapnik, V. N. (1998). *Statistical learning theory*, John Wiley, New York. ISBN: 0-471-03003-1.
- Viswanathan, M., Beigi, H. S. & Maali, F. (2000). Information access using speech, speaker and face recognition, *IEEE International Conference on Multimedia and Expo (ICME2000)*.
- Zhang, J.-P., Li, Z.-W. & Yang, J. (2005). A parallel svm training algorithm on large-scale classification problems, *International Conference on Machine Learning and Cybernetics*, Vol. 3, pp. 1637–1641.

Finger Vein Recognition

Kejun Wang, Hui Ma, Oluwatoyin P. Popoola and Jingyu Li
*Pattern Recognition & Intelligent Systems Department, College of Automation,
Harbin Engineering University
China*

1. Introduction

Smart recognition of human identity for security and control is a global issue of concern in our world today. Financial losses due to identity theft can be severe, and the integrity of security systems compromised. Hence, automatic authentication systems for control have found application in criminal identification, autonomous vending and automated banking among others. Among the many authentication systems that have been proposed and implemented, finger vein biometrics is emerging as the foolproof method of automated personal identification. Finger vein is a unique physiological biometric for identifying individuals based on the physical characteristics and attributes of the vein patterns in the human finger. It is a fairly recent technological advance in the field of biometrics that is being applied to different fields such as medical, financial, law enforcement facilities and other applications where high levels of security or privacy is very important. This technology is impressive because it requires only small, relatively cheap single-chip design, and has a very fast identification process that is contact-less and of higher accuracy when compared with other identification biometrics like fingerprint, iris, facial and others. This higher accuracy rate of finger vein is not unconnected with the fact that finger vein patterns are virtually impossible to forge thus it has become one of the fastest growing new biometric technology that is quickly finding its way from research labs to commercial development.

Historically, R&D at Hitachi of Japan (1997-2000) discovered that finger vein pattern recognition was a viable biometric for personal authentication technology and by 2000-2005 were the first to commercialize the technology into different product forms, such as ATMs. Their research reports false acceptance rate (FAR) of as low as 0.0001 % and false reject rate (FRR) of 0.1%. Today 70% of major financial institutions in Japan are using finger vein authentication.



Fig. 1. Hitachi of Japan history of research & development on finger-vein recognition technology

Fingerprints have been the most widely used and trusted biometrics. The reasons being: the ease of acquiring fingerprints, the availability of inexpensive fingerprint sensors and a long history of its use. However, limitations like the deterioration of the epidermis of the fingers, finger surface particles etc result in inaccuracies that call for more accurate and robust methods of authentication. Vein recognition technology however offers a promising solution to these challenges due the following characteristics. (1) Its universality and uniqueness. Just as individuals have unique fingerprints, so also they do have unique finger vein images. The vein images of most people remain unchanged despite ageing. (2) Hand and finger vein detection methods do not have any known negative effects on body health. (3) The condition of the epidermis has no effect on the result of vein detection. (4) Vein features are difficult to be forged and changed even by surgery [1]. These desirable properties make vein recognition a highly reliable authentication method.

Vein object extraction is the first crucial step in the process. The aim is to obtain vein ridges from the background. Recognition performance relates largely to the quality of vein object extraction. The standard practice is to acquire finger vein images by use of near-infrared spectroscopy. When a finger is placed across near infra-red light rays of 760 nm wavelength, finger vein patterns in the subcutaneous tissue of the finger are captured because deoxygenated hemoglobin in the vein absorb the light rays. The resulting vein image appears darker than the other regions of the finger, because only the blood vessels absorb the rays. The extraction method has a direct impact on feature extraction and feature matching [2]. Therefore, vein object extraction significantly affects the effectiveness of the entire system.

2. Processing

After vein image extraction, comes segmentation. The traditional vein extraction technology can be classified into three broad categories according to their approach to segmentation i.e separating the actual finger veins from the background and noise. There are those based on region information, those based on edge information, and those based on particular theories and tools. However, the application of the traditional single-threshold segmentation methods such as fixed threshold, total mean, total Otsu to perform segmentation, faces limitations in obtaining the desired accurate segmentation results. Using multi-threshold methods like local mean and local Otsu, improve these results but still cannot effectively deal with noise and over-segmentation effects [3], [4], [5], [6], [7],[8]. In a related research, reference [9] proposed an oriented filter method to enhance the image in order to eliminate noise and enhance ridgeline. Authors in [10] used the directionality feature of fingerprint to present a fingerprint image enhancement method based on orientation field. These two methods take the directionality characteristic of fingerprints into account, so they can enhance and de-noise fingerprint images effectively. Finger vein pattern also has textural and directionality features, with directionality being consistent within the local area. Inspired by methods in [9] and [10], we discuss in this chapter, finger vein pattern extraction methods using oriented filtering from the directionality feature of veins. These utilize the directionality feature of finger vein images using a group of oriented filters, and then extracting the vein object from an enhanced oriented filter image.

2.1 Normalization

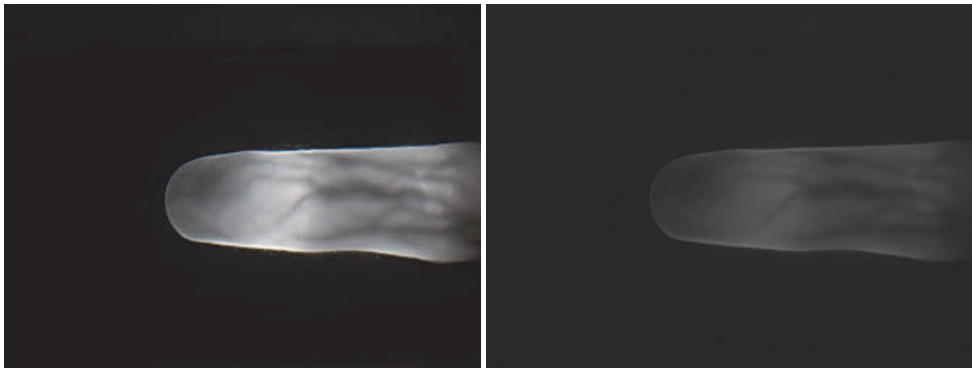
Normalization is a pixel-wise operation often used in image processing. The main purpose of normalization is to get an output image with desirable mean and variance, which facilitates the subsequent processing. The uniformly illuminated image becomes normalized by this formula:

$$M = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j) \quad (1)$$

$$VAR = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^2 \quad (2)$$

$$G(i, j) = \begin{cases} M_0 + \sqrt{\frac{VAR_0}{VAR}} \times (I(i, j) - M)^2, I(i, j) \geq M \\ M_0 - \sqrt{\frac{VAR_0}{VAR}} \times (I(i, j) - M)^2, I(i, j) < M \end{cases} \quad (3)$$

Where M and VAR denote the estimated mean and variance of input image and $M_0 = 150, VAR_0 = 255$ are desired mean and variance values respectively. After normalization, the output image is ready for next processing step. The result of the above-mentioned process is shown in Fig.2. We note that Fig.2 image has lost some of its contrast but now has uniform illumination.



(a) Original finger vein image (b) Normalized finger vein image

Fig. 2. Original image and Normalized image

2.2 Oriented filter enhancement

Finger vein pattern has directionality characteristics, which the traditional filter methods do not take into account; therefore, its resultant filtering enhancement is not satisfactory. We propose a vein pattern extraction method using oriented filtering technology that takes into account the directionality feature of the veins. This algorithm utilizes a group of oriented filters to filter the image depending on the orientation of the local ridge.

a. Calculation of directional image

The directional image is an image transform, where we use the direction of every pixel of an image to represent the original vein image. Pixels' direction refers to the orientation of continuous gray value. We can determine the direction of pixel according to the gray distribution of the neighborhood. Pixels along the vein ridge have minimum gray level difference while pixels perpendicular to the vein ridge have the maximum gray level difference.

To estimate the orientation field of vein image, the direction of the vein is quantized into eight directions. Using a 9×9 template window as shown in Fig.2, we choose reference pixel $p(i, j)$ the center of the direction template. The values 1-8 of the template correspond to eight directions rotating from 0 to π . Each interval has and angle of $\pi/8$ in an anti-clockwise direction from the horizontal axis.

The main steps of the method are:

1. For every pixel, use the 9×9 rectangular window (shown as Fig.2) to obtain the pixel gray value average \overline{M}_i ($i = 1, 2, \dots, 8$).
2. Divide \overline{M}_i ($i = 1, 2, \dots, 8$) into four groups. \overline{M}_1 and \overline{M}_5 are perpendicular in direction, so they belong to the same group. For the same reason, \overline{M}_2 and \overline{M}_6 , \overline{M}_3 and \overline{M}_7 , \overline{M}_4 and \overline{M}_8 belong to the same group. In each group, calculate ΔM - the absolute difference between two gray value averages \overline{M}_j and \overline{M}_{j+4} .

$$\Delta M = \left| \overline{M}_j - \overline{M}_{j+4} \right|, j = 1, \dots, 4 \quad (4)$$

where j is the direction of the vein.

3. Choose the maximum ΔM to determine the pixel's possible directions j_{\max} and $j_{\max} + 4$.
4. Determine the actual direction of $p(i, j)$ by comparing its gray value with the gray value averages of j_{\max} and $j_{\max} + 4$. The closer value is its direction. Therefore, the pixels direction is given by:

$$D(x, y) = \begin{cases} j_{\max}, & \text{if } \left| M - \overline{M}_j \right| < \left| M - \overline{M}_{j+4} \right| \\ j_{\max} + 4, & \text{otherwise} \end{cases} \quad (5)$$

When the above process is performed on each pixel in the image, we can obtain the directional image $D(x, y)$ of the vein image. Due to the presence of noise in the vein image, the estimated orientation field may not always be correct. In a small local neighborhood, the pixels' orientations are generally uniform; and so a local ridge orientation is specified for a block rather than at every pixel. Using a smoothing process on the point directional map, a continuous directional map is obtained. A continuous $w \times w$ window can be used to modify the incorrect ridge orientation and smooth the point directional map. The experiments show that $w = 8$ is very good. We obtain each window's directional histogram and choose the peak value of the direction histogram as $P(x, y)$'s orientation. The continuous directional map $O(x, y)$ is defined as:

$$O(x, y) = \text{ord}(\max(N_i)) \quad (6)$$

7		6		5		4		3
8		7	6	5	4	3		2
		8				2		
1		1		$p(i, j)$		1		1
		2				8		
2		3	4	5	6	7		8
3		4		5		6		7

Fig. 3. 9×9 rectangular window

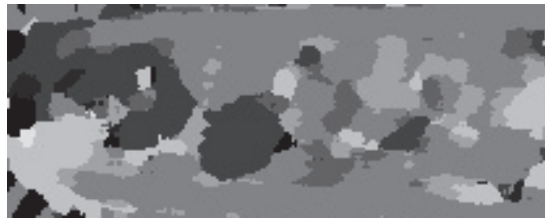


Fig. 4. Directional image of finger vein

Where $i = 1, 2, \dots, 8$, function $ord(*)$ is used to obtain the subscript of element $*$.

The directional image of finger vein is as shown in Fig.3 .Each color of directional image corresponds to a direction. As can be seen from the Fig.4, vein ridge has the feature of directionality, and the vein ridge orientation varies slowly in a local neighborhood.

b. Oriented filter

The vein directional image is a kind of textured pattern generated by using oriented filters based on directional map to enhance the original images. We designed eight filter masks, each one associated with the discrete ridge orientation of finger vein pixels. From the direction determined for a specific block (from the original image), a corresponding filter is selected to enhance this block image. The template coefficients of horizontal mask are designed first. To generate the seven other masks, the horizontal filter mask is rotated according to the direction of the vein. O’Gorman’s rules for filter design which is described for enhancing fingerprint images consists of four key points:

1. An appropriate filter template size.
2. The width of the filter template should be odd in order that the template is symmetric in the direction of horizontal and vertical.
3. In the vertical direction, central part of the filter template coefficient should be positive while both sides of the coefficient negative.
4. The sum of all template coefficients should be zero.

Applying the above rules based on the direction of the finger vein, we modify the filter’s coefficients so that they decay from the center to both ends of the template. The oriented filter’s size is decided according to vein ridge width. From experiments, a filter template of size 7 has been shown to be quite effective.

$-c/3$	$-2c/3$	$-c$	$-c$	$-c$	$-2c/3$	$-c/3$	-8	-16	-24	-24	-24	-16	-8
$b/3$	$2b/3$	b	b	b	$2b/3$	$b/3$	0	0	0	0	0	0	0
$a/3$	$2a/3$	a	a	a	$2a/3$	$a/3$	3	6	9	9	9	6	3
$d/3$	$2d/3$	d	d	d	$2d/3$	$d/3$	10	20	30	30	30	20	10
$a/3$	$2a/3$	a	a	a	$2a/3$	$a/3$	3	6	9	9	9	6	3
$b/3$	$2b/3$	b	b	b	$2b/3$	$b/3$	0	0	0	0	0	0	0
$-c/3$	$-2c/3$	$-c$	$-c$	$-c$	$-2c/3$	$-c/3$	-8	-16	-24	-24	-24	-16	-8

(a) Template coefficient of horizontal oriented filter (b) An example of oriented filter

Fig. 5. Template coefficient of horizontal oriented filter and an example of oriented filter

The coefficient spatial arrangement of the horizontal mask is shown in Fig.5.

- a. The coefficients of the filter template should meet the following given conditions: $d + 2a + 2b - 2c = 0$, where $d > a > d \geq 0$, $c > 0$; . An example of oriented filter is shown as Fig.5 (b). Now, aiming at every pixel (i, j) in the input image, we select a 3 neighborhood that takes (i, j) as the center. It is filtered with the mask that corresponds to the block orientation of the center (i, j) . This filtering technique is given by:

$$f(i, j) = \sum_{x=-3}^3 \sum_{y=-3}^3 G(i+x, j+y) g_{\theta}(x, y) \quad (7)$$

where (i, j) represents the pixel of original image, and x, y represents the size of oriented filter template and $g_{\theta}(x, y)$ represents the corresponding coefficients of the template. T_d represents the filtered image. It is possible that some gray values fall outside the $[0, 255]$ range. Equation (8) is used to adjust the gray values to fall within the range.

$$f'(i, j) = \text{Round}\left(\frac{f(i, j) - f_{\min}(i, j)}{f_{\max}(i, j) - f_{\min}(i, j)} \times 255\right) \quad (8)$$

Where $f(i, j)$ represents the original gray of (i, j) , $f_{\min}(i, j)$ represents smallest gray value of original image, $f_{\max}(i, j)$ represents biggest gray value of original image, $f'(i, j)$ represents transformed gray of (i, j) and $\text{Round}(\cdot)$ is a rounding function.

To generate the other seven masks, we rotate the horizontal filter mask according to the following equation: Where (i, j) represents the coordinates in the horizontal mask, and (i', j') represents the ones in the rotated mask.

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} i' \\ j' \end{bmatrix} \quad (9)$$

Where $\theta = (d-1)/\pi$, $d = 1, 2, \dots, 8$, θ represents the rotation angel, and d represents the direction value of (i, j) .

To use this method, (i, j) is usually not an integer, so we need to use nearest neighbor interpolation to get the coefficients of the rotated mask. $g_{\theta}(i', j')$ (the coefficients in the rotated mask) is equal to $g_1(i, j)$ (the coefficients in the horizontal mask).

Suppose the four points $(i_m, j_m), (i_m, j_n), (i_n, j_m), (i_n, j_n)$ compose a square centered at pixel (i, j) , and $g_\theta(i_m, j_m), g_\theta(i_m, j_n), g_\theta(i_n, j_m), g_\theta(i_n, j_n)$ are their corresponding coefficients in the horizontal mask, where $i_m < i < i_n, j_m < j < j_n$.

First, make interpolation between (i_m, j_n) and (i_n, j_n) :

$$g(i, j_n) = g_1(i_m, j_n) + (i - j_m) \times [g_1(i_n, j_n) - g_1(i_m, j_n)] \quad (10)$$

Then make interpolation between (i_m, j_m) and α :

$$g(i, j_m) = g_1(i_m, j_m) + (i - i_m) \times [g_1(i_n, j_m) - g_1(i_m, j_m)] \quad (11)$$

Finally, make interpolation between a_0 and b_0 :

$$g_\theta(i', j') = g(i, j_m) + (j - j_m) \times [g(i, j_n) - g(i, j_m)] \quad (12)$$

Once we have all eight filter masks, their coefficients can be used to enhance vein image.

2.3 Image segmentation

NiBlack segmentation method is a commonly used simple and effective local dynamic threshold algorithm, so we choose this method to segment the image. The basic idea of this algorithm is that for every pixel in the input image; calculate their mean and variance of its $r \times r$ neighborhood. Then the result of the following formula is used as threshold to segment image.

$$T(x, y) = m(x, y) + k \times s(x, y) \quad (13)$$



Fig. 6(a). Image enhanced by oriented filter



Fig. 6(b). Image after segmentation and noise removal

Where (x, y) represents pixel in the input image, $T(x, y)$ represents the threshold of $\forall a, a' \in M', a = H(x) = T_a(T_a^{-1}(a_0))$ represents the mean value of $a_0 = T_a^{-1}(T_a^{-1}(a))$'s $r \times r$ neighborhood, and $s(x, y)$ represents the standard deviation of $v(a') = v(a)$'s $a_0 = T_a(v(a'))$ neighborhood, k is a correction coefficient. The experiments show that $v(a) = T_a(v(a_0))$ equal to 9 and $v(a_0) = T_a^{-1}(v(a'))$ equal to 0.01 is excellent. There are some

small black blocks in the background and some small white holes on the target object in the segmented image. Such noise can be removed with area elimination method. Because of variability in image acquisition and the inherent differences in individual samples, the size and ratios of extracted finger veins are often inconsistent. In order to facilitate further research there is a need for standardization of the segmented vein image height (and width). The normalized image is shown in Fig.7. In the experiment, we have standardized the height of the image to 80 pixels.



Fig. 7. Standardized finger vein image width

Accurate extraction of finger vein pattern is a fundamental step in developing finger vein based biometric authentication systems. The finger vein pattern extraction method proposed and discussed above extends traditional image segmentation methods, by extracting vein object from the oriented filter enhanced image. The addition of oriented filter operation extracts smooth and continuous vein features from not only high quality vein images but also noisy low quality images and does not suffer from the over-segmentation problem.

3. Feature extraction, fusion and matching

Finger vein recognition as a feature for biometric recognition has excellent advantages such as being stable, contactless, unique, immune to counterfeiting, highly accurate etc. This makes finger-vein recognition widely considered as the most promising biometric technology for the future.

Naoto Miura [14] proposed one method for finger-vein recognition based on template matching. In the experiment, the finger-vein image is first binarized, and then using a distance transform noise is removed, and embedded hidden Markov model is used for finger-vein recognition. This approach is time intensive, and another major limitation is that it cannot recognize distorted finger-vein images correctly. Kejun Wang [15] combined wavelet moment, PCA and LDA transform for finger-vein recognition. Here the metric of finger-vein image is converted to a one-dimensional vector, which has been reduced dimensionally. To deal with the problem of high dimensionality, researchers usually first partition the finger-vein image and then principal component analysis (PCA) is applied. To date, this has been the most popular method for dimensionality reduction in finger-vein recognition research. Xueyan Li [16] proposed a method, which combines two-dimensional wavelet and texture characteristic, to recognize the finger vein while Xiaohua Qian [17] used seven moment invariant finger vein features. Euclidean distance and a pre-defined threshold were used as the classifying criterion for matching and recognition. Chengbo Yu [13] defined valley regions as finger vein features such that real features could not be missed and the false features would not be extracted. Zhongbo Zhang [18] proposed an algorithm based wavelet and neural network, which extracts features at multi-scale. Zhang's algorithm can capture features from degraded images.

3.1 Novel finger vein recognition methods based using fusion approach

The above-mentioned algorithms have different advantages for different problems in finger-vein recognition. However, because fingers have curved surfaces, finger vein diameter is not consistent and the texture characteristic is aperiodic. When near infrared light is used to acquire the image, the gray-scale is uneven and contrast is low; besides, finger veins are tiny and few in number, such that only very few features can be extracted. What is more, a change in the finger position can cause image translation and rotation and influence recognition negatively. To deal with these problems some novel fusion methods are used. First, we discuss a method based on relative distance and angle. This approach makes full use of the uniqueness of topology, the varied distances between the intersection points of two different vein images, and the differences in angles produced by these intersection points connections, all combined for recognition. This method overcomes the influence of image translation and rotation, because relative distance and angle don't change. Therefore, the method based on these identified characteristics has great use in practice.

3.1.1 Theoretical basis

Let a , b , and c be non-zero vectors. θ is the angle between two vectors, and the length of line segment is written as $\|a\|$. The thinned finger-vein image is illustrated as a function denoted as M , which is defined in field D ; where M is a subset of D . Image translation and rotation occurs in D . Image translation and rotation implies that every point of the image is translated by a vector and rotated by the some angle. The relative distances and angles remain constant before and after translation and rotation, which is proved as follows. The topology produced by all the character point connections can be called the image M . Then M can be shown by the vectors: $a = \{a_0, a_1, \dots, a_{n-1}\}$ $n \in N$.

Let $a[s] = \{a_s, a_{s+1}, \dots, a_{s+n-1}\}$, where $a[s]$ denotes a vector a , translated by s unit distances.

If $g_s = \langle a, a[s] \rangle$, where $\langle a, b \rangle$ denotes the inner product of a and b ; and $v(a) = (g_0, g_1, \dots, g_{n-1})$, where $v(a)$ is the convolution of a .

Now after image M is translated by s unit distances in the plane D , we get the image $M' = M[s]$.

A random translation of image M is translated by T_d , is $T_d(a) = a[s]$, $(0 \leq s \leq n)$.

Theorem 1. Suppose T_d is a translation in D , then $v(a) = v(T(a))$.

Proof: $\forall s, t \in N$, there is

$$\langle a[t], a[s+t] \rangle = \langle a, a[s] \rangle \quad (14)$$

$$\exists \langle a[t], a[s+t] \rangle = \sum_i a_i[t] a_i[s+t] = \sum_i a_{i+t}[0] a_{i+t}[s] = \sum_i a_i[0] a_i[s] = \langle a[0], a[s] \rangle$$

From formula (14), we know $\forall s, t \in N$, $\exists g_s(a[t]) = g_s(a)$ which satisfies

$$v(a[t]) = v(a) \quad (15)$$

For every T_d , $\exists T_d(a) = a(s)$ which leads to $v(a) = v(a[s]) = v(T(a))$.

Theorem 2. After transformation, the relative distances and angles produced by the character point connections are in unchanged

Proof: In the plane D , $\forall a_0, b_0, c_0 \in M$ denotes the line segment vector produced by character point connections. θ_0 is the angle of a_0 and b_0 . Any angle α , makes the image M rotate around its ordinate origin by α . (α is positive while clockwise, and negative otherwise), which is a linear transform T_α . This results in image \bar{M} . For $\forall a$ and b , there is\

$$[a, b] = [a_0, b_0] T_\alpha = [a_0, b_0] \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \quad (16)$$

Here $[a, b]$ is a homogeneous orthogonal rotated matrix, and $\|T_\alpha\| = \sqrt{\rho(T_\alpha^T T_\alpha)} = 1$, $\rho(\cdot)$ is the matrix radius. Then we can write

$$\|a\| = \sqrt{a^T a} = \sqrt{a_0^T T_\alpha^T T_\alpha a_0} = \|a_0\|.$$

Accordingly,

$$\|b\| = \|b_0\| \quad (17)$$

$$\langle a, b \rangle = \langle T_\alpha a_0, T_\alpha b_0 \rangle = \|T_\alpha\|^2 \langle a_0, b_0 \rangle = \langle a_0, b_0 \rangle$$

This leads to

$$\theta = \arccos \frac{\langle a, b \rangle}{\|a\| \cdot \|b\|} = \theta_0 \quad (18)$$

Theorem 3. The relative distances and angles are invariable after the translation and rotation

Proof: Suppose the image M converts to M' after translation by T_d and rotation by T_α . $\forall a, a' \in M'$, suppose $a = H(x) = T_d(T_\alpha(a_0))$, then $a_0 = T_\alpha^{-1}(T_d^{-1}(a))$. Suppose, too, $a' = T_\alpha^{-1}(a)$, then $a_0 = T_\alpha^{-1}(a')$.

From theorem 1, we know $v(a') = v(a)$, further because of $a_0 = T_\alpha(v(a'))$ and $v(a) = T_\alpha(v(a_0))$, so $v(a_0) = T_\alpha^{-1}(v(a'))$. All this leads up to

$$v(a_0) = v(H(a_0)) \quad (19)$$

3.1.2 Method description

Extract the intersecting points from the repaired thinned finger-vein image and connect all the points with each other. Compute the relative distances and angles to get the relative distance feature M , and the angle feature θ . Fuse these two features by "Logical And", and on this basis, match any two images to get the number of relative distances and angles that correlate. Only when both features are approximately the same, is the matching successful. Otherwise, the matching has failed.

A. Finger vein topology

Using Kejun Wang's method [19] for pre-processing hand-back vein image, combined with region merging and watershed algorithm, the finger-vein skeleton is extracted, thinned and further repaired. A fully meshed topology is formed by selecting the intersecting points on the thinned finger-vein image as character points and connecting these points to each other with straight lines, partitioning the image into several regions as shown in Fig.8.

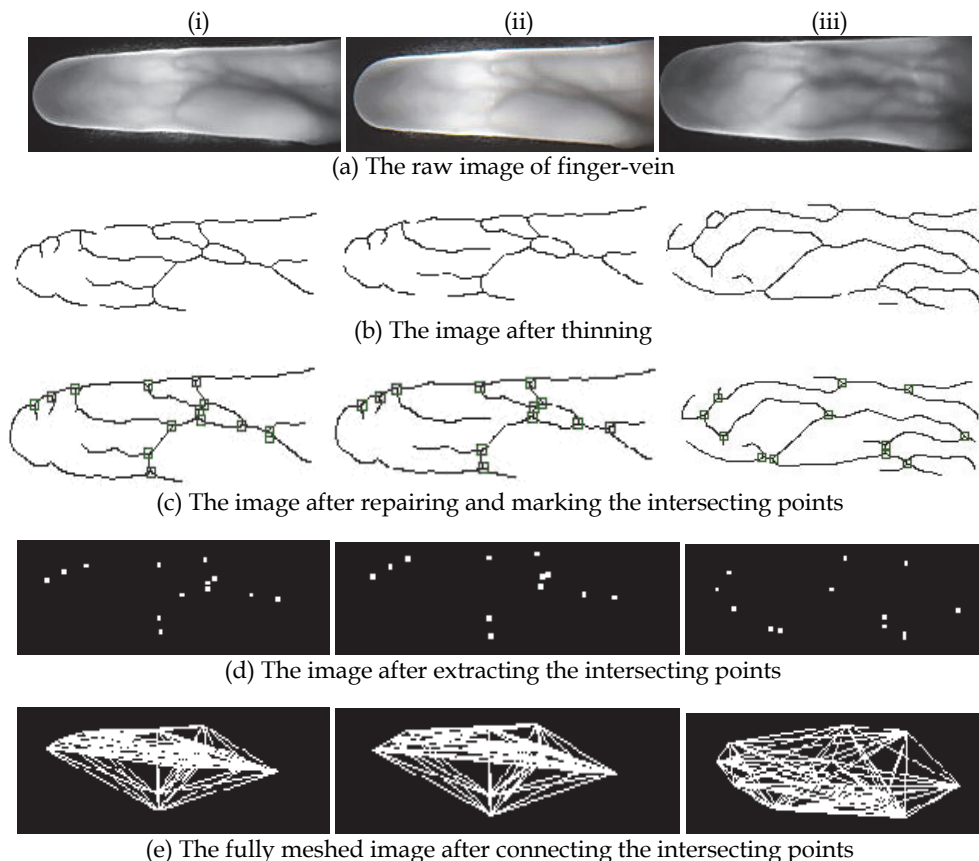


Fig. 8. The finger-vein image feature extraction process

In Fig.8, (i) and (ii) are two finger-vein images from same source, so their topology is similar. However, (iii) is of a different source and its topology is obviously different from (i) and (ii). Specifically, the topology expresses an integral property and peculiarity of finger-veins, the relationship between corresponding character points is of importance.

3.1.3 Matching finger-vein images using relative distance and angles

From the thinned finger-vein image of Fig.8 (b), we can see the random finger-vein pattern and inner structure. The inner characteristic points produced by the intersecting vein crossings reflect the unique property of the finger-vein. However, those breakpoints may be thought as finger-vein endpoints, which would influence recognition results. For this reason, the more reliable intersecting points are chosen to characterize finger-veins. Considering that different line segments are produced by intersecting points from different finger-vein images, the two features -relative distance and angle - are combined for matching.

Relative distance and angle are essential attributes of finger-veins, which ensure the feature uniqueness and reflect different characteristics of finger-vein structure. Fusion of the two

features with “Logical And” make the recognition results more reliable. Thus, matching two finger-vein images is converted into matching the similarity of topologies.

The detailed steps are as follows.

1. Calculate the relative distances and angles of finger-vein image. Suppose, there are d points of intersection in one image, then the number of relative distance is $d(d-1)/2$. The number of angles produced by the point connections is $d(d-1)(d-2)/2$. Here a set of finger-vein image features is defined as $R = (l_m, \theta_u)$, where l is the distance of any two intersecting points, θ is the angle produced by the point connections, m and u are the number index respectively. Suppose, $R_1 = (l_m, \theta_u)$ and $R_2 = (l_n, \theta_v)$ are two sets of finger-vein image features.
2. Compare m relative distances from R_1 with n relative distances from R_2 , by calculating the number of approximately similar relative distances. If the number is greater than the pre-defined threshold, go to next step; else, the matching is assumed to have failed. To take care of position error of those points, we define $\|l_m - l_n\| < e$ to show the extent of similarity between any two Eigen values (e is the allowable error range). From experimental analysis, $e = 0.0005$ is very appropriate.
3. Suppose there are q eigenvalues of approximately equivalent relative distances, connect the q character points in the two sets respectively, with each other. Thus, z angles are produced, which are denoted as θ_{z1} and θ_{z2} in R_1 and R_2 respectively. On this basis, calculate the number of approximately equivalent angles. If the number is greater than the pre-defined threshold, the matching is successful; else, the matching is thought to have failed. Similarly, $\|\theta_m - \theta_n\| < e'$ is used to show the relationship of two approximately equivalent. From experimental analysis, $e' = 0.006^\circ$ is very appropriate.

3.2 Finger vein recognition based on wavelet moment fused with PCA transform

3.2.1 Finger vein feature extraction

Different people have different finger lengths. Also, there can be variation in the image captured for the same person due to positioning during the image capture process. Thus if image sizes are not standardized, there is bound to be representation error which leads to a decrease in the recognition rate. In this part, we resize the vein image into a specific image block size to facilitate further processing. The original image is standardized to a height of 80 pixels and split along the width into 80×80 sub-image block size. If the image is split evenly (given that the image width is generally about 200 pixels) there will be loss of information that will affect recognition. Therefore, the sub-blocks are created with an overlap of 60 pixels for every 80×80 image sub-block. The original image can thus be split into 6-7 sub-images, with sufficient characteristic quantities for identification.

Set a matrix $A_{m \times n}$ to represent the standardized images $f(x, y)$.

$$A_{mn} = [A_0, A_1, A_2, \dots, A_{n-1}] \quad (20)$$

Which: A_i is a column vector, $i \in [0, n-1]$

Here we define the sub-block of the image width w , standardized image height h (in experiment $w=80, h=80$). Sub-images are extracted at interval r when (in this experiment $r=20$).

$$\begin{aligned}
 B_1 &= [A_0, \dots, A_{w-1}] \\
 B_2 &= [A_r, \dots, A_{w-1+r}] \\
 &\dots \\
 B_k &= [A_{kr}, \dots, A_{w-1+kr}]
 \end{aligned}$$

$[x]$ takes the maximum integer less than x .

$$k = \left\lceil \frac{n-w}{r} + 1 \right\rceil \tag{21}$$

Thus we get a total of B_1, B_2, \dots, B_k k sub-image, and the size of each sub-image is $w \times h$. Then we extract features for each sub-image B_i . The finger block, feature extraction and recognition process shown in Fig.12.

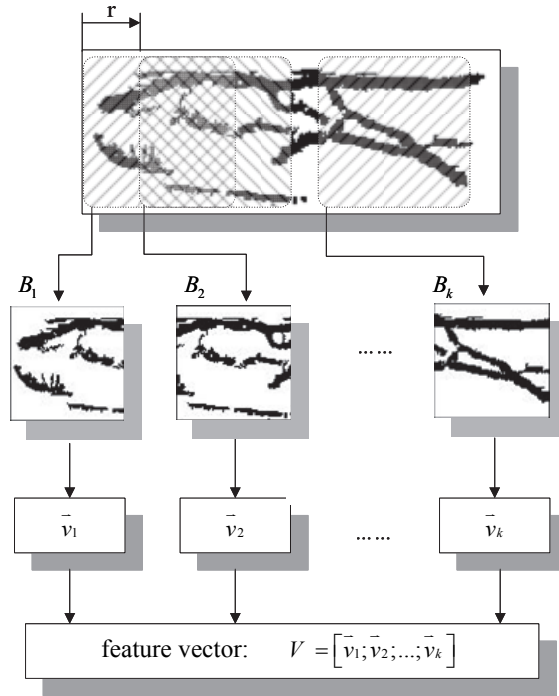


Fig. 12. The sketch map of sub-image extraction

3.2.2 Wavelet transform and wavelet moments extraction

Wavelet moment is an invariant descriptor for image features. A wavelet moment feature is invariant to image rotation, translation and scaling so it is successfully applied in the pattern recognition.

For each sub-image $B_i(x,y)$, its size is $w \times h$. Applying two dimensional Mallat decomposition algorithm, we can make wavelet decomposed image $B_i(x,y)$.

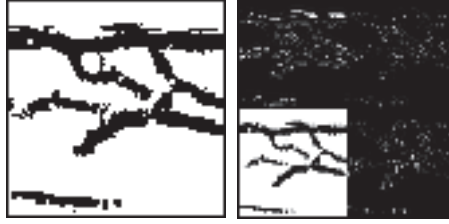


Fig. 13. The sub-image and result of wavelet decomposition

Setting $f(x,y) = B_i(x,y) \in L^2(R^2)$ to be the analyzed sub-image vein blocks, the wavelet decomposed layer is

$$f(x,y) = A_1 + D_1^1 + D_1^2 + D_1^3 \quad (22)$$

where A_1 is the scale for the low frequency component (i.e. approaching component), and D_1^1, D_1^2, D_1^3 are the scales for the horizontal, vertical and diagonal components respectively.

$$A_1 = \sum_{(m,n)} c_1(m,n) \phi_1(m,n), \quad (23)$$

$$c_1(m,n) = \langle f(x,y), \phi_1(m,n) \rangle$$

$$D_1^k = \sum_{(m,n)} d_1^k(m,n) \psi_1^k(m,n), \quad (24)$$

$$d_1^k(m,n) = \langle f(x,y), \psi_1^k(m,n) \rangle$$

Where $k=1,2,3$ $c_1(m,n)$ is the coefficient of A_1

d_1^k is the coefficient of the three high frequency components.

$\phi_1(m,n)$ is the scale function

$\psi_1^k(m,n)$ is the wavelet function

Daub4 was chosen for wavelet decomposition, as it produced better identification results from several experimental compared with other wavelets. We use the approximation wavelet coefficients c_j to compute the wavelet moment [4]. Set $w_{p,q}$ expressed as $(p+q)$ order central moments of $f(x,y)$. The wavelet moment approximation is:

$$\begin{cases} w_{p,q}^0 = 2^{-(p+q+1)j} \sum_{m,n \in Z} m^p n^q c_1(m,n) \\ w_{p,q}^k = 2^{-(p+q+1)j} \sum_{m,n \in Z} m^p n^q d_1^k(m,n) \end{cases} \quad (25)$$

Here we access the wavelet moment w_{22} .

3.2.3 PCA transformation

The advantage of using wavelet transform to reduce computation is explored here. Using each sub-image B_i directly without the PCA transform not only leads to poor classification of extracted features but also huge computational cost. After the low-frequency wavelet sub-

images compression of the original image to about one-fourth of the original size, PCA decomposition is applied on the sub-image which greatly reduces computation.

3.2.4 The transformation matrix

Here we analyze a layer of B_i wavelet decomposition of the low-frequency sub-image. For PCA, A_1 is transformed to a separate $wh / 4$ dimension of image vector $\xi = Vec(A_1)$.

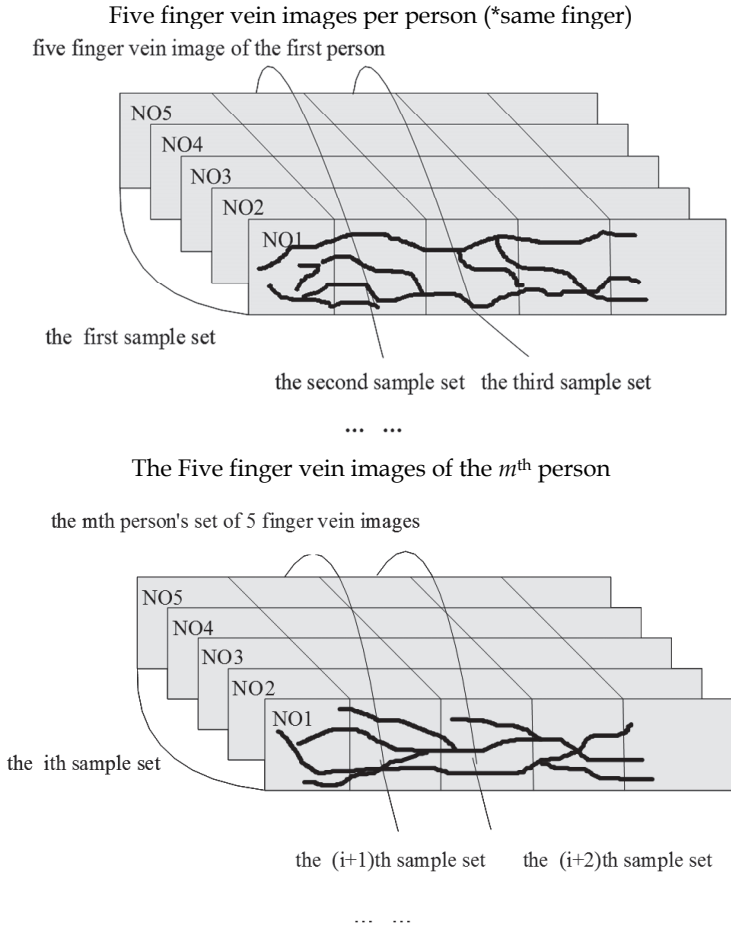


Fig. 14. The sketch map of sample classes after PCA transform

To illustrate the problem, we take finger vein samples from a total set of c people. Each sample of the same finger has five images as shown in Fig.14. (Note that Fig.14 is only for illustration purpose. In practice, there is an interval of 20pixels between two adjacent sub-blocks, and an overlap of 60pixels as earlier described).

The n -th sub-block set of the m -th person is indexed as $k_{m,n}$; where $n = (1,2,3,\dots, L)$; L_m is the total number of sub-blocks for person m .

To compare images we only use the k_{\min} sub-image set where

$k_{\min} = L = \min(L_1, L_2, L_3, \dots, L_m)$. when $k_{\min} = 5$, then

$$k_{\min} = \min(k_{1,1}, k_{1,2}, \dots, k_{1,5}, \dots, k_{c,5}) \quad (26)$$

Thus, a total of $C = c \times k_{\min}$ of the available pattern classes, i.e. $\omega_1, \omega_2, \dots, \omega_C$.

Four of the corresponding samples in the i -th class ($i = 5m + k$), $k = 1, 2, \dots, k_{\min}$ is the number of sub-image of each finger image, for simplicity, we write: $\xi_{i,1}, \xi_{i,2}, \xi_{i,3}, \xi_{i,4}, \xi_{i,5}$, and all are $wh/4$ dimension column vectors. The total number of training samples is $N = 5C$.

Mean of the i -th class training sample:

$$\bar{\xi}_i = \frac{1}{5} \sum_{j=1}^5 \xi_{i,j} \quad (27)$$

Mean of all training samples:

$$\bar{\xi} = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^5 \xi_{ij} \quad (28)$$

The scatter matrix is:

$$S_i = \sum_{i=1}^C P(\omega_i) (\bar{\xi}_i - \bar{\xi})(\bar{\xi}_i - \bar{\xi})^T \quad (29)$$

where $P(\omega_i)$ is prior probability of the i -th class of training samples. Then we can obtain the characteristic value $\lambda_1, \lambda_2, \dots$, of S_i (the value of these features have been lined up in sequence by order of $\lambda_1 \geq \lambda_2 \geq \dots$) and its corresponding eigenvector $\varphi_1, \varphi_2, \dots$. Take d before the largest eigenvalue corresponding to the standard eigenvectors orthogonal transformation matrix $P = [\varphi_1, \varphi_2, \dots, \varphi_d]$. For each sub-image blocks B_i , through the wavelet decomposition of the low-frequency sub-image A_1 , A_1 in accordance with the preceding method into a column vector ξ , to extract the features use the transformation matrix P obtained in the previous section, the following formula:

$$e = P^T \xi \quad (30)$$

This $e = [e_1, e_2, \dots, e_d]$ is the PCA extraction of feature vectors from sub-image blocks. After several experiments, we found that when $d = 200$ we can get a good result, and when $d = 300$ i.e a 300-dimensional compression, we get the best recognition results.

3.2.5 LDA map

In general, PCA method is the best for describing feature characteristics, but not the best for feature classification. In order to get better classification results, we use the LDA method for further classification of PCA features.

Each sample is transformed into a lower d -dimensional space in the post-dimensional feature vector $e_i = [e_1^i, e_2^i, \dots, e_d^i]$. Using PCA projection matrix P , $i = 1, 2, \dots, N$ is the sample number. Our classifier design follows dimension reduction to get PCA feature vectors

e_1, e_2, \dots, e_N and form the class scatter matrix S_w and the within class scatter matrix S_b . Calculate the corresponding matrix $S_w^{-1}S_b$ of the l largest eigenvalue eigenvectors $\alpha_1, \alpha_2, \dots, \alpha_l$. The l largest eigenvectors corresponding to the LDA transformation matrix $W_{LDA} = [\alpha_1, \alpha_2, \dots, \alpha_l]$. Then we use the LDA transformation matrix W_{LDA} as

$$z_i = [z_1^i, z_2^i, \dots, z_l^i] = W_{LDA}^T e_i, \quad i = 1, 2, \dots, N \text{ is the sample number.} \quad (31)$$

Thus, we can use the best classification feature z vector to replace the feature vectors e for identification and classification.

3.2.6 Matching and recognition

Through the above wavelet decomposition and PCA transform for each sub-image B_i , we obtain wavelet moments w_{22} and extract feature vector z of PCA and LDA. B_i is characterized by $v_i = [w_{22}; z]$. Matching feature vectors of finger1 $v_i = [w_{22}; z]$ and of finger 2 can be done as follows.

The first step is the length of V and V' and may not be the same, that is, k and k' is not necessarily the same. Here we define:

$$K = \min(k, k') \quad (32)$$

Taking the K vectors of V and V' for comparison, first analyze the corresponding sub-image blocks v_i and v'_i .

$$v_i = [w_{22}; z], \quad v'_i = [w'_{22}; z'] \quad (33)$$

From several experiments, we set two threshold vectors w_t, z_t . Euclidean distance between B_i sub-images, δ_i is defined for two feature vectors w and w' from V and V' . A matching score defined for V and V' feature w_{22} of wavelet moment of corresponding sub-image B_i matching score:

$$w_mark_i = \begin{cases} \frac{w_t - \delta_i}{w_t} & \text{if } \delta_i < w_t \\ 0 & \text{else} \end{cases} \quad (34)$$

Finally, we obtain wavelet moment feature of the finger matching score:

$$w_mark = \sum_{i=0}^K w_mark_i \quad (35)$$

Similarly, we create a match V and V' , of the feature vector z scores z_mark .

Finally, combining the scores:

$$total_mark = s_1 \times w_mark + s_2 \times z_mark \quad (36)$$

s_1, s_2 are the share of feature matching scores, and $s_1 \geq 0, s_2 \geq 0, s_1 + s_2 = 1$.

Thus, if the finger vein 1 and 2 match *total_mark* value score is greater than a given threshold, the two fingers match, otherwise they do not match. A minimum distance classifier can also be used for the recognition task.

4. Experimental results

4.1 Processing experimental results

To verify the effectiveness of the proposed method, we test the algorithm using images from a custom finger vein image database. The database includes five images each of 300 individuals' finger veins. Each image size is 320*240.

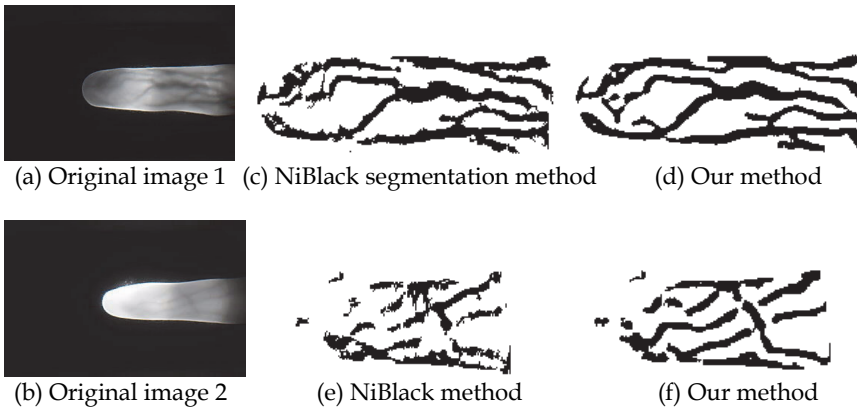


Fig. 15. Experimental results

We have used a variety of traditional segmentation algorithms and their improved algorithms to segment vein image. But segmentation results of vein image by these algorithms aren't ideal. Because the result of NiBlack segmentation method is better than other methods [13], we use NiBlack segmentation method as the benchmark for comparison. Segmentation was done for all the images in our database using NiBlack segmentation method and using our method. Experimental results show that our method has better performance. To take full account of the original image quality factor, we select two typical images from our database with one from high quality images and the other from poor quality images to show the results of comparative test. Where Fig. 15(a) is the high-quality vein image in which veins are clear and the background noise is small. Fig. 15(b) is the low-quality vein image. The uneven illumination caused the finger vein image to be fuzzy, which seriously affects image quality. We extract veins feature by using our method and compare with results of the NiBlack segmentation method. Experimental results shown in Fig. 15(c) and Fig. 15(e) are obtained from the NiBlack method applied in [9]. This algorithm extracts smooth and continuous vein features of high-quality image. There are a few pseudo-vein characteristics in Fig. 15(c). But in Fig. 15(e), there is much noise in the segmentation results. Segmented image features have poor continuity and smoothness, and there is the effect of the over-segmentation. Experimental results show that apart from smoothness and continuity or removal of noise and pseudo-vein characteristics, the method proposed in this paper extracts vein features effectively not only from the high-quality

images but also from the low-quality vein images as shown in Fig.15(d) and Fig.15(f). We show that the algorithm proposed in this paper performs better than the traditional NiBlack method.

4.2 Relative distance and angle experimental results

Finger-vein images (size 320×240) of 300 people were selected randomly from Harbin Engineering University finger-vein database. One forefinger vein image of each person was acquired, so there are 300 training images.

Generally, a good recognition algorithm can be successfully trained on a small dataset to get the required parameters and achieve good performance on a large test dataset. Therefore four more images from the forefinger of those 300 people were acquired giving a total sum of 1200 images to be used as verification dataset.

When matching, every sample is matched with others, so there are $(300 \times 299) / 2 = 44850$ matching times; 300 of which are legal, while the others are illegal matches. Two different verifying curves are shown in Fig.16. The horizontal axis stands for the matching threshold, and vertical axis stands for the corresponding probability density. The solid curve is legal matching curve, while the dashed is illegal. Both curves are similar to the Gaussian distribution. The two curves intersect, at a threshold of 0.41. The mean legal matching distance corresponds to the wave crest near to 0.21 on the horizontal axis, and the mean of illegal matching distance corresponds to the wave crest near to 0.62 on the horizontal axis. The two wave crests are far from each other with very small intersection. So this method can recognize different finger-veins, especially when the threshold is in the range $[0.09-0.38]$, where the GAR is highest.

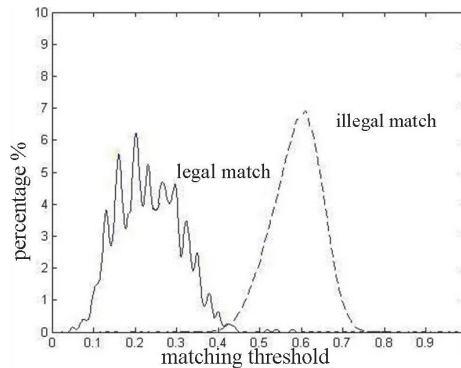


Fig. 16. Legal matching curve and illegal matching curve

The relationship between FRR and FAR is shown in Fig.17. For this method, the closer the ROC curve is to the horizontal axis, the higher the Genuine Acceptance Rate (GAR). Besides, the threshold should be set suitably according to the fact, when FRR and FAR are equivalent, the threshold is 0.47, that is to say, EER is 13.5%. In this case, GAR of the system is 86.5%. The result indicates that this method is reasonable, giving accurate finger-vein recognition.

The method above compares the numbers of relative distances and angles which are approximately equivalent from two finger-vein images. In the second step, only the

intersecting points which are matched successfully on the first step are used and thus, computation of superfluous information is avoided and only information vital to decision making is used. Only when the two matching steps are successful is the recognition successful. According to Theorem 3, the relative distance and angle would not change when even after image translation and rotation. So the proposed algorithm is an effective method for finger-vein recognition.

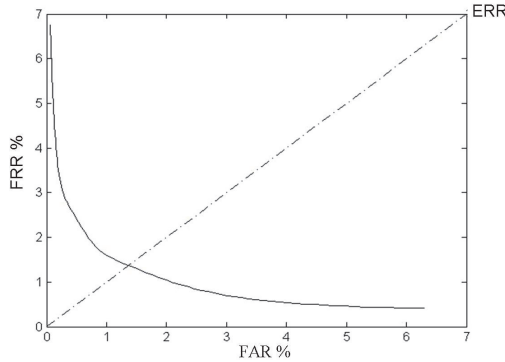


Fig. 17. ROC curve of the method

In 1:1 verifying mode, compare the one image out of 1200 samples in verifying set with the image, which has the same source with the former one, in training set to verify. The experiment result is shown in Tab.1, the times of success are 1120, and the rate of success is 93.33%. In 1:n recognition mode, compare the 1200 images with all images in training set, 360000 times in sum. The result is as Tab.2, the times of FAR is 25488, and GAR is 92.92%.

Total matching times	Number of successes	Number of failures	Success rate (%)	FRR(%)
1200	1120	80	93.33	6.67

Table 1. Test result of FRR in 1:1 mode

Total matching times	Total false acceptance	GAR (%)	FAR(%)
360000	25488	92.92	7.08

Table 2. Test result of FAR in 1:n mode

To test the ability to overcome image translation and rotation, translate randomly in the range $[-10,+10]$ and rotate the image randomly in the range $[-10^{\circ},+10^{\circ}]$, in order to establish the translation and rotation test sets. Then verify and recognize the two sets respectively. The samples which have the same source are compared in a 1:1 experiment; matching each sample from the two sets with the samples from the training set to accomplish 1:n experiment. The result is shown in Tab.3, Tab.4, Tab.5 and Tab.6.

	Total matching times	Number of successes	Number of failures	Success rate (%)	FRR(%)
1:1	1200	1111	89	92.58	7.42

Table 3. Translation test set verification result (1:1)

	Total matching times	No. of genuine acceptance	GAR(%)	FAR(%)
1:n	360000	27900	92.25	7.75

Table 4. Translation test set recognition result (1:n)

As the experiment shows, in 1:1 mode the rate of success can reach 92.58% even though the finger-vein image is translated, and can reach 91.75% when rotated. In 1: n mode, GAR can reach 92.25% and 91.17% respectively, which implies a robust recognition system. Further, the method can overcome the influence caused by image translation and rotation, thus it can meet practical requirements.

	Total matching times	Number of successes	Number of failures	Success rate (%)	FRR(%)
1:1	1200	1101	99	91.75	8.25

Table 5. Rotation test set verification result (1:1)

	Total matching times	No. of genuine acceptance	GAR(%)	FAR(%)
1:n	360000	31788	91.17	8.83

Table 6. Rotation test set recognition result (1:n)

4.5.5 Experimental results analysis

The algorithm was implemented on a Windows XP platform using Visual C + +6.0. Finger vein image capture was performed taking into account the convenience of users, while collecting index finger and middle finger vein images. A total of 287 finger vein images collected for each finger 5 times, and in all, a total of $287 \times 5 = 1435$ were collected to form finger vein library. Two sets $287 \times 2 = 574$ were taken for verification.

The identification results based on template matching:

We first according to the method proposed in reference [1], recognition for the veins of our library of images. In 1:1 verification mode, we use the validation library of 574 samples to verify the experimental results shown in Tab.7.

Matching times	Pass times	Reject recognition times	Correct recognition rate (%)	Reject recognition rate (%)
574	559	15	97.4	2.6

Table 7. The results of refusing ratio of 1:1

For 1: n identification model, we use the validation library to identify 574 samples of the experimental results shown in Tab.8.

Matching times	False recognition times	False recognition rate (%)
574	7	1.2

Table 8. The result of mistaken identifying ratio of 1: n

For a number of reasons, we realize that the algorithm recognition rate is not as high as the reference [1], perhaps due to the acquisition of image and acquisition machine quality problems.

The identification results based on wavelet moment:

We first decompose the predetermined wavelet sample in the vein sample database, and then construct the wavelet moment features using identified wavelet coefficients.

A few typical experimental results of 1:1 verification mode shown in Tab.9.

	Matching times	Pass times	Reject recognition times	Correct recognition rate (%)	Reject recognition rate (%)
Hear	574	536	38	93.4	6.6
Daub4	574	547	27	95.3	4.7
Daub8	574	543	31	94.6	5.4
coif2	574	534	40	93.04	6.96
sym	574	529	45	92.2	7.8

Table 9. The results of rejection ratio with different wavelet base in the 1:1 case

For 1: n identification pattern, the experimental results shown in Tab.10.

	Matching times	False recognition times	false recognition rate
Haar	574	27	4.7
Daub4	574	11	1.9
Daub8	574	19	3.3
coif2	574	30	5.2
sym	574	33	5.7

Table 10. Results of mistaken identifying with different wavelet base in the case of 1: n

We chose Daub4 to carry out wavelet decomposition, identification was better than other wavelets.

Identification results of wavelet moment integration of PCA.

When PCA is used for dimension reduction, the relationship of selection of the compressed dimension k and the proportion of it represent components shown in Tab.11:

$$w_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^N \lambda_i} \tag{37}$$

k	215	240	276	299	338	368	380	392
w_k	0.75	0.80	0.86	0.90	0.95	0.98	0.99	1.00

Table 11. The compressed dimension and its proportion

To balance computation and weighting, we use 300 as the dimension for decomposition. Authentication in 1:1 mode and 1: n identification pattern, we use 574 samples in the validation library for the experimental, results shown in Tab.12, 13.

Matching times	Pass times	reject recognition times	recognition rate (%)	reject recognition rate (%)
574	568	6	98.95	1.05

Table 12. Results of rejection ratio of 1:1

Matching times	False recognition times	False recognition rate (%)
574	4	0.7

Table 13. The results of mistaken identifying ratio of 1: n

As seen from the recognition results, recognition rate and rejection rate in the method based on wavelet PCA can meet the requirements of practical applications. In recognition speed, that is, 1: n of the mode can meet the requirements.

5. Conclusion

Accurate extraction of finger vein pattern is a fundamental step in developing finger vein based biometric authentication systems. Finger veins have textured patterns, and the directional map of a finger vein image represents an intrinsic nature of the image. The finger vein pattern extraction method using oriented filtering technology. Our method extends traditional image segmentation methods, by extracting vein object from the oriented filter enhanced image. Experimental results indicate that our method is a better enhancement over the traditional NiBlack method [11], [12], [13], and has good segmentation results even with low-quality images. The addition of oriented filter operation, extracts smooth and continuous vein features not only from high quality vein images but also handles noisy low quality images and does not suffer from the over-segmentation problem. However, it requires a little more processing time because of the added oriented filter operation.

Topology is an essential image property and usually, even an inflection point may contain plenty of accurate information. Finger-vein recognition is faced with some basic challenges, like positioning, the influence of image translation and rotation etc. To address these problems, essential topology attributes of individual finger veins are utilized in a novel method. Particularly, the relative distance and angles of vein intersection points are used to characterize a finger-vein for recognition, since the topology of finger-vein is invariant to image translation and rotation. The first step is to extract those intersecting points of the thinned finger-vein image, and connect them with line segments. Then relative distances and angles are calculated. Finally combine the two features for matching and recognition. Experimental results indicate that the method can accurately recognize finger-vein, and to a certain degree, overcome the influence image translation and rotation. Furthermore, the method resolves the difficult problem of finger-vein positioning. It is also computationally efficient with minimal storage requirement, which makes the method of practical significance. However there are still problems of non- recognition and false recognition. Besides, pre-processing is an import requirement for this method and the accuracy of pre-processing influences recognition result significantly. In view of this, further research will be done on the pre-processing method, to improve the image quality and the accuracy of feature extraction, and subsequently improve system reliability.

This chapter discussed recent approaches to solving the problem of varying finger lengths and proposed using a set of images of same size interval in a selected sub-block approach. For each image sub-block, wavelet moment was performed and PCA features extracted. LDA transform is performed, and the two features were combined for recognition. For

matching and identification, we proposed a method of fuzzy matching scores. Experimental results show that wavelet moment PCA fusion method achieved good recognition performance; error rate FAR was 0.7%, rejection rate FRR of 1.05%. In future research, we are committed to further study finger vein feature fusion with fingerprint and other features to improve system reliability.

6. References

- [1] N. Miura, A. Nagasaka and T. Miyatake, Extraction of finger-vein patterns using maximum curvature points in image profiles, *IEICE Trans. Inf. & Syst.* Vol.90, no.D(8),pp. 1185-1194, 2007.
- [2] Shahin M, Badawi A, Kamel M. Biometric authentication using fast correlation of near infrared hand vein patterns, *J. International Journal of Biometrical Sciences*, vol.2,no.3,pp.141-148,2007.
- [3] Lin Xirong,Zhuang Bo,Su Xiaosheng, ZhouYunlong, Bao Guiqiu. Measurement and matching of human vein pattern characteristics, *J.Journal of Tsinghua University (Science and Technology)*.vol. 43no. 2 pp.164-167, 2003. (In Chinese.)
- [4] H.Tian, S.K.Lam, T. Srikanthan. Implementing OTSU's Thresholding Process Using Area-time Efficient Logarithmic Approximation Unit, *J. Circuits And Systems*, vol.5, pp. 21-24, 2003.
- [5] Zhongbo Zhang, Siliang Ma, Xiao Han. Multiscale Feature Extraction of Finger-Vein Patterns Based on Curvelets and Local Interconnection Structure Neural Network, *IEEE Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China,vol 4, pp.145-148,2006.
- [6] M.Naoto, A.Nagasaka, iM.Takafm Feature Extraction of Finger-vein Patterns Based on Repeated Line Tracking and Its Application to Personal Identification, *J. Machine Vision and Application*, vol. 15, no.4, pp. 194-203, 2004.
- [7] Rigau J. Feixas, M.Sbert. Metal Medical image segmentation based on mutual information maximization, *In Proceedings of MICCAI 2004, Saint-Malo, France pp.135-142, 2004.*
- [8] Yuhang Ding, Dayan Zhuang and Kejun Wang, A Study of Hand Vein Recognition Method, *Mechatronics and Automation, 2005 IEEE International Conference*, vol. 4 no.29pp.2106-2110, 2005.
- [9] O'Gorman, L. Lindeberg, J.V. Nickerson. An approach to fingerprint filter design, *Pattern Recognition*, vol. 22 no.1pp. 29-38, 1989.
- [10] Xiping Luo; Jie Tian. Image Enhancement and Minutia Matching Algorithms in Automated Fingerprint Identification System, *J Journal of Software*, vol. 13 no.5pp. 946-956. 2002. (In Chinese.)
- [11] W. Niblack. *An Introduction to Digital Image Processing*, Prentice Hall, ISBN 978-0134806747, Englewood Cliffs, NJ, pp.115-116, 1986
- [12] Kejun Wang Zhi Yuan. Finger vein recognition based on wavelet moment fused with PCA transform, *J Pattern Recognition and Artificial Intelligence*, vol. 20 no.5 pp. 692-697, 2007. (In Chinese.)
- [13] Chengbo Yu, Huaifeng Qing, *Biometric Identification Technology Finger Vein Identification Technology*: Tsinghua University Press, 2009, pp: 81-87. (In Chinese.)

-
- [14] Naoto Miura, Akio Nagasaka, Takafumi Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification [J]. *Machine Vision and Applications*, 2004,15(4):194 -203
- [15] Kejun Wang, Yuan Zhi. Finger Vein Recognition Based on Wavelet Moment Fused with PCA Transform. [J] *Pattern Recognition and Artificial Intelligence*. 2007
- [16] Xueyan Li. Study of Multibiometrics System Based on Fingerprint and Finger Vein[D]. the doctorate dissertations of Jilin University. 2008.
- [17] Xiaohua Qian. Research of Finger-vein Recognition Algorithm[D]. MA Dissertation of Jilin University. 2009.
- [18] Zhong Bo Zhang, Dan Yang Wu, Si Liang Ma. Pattern Recognition, 2006. ICPR 2006. 18th International Conference on Volume: 4 Digital Object Identifier: 10.1109/ICPR.2006.848. Publication Year: 2006, Page(s): 145 - 148
- [19] Kejun Wang, Yuhang Ding, Dazhen Wang. A Study of Hand Vein-based Identity Authentication Method [J]. *Science & Technology Review*. 2005, 23(1) :35-37.
- [20] Ji Hu, SunJixiang, YaoWei. Wavelet Moment for Images. *Journal of Circuits and Systems*, 2005, 10(6):132-136 (inChinese)

Minutiae-based Fingerprint Extraction and Recognition

Naser Zaeri
Arab Open University
Kuwait

1. Introduction

In our electronically inter-connected society, reliable and user-friendly recognition and verification system is essential in many sectors of our life. The person's physiological or behavioral characteristics, known as biometrics, are important and vital methods that can be used for identification and verification. Fingerprint recognition is one of the most popular biometric techniques used in automatic personal identification and verification.

Many researchers have addressed the fingerprint classification problem and many approaches to automatic fingerprint classification have been presented in the literature; nevertheless, the research on this topic is still very active. Although significant progress has been made in designing automatic fingerprint identification systems over the past two decades, a number of design factors (lack of reliable minutia extraction algorithms, difficulty in quantitatively defining a reliable match between fingerprint images, poor image acquisition, low contrast images, the difficulty of reading the fingerprint for manual workers, etc.) create bottlenecks in achieving the desired performance. Nowadays, investigating the influence of the fingerprint quality on recognition performances also gains more and more attention.

A fingerprint is the pattern of ridges and valleys on the surface of a fingertip. Each individual has unique fingerprints. Most fingerprint matching systems are based on four types of fingerprint representation schemes (Fig. 1): grayscale image (Bazen et al., 2000), phase image (Thebaud, 1999), skeleton image (Feng, 2006; Hara & Toyama, 2007), and minutiae (Ratha et al., 2000; Bazen & Gerez, 2003). Due to its distinctiveness, compactness, and compatibility with features used by human fingerprint experts, minutiae-based representation has become the most widely adopted fingerprint representation scheme.

The uniqueness of a fingerprint is exclusively determined by the local ridge characteristics and their relationships. The ridges and valleys in a fingerprint alternate, flowing in a local constant direction. The two most prominent local ridge characteristics are: 1) ridge ending and, 2) ridge bifurcation. A ridge ending is defined as the point where a ridge ends abruptly. A ridge bifurcation is defined as the point where a ridge forks or diverges into branch ridges. Collectively, these features are called *minutiae*. Detailed description of fingerprint minutiae will be given in the next section.

The widespread deployment of fingerprint recognition systems in various applications has caused concerns that compromised fingerprint templates may be used to make fake fingers, which could then be used to deceive all fingerprint systems the same person is enrolled in.

Once compromised, the grayscale image is the most at risk. Leakage of a phase image or skeleton image is also dangerous since it is a trivial problem to reconstruct a grayscale fingerprint image from the phase image or the skeleton image. In contrast to the above three representations, leakage of minutiae templates has been considered to be less serious as it is not trivial to reconstruct a grayscale image from the minutiae (Feng & Jain, 2011).

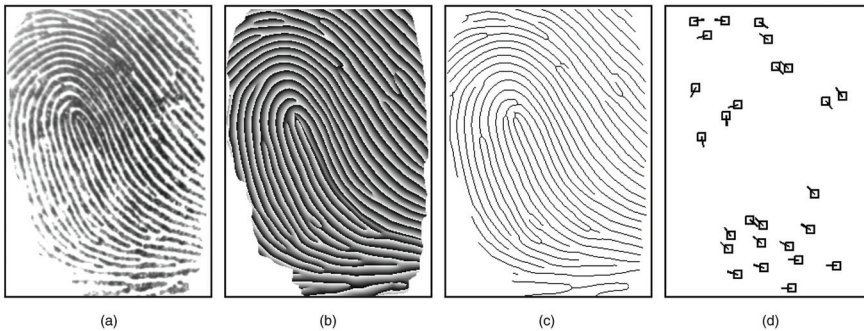


Fig. 1. Fingerprint representation schemes. (a) Grayscale image (FVC2002 DB1, 19_1), (b) phase image, (c) skeleton image, and (d) minutiae (Feng & Jain, 2011)

In this chapter, we study the recent advancements in the field of minutia-based fingerprint extraction and recognition, where we give a comprehensive idea about some of the well-known methods that were presented by researchers during the last two decades. Further, we provide a special focus on the recent techniques presented in the last few years. A close analysis of the fingerprint image will be discussed and the various minutiae features shall be described, as well.

2. Fingerprint minutiae description

The first scientific studies on fingerprint classification were made by (Galton, 1892), who divided the fingerprints into three major classes. Later, (Henry, 1900) refined Galton's classification by increasing the number of the classes. All the classification schemes currently used by police agencies are variants of the so-called *Henry's classification scheme*.

As mentioned in the previous section, the uniqueness of a fingerprint is exclusively determined by the local ridge characteristics and their relationships (Kamijo, 1993; Kawagoe & Tojo, 1984). The ridges and valleys in a fingerprint alternate, flowing in a local constant direction (Fig. 2). Eighteen different types of fingerprint features have been enumerated by (Federal Bureau of Investigation, 1984). Further, a total of 150 different local ridge characteristics (islands, short ridges, enclosure, etc.) have been identified by (Kawagoe & Tojo, 1984). These local ridge characteristics are not evenly distributed. Most of them depend heavily on the impression conditions and quality of fingerprints and are rarely observed in fingerprints. The two most prominent local ridge characteristics are: 1) ridge ending and, 2) ridge bifurcation.

A ridge ending is defined as the point where a ridge ends abruptly. A ridge bifurcation is defined as the point where a ridge forks or diverges into branch ridges. Collectively, these features are called *minutiae*. Most of the fingerprint extraction and matching techniques restrict the set of features to two types of minutiae: *ridge endings* and *ridge bifurcations*, as shown in Fig. 3. A good quality fingerprint typically contains about 40–100 minutiae.

In a latent or partial fingerprint, the number of minutiae is much less (approximately 20 to 30). More complex fingerprint features can be expressed as a combination of these two basic features. For example, an *enclosure* can be considered a collection of two bifurcations and a *short ridge* can be considered a pair of ridge endings as shown in Fig. 4.

Each of the *ridge endings* and *ridge bifurcations* types of minutiae has three attributes, namely, the x-coordinate, the y-coordinate, and the local ridge direction (θ) as shown in Fig. 5. Many other features have been derived from this basic three-dimensional feature vector. Given the minutiae representation of fingerprints, matching a fingerprint against a database reduces to the problem of point matching.

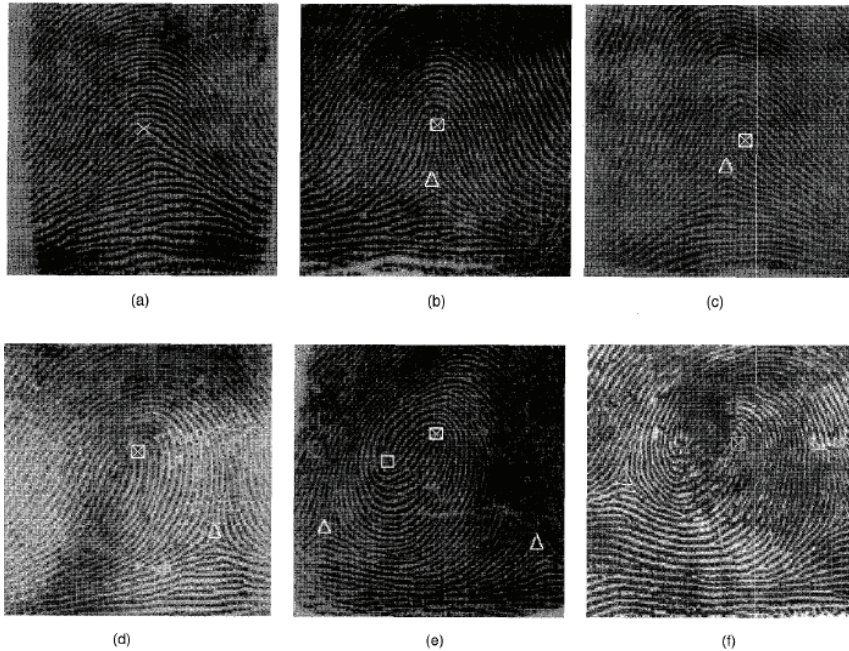


Fig. 2. Gray level fingerprint images of different types of patterns with core (□) and delta (Δ) points: (a) arch; (b) tented arch; (c) right loop; (d) left loop; (e) whorl; (f) twin loop (Ratha et al., 1996)

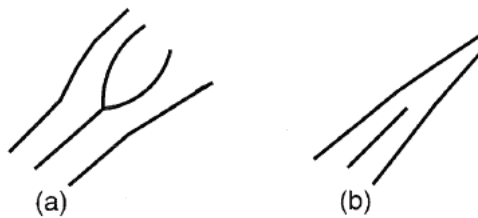


Fig. 3. Two commonly used fingerprint features: (a) ridge bifurcation; (b) ridge ending (Ratha et al., 1996)

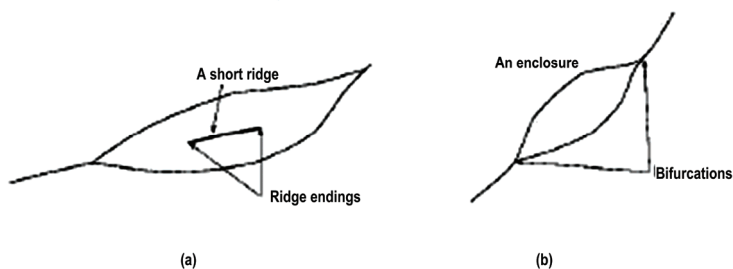


Fig. 4. Complex features as a combination of simple features: (a) short ridge; (b) enclosure (Ratha et al., 1996)

The matching problem can be defined as finding a degree of match between a query and reference fingerprint feature set. The minutiae sets can be matched using many techniques, where some of them will be addressed in following sections. The large computational requirement of matching is primarily due to the following three factors: 1) a query fingerprint is usually of poor quality, 2) the fingerprint database is very large, and 3) structural distortion of the fingerprint images requires powerful matching algorithms.

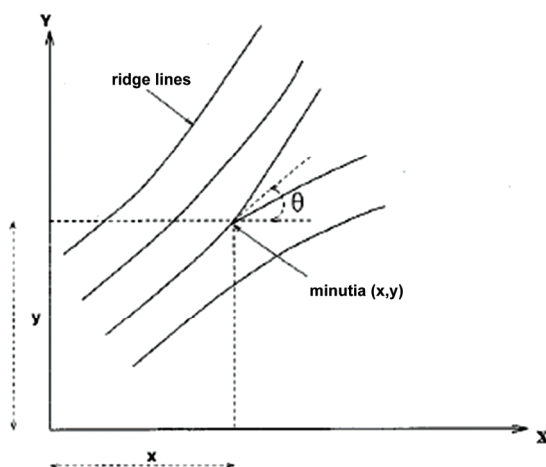


Fig. 5. Components of a minutiae feature (Hong et al., 1998)

In addition to minutiae features described above, there are other high-level features that can be used in reducing the search space during a match. A very important feature for this purpose is the pattern class of a fingerprint. Fingerprints are classified into five main categories:

- arch,
- tented arch,
- left loop,
- right loop, and
- whorl.

The pattern class may be ambiguous in partial fingerprints and indeterminate for noisy fingerprints. Yet another high-level feature is the ridge density in a fingerprint. Ridge density can be defined as the number of ridges per unit distance. In order to make it invariant to position, the ridge density between two singular points in a fingerprint is computed. Some singular points of interest are defined as the *core* and *delta* points (Ratha et al., 1996). The core point is the top most point on the inner most ridge and a delta point is the tri-radial point with three ridges radiating from it (Fig. 2 and Fig. 6).

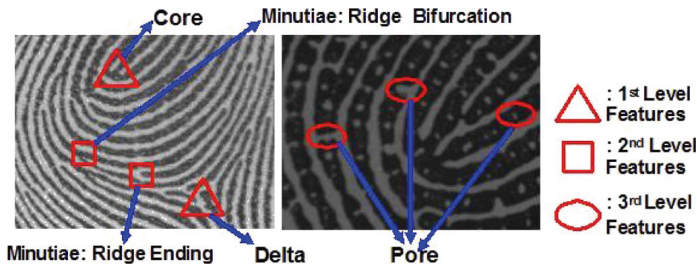


Fig. 6. Three levels of fingerprint features (Zhang et al., 2011)

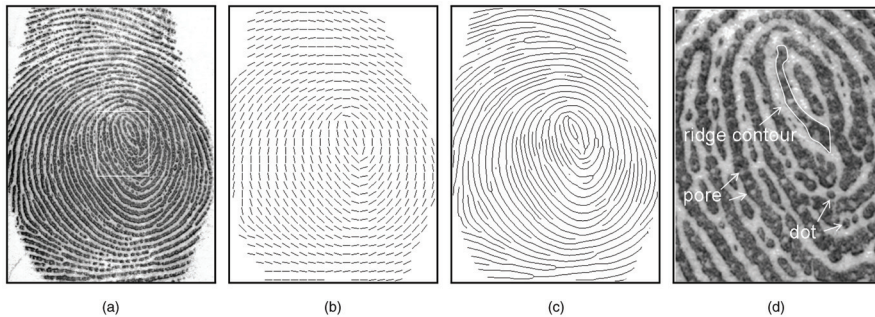


Fig. 7. Features at three levels in a fingerprint. (a) Grayscale image (NIST SD30, A067_11), (b) Level 1 feature (orientation field), (c) Level 2 feature (ridge skeleton), and (d) Level 3 features (ridge contour, pore, and dot) (Feng & Jain, 2011)

Recently, fingerprint features have been classified at three distinctive levels of detail (Feng & Jain, 2011; Zhang et al., 2011), as shown in Figs. 6 and 7. Although their definitions for level-1 and level-2 are different, they agree on the definition of level-3. In (Zhang et al., 2011), level-1 features are the macro details of fingerprints, such as singular points and global ridge patterns, e.g., deltas and cores (indicated by red triangles in Fig. 6). They are not very distinctive and are thus mainly used for fingerprint classification rather than recognition. The level-2 features (red rectangles) primarily refer to the minutiae, namely, ridge endings and bifurcations. Level-2 features are the most distinctive and stable features, which are used in almost all automated fingerprint recognition systems and can reliably be extracted from low-resolution fingerprint images (~500 dpi). A resolution of 500 dpi is also the standard fingerprint resolution of the Federal Bureau of Investigation for automatic fingerprint recognition systems using minutiae (Jain et al., 2007). Level-3 features (red circles) are often defined as the dimensional attributes of the ridges and include sweat pores,

ridge contours, and ridge edge features, all of which provide quantitative data supporting more accurate and robust fingerprint recognition. Among these features, recent researches are focusing on pores (International Biometric Group, 2008; Jain et al., 2006; Jain et al., 2007; Parsons et al., 2008; Zhao et al., 2008; Zhao et al., 2009), where they are considered to be reliably available only at a resolution higher than 500 dpi.

3. Structural approach

One of the early attempts to automate fingerprint recognition was proposed by (Liu & Shelton, 1970). The fundamental concept underlying the proposed system is to use an operator to recognize the ridge characteristics and to impart to a computer the ability to manipulate and compare the digitized locations and directions of these characteristics for single-fingerprint classification. In (Moayer & Fu, 1975) and (Rao & Balck, 1980), patterns were described by means of terminal symbols and production rules. Terminal symbols are associated to small groups of directional elements within the fingerprint directional image. A grammar is defined for each class and a parsing process is responsible for classifying each new pattern. (Moayer & Fu, 1976) demonstrated how a tree system may be used to represent and classify fingerprint patterns. The fingerprint impressions are subdivided into sampling squares which are preprocessed and postprocessed for feature extraction. A set of regular tree languages is used to describe the fingerprint patterns. In order to infer the structural configuration of the encoded fingerprints, a grammatical inference system is developed.

In (Maio & Maltoni, 1996), a well-defined structural approach for fingerprint classification was presented. The basic idea is to perform a directional image partitioning into several homogeneous regular-shaped regions, which are used to build a relational graph summarizing the fingerprint macro-features. The whole approach can be divided into four main steps: computation of the directional image, segmentation of the directional image, construction of the relational graph, and inexact graph matching. The directional image is computed over a discrete grid by means of a robust technique proposed by (Donahue & Rokhlin, 1993). A dynamic clustering algorithm (Maio et al., 1996) is adopted to segment the directional image according to well-suited optimality criteria. In particular, with the aim of creating regions as homogeneous as possible, the algorithm works by minimizing the variance of the element directions within the regions and, simultaneously, by maintaining the regularity of the region shape. Starting from the segmentation of the directional image, a relational graph is built by creating a node for each region and an arc for each pair of adjacent regions. By appropriately labeling the nodes and arcs of the graph, the authors obtained a structure which summarizes the topological features of the fingerprint and is invariant with respect to displacement and rotation.

The PCASYS approach (Pattern-level Classification Automation SYStem) proposed by (Candela & Chellappa, 1993) and (Candela et al., 1995) assigns fingerprints to six non-overlapping classes. Before computing the directional images, the ridge-line area is separated from the background and an enhancement is performed in the frequency domain. The computation of the directions is carried out by the method reported in (Stock & Swonger, 1969). The directional image is then registered with respect to the core position which corresponds to the fingerprint center. The dimensionality of the directional image, considered as a vector of 1,680 elements, is reduced to 64 elements by using the principal component analysis (Jolliffe, 1986). At this stage, a PNN (Probabilistic Neural Network) (Specht, 1990) is used for assigning each 64-element vector to one class of the classification

scheme. In order to improve the classification reliability, especially for whorl fingerprints, the authors also implemented an auxiliary module (called pseudoridge tracer), which works by analyzing the ridge-line concavity under the core position.

(Wahab et al., 1998) described an enhanced fingerprint recognition system consisting of image preprocessing, feature extraction and matching that runs accurately on a personal computer platform. The image preprocessing includes histogram equalization, modification of directional codes, dynamic thresholding and ridgeline thinning. Only the extracted features are stored in a file for fingerprint matching. The matching algorithm presented is a modification and improvement of the structural approach. In their approach, they first divided the original image (320 x 240) into 40 x 30 small areas. Next, each area is assigned a directional code to represent the direction of the ridgeline in that area. To reduce computational time, a total of eight directional codes are used. The eight directional windows w_d ($d = 0, 1, 2, \dots, 7$), each having a length of 16 pixels are shown in Fig. 8. To find the ridge direction of a given area, each of the directional windows, w_d is moved in the direction tangential to the direction of the window. Each of the directional windows will have to move eight times to cover the entire area. At each location when the window moves, the mean value $M(Wd)$ of the grey level of the pixels in the window is calculated. The fluctuation of $M(Wd)$ is expected to be the largest when the movement of the directional window is orthogonal to the direction of the ridges. Therefore this area will be assigned to have ridges in the direction d such that the fluctuation of $M(Wd)$ is the largest.

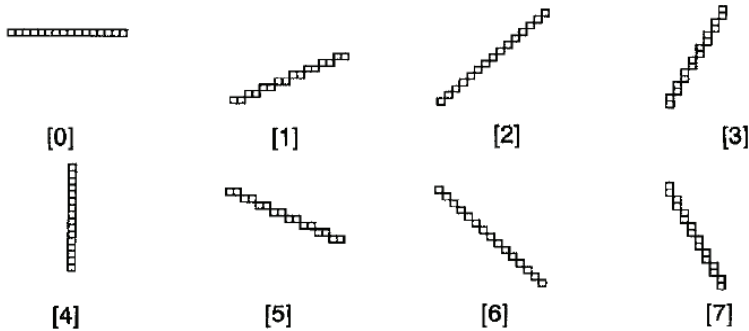


Fig. 8. Eight directional windows Wd for extraction of ridge direction (Wahab et al., 1998)

4. Ridge orientation approach

Since the performance of a minutiae extraction algorithm relies heavily on the quality of the input fingerprint images, it is essential to incorporate a fingerprint enhancement algorithm in the minutiae extraction module to ensure that the performance of the system is robust with respect to the quality of input fingerprint images. In practice, due to variations in impression conditions, ridge configuration, skin conditions (aberrant formations of epidermal ridges of fingerprints, postnatal marks, and occupational marks), acquisition devices, and non-cooperative attitude of subjects, etc., a significant percentage of acquired fingerprint images is of poor quality. The ridge structures in poor-quality fingerprint images are not always well-defined and, hence, they cannot be correctly detected. This leads to following problems:

1. a significant number of spurious minutiae may be created,
2. a large percent of genuine minutiae may be ignored, and
3. large errors in their localization (position and orientation) may be introduced.

In order to ensure that the performance of the minutiae extraction algorithm is robust with respect to the quality of the input fingerprint images, an enhancement algorithm that improves the clarity of the ridge structures is necessary. Fingerprint enhancement can be conducted on either: 1) binary ridge images or, 2) gray-level images.

A binary ridge image is an image where all the ridge pixels are assigned a value one and nonridge pixels are assigned a value zero. The binary image can be obtained by applying a ridge extraction algorithm on a gray-level fingerprint image. Since ridges and valleys in a fingerprint image alternate and run parallel to each other in a local neighborhood, a number of simple heuristics can be used to differentiate the spurious ridge configurations from the true ridge configurations in a binary ridge image. However, after applying a ridge extraction algorithm on the original gray-level images, information about the true ridge structures is often lost depending on the performance of the ridge extraction algorithm. Therefore, enhancement of binary ridge images has its inherent limitations. In a gray-level fingerprint image, ridges and valleys in a local neighborhood form a sinusoidal-shaped plane wave which has a well-defined frequency and orientation.

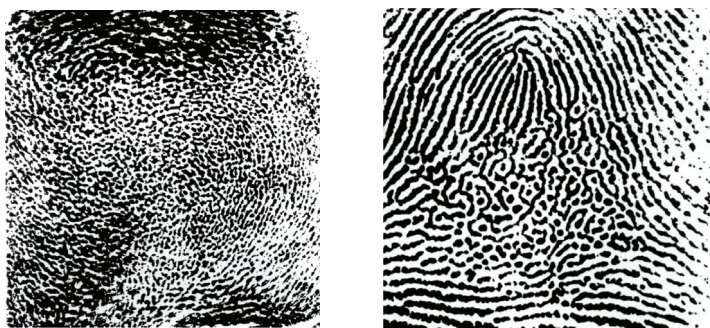


Fig. 9. Fingerprint images of very poor quality (Hong et al., 1998)

(Hong et al., 1998) presented a fast fingerprint enhancement algorithm, which can adaptively improve the clarity of ridge and valley structures of input fingerprint images based on the estimated local ridge orientation and frequency using both the local ridge orientation and local frequency information. (Vaikol et al., 2009) presented a reliable method of computation for minutiae feature extraction from fingerprint images. The scheme relies on describing the orientation field of the fingerprint pattern with respect to each minutia detail. A fingerprint image is treated as a textured image, where an orientation flow field of the ridges is computed. To accurately locate ridges, a ridge orientation based computation method is used. After ridge segmentation, smoothing is done using morphological operators.

(Choi et al., 2010) introduced a novel fingerprint matching algorithm using both ridge features and the conventional minutiae features to increase the recognition performance against nonlinear deformation in fingerprints. The proposed ridge features are composed of four elements: ridge count, ridge length, ridge curvature direction, and ridge type. These ridge features have some advantages in that they can represent the topology information in

entire ridge patterns that exist between two minutiae and are not changed by non-linear deformation of the finger. For extracting ridge features, they have also defined the ridge-based coordinate system in a skeletonized image. With the proposed ridge features and conventional minutiae features (minutiae type, orientation, and position), they have proposed a novel matching scheme using a breadth first search to detect the matched minutiae pairs incrementally (Fig. 10).

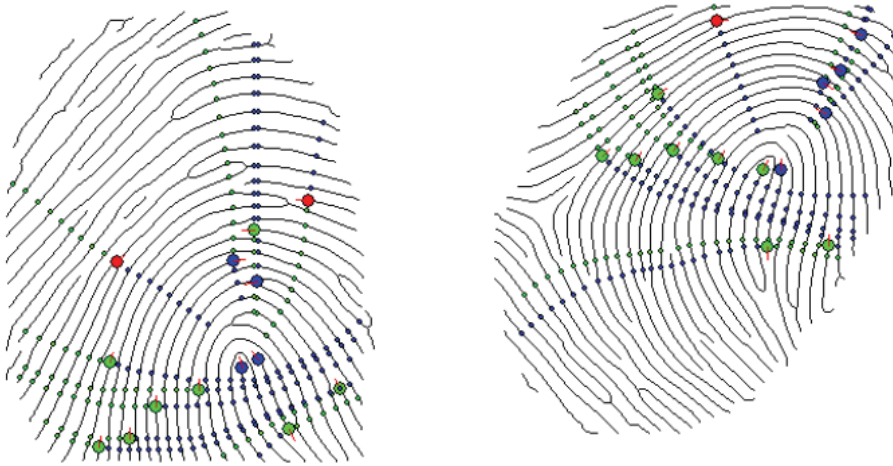


Fig. 10. Example of matched minutiae using the proposed ridge feature vectors (solid circles represent matched minutiae and dotted lines represent the vertical axis of each minutia) (Choi et al., 2010)

5. Pixel-level approach

(Abutaleb & Kamel, 1999) used the fact that a fingerprint is made of white followed by black lines of bounded number of pixels. This enabled the problem formulation to be cast as a parametric optimization problem. The parameters are the widths of the black and white lines in the scanned line in the fingerprint. The proposed adaptive genetic algorithm proved to be effective in determining the ridges or edges in the fingerprint. Further, (Ceguema & Koprinska, 2002) presented an approach for combining local and global recognition schemes for automatic fingerprint verification by using matched local features as the reference axis for generating global features. In their implementation, minutia-based and shape-based techniques were combined. The first one matches local features (minutiae) by a point-pattern matching algorithm. The second one generates global features (shape signatures) by using the matched minutiae as its frame of reference. Shape signatures are then digitized to form a feature vector describing the fingerprint. Finally, a Learning Vector Quantization neural network was trained to match the fingerprints using the difference between a pair of feature vectors.

In (Zhang et al., 2010), investigation has been conducted on analyzing the mechanisms of fingerprint image rotation processing and its potential effects on the major features, mainly minutiae and singular point, of the rotation transformed fingerprint. It was observed that

the information integrity of the original fingerprint image can be significantly compromised by the image rotation transformation process, which can cause noticeable singular point change and produce non-negligible number of fake minutiae. It is found that the quantization and interpolation process can change the fingerprint features significantly though they may not change the image visually. Their experimental results have shown that up to 7% of the minutiae can be mis-matched. For the matched ones, their positions deviate up to 16 pixels. The position of singular point can change up to 55 pixels while the orientation angle change can be up to 90 degrees. (Kaur et al., 2010) proposed an approach for feature extraction based on dividing the image into equal sized blocks. Each block is processed independently. The gray level projection along a line perpendicular to the local orientation field provides the maximum variance. Then the ridges are located using the peaks and the variance in this projection. The ridges are thinned and the resulting image is enhanced using an adaptive morphological filter.

Square-based method was presented in (Gamassi et al., 2005) and (Alibeigi et al., 2009). The Square-based method is composed of the following steps, repeated for each pixel of the binary image:

1. Create a 3x3 square mask around the (x, y) pixel and compute the average of the pixels. If the average is less than 0.25 the pixel is preliminary identified as a ridge termination minutiae, otherwise if the average is greater than 0.75 the pixel is treated like a bifurcation minutiae.
2. Create a square perimeter P around the (x, y) pixel of size $W \times W$.
3. Compute the number of the logic commutations present in the perimeter P without considering isolated pixels as shows in Fig. 11.

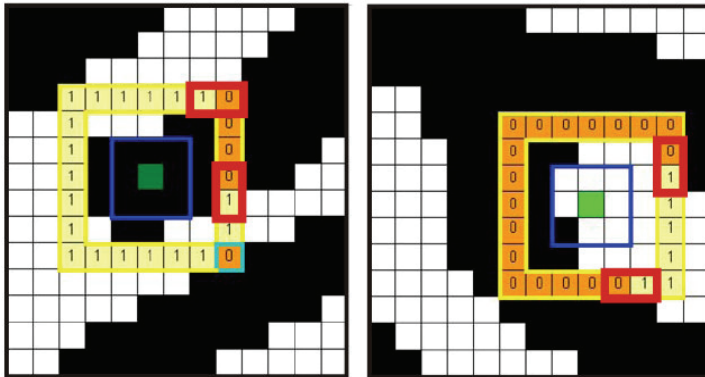


Fig. 11. Processing the commutations (Gamassi et al., 2005)

4. The algorithm continues if there are two logic commutations, otherwise it jumps to step 1 processing another pixel.
5. Compute the average of the pixels in the perimeter P . If the pixel has been defined as a termination minutiae in step 1, it checks if the average is greater than the threshold K . (in bifurcation minutiae, the average must be less than $1-K$) otherwise it jumps to step 1 processing another pixel.
6. Estimate the orientation angle α in the minutiae point.
7. False detection removal (Fig. 12).

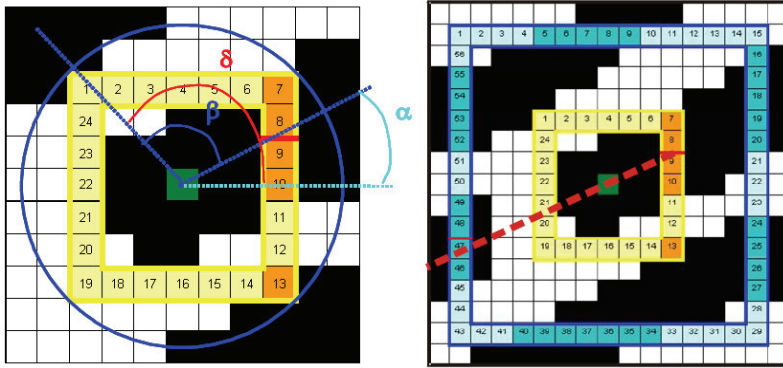


Fig. 12. Estimating the orientation angle (left) and removing the false terminations detections (right) (Gamassi et al., 2005)

6. Filtering and wavelet approach

(Lee & Wang, 1999) have developed a one-step method using Gabor filters for directly extracting fingerprint features for a small-scale fingerprint recognition system. From the experimental results, the use of magnitude Gabor features with eight orientations as fingerprint features led to good shift-invariant properties and an accuracy of 97.2% with 3-NN classifiers, on a database of 192 inked fingerprint images from 16 persons. In (Watson et al., 1994) and (Willis & Myers, 2001), the fingerprint's blockwise Fourier transform is multiplied by its power spectrum raised to a power, thus magnifying the dominant orientation.

Also, a Laplacian-like image pyramid is used to decompose the original fingerprint into sub-bands corresponding to different spatial scales for fingerprint enhancement. The Laplacian pyramid (Adelson et al., 1984), (Simoncelli & Freeman, 1995) is equivalent to bandpass filtering in the spatial domain. In a further step, contextual smoothing is performed on these pyramid levels, where the corresponding filtering directions stem from the frequency-adapted structure tensor. For minutiae extraction, parabolic symmetry is added to the local fingerprint model which allows to accurately detecting the position and direction of a minutia simultaneously.

Also, (Cappelli et al., 1999) have implemented the directional approach. The directional image is partitioned into "homogeneous" connected regions according to the fingerprint topology, thus giving a synthetic representation which can be exploited as a basis for the classification. A set of dynamic masks, together with an optimization criterion, was used to guide the partitioning. The adaptation of the masks produces a numerical vector representing each fingerprint as a multidimensional point, which can be conceived as a continuous classification. Different search strategies were discussed to efficiently retrieve fingerprints both with continuous and exclusive classification. A directional image is a discrete matrix whose elements represent the local average directions of the fingerprint ridge lines.

(Hong et al., 1998) proposed an algorithm using Gabor bandpass filters tuned to the corresponding ridge frequency and orientation to remove undesired noise while preserving the true ridge-valley structures. All operations are performed in the spatial domain, whereas

the contextual filtering in (Sherlock et al., 1994) and (Chikkerur & Govindaraju, 2005) is done in the Fourier domain. From (Tico et al., 2001), the discrete wavelet transform (DWT) coefficients have been used as the ridge pattern. The authors discussed that the middle frequency has an oscillated pattern corresponding to the ridge pattern. Then, to extract wavelet features from a gray-scale fingerprint image, the image was first cropped to the size of 64×64 pixels, where the center point in the image is referred to as a *reference point*. The cropped image was then quartered, centered at the reference point, to obtain four non-overlapping images of size 32×32 pixels. After applying the DWT to each non-overlapping image four times, twelve sub-images in the wavelet domain at each decomposition level are created as shown in Fig. 13. Next, the standard deviation of the DWT coefficients from each sub-image is computed to create a feature vector of length 48 (12 DWT sub-images from 4 non-overlapping images). The resulted feature vector is then used as a representation of that fingerprint image.

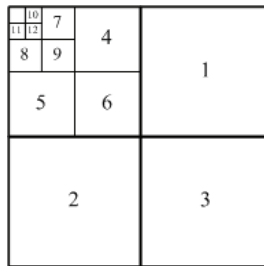


Fig. 13. Arrangements of twelve sub-images in the wavelet domain (Tico et al., 2001)

On the other hand, it was shown in (Tachaphetpiboon & Amornraksa, 2005, 2007) that the discrete cosine transform (DCT) is better suited in extracting informative features than the DWT. The results have shown that the fingerprint matching system based on the DCT obtained a high recognition rate and a lower complexity. (Tachaphetpiboon & Amornraksa, 2005) proposed to divide all the DCT coefficients containing the oscillate pattern in a zigzag-scanned fashion, extract DCT features from the divided DCT coefficients, and then use them for fingerprint matching. Accordingly, all the DCT coefficients in each non-overlapping image were divided into 12 areas, where one feature was extracted from each one of these areas, generating 12 features for each non-overlapping image.

(Fronthaler et al., 2008) proposed the use of an image-scale pyramid and directional filtering in the spatial domain for fingerprint image enhancement to improve the matching performance as well as the computational efficiency. Image pyramids or multiresolution processing is especially known from image compression and medical image processing. (Fronthaler et al., 2008) expected that all the relevant information to be concentrated within a few frequency bands. Furthermore, they have proposed Gaussian directional filtering to enhance the ridge-valley pattern of a fingerprint image using computationally cheap 1-D filtering on higher pyramid levels (lower resolution) only. The filtering directions are recovered from the orientations of the structure tensor (Bigun, 2006) at the corresponding pyramid level. Linear symmetry features are thereby used to extract the local ridge-valley orientation (angle and reliability).

(On et al., 2006) presented a filtering strategy that can solve the problem of rotated scanned input images. The fingerprint image is scanned with an optical fingerprint scanner. The

scanned fingerprint image is saved in bitmap format with black and white colour. The scanned fingerprint image is then enhanced for quality improvement. Further, the enhanced fingerprint image is applied for binarization. The conversion is needed to reduce the computation and analysis time for filtering and thinning process. The noise produced from the binarized fingerprint image is then removed using median filtering and the filtered fingerprint image is further thinned. After that, the bifurcation minutiae extraction method is applied for the thinned fingerprint image. The extracted feature data are then used for neural network training.

The concept of spectral minutiae representation was used by (Xu & Veldhuis, 2009). The spectral minutiae representation is based on the shift, scale and rotation properties of the two-dimensional continuous Fourier transform. Assume a fingerprint with Z minutiae. In location-based spectral minutiae representation (SML), with every minutia, a function $m_i(x, y) = \delta(x - x_i, y - y_i)$, $i = 1, \dots, Z$ is associated where (x_i, y_i) represents the location of the i -th minutia in the fingerprint image. Thus, in the spatial domain, every minutia is represented by a Dirac pulse. The Fourier transform of $m_i(x, y)$ is given by:

$$\mathcal{F}\{m_i(x, y)\} = \exp(-j(\omega_x x_i + \omega_y y_i)) \quad (1)$$

and the location-based spectral minutiae representation is defined as

$$\mathcal{M}_L(\omega_x, \omega_y) = \sum_{i=1}^Z \exp(-j(\omega_x x_i + \omega_y y_i)). \quad (2)$$

In order to reduce the sensitivity to small variations in minutiae locations in the spatial domain, a Gaussian low-pass filter is used to attenuate the higher frequencies. This multiplication in the frequency domain corresponds to a convolution in the spatial domain where every minutia is now represented by a Gaussian pulse. Following the shift property of the Fourier transform, the magnitude of M is taken in order to make the spectrum invariant to translation of the input, and we obtain

$$\left| \mathcal{M}_L(\omega_x, \omega_y; \sigma_L^2) \right| = \left| \exp\left(-\frac{\omega_x^2 + \omega_y^2}{2\sigma_L^2}\right) \sum_{i=1}^Z \exp(-j(\omega_x x_i + \omega_y y_i)) \right|. \quad (3)$$

Then, the orientation information in the spectral representation is included. The orientation θ of a minutia can be incorporated by using the spatial derivative of $m(x, y)$ in the direction of the minutia orientation. Thus, to every minutia in a fingerprint, a function $m_i(x, y, \theta)$ is assigned being the derivative of $m_i(x, y)$ in the direction θ_i , such that

$$\mathcal{F}\{m_i(x, y, \theta)\} = j(\omega_x \cos \theta_i + \omega_y \sin \theta_i) \cdot \exp(-j(\omega_x x_i + \omega_y y_i)). \quad (4)$$

As with the SML algorithm, using a Gaussian filter and taking the magnitude of the spectrum yields

$$|M_O(\omega_x, \omega_y; \sigma_O^2)| = \left| \exp\left(-\frac{\omega_x^2 + \omega_y^2}{2\sigma_O^2}\right) \sum_{i=1}^Z j(\omega_x \cos \theta_i + \omega_y \sin \theta_i) \cdot \exp(-j(\omega_x x_i + \omega_y y_i)) \right|. \quad (5)$$

Recently, (Xu & Veldhuis, 2010) have further discussed the objective of the spectral minutiae representation in representing a minutiae set as a fixed-length feature vector that is invariant to translation, rotation and scaling. Fig. 14 illustrates a general procedure of the spectral minutiae representation discussed by (Xu & Veldhuis, 2010).

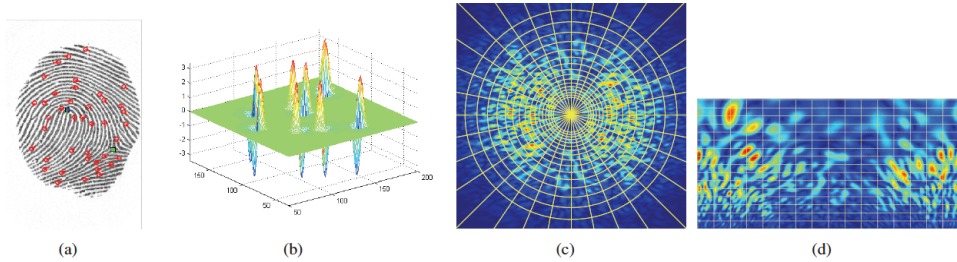


Fig. 14. Illustration of the general spectral minutiae representation procedure. (a) a fingerprint and its minutiae; (b) representation of minutiae points as real (or complex) valued continuous functions; (c) the 2D Fourier spectrum of ‘b’ in a Cartesian coordinate and a polar-logarithmic sampling grid; (d) the Fourier spectrum sampled on a polar-logarithmic grid (Xu & Veldhuis, 2010)

Moreover, based on the spectral minutiae feature, (Xu et al., 2009a, 2009b) introduced two feature reduction methods: the Column-Principal Component Analysis (PCA) and the Line-Discrete Fourier Transform feature reduction algorithms. The experiments demonstrated that these methods decrease the minutiae feature dimensionality with a reduction rate of 94%, while at the same time, the recognition performance of the fingerprint system is not degraded. On the other hand, (Dadgostar et al., 2009) presented a novel feature extraction method based on Gabor filter and Recursive Fisher Linear Discriminate (RFLD) algorithm, for fingerprint identification. The proposed method was assessed on images from the biolab database (Biometric System Lab). Experimental results have shown that applying RFLD to a Gabor filter in four orientations, in comparison with Gabor filter and PCA transform, increases the identification accuracy from 85.2% to 95.2% by nearest cluster center point classifier with Leave-One-Out method. Also, it has been shown that applying RFLD to a Gabor filter in four orientations, in comparison with Gabor filter and PCA transform, increases the identification accuracy from 81.9% to 100% by 3NN classifier.

7. Geometric approach

(Chen et al., 2009) proposed an algorithm to use minutiae for fingerprint recognition, in which the fingerprint’s orientation field is reconstructed from minutiae and further utilized in the matching stage to enhance the system’s performance. First, they have produced “virtual” minutiae by using interpolation in the sparse area, and then used an orientation model to reconstruct the orientation field from all “real” and “virtual” minutiae (Fig. 15). A

decision fusion scheme is used to combine the reconstructed orientation field matching with the conventional minutiae-based matching. (Min & Thein, 2009) have presented a recognition system which combines both the statistical and geometry approaches. The core point (CP) of the input fingerprint is detected and located in the centre. Then, the fingerprint image is cropped around the based point. Fingerprint features such as minutiae points' determination, their coordinates location, and radius of arcs for each ridge are stored in different databases. For a testing fingerprint image, the features are compared with these pre-defined databases and the decision is made by a voting system.

In (Wei-bo et al., 2008), each minutia was defined by the type and the relative topological relationship among the minutia and its 5 nearest neighbors. (Qi et al., 2008) proposed a fingerprint matching algorithm using the elaborate combination of minutiae and curvature maps from fingerprint images. First, they computed the curvature in a simple way based on orientation field, and then performed the sampling operation on the curvature map around each minutia to get the fixed length minutiae specifiers. Second, a similarity measurement was defined between two specifiers. Third, they found the reference points pair based on computing the least squared error of Euclidean distance between these two specifiers. Finally, they completed the matching task by aligning the two fingerprint minutiae sets and accounting the number of overlapping minutiae.

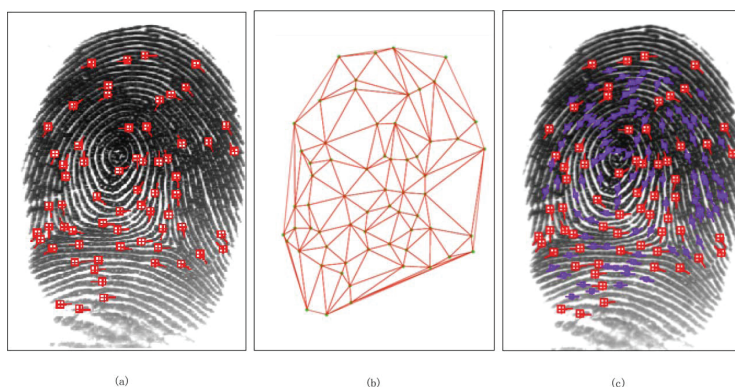


Fig. 15. Interpolation step: (a) the minutiae image; (b) the triangulated image; (c) virtual minutiae by interpolation (the bigger red minutiae are “real,” while the smaller purple ones are “virtual”) (Chen et al., 2009)

8. Singularity approach

The global features of a fingerprint are singularity points, namely core and delta. Generally, singularity points are used to classify fingerprint images to reduce the search space. (Kryszczuk & Drygajlo, 2006) presented a method in which the singular point detection is performed by analyzing the local quadrant change of the ridge gradient vectors. Singular points (SP) are defined as discontinuities in the directional field (Liu et al., 2005). In formally, this can be stated as the area where ridges oriented rightwards change to leftwards and those that were oriented upwards turn downwards, and opposite. Their algorithm performs a robust estimation of the local ridge gradient. They employed a modified version of the “*squared average gradients*” to estimate the direction of the smoothed gradient vectors.

Also, they allowed cancelling out the opposite local gradients, achieving a more robust average local ridge gradient estimation. Moreover, (Militello et al., 2008) proposed a fingerprint recognition approach based on core and delta singularity points detection. The singularity points extraction is performed using three sequential steps: directional image extraction, Poincaré indexes computation and core and delta extraction. The approach has shown a good accuracy level in the singularity points detection and extraction and a low computational cost.

(Conti et al., 2010) proposed another fingerprint recognition that is based on singularity points detection and singularity regions analysis. Despite to the classical minutiae-based fingerprint recognition system, the proposed system is based on core and delta position, their relative distance and orientation to perform both classification and matching tasks. The proposed approach enhances the performance of singularity points based methods by introducing *pseudo-singularity points* when the standard singularity points (core and delta) cannot be extracted. As a result, singularity points and/or *pseudosingularity-points* are detected and extracted to make possible successful fingerprint classification and recognition. After singularity points extraction, a roto-translation operation is applied for fingerprint image registration. Finally, a matching algorithm based on morphological operation, such as dilation and erosion, on two considered regions of interests (singularity regions or pseudo singularity regions) around core and delta is performed. The obtained similarity degree considering the regions of interest gives the matching result. The experimental results have shown good accuracy levels, reaching a FAR=1.22% and a FRR=9.23% using FVC2002 DB2-A database, and in the best of case, a FAR=0.26% and a FRR=7.36% using FVC2000 DB1-B database.

9. Pore approach

Recently, researchers have focused on pores as a distinctive fingerprint features (International Biometric Group, 2008; Jain et al., 2006, 2007; Parsons et al., 2008; Zhang et al., 2011; Zhao et al., 2008; Zhao et al., 2009). Focusing on this kind of features depends heavily on the quality of the digital fingerprint image. Resolution is one of the main parameters affecting the quality of a digital fingerprint image, and so, it has an important role in the design and deployment of fingerprint recognition systems and impacts both their cost and recognition performance. Despite this, the field of fingerprint recognition does not currently have a well-proven reference resolution or standard resolution that can be used interoperably between different systems. For example, (Jain et al., 2007) chose a resolution of 1,000 dpi based on the 2005 ANSI/NIST fingerprint standard update workshop. (Zhao et al., 2008, 2009) proposed some pore extraction and matching methods at a resolution of 902 dpi x 1200 dpi. Finally, the International Biometric Group analyzed level-3 features at a resolution of 2000 dpi.

(Zhao & Jain, 2010) have studied the utility of pores on rolled ink fingerprint images which are widely used in forensic applications. Fingerprint images of three different qualities at two different resolutions (500ppi and 1000ppi) were considered in their experiments. By using NIST SD30 database, and a commercial minutiae matcher, they have investigated the impact of fingerprint image quality on the accuracy of automatic pore extraction, and the effectiveness of pores in improving fingerprint recognition accuracy. The experimental results have shown that the (i) pores do not provide any significant improvement to the fingerprint recognition accuracy on 500ppi fingerprint images, and (ii) fusion of pore and

minutiae matchers is effective only for high resolution (1000ppi) fingerprint images of good quality.

(Zhang et al., 2011) have taken further steps toward establishing a reference resolution, assuming a fixed image size and making use of the two most representative fingerprint features, i.e., minutiae and pores, and providing a minimum resolution for pore extraction that is based on anatomical evidence. They conducted experiments on a set of fingerprint images of different resolutions (from 500 to 2000 dpi). By evaluating these resolutions in terms of the number of minutiae and pores, their results have shown that 800 dpi would be a good choice for a reference resolution. (Malathi & Meena, 2010) presented a suitable technique for partial fingerprint matching based on pores and its corresponding Local Binary Pattern (LBP) features. The first step involves extracting the pores from the partial image. These pores act as anchor points where a sub window (32x32) is formed to surround them. Then rotation invariant LBP histograms are obtained from this surrounding window. Finally, a chi-square formula is used to calculate the minimum distance between two histograms to find the best matching score.

10. Other approaches

In this section we present an overview of some other general techniques proposed for fingerprint recognition. Neural network approaches are mostly based on multilayer perceptrons or Kohonen self-organizing networks (Bowen, 1992; Hughes & Green, 1991; Kamijo, 1993; Moscinska & Tyma, 1993). In particular, (Kamijo, 1993) presented an interesting pyramidal architecture constituted by several multilayer perceptrons, each of which was trained to recognize fingerprints belonging to different classes. (Wang et al., 2008) proposed the cellular neural/nonlinear network (CNN) as a powerful tool for fingerprint feature extraction. They presented two theorems for designing two kinds of CNN templates. These two theorems provided the template parameter inequalities to determine parameter intervals for implementing the corresponding functions. (Senior, 1997) proposed a hidden Markov model classifier whose input features are the measurements (ridge angle, separation, curvature, etc.) taken at the intersection points between some horizontal- vertical fiducial lines and the fingerprint ridge lines. (Yang et al., 2008) proposed a fingerprint matching algorithm based on invariant moments. (Montesanto et al., 2007) have studied the fingerprint verification based on the fuzzy logic, where they combined the results obtained using three different methods of minutiae extraction: the sequential method, the reactive agent and the neural classification system. (Puertas et al., 2010) studied the performance of a fingerprint recognition technology, in several practical scenarios of interest in forensic casework. First, the differences in performance between manual and automatic minutiae extraction for latent fingerprints were presented. Then, automatic minutiae extraction was analyzed using three different types of fingerprints: latent, rolled and plain. The experiments were carried out using a database of latent fingermarks and fingerprint impressions from real forensic cases.

11. Quality assessment methods

Fingerprint quality is usually defined as a measure of the clarity of ridges and valleys and the extractability of the features used for identification such as minutiae, core and delta points, etc. Therefore, it is important to estimate the quality and validity of the fingerprint

images in order to improve recognition performance. A number of factors can affect the quality of fingerprint images (Joun et al., 2003): occupation, motivation/collaboration of users, age, temporal or permanent cuts, dryness/wetness conditions, temperature, dirt, residual prints on the sensor surface, etc. Unfortunately, many of these factors cannot be controlled and/or avoided. For this reason, assessing the quality of captured fingerprints is important for a fingerprint recognition system.

(Qi et al., 2005) proposed a hybrid scheme to measure quality by considering local and global features. They used seven quality indices and analyzed the correlation between the quality value and each quality index. (Chen et al., 2005) suggested quality estimation methods based on the power spectrum analysis in the 2-D Fourier domain and coherence in the spatial domain.

ISO/INCITS-M1 (International Standards Organization/International Committee for Information Technology Standards) has established a biometric sample-quality draft standard (Int. Com. Inf. Technol. Standards, 2005), in which a biometric sample quality is considered from three different points of view: 1) character, which refers to the quality attributable to inherent physical features of the subject; 2) fidelity, which is the degree of similarity between a biometric sample and its source, attributable to each step through which the sample is processed; and 3) utility, which refers to the impact of the individual biometric sample on the overall performance of a biometric system, where the concept of sample quality is a scalar quantity that is related monotonically to the performance of the system. The character of the sample source and the fidelity of the processed samples contribute to, or similarly detract from, the utility of the sample. It is generally accepted that the utility is most importantly mirrored by a quality metric (Grother & Tabassi, 2007), so that images assigned higher quality shall necessarily lead to better identification of individuals (i.e., better separation of genuine and impostor match score distributions).

A theoretical framework for a biometric sample quality has been developed by (Youmaran & Adler, 2006), where they relate biometric sample quality with the identifiable information contained. They defined "Biometric information" (BI) as the decrease in uncertainty about the identity of a person due to a set of biometric measurements. BI is calculated by the relative entropy between the population feature distribution and the person's feature distribution. The results reported by (Youmaran & Adler, 2006) show that degraded biometric samples result in a decrease in BI. In (Van derWeken et al., 2007), a number of quality metrics can be found aimed at objectively assessing the quality of an image in terms of the similarity between a reference image and a degraded version of it.

Finally, (Lee et al., 2008) proposed a fingerprint-quality measurement method based on the shapes of several probability density functions (PDFs). The 2-D gradients of the fingerprint images are first separated into two sets of 1-D gradients. Then, the shapes of the PDFs of these gradients are measured in order to determine the fingerprint quality.

12. Conclusion

In this chapter, we have presented a study covering different automatic fingerprint recognition techniques, presented by the experts in this field. Although many academic and commercial systems for fingerprint recognition exist, there is a necessity for further research in this topic in order to improve the reliability and performances of the current systems. Many unresolved problems still need to be explored and investigated. For example, for a

large automated fingerprint identification system, the recognition accuracy, matching speed and its robustness to poor image quality are normally regarded as the most critical elements of system performance. Also, fast comparison algorithm is necessary since most minutiae-based matching algorithms will fail to meet the high speed requirement. Further, matching partial fingerprints still needs lots of improvement. The major challenges faced in partial fingerprint matching are the absence of sufficient level 2 features (minutiae) and other structures such as core and delta. Thus, common matching methods based on alignment of singular structures would fail in case of partial prints. Pores (level 3 features) on fingerprints have proven to be discriminative features and have recently been successfully employed in automatic fingerprint recognition systems. Finally, there is still a lot of research to be done when dealing with latent fingermarks. Low quality, incompleteness and distortion are typical problems that forensic fingerprint recognition systems have to face when extracting features from latent fingermarks.

13. Acknowledgments

The author would like to acknowledge and thank Kuwait Foundation for the Advancement of Sciences (KFAS) for financially supporting this work.

14. References

- Abutaleb, A. S. & Kamel, M. (1999). A Genetic Algorithm for the Estimation of Ridges in Fingerprints, *IEEE Trans. On Image Processing*, Vol. 8, No. 8, pp. 1134-1139, AUGUST 1999
- Adelson, E. H.; Anderson, C. H.; Bergen, J. R.; Burt, P. J. & Ogden, J. M. (1984). Pyramid methods in image processing, *RCA Eng.*, Vol. 29, No. 6, pp. 33-41, 1984
- Alibeigi, E.; Rizi M. T. & Behnamfar, P. (2009). Pipelined Minutiae Extraction From Fingerprint Images, *Proceedings of the IEEE*, 2009
- Bazen, A.M. & Gerez, S.H. (2003). Fingerprint Matching by Thin-Plate Spline Modelling of Elastic Deformations, *Pattern Recognition*, vol. 36, no. 8, pp. 1859-1867, Aug. 2003
- Bazen, A.M.; Verwaaijen, G.T.B.; Gerez, S.H.; Veelenturf, L.P.J. & Van der Zwaag, B.J. (2000). A Correlation-Based Fingerprint Verification System, *Proc. 11th Ann. Workshop Circuits Systems and Signal Processing*, pp. 205-213, Nov. 2000
- Bigun, J. (2006). *Vision With Direction*. New York: Springer, 2006 *Biometric System Lab.*, University of Bologna, Cesena-Italy. (www.csr.unibo.it/research/biolab/)
- Bowen, J.D. (1992). The Home Office Automatic Fingerprint Pattern Classification Project, *Proc. IEE Colloquium Neural Network for Image Processing Applications*, 1992
- Candela, G.T. & Chellappa, R. (1993). *Comparative Performance of Classification Methods for Fingerprints*, NIST Technical Report NISTIR 5163, Apr. 1993
- Candela, G.T. & et al. (1995). *PCASYS – A Pattern-Level Classification Automation System for Fingerprints*, NIST Technical Report NISTIR 5647, Aug. 1995
- Cappelli, R.; Lumini, A.; Maio, D. & Maltoni, D. (1999). Fingerprint Classification by Directional Image Partitioning, *IEEE Trans. On Pattern Analysis And Machine Intelligence*, Vol. 21, No. 5, pp. 402-421, MAY 1999

- Ceguema, A. V. & Koprinska, I. (2002). Integrating Local and Global Features in Automatic Fingerprint Verification, *Proceedings of the 16th International Conference on Pattern Recognition*, 2002
- Chen, Y.; Dass, S. & Jain, A. (2005). Fingerprint quality indices for predicting authentication performance, *Proceedings 5th Int. Conf. Audioand Video-based Biometric Person Authentication*, Rye Brook, NY, Jul. 2005, pp. 160-170
- Chen, F.; Zhou, J. & Yang, C. (2009). Reconstructing Orientation Field From Fingerprint Minutiae to Improve Minutiae-Matching Accuracy, *IEEE Transactions On Image Processing*, Vol. 18, No. 7, pp. 1665- 1670, JULY 2009
- Chikkerur, S. & Govindaraju, V. (2005). Fingerprint image enhancement using STFT analysis, *Proc. Int. Workshop on Pattern Recognition for Crime Prevention, Security and Surveillance*, 2005, pp. 20-29
- Choi, H.; Choi, K. & Kim, J. (2010). Fingerprint Matching Incorporating Ridge Features with Minutiae, *Prodeedings of the IEEE*, 2010
- Conti, V.; Militello, C.; Sorbello, F. & Vitabile, S. (2010). Introducing Pseudo-Singularity Points for Efficient Fingerprints Classification and Recognition, *International Conference on Complex, Intelligent and Software Intensive Systems*, 2010
- Dadgostar, M.; Tabrizi, P. R.; Fatemizadeh, E. & Soltanian-Zadeh, H. (2009). Feature Extraction Using Gabor-Filter and Recursive Fisher Linear Discriminant with Application in Fingerprint Identification, *7th Int'l Conf. on Advances in Pattern Recognition*, 2009
- Donahue, M.J. & Rokhlin, S.I. (1993). On the Use of Level Curves in Image Analysis, *Image Understanding*, Vol. 57, no. 3, pp. 185-203, 1993
- Feng, J. & Jain, A. K. (2011). Fingerprint Reconstruction: From Minutiae to Phase, *IEEE Trans. On Pattern Analysis And Machine Intelligence*, Vol. 33, No. 2, pp. 209-223, FEB. 2011
- Feng, J.; Ouyang, Z. & Cai, A. (2006). Fingerprint Matching Using Ridges, *Pattern Recognition*, vol. 39, no. 11, pp. 2131-2140, 2006
- Fronthaler, H.; Kollreider, K. & Bigun, J. (2008). Local Features for Enhancement and Minutiae Extraction in Fingerprints, *IEEE Transactions On Image Processing*, Vol. 17, No. 3, pp. 354-363, MARCH 2008
- Galton, F. (1892). *Finger Prints*. London: McMillan, 1892
- Gamassi, M.; Piuri, V. & Scotti, F. (2005). Fingerprint local analysis for high-performance minutiae extraction, *IEEE ICIP*, Vol. 3, pp. 265-268, Sept 2005
- Grother, P. & Tabassi, E. (2007). Performance of biometric quality measures, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 4, pp. 531-543, Apr. 2007
- Hara, M. & Toyama, H. (2007). *Method and Apparatus for Matching Streaked Pattern Image*, US Patent No. 7,295,688, 2007
- Henry, E.R. (1900). *Classification and Uses of Finger Prints*. London: Routledge, 1900
- Hong, L.; Wan, Y. & Jain, A. (1998). Fingerprint Image Enhancement: Algorithm and Performance Evaluation, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 20, No. 8, pp. 777-789, AUGUST 1998
- Hughes, P.A. & Green, A.D.P. (1991). The Use of Neural Network for Fingerprint Classification, *Proc. Second Int'l Conf. Neural Network*, pp. 79-81, 1991
- International Biometric Group. (2008). *Analysis of Level 3 Features at High Resolutions (Phase II)*, 2008

- International Com. Inf. Technol. Standards. (2005). *Biometric Sample Quality Std. Draft (Rev. 4)*, Document M1/06-0003, 2005
- Jain, A.; Chen, Y. & Demirkus, M. (2006). Pores and ridges: Fingerprint matching using level 3 features, *Proc. 18th Int. Conf. Pattern Recog.*, 2006, pp. 477-480
- Jain, A.; Chen, Y. & Demirkus, M. (2007). Pores and ridges: High-resolution fingerprint matching using level-3 features, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 15-27, Jan. 2007
- Jolliffe, I.T. (1986). *Principle Component Analysis*. New York: Springer-Verlag, 1986
- Joun, S.; Kim, H.; Chung, Y. & Ahn, D. (2003). An experimental study on measuring image quality of infant fingerprints, *Proceedings of the KES*, 2003, pp. 1261-1269
- Kamijo, M. (1993). Classifying Fingerprint Images Using Neural Network: Deriving the Classification State, *Proc. Third Int'l Conf. Neural Network*, pp. 1,932-1,937, 1993
- Kaur, R.; Sandhu, P. S. & Kamra, A. (2010). A Novel Method For Fingerprint Feature Extraction, *International Conference on Networking and Information Technology*, 2010
- Kawagoe, M. & Tojo, A. (1984). Fingerprint Pattern Classification, *Pattern Recognition*, vol 17, no. 3, pp. 295-303, 1984
- Kryszczuk, K. & Drygajlo, A. (2006). Singular point detection in fingerprints using quadrant change information, *18th International Conf. on Pattern Recognition*, 2006
- Lee, C.-J. & Wang, S.-D. (1999). Fingerprint feature extraction using Gabor filters, *Electronics Letters*, 18th February, Vol. 35 No. 4, pp. 288-290, 1999
- Lee, S.; Choi, H.; Choi, K. & Kim, J. (2008). Fingerprint-Quality Index Using Gradient Components, *IEEE Tran. On Information Forensics And Security*, Vol. 3, No. 4, pp. 792-800, DEC. 2008
- Liu, C. N. & Shelton, G. L. Jr. (1970). Computer-Assisted Fingerprint Encoding and Classification, *IEEE Tran. on Man-Machine Systems*, , pp. 156-160, SEPT. 1970
- Liu, T.; Hao, P. & Zhang, C. (2005). Fingerprint Singular Points Detection and Direction Estimation with a "T" Shape Model, *Proceedings of the AVBPA 2005*, Springer, Hilton Rye Town, NY, USA, July 2005 pp. 201-207
- Maio, D. & Maltoni, D. (1996). A Structural Approach to Fingerprint Classification, *Proc. 13th ICPR*, Vienna, Aug. 1996
- Maio, D.; Maltoni, D. & Rizzi, S. (1996). Dynamic Clustering Of Maps In Autonomous Agents, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 18, No. 11, pp. 1,080-1,091, Nov. 1996
- Malathi, S. & Meena, C. (2010). An efficient method for partial fingerprint recognition based on Local Binary Pattern, *ICCCCT'2010*
- Militello, C.; Conti, V.; Sorbello, F. & Vitabile, S. (2008). A Novel Embedded Fingerprints Authentication System Based on Singularity Points, *International Conference on Complex, Intelligent and Software Intensive Systems*, 2008
- Min, M. M. & Thein, Y. (2009). Intelligent Fingerprint Recognition System by Using Geometry Approach, *Proceedings of the IEEE*, 2009
- Moayer, B. & Fu, K.S. (1975). A Syntactic Approach to Fingerprint Pattern Recognition, *Pattern Recognition*, vol. 7, pp. 1-23, 1975
- Moayer, B. & FU, K.-S. (1976). A Tree System Approach for Fingerprint Pattern Recognition, *IEEE Trans. On Computers*, Vol. C-25, NO. 3, pp.262-274, March 1976

- Montesanto, A.; Baldassarri, P.; Vallesi, G. & Tascini, G. (2007). Fingerprints Recognition Using Minutiae Extraction: a Fuzzy Approach, *14th International Conference on Image Analysis and Processing*, 2007
- Moscinska, K. & Tyma, G. (1993). Neural Network Based Fingerprint Classification, *Proc. Third Int'l Conf. Neural Network*, pp. 229-232, 1993
- On, C. K.; Pandiyan, P. M.; Yaacob, S. & Saudi, A. (2006). Fingerprint Feature Extraction Based Discrete Cosine Transformation (DCT), *Proceedings of the ICOCI*, 2006
- Parsons, N. R.; Smith, J. Q.; Thonnes, E.; Wang, L. & Wilson, R. G. (2008). Rotationally invariant statistics for examining the evidence from the pores in fingerprints, *Law, Probability and Risk*, vol. 7, no. 1, pp. 1-14, Mar. 2008
- Puertas, M.; Ramos, D.; Fierrez, J.; Ortega-Garcia, J. & Exposito, N. (2010). Towards a Better Understanding of the Performance of Latent Fingerprint Recognition in Realistic Forensic Conditions, *International Conference on Pattern Recognition*, 2010
- Qi, J.; Abdurrachim, D.; Li, D. & Kunieda, H. (2005). A hybrid method for fingerprint image quality calculation, *Proceedings of the 4th IEEE Workshop Automatic Identification Advanced Technologies*, 2005, pp. 124-129
- Qi, J.; Xie, M. & Wang, W. (2008). A Novel Fingerprint Matching Method Using a Curvature-Based Minutia Specifier, *Proceedings of the 15th International IEEE Conference on Image Processing*, 2008, pp.1488-1491
- Rao, K. & Balck, K. (1980). Type Classification of Fingerprints: A Syntactic Approach, *IEEE Trans. Anal. Mach. Intell*, vol. 2, no. 3, pp. 223-231, 1980
- Ratha, N.K.; Bolle, R.M.; Pandit, V.D. & Vaish, V. (2000). Robust Fingerprint Authentication Using Local Structural Similarity, *Proc. Fifth IEEE Workshop Applications of Computer Vision*, pp. 29-34, 2000
- Ratha, N. K.; Karu, K.; Chen, S. & Jain, A. K. (1996). A Real-Time Matching System for Large Fingerprint Databases, *IEEE Tran. On Pattern Analysis And Machine Intelligence*, VOL. 18, NO. 8, AUGUST 1996
- Senior, A. (1997). A Hidden Markov Model Fingerprint Classifier, *Proc. 31st Asilomar Conf. Signals, Systems, and Computers*, pp. 306-310, 1997
- Simoncelli, E. P. & Freeman, W. T. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation, *Proc. Int. Conf. Image Processing*, Washington, DC, Oct. 1995, vol. 3, pp. 23-26
- Sherlock, B. G.; Monro, D. M. & Millard, K. (1994). Fingerprint enhancement by directional fourier filtering, *Vision Image and Signal Processing*, Vol. 141, No. 2, pp. 87-94, 1994
- Specht, D.F. (1990). Probabilistic Neural Network, *Neural Networks*, vol. 3, no. 1, pp. 109-118, 1990
- Stock, R.M. & Swonger, C.W. (1969). *Development and Evaluation of a reader of Fingerprint Minutiae*, Cornell Aeronautical Laboratory, Technical Report CAL no. XM-2478-X-1:13-17, 1969
- Stosz, J. & Alyea, L. (1994). Automated system for fingerprint authentication using pores and ridge structure, *Proceedings of SPIE*, Vol. 2277, pp. 210-223, 1994
- Tachaphetpiboon, S. & Amornraksa, T. (2007). Fingerprint Features Extraction Using Curve-scanned DCT Coefficients, *Proceedings of Asia-Pacific Conference on Communications*, 2007
- Tachaphetpiboon, S. & Amornraksa, T. (2005). Fingerprint matching method using zigzag-scanned DCT coefficients, *Proc. of ITC-CSCC'05*, pp. 1171-1172, July 2005

- Thebaud, L.R. (1999). *Systems and Methods with Identity Verification by Comparison and Interpretation of Skin Patterns Such as Fingerprints*, US Patent No. 5,909,501, 1999
- The Science of Fingerprints: Classification and Uses*, Federal Bureau of Investigation. Washington, D.C.: US. Government Printing Office, 1984
- Tico, M.; Immonen, E.; Ramo, P.; Kuosmanen, P. & Saarinen, J. (2001). Fingerprint recognition using wavelet features, *Proc. of ISCS'01*, Vol. 2, pp. 21-24, May 2001
- Vaikol, S.; Sawarkar, S.D.; Hivrale, S. & Sharma, T. (2009). Minutiae Feature Extraction From Fingerprint Images, *IEEE Int'l Advance Comp. Conf.*, India, 6-7 March 2009
- Van derWeken, D.; Nachtegael, M. & Kerre, E. (2007). Combining neighborhood- based and histogram similarity measures for the design of image quality measures, *Image and Vision Computing*, Vol. 25, pp. 184-195, 2007
- Wahab, A.; Chin, S.H. & Tan, E. C. (1998). Novel approach to automated fingerprint recognition, *Prodeedings of the IEE*, 1998
- Wang, H.; Min, L.Q. & Liu, J. (2008). Robust designs for Fingerprint Feature Extraction CNN with Von Neumann Neighborhood, *International Conference on Computational Intelligence and Security*, 2008
- Watson, C. I.; Candela, G. T. & Grother, P. J. (1994). Comparison of FFT fingerprint filtering methods for neural network classification, *NISTIR*, vol. 5493, 1994
- Wei-bo, Z.; Xin-bao, N. & Chen-jian, W. (2008). A Fingerprint Matching Algorithm Based on Relative Topological Relationship among Minutiae, *Proceedings of the of International IEEE Conference Neural Networks & Signal Processing*, pp. 225-228, 2008
- Willis, A. & Myers, L. (2001). A cost-effective fingerprint recognition system for use with low-quality prints and damaged fingerprint, *Pattern Recognition*, Vol. 34, No. 2, pp. 255-270, Feb. 2001
- Xu, H. & Veldhuis, R. N.J. (2009). Spectral Minutiae Representations of Fingerprints Enhanced by Quality Data, *Proceedings of the IEEE*, 2009
- Xu, H. & Veldhuis, R. N.J. (2010). Complex Spectral Minutiae Representation For Fingerprint Recognition, *Proceedings of the IEEE*, 2010
- Xu, H.; Veldhuis, R. N. J.; Bazen, A. M.; Kevenaar, T. A. M.; Akkermans, T. A. H. M. & Gokberk, B. (2009). Fingerprint Verification Using Spectral Minutiae Representations, *IEEE Transactions On Information Forensics And Security*, Vol. 4, No. 3, pp. 397-409, SEP. 2009
- Xu, H.; Veldhuis, R. N. J.; Kevenaar, T. A. M. & Akkermans, T. A. H. M. (2009). A Fast Minutiae-Based Fingerprint Recognition System, *IEEE SYSTEMS JOURNAL*, Vol. 3, No. 4, pp. 418-427, December 2009
- Yahagi, H.; Igaki, S. & Yamagishi, F. (1990). *Moving-Window Algorithm For Fast Fingerprint Verification*, Fujitsu Laboratories LTD., Atsugi 243-01, Japan, 1990
- Yang, J.C.; Shin, J.W.; Min, B.J.; Lee, J.W.; Park, D.S. & Yoon, S. (2008). Fingerprint Matching using Global Minutiae and Invariant Moments, *Congress on Image and Signal Processing*, 2008
- Youmaran, R. & Adler, A. (2006). Measuring biometric sample quality in terms of biometric information, *Proceedings of the Biometrics Symp.*, Baltimore, MD, Sep. 2006.
- Zhang, P.; Li, C. & Hu, J. (2010). A Pitfall in Fingerprint Features Extraction, *11th Int. Conf. Control, Automation, Robotics and Vision*, Singapore, 7-10th December 2010

- Zhang, D.; Liu, F.; Zhao, Q.; Lu, G. & Luo, N. (2011). Selecting a Reference High Resolution for Fingerprint Recognition Using Minutiae and Pores, *IEEE Trans. On Instrumentation And Measurement*, Vol. 60, No. 3, pp. 863-871, MARCH 2011
- Zhao, Q. & Jain, A. K. (2010). On the Utility of Extended Fingerprint Features:A Study on Pores, *Proceedings of the IEEE*, 2010
- Zhao, Q.; Zhang, D.; Zhang, L. & Luo, N. (2009). High resolution partial fingerprint alignment using pore-valley descriptors, *Pattern Recognition*, vol. 43, no. 3, pp. 1050-1061, Aug. 2009
- Zhao, Q.; Zhang, L.; Zhang, D.; Luo, N. & Bao, J. (2008). Adaptive pore model for fingerprint pore extraction, *Proc. 18th Int. Conf. Pattern Recog.*, 2008, pp. 1-4

Non-minutiae Based Fingerprint Descriptor

Jucheng Yang

*School of information technology, Jiangxi University of Finance and Economics
Ahead Software Company Limited, Nanchang
China*

1. Introduction

Fingerprint recognition refers to the techniques of identifying or verifying a match between human fingerprints. Fingerprint recognition has been one of the hot research areas in recent years, and it plays an important role in personal identification (Maio et al., 2003). A general fingerprint recognition system consists of some important steps, such as fingerprint pre-processing, feature extraction, matching, and so on. Usually, a descriptor is defined to identify an item with information storage. A fingerprint descriptor is used to describe and represent a fingerprint image for personal identification.

Various fingerprint descriptors have been proposed in the literature. Two main categories for fingerprint descriptors can be classified into minutiae based and non-minutiae based. Minutiae based descriptors (Jain et al. 1997a; Jain et al. 1997b; Liu et al. 2000; Ratha et al. 1996; He et al. 2007; Cappelli et al. 2011) are the most popular algorithms for fingerprint recognition and are sophisticatedly used in fingerprint recognition systems. The major minutia features of fingerprint ridges are: ridge ending, ridge bifurcation and so on (Maio et al., 2003). Minutiae based descriptor use a feature vector extracted from fingerprints as sets of points in a multi-dimensional space, which comprise several characteristics of minutiae such as type, position, orientation, etc. The matching is to essentially search for the best alignment between the template and the input minutiae sets. However, due to the poor image quality and complex input conditions, minutiae are not easy to be accurately determined, thus it may result in low matching accuracy. In addition, minutiae based descriptors may not fully utilize the rich discriminatory information available in the fingerprints with high computational complexity.

Non-minutiae based descriptors (Amornraksa & Tachaphetpiboon, 2006; Benhammedi, et al. 2007; Jain et al., 2000; Jin et al., 2004; Nanni and Lumini, 2008; Nanni & Lumini, 2009; Ross, et al. 2003; Sha et al. 2003; Tico et al. 2001; Yang et al., 2006; Yang & Park, 2008a; Yang & Park, 2008b), however, overcome the demerits of the minutiae based method. It uses features other than characteristics of minutiae from the fingerprint ridge pattern, such as local orientation and frequency, ridge shape, and texture information. It can extract more rich discriminatory information and abandon the pre-processing process such as binarization and thinning and post minutiae processing. Other merits are listed by using the non-minutiae based methods, such as a high accuracy; a fast processing speed; a fixed length feature vector; easily coupled with other system; being combined with Biohashing and so on.

Among various non-minutiae based descriptors, Gabor feature-based ones (Jain et al., 2000; Sha et al. 2003) present a relatively high matching accuracy by using a bank of Gabor filters to capture both the local and global details in a fingerprint, and represent them as a compact fixed-length fingerCode. Ross et al. (2003) describes a hybrid fingerprint descriptor that uses both minutiae and a ridge feature map constructed by a set of eight Gabor filters. Benhammedi et al. (2007) also propose a new hybrid fingerprint descriptor based on minutiae texture maps according to their orientations. It exploits the eight fixed directions of Gabor filters to generate its weighting oriented Minutiae Codes. Tico et al. (2001) propose a transform-based descriptor using digital wavelet transform (DWT) features. The features are obtained from the standard deviations of the DWT coefficients of the image details at different scales and orientations. Amornraksa & Tachaphetpiboon (2006) propose another transform-based descriptor using digital cosine transform (DCT) features. The transform methods show a high matching accuracy for inputs identical to one in its database. Jin et al. (2004) propose an improved transform-based descriptor based on the features extracted from the integrated wavelet and Fourier-Mellin transform (WFMT) framework. Multiple WFMT features can be used to form a reference invariant feature through the linearity property of FMT and hence to reduce the variability of the input fingerprint images. Nanni & Lumini (2008) proposed a hybrid fingerprint descriptor based on local binary pattern (LBP). Nanni & Lumini (2009) proposed another descriptor based on histogram of oriented gradients (HoG). Yang et al. (2006) propose a fingerprint descriptor using invariant moments (IM) with the learning vector quantization neural network (LVQNN) for matching, which use a fixed-size ROI (Region of Interest) to extract seven invariant moments as a feature vector. Its improved ones (Yang & Park, 2008a; Yang & Park, 2008b;) using tessellated invariant moments (Tessellated IM) or sub-regions IM with eigenvalue-weighted cosine (EWC) distance or nonlinear back-propagation Neural Networks (BPNN) to handle the various input conditions for fingerprint recognition.

In this chapter, some state of the art non-minutiae based descriptors are first reviewed, and a non-minutiae based fingerprint descriptor with tessellated invariant moment features, feature selection with PCA (Principle Component Analysis), and a Support Vector Machine (SVM) for classification is proposed. The proposed descriptor basically uses moment features invariant to scale, position and rotation to increase the matching accuracy with a low computational load. It further pursues an improved performance by using the alignment and rotation after a sophisticated, reliable detection of a reference point. Having invariant characteristics in the proposed algorithm can significantly improve the performance for input images under various conditions. To fully utilize both the global and local ridge information while removing unwanted noises, the algorithm extracts features from tessellated cells around the reference point. Combining with the PCA for feature selection to reduce the dimension of feature vector and choose the distinct features. Matching with a SVM also contributes to the performance enhancement by simply assigning weights on different cells and classification with high accuracy.

The chapter is organized as follows: some state of the art non-minutiae based descriptors are briefly reviewed in section 2. In section 3, a proposed non-minutiae based fingerprint descriptor with tessellated invariant moments and SVM is explained. And experimental results are illuminated in section 4. Finally, conclusion remarks are given in section 5.

2. Non-minutiae based descriptors

It is important to establish descriptors to extract reliable, independent and discriminate fingerprint image features. Exception of the widely used minutiae descriptors, the non-minutiae based descriptors use features other than characteristics of minutiae from the fingerprint ridge pattern are able to achieve the characters of the mentioned fine traits. The features of these descriptors may be extracted more reliably than those of minutiae. The next sub-sections will introduce some classical and state of the art non-minutiae based descriptors, such as Gabor filters, DWT, DCT, WFMT, LBP, HOG, IM based.

2.1 Gabor filters based descriptor

The Gabor filters based descriptors (Jain et al.,2000; Sha et al. 2003) have been proved with their effectiveness to capture the local ridge characteristics with both frequency-selective and orientation-selective properties in both spatial and frequency domains. They describe a new texture descriptor scheme called fingerCode which is to utilize both global and local ridge descriptions to represent a fingerprint image. The features are extracted by tessellating the image around a reference point (the core point) determined in advance. The feature vector consists of an ordered collection of texture descriptors from some tessellated cells. Since the scheme assumes that the fingerprint is vertically oriented, to achieve invariance, image rotation is compensated by computing the features at various orientations. The texture descriptors are obtained by filtering each sector with 8 oriented Gabor filters and then computing the AAD (Average Absolute Deviation) of the pixel values in each cell. The features are concatenated to obtain the fingerCode. Fingerprint matching is based on finding the Euclidean distance between the two corresponding FingerCodes.

However, the Gabor filters based descriptors are not rotation invariant. To achieve approximate rotation invariance, each fingerprint has to be represented with ten associated templates stored in the database, and the template with the minimum score is considered as the rotated version of the input fingerprint image. So these methods require a larger storage space and a significantly high processing time.

Recently, some hybrid descriptors combined with Gabor filters are proposed. Ross et al. (2003) describes a hybrid fingerprint descriptor that uses both minutiae and a ridge feature map constructed by a set of eight Gabor filters. The ridge feature map along with the minutiae set of a fingerprint image is used for matching purposes. The hybrid matcher is proved to perform better than a minutiae-based fingerprint matching system by the author.

Benhammedi et al. (2007) also propose a hybrid fingerprint descriptor based on minutiae texture maps according to their orientations. Rather than exploiting the eight fixed directions of Gabor filters for all original fingerprint images filtering process, they construct absolute images starting from the minutiae localizations and orientations to generate the Weighting Oriented Minutiae Codes. The extracted features are invariant to translation and rotation, which avoids the fingerprint pair relative alignment stage.

Another Gabor filters based descriptor is proposed by Nanni & Lumini (2007), where the minutiae are used to align the images and a multi-resolution analysis performed on separate regions or sub-windows of the fingerprint pattern is adopted for feature extraction and classification. The features extracted are the standard deviation of the image convolved with 16 Gabor filters. The similarity measurement is done by the weighed Euclidean distance matchers with a sequential forward floating scheme.

However, the matching accuracy of these hybrid approaches may be degraded for low-quality inputs, since the performance highly depends on extracting all minutiae precisely and reliably.

2.2 DWT based descriptor

Tico et al. (2001) proposed a method using DWT features. The features are obtained from the standard deviations of the DWT coefficients of the entire image details at different scales and orientations to form a feature vector of length 12 (48 in total from four sub-images). The normalized l_2 -norm of each wavelet sub-image is computed in order to create a feature vector. The feature vector represents an approximation of the image energy distribution. Fingerprint matching is based on k-Nearest Neighbor (KNN) with finding the Euclidean distance between the corresponding feature vectors.

However, the features extraction from frequency domain with corresponding transforms are not rotation-invariant, so if the image input with rotation, the features from the same fingerprint image can be judged into the different ones.

2.3 DCT based descriptor

Amornraksa & Tachaphetpiboon (2006) also propose a method using digital cosine transform (DCT) features for fingerprint matching. The standard deviations of the DCT coefficients located in six predefined areas from a 64×64 ROI are used as a feature vector of length 6 (24 in total from four sub-images) for fingerprint matching.

To extract the informative features from a fingerprint image, the image is first cropped to a 64×64 pixel region, centred at the reference point, and then quartered to obtain four non-overlapping sub-images of size 32×32 pixels. Next, the DCT is applied to each sub-image to obtain a block of 32×32 DCT coefficients. Finally, the standard deviations of the DCT coefficients located in six predefined areas are calculated and used as a feature vector of length 6 (24 in total from four sub-images) for fingerprint matching. Fingerprint matching is also based on KNN with the Euclidean distance.

However, the features are not rotation-invariant, too, to achieve rotation-invariant, so an improved method needs.

2.4 Integrated wavelet and the Fourier–Mellin transform (WFMT) based descriptor

Jin et al. (2004) propose an improved transform-based descriptor extracted features from the integrated wavelet and Fourier-Mellin transform (WFMT) framework. Wavelet transform, with its energy compacted feature is used to preserve the local edges and reduce noise in the low frequency domain, is able to make the fingerprint images less sensitive to shape distortion. The Fourier–Mellin transform (FMT) serves to produce a translation, rotation and scale invariant feature. And multiple WFMT features can be used to form a reference invariant feature through the linearity property of FMT and hence reduce the variability of the input fingerprint images. Multiple m WFMT features can be used to form a reference WFMT feature and just only one representation per user needs to be stored in the database. Fingerprint matching is based on finding the Euclidean distance between the corresponding multiple WFMT feature vectors.

However, the main disadvantage of this descriptor is that the reference point is based on a non-precise core point, and the descriptor requires a high time-consuming process if using the multiple WFMT features with training images.

2.5 Local binary pattern (LBP) based descriptor

Nanni & Lumini (2008) proposed a fingerprint descriptor based on LBPs. A LBP proposed by Ojala et al. (2002) is a grayscale local texture operator with powerful discrimination and low computational complexity. Moreover, it is invariant to monotonic grayscale transformation, hence the LBP representation may be less sensitive to changes in illumination. The two fingerprints to be matched are first aligned using their minutiae, then the images are decomposed in several overlapping sub-windows; and a Gabor filter with LBP hybrid method (GLBP) also proposed by Nanni & Lumini (2008), instead of from the original sub-windows, each sub-window is convolved with a bank of Gabor filters and the invariant LBPs histograms are extracted from the convolved images. The matching value between two fingerprint images is performed by a complex distance function that takes into account the presence of different types of descriptors and different regions.

However, the matching accuracy depends on extracting all minutiae precisely and reliably, and it may be degraded for low-quality inputs.

2.6 Histogram of oriented gradients (HoG) based descriptor

Nanni & Lumini (2009) recently proposed a hybrid fingerprint descriptor based on HoG. HoG has been first proposed by Dalal & Triggs (2005) as an image descriptor for localizing pedestrians in complex images and has reached increasingly popularity. The aim of this descriptor is to represent an image by a set of local histograms which count occurrences of gradient orientation in a local cell of the image. The implementation of the HoG descriptors can be achieved by computing gradients of the image; dividing the image into small sub-regions; building a histogram of gradient directions and normalizing histograms within some groups of sub-regions for each sub-region to achieve a better invariance to changes in illumination or shadowing. The matching value between two fingerprint images is performed by a complex distance function, too.

However, similar with the above minutiae based alignment methods, the matching accuracy depends on extracting all minutiae precisely and reliably.

2.7 IM based descriptor

Except for the above non-minutiae based descriptor, recently, Yang et al. (2006) propose an IM based descriptor with the LVQNN for fingerprint matching, which use a fixed-size ROI to extract seven invariant moments as a feature vector. And its improved ones (Yang & Park, 2008a; Yang & Park, 2008b;) using tessellated IM or sub-regions IM with EWC distance or nonlinear BPNN to handle the various input conditions for fingerprint recognition. The details of the invariant moments are introduced as below:

2.7.1 Raw moments

For a 2-dimensional continuous function $f(x,y)$ the moment (sometimes called 'raw moment') of order $(p + q)$ is defined as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy \quad (1)$$

for $p, q = 0, 1, 2, \dots$

Adapting this to scalar (grey-tone) image with pixel intensities $I(x,y)$, raw image moments are calculated by

$$m_{pq} = \sum_x \sum_y x^p y^q I(x,y) \quad (2)$$

In some cases, this may be calculated by considering the image as a probability density function, i.e., by dividing the above by

$$\sum_x \sum_y I(x,y) \quad (3)$$

A uniqueness theorem (Papoulis, 1991) states that if $f(x,y)$ is piecewise continuous and has nonzero values only in a finite part of the xy plane, moments of all orders exist, and the moment sequence (M_{pq}) is uniquely determined by $f(x,y)$. Conversely, (M_{pq}) uniquely determines $f(x,y)$. In practice, the image is summarized with functions of a few lower order moments.

Simple image properties derived via raw moments include:

1. Area (for binary images) or sum of grey level (for grey-tone images):

The Area (A) of the image can be extracted by the formula:

$$A = m_{00} \quad (4)$$

2. Centroid point of the image:

The centroid point $\{\bar{x}, \bar{y}\}$ of the image can be extracted by the formula:

$$\{\bar{x}, \bar{y}\} = \{m_{10}/m_{00}, m_{01}/m_{00}\} \quad (5)$$

2.7.2 Central moments

- a. The central moments are defined as

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) dx dy \quad (6)$$

Where

$$\bar{x} = m_{10}/m_{00} \quad \text{and} \quad \bar{y} = m_{01}/m_{00} \quad (7)$$

If $f(x,y)$ is a digital image, then Eq.(6) becomes

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y) \quad (8)$$

- b. The meanings of the central moments are listed as below:

1. μ_{20} represents the horizontal extension parameter of the image;
2. μ_{02} represents vertical extension of the image;
3. μ_{11} represents the gradient of the image, if $\mu_{11} > 0$, it means the image inclines towards up-left; if $\mu_{11} < 0$, it means the image inclines towards up-right;

4. μ_{30} represents the excursion degree of the image's barycenter on horizontal direction, when $\mu_{30} > 0$, the barycenter inclines towards left; then if $\mu_{30} < 0$, the barycenter inclines towards right;
 5. μ_{03} represents the vertical excursion degree of the image's barycenter, if $\mu_{03} > 0$, the barycenter inclines towards upward, and inclines towards downward while $\mu_{03} < 0$;
 6. μ_{21} represents the equilibrium degree about the image on the vertical direction, if $\mu_{21} > 0$, the extension of the downside image is greater than the upside; then smaller when $\mu_{21} < 0$;
 7. μ_{12} represents the equilibrium degree of the image about the vertical direction, when $\mu_{12} > 0$, the image's extension on the right side is greater than the left side, then smaller when $\mu_{12} < 0$.
- c. Normalized moments:

Moments where $p+q \geq 2$ can also be invariant to both translation and changes in scale by dividing central moments by the properly scaled μ_{00} moment, as the normalized central moments, denoted η_{pq} , are defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (9)$$

where $\gamma = \frac{p+q}{2} + 1$ for $p+q = 2, 3, \dots$

2.7.3 Invariant moments

As a region description method, invariant moments are used for texture analysis and pattern identification. A set of seven invariant moments derived from the second and the third moments were proposed by Hu (1962). As the equation shown below, Hu derived the expressions from algebraic invariants applied to the moment generating function under a rotation transformation. The set of moment invariants consist of groups of nonlinear centralized moment expressions. And they are scale, position, and rotation invariant (Gonzalez and woods, 2002).

$$\begin{aligned}
 \phi_1 &= \eta_{20} + \eta_{02} \\
 \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
 \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - 3\eta_{03})^2 \\
 \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
 \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 \phi_6 &= (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
 \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
 \end{aligned} \quad (10)$$

The values of the computed moment invariants are usually small values of higher order moment invariants are close to zero in some cases. So we reset the value range using the logarithmic function, the outputs of the moment invariants are mapped into $\Phi_i = \lceil \lg(|\phi_i|) \rceil$ $i=1,2,\dots,7$, which can take the values to the large dynamic range with a nonlinear scale.

Figure 1 and Figure 2 show the invariant moments analysis on the four sub-images with 2D and 3D views, respectively. The experiments here are designed to check the characteristics of the invariant moment features' invariance to rotation and scale. We choose four sub-images from a ROI of a fingerprint image, and divide the ROI into four sub-images. From the figures, we can see that the ridge valleys of the sub-images are totally different.

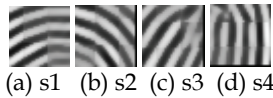


Fig. 1. Sub-images of sub-images (a) s1 (b) s2 (c) s3 (d) s4

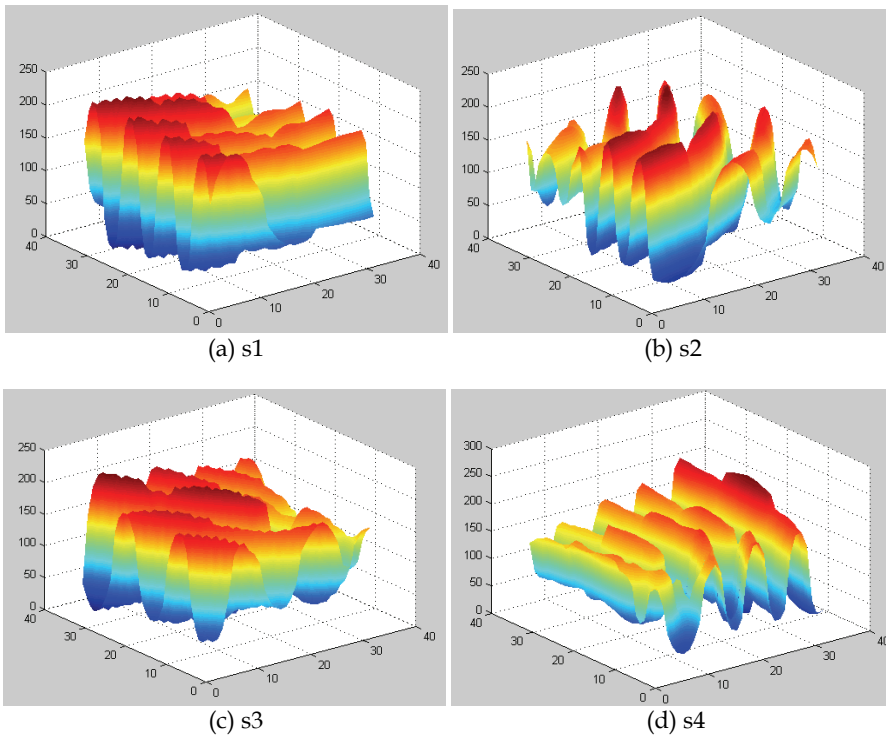


Fig. 2. 3D images of sub-images (a) s1 (b) s2(c) s3 (d) s4

And Table 1 and Table 2 show seven invariant moments of these four sub-images, and scale, rotation invariance of sub-image S1, respectively. As we can see from the tables, the invariant moments are nearly invariant to the scale and rotation invariance.

Sub-image	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7
S1	6.660157	22.792705	27.607114	29.742121	59.753254	41.222497	58.474584
S2	6.664572	20.127298	29.650909	29.742485	59.945926	39.811898	59.811871
S3	6.676695	21.271717	28.456694	29.916016	60.753962	41.309514	59.528609
S4	6.674009	20.209754	28.826501	29.101012	58.682184	39.417999	58.701581

Table 1. Seven invariant moments of the four sub-images

Scale	Rotation	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7
0.8	0	6.659508	21.450452	27.459971	30.169709	59.518255	41.693181	59.301135
	90	6.657605	21.641931	27.433996	31.337310	63.211065	43.918130	61.111226
	180	6.657625	21.654452	27.376333	31.318065	61.275622	42.931670	60.714937
	270	6.659334	21.473568	27.400168	30.162576	60.756198	41.433472	59.053503
1	0	6.660157	22.792705	27.607114	29.742121	59.753254	41.222497	58.474584
	90	6.660157	22.792705	27.607114	29.742121	58.901367	41.222497	58.474584
	180	6.660157	22.792705	27.607114	29.742121	59.753254	41.222497	58.474584
	270	6.660157	22.792705	27.607114	29.742121	58.901367	41.222497	58.474584
2	0	6.659431	22.792705	27.607114	29.742121	59.753254	41.222497	58.474584
	90	6.659431	22.792705	27.607114	29.742121	58.901367	41.222497	58.474584
	180	6.659431	22.792705	27.607114	29.742121	59.753254	41.222497	58.474584
	270	6.659431	22.792705	27.607114	29.742121	58.901367	41.222497	58.474584

Table 2. Scale, rotation invariance of sub-image S1

2.8 Tessellated invariant moments based descriptors

In order to reduce the effects of noise and non-linear distortions, and speed up the processing time, tessellated IM based descriptors (Yang & Park, 2008a) using only a certain area (ROI) around the reference point at the centre as the feature extraction area instead of using the entire fingerprint is proposed. By adjusting the size of ROI and the number of the cells, we can capture both the local and global structure information around the reference point (see Figure 3 (a-b)). The ROI is partitioned into some non-overlapping rectangular cells as depicted in Figure 3 (a), e.g. the size of ROI is 64×64 pixels, 16 rectangular cells, and each cell has a size of 16×16 pixels.

Invariant moment analysis introduced in section 2 is used as the features for the fingerprint recognition system. For each cell, a set of seven invariant moments are computed. Suppose $\phi = \{\phi_n\}_{n=1}^7$ is a set of invariant moments, and $s = \{s_n\}_{n=1}^N$ is the collection of all the cells,

$$s = \{s_n\}_{n=1}^N = \{\{\phi_{nk}\}_{k=1}^7\}_{n=1}^M \quad (11)$$

where M is the number of the cells, N=7M is the total length of the collection.

Furthermore, in order to improve the matching accuracy, weights w is assigned to each tessellated cell to distinguish the cell from the foreground or background, which can be used to resolve the embarrassment when the reference point is nearby the border of the

fingerprint image. If the tessellated cells that contain a certain proportion of the background pixels, it will be labelled as background cells and the corresponding w is set to 0, else w equals 1. Therefore, a fingerprint can be represented by a fixed-length feature vector as

$$f = \{f_n\}_{n=1}^N = \{ws_n\}_{n=1}^N \quad (12)$$

The length of the total feature vectors is N .

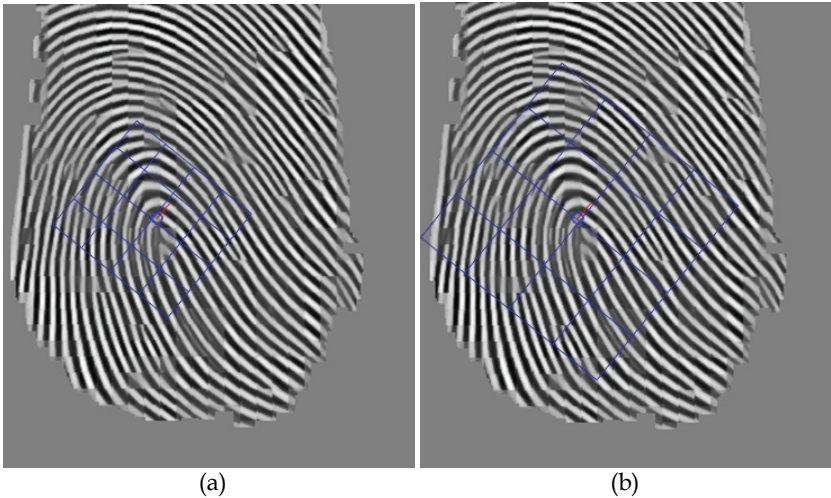


Fig. 3. (a) Tessellated cells (local) or (b) tessellated cells (global) (Yang & Park, 2008a)

2.9 Summarization

As the above analysis, all kinds of non-minutiae descriptor has proposed to hand the shortcomings of minutiae methods in the fingerprint recognition systems. Table 3 gives a summarization of some classical and state-of-art non-minutiae based descriptors in these recent years, such as Gabor filters, DWT, DCT, WFMT, LBP, HOG, IM based, and analyzes their feature extraction, alignment, and matching methods.

Table 3 summarizes the classical and state of the art non-minutiae based method (EER is the equal error rate, FAR is the false acceptance rate, FRR is the false reject rate, and GAR is the genuine acceptance rate (GAR = 1-FRR)). From the table, we can see that based on the public database of FVC2002 DB2, the sub-regions IM method with BPNN has the lowest EER of 3.26%, which means its performance is the best among all the non-minutiae descriptors on the same public database listed in the table.

3. Proposed non-minutiae based descriptor

Although the IM descriptor proposed in Yang & Park (2008a) and Yang & Park (2008b) use the tessellated or sub-regions IM to extract both the local and global features, comparing to Yang et al. (2006) just using only a fixed ROI, the tessellated or sub-regions ROIs are delineated from the same area of a fingerprint image, strong correlation may exist among the extracted features, and the dimension of the features are high. An improvement way is

proposed here to reduce the dimension of feature vector and choose the distinct features before matching.

PCA is one of the oldest and best known techniques in multivariate analysis (Fausett, 1994). So we reduce the dimension of feature vector by examining feature covariance matrix and then selecting the most distinct features and using them to improve the verification performance. Besides, as a powerful classification tool, a SVM is proposed for fingerprint matching.

Approaches	Features	Alignment methods	Matching methods	Performance analysis	
				Databases	Results
Jain et al.(2000)	Gabor filters	Core point	Euclidean distance	Self	FAR=4.5 FRR=2.8
Ross et al. (2003)	Gabor filters& Minutiae	Minutiae	Euclidean distance	Self	EER=4
Benhammadi et al. (2007)	Minutiae Gabor filters maps	Minutiae	Normalized distance	FVC2002 DB2	EER=5.19
Nanni & Lumini (2007)	Gabor filters	Minutiae	Weighed Euclidean distance	FVC2002 DB 2	EER=3.6
Tico et al. (2001)	DWT	NO	KNN with Euclidean distance	Self &small	Recognition rates 95.2
Amornraksa & Tachaphetpiboon (2006)	DCT	NO	KNN with Euclidean distance	Self &small	Recognition rates 99.23
Jin et al. (2004)	WFMT	Core point	Euclidean distance	FVC2002 DB 2	EER=5.309
Nanni & Lumini (2008)	LBP	Minutiae	Complex distance function	FVC2002 Db2	EER=6.2
Nanni & Lumini (2009)	HoG	Minutiae	Complex distance function	FVC2002 DB 2	EER=3.8
Yang et al. (2006)	Seven IM	Reference point	LVQNN	FVC2002 DB 2	FAR=0.6G AR=96.1
Yang & Park (2008a)	Tessellated IM	Reference point	EWC distance	FVC2002 DB 2	EER=3.78
Yang & Park (2008b)	Sub-regions IM	Reference point	BPNN	FVC2002 DB 2	EER=3.26

Table 3. Summarization of the classical and state of the art non-minutiae based method

3.1 Feature vector construction

After the pre-processing steps of fingerprint enhancement, reference point determination and image alignment, ROI determination (Yang & Park, 2008a; Yang & Park, 2008b; Yang et

al, 2008c), due to the good characteristics of reducing the effects of noise and non-linear distortions, we use the tessellated IM as the features, and a fingerprint can be represented by a fixed-length feature vector as $f = \{f_n\}_{n=1}^N = \{ws_n\}_{n=1}^N = \{w\{\phi_{nk}\}_{k=1}^7\}_{n=1}^N$, as described in section 2.8, where $\phi = \{\phi_i\}_{i=1}^7$ is a set of invariant moments, and the length of the total feature vectors is N , w is the weight of distinguishing the cell from the foreground or background. For example, the ROI is partitioned into some non-overlapping rectangular cells with the size of ROI is 64×64 pixels, 16 rectangular cells, and each cell has a size of 16×16 pixels. Then the feature-vector with the elements consists of a sets of moments derived from tessellated ROIs. So the total length of the feature vector is $16 \times 7 = 112$.

3.2 Feature selection with PCA

Here, the feature selection with PCA is briefly introduced. The objective of PCA is to reduce the dimensionality of the data set, while retaining as much as possible variation in the data set, and to identify new meaningful underlying variables.

The basic idea in PCA is to find the orthonormal features. Let $x \in R^n$ be a random vector, where n is the dimension of the input space. The first principal component is the projection in the direction in which the variance of the projection is maximized. The m th principal component is determined as the principal component of the residual based on the covariance matrix.

The covariance matrix of x is defined as $\Xi = E\{[x-E(x)][x-E(x)]^T\}$. Let u_1, u_2, \dots, u_n and $\lambda_1, \lambda_2, \dots, \lambda_n$ be eigenvectors and eigenvalues of Ξ , respectively, and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

Then, PCA factorizes Ξ into $\Xi = U\Lambda U^T$, with $U = [u_1, u_2, \dots, u_n]$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. We need to note that once the PCA of Ξ is available, the best rank- m approximation of Ξ can be readily computed.

Let $P = [u_1, u_2, \dots, u_m]$, where $m < n$. Then $y = P^T x$ will be an important application of PCA in dimensionality reduction.

3.3 Matching with SVM

SVMs (John & Nello, 2000) are a set of related supervised learning methods used for classification and regression. Viewing input data as two sets of vectors in an n -dimensional space, a SVM will construct a separating hyperplane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are 'pushed up against' the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the better the generalization error of the classifier.

SVM is a powerful tool for solving small sample learning problem that offers favorable performance using linear or nonlinear function estimators. It is a type of neural network that automatically determines the structural components. In our study, we used a two-class SVM to classify the input fingerprints into the corresponding class. Let the training set be $\{(x_i, y_i)\}_{i=1}^m$, with input vector x_i (n components), $y_i = \pm 1$ indicates two different classes and $i = 1, 2, \dots, m$. The decision function of SVM is:

$$y = \text{sign}(\sum_{i=1}^m \alpha_i y_i K(x_i, x) + b) = \text{sign}(\sum_{i=1}^m \omega_i K(x_i, x) + b) \quad (13)$$

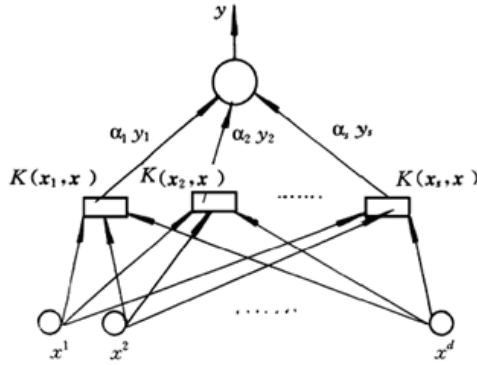


Fig. 4. Illustration of a SVM structure

Figure 4 illustrates the structure of a SVM, where $K(x_i, \mathbf{x})$ is the output of the i th hidden node with respect to the input vector \mathbf{x} , it is a mapping of the input \mathbf{x} and the support vector \mathbf{x}_i in an alternative space (the so-called feature space), by choosing the kernel $K(x_i, \mathbf{x}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})$. α_i and b are the learning parameters of the hidden nodes, and $\omega_i = \alpha_i y_i$ is the weight of the i th hidden node connecting with the output node. The learning task of a SVM is to find the optimal α_i by solving the maximization of the Lagrangian

$$W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (14)$$

subject to the constraints

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq c \quad (15)$$

The common kernel functions $K(\mathbf{x}_i, \mathbf{x})$ are defined as :

$$K(\mathbf{x}_i, \mathbf{x}) = \begin{cases} \mathbf{x}_i^T \cdot \mathbf{x} & \text{Linear} \\ (\gamma \cdot \mathbf{x}_i^T \cdot \mathbf{x} + c)^{\text{Degree}} & \text{Polynomial} \\ \exp(-\gamma \cdot \|\mathbf{x}_i - \mathbf{x}\|^2) & \text{Radial - basis} \\ \tanh(\gamma \cdot \mathbf{x}_i^T \cdot \mathbf{x} + c) & \text{Sigmoid} \end{cases} \quad (16)$$

Where γ, c, Degree are the kernel parameters.

4. Experimental results

The proposed algorithm was evaluated on the fingerprint images taken from the public FVC2002 database (Maio et al. 2002), which contains 4 distinctive databases: DB1_a, DB2_a, DB3_a and DB4_a. The resolution of DB1_a, DB3_a, and DB4_a is 500 dpi, and that of DB2_a is 569 dpi. Each database consists of 800 fingerprint images in 256 gray scale levels (100 persons, 8 fingerprints per person). Fingerprints from 101 to 110 (set B) have been made available to the participants to allow parameter tuning before the submission of the

algorithms; the benchmark is then constituted by fingers numbered from 1 to 100 (set A). In our experiments, the used FVC2002 databases are set A.

4.1 Performances with a SVM matching

In the experiments, a SVM is used to verify a matching between feature vectors of input fingerprint and those of template fingerprint, the number of the output class is the same with the verifying persons. For each input fingerprint and its template fingerprint, we compute the IM features. Since the output is to judge whether the input fingerprint is match or non-match according to the identity ID, we can take the matching process as a two-class problem.

In the training stage, training samples after normalization and scale processing (Hao, et al. 2007) are fed to a SVM with indicating their corresponding class. The features are computed from the training data, each contains vector from the training fingerprint, and the identity ID of the corresponding class is used to guide the classifying results through the SVM. While in the testing stage, test samples with the same normalization and scale processing are fed to the SVM to produce the output values. Similarly, the features are computed from the testing data, each contains vector from the test fingerprint with the corresponding identity ID. The element of the output values is restricted in the class number. If the output number is equal the corresponding ID, then it means match, vice visa.

To evaluate the performance of the verification rate of the proposed method, the receiver operating characteristic (ROC) curve is used. An ROC curve is a plot of false reject rate (FRR) against false acceptance rate (FAR). The FRR and FAR are defined as follows:

$$FRR = \frac{\text{Number of rejected genuine claims}}{\text{Total number of genuine accesses}} \times 100\% \quad (17)$$

$$FAR = \frac{\text{Number of accepted imposter claims}}{\text{Total number of imposter accesses}} \times 100\% \quad (18)$$

The equal error rate (EER) is also used as a performance indicator. The EER indicates the point where the FRR and FAR are equal, as below.

$$EER = (FAR + FRR)/2, \text{ if } FAR = FRR \quad (19)$$

For evaluating the recognition rate performance of the proposed descriptor, we did experiments on FVC2002 databases DB2_a. We divided the database into a training set and a testing set. Six out of eight fingerprints from each person were chosen for training and all the eight fingerprints for testing. For a database, therefore, 600 patterns were used for training, and 800 for testing. In the experiments, the 64×64 pixels size of ROI was adopted, and the ROI was tessellated into 16 rectangular cells with each cell had a size of 16×16 pixels. So the length of the feature vector was 16×7=112. In the feature selection experiments, the feature-vector with the elements consisting of a sets of moments derived from tessellated ROIs was processed with PCA. The new dimensional features are uncorrelated from original features due to the PCA analysis. Figure 5 describes the eigenvalues spectrum with different elements by using PCA for feature selection. Since our feature vector contains 112 elements, from the figure, we can see that almost all the eigenvalues spectrum with the eigenvectors elements are below than 70, if we keep 95% energy, then eigenvectors element of 50 was chosen.

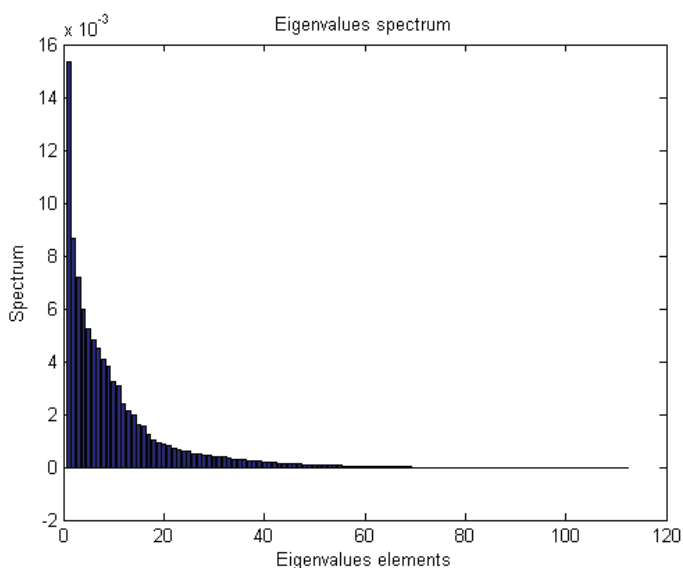


Fig. 5. The eigenvalues spectrum with different elements by using PCA for feature selection

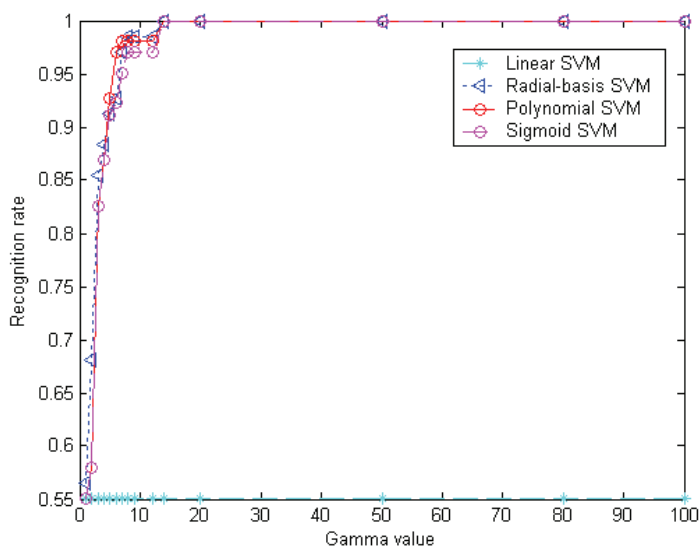


Fig. 6. The curves of the recognition rate of the proposed descriptor by different types of SVM with different γ value on the database FVC2002 DB2_a

As we know from the section 3.3, the linear SVM has no parameters, the Radial-basis SVM has only one parameter γ , the Polynomial SVM has three parameters of c, γ , Degree and

the Sigmoid SVM has two parameters of c, γ . For evaluating the recognition rate performance of the proposed tessellated IM based descriptor with different types of SVM, Figure 6 describes the curves of the recognition rate of the proposed descriptor by different types of SVM with different γ value on the database FVC2002 DB2_a. In our experiments, the parameters of c and Degree are fixed and determined by experiments, and the eigenvalues element is 50. From the figure, we can see that the recognition rates of the proposed descriptor of all nonlinear types of SVM are growing up with the γ value, and the proposed descriptor can achieve high recognition rates with the nonlinear types of SVM.

Figure 7 shows the curves of the recognition rate of the proposed descriptor by different γ value of the radial-basis SVM with different PCA eigenvalues elements on the database FVC2002 DB2_a. From the figure, we can see that the recognition rate is growing up with the γ value of the Radial-basis SVM. And the descriptor with different eigenvalue elements PCA has different recognition rate, with small eigenvalue (such as 10) the recognition rate is lower than the PCA with large eigenvalue (such as 20) under the same γ value, and the best recognition rate achieves by the eigenvalue element equals 50. However, all the recognition rate of the descriptor with PCA may have better performances comparing with the descriptor without PCA.

4.2 Comparing with other descriptors

The performances of several descriptors aimed at evaluating the usefulness of our descriptor were compared; all the descriptors were based on the two-stage enhanced image:

1. Gabor filters based, a descriptor using both minutiae and a ridge feature map with eight directions for fingerprint matching according to the method of Ross et al. (2003), and the same parameters as their paper are used;
2. DCT based, a descriptor using DCT features for fingerprint matching according to the method of Amornraksa & Tachaphetpiboon (2006), and the same parameters as their paper are used;
3. WFMT based, a descriptor using WFMT features for fingerprint matching according to the method of Jin et al. (2004), and the same parameters as their paper are used;
4. LBP based, a descriptor using the invariant local binary patterns features ($P = 8; R=1; n=10$) according to the method of Nanni & Lumini (2008).
5. Tessellated IM based, the 64×64 pixels ROI was tessellated into 16 rectangular cells with each cell had a size of 16×16 pixels and matching with SVM.

In our experiment, to compute the FAR and the FRR, the genuine match and impostor match were performed on the four sub-databases of FVC2002 database. We divided each database into a training set and a testing set using a 25% jackknife method. Six out of eight fingerprints from each person were chosen for training and the remaining two for testing. For a database, therefore, 600 patterns are used for training, and 200 for testing. For genuine match, each fingerprint of each person is compared with other fingerprints of the same person. And for impostor match, each test fingerprint is compared with the fingerprints belonged to other persons. Since there are 200 test patterns for an experiment, the number of matches for genuine and impostor are $6 \times 200 = 1200$ and $99 \times 200 = 19800$ for each database. The same experiments are repeated four times by selecting different fingerprints for training and testing, and then the average of four experimental results becomes the final performance.

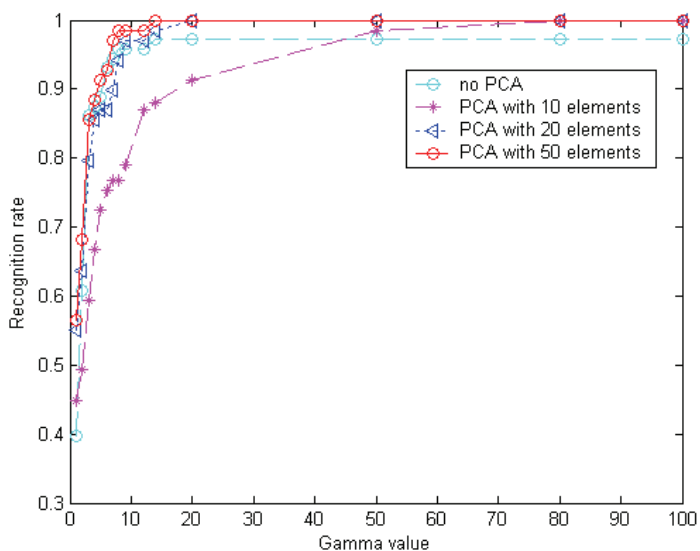


Fig. 7. The curves of the recognition rate of the proposed descriptor by different γ values of the radial-basis SVM with different PCA eigenvalues elements on the database FVC2002 DB2_a

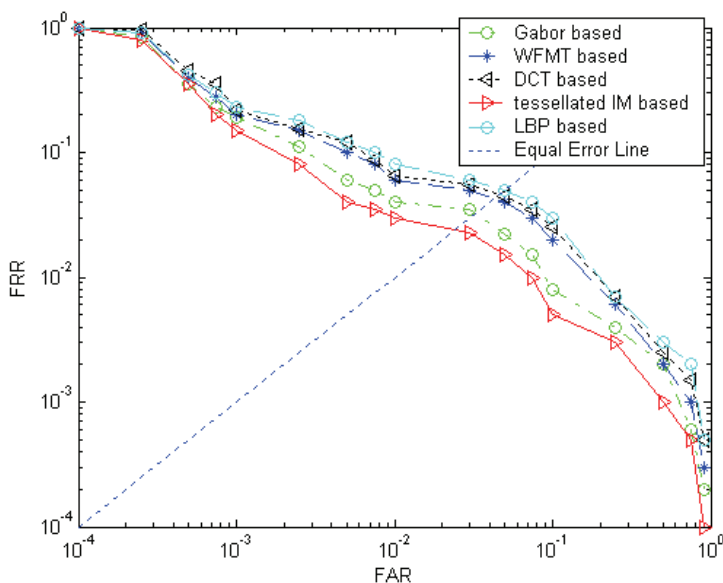


Fig. 8. ROC curves of different methods on database FVC2002 DB2_a

Figure 8 compares the ROC curves of the descriptors of Gabor filters, DCT, WFMT, LBP based with tessellated IM based descriptor on the database FVC2002 DB2_a. On database containing most poor quality images such as FVC2002 DB2_a, the FRRs of all the descriptors slowly drop down with respect to their FARs; The ROC curves of the Figure 8 prove the facts. On the other hand, we can see that the ROC curve of tessellated IM based descriptor (solid line) is below those of the descriptors of Gabor filters, DCT, WFMT, LBP based (dashed lines), which means that tessellated IM based descriptor outperforms the compared descriptors.

And Table 4 illustrates the EER (%) performances of the descriptors of Gabor filters based, DCT based, WFMT based and LBP based with Tessellated IM based over the databases of FVC2002. From the table, we can see that the Tessellated IM based descriptor has the best of results with an average EER of 2.36% over the four sub-databases of FVC2002, while those of the descriptors of Gabor filters, DCT, WFMT, LBP based are 4.17%, 5.68%, 4.66%, 5.97%, respectively.

	DB1_a	DB2_a	DB3_a	DB4_a	Average
Gabor filters based	1.87	3.98	4.64	6.21	4.17
DCT based	2.96	5.42	6.79	7.53	5.68
WFMT based	2.43	4.41	5.18	6.62	4.66
LBP based	3.21	5.62	6.82	8.23	5.97
Tessellated IM based	1.42	2.23	2.48	3.31	2.36

Table 4. EER(%) performance of the descriptors of Gabor filters, DCT, WFMT, LBP based and tessellated IM based over the databases of FVC2002

5. Conclusion

In this chapter, we emphasize on the introduction of the designing approaches and implementing technologies for non-minutiae based fingerprint descriptors. We firstly review some classical and state of the art fingerprint descriptor, analyze their feature extraction, alignment method and matching methods, and propose a non-minutiae based fingerprint descriptor by using tessellated IM features with a feature selection of PCA and a SVM classifier. The scheme consists of other three important steps after the previous pre-processing: feature vector construction, feature selection with PCA, and matching with a SVM. Experimental results show that our proposed descriptor has better performances of matching accuracy comparing to other prominent descriptors on public databases.

The contributions of this chapter are that we analyze and summarize some classical and state of the art non-minutiae based descriptors and propose a new improved one with tessellated IM features, feature selection by PCA and intelligent SVM to represent fingerprint image in order to effectively handle various input conditions. Further works need to be explored for robustness and reliability of the system.

6. Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 61063035), and it is also supported by Merit-based Science and Technology Activities Foundation for Returned Scholars, Ministry of Human Resources of China.

7. References

- Amornraksa, T. ; Tachaphetpiboon, S. (2006). Fingerprint recognition using DCT features, *Electron. Lett.*, 42(9), pp. 522-523
- Benhammadi, F. ;Amirouche, M.N. ;Hentous, H. ; Beghdad, K.B ; Aissani, M. (2007). Fingerprint matching from minutiae texture maps, *Pattern Recognition*,40(1), pp.189-197
- Cappelli, R.; Ferrara, M.; Maltoni, D.(2011). Minutia Cylinder-Code: A New Representation and Matching Technique for Fingerprint Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12), pp: 2128 - 2141
- Dalal, N.; Triggs, B. (2005). Histograms of oriented gradients for human detection. *In Proceedings of the ninth European conference on computer vision*, San Diego,CA
- Fausett, L. (1994), *Fundamentals of Neural Networks architecture, algorithm and application* (Prentice Hall) 289-304
- Gonzalez, R. C. ; Woods, R E.(2002), *Digital Image Processing (second edition)*, Prentice Hall, pp. 672-675
- Hao, P.Y. ; Chiang, J.H. ; Lin Y.H. (2007). A new maximal-margin spherical-structured multi-class support vector machine. *Applied Intelligence*, 30(2), pp.98-111
- He, X. ; Tian, J. ; Li, L. ; He, Y ; Yang X.(2007). Modeling and Analysis of Local Comprehensive Minutia Relation for Fingerprint Matching, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 4, 37(5), pp.1204-1211
- Hu, M.K.(1962). Visual pattern recognition by moment invariants, *IRE Trans. Info. Theory*. IT-8,pp. 179-187
- Jain, A. K.; Hong, L. ; Bolle, R. (1997a). On-line fingerprint verification, *IEEE Trans.Pattern Analysis and Machine Intelligence*, 19, pp. 302-313
- Jain, A.K. ; Hong, L. ; Pankanti, S. ; Bolle, R. (1997b) .An identity-authentication system using fingerprints, *Proc. IEEE*, 85(9) pp. 1365-1388
- Jain, A.K. ; Prabhakar, S. ; Hong, L. ; Pankanti, S. (2000). Filterbank-based fingerprint matching, *IEEE Trans. Image Processing*,9(5),pp. 846-859
- Jiang, X. ; Yau, W.(2000). Fingerprint minutiae matching based on the local and global structures, *International Conference on Pattern Recognition*, pp. 1038-1041
- Jin, A.T.B. ; Ling, D.N.C. ; Song, O.T. (2004). An efficient fingerprint verification system using integrated wavelet and Fourier-Mellin invariant transform, *Image and Vision Computing*, 22(6),pp. 503-513
- John, S.T. ; Nello, C. (2000), *Support Vector Machines and other kernel-based learning methods*, Cambridge University Press
- Kenneth, N. ; Josef, B. (2003). Localization of corresponding points in fingerprints by complex filtering, *Pattern Recognition Letters*, 24 ,pp.2135-2144
- Liu, J. ; Huang, Z. ; Chan K.(2000). Direct minutiae extraction from gray-level fingerprint image by relationship examination, *Int. Conf.. Image Processing*, 2,pp. 427-430
- Maio, D. ; Maltoni, D.(1997). Direct gray scale minutia detection in fingerprints, *IEEE Trans.Pattern Analysis and Machine Intelligence*,19(1), pp. 27-39
- Maio, D. ; Maltoni, D. ; Cappelli, R. ; Wayman, J.L. ; Jain, A.K.(2002). FVC2002: Second Fingerprint Verification Competition, *Proc. Internationl Conference on Pattern Recognition*,(3), pp.811-814
- Maltoni, D. ;Maio, D. ;Jain, A.K. ; Prabhakar S. (2003). *Handbook of Fingerprint Recognition*, Springer, Berlin, pp.135-137

- Nagaty, K.A.(2005). An adaptive hybrid energy-based fingerprint matching technique, *Image and Vision Computing*, 23,pp.491-500
- Nanni, L. ; Lumini, A. (2007). A hybrid wavelet-based fingerprint matcher, *Pattern Recognition*, 40(11),pp. 3146-3151
- Nanni, L. ; Lumini, A. (2008). Local Binary Patterns for a hybrid fingerprint matcher, *Pattern Recognition*, 41(11), pp.3461-3466
- Nanni, L. ; Lumini, A.(2009). Descriptors for image-based fingerprint matchers, *Expert Systems With Applications*, 36(10), pp.12414-12422
- Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp.971-987.
- Papoulis, A. (1991). *Probability, random variables and stochastic processes*, Boston: McGraw-Hill.
- Ratha, N. K. ; Karu, K. ; Chen, S. ; Jain, A. K.(1996). A real-time matching system for large fingerprint databases, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), pp.799-813
- Ross, A., Jain, A.K., Reisman, J.(2003). A hybrid fingerprint matcher, *Pattern Recognition*, 36(7), pp. 1661-1673
- Sha, L.F. ; Zhao F. ; Tang X.O.(2003). Improved fingerprint code for filterbank-based fingerprint matching, *Int. Conf. Image Processing*, pp.895-898
- Tico, M. ; Kuosmanen, P. ; Saarinen, J. (2001) . Wavelet domain features for fingerprint recognition, *Electron. Lett.*, 37(1), pp. 21-22
- Yang, J.C. ; Yoon, S. ; Park, D.S. (2006). Applying learning vector quantization neural network for fingerprint matching, *Lecture Notes in Artificial Intelligence (LNAI 4304)* (Springer, Berlin) , pp.500-509
- Yang, J.C. ; Park, D. S. (2008a). A fingerprint verification algorithm using tessellated invariant moment features, *Neurocomputing*, 71(10-12), pp.1939-1946
- Yang, J.C. ; Park, D. S. (2008b). Fingerprint Verification Based on Invariant Moment Features and Nonlinear BPNN, *International Journal of Control, Automation, and Systems*, 6(6), pp.800-808
- Yang, J.C. ; Park, D.S. ; Hitchcock, R. (2008c). Effective Enhancement of Low-Quality Fingerprints with Local Ridge Compensation, *IEICE Electronics Express*, 5(23), pp.1002-1009

Retinal Identification

Mikael Agopov
University of Heidelberg
Germany

1. Introduction

Since the pioneering studies of Drs. Carleton Simon and Isidore Goldstein in 1935 [Simon & Goldstein (1935)], it has been known that every eye has its own unique pattern of blood vessels, and that retinal photographs can be used for identifying people. In the 1950's, this was proven to hold even for identical twins [Tower (1955)]. Hence the idea of using the retinal blood vessel pattern for identification. Eye fundus photography for the purpose of identification is impractical, however. An optical device which scans the retinal blood vessel pattern is required.

Such an optical Retinal Identification (RI) device was originally patented in 1978 [Hill (1978)]; after several subsequent patents, it developed into a commercial product in the 1980s and 1990s. As the patent of retinal identification (opposed to the actual design of the device) has now worn off, new developments in the field have taken place. In this chapter, the history, technique and recent developments of RI are discussed.

1.1 The anatomy and optical properties of the human eye

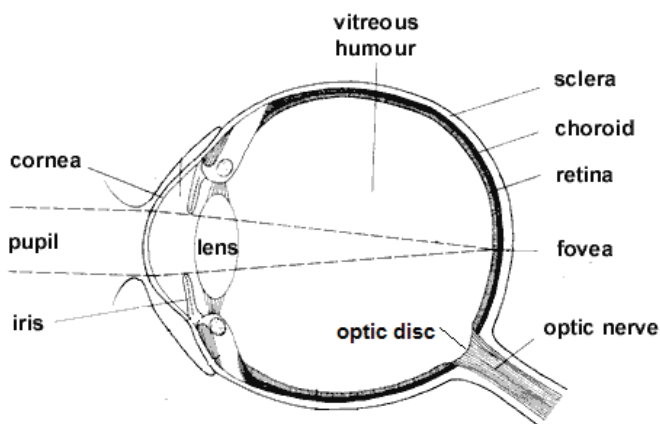


Fig. 1. A schematic picture of the human eye.

A schematic picture of the human eye is shown in Figure 1. The eyeball is of about 24 mm in diameter and filled with *vitreous humor*, jelly-like substance similar to water; its outer shell, the *sclera*, is made of rigid proteins called collagen. The light entering the eye first passes through the *pupil*, an aperture-like opening in the *iris*. The size of the pupil limits the amount of light entering the eye. The light is focused by the *cornea* and the *crystalline lens* onto the *retina*. The retina converts the photon energy into an electric signal, which is transferred to the brain through the *optic nerve*.

The cornea is about 11 mm in diameter and only 0.5 mm thick. It accounts for most of the refractive power of the eye (about 45 D). The remaining 18 D come from the crystalline lens, which is also - through deformation - able to change its refractive power, thus partly compensating for the refractive error and helping to focus the eye.

The retina is a curved surface in the back of the eye. The point of sharpest vision is called *fovea* - here the light-sensing *photoreceptor cells* are only behind a small number of other cells. Elsewhere on the retina, the light has to travel through a multi-layered structure of different cells. These various cells are responsible for the eye's 'signal processing', i.e. turning the incoming photons first into a chemical and then to an electric signal.

After being amplified and pre-processed, the signal is transferred to the nerve fibers, which reside on the peripheral area of the retina around the *optic disc*, where they form the *retinal nerve fiber layer* (RNFL). The optic disc is an approximately $5^\circ \times 7^\circ$, ellipse-like opening in the eye fundus, through which the nerve fibers and blood vessels enter the eyeball. It is about 15° away from the fovea in the nasal direction. The *choroid* is the utmost layer behind the retina just in front of the sclera. It has a bunch of small blood vessels, and is responsible for the retina's metabolism.

1.2 The birefringence properties of the eye

Birefringence is a form of optical anisotropy in a material, in which the material has different indices of refraction for p- and s-polarization components of the incoming light beam. The components are thus refracted differently, which in general results the beam being divided into two parts. If the parts are then reflected back by a diffuse reflector (such as the eye fundus), a small portion of the light will travel the same way as it came, joining the polarization components into one again, but having changed the beam's polarization state in process¹.

The birefringence of the eye is well documented (Cope et al. (1978), Klein Brink et al. (1988), Weinreb et al. (1990), Dreher et al. (1992)). The birefringence of the corneal collagen fibrils constitutes the main part of the total birefringence of the eye. Its amount and orientation changes throughout the cornea. In the retina, the main birefringent component is the retinal nerve fiber layer (RNFL), which consists of the axons of the nerve fibers. The thickness of RNFL is not constant over the retina; the amount of birefringence varies according to the RNFL thickness and also drops steeply if a blood vessel (which is non-birefringent) is encountered. The most successful application of measuring the RNFL thickness around the optic disc is probably the GDx glaucoma diagnostic device (Carl Zeiss Meditec, Jena, Germany). It uses scanning laser polarimetry to topograph the RNFL thickness on the retina. A reduced RNFL thickness means death of the nerve fibers and thus advancing glaucoma. A typical GDx image is shown in Figure 2.

¹ See Appendix about how the polarization change can be measured.

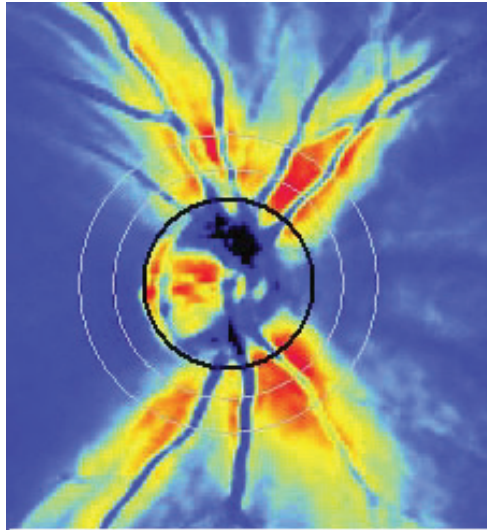


Fig. 2. A typical GDx image of a healthy eye. The birefringent nerve fiber layer is seen brightly in the image, as well as the blood vessels which displace the nerve fibers, thus resulting in a weaker measured signal (darker in the image).

2. RI using retinal blood vessel absorption

The first patent of the biometric identification using the retinal blood vessel pattern dates back to 1978 [Hill (1978)]. Soon afterwards, the author of the patent founded the company EyeDentity (then Oregon, Portland, USA) and began full-time efforts to develop and commercialize the technique.

In the original patent, the retinal blood vessel pattern is scanned with the help of two rings of LEDs. The amount of light reflected back from the retina is measured - when the beam hits a blood vessel, it is absorbed to a bigger extent than when it hits other tissue. In the original retina scan, green laser light was used - it was strongly absorbed by the red blood vessels. However, it was found out that visible light causes discomfort to the identified individual, as well as pupil constriction, causing loss of signal intensity.

Since the first working prototype RI, patented in 1981 [Hill (1981)], near-infrared (NIR) light has been used for illumination. The infrared light is not absorbed by the photoreceptors (the absorption drops steeply above 730 nm); however, the retinal blood vessels are fairly transparent to the NIR wavelengths as well - the light is absorbed by the smaller choroidal blood vessels instead (thus, considering this technique, the term *Retinal Identification* is slightly misleading) before being reflected back from the eye fundus. The image acquisition technique has also been changed: the LEDs are given up in favour of scanning optics. A circular scan is preferred over a raster, which suffers from the problem of reflections from the cornea.

In the patent of 1986 [Hill (1986)], the scan is centered around the fovea instead of the optic disc. The fovea is on the optical axis of the eye, so this arrangement has the definite advantage that no fixation outside the normal line of sight is required, unlike when a ring around the optic disc is scanned, when the subject has to look 15 degrees off-axis. The downside is that

the choroidal blood vessels are much thinner than around the optic disc, and they don't form a clear pattern. Thus the price paid for easier fixation is the quality of the signal.

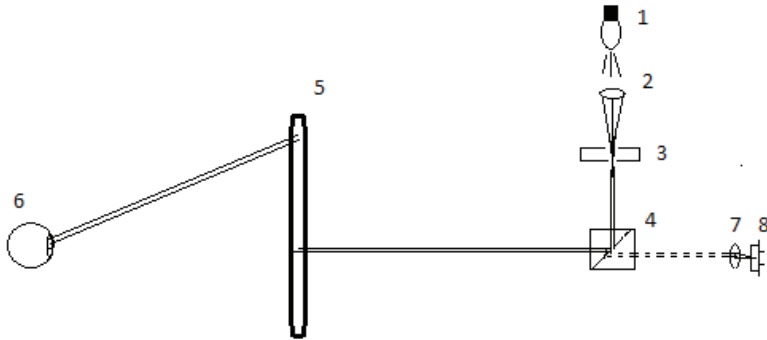


Fig. 3. A schematic drawing of the current Retinal Identification technique, based on the patent from 1996. A detailed explanation is in the text.

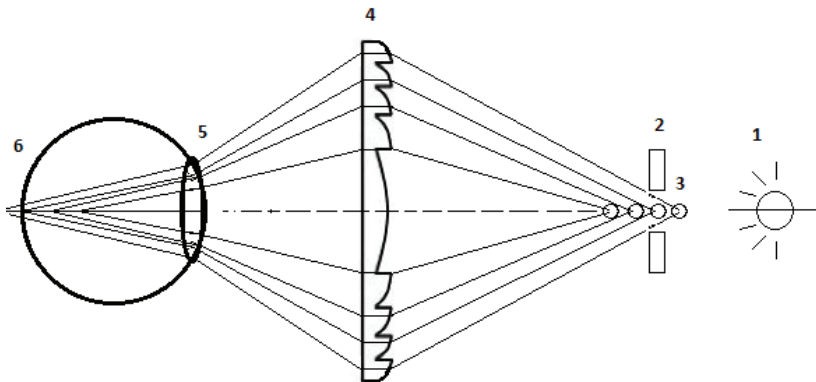


Fig. 4. A schematic drawing fixation-alignment technique using a Fresnel lens. A detailed explanation is in the text.

Current RI technology is based on an active US patent [Johnson & Hill (1996)]. A schematic drawing of the measurement setup is shown in Figure 3. An infrared light source, for example a krypton lamp (1), is focused by a lens (2) through an optical mirror (explained below) via a pinhole (3). The light enters a beam splitter (4) which reflects it into a Fresnel optical scanner (5). The rotating optical scanner scans a ring on the cornea, which - if properly focused - will hit the retina at the same angle. A multifocus Fresnel lens, cemented in the scanner, creates a fan of nearly-collimated light beams which hit the eye of the tested subject. One of the

beams will be focused on the retina by the eye's own optical apparatus, thus compensating for refractive error (explained in detail below).

The light reflected back from the eye fundus travels the same way through the scanner and into the beam splitter; a part of it is transmitted into a photodetector (8) through a focusing lens (7). After being measured by the photodetector, the signal is A/D-converted, amplified and processed. The processed signal is converted into points, which are stored in an array, which is used for matching. A similar process is used in all RI techniques.

Fixation and alignment of the subject's eye in a RI measurement is critical; it is almost impossible to scan a non-willing subject. The fixation system of the RI technique, which was also patented (Arndt (1990)), is illustrated in Figure 4. A Light-Emitting Diode (LED) is situated next to the Krypton lamp. It illuminates the optical double-surface mirror, creating several reflections, 'ghost images', of the LED on the optical axis of the system. These images function as targets for the test subject's eye. The eye looks at them through a multifocal Fresnel lens. The lens, which consists of several focusing parts with different focal lengths, focuses the ghost images on different points on the eye's optical axis. Regardless of whether the test subject is emmetropic (normal visual acuity), myopic (near-sighted) or hyperopic (far-sighted), one of the images will almost certainly end up on the retina and will thus result a sharp image and effectively compensate for the refractive error of the subject's eye. However, this happens at the cost of optical image quality; the measured pattern is a sum of contributions from choroidal blood vessels and other structures.

2.1 Matching

At first, a reference measurement, which the further measurements will be compared against, has to be taken from each tested subject. Any further measurement will be compared against the reference. As the subject's eye can rotate around the optical axis (due to different head position, i.e. head tilt, between the measurements), the best possible match is found by 'rotating' the measurement points in the array. The matching is done using a Fourier-based correlation; the match is measured on a scale of +1,0 (a perfect match) to -1,0 (a complete mismatch). User experience has shown that a match above 0,7 can be considered a matching identification.

3. RI using the RNFL birefringence

The blood vessels emerging into the retina through the optic disc often displace the nerve fibers in the retinal nerve fiber layer. Unlike the RNFL nerve fiber axons, the blood vessels are not birefringent - thus, if the birefringence (change in the state of polarization) of the scanning laser beam is measured around the optic disc, a steep signal drop proportional to the blood vessel size is measured wherever one is encountered. This can be seen in the GDx-pictures, where the blood vessels are seen as dark lines on the otherwise bright nerve fiber layer.

The author and his co-workers studied the possibility of using blood vessel-induced RNFL birefringence changes for biometric purposes [Agopov et al. (2008)]. A measurement device was built to scan a circle of 20° around the optic disc. The measured birefringence would drop steeply where a blood vessel is encountered, creating a sharp drop, or 'blip', in the measured signal. The scanning angle is big enough to catch the major blood vessels which enter the fundus through the disc.

3.1 Apparatus and method

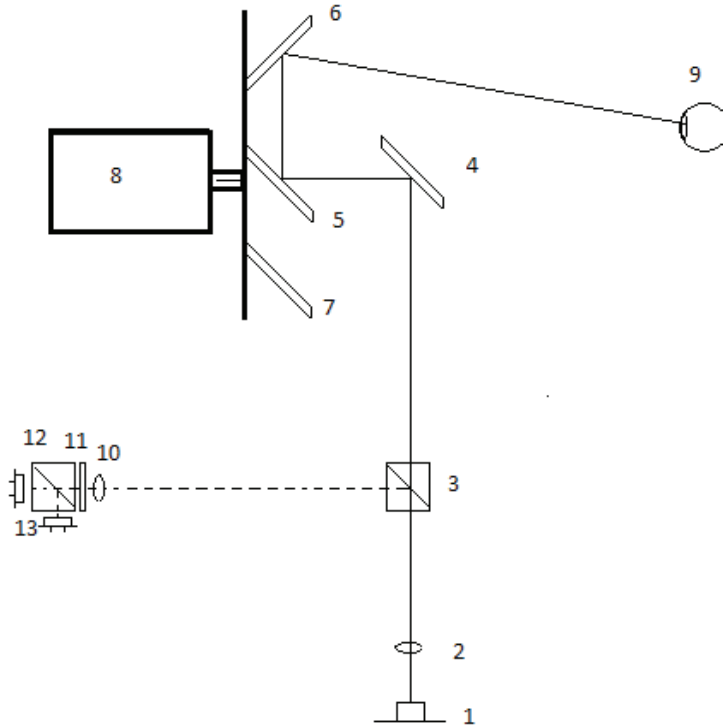


Fig. 5. A schematic drawing of the RI measurement setup using the RNFL birefringence. A detailed explanation is in the text.

The measurement setup is explained in detail in [Agopov et al. (2008)]; here it is explained briefly. The apparatus and the light paths within it are shown in Figure 5. The light path is drawn with a solid line. A 785 nm laser diode (1) was used as a light source. The beam was collimated by a lens (2); then it passed through a non-polarizing beam splitter (3) and was reflected by a mirror (4) further into the optical scanner (5-8). The two scanning mirrors (5 and 6) and a counterweight (7) were cemented on an aluminium plate which was spun by a DC motor (8). The center mirror was (5) tilted 45° from the disc plane rotated clockwise and reflected the laser beam onto the edge mirror. The edge mirror (6) was tilted 50° , creating a circular scan subtending a 10° radius of visual angle (20° diameter) in the tested subject's eye (9). The reflection from the ocular fundus traveled back the same optical path through the scanner; now, however, the beam splitter (3) reflected the useful half of the beam (drawn with dashed line) into the detection system (10-13). A lens (10) focused the beam into a polarizing beam splitter (12) through a quarter wave plate (11) which had its fast axis 45° to the original plane of polarization. The polarizing beam splitter separated the p- and s-polarization components; two avalanche photodiodes (13) were placed right after it, measuring the two signals which corresponded the two polarization components. Amplified by the detection electronics, the signals were added and subtracted

respectively. The polarization was manipulated so that the Stokes parameters S_0 and S_3 were measured - it was decided to measure S_3 instead of S_1 for birefringence-based changes as it appeared to suffer less from various amounts and orientations of corneal birefringence in our computerized model.

As the alignment of the eye is critical, special care was taken to properly align the test subject's eye. The measurement apparatus included three eyepieces: the measurement was taken through the fixed central piece; in addition there were two horizontally movable ones; the subject could look through the central piece with either eye while having a 'dummy' eyepiece available for the other eye. Thus possible head tilt was reduced to almost zero.

Because the fixation point of the retina, the fovea, is approximately 15 degrees away from the optic disc, the measured eye had to look 15° away in the nasal direction to center the scan on the optic disc. To achieve this, two fixation LEDs were set at 15° angle to the central axis of the scan. Because the human eye is about 0,75 D myopic at the wavelength 785 nm (see [Fernandez et al. (2005)] for details) the fixation LEDs were placed at 130 cm distance, so that the eye's fixation would compensate for this.

4. RI using the optic disc structure

Another interesting RI technology was patented in 2004 [Marshall & Usher (2004)]. The idea is to use the image of the optic disc - taken by a scanning laser ophthalmoscope (SLO) - for identification. A company Retinal Technologies (Boston, MA, USA) was founded for developing the technique.

4.1 The principle of an SLO

The best-known application of the SLO is probably the Heidelberg Retina Tomograph (Heidelberg Engineering, Heidelberg, Germany), which is used for glaucoma diagnostics. The principle of an SLO is illustrated in 6. A low-intensity laser diode (1) is used for illumination. A collimated laser beam goes through a beam splitter (2) into an optical scanner (3). The scanner consists of two rotating mirrors, a fast and a slow one, creating a raster scan. The scanning beam is imaged through two lenses (4 and 5) onto exactly one point called the *conjugated plane*. If the imaged subject's cornea is at this point(6), the eye's optics focus the scanning beam onto the retina (7). The reflection from the eye fundus travels back through the system, but is reflected into the detection system (7-9) by the beam splitter. A lens (7) focuses the beam through a pinhole (8) onto a photodiode (9). The pinhole is very important - only the light which comes exactly from the conjugated plane reaches the photodetector. Thus the SLO creates a high-resolution microscopic image of the retina. The scan is usually centered around the optic disc (using an off-axis fixation target - as in the previous setup). A typical SLO image of the optic disc (taken with the HRT) is shown in Figure 7.

4.2 Image analysis

The boundary of the optic disc is found from the image taken by the SLO. There is a clear boundary between the disc and surrounding tissue (as seen in Figure 7); an ellipse is fit onto the image by analyzing the average intensity of the pixels around the boundary.

Once the disc is identified, a recognition pattern is created from the its structures. This fairly complicated procedure is explained in detail in the patent. The patterns of recognized

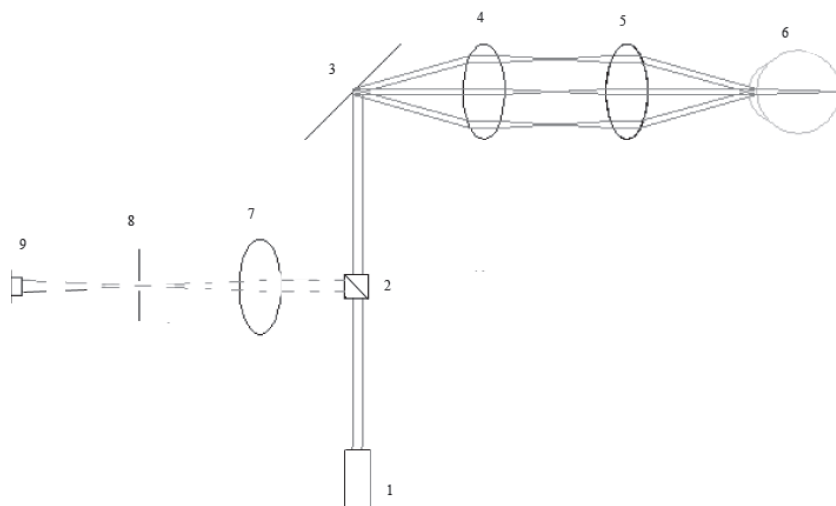


Fig. 6. An schematic drawing of a scanning laser ophthalmoscope. A detailed explanation is in the text.

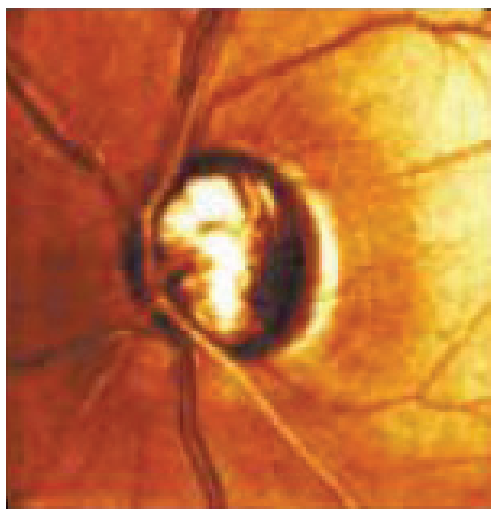


Fig. 7. A typical SLO image of the optic disc (taken with an HRT).

individuals are stored in a database for comparing with the pattern of a person wishing to be identified.

5. Combined retina and iris identification

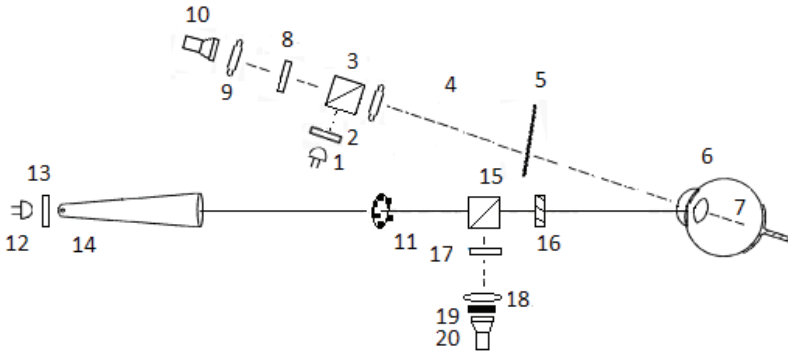


Fig. 8. The measurement setup of a combined retina and iris identification device. The details are in the text.

In a fairly recent patent [Muller et al. (2007)], a combination of retina and iris identification was suggested (about iris identification, see the previous chapter). The measurement system is constructed so that both biometrics can be recorded with one scan. This is obviously advantageous, as now two biometrics are available simultaneously. The measurement setup is shown in Figure 8. The setup has two optical axes: one for the retinal image (dashed line) and one for alignment and the iris image (solid line). They are 15° apart; as explained earlier, the incoming light is focused on the area around the optic disc only if the eye is looking 15° in the nasal direction. To achieve this, the test subject has to be faced towards the dashed line but look in the direction of the solid line. As a fixation target, a ring of green LED's (11) is placed around the optical axis of the iris scanning system.

Two illuminating LED's are used: One infrared ($\lambda \approx 800$ nm) LED (1), and one red ($\lambda \approx 660$ nm) LED (12). The interference from the 'wrong' light source is blocked a dichroic mirror (5). The light from the infrared LED (1) first goes through a polarizer (2), which ensures that the outgoing light is linearly polarized. The outgoing beam is then divided into two parts by a beam splitter (3). The transmitted (useless) part crosses the beam splitter and is preferably absorbed by a light trap. The reflected (useful) beam part is collimated by a lens (4) and goes through the dichroic mirror (5) into the eye. The beam is focused by the eye's optics (6) onto the area around the optic disc; after being reflected from the eye fundus (7), a reflection of the beam returns back the same way as the beam came; however, coming from the other direction, a part of the reflected beam now passes through the beam splitter into the detection system. The polarizer (8) is set so that it absorbs the illuminating LED's polarization direction. However - having changed its state of polarization while passing through the eye tissue - a part of the beam is now able to pass through the polarizer. The beam is focused by a lens (9) onto the detector (10), which can be for example a CCD camera. It should be noted that this setup has no optical scanner (nor it is confocal) - a wide-field image around the optic disc is

captured, thus resulting in lower resolution than a confocal scanning system would achieve. For the iris scan, the illuminating light from the red LED (12) is first linearly polarized (13)², and then guided into the eye through an alignment tube (14). The function of the alignment tube is not explained in the patent; preferably it would consist of at least one lens with a long focal length, which would focus the light onto the iris (the beam entering the eye should not be parallel, otherwise it will be focused on the retina). The distance between the lens and the beam splitter should be much bigger than that between the beam splitter and the eye, so that the slightly de-focused reflection image of the iris can be caught. On its way to the iris and back, the light double-passes a quarter wave plate (16). The wave plate's axes are set so that both the fast and the slow axis of it are 45° to the original polarization direction - when the beam passes through it, its polarization becomes circular; the second pass (the reflection from the iris) turns the circular polarization into linear again, but having turned the polarization direction by 90° in the process. After entering the detection system (17-20), the light can now pass through the polarizer (17), which is set to absorb the polarizing direction of the initial polarizer (13). The beam is focused by a lens onto the detector (a CCD camera); the possible reflections of the green LED are filtered out by a red band pass filter (19).

The two detection systems are electronically synchronized so that one scan records both images simultaneously. The recording is triggered by a switch, which is turned on when the eye is correctly aligned.

The eye is at a crossing of two optical axes - its distance and orientation are critical. In the setup suggested in the patent, the distance is controlled by an ultrasound transducer. It sends and receives ultrasound pulses which are reflected back from the surface of the cornea. When the distance is right, and the optic disc is seen on the CCD (10), the eye is aligned properly, and the recording can be taken.

6. Results of performance tests and limitations of the RI techniques

6.1 RI using absorption

In a performance test by Sandia National Laboratories [Sandia Laboratories (1991)], the EyeDentity RI device recognized >99% of the tested subjects in a three-attempt measurement, with no false positives.

6.1.1 Limitations

As the light has to pass twice through the pupil during the measurement, a constricted pupil can increase the number of false negative scan results. Thus dim light conditions are preferable for the RI (of course, this is true for almost any optical measurement); the technique has difficulties in broad daylight. In addition, various eye conditions can disturb the light's passing through the eye, compromising vision; this also affects measuring the eye's properties optically, including the RI.

² In the patent, the device is described without the polarizers (13 and 17), the quarter wave plate (16), the filter (19); instead of a beam splitter (15), a dichroic mirror is suggested. The fixation LEDs are also placed together with the red illuminating LED. However, this leads to unsurmountable difficulties. First of all, the IR light and the green light used for fixation cannot *both* pass through the dichroic mirror; moreover, the IR light would first have to *pass* through the mirror - the reflection from the eye would then have to be *reflected* by it. Therefore, the author suggests slight modifications in the setup.

1. Severe astigmatism: An astigmatic eye's optics image dots as lines. This results in problems with focusing and also bad optical quality of measurements.
2. Cataracts: A cataract is an eye condition in which clouding develops in the crystalline lens. The lens becomes opaque so seeing becomes compromised. Obviously any optical measurement in the eye, including the RI, becomes increasingly difficult.
3. Severe eye diseases, such as the age-related macula degeneration (AMD), can change the structure of the retina, either by destroying retinal tissue or by stimulating growth of new blood vessels.

6.2 RI using birefringence

Eight eyes of four volunteer subjects were measured. Both absorption- and birefringence-based signals were recorded two times for each eye. For verifying purposes, fundus photographs were taken from all the eyes. The measured peaks and the blood vessels on the fundus photos were compared as follows:

A 20° diameter circle was drawn on a transparent overlay, on which the measured peaks were marked at corresponding angles on the perimeter of the circle. The transparency was then placed on the fundus photo to compare the marked signal peaks with the blood vessels on the photo. Only the vessels larger than a certain threshold size (set for each eye individually) were taken into account, i.e. the smallest vessels were ignored. In this way, the numbers of blood vessels corresponding to the peaks in the measured absorption/reflectance and birefringence-based tracings were calculated. If a confirmed peak did not correspond to a vessel above the threshold size on the fundus photo, it was considered a false positive.

Altogether 55 blood vessels were located on the fundus photos of the 'better' eyes (the ones which yielded clearer signals) of the four volunteers, of which 34 could be correlated with the 'peaks' in the measured signal. The calculated sensitivities (number of vessels identified / number of vessels altogether) and specificities (number of positive recognitions / number of positive + false positive recognitions) - are presented in Table 1. The columns in the tables represent total percentage of blood vessel 'peaks' correlating with vessels in the fundus photo (Total), from the two reflection/absorption measurements (Sum) and from the two birefringence measurements (Diff).

	Total	Sum	Diff
Average sensitivity	69%	51%	33%
Average specificity	78%	76%	60%

Table 1. Summarized results of the measurements taken from the eye with a better signal of each subject.

6.2.1 Limitations

Our system was of 'proof-of-principle' -nature and - unlike the conventional RI technique - had no inherent defocus compensation; the test subjects were required to have a refractive error of less than about ± 2 diopters or a good corrected vision using contact lenses. The eye conditions disturbing the measurement include cataracts and astigmatism as well as a severe glaucoma, which damages the RNFL by killing the nerve fibers going through the optic nerve.

6.3 Other techniques

The author is unaware of any scientific studies on the accuracy of the other techniques mentioned here.

7. Discussion

Retinal Identification remains the most reliable and secure biometric. Falsifying an image of a retinal blood vessel pattern appears impossible. The eye scans are generally considered invasive or even harmful, especially if a laser (even if the light is weak intensity and harmless to the eye, as in our system). However, the enrollees should be able to overcome their fear of these scans once they acquire more user experience.

The original RI technique uses the choroidal blood vessel pattern for identification; however, several newer RI techniques center the scan around the optic disc. The main drawback in such a device is that the eye has to fixate 15 degrees off-axis while being measured. However, if this is achieved, the blood vessels around the optic disc are easier to detect - in addition, the - as the authors and his co-workers proved - birefringence-based blood vessel detection can help detecting more blood vessels, only at a small cost of specificity. The measured signal was also directly linked to blood vessels and not to other retinal structures, as in the original RI technique.

The newer techniques - imaging the optic disc using an SLO, or the combined retina- and iris scan - appear very promising. However, the author is unaware either of any commercial devices or of any scientific studies on the accuracy of these identifying method.

The use of retinal identification is not limited to humans. In 2004, a patent was filed on identifying various animals using their retinal blood vessel pattern [Golden et al. (2004)]. To develop the technique for animals, a company OptiBrand was started (Ft. Collins, Colorado, USA). The company produces and develops hand-held video camera -based devices which provide an acceptable image an animal's eye fundus to allow identification. This method is certainly preferable over the traditional hot iron branding, which is not only painful to the animal, but also costly in the lost hide value. When a false identification is not disastrous (as opposed to some high security installations), a simple video camera based device provides an accurate enough identification.

8. Appendix: The Stokes parameters

The modern treatment of polarization was first suggested by Stokes in the mid-1800's. The polarization state of the light can be completely represented with four quantities, the *Stokes parameters*:

$$\begin{aligned}
 S_0 &= \langle E_x^2 \rangle_T + \langle E_y^2 \rangle_T \\
 S_1 &= \langle E_x^2 \rangle_T - \langle E_y^2 \rangle_T \\
 S_2 &= \langle 2E_x E_y \cos \epsilon \rangle_T \\
 S_3 &= \langle 2E_x E_y \sin \epsilon \rangle_T,
 \end{aligned} \tag{1}$$

where $\epsilon = \epsilon_y - \epsilon_x$ is the phase difference between x- and y-polarization and the $\langle \rangle_T$ denote time averages.

For totally unpolarized light $S_0 > 0$ and $S_1 = S_2 = S_3 = 0$. For completely polarized light, on

the other hand, $S_0^2 = S_1^2 + S_2^2 + S_3^2$. The polarization state of the light can be defined as

$$V = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0}. \quad (2)$$

The importance of the Stokes parameters lies in the fact that they are connected to easily measurable intensities:

$$\begin{aligned} S_0(\theta, \rho) &= I(0^\circ, 0) + I(90^\circ, 0) \\ S_1(\theta, \rho) &= I(0^\circ, 0) - I(90^\circ, 0) \\ S_2(\theta, \rho) &= I(45^\circ, 0) - I(135^\circ, 0) \\ S_3(\theta, \rho) &= I(45^\circ, \pi/2) - I(135^\circ, \pi/2), \end{aligned} \quad (3)$$

where θ is the angle of the azimuth vector measured from the x-plane and ρ is the birefringence. These can be easily measured: for example, S_1 can be measured with polarizing beam splitter and two detectors, which are placed so that they measure the intensities coming out of the beam splitter, and S_3 can be measured by adding a quarter wave plate before the aforementioned system.

9. References

- Agopov, M.; Gramatikov, B.I.; Wu, Y.K.; Irsch, K; Guyton, D.L. (2008). Use of retinal nerve fiber layer birefringence as an addition to absorption in retinal scanning for biometric purposes. *Applied Optics* 47: 1048-1053
- Bettelheim, F. (1975). On optical anisotropy of the lens fiber cells. *Exp. Eye Res.* 21: 231-234
- Cope, W.; Wolbarsht; Yamanashi, B. (1978). The corneal polarization cross. *J. Opt Soc Am* 68: 1149-1140
- Dreher, A.; Reiter, K. (1992). Scanning laser polarimetry of the retinal nerve fiber layer. *SPIE* 1746:34-41
- Fernandez, E. J.; Unterhuber, A; Pioto, P.M.; Hermann, B.; Drexler, W.; Artal, P. (2005). Ocular aberrations as a function of wavelength in the near infrared measured with a femtosecond laser. *Optics Express* 13:400-409
- Hill, R. (1978). Apparatus and method for identifying individuals through their retinal vasculature patterns, *United States Patent* 4109237
- Hill, R. (1981). Rotating beam ocular identification apparatus and method. *United States Patent* 4,393,366
- Hill, R. (1986). Fovea-centered eye fundus scanner. *United States Patent* 4620318
- Johnson, J.C.; Hill, R. (1996). Eye fundus optical scanner system and method, *United States Patent* 5532771
- Golden, B.L.; Rollin, B.E.; Switzer, R.; Comstock, C.R. (2004). Retinal vasculature image acquisition apparatus and method, *United States Patent* 6766041
- Arndt, J.H. (1990). Optical alignment system. *United States Patent* 4923297
- Klein Brink, H. B.; Van Blockland, G.J. (1988). Birefringence of the human foveal area assessed in vivo with the Mueller-Matrix ellipsometry. *J. Opt. Soc. Amer. A* 5:49-57
- Marshall, J.; Usher, D. (2004). Method for generating a unique and consistent signal pattern for identification of an individual, *United States Patent* 6757409

- Muller, D. F.; Heacock G. L.; Usher D. B. (2007). Method and system for generating a combined retina/iris patten biometric, *United States Patent 7248720*
- Sandia National Laboratories (1991). Performance Evaluation of Biometric Identification Devices, *Technical Report SAND91-0276, UC-906*
- Simon C.; Goldstein I. (1935). A New Scientific Method of Identification, *New York State Journal of Medicine*, Vol. 35, No. 18, pp. 901-906
- Tower, P. (1955). The fundus Oculi in monozygotic twins: Report of six pairs of identical twins, *Archives of Ophthalmology*, Vol. 54, pp. 225-239
- Weinreb, R.; Dreher A.; Coleman, A.; Quigley H.; Shaw, B.; Reiter, K. (1990). Histopatologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness, *Arch Ophthalmol* 108:557-60

Retinal Vessel Tree as Biometric Pattern

Marcos Ortega and Manuel G. Penedo
University of Coruña
Department of Computer Science
Spain

1. Introduction

In current society, reliable authentication and authorization of individuals are becoming more and more necessary tasks for everyday activities or applications. Just for instance, common situations such as accessing to a building restricted to authorized people (members, workers,...), taking a flight or performing a money transfer require the verification of the identity of the individual trying to perform these tasks. When considering automation of the identity verification, the most challenging aspect is the need of high accuracy, in terms of avoiding incorrect authorizations or rejections. While the user should not be denied to perform a task if authorized, he/she should be also ideally inconvenienced to a minimum which further complicates the whole verification process Siguenza Pizarro & Tapiador Mateos (2005).

With this scope in mind, the term *biometrics* refers to identifying an individual based on his/her distinguished intrinsic characteristics. Particularly, this characteristics usually consist of physiological or behavioral features. Physiological features, such as fingerprints, are physical characteristics usually measured at a particular point of time. Behavioral characteristics, such as speech or handwriting, make reference to the way some action is performed by every individual. As they characterize a particular activity, behavioral biometrics are usually measured over time and are more dependant on the individual's state of mind or deliberated alteration. To reinforce the active versus passive idea of both paradigms, physiological biometrics are also usually referred to as *static biometrics* while behavioral ones are referred to as *dynamic biometrics*.

The traditional authentication systems based on possessions or knowledge are widely spread in the society but they have many drawbacks that biometrics try to overcome. For instance, in the scope of the knowledge-based authentication, it is well known that password systems are vulnerable mainly due to the wrong use of users and administrators. It is not rare to find some administrators sharing the same password, or users giving away their own to other people. One of the most common problems is the use of easily discovered passwords (child names, birth dates, car plate,...). On the other hand, the use of sophisticated passwords consisting of numbers, upper and lower case letters and even punctuation marks makes it harder to remember them for an user.

Nevertheless, the password systems are easily broken by the use of *brute force* where powerful computers generate all the possible combinations and test it against the authentication system. In the scope of the possession-based authentication, it is obvious that the main concerns are

related to the loss of the identification token. If the token was stolen or found by another individual, the privacy and/or security would be compromised. Biometrics overcome most of these concerns while they also allow an easy entry to computer systems to non expert users with no need to recall complex passwords. Additionally, commercial webs on the Internet are favored not only by the increasing trust being transmitted to the user but also by the possibility of offering a customizable environment for every individual along with the valuable information on personal preferences for each of them.

Many different human biometrics have been used to build a valid template for verification and identification tasks. Among the most common biometrics, we can find the fingerprint Bolle et al. (2002); Maio & Maltoni (1997); Seung-Hyun et al. (1995); Venkataramani & Kumar (2003), iris Chou et al. (2006); He et al. (2008); Kim, Cho, Choi & Marks (2004); Ma et al. (2002); Nabti & Bouridane (2007) or face Kim, Kim, Bang & Lee (2004); Kisku et al. (2008); Mian et al. (2008); Moghaddam & Pentland (1997); Yang et al. (2000) or hand geometry Jain et al. (1999); Sidlauskas (1988); systems Lab (n.d.); Zunkel (1999). However, there exist other emerging biometrics where we can find retina biometrics. Identity verification based on retina uses the blood vessels pattern present in the retina (Figure1).

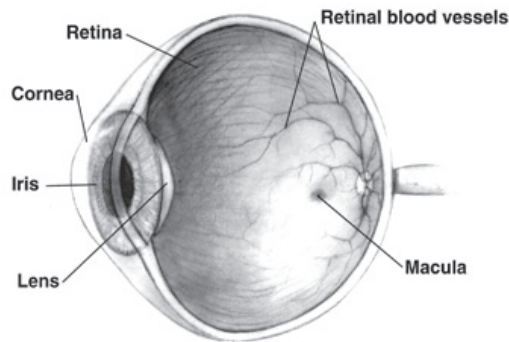


Fig. 1. Schema of the retina in the human eye. Blood vessels are used as biometric characteristic.

Retinal blood vessel pattern is unique for each human being even in the case of identical twins. Moreover, it is a highly stable pattern over time and totally independent of genetic factors. Also, it is one of the hardest biometric to forge as the identification relies on the blood circulation along the vessels. These property make it one of the best biometric characteristic in high security environments. Its main drawback is the acquisition process which requires collaboration from the user and it is sometimes perceived as intrusive. As it will be further discussed, some advances have been done in this field but, in any case, this continues to be the weak point in retinal based authentication.

Robert Hill introduced the first identification system based on retina Hill (1999). The general idea was that of taking advantage of the inherent properties of the retinal vessel pattern to build a secure system. The system acquired the data via a scanner that required the user to be still for a few seconds. The scanner captured a band in the blood vessels area similar to the one employed in the iris recognition as shown in Figure 2.

The scanned area is a circular band around blood vessels. This contrast information of this area is processed via fast Fourier transform. The transformed data forms the final biometric pattern

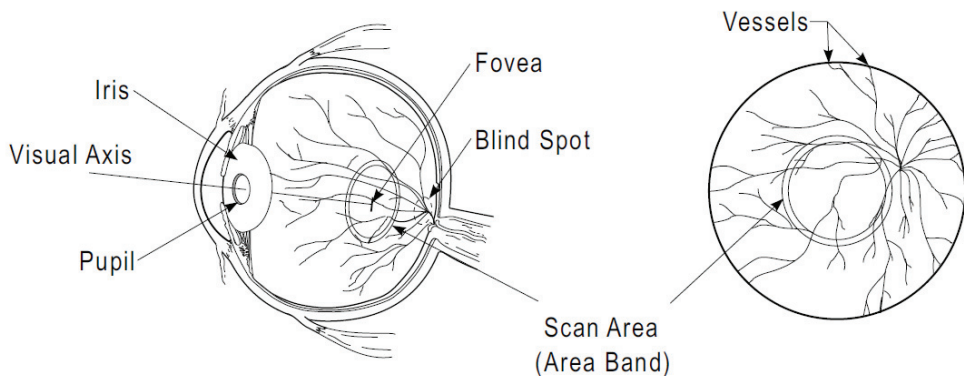


Fig. 2. Illustration of the scan area in the retina used in the system of Robert Hill.

considered in this system. This pattern worked good enough as the acquisition environment was very controlled. Of course, this is also the source of the major drawbacks present in the device: the data acquisition process. This process was both slow and uncomfortable for the user. Moreover, the hardware was very expensive and, therefore, it rendered the system hardly appealing. Finally, the result was that the use of retinal pattern as a biometric characteristic, despite all its convenient properties, was discontinued.

Nowadays, retinal image cameras (Figure 3) are capable of taking a photograph of the retina area in a human eye without any intrusive or dangerous scanning. Also, currently, the devices are cheaper and more accessible in general. This technology reduces the perception of danger by the user during the retina acquisition process but also brings more freedom producing a more heterogeneous type of retinal images to work with. The lighting conditions and the movement of the user's eye vary between acquisitions. This produces as a result that previous systems based on contrast information of reduced areas may lack the required precision in some cases, increasing the false rejection rate.



Fig. 3. Two retinal image cameras. The retinal image is acquired by taking an instant photograph.

In Figure 4 it can be observed two images from the same person acquired at different times by the same retinograph. There are some zones in the retinal vessels that can not be compared

because of the lack of information in one of the images. Thus, to allow the retinal biometrics to keep and increase the acquisition comfortability, it is necessary to implement a more robust methodology that, maintaining the extremely low error rates, is capable to cope with a more heterogeneous range of retinal images.

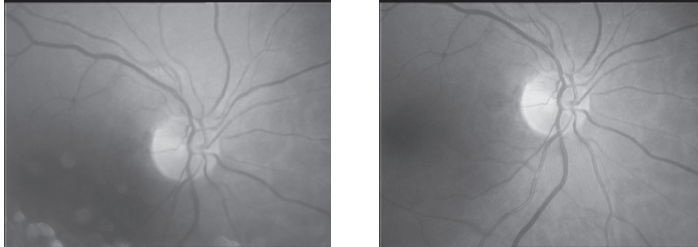


Fig. 4. Example of two digital retina images from the same individual acquired by the same retinal camera at different times.

This work is focused on the proposal of a novel personal authentication system based on the retinal vessel tree. This system deals with the new challenges in the retinal field where a more robust pattern has to be designed in order to increase the usability for the acquisition stage. In this sense, the approach presented here to the retinal recognition is closer to the fingerprint developments than to the iris ones as the own structure of the retinal vessel tree suggests. Briefly, the objectives of this work are enumerated:

- Empirical evaluation of the retinal vessel tree as biometric pattern
- Design a robust, easy to store and process biometric pattern making use of the whole retinal vessel tree information
- Development of an efficient and effective methodology to compare and match such retinal patterns
- Analysis on similarity metrics performance to establish reliable thresholds in the authentication process

To deal with the suggested goals, the rest of this document is organized as follows. Second section introduces previous works and research on the retinal vessel tree as biometric pattern. *Section 3* presents the methodology developed to build the authentication system, including biometric template construction and template matching algorithms. *Section 4* discusses the experiments aimed to test the proposed methodologies, including an analysis of similarity measures. Finally, *Section 5* offers some conclusions and final discussion.

2. Related work

Awareness of the uniqueness of the retinal vascular pattern dates back to 1935 when two ophthalmologists, Drs. Carleton Simon and Isodore Goldstein, while studying eye disease, realized that every eye has its own unique pattern of blood vessels. They subsequently published a paper on the use of retinal photographs for identifying people based on their blood vessel patterns Simon & Goldstein (1935). Later in the 1950s, their conclusions were supported by Dr. Paul Tower in the course of his study of identical twins. He noted that, of any two persons, identical twins would be the most likely to have similar retinal vascular

patterns. However, Tower showed that, of all the factors compared between twins, retinal vascular patterns showed the least similarities Tower (1955).

Blood vessels are among the first organs to develop and are entirely derived from the mesoderm. Vascular development occurs via two processes termed vasculogenesis and angiogenesis. Vasculogenesis, this is, the blood vessel assembly during embryogenesis, begins with the clustering of primitive vascular cells or hemangioblasts into tube-like endothelial structures, which define the pattern of the vasculature. In angiogenesis, new vessels arise by sprouting of budlike and fine endothelial extensions from preexisting vessels Noden (1989).

In a more recent study Whittier et al. (2003), retinal vascular pattern images from livestock were digitally acquired in order to evaluate their pattern uniqueness. To evaluate each retinal vessel pattern, the dominate trunk vessel of bovine retinal images was positioned vertically and branches on the right and left of the trunk and other branching points were evaluated. Branches from the left (mean 6.4 and variance 2.2) and the right (mean 6.4 and variance 1.5) of the vascular trunk; total branches from the vascular trunk (mean 12.8 and variance 4.3), and total branching points (mean 20.0 and variance 13.2) showed differences across all animals (52). A paired comparison of the retinal vessel patterns from both eyes of 30 other animals confirmed that eyes from the same animal differ. Retinal images of 4 cloned sheep from the same parent line were evaluated to confirm the uniqueness of the retinal vessel patterns in genetically identical animals. This would be confirming the uniqueness of animal retinal vascular pattern suggested earlier in the 1980s also by De Schaepe-drijver et al. (1989).

In general, retinal vessel tree his is a unique pattern in each individual and it is almost impossible to forge that pattern in a false individual. Of course, the pattern does not change through the individual's life, unless a serious pathology appears in the eye. Most common diseases like diabetes do not change the pattern in a way that its topology is affected. Some lesions (points or small regions) can appear but they are easily avoided in the vessels extraction method that will be discussed later. Thus, retinal vessel tree pattern has been proved a valid biometric trait for personal authentication as it is unique, time invariant and very hard to forge, as showed by Mariño et al. C.Mariño et al. (2003); Mariño et al. (2006), who introduced a novel authentication system based on this trait. In that work, the whole arterial-venous tree structure was used as the feature pattern for individuals. The results showed a high confidence band in the authentication process but the database included only 6 individuals with 2 images for each of them. One of the weak points of the proposed system was the necessity of storing and handling a whole image as the biometric pattern. This greatly difficults the storing of the pattern in databases and even in different devices with memory restrictions like cards or mobile devices. In Farzin et al. (2008) a pattern is defined using the optic disc as reference structure and using multi scale analysis to compute a feature vector around it. Good results were obtained using an artificial scenario created by randomly rotating one image per user for different users. The dataset size is 60 images, rotated 5 times each. The performance of the system is about a 99% accuracy. However, the experimental results do not offer error measures in a real case scenario where different images from the same individual are compared.

Based on the idea of fingerprint minutiae, a robust pattern is introduced where a set of landmarks (bifurcations and crossovers of retinal vessel tree) were extracted and used as feature points. In this scenario, the pattern matching problem is reduced to a point pattern matching problem and the similarity metric has to be defined in terms of matched points. A common problem in previous approaches is that the optic disc is used as a reference structure in the image. The detection of the optic disc is a complex problem and in some individuals

with eye diseases this cannot be achieved correctly. In this work, the use of reference structures is avoided to allow the system to cope with a wider range of images and users.

3. Retinal verification based on feature points

Figure 5 illustrates the general schema for the new feature point based authentication approach. The newly introduced stages are the feature point extraction and the feature point matching. The following chapter sections will discuss the methodology on these new stages of the system.

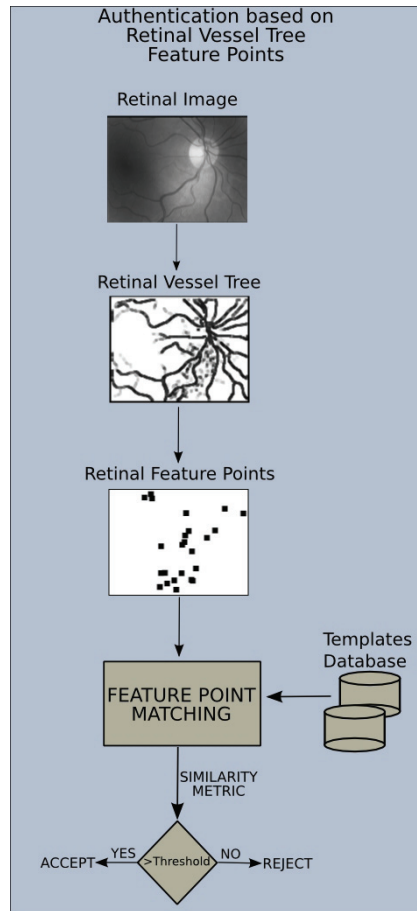


Fig. 5. Schema of the main stages for the authentication system based in the retinal vessel tree structure.

3.1 Feature points extraction

Following the idea that vessels can be thought of as creases (ridges or valleys) when images are seen as landscapes (see Figure 6), curvature level curves will be used to calculate the creases

(ridge and valley lines). Several methods for crease detection have been proposed in the literature (see López et al. (1999) for a comparison between methods), but finally a differential geometry based method López et al. (2000) was selected because of its good performance in similar images Lloret et al. (1999; 2001), producing very good results.

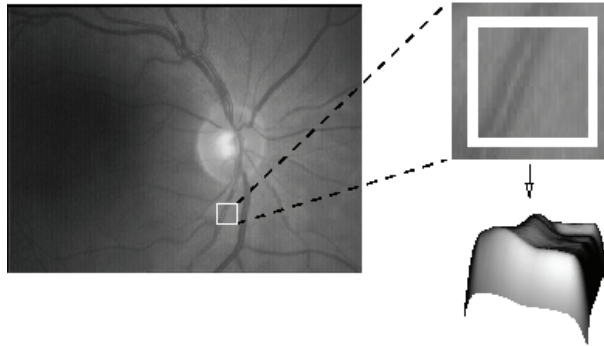


Fig. 6. Picture of a region of the retinal image as landscape. Vessels can be represented as creases.

Among the many definitions of crease, the one based on Level Set Extrinsic Curvature, LSEC López et al. (1998), has useful invariance properties. The geometry based method named LSEC gives rise to several problems, solved through the improvement of this method by a multilocal solution, the MLSEC López et al. (2000). But results obtained with MLSEC can still be improved by pre-filtering the image gradient vector field using structure tensor analysis and by discarding creaseness at isotropic areas by means of the computation of a confidence measure. The methodology allows to tune several parameters to apply such filters as for creases with a concrete width range or crease length. In Caderno et al. (2004) a methodology was presented for automatic parameter tuning by analyzing contrast variance in the retinal image.

One of the main advantages of this method is that it is invariant to changes in contrast and illumination, allowing the extraction of creases from arteries and veins independently of the characteristics of the images, avoiding a previous normalization of the input images. The final result is an image where the retinal vessel tree is represented by its crease lines. Figure 7 shows several examples of the creases obtained from different retinal images.

The landmarks of interest are points where two different vessels are connected. Therefore, it is necessary to study the existing relationships between vessels in the image. The first step is to track and label the vessels to be able to establish their relationships between them.

In Figure 8, it can be observed that the crease images show discontinuities in the crossovers and bifurcations points. This occurs because of the two different vessels (valleys or ridges) coming together into a region where the crease direction can not be set. Moreover, due to some illumination or intensity loss issues, the crease images can also show some discontinuities along a vessel (Figure 8). This issue requires a process of joining segments to build the whole vessels prior to the bifurcation/crossover analysis.

Once the relationships between segments are established, a final stage will take place to remove some possible spurious feature points. Thus, the four main stages in the feature point extraction process are:

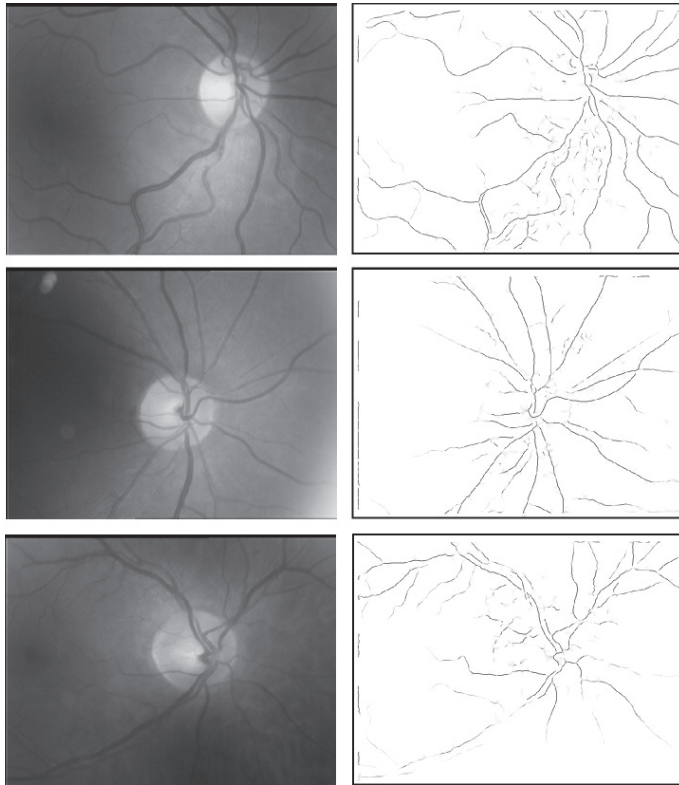


Fig. 7. Three examples of digital retinal images, showing the variability of the vessel tree among individuals. Left column: input images. Right column: creases of images on the left column representing the main vessels.

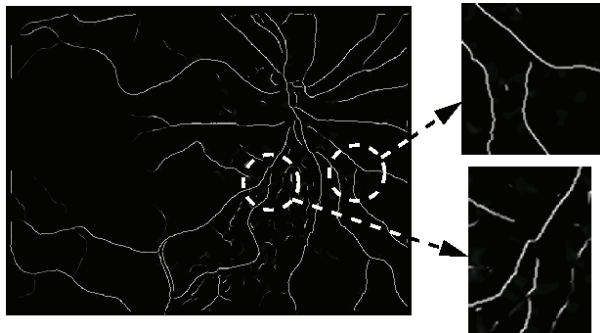


Fig. 8. Example of discontinuities in the creases of the retinal vessels. Discontinuities in bifurcations and crossovers are due to two creases with different directions joining in the same region. But, also, some other discontinuities along a vessel can happen due to illumination and contrast variations in the image.

1. Labelling of the vessels segments
2. Establishing the joint or union relationships between vessels
3. Establishing crossover and bifurcation relationships between vessels
4. Filtering of the crossovers and bifurcations

3.1.1 Tracking and labelling of vessel segments

To detect and label the vessel segments, an image tracking process is performed. As the crease images eliminate background information, any non-null pixel (intensity greater than zero) belongs to a vessel segment. Taking this into account, each row in the image is tracked (from top to bottom) and when a non-null pixel is found, the segment tracking process takes place. The aim is to label the vessel segment found, as a line of 1 pixel width. This is, every pixel will have only two neighbors (previous and next) avoiding ambiguity to track the resulting segment in further processes.

To start the tracking process, the configuration of the 4 pixels which have not been analyzed by the initially detected pixel is calculated. This leads to 16 possible configurations depending on whether there is a segment pixel or not in each one of the 4 positions. If the initial pixel has no neighbors, it is discarded and the image tracking continues. In the other cases there are two main possibilities: either the initial pixel is an endpoint for the segment, so this is tracked in one way only or the initial pixel is a middle point and the segment is tracked in two ways from it. Figure 9 shows the 16 possible neighborhood configurations and how the tracking directions are established in any case.

Once the segment tracking process has started, in every step a neighbor of the last pixel flagged as segment is selected to be the next. This choice is made using the following criterion: the best neighbor is the one with the most non-flagged neighbors corresponding to segment pixels. This heuristic contains the idea of keeping the 1-pixel width segment to track along the middle of the crease, where pixels have more segment pixel neighbors. In case of a tie, the heuristic tries to preserve the most repeated orientation in the last steps. When the whole image tracking process finishes, every segment is a 1 pixel width line with its endpoints defined. The endpoints are very useful to establish relationships between segments because these relationships can always be detected in the surroundings of a segment endpoint. This avoids the analysis of every pixel belonging to a vessel, considerably reducing the complexity of the algorithm and therefore the running time.

3.1.2 Union relationships

As stated before, the union detection is needed to build the vessels out of their segments. Aside the segments from the crease image, no additional information is required and therefore is the first kind of relationship to be detected in the image. An union or joint between two segments exists when one of the segments is the continuation of the other in the same retinal vessel. Figure 10 shows some examples of union relationships between segments.

To find these relationships, the developed algorithm uses the segment endpoints calculated and labelled in the previous subsection. The main idea is to analyze pairs of close endpoints from different segments and quantify the likelihood of one being the prolongation of the other. The proposed algorithm connects both endpoints and measures the smoothness of the connection.

An efficient approach to connect the segments is using an straight line between both endpoints. In Figure 11, a graphical description of the detection process for an union is

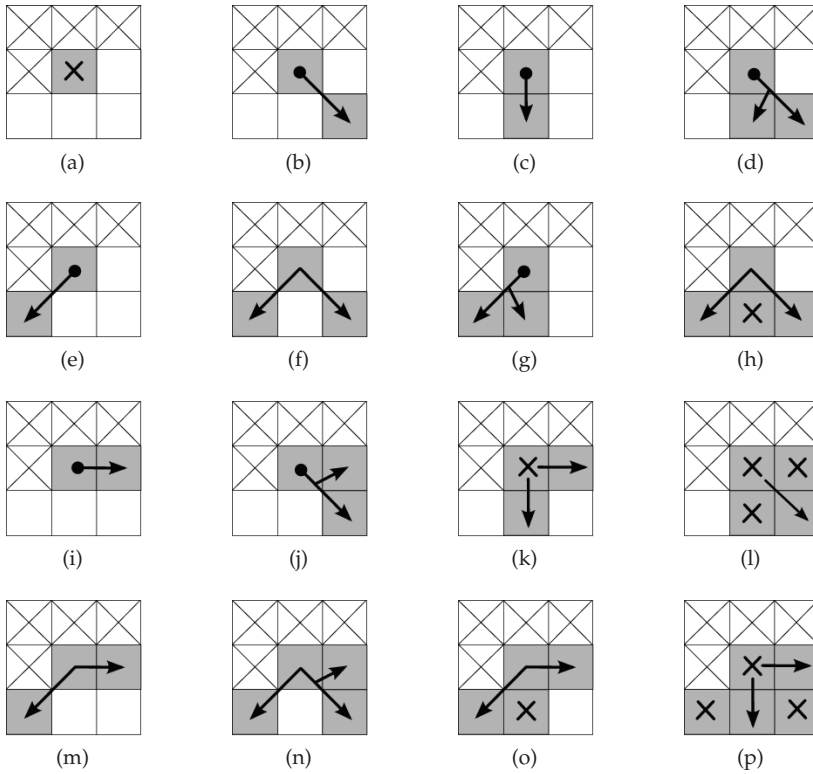


Fig. 9. Initial tracking process for a segment depending on the neighbor pixels surrounding the first pixel found for the new segment in a 8-neighborhood. As there are 4 neighbors not tracked yet (the bottom row and the one to the right), there are a total of 16 possible configurations. Gray squares represent crease (vessel) pixels and the white ones, background pixels. The upper row neighbors and the left one are ignored as they have already been tracked due to the image tracking direction. Arrows point to the next pixels to track while crosses flag pixels to be ignored. In 9(d), 9(g), 9(j) and 9(n) the forked arrows mean that only the best of the pointed pixels (i.e. the one with more new vessel pixel neighbors) is selected for continuing the tracking. Arrows starting with a black circle flag the central pixel as an endpoint for the segment (9(b), 9(c), 9(d), 9(e), 9(g), 9(i), 9(j)).

shown. The smoothness measurement is obtained from the angles between the straight line and the segment direction. The segment direction is calculated by the endpoint direction. The maximum smoothness occurs when both angles are π rad., i.e. both segments are parallel and belong to the straight line connecting it. The smoothness decreases as both angles decrease. A criterion to accept the candidate relationship must be established. A minimum angle θ_{min} is set as the threshold for both angles. This way, the criterion to accept an union relationship is defined as

$$Union(r, s) = (\alpha > \theta_{min}) \wedge (\beta > \theta_{min}) \tag{1}$$

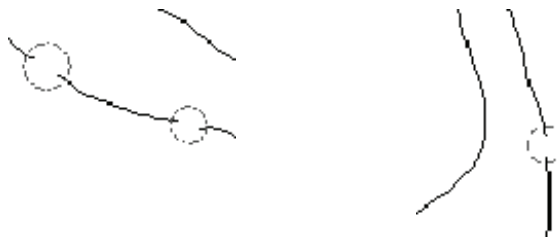


Fig. 10. Examples of union relationships. Some of the vessels present discontinuities leading to different segments. These discontinuities are detected in the union relationships detection process.

where r, s are the segments involved in the union and α, β their respective endpoint directions. It has been observed that for values of θ_{min} close to $\frac{3}{4}\pi$ rad. the algorithm delivers good results in all cases.

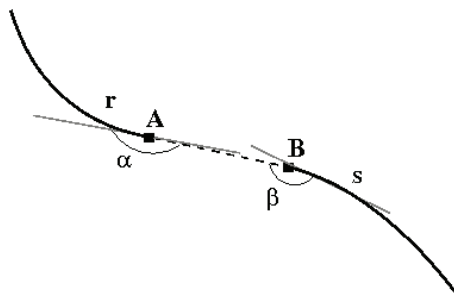


Fig. 11. Union of the crease segments r and s . The angles between the new segment \overline{AB} and the crease segments r (α) and s (β) are near π rad, so they are above the required threshold ($\frac{3}{4}\pi$) and the union is finally accepted.

3.1.3 Bifurcation/crossover relationships

Bifurcations and crossovers are the feature interest points in this work for characterizing individuals by a biometric pattern. A crossover is an intersection between two segments. A bifurcation is a point in a segment where another one starts from. While unions allow to build the vessels, bifurcations allow to build the vessel tree by establishing relationships between them. Using both types, the retinal vessel tree can be reconstructed by joining all segments. An example of this is shown in Figure 12.

A crossover can be seen in the segment image, as two close bifurcations forking from the same segment. Therefore, finding bifurcation and crossover relationships between segments can be initially reduced to find only bifurcations. Crossovers can then be detected analyzing close bifurcations.

In order to find bifurcations in the image, an idea similar to the union algorithm is followed based on the search of the bifurcations from the segments endpoints. The criterion in this case is finding a segment close to an endpoint whose segment can be assumed to start in the found one. This way, the algorithm does not require to track the whole segments, bounding complexity to the number of segments and not to their length.

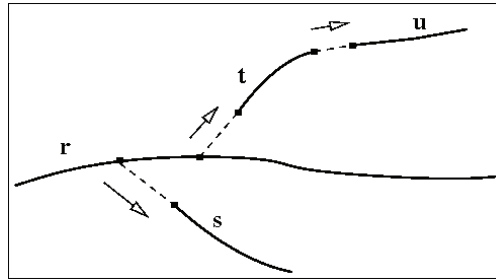


Fig. 12. Retinal Vessel Tree reconstruction by unions (t, u) and bifurcations (r, s) and (r, t) .

For every endpoint in the image, the process is as follows (Figure 13):

1. Compute the endpoint direction.
2. Extend the segment in that direction a fixed length l_{max} .
3. Analyze the points in and nearby the prolongation segment to find candidate segments.
4. If a point of a different segment is found, compute the angle (α) associated to that bifurcation, defined by the direction of this point and the extreme direction from step 1.

The parameter l_{max} is inserted in the model to avoid indefinite prolongation of the segments. If it follows that $l \leq l_{max}$, the segments will be joined and a bifurcation will be detected, being l the distance from the endpoint of the segment to the other segment.

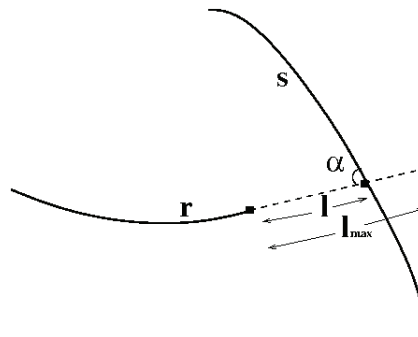


Fig. 13. Bifurcation between segment r and s . The endpoint of r is prolonged a maximum distance l_{max} and eventually a point of segment s is found.

Figure 14 shows an example of results after this stage where feature points are marked. Also, spurious detected points are identified in the image. These spurious points may occur for different reasons such as wrongly detected segments. In the image test set used (over 100 images) the approximate mean number of feature points detected per image was 28. The mean of spurious points corresponded to 5 points per image. To improve the performance of the matching process is convenient to eliminate as spurious points as possible. Thus, the last stage in the biometric pattern extraction process will be the filtering of spurious points in order to obtain an accurate biometric pattern for an individual.

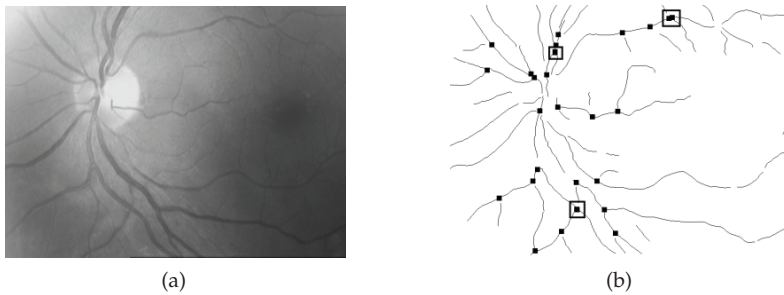


Fig. 14. Example of feature points extracted from original image after the bifurcation/crossover stage. (a) Original Image. (b) Feature points marked over the segment image. Spurious points corresponding to the same crossover (detected as two bifurcations) are signalled in squares.

3.1.4 Filtering of feature points

A segment filtering process takes place in the tracking stage, filtering detected segments by their length using a threshold, T_{min} . This leads to images with minimum false segments and with only important segments in the vessel tree.

Finally, since crossover points are detected as two bifurcation points, as Figure 14(b) shows, these bifurcation points are merged into an unique feature point by calculating the midpoint between them.

Figure 15 shows an example of the filtering process result, i.e. the biometric pattern obtained from an individual. Briefly, in the initial test set of images used to tune the parameters, the reduction of false detected points was about from 5 to 2 in the average.

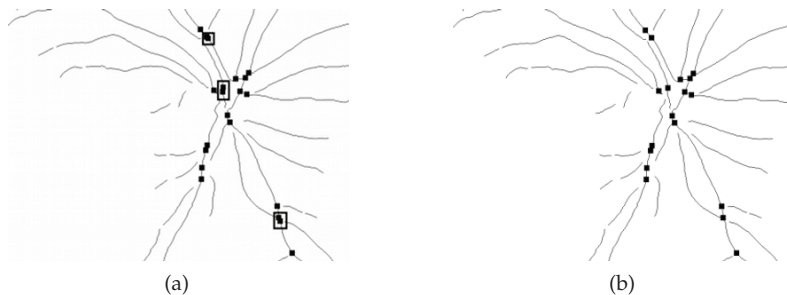


Fig. 15. Example of the result after the feature point filtering. (a) Image containing feature points before filtering. (b) Image containing feature points after filtering. Spurious points from duplicate crossover points have been eliminated.

3.2 Biometric pattern matching

In the matching stage, the stored reference pattern, ν , for the claimed identity is compared to the pattern extracted, ν' , during the previous stage. Due to the eye movement during the image acquisition stage, it is necessary to align ν' with ν in order to be matched L.G.Brown (1992); M.S.Markov et al. (1993); Zitová & Flusser (2003). This fact is illustrated in Figure 16

where two images from the same individual, 16(a) and 16(c), and the obtained results in each case, 16(b) and 16(d), are shown using the crease approach.

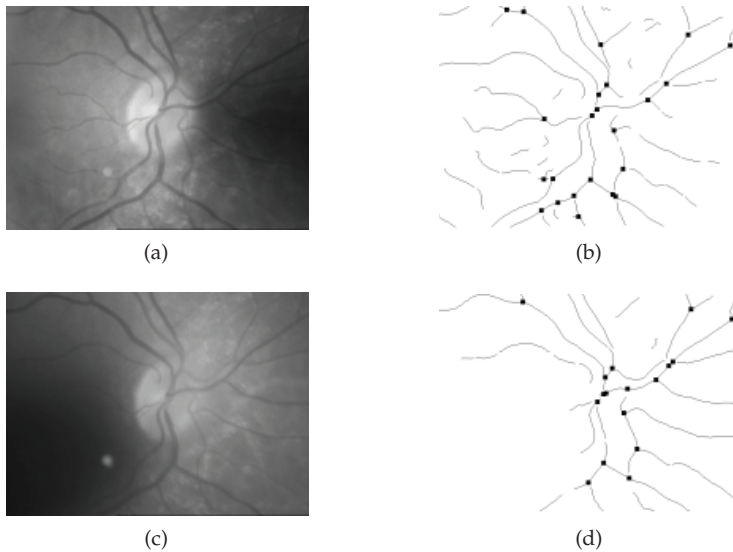


Fig. 16. Examples of feature points obtained from images of the same individual acquired in different times. (a) and (c) original images. (b) Feature point image from (a). A set of 23 points is obtained. (d) Feature point image from (c). A set of 17 points are obtained.

Depending on several factors, such as the eye location in the objective, patterns may suffer some deformations. A reliable and efficient model is necessary to deal with these deformations allowing to transform the candidate pattern in order to get a pattern similar to the reference one. The movement of the eye in the image acquisition process basically consists in translation in both axis, rotation and sometimes a very small change in scale. It is also important to note that both patterns v and v' could have a different number of points even being from the same individual. This is due to the different conditions of illumination and orientation in the image acquisition stage.

The transformation considered in this work is the Similarity Transformation (ST), which is a special case of the Global Affine Transformation (GAT). ST can model translation, rotation and isotropic scaling using 4 parameters Ryan et al. (2004). The ST works fine with this kind of images as the rotation angle is moderate. It has also been observed that the scaling, due to eye proximity to the camera, is nearly constant for all the images. Also, the rotations are very slight as the eye orientation when facing the camera is very similar. Under these circumstances, the ST model appears to be very suitable.

The ultimate goal is to achieve a final value indicating the similarity between the two feature points set, in order to decide about the acceptance or the rejection of the hypothesis that both images correspond to the same individual. To develop this task the matching pairings between both images must be determined. A transformation has to be applied to the candidate image in order to register its feature points with respect to the corresponding points in the reference image. The set of possible transformations is built based on some restrictions and

a matching process is performed for each one of these. The transformation with the highest matching score will be accepted as the best transformation.

To obtain the four parameters of a concrete ST, two pairs of feature points between the reference and candidate patterns are considered. If M is the total number of feature points in the reference pattern and N the total number of points in the candidate one, the size of the set T of possible transformations is computed using Eq.(2):

$$T = \frac{(M^2 - M)(N^2 - N)}{2} \quad (2)$$

where M and N represent the cardinality of ν and ν' respectively.

Since T represents a high number of transformations, some restrictions must be applied in order to reduce it. As the scale factor between patterns is always very small in this acquisition process, a constraint can be set to the pairs of points to be associated. In this scenario, the distance between both points in each pattern has to be very similar. As it cannot be assumed that it will be the same, two thresholds are defined, S_{min} and S_{max} , to bound the scale factor. This way, elements from T are removed where the scale factor is greater or lower than the respective thresholds S_{min} and S_{max} . Eq.(3) formalises this restriction:

$$S_{min} < \frac{distance(p, q)}{distance(p', q')} < S_{max} \quad (3)$$

where p, q are points from ν pattern, and p', q' are the matched points from the ν pattern. Using this technique, the number of possible matches greatly decrease and, in consequence, the set of possible transformations decreases accordingly. The mean percentage of not considered transformations by these restrictions is around 70%.

In order to check feature points, a similarity value between points (SIM) is defined which indicates how similar two points are. The distance between these two points will be used to compute that value. For two points A and B , their similarity value is defined by Eq.(4):

$$SIM(A, B) = 1 - \frac{distance(A, B)}{D_{max}} \quad (4)$$

where D_{max} is a threshold that stands for the maximum distance allowed for those points to be considered a possible match. If $distance(A, B) > D_{max}$ then $SIM(A, B) = 0$. D_{max} is a threshold introduced in order to consider the quality loss and discontinuities during the creases extraction process leading to mislocation of feature points by some pixels.

In some cases, two points B_1, B_2 could have both a good value of similarity with one point A in the reference pattern. This happens because B_1 and B_2 are close to each other in the candidate pattern. To identify the most suitable matching pair, the possibility of correspondence is defined comparing the similarity value between those points to the rest of similarity values of each one of them:

$$P(A_i, B_j) = \frac{SIM(A_i, B_j)^2}{\left(\sum_{i'=1}^M SIM(A_{i'}, B_j) + \sum_{j'=1}^N SIM(A_i, B_{j'}) - SIM(A_i, B_j) \right)} \quad (5)$$

A $M \times N$ matrix Q is constructed such that position (i, j) holds $P(A_i, B_j)$. Note that if the similarity value is 0, the possibility value is also 0. This means that only valid matchings will have a non-zero value in Q . The desired set C of matching feature points is obtained from P using a greedy algorithm. The element (i, j) inserted in C is the position in Q where the maximum value is stored. Then, to prevent the selection of the same point in one of the images again, the row (i) and the column (j) associated to that pair are set to 0. The algorithm finishes when no more non-zero elements can be selected from Q .

The final set of matched points between patterns is C . Using this information, a similarity metric must be established to obtain a final criterion of comparison between patterns.

4. Similarity metrics analysis

The goal in this stage of the process is to define similarity measures on the aligned patterns to correctly classify authentications in both classes: attacks (unauthorised accesses), when the two matched patterns are from different individuals and clients (authorised accesses) when both patterns belong to the same person.

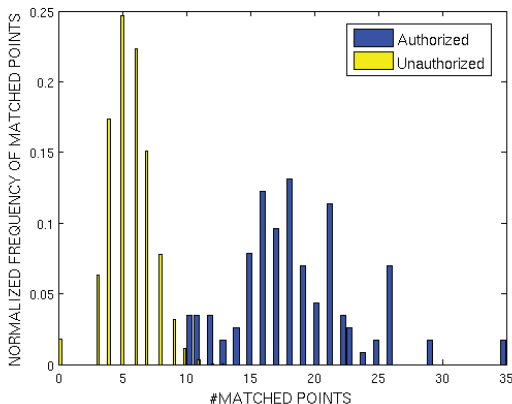
For the metric analysis a set of 150 images (100 images, 2 images per individual and 50 different images more) from VARIA database VARIA (2007) were used. The rest of the images will be used for testing in the next section. The images from the database have been acquired with a TopCon non-mydratic camera NW-100 model and are optic disc centred with a resolution of 768x584. There are 60 individuals with two or more images acquired in a time span of 6 years. These images have a high variability in contrast and illumination allowing the system to be tested in quite hard conditions. In order to build the training set of matchings, all images are matched versus all the images (a total of 150x150 matchings) for each metric. The matchings are classified into attacks or clients accesses depending if the images belong to the same individual or not. Distributions of similarity values for both classes are compared in order to analyse the classification capabilities of the metrics.

The main information to measure similarity between two patterns is the number of feature points successfully matched between them. Fig.17(a) shows the histogram of matched points for both classes of authentications in the training set. As it can be observed, matched points information is by itself quite significative but insufficient to completely separate both populations as in the interval [10, 13] there is an overlapping between them.

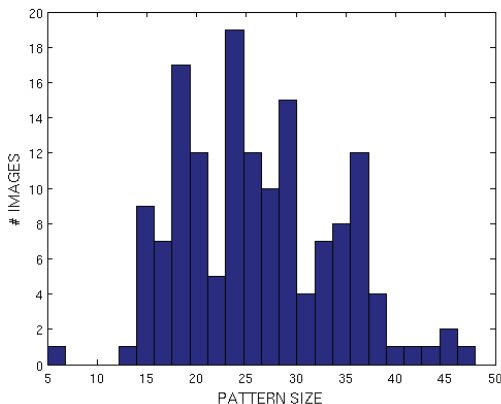
This overlapping is caused by the variability of the patterns size in the training set because of the different illumination and contrast conditions in the acquisition stage. Fig.17(b) shows the histogram for the biometric pattern size, i.e. the number of feature points detected. A high variability can be observed, as some patterns have more than twice the number of feature points of other patterns. As a result of this, some patterns have a small size, capping the possible number of matched points (Fig. 18). Also, using the matched points information alone lacks a well bounded and normalised metric space.

To combine information of patterns size and normalise the metric, a function f will be used. Normalised metrics are very common as they make easier to compare class separability or establishing valid thresholds. The similarity measure (S) between two patterns will be defined by

$$S = \frac{C}{f(M, N)} \quad (6)$$



(a)



(b)

Fig. 17. (a) Matched points histogram in the attacks (unauthorised) and clients (authorised) authentications cases. In the interval [10, 13] both distributions overlap. (b) histogram of detected points for the patterns extracted from the training set.

where C is the number of matched points between patterns, and M and N are the matching patterns sizes. The first f function defined and tested is:

$$f(M, N) = \min(M, N) \tag{7}$$

The \min function is the less conservative as it allows to obtain a maximum similarity even in cases of different sized patterns. Fig.19(a) shows the distributions of similarity scores for clients and attacks classes in the training set using the normalisation function defined in Eq.(7), and Fig.19(b) shows the FAR and FRR curves versus the decision threshold.

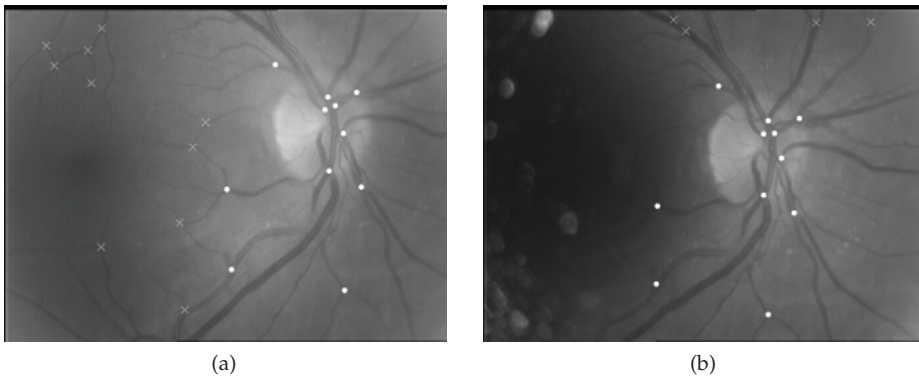


Fig. 18. Example of matching between two samples from the same individual in VARIA database. White circles mark the matched points between both images while crosses mark the unmatched points. In (b) the illumination conditions of the image lead to miss some features from left region of the image. Therefore, a small amount of detected feature points is obtained capping the total amount of matched points.

Although the results are good when using the normalisation function defined in Eq.(7), a few cases of attacks show high similarity values, overlapping with the clients class. This is caused by matchings involving patterns with a low number of feature points as $\min(M, N)$ will be very small, needing only a few points to match in order to get a high similarity value. This suggests, as it will be reviewed in section 5, that some minimum quality constraint in terms of detected points would improve performance for this metric.

To improve the class separability, a new normalisation function f is defined:

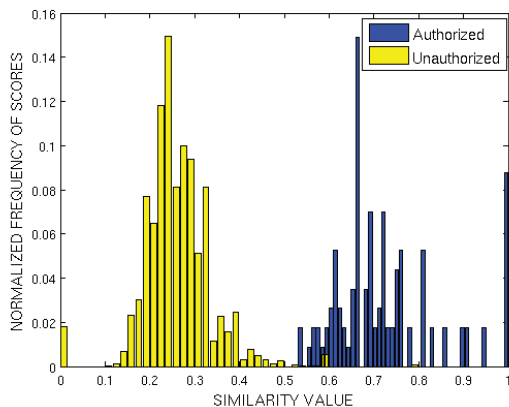
$$f(M, N) = \sqrt{MN} \quad (8)$$

Fig.20(a) shows the distributions of similarity scores for clients and attacks classes in the training set using the normalisation function defined in Eq.(8) and Fig.20(b) shows the FAR and FRR curves versus the decision threshold.

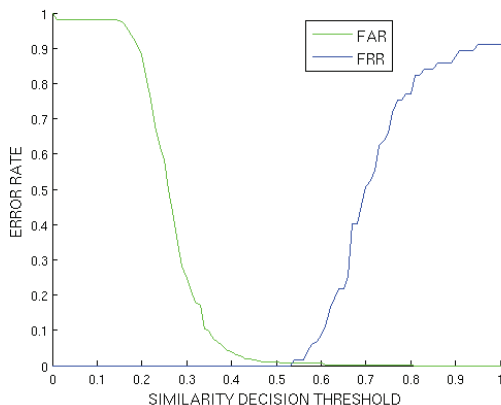
Function defined in Eq.(8) combines both patterns size in a more conservative way, preventing the system to obtain a high similarity value if one pattern in the matching process contains a low number of points. This allows to reduce the attacks class variability and, moreover, to separate its values away from the clients class as this class remains in a similar values range. As a result of the new attacks class boundaries, a decision threshold can be safely established where $FAR = FRR = 0$ in the interval $[0.38, 0.5]$ as Fig.20(b) clearly exposes. Although this metric shows good results, it also has some issues due to the normalisation process which can be corrected to improve the results as showed in next subsection.

4.1 Confidence band improvement

Normalising the metric has the side effect of reducing the similarity between patterns of the same individual where one of them had a much greater number of points than the other, even in cases with a high number of matched points. This means that some cases easily distinguishable based on the number of matched points are now near the confidence band borders. To take a closer look at this region surrounding the confidence band, the cases of



(a)



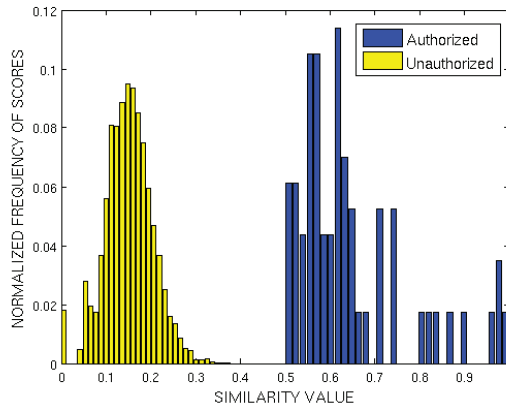
(b)

Fig. 19. (a) Similarity values distribution for authorised and unauthorised accesses using $f = \min(M, N)$ as normalisation function for the metric. (b) False Accept Rate (FAR) and False Rejection Rate (FRR) for the same metric.

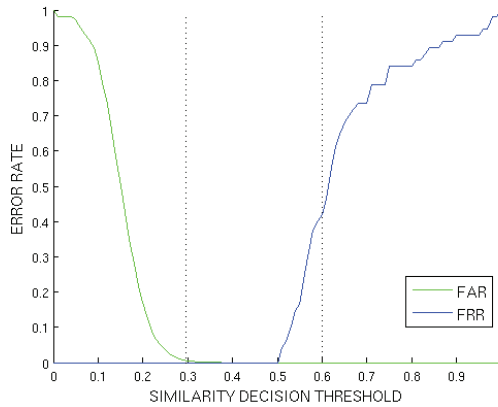
unauthorised accesses with the highest similarity values (S) and authorised accesses with the lowest ones are evaluated. Fig.21 shows the histogram of matched points for cases in the marked region of Fig.20(b). It can be observed that there is an overlapping but both histograms are highly distinguishable.

To correct this situation, the influence of the number of matched points and the patterns size have to be balanced. A correction parameter (γ) is introduced in the similarity measure to control this. The new metric is defined as:

$$S_\gamma = S \cdot C^{\gamma-1} = \frac{C^\gamma}{\sqrt{MN}} \tag{9}$$



(a)



(b)

Fig. 20. (a) similarity values distribution for authorised and unauthorised accesses using $f = \sqrt{MN}$ as normalisation function for the metric. (b) False Accept Rate (FAR) and False Rejection Rate (FRR) for the same metric. Dotted lines delimit the interest zone surrounding the confidence band which will be used for further analysis.

with S, C, M and N the same parameters from Eq.(8). The γ correction parameter allows to improve the similarity values when a high number of matched points is obtained, specially in cases of patterns with a high number of points.

Using the gamma parameter, values can be higher than 1. In order to normalise the metric back into a $[0, 1]$ values space, a sigmoid transference function, $T(x)$, is used:

$$T(x) = \frac{1}{1 + e^{S \cdot (x-0.5)}} \tag{10}$$

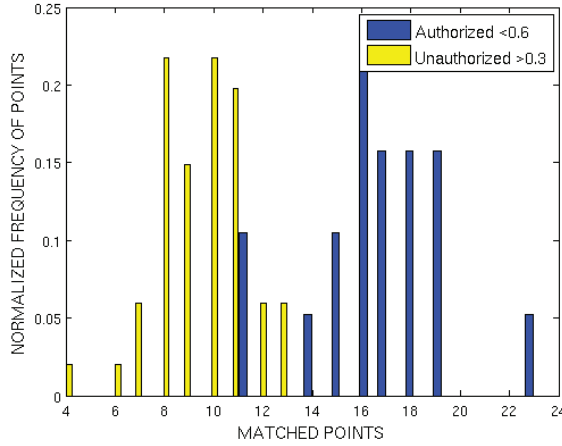


Fig. 21. Histogram of matched points in the populations of attacks whose similarity is higher than 0.3 and clients accesses whose similarity is lower than 0.6.

where s is a scale factor to adjust the function to the correct domain as S_γ does not return negatives or much higher than 1 values when a typical $\gamma \in [1, 2]$ is used. In this work, $s=6$ was chosen empirically. The normalised gamma-corrected metric, $S'_\gamma(x)$, is defined by:

$$S'_\gamma = T(S_\gamma) \tag{11}$$

Finally, to choose a good γ parameter, the confidence band improvement has been evaluated for different values of γ (Fig.22(a)). The maximum improvement is achieved at $\gamma = 1.12$ with a confidence band of 0.3288, much higher than the original from previous section. The distribution of the whole training set (using $\gamma = 1.12$) is showed in Fig.22(b) where the wide separation between classes can be observed.

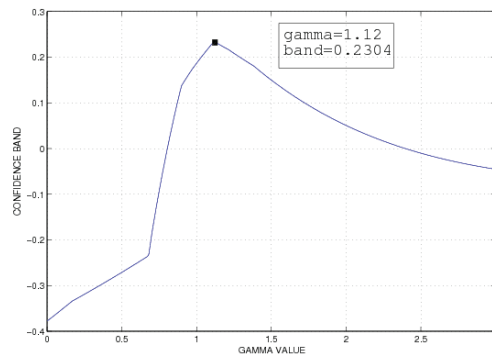
5. Results

A set of 90 images, 83 different from the training set and 7 from the previous set with the highest number of points, has been built in order to test the metrics performance once their parameters have been fixed with the training set. To test the metrics performance, the False Acceptance Rate and False Rejection Rate were calculated for each of them (the metrics normalised by Eq.(7), Eq.(8) and the gamma-corrected normalised metric defined in Eq.(11).

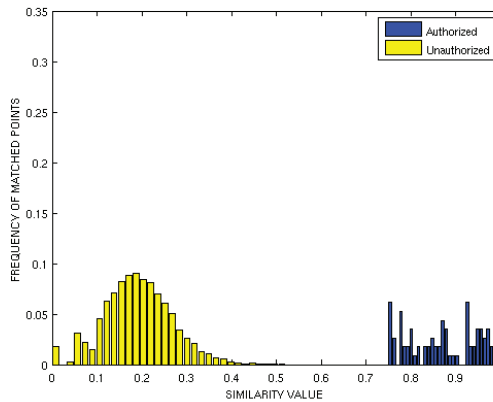
A usual error measure is the Equal Error Rate (EER) that indicates the error rate where FAR curve and FRR curve intersect. Fig.23(a) shows the FAR and FRR curves for the three previously specified metrics. The EER is 0 for the normalised by geometrical mean (MEAN) and gamma corrected (GAMMA) metrics as it was the same case in the training set, and, again, the gamma corrected metric shows the highest confidence band in the test set, 0.2337.

The establishment of a wide confidence band is specially important in this scenario of different images from users acquired on different times and with different configurations of the capture hardware.

Finally, to evaluate the influence of the image quality, in terms of feature points detected per image, a test is run where images with a biometric pattern size below a threshold are removed



(a)

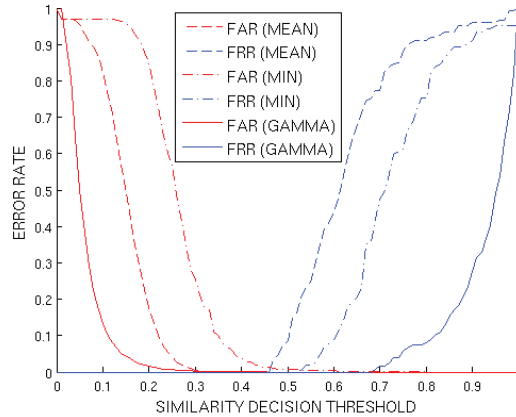


(b)

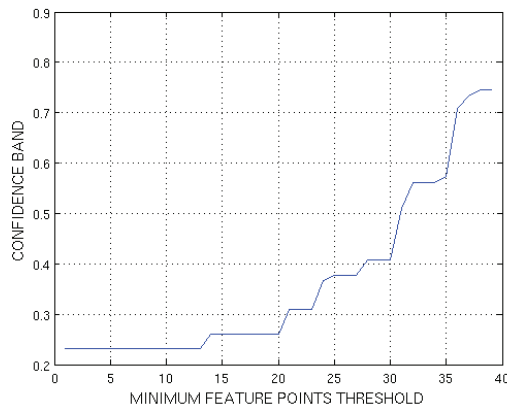
Fig. 22. (a) Confidence band size vs gamma (γ) parameter value. Maximum band is obtained at $\gamma = 1.12$. (b) Similarity values distributions using the normalised metric with $\gamma=1.12$.

for the set and the confidence band obtained with the rest of the images is evaluated. Fig.23(b) shows the evolution of the confidence band versus the minimum detected points constraint. The confidence band does not grow significantly until a fairly high threshold is set. Taking as threshold the mean value of detected points for all the test set, 25.2, the confidence band grows from 0.2337 to 0.3317. So removing half of the images, the band is increased only by 0.098 suggesting that the gamma-corrected metric is very robust to low quality images.

The mean execution time on a 2.4Ghz. Intel Core Duo desktop PC for the authentication process, implemented in C++, was 155ms: 105ms in the feature extraction stage and 50ms in the registration and similarity measure estimation, so that the method is very well-fitted to be employed in a real verification system.



(a)



(b)

Fig. 23. (a) FAR and FRR curves for the normalised similarity metrics (min: normalised by minimum points, mean: normalised by geometrical mean and gamma: gamma corrected metric). The best confidence band is the one belonging to the gamma corrected metric corresponding to 0.2337. (b) Evolution of the confidence band using a threshold of minimum detected points per pattern.

6. Conclusions and future work

In this work a complete identity verification method has been introduced. Following the same idea as the fingerprint minutiae-based methods, a set of feature points is extracted from digital retinal images. This unique pattern will allow for the reliable authentication of authorised users. To get the set of feature points, a creases-based extraction algorithm is used. After that, a recursive algorithm gets the point features by tracking the creases from the localised optic disc. Finally, a registration process is necessary in order to match the reference pattern from the database and the acquired one. With the patterns aligned, it is possible to measure the degree of similarity by means of a similarity metric. Normalised metrics have been defined and analysed in order to test the classification capabilities of the system. The results are very good and prove that the defined authentication process is suitable and reliable for the task. The use of feature points to characterise individuals is a robust biometric pattern allowing to define metrics that offer a good confidence band even in unconstrained environments when the image quality variance can be very high in terms of distortion, illumination or definition. This is also possible as this methodology does not rely on the localisation or segmentation of some reference structures, as it might be the optic disc. Thus, if the user suffers some structure distorting pathology and this structure cannot be detected, the system works the same with the only problem being a possible loss of feature points constrained to that region. Future work includes the use of some high-level information of points to complement metrics performance and new ways of codification of the biometric pattern allowing to perform faster matches.

7. References

- Bolle, R. M., Senior, A. W., Ratha, N. K. & Pankanti, S. (2002). Fingerprint minutiae: A constructive definition, *Biometric Authentication*, pp. 58–66.
- Caderno, I. G., Penedo, M. G., Mariño, C., Carreira, M. J., Gómez-Ulla, F. & González, F. (2004). Automatic extraction of the retina av index, *ICIAR (2)*, pp. 132–140.
- Chou, C.-T., Shih, S.-W. & Chen, D.-Y. (2006). Design of gabor filter banks for iris recognition, *IJH-MSP*, pp. 403–406.
- C.Mariño, M.G.Penedo, M.J.Carreira & F.Gonzalez (2003). Retinal angiography based authentication, *Lecture Notes in Computer Science 2905*: 306–313.
- De Schaepprijver, L., Simoens, L., Lauwers, H. & DeGesst, J. (1989). Retinal vascular patterns in domestic animals, *Res. Vet. Sci.* 47: 34–42.
- Farzin, H., Abrishami-Moghaddam, H. & Moin, M.-S. (2008). A novel retinal identification system, *EURASIP Journal on Advances in Signal Processing* ID 280635: 10 pp.
- He, Z., Sun, Z., Tan, T., Qiu, X., Zhong, C. & Dong, W. (2008). Boosting ordinal features for accurate and fast iris recognition, *CVPR*.
- Hill, R. (1999). Retina identification, in A. Jain, R. Bolle & S. Pankanti (eds), *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Press, Boston, pp. 123–142.
- Jain, A. K., Ross, A. & Pankanti, S. (1999). A prototype hand geometry-based verification system, *AVBPA*, pp. 166–171.
- Kim, H.-C., Kim, D., Bang, S. Y. & Lee, S.-Y. (2004). Face recognition using the second-order mixture-of-eigenfaces method, *Pattern Recognition* 37(2): 337–349.
- Kim, J., Cho, S., Choi, J. & Marks, R. J. (2004). Iris recognition using wavelet features, *VLSI Signal Processing* 38(2): 147–156.

- Kisku, D. R., Rattani, A., Tistarelli, M. & Gupta, P. (2008). Graph application on face for personal authentication and recognition, *ICARCV*, pp. 1150–1155.
- L.G.Brown (1992). A survey of image registration techniques, *ACM Computer Surveys* 24(4): 325–376.
- Lloret, D., López, A., Serrat, J. & Villanueva, J. (1999). Creaseness-based CT and MR registration: comparison with the mutual information method, *Journal of Electronic Imaging* 8(3): 255–262.
- Lloret, D., Mariño, C., Serrat, J., A.M.López & Villanueva, J. (2001). Landmark-based registration of full SLO video sequences, *Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis*, Vol. I, pp. 189–194.
- López, A., Lloret, D., Serrat, J. & Villanueva, J. (2000). Multilocal creaseness based on the level-set extrinsic curvature, *Computer Vision and Image Understanding* 77(1): 111–144.
- López, A., Lumbreras, F., Serrat, J. & Villanueva, J. (1999). Evaluation of methods for ridge and valley detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 21(4): 327–335.
- López, A. M., Lumbreras, F. & Serrat, J. (1998). Creaseness from level set extrinsic curvature, *ECCV*, pp. 156–169.
- Ma, L., Wang, Y. & Tan, T. (2002). Iris recognition using circular symmetric filters, *ICPR* (2), pp. 414–417.
- Maio, D. & Maltoni, D. (1997). Direct gray-scale minutiae detection in fingerprints, *IEEE Trans. Pattern Anal. Mach. Intell.* 19(1): 27–40.
- Mariño, C., Penedo, M. G., Penas, M., Carreira, M. J. & Gonzalez, F. (2006). Personal authentication using digital retinal images, *Pattern Analysis and Applications* 9: 21–33.
- Mian, A. S., Bennamoun, M. & Owens, R. A. (2008). Keypoint detection and local feature matching for textured 3d face recognition, *International Journal of Computer Vision* 79(1): 1–12.
- Moghaddam, B. & Pentland, A. (1997). Probabilistic visual learning for object representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7): 696–710.
- M.S.Markov, H.G.Rylander & A.J.Welch (1993). Real-time algorithm for retinal tracking, *IEEE Trans. on Biomedical Engineering* 40(12): 1269–1281.
- Nabti, M. & Bouridane, A. (2007). An improved iris recognition system using feature extraction based on wavelet maxima moment invariants, *ICB*, pp. 988–996.
- Noden, D. (1989). Embryonic origins and assembly of blood vessels, *Am. Rev. Respir. Dis.* 140: 1097–1103.
- Ryan, N., Heneghan, C. & de Chazal, P. (2004). Registration of digital retinal images using landmark correspondence by expectation maximization, *Image and Vision Computing* 22: 883–898.
- Seung-Hyun, L., Sang-Yi, Y. & Eun-Soo, K. (1995). Fingerprint identification by use of a volume holographic optical correlator, *Proc. SPIE Vol. 3715, Optical Pattern Recognition* pp. 321–325.
- Sidlauskas, D. (1988). 3d hand profile identification apparatus, United States Patent No.4,736,203.
- Siguenza Pizarro, J. A. & Tapiador Mateos, M. (2005). *Tecnologías Biométricas Aplicadas a Seguridad*, Ra–Ma.
- Simon, C. & Goldstein, I. (1935). A new scientific method of identification, *J. Medicine* 35(18): 901–906.
- systems Lab, B. (n.d.). Hasis, a hand shape identification system.

- Tower, P. (1955). The fundus oculi in monozygotic twins: Report of six pairs of identical twins, *Arch. Ophthalmol.* 54: 225–239.
- VARIA (2007). VARPA Retinal Images for Authentication. <http://www.varpa.es/varia.html>.
- Venkataramani, K. & Kumar, V. (2003). Fingerprint verification using correlation filters, *AVBPA*, pp. 886–894.
- Whittier, J. C., Doubet, J., Henrickson, D., Cobb, J., Shaddock, J. & Golden, B. (2003). Biological considerations pertaining to use of the retinal vascular pattern for permanent identification of livestock, *J. Anim. Sci* 81: 1–79.
- Yang, M.-H., Ahuja, N. & Kriegman, D. J. (2000). Face recognition using kernel eigenfaces, *ICIP*.
- Zitová, B. & Flusser, J. (2003). Image registration methods: a survey, *Image Vision and Computing* 21(11): 977–1000.
- Zunkel, R. (1999). Hand geometry based verification, *BIOMETRICS: Personal Identification in Networked Society*, Kluwert Academic Publishers.

DNA Biometrics

Masaki Hashiyada

*Division of Forensic Medicine, Department of Public Health and Forensic Medicine,
Tohoku University Graduate School of Medicine
Japan*

1. Introduction

The biometric authentication technologies, typified by fingerprint, face recognition and iris scanning, have been making rapid progress. Retinal scanning, voice dynamics and handwriting recognition are also being developed. These methods have been commercialized and are being incorporated into systems that require accurate on-site personal authentication. However, these methods are based on the measurement of similarity of feature-points. This introduces an element of inaccuracy that renders existing technologies unsuitable for a universal ID system. Among the various possible types of biometric personal identification system, deoxyribonucleic acid (DNA) provides the most reliable personal identification. It is intrinsically digital, and does not change during a person's life or after his/her death. This chapter addresses three questions: First, how can personally identifying information be obtained from DNA sequences in the human genome? Second, how can a personal ID be generated from DNA-based information? And finally, what are the advantages, deficiencies, and future potential for personal IDs generated from DNA data (DNA-ID)?

2. Human identification based on DNA polymorphism

A human body is composed of approximately of 60 trillion cells. DNA, which can be thought of as the blueprint for the design of the human body, is folded inside the nucleus of each cell. DNA is a polymer, and is composed of nucleotide units that each has three parts: a base, a sugar, and a phosphate. The bases are adenine, guanine, cytosine and thymine, abbreviated A, G, C and T, respectively. These four letters represent the informational content in each nucleotide unit; variations in the nucleotide sequence bring about biological diversity, not only among human beings but among all living creatures. Meanwhile, the phosphate and sugar portions form the backbone structure of the DNA molecule. Within a cell, DNA exists in the double-stranded form, in which two antiparallel strands spiral around each other in a double helix. The bases of each strand project into the core of the helix, where they pair with the bases of the complementary strand. A pairs strictly with T, and C with G (Alberts, 2002; Watson, 2004).

Within human cells, DNA found in the nucleus of the cell (nuclear DNA) is divided into chromosomes. The human genome consists of 22 matched pairs of autosomal chromosomes and two sex-determining chromosomes, X and Y. In other words, human cells contain 46 different chromosomes. Males are described as XY since they possess a single copy of the X

chromosome and a single copy of the Y chromosome, while females possess two copies of the X chromosome and are described as XX.

The regions of DNA that encode and regulate the synthesis of proteins are called genes; these regions consist of exons (protein-coding portions) and introns (the intervening sequences) and constitute approximately 25% of the genome (Jasinska & Krzyzosiak, 2004). The human genome contains only 20,000–25,000 genes (Collins et al., 2004; Lander et al., 2001; Venter et al., 2001). Therefore, most of the genome, approximately 75%, is extragenic. These regions are sometimes referred to as ‘junk’ DNA; however, recent research suggests that they may have other essential functions. Markers commonly used to identify individual human beings are usually found in the noncoding regions, either between genes or within genes (i.e., introns).

2.1 Sort tandem repeat (STR)

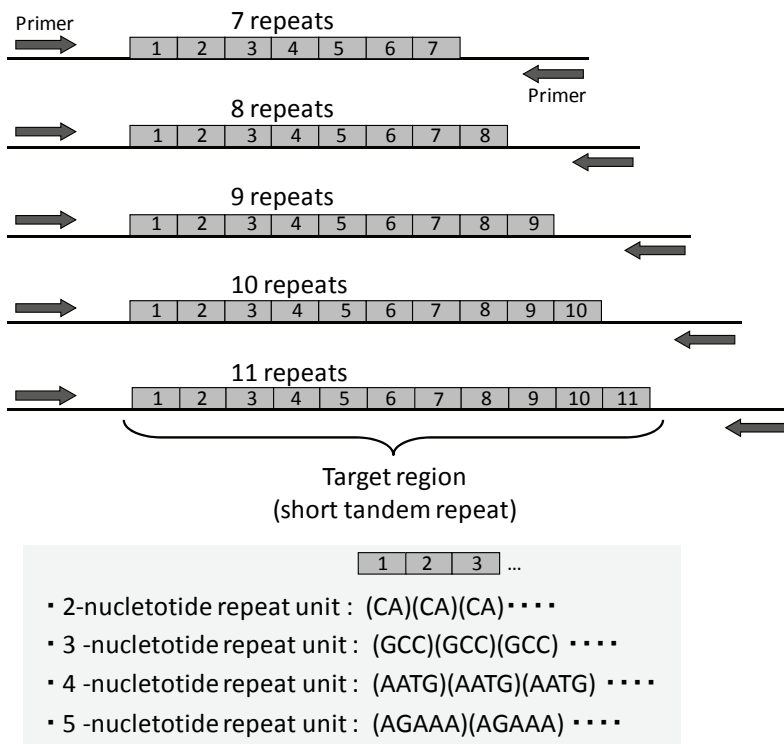


Fig. 1. The structure of Short Tandem Repeat (STR)

In the extragenic region of eukaryotic genome, there are many repeated DNA sequences (approximately 50% of the whole genome). These repeated DNA sequences come in all sizes, and are typically designated by the length of the core repeat unit and either the number of contiguous repeat units or the overall length of the repeat region. These regions are referred to as satellite DNA (Jeffreys et al., 1995). The core repeat unit for a medium-length repeat, referred to as a minisatellite or VNTR (variable number of tandem repeats), is in the range of approximately 8–100 bases in length (Jeffreys et al., 1985). DNA regions with

repeat units that are 2–7 base pairs (bp) in length are called microsatellites, simple sequence repeats (SSRs), or most commonly short tandem repeats (STRs) (Clayton et al., 1995; Hagelberg et al., 1991; Jeffreys et al., 1992) (Fig. 1). STRs have become popular DNA markers because they are easily amplified by the polymerase chain reaction (PCR) and they are spread throughout the genome, including both the 22 autosomal chromosomes and the X and Y sex chromosomes. The number of repeats in STR markers can vary widely among individuals, making the STRs an effective means of human identification in forensic science (Ruitberg et al., 2001). The location of an STR marker is called its “locus.” The type of STR is represented by the number of repeat called ‘allele’ which is taken from biological father and mother. When an individual has two copies of the same allele for a given marker, they are homozygous; when they have two different alleles, they are heterozygous.

2.1.1 DNA sample collection

DNA can be easily obtained from a variety of biological sources, not only body fluid but also nail, hair and used razors (Anderson et al., 1999; Lee et al., 1998; Lee & Ladd, 2001). For biometric applications, a buccal swab is the most simple, convenient and painless sample collection method (Hedman et al., 2008). Buccal cell collection involves wiping a small piece of filter paper or a cotton swab against the inside of the subject’s cheek, in order to collect shed epithelial cells. The swab is then air dried, or can be pressed against a treated collection card in order to transfer epithelial cells for storage purposes.

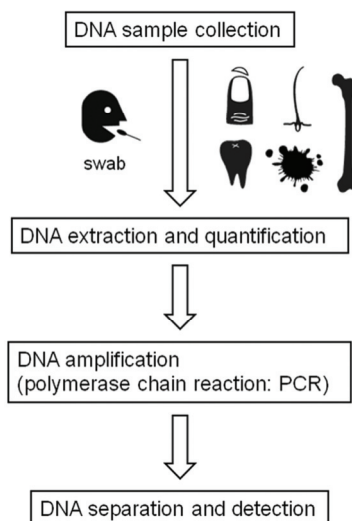


Fig. 2. The flow of DNA polymorphism analysis

2.1.2 DNA extraction and quantification

There are many methods available for extracting DNA (Butler, 2010). The choice of which method to use depends on several factors, especially the number of samples, cost, and speed. Extraction time is the critical factor for biometric applications. The author has already reported the “5-minute DNA extraction” using an automated procedure (Hashiyada, 2007a). The use of large quantities of fresh buccal cells made it possible to extract DNA in a short time.

In forensic cases, DNA quantitation is an important step (Butler, 2010). However, this step can be omitted in biometrics because a relatively large quantity of DNA can be recovered from fresh buccal swab samples.

2.1.3 DNA amplification (polymerase chain reaction: PCR)

The field of molecular biology has greatly benefited from the discovery of a technique known as the polymerase chain reaction, or PCR (Mullis et al., 1986; Mullis & Faloona, 1987; Saiki et al., 1986). First described in 1985 by Kary Mullis, who received the Nobel Prize in Chemistry in 1993, PCR has made it possible to make hundreds of millions of copies of a specific sequence of DNA in a few hours. PCR is an enzymatic process in which a specific region of DNA is replicated over and over again to yield many copies of a particular sequence. This molecular process involves heating and cooling samples in a precise thermal cycling pattern for approximately 30 cycles. During each cycle, a copy of the target DNA sequence is generated for every molecule containing the target sequence. In recent years, it has become possible to PCR amplify 16 STRs, including the gender assignment locus called 'amelogenin,' in one tube (Kimpton et al., 1993; Kimpton et al., 1996). Such multiplex PCR is enabled by commercial typing kits, such as AmpFlSTR® Identifiler® (Applied Biosystems, Foster City, CA, USA) and PowerPlex® 16 (Promega, Madison, WI, USA).

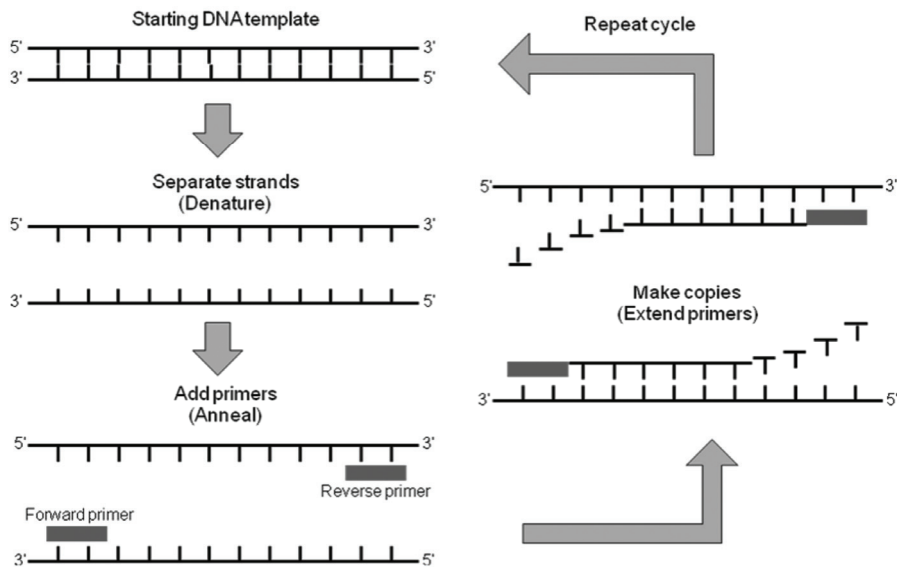


Fig. 3. DNA amplification with polymerase chain reaction (PCR)

2.1.4 DNA separation and detection

After STR polymorphisms have been amplified using PCR, the length of products must be measured precisely; some STR alleles differ by only 1 base-pair. Electrophoresis of the PCR products through denaturing polyacrylamide gels can be used to separate DNA molecules from 20–500 nucleotides in length with single base pair resolution (Slater et al., 2000). Recently, the fluorescence labelling of PCR products followed by multicolour detection has

been adopted by the forensic science field. Up to five different dyes can be used in a single analysis. Electrophoresis platforms have evolved from slab-gels to capillary electrophoresis (CE), which use a narrow glass filled with an cross-linked polymer solution to separate the DNA molecules (Butler et al., 2004). After data collection by the CE, the alleles (i.e., the type or the number of STR repeat units), are analyzed by the software that accompanies the CE machine.

It takes around four hours, starting with DNA extraction, to obtain data from 16 STRs including the sex determination locus.

2.2 Single nucleotide polymorphism (SNP)

The simplest type of polymorphism is the single nucleotide polymorphism (SNP), a single base difference at a particular point in the sequence of DNA (Brookes, 1999). SNPs normally have just two alleles, e.g., one allele is a cytosine (C) and the other is a thymine (T) (Fig. 4). SNPs therefore are not highly polymorphic and do not possess ideal properties for DNA polymorphism to be used in forensic analysis. However, SNPs are so abundant throughout the genome that it is theoretically possible to type hundreds of them. Furthermore, sample processing and data analysis may be more fully automated because size-based separation is not required. Thus, SNPs are prospective new bio-markers in clinical medicine (Sachidanandam et al., 2001; Stenson et al., 2009).

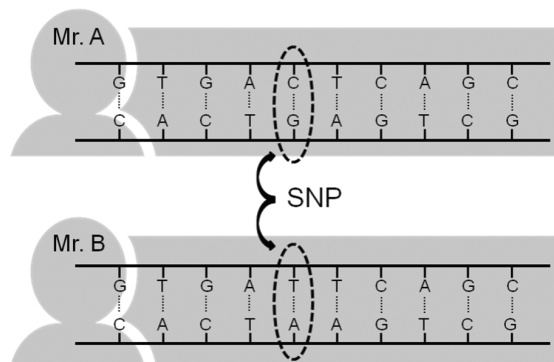


Fig. 4. The schema of Single nucleotide polymorphism (SNP)

2.2.1 SNP detection methods

Several SNP typing methods are available, each with its own strengths and weaknesses, unlike the STR analysis (Butler, 2010). In order to achieve the same power of discrimination as that provided by STRs, it is necessary to analyse many more SNPs. 40 to 50 SNPs must be analyzed in order to obtain reasonable powerful discrimination and define the unique profile of an individual (Gill, 2001). Importantly, however, we can count on the development of new SNP detection technologies, capable of high-throughput analysis, in the near future.

2.3 Lineage markers

Autosomal DNA markers are shuffled with each generation, which means that half of an individual's genetic information comes from his or her father and the other half from his or her mother. However, the Y chromosome (Chr Y) and mitochondrial DNA (mtDNA)

markers are called "lineage markers" because they are passed down from generation to generation without changing (except for mutational events). Maternal lineages can be followed using mtDNA sequence information (Anderson et al., 1981; Andrews et al., 1999) and whereas paternal lineages can be traced using Chr Y markers (Jobling & Tyler-Smith, 2003; Kayser et al., 2004). The analysis of lineage markers does not have the discriminatory power of autosomal markers. Even so, there are some features of both Chr Y and mtDNA that make them valuable forensic tools.

3. DNA polymorphism for biometric source

The most commonly studied or implemented biometrics are fingerprinting, face, iris, voice, signature, retina and the patterns of vein and hand geometry (Shen & Tan, 1999; Vijaya Kumar et al., 2004). No one model is best for all situations. In addition, these technologies are based on the measurement of similarity of features. This introduces an element of inaccuracy that renders the existing technologies unsuitable for a universal ID system. However, DNA polymorphism information, such as STRs and SNPs, could provide the most reliable personal identification. This data can be precisely defined the most minute level, is intrinsically digital, and does not change during a person's life or after his/her death. Therefore, DNA identification data is utilized in the forensic sciences. On the negative side, the biggest problem in using DNA is the time required for the extraction of nucleic acid and the evaluation of STR or SNP data. In addition, there are several other problems, such as the high cost of analysis, issues raised by monozygotic twins, and ethical concerns.

This section describes a method for generation of DNA personal ID (DNA-ID) based on STR and SNP data, specifically. In addition, by way of example, the author proposes DNA INK for authentic security.

3.1 DNA personal ID using STR system

We will refer to repeat counts of alleles obtained by STR analysis, as described in section 2.1, as (j, k) . Each locus is associated with two alleles with distinct repeat counts (j, k) , as shown in Fig. 2: one allele is inherited from the father, and the other from the mother. Before (j, k) can be applied to a DNA personal ID, it is necessary to statistically analyze how the distribution of (j, k) varies at a given locus based on actual data.

We can generate a DNA-ID, α_x , that includes allelic information about STR loci. The loci are incorporated in the following sequence. The repeat counts for the pair of alleles at each locus are arranged in ascending order.

Step 1. Measure the STR alleles at each locus.

Step 2. Obtain STR count values for each locus; express these in ascending order.

$$L : j \parallel k, j \leq k$$

Depending on the measurement, the same person's STR count may appear as (j, k) or (k, j) . Therefore, j and k are expressed in an ascending order, i.e., using $(j, k | j \leq k)$, in order to establish a one-to-one correspondence for each individual. This step is referred to as a ordering operation.

Step 3. Generate a DNA-ID α_x according to the following series, $L_i(j, k)$:

$$\alpha_x = L_1 \parallel L_2 \parallel L_3 \parallel \dots \parallel L_n$$

where L_i indicates the i th STR count (j, k) .

For example, suppose that Mr. M has the following alleles at the respective loci;

$$\begin{aligned}\alpha_x &= \text{D3S1358} \parallel \text{D13S317} \parallel \text{D18S51} \parallel \text{D21S11} \parallel \dots \parallel \text{D16S539} \\ &= (12,14) \parallel (8,11) \parallel (13,15) \parallel (29,32.2) \parallel \dots \parallel (10,10)\end{aligned}$$

The α_x was thus defined as follows.

$$\alpha_x = 1214811131529322 \dots 1010$$

When the STR number of an allele had a fractional component, such as allele 32.2 in D21S11, the decimal point was removed, and all of the numbers, including those after the decimal point, were retained.

Finally, α_x is generated number with several tens of digits, and becomes a personal identification information that is unique with a certain probability predicted by statistical and theoretical analysis.

3.1.1 Establishment of the identification format

Because α_x contains personal STR information, it must be encrypted to protect privacy. This can be achieved using a one-way function that also reduces the data length of the DNA-ID. This one-way function, the secure hash algorithm-1 (SHA-1), produces an ID with a length δ_x of 160 bits, according to the following transformation:

$$\delta_x = h(\alpha_x)$$

3.2 Statistical and theoretical analysis of DNA-ID

3.2.1 Matching probability at locus L

The probability that a STR allele (j, k) at locus L will occur in this combination is denoted as p_{jk} . The individual occurrence probabilities of j and k are denoted as p_j and p_k , respectively. Here, j and k are sequenced in ascending order to make the choice of generated ID unambiguous, using the STR analysis system described above. After the STR analysis, the probability that (j, k) occurs is p_{jk} plus the probability p_{kj} that (k, j) occurs. The reason for this is as follows. Even if (k, j) occurs in the same person during measurement, it is treated as (j, k) by rewriting it as (j, k) if $k > j$.

Therefore, p_{jk} is expressed as follows when $j \neq k$ ($j < k$):

$$\begin{aligned}p_{jk} &= p_j \cdot p_k + p_k \cdot p_j \\ &= 2p_j p_k\end{aligned}$$

$$\text{If } j = k,$$

$$p_{jk} = p_j \cdot p_j$$

3.2.2 Probability of a match between any two persons' DNA-ID

Probability p that the STR count at the same locus is identical for any two persons can be expressed as follows:

$$\begin{aligned} \text{When } j = k, & \sum_{j=1}^m (p_j \cdot p_k)^2 \\ \text{When } j \neq k, & \sum_{1 \leq j < k \leq m} (2p_j \cdot p_k)^2 \\ \therefore p = & \sum_{j=1}^m (p_j)^4 + \sum_{1 \leq j < k \leq m} 4(p_j \cdot p_k)^2 \end{aligned}$$

Here, m is the upper limit of j and k , and the information reported so far indicates $m = 60$. Next, a determination is made of the DNA-ID matching probability p_n , where n loci were used to generate the ID. The probability that the STR counts at the i^{th} locus will match for any two persons is denoted as p_i . When n loci are used, the probability p_n that the DNA-IDs of any two persons will match (the DNA-ID matching probability) is as follows:

$$p_n = \prod_{i=1}^n p_i$$

Here, it is assumed that there is no correlation among the STR loci.

3.2.3 Verification using validation experiment (STR)

As a validation experiment, we studied the genotype and distribution of allele frequencies at 18 STRs in 526 unrelated Japanese individuals. Data was obtained using three commercial STR typing kits: PowerPlex™ 16 system (Promega), PowerPlex SE33 (Promega), and AmpFISTR Identifiler™ (Applied biosystems) (Hashiyada, 2003a; 2003b). Information about the 18 target STRs is described in Table 1.

Step 1. Perform DNA extraction, PCR amplification and STR typing

Step 2. Perform the exact test (the data were shuffled 10,000 times), the homozygosity, and likelihood ratio tests using STR data for each STR locus in order to evaluate Hardy-Weinberg equilibrium (HWE). HWE provides a simple mathematical representation of the relationship among genotype and allele frequencies within an ideal population, and is central to forensic genetics. Importantly, when a population is in HWE, the genotype frequencies can be predicted from the allele frequencies.

Step 3. Calculate parameters, the matching probability, the expected and observed heterozygosity, the power of discrimination, the polymorphic information content, the mean exclusion chance, in order to estimate the polymorphism at each STR locus.

There are some loci on the same chromosomes (chr) such as D21S11 and Penta D on chr 21, D5S818 and CSF1PO on chr 5, and TPOX and D2S1338 on chr 2. No correlation was found between any sets of loci on the same chromosome, which means they are statistically independent. In addition, the statistical data for the 18 analyzed STRs, excluding the Amelogenin locus, were analyzed and showed a relatively high rate of matching probability; no significant deviation from HWE was detected. The combined mean exclusion chance was 0.999998995 and the combined matching probability was 1 in 9.98×10^{21} , i.e., 1.0024×10^{-22} . These values were calculated using polymorphism data from Japanese subjects; it is likely that different values would be obtained using data compiled from different ethnic groups, e.g., Caucasian or African.

Locus	Chromosome Location	Repeat Motif*	Locus	Chromosome Location	Repeat Motif*
TPOX	2 q 25.3	GAAT	TH01	11 p 15.5	TCAT
D2S1338	2 q 35	TGCC/TTCC	VWA	12 p 13.31	TCTG/TCTA
D3S1358	3 p 21.31	TCTG/TCTA	D13S317	13 q 31.1	TATC
FGA	4 q 31.3	CTTT/TTCC	Penta E	15 q 26.2	AAAGA
D5S818	5 q 23.2	AGAT	D16S539	16 q 24.1	GATA
CSF1PO	5 q 33.1	TAGA	D18S51	18 q 21.33	AGAA
SE33	6 q 14	AAAG	D19S433	19 q 12	AAGG/TAGG
D7S820	7 q 21.11	GATA	D21S11	21 q 21.1	TCTA/TCTG
D8S1179	8 q 24.13	TCTA/TCTG	Penta D	21 q 22.3	AAAGA

* Two types of motif means a compound or complex repeat sequence

Table 1. Information about autosomal STR loci

3.2.4 The “Birthday Paradox” of DNA-ID

In principle, the low matching probability of STR-based IDs would allow absolute and unequivocal discrimination between individuals. However, if STRs are to be used as an authentication system in our society, we must investigate the probability of two or more randomly selected people having an identical DNA-ID. The most well-known simulation of this probability is “the birthday paradox”. Of 40 students in a class, the probability that at least two students have the same birthday is approximately 0.9. This result seems counterintuitive, and is called a “paradox,” because for any single pair of students, the probability that they have the same birthday is $1/365$ (0.0027). The paradox arises when we forget to consider that we are selecting samples randomly out of the members in a group.

In two randomly selected individuals, the probability that one STR locus is different and that all STR loci are identical is $(1-P_M)^{L(L-1)/2}$ and $1-(1-P_M)^{L(L-1)/2}$, respectively, where L is the population size. However, the formula, $1-(1-P_M)^{L(L-1)/2}$, is beyond the ability of personal computers, so we use the expected value, $L(L-1)/2 \cdot P_M$, to estimate two persons having the same STR genotype. This formula can use an approximate value of $1-(1-P_M)^{L(L-1)/2}$. This is because L^2 is much smaller than $1/P_M$ when L is small, and because $1-(1-P_M)^{L(L-1)/2}$ is smaller than $L(L-1)/2 \cdot P_M$ when L is not small. In this report, the value, $L(L-1)/2 \cdot P_M$, is defined as the practical matching probability (P_{PM}). The matching probability (P_M) for 18 STRs is 1.0024×10^{-22} , as described above. When P_{PM} multiplied by the population size is less than 1, each person in the population could have a unique DNA-ID. Therefore, when using 18 loci, a population of tens of millions could be expected to include pairs of individuals with identical STR alleles. If the frequencies of STR alleles are similar among all ethnic groups, each person in Japan (or the world) could have a unique DNA-ID if the P_{PM} of the STR system were approximately 10^{-24} and 10^{-30} , respectively. As the number of people in a community increases, the more the practical matching probability increases.

This number can be applied for unrelated persons; however, we also need to consider P_{PM} between related individuals. For instance, between two first cousins, if 41 STR loci are analyzed, we can obtain a unique DNA-ID. In addition, discrimination between half siblings requires analysis of 57 STR loci guarantee a unique DNA-ID. Thus, when using DNA identification systems such as STR systems for DNA-personal-IDs, the P_{PM} should be considered for both related and unrelated individuals (Hashiyada, 2007b).

3.3 DNA personal ID using SNP system

The vast majority of SNPs are biallelic, meaning that they have two possible alleles and therefore three possible genotypes. For example, if the alleles for a SNP locus are R and S (where 'R' and 'S' could represent a A(adenine), G(guanine), C(cytosine) and T(thymine) nucleotide), three possible genotypes would be RR, RS (SR) or SS. Because a single biallelic SNP by itself yields less information than a multiallelic STR marker, it is necessary to analyze a larger number of SNPs in order to obtain a reasonable power of discrimination to define a unique profile. Computational analysis have shown that on average, 25 to 45 SNP loci are needed in order to yield equivalent random match probabilities comparable to those obtained with the 13 core STR loci that have been adopted by the FBI's DNA database (COmbined DNA Index System, CODIS).

The steps of creating a DNA-ID using SNPs are as follows;

Step 1. Define alleles 1 and 2 for each SNP locus. Since DNA has a double helix structure, the single nucleotide polymorphism of A or G is the same polymorphism of T or C, respectively (Fig. 4). In other words, it is important to specify which strand of the double helix is to be analyzed, and to define allele 1 and allele 2 at the outset.

Step 2. Analyze the SNP loci and place them in the following order.

$$L : \text{allele 1} \parallel \text{allele 2}$$

Step 3. Generate the DNA-ID α_x according to the following series of L_i (allele1, allele2):

$$\alpha_x = L_1 \parallel L_2 \parallel L_3 \parallel \dots \parallel L_n$$

where L_i indicates the i^{th} SNP nucleotide (allele1, allele2).

For example, suppose that a person has the following alleles at the respective loci;

$$\begin{aligned} \alpha_x &= \text{SNP 1} \parallel \text{SNP 2} \parallel \text{SNP 3} \parallel \text{SNP 4} \parallel \dots \parallel \text{SNP 50} \\ &= (A,A) \parallel (C,T) \parallel (T,C) \parallel (C,C) \parallel \dots \parallel (G,A) \end{aligned}$$

Then α_x would be defined as follows.

$$\alpha_x = \text{AACTTCCC.....GA}$$

Next, the four types of nucleotide, A, G, C and T, are translated into binary notation.

$$A=00, G=01, C=10, T=11$$

Finally, the α_x is described as a string of 100 bits (digits of value 0 or 1).

$$\alpha_x = 0000101111101010.....0100$$

This α_x must be encrypted for privacy protection using the secure hash algorithm-1 (SHA-1) for the same reasons as described above for STRs. The resulting DNA-ID (SNP) has a length δ_x of 160 bits, according to the following transformation:

$$\delta_x = h(\alpha_x)$$

3.3.1 Verification using validation experiment (SNP)

As a validation experiment, the author analyzed 120 autosomal SNPs in 100 unrelated Japanese subjects using the TaqMan® method (Applied Biosystems), and built a Japanese SNP database for identification. Although several SNPs were located on the same autosomal chromosome, no correlation was found between alleles at any SNP loci. Furthermore, no significant deviation from Hardy–Weinberg Equilibrium (HWE) was detected. The matching probability (MP) of each SNP ranged from 0.375–0.465 (Hashiyada, 2007a). The MP for 41 SNPs (3.63×10^{-18}), which have high MP in each loci, was very similar to the MPs obtained with the current STR multiplex kits, PowerPlex™ 16 System (Promega) and AmpFISTR Identifier (Applied Biosystems), which were 5.369×10^{-18} and 1.440×10^{-17} , respectively in Japanese population.

3.4 Rapid analysis system of SNP

A reduction of the time required for DNA analysis is necessary in order to make practical use of DNA biometrics. In the STR system, it is difficult to decrease the analysis time because it is necessary to perform electrophoresis after PCR amplification. From DNA extraction to STR typing, the entire process takes 4–5 hours. However, there are many methods for analyzing SNPs that do not demand such a lengthy process. The author developed the SNP typing methodology using the modified TaqMan® method, which is capable of amplifying the DNA and typing the SNPs at the same time. The author modified the number of PCR cycles and the annealing/extension time, and selected SNP loci that yield successful results under the modified PCR conditions. This new method is capable of detecting and typing 96 SNPs within 30 minutes (Hashiyada et al., 2009).

3.5 DNA INK

In this paragraph, the author demonstrates an example of an application of STR polymorphism information, specifically the authentication of rare or expensive goods using the DNA-ID. The author outlines the development of biometric ink containing DNA whose sequence is based on personal STR information. The “DNA INK” is made of synthetic DNA and printing ink.

Step 1. Perform STR analysis by the method described above.

Step 2. Generate the DNA-ID, δ_x , consisting of 160 bits, as described above.

Step 3. Extract one-quarter of the data in the DNA-ID (δ_x) in order to reduce costs and improve practicality. The original 160-bit length was defined as

$$\delta_{x_i} = \delta_{x_1} | \delta_{x_2} | \delta_{x_3} | \delta_{x_4}$$

where δ_{x_1} , δ_{x_2} , δ_{x_3} and δ_{x_4} refer to the identification, ID, containing of the first, second, third and fourth 40 bits of δ_x . Each set of 8 data bits is extended by two redundant bits known as the shift and check bits, which serve not only as check but also as limiting factors in the latter stages of DNA sequence generation. These limiting factors are necessary in DNA sequence analysis in order to exclude five or more repetitions of the same base. The extracted 40-bit data as follows;

$$\delta_{x_1} = 1001100110011101011010101001011110100010$$

$$\delta_{x_1} = 10011001 \ [10] \ 10011101 \ [00] \ 01101010 \ [01] \ 10010111 \ [00] \ 10100010 \ [11]$$

(Shift and check bits show as square brackets with underlines.)

Step 4. Transform the bit series generated above into base sequences according to the following scheme. We called this step the “Encoded Base Array” method.

00=A(adenine), 01=C(cytosine), 10=G(guanine), 11=T(thymine)

δ_{x_1} =10011001 [10]10011101 [00] 01101010 [01] 10010111 [00] 10100010 [11]
 =GCGC [C] GCTC [A] CGGG [C] GCCT [A] GGAG [T]

Step 5. Define the identification data format by adding a header (H, 10 bits) and a serial number (N, 30 bits) to δ_{x_1} (40 + 10 = 50 bits). The resulting DNA sequence, consisting of H (5-bp), N (15-bp) and δ_{x_1} (25-bp) would then be flanked by two 20-bp-long primer sequences. This synthetic DNA could be amplified by PCR, and only those who know the primer sequences would be able to analyze the intervening sequence. Figure 5 shows the structure of the 85-bp synthetic DNA sequence.

Step 6. Synthesize the complementary strand. Synthetic single-strand DNA is more economical to produce than double-strand DNA, but much less physically stable; therefore, double-strand PCR-amplified DNA should be used for incorporation into the DNA ink.

Step 7. Mix 3 mg of double-strand DNA with 100 ml of ink. The ink itself is composed of a colorless transparent pigment, so that it is invisible to the naked eye, but contains an IR color former that enables easy detection of the printed mark. In addition, add dummy DNA in order to make the DNA-ID sequence difficult to analyze by someone who does not know the primer sequences.

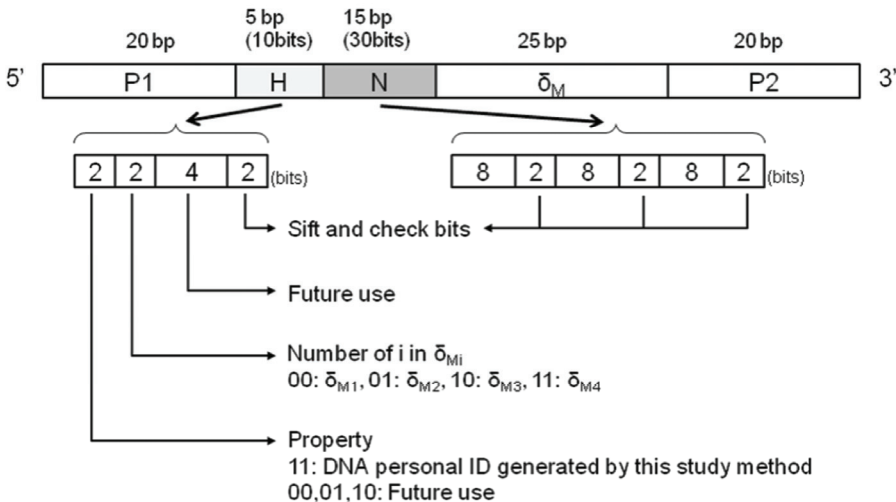


Fig. 5. Sequence structure of the 85-bp single-strand DNA-ID
 P1, P2: Primer sequences are designed so as not to anneal to the human genome
 H: Header, N: Serial number

The several types of resistance tests, by heat, acids, alkalis, alcohol, ultraviolet (UV) and sunlight, were used to ascertain the durability of DNA ink for practical use. Samples printed using DNA ink were covered with zinc oxide (ZnO) on the surface in order to enhance resistance to UV light, which is the major cause of DNA degradation.

The target DNA sequence was detected successfully in all resistance tests except for the UV exposure test. However, the durability improved when the ink was covered by ZnO, allowing successful amplification even after 40 hours of UV exposure. Finally, the DNA ink was proved as a sort of biological memory which could print the polymorphism information created by DNA, on the surface of everything excluding the air and water.

4. Problems of DNA biometrics

There can be no doubt that DNA-ID is potentially useful as a biometric. It has many advantages, including accuracy, strictness, discriminatory power (and ease of increasing this power), and the ability to use the same analysis platform all over the world. However, DNA polymorphism information is not widely used in biometrics at this point. The weak points of DNA-ID are discussed below.

4.1 Time required for DNA analysis

The most serious flaw is that DNA analysis is time-consuming compared to other authentication methods. It takes at least 4 hours to get STR identification data by common methods used in forensic science. Most of the time required for DNA analysis is taken up by PCR amplification and electrophoresis. It is impossible to dramatically shorten the duration of these steps using existing technologies. SNP analysis may be faster, however: it is possible to analyze 96 SNPs within 30 minutes (Hashiyada, Itakura et al., 2009). Thus, a SNP system could use a specific usage, for example in passports or in very large-scale mercantile transactions.

4.2 Ethical concerns

The polymorphic target region in DNA used to create the DNA-ID does not relate to a person's physical characteristics or disease factors, since the STRs and the SNP loci were selected from the extragenic regions. However, because the DNA-ID system involves handling information that can identify each individual, it should be strictly supervised in order to protect privacy. Once the DNA-ID has been generated, the one-way encryption described above makes it impossible to recover any of the original DNA information (3.1, 3.3). Therefore, raw materials like buccal swab should be especially tightly controlled in order to prevent spoofing.

4.3 Monozygotic twins and DNA chimeras

Monozygotic twins, or more commonly referred to as identical twins, begin life as a single egg, which is fertilized by one sperm but then splits into two eggs early in the gestational period. Therefore, the twins share a precisely duplicated whole genome, and can't be distinguished by DNA polymorphism. However, sometimes one member of a pair of identical twins can develop cancer or schizophrenia while the other does not (Zwijnenburg et al., 2010). A recent "twin study" has revealed that twin pairs have significant differences in their DNA sequence, and furthermore that environmental factors can change gene expression and susceptibility to disease by affecting epigenetics, i.e., changes in the DNA

that do not alter its sequence (Haque et al., 2009). Such data will hopefully aid development of tools that allow discrimination between the identical twins in the near future.

A DNA chimera refers to a recombinant molecule of DNA composed of segments from more than one source. The author has observed chimerism in a case of allogeneic bone marrow transplantation (BMT). The recipient had suffered from acute promyelocytic leukemia and received a BMT from a healthy donor, resulting in complete remission of the leukemia. Samples of peripheral blood leukocytes (PBL), buccal mucosa, hair follicles and fingernails were collected from the transplant recipient. DNA analysis revealed that the STR profile of PBL of the recipient had completely converted to donor type, whereas the hair follicles and fingernails were recipient-derived. DNA patterns of the buccal mucosa appeared chimeric, i.e., they had qualities of both the recipient and donor. Neutrophilic leukocytes were observed in smear specimens from buccal swabs of the recipient, indicating that the buccal cells were not truly chimeric but were instead merely contaminated with leukocytes.

4.4 Cost

DNA analysis requires a high capital cost in order to buy and maintain equipment as well as purchase commercial kits. In addition, it is necessary to equip a laboratory and employ specialists in molecular biology. These high costs may pose a barrier to entry of venture capitals. The more popular such DNA techniques become, however, the lower the unit costs of the apparatus and reagents will become.

5. Conclusion

Development of biometric authentication technologies has progressed rapidly in the last few years. Personal identification devices based on unique patterns of fingerprints, iris, or subcutaneous veins in the finger have all been commercialized. All of these methods of verification are based on matching analog patterns or feature-point comparisons. Because they lack absolute accuracy, they have not yet achieved a universal standard. Among the various types of biometric information source, the DNA-ID is thought to be the most reliable method for personal identification. DNA information is intrinsically digital, and does not change either during a person's life or after his/her death. The discriminatory power of the data can be enhanced by increasing the number of STR or SNP loci. The DNA-ID could be encrypted via the one-way function (SHA-1) to protect privacy and to reduce data length. Using the STR system, it is currently difficult to complete analysis within 3 hours; however, using the SNP system, it is possible to analyse 96 SNPs within 30 minutes. Both systems yielded verifiable results in validation experiments. The author also introduced the idea of DNA-INK as a practical application of DNA-ID.

DNA-ID has some disadvantages, as well, including long analysis time, ethical concerns, high cost, and the impossibility of discrimination of monozygotic twins. However, the author believes that the DNA-ID must be employed as a biometric methodology, using breakthrough methods developed in the near future.

6. Acknowledgments

I am grateful to Dr. Yukio Itakura for his extensive support, and I give special thanks to my colleagues at Div. Forensic Medicine, Tohoku University. I also thank Prof. M. Funayama for reading the manuscript and giving me helpful advice.

7. References

- Alberts, B., Jhonson, A., Lewis, J., Raff, M., Roberts, K., Walter P. (2002). *Molecular biology of THE CELL* NY, USA: Garland Science.
- Anderson, S., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806): p. 457-65.
- Anderson, T.D., et al. (1999). A validation study for the extraction and analysis of DNA from human nail material and its application to forensic casework. *J Forensic Sci*, 44(5): p. 1053-6.
- Andrews, R.M., et al. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*, 23(2): p. 147.
- Brookes, A.J. (1999). The essence of SNPs. *Gene*, 234(2): p. 177-86.
- Butler, J.M., et al. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 25(10-11): p. 1397-412.
- Butler, J.M. (2010). *Fundamemntals of Forensic DNA Typipng*: ELSERVIER.
- Clayton, T.M., et al. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci Int*, 76(1): p. 7-15.
- Collins, F.S., et al. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011): p. 931-45.
- Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med*, 114(4-5): p. 204-10.
- Hagelberg, E., et al. (1991). Identification of the skeletal remains of a murder victim by DNA analysis. *Nature*, 352(6334): p. 427-9.
- Haque, F.N., et al. (2009). Not really identical: epigenetic differences in monozygotic twins and implications for twin studies in psychiatry. *Am J Med Genet C Semin Med Genet*, 151C(2): p. 136-41.
- Hashiyada, M., et al. (2009). Development of a spreadsheet for SNPs typing using Microsoft EXCEL. *Leg Med (Tokyo)*, 11 Suppl 1: p. S453-4.
- Hashiyada, M., Itakura, Y., Nagashima, T., Nata, M., Funayama, M. (2003a). Polymorphism of 17 STRs by multiplex analysis in Japanese population. *Forensic Sci Int*, 133(3): p. 250-3.
- Hashiyada, M., Itakura, Y., Nagasima, T., Sakai, J., Funatyama, M. (2007a). High-throughput SNP analysis for human identification. *DNA Polymorphism Official Journal of Japanese Society for DNA Polymorphism Resarch*, 15: p. 3.
- Hashiyada, M., Matsuo, S., Takei, Y., Nagasima, T., Itakura, Y., Nata, M., Funayama, M. (2003b). The length polymorphism od SE33(ACTBP2) locus in Japanese population. *Practice in Forensuic Medicine*, 46: p. 4.
- Hashiyada, M., Sakai, J., Nagashima, T., Itakura Y., Kanetake, J., Takahashi, S., Funayama, M. (2007b). The birthday paradox in the biometric personal authentication system using STR polymorphism -Practical matching probabilities evaluate the DNA psonal ID system-. *The Research and practice in forensic medicine*, 50: p. 5.
- Hedman, J., et al. (2008). A fast analysis system for forensic DNA reference samples. *Forensic Sci Int Genet*, 2(3): p. 184-9.
- Jasinska, A.&W.J. Krzyzosiak (2004). Repetitive sequences that shape the human transcriptome. *FEBS Lett*, 567(1): p. 136-41.

- Jeffreys, A.J., et al. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314(6006): p. 67-73.
- Jeffreys, A.J., et al. (1992). Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Sci Int*, 56(1): p. 65-76.
- Jeffreys, A.J., et al. (1995). Mutation processes at human minisatellites. *Electrophoresis*, 16(9): p. 1577-85.
- Jobling, M.A.&C. Tyler-Smith (2003). The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*, 4(8): p. 598-612.
- Kayser, M., et al. (2004). A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet*, 74(6): p. 1183-97.
- Kimpton, C.P., et al. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl*, 3(1): p. 13-22.
- Kimpton, C.P., et al. (1996). Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis*, 17(8): p. 1283-93.
- Lander, E.S., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822): p. 860-921.
- Lee, H.C., et al. (1998). Forensic applications of DNA typing: part 2: collection and preservation of DNA evidence. *Am J Forensic Med Pathol*, 19(1): p. 10-8.
- Lee, H.C.&C. Ladd (2001). Preservation and collection of biological evidence. *Croat Med J*, 42(3): p. 225-8.
- Mullis, K., et al. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1: p. 263-73.
- Mullis, K.B.&F.A. Faloona (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, 155: p. 335-50.
- Ruitberg, C.M., et al. (2001). STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res*, 29(1): p. 320-2.
- Sachidanandam, R., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822): p. 928-33.
- Saiki, R.K., et al. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature*, 324(6093): p. 163-6.
- Shen, W.&T. Tan (1999). Automated biometrics-based personal identification. *Proc Natl Acad Sci U S A*, 96(20): p. 11065-6.
- Slater, G.W., et al. (2000). Theory of DNA electrophoresis: a look at some current challenges. *Electrophoresis*, 21(18): p. 3873-87.
- Stenson, P.D., et al. (2009). The Human Gene Mutation Database: 2008 update. *Genome Med*, 1(1): p. 13.
- Venter, J.C., et al. (2001). The sequence of the human genome. *Science*, 291(5507): p. 1304-51.
- Vijaya Kumar, B.V., et al. (2004). Biometric verification with correlation filters. *Appl Opt*, 43(2): p. 391-402.
- Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., Losick R. (2004). *Molecular Biology of the Gene*, San Francisco, CA, USA: Benjamin Cummings, Cold Spring Harbor Laboratory Press.
- Zwijnenburg, P.J., et al. (2010). Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet*, 153B(6): p. 1134-49.

Part 2

Behavioral Biometrics

Keystroke Dynamics Authentication

Romain Giot, Mohamad El-Abed and Christophe Rosenberger
GREYC Research Lab
Université de Caen Basse Normandie, CNRS, ENSICAEN
France

1. Introduction

Everybody needs to authenticate himself on his computer before using it, or even before using different applications (email, e-commerce, intranet, . . .). Most of the times, the adopted authentication procedure is the use of a classical couple of *login* and *password*. In order to be efficient and secure, the user must adopt a strict management of its credentials (regular changing of the password, use of different credentials for different services, use of a strong password containing various types of characters and no word contained in a dictionary). As these conditions are quite strict and difficult to be applied for most users, they do not respect them. This is a big security flaw in the authentication mechanism (Conklin et al., 2004). According to the 2002 NTA Monitor Password Survey¹, a study done on 500 users shows that there is approximately 21 passwords per user, 81% of them use common passwords and 30% of them write their passwords down or store them in a file. Hence, password-based solutions suffer from several security drawbacks.

A solution to this problem, is the use of *strong authentication*. With a strong authentication system, you need to provide, at least, two different authenticators among the three following: (a) *what you know* such as passwords, (b) *what you own* such as smart cards and (c) *what you are* which is inherent to your person, such as biometric data. You can adopt a more secure password-based authentication by including the *keystroke dynamics* verification (Gaines et al., 1980; Giot et al., 2009c). In this case, the strong authentication is provided by what we know (the password) and what we are (the way of typing it). With such a scheme, during an authentication, we verify two issues: (i) is the credential correct? (ii) is the way of typing it similar? If an attacker is able to steal the credential of a user, he will be rejected by the verification system because he will not be able to type the genuine password in a same manner as its owner. With this short example, we can see the benefits of this behavioral modality. Figure 1 presents the enrollment and verification schemes of keystroke dynamics authentication systems.

We have seen that keystroke dynamics allows to secure the authentication process by verifying the way of typing the credentials. It can also be used to secure the session after its opening by detecting the changing of typing behavior in the session (Bergadano et al., 2002; Marsters, 2009). In this case, we talk about continuous authentication (Rao, 2005), the computer knows how the user interacts with its keyboard. It is able to recognize if another individual uses the

¹ <http://www.nta-monitor.com/>

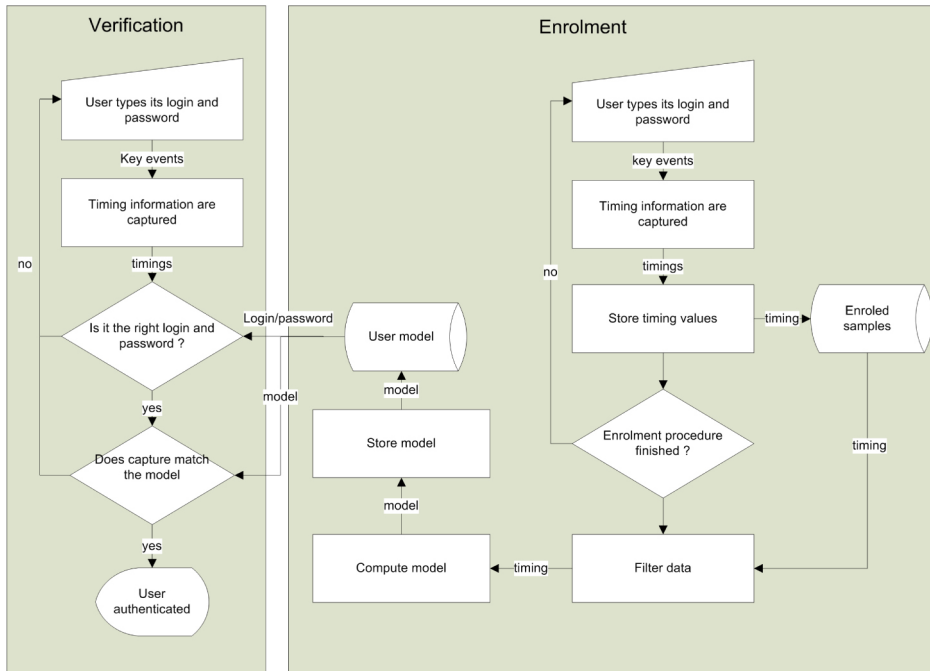


Fig. 1. Keystroke dynamics enrolment and authentication schemes: A password-based authentication scenario

keyboard, because the way of interacting with it is different. Moreover, keystroke dynamics can also prevent the steal of data or non authorized computer use by attackers.

In this chapter, we present the general research field in keystroke dynamics based methods. Section 2 presents generalities on keystroke dynamics as the topology of keystroke dynamics methods and its field of application. Even if it has not been studied a lot comparing to other biometric modalities (see Table 1), keystroke dynamics is a biometric modality studied for many years. The first reference to such system dates from 1975 (Spillane, 1975), while the first real study dates from 1980 (Gaines et al., 1980). Since, new methods appeared all along the time which implies the proposal of many keystroke dynamics systems. They can be static, dynamic, based on one or two classes pattern recognition methods. The aim of this section is to explain all these points.

Modality	keystroke dynamics	gait	fingerprint	face	iris	voice
Nb doc.	2,330	1,390	17,700	18,300	10,300	14,000

Table 1. Number of documents referenced by Google Scholar per modality. The query is "modality biometric authentication"

In section 3, we present the acquisition and features extraction processes of keystroke dynamics systems. Section 4 presents the authentication process of such keystroke dynamics based methods. These methods can be of different types: one class based (in this case, the model of a user is only built with its own samples), or two classes based (in this case, the model of a user is built also with samples of impostors). For one class problems, studies are based

on distance measures Monroe & Rubin (1997), others on statistical properties (de Magalhaes et al., 2005; Hocquet et al., 2006) or bioinformatics tools Revett (2009). Concerning two classes problems, neural networks (Bartmann et al., 2007) and Support Vectors Machines (SVM) (Giot et al., 2009c) have been used. Section 5 presents the evaluation aspects (performance, satisfaction and security) of keystroke dynamics systems. A conclusion of the chapter and some emerging trends in this research field are given in section 6.

2. Generalities

2.1 Keystroke dynamics topology

Keystroke dynamics has been first imagined in 1975 (Spillane, 1975) and it has been proved to work in early eighties (Gaines et al., 1980). First studies have proved that keystroke dynamics works quite well when providing a lot of data to create the model of a user. Nowadays, we are able to perform good performance without necessitating to ask a user to give a lot of data. "A lot of data" means typing a lot of texts on a computer. This possibility of using, or not, a lot of data to create the model allows us to have two main families of keystroke dynamics methods (as illustrated in Figure 2):

- The *static families*, where the user is asked to type several times the same string in order to build its model. During the authentication phase, the user is supposed to provide the same string captured during his enrollment. Such methodology is really appropriate to authenticate an individual by asking him to type its own password, before login to its computer session, and verifying if its way of typing matches the model. Changing the password implies to enroll again, because the methods are not able to work with a different password. Two main procedures exist: the use of a real *password* and, the use of a *common secret*. In the first case, each user uses its own password, and the pattern recognition methods which can be applied can only use one class classifiers or distance measures. In the second case, all users share the same password and we have to address a two classes problem (genuine and impostor samples) (Bartmann et al., 2007; Giot et al., 2009c). Such systems can work even if all the impostors were not present during the training phase (Bartmann et al., 2007).
- The *dynamic families* allow to authenticate individuals independently of what they are typing on the keyboard. Usually, they are required to provide a lot of typing data to create their model (directly by asking them to type some long texts, or indirectly by monitoring their computer use during a certain period). In this solution, the user can be verified on the fly all the time he uses its computer. We can detect a changing of user during the computer usage. This is related as continuous authentication in the literature. When we are able to model the behavior of a user, whatever the thing he types, we can also authenticate him through a challenge during the normal login process: we ask the user to type a random phrase, or a shared secret (as a one-time password, for example).

2.2 Applications and interest

From the topology depicted in Figure 2, we can imagine many applications. Most of them have been presented in scientific papers and some of them are proposed by commercial applications.

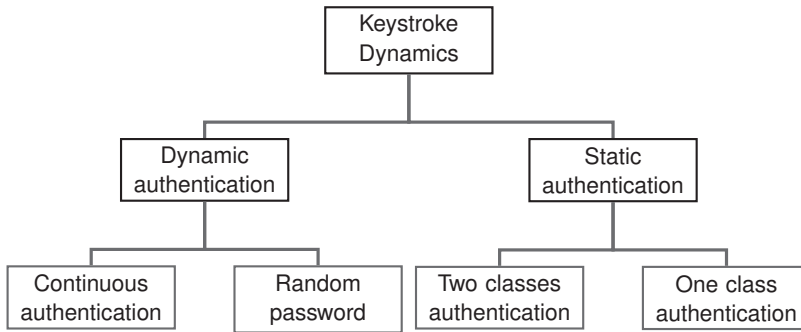


Fig. 2. Topology of keystroke dynamics families

2.2.1 Authentication for logical access control

Most of commercial softwares are related to static keystroke dynamics authentication by modifying the Operating System login procedure. The authentication form is modified to include the capture of the timing information of the password (see Section 3.2.1), and, in addition of verifying the password, the way of typing is also verified. If it matches to the user profile, he is authenticated. Otherwise, he is rejected and considered as an impostor. By this way, we obtain two authentication factors (strong authentication): (i) *what we know*, which is the password of the user; (ii) *what we are*, which is the way of typing the password. The best practices of password management are rarely (even never) respected (regular change of password, use of a complex password, forbid to write the password on a paper, ...), because they are too restrictive. Moreover, they can be easily obtained by sniffing network, since a wide range of websites or protocols do not implement any protection measures on the transmission links. That is here, where keystroke dynamics is interesting, since it allows to avoid impostors which were able to get the password to authenticate instead of the real user. In addition, some studies showed that keystroke dynamics holds better performance when using simple passwords, than more complicated ones. If the user keeps a simple password, he remembers it more easily, and, administrators lost less time by giving new passwords.

When used in a logical access control, the keystroke dynamics process uses different information such as the name of the user, the password of the user, the name and the password of the user, an additional passphrase (common for all the users, unique to the user). Modi & Elliott (2006) show that, sadly, using spontaneously generated password does not give interesting performance. This avoids the use of one time passwords associated to keystroke dynamics (when we are not in a monitoring way of capturing biometric data).

2.2.2 Monitoring and continuous authentication

Continuously monitoring the way the user interacts with the keyboard is interesting (Ahmed & Traore, 2008; Rao, 2005; Song et al., 1997). With such a mechanism, the system is able to detect the change of user during the session life. By this way, the computer is able to lock the session if it detects that the user is different than the one which has previously been authenticated on this computer. Such monitoring can also be used to analyse the behavior of the user (instead its identity), and, detect abnormal activities while accessing to highly restricted documents or executing tasks in an environment where the user must be alert at all the times (Monrose & Rubin, 2000).

Continuous authentication is interesting, but has a lot of privacy concerns, because the system monitors all the events. Marsters (2009) proposes a solution to this problem of privacy. His keystroke dynamics system is not able to get the typed text from the biometric data. It collects quadgraphs (more information on ngraphs is given later in the chapter) for latency and trigraphs for duration. Instead of storing this information in an ordered log, it is stored in a matrix. By this way, it is impossible to recover the chronological log of keystroke, and, improve the privacy of the data.

2.2.3 Ancillary information

Keystroke dynamics can also be used in different contexts than the authentication. Monroe & Rubin (2000) suggest the use of keystroke dynamics to verify the state of the user and alert a third party if its behavior is abnormal. But, this was just a suggestion, and not a verification. Hocquet et al. (2006) show that keystroke dynamics users can be categorised into different groups. They automatically assign each user to a group (authors empirically use 4 clusters). The parameters of the keystroke dynamics system are different for each group (and common for each user of the group), which allows to improve the performance of the system. However, there is no semantic information on the group, as everything is automatic. Giot & Rosenberger (2011) show that it is possible to recognize the gender of an individual who types a predefined string. The gender recognition accuracy is superior to 91%. This information can be useful to automatically verify if the gender given by an individual is correct. It can be also used as an extra feature during the authentication process in order to improve the performance. Authors achieved an improvement of 20% of the Error Equal Rate (EER) when using the guessed gender information during the verification process. Epp (2010) shows that it is possible to get the emotional state of an individual through its keystroke dynamics. The author argues that if the computer is able to get the emotional state of the user, it can adapt its interface depending on this state. Such ability facilitates computer-mediated communication (communication through a computer). He respectively obtains 79.5% and 84.2% of correct classification for the relaxed and tired states. Khanna & Sasikumar (2010) show that 70% of users decrease their typing speed while there are in a negative emotional state (compared to a neutral emotional state) and 83% of users increase their typing speed when they are in a positive emotional state. Keystroke dynamics is also used to differentiate human behavior and robot behavior in keyboard use. This way, it is possible to detect a bot which controls the computer, and, intercepts its actions (Stefan & Yao, 2008).

3. Keystroke dynamics capture

The capture phase is considered as an important issue within the biometric authentication process. The capture takes place at two different important times:

- *The enrollment*, where it is necessary to collect several samples of the user in order to build its model. Depending of the type of keystroke dynamics systems, the enrollment procedure can be relatively different (typing of the same fixed string several times, monitoring of the computer usage, ...), and, the quantity of required data can be totally different between the studies (from five inputs (Giot et al., 2009c) to more than one hundred Obaidat & Sadoun (1997)).
- *The verification*, where a single sample is collected. Various features are extracted from this sample. They are compared to the biometric model of the claimant.

This section first presents the hardware which must be used in order to capture the biometric data, and, the various associated features which can be collected from this data.

3.1 Mandatory hardware and variability

Each biometric modality needs a particular hardware to capture the biometric data. The price of this hardware, as well as the number of sensors to buy, can be determinant when choosing a biometric system supposed to be used in a large infrastructure with number of users (*e.g.*, necessity to buy a fingerprint sensor for each computer, if we choose a logical access control for each machine). Keystroke dynamics is probably the biometric modality with the cheapest biometric sensor : it uses only a simple keyboard of your computer. Such keyboard is present in all the personal computers and in all the laptops. If a keyboard is broken and it is necessary to change it, it would cost no more than 5\$. Table 2 presents the sensor and its relative price for some modalities, in order to ease the comparison of these systems.

Modality	<i>keystroke</i>	<i>fingerprint</i>	<i>face</i>	<i>iris</i>	<i>hand veins</i>
Sensor	keyboard	fingerprint sensor	camera	infrared camera	near infra red camera
Price	very cheap	normal	normal	very expensive	expensive

Table 2. Price comparison of hardware for various biometric modalities

Of course, each keyboard is different on various points:

- The shape (straight keyboard, keyboard with a curve, ergonomic keyboard, ...)
- The pressure (how hard it is to press the key)
- The position of keys (AZERTY, QWERTY, ...). Some studies only used the numerical keyboard of a computer (Killourhy & Maxion, 2010; Rodrigues et al., 2006).

Hence, changing a keyboard may affect the performances of the keystroke recognition. This problem is well known in the biometric community and is related as *cross device matching* (Ross & Jain, 2004). It has not been treated a lot in the keystroke dynamics literature. Figure 3 presents the shape of two commonly used keyboards (laptop and desktop). We can see that they are totally different, and, the way of typing on it is also different (maybe mostly due by the red ball on the middle of the laptop keyboard).

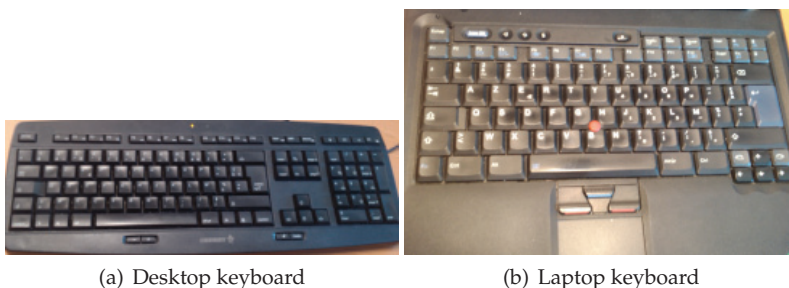


Fig. 3. Difference of shape of two classical keyboards

Having this sensor (the keyboard) is not sufficient, because (when it is a classical one), the only information it provides is the code of the key pressed or released. This is not at all a biometric information, all the more we already know if it is the correct password or not, whereas we

are interested in if it is the right individual who types it. The second thing we need is an accurate timer, in order to capture at a sufficient precision the time when an event occurs on the keyboard. Once again, this timer is already present in every computer, and, each operating system is able to use it. Hence, we do not need to buy it. There is a drawback with this timer: its resolution can be different depending on the chosen programming language or the operating system. This issue has been extensively discussed by Killourhy & Maxion (2008), where it is shown that better performance are obtained with higher accuracy timer. Some researchers have also studied the effect of using an external clock instead of the one inside the computer. Pavaday. et al. (2010) argue that it is important to take into consideration this timer, especially when comparing algorithms, because it has an impact on performance. They also explain how to configure the operating system in order to obtain the best performances. Even on the same machine, the timer accuracy can be different between the different languages used (by the way, keep in mind, that web based keystroke dynamics implementation use interpreted languages –java or javascript– which are known to not have a precise timer on all the architectures).

Historically, keystroke dynamics works with a classical keyboard on a computer, and avoids the necessity to buy a specific sensor. However, some studies have been done by using other kinds of sensors in order to capture additional information and improve the recognition. Some works (Eltahir et al., 2008; Grabham & White, 2008) have tested the possibility of using a pressure sensor inside each key of the keyboard. In this case, we can exploit an extra information in order to discriminate more easily the users: the pressure force exerted on the key. Lopatka & Peetz (2009) propose to use a keyboard incorporating a Sudden Motion Sensor (SMS)². Such sensor (or similar ones) is present in recent laptops and is used to detect sudden motion of the computer in order to move the writing heads of the hard drive when a risk of damage of the drive is detected. Lopatka & Peetz use the movement in the *z* axis as information. From these preliminary study, it seems that this information is quite efficient. Sound signals produced by the keyboard typing have also been used in the literature. Nguyen et al. (2010) only use sound signals when typing the password, and obtain indirectly through the analysis of this signal, key-pressed time, key-released time and key-typed forces. Performance is similar to classical keystroke dynamics systems. Dozono et al. (2007) use the sound information in addition to the timing values (*i.e.*, it is a feature fusion) which held better performance than the sound alone, or the timing information alone. Of course, as keystroke dynamics can work with any keyboard, it can also work with any machine providing a keyboard, or something similar to a keyboard. One common machine having a keyboard and owned by a lot of people is the mobile phone where we can use keystroke dynamics on it. We have three kinds of mobile phones:

- Mobile phone with a numerical keyboard. In this case, it is necessary to press several times the same key in order to obtain an alphabetical character. Campisi et al. (2009) present a study on such a mobile phone. They argue that such authentication mechanism must be coupled with another one.
- Mobile phone with all the keys (letters and numbers) accessible with the thumbs. This is a kind of keyboard quite similar to a computer's keyboard. Clarke & Furnell (2007) show its feasibility and highlight the fact that such authentication mechanism can only be used by regular users of mobile phones.

² <http://support.apple.com/kb/HT1935>

- Mobile phone without any keyboard, but a touch screen. We can argue that the two previous mobile phones are already obsolete and will be soon replaced by such kind of mobile phones. Although, there are few studies on this kind of mobile phone, we think the future of keystroke dynamics is on this kind of material. With such a mobile phone, we can capture the pressure information and position of the finger on the key which could be discriminating.

Figure 4 presents the topology of the different keystroke dynamics sensors, while the Figure 5 presents the variability on the timer.

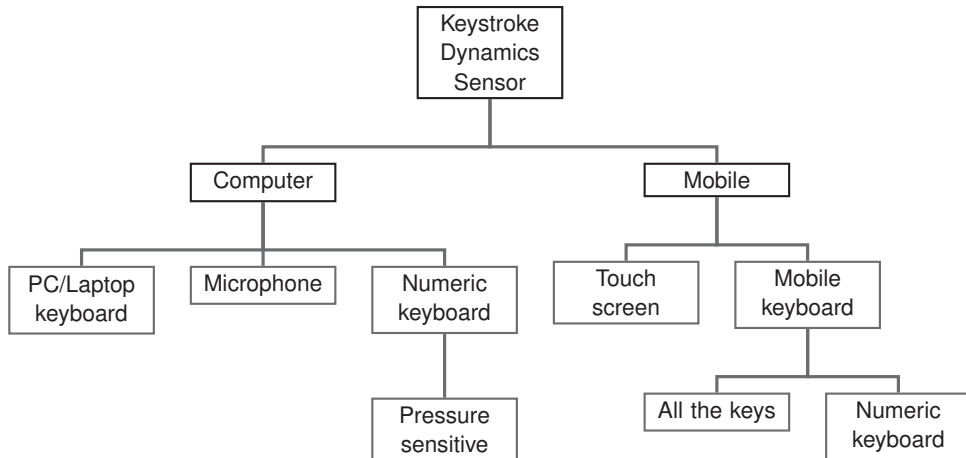


Fig. 4. Topology of keystroke dynamics sensors of the literature

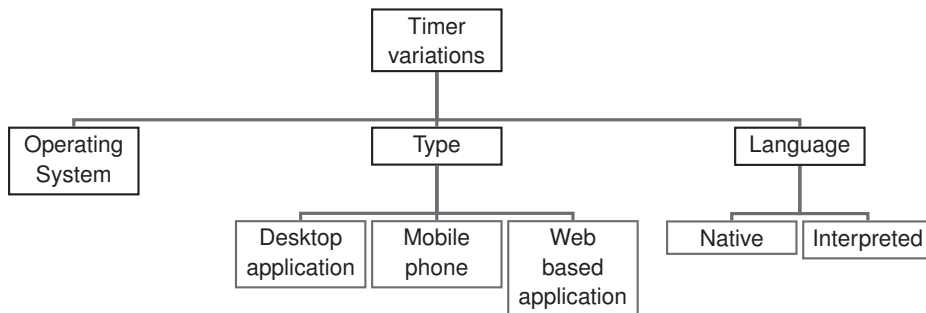


Fig. 5. Topology of factors which may impact the accuracy of the timer

3.2 Captured information

As argued before, various kinds of information can be captured. They mainly depend on the kind of used sensors. Although, we have presented some sensors that are more or less advanced in the previous subsection, we only emphasize, in this chapter, on a classic keyboard.

3.2.1 Raw data

In all the studies, the same raw data is captured (even if they are not manipulated as explained here). We are interested by events on the keyboard. These events are initiated by its user. The raw biometric data, for keystroke dynamics, is a chronologically ordered list of events: the list starts empty, when an event occurs, it is appended at the tail of the list with the following information:

- **Event.** It is generated by an action on the key. There are two different events:
 - **press** occurs when the key is pressed.
 - **release** occurs when the key is released.
- **Key code.** It is the code of the key from which the event occurs. We can obtain the character from this code (in order to verify if the list of characters corresponds to the password, for example). The key code is more interesting than the character, because it gives some information on the location of the key on the keyboard (which can be used by some keystroke dynamics recognition methods) and allows to differentiate different keys giving the same character (which is a discriminant information (Araujo et al., 2005)). This key code may be dependant of the platform and the language used.
- **Timestamp.** It encodes the time when the event occurs. Its precision influence greatly the recognition performance. Pavaday. et al. (2010) propose to use the Windows function *QueryPerformanceCounter*³ with the highest priority enabled for Windows computers, and, changing the scheduler policy to FIFO for Linux machines. It is usually represented in milliseconds, but this is not mandatory.

The raw data can be expressed as (with n the number of events on the form $n = 2 * s$ with s the number of keys pressed to type the text):

$$\begin{cases} (keycode_i, event_i, time_i), \forall i, 0 \leq i < n \\ keycode_i \in \mathbb{Z} \\ event_i \in \{PRESS, RELEASE\} \\ time_i \in \mathbb{N} \end{cases} \quad (1)$$

Umphress & Williams (1985) only use the six first time values of each word (so $s \leq 6$). Depending on the kind of keystroke dynamics application, the raw data is captured in different kind of scenarios: in the authentication form to type the login and password, in a form asking to type a predefined or random text different than the login and password, or in continuous capture during the use of the computer.

3.2.2 Extracted features

Various features can be extracted from this raw data, we present the most commonly used in the literature.

3.2.2.1 First order

The most often extracted features are local ones, computed by subtracting timing values.

- **Duration.** The duration is the amount of time a key is pressed. For the key i (i is omitted for sake of readability) it is computed as following:

$$duration = time\{event = RELEASE\} - time\{event = PRESS\} \quad (2)$$

³ <http://msdn.microsoft.com/en-us/library/ms644904%28v=VS.85%29.aspx>

We then obtain a timing vector (of the size of the typed text), also named **PR** in the literature, containing the duration of each key press (by order of press).

$$\forall i, 1 \leq i \leq n, \quad PR_i = \text{duration}_i \quad (3)$$

- **Latencies.** Different kinds of latencies can be used. They are computed by getting the differences of time between two keys events. We can obtain the **PP** latencies which are the difference of time between the pressure of each key:

$$\forall i, 1 \leq i < n, \quad PP_i = \text{time}_{i+1}\{\text{event}_{i+1} = \text{PRESS}\} - \text{time}_i\{\text{event}_i = \text{PRESS}\} \quad (4)$$

We can obtain the **RR** latencies which are the difference of time between the release of each key:

$$\forall i, 1 \leq i < n, \quad RR_i = \text{time}_{i+1}\{\text{event}_{i+1} = \text{RELEASE}\} - \text{time}_i\{\text{event}_i = \text{RELEASE}\} \quad (5)$$

We can obtain the **RP** latencies which are the difference of time between the release of one key and the pressure of the next one:

$$\forall i, 1 \leq i < n, \quad RP_i = \text{time}_{i+1}\{\text{event}_{i+1} = \text{PRESS}\} - \text{time}_i\{\text{event}_i = \text{RELEASE}\} \quad (6)$$

Most of the time, a feature fusion is operated by concatenating the duration vector with, at least, one of the latency vector (it seems that most of the time, the selected latency vector is the *PP* one, but it is not always indicated in the papers). A recent paper Balagani et al. (2011) discusses on the way of using these extracted features in order to improve the recognition rate of keystroke dynamics systems. Other kinds of data can be encountered in various papers Ilonen (2003). They are mainly global types of information:

- **Total typing.** The total time needed to type the text can also be used. The information can be used as an extra feature to append to the feature vectors, or as a normalisation factor.
- **Middle time.** The time difference between the time when the user types the character at the middle of the password, and the time at the beginning of the input.
- **Mistake ratio.** When the user is authorised to do typing mistakes (this is always the case in continuous authentication, but almost never the case in static authentication), counting the number of times the backspace key is hit gives an interesting feature.

Another concept that is often encountered in the literature, is the notion of *digraph*. A digraph represents the time necessary to hit two keys. The digraph features **D** of a password is computed as following:

$$\forall i, 1 \leq i < n, \quad D_i = \text{time}_{i+1}\{\text{event}_{i+1} = \text{RELEASE}\} - \text{time}_i\{\text{event}_i = \text{PRESS}\} \quad (7)$$

This notion has been extended to *ngraph*, with *n* taking different values. *trigraph* are heavily used in (Bergadano et al., 2002). de Ru & Eloff (1997) use a concept of typing difficulty based on the fact that certain key combinations are more difficult to type than other. The typing difficulty is based on the distance (on the keyboard) between two successive characters (to type), and if several keys are needed to create a character (*i.e.*, use of shift key).

3.2.2.2 Second order

Some features are not extracted from the raw biometric data, but from the first order features.

- **min/max.** It consists to get the minimum and maximum value of each type of data (latency and duration).
- **mean/std.** It consists to get the mean value and its standard deviation of each type of data (latency and duration).
- **Slope.** By using the slope of the biometric sample, we are interested in the global shape of the typing. We expect that users type in the same way even if the speed may be different (Modi & Elliott, 2006). The new features (*result*) set is computed as following (with *source*):

$$\forall i, 1 \leq i < n, \quad result_i = source_{i+1} - source_i \quad (8)$$

- **Entropy.** The entropy inside a sample has been only studied in (Monrose et al., 2002).
- **Spectral information.** Chang (2006a) applies a discrete wavelet transformation to the original extracted features. All the operations are done with the wavelet transformed data.

We can imagine more complicated features, but the final biometric data is always a single vector composed of various features. While computing the model with several samples (see next section) feature selection mechanisms can remove non informative features. We do not insist on papers using other information than timing values in the rest of this chapter (pressure force, movements, ...). We have seen in this section that several features can be extracted. Verification procedures performance greatly depends on the chosen features, but, most of the time, papers only use one latency and the duration.

4. Authentication framework

Once the different biometric data during enrolment procedure have been captured, it is time to build the model of each user. The way of computing it greatly depends on the used verification methods. During an authentication, the verification method compares the query sample (the biometric data captured during the authentication) to the model. Based on the result of this comparison (which is commonly a distance), the decision module accepts or rejects the user.

4.1 Enrolment

The enrolment step allows to create the model of each user, thanks to its enrolled samples. Most of the time, the number of samples used during the enrolment is superior to 20. Such a high quantity of data can be really boring for the users to provide.

4.1.1 Outliers detection

It is known that the classifier performance greatly depends on outliers presence in the learning dataset. Most keystroke dynamics studies do not take care of the presence of outliers in the learning set. Some studies (mainly in free text) remove times superior to a certain threshold. In (Gaines et al., 1980), filtering is done by removing timing values superior to 500ms, while in (Umphress & Williams, 1985) it is timing values superior to 750ms. Rogers & Brown (1996) cleanup data with using a Kohonen network (Kohonen, 1995) using impostors samples. They also use a statistical method. Killourhy & Maxion (2010) also detect outliers in biometric samples. An outlier feature is detected in the following way (for each feature): the feature is more than 1.5 inter-quartile range greater than the third quartile, or more than 1.5

inter-quartile less than the first quartile. When a feature is detected as being an outlier, it is replaced by a random sample (which is not an outlier) selected among possible values of this feature for this user. The procedure is operated for each feature of each sample. By this way, the number of samples is always the same.

It seems that, most of the time, the outlier detection and correction is operated on the whole dataset, and not on the learning set. This allows to cleanup the used dataset to compute the algorithm performance (and obtain better performance), but not the enrolled samples of the user.

4.1.2 Preprocessing

Biometric data may be normalized before being used. Such pre-processing allows to get better performance by using a normalisation function (Filho & Freire (2006) observed that the timing distribution is roughly Log-Normal) :

$$g(x) = \frac{1}{1 + \exp\left(-\frac{K(\log_e(x) - \mu)}{\sigma}\right)} \quad (9)$$

We did not find other references to other pre-processing approaches in the literature. The parameters K (k is chosen in order to minimise the squared error between the approximated function and the cumulative distribution function of the logarithm of timings distribution), μ and σ respectively represent an optimisation factor, the mean of the logarithm of the timing values, the standard deviation of the logarithm of the timing values.

4.1.3 Feature selection

A feature selection mechanism can be applied to remove irrelevant features. It seems that this point has also been rarely tested. The aim of the feature selection is to reduce the quantity of data and speed up the computation time, and, eventually to improve the performance. Very few studies have applied such kind of mechanism. Two different kinds of feature extraction systems can be used:

- *Filter* approach which does not depend on the verification algorithm. The aim is to remove irrelevant features based on different measures (e.g., the variance);
- *Wrapper* approach which depends on the verification algorithm. Different feature subsets are generated and evaluated. The best one is kept.

Boechat et al. (2006) select a subset of N features with the minors of standard deviation, which allows to eliminate less significant features. Experiments are done at Zero False Acceptance Rate. False Rejection Rate reduces when the number of selected features increases. Keeping 70% of the features gives interesting results. Azevedo et al. (2007) use a wrapper system based on Particle Swarm Optimization (PSO) to operate the feature selection. The PSO gives better results than a Genetic Algorithm. Bleha & Obaidat (1991) use a reduction technique based on Fisher analysis. However, the technique consists in keeping $m - 1$ dimension for each vector, with m the number of users in the system (they have only 9 users in their system). Yu & Cho (2004) use an algorithm based on Support Vector Machines (SVM) and Genetic Algorithms (GA) to reduce the size of samples and keep only key values for each user. Other similar methods are present in the literature (Chen & Lin, 2005).

4.1.4 Model computation

There are numbers of methods to verify if a query corresponds to the expected user. Some of them are based on statistical methods, other on data mining methods. Some methods use one-class assumption (they only use the enrolment samples of the user), while other use two-class or multi-class assumption (they also use impostors enrollment samples to compute the model). When impostors samples are needed, they may be automatically generated (Sang et al., 2004), instead of being collected with real impostors (Clarke & Furnell, 2006; Obaidat & Sadoun, 1997). Generally, data mining methods use a really huge number of enrolled samples to compute the model (several hundred of samples in some neural network methods) which is not realistic at all. Most used way of model computing are:

- Computing the mean vector and standard deviation of enrolled samples (Umphress & Williams, 1985);
- Store the enrolled vectors in order to use them with k nearest neighbour methods (Rao, 2005) (variations being in the distance computing method (Kang & Cho, 2009));
- Learning of bayesian classifiers (Janakiraman & Sim, 2007; Rao, 2005);
- Learning clusters with k -mean (Hwang et al., 2006; Obaidat & Sadoun, 1997) ;
- Learning parameters of generative functions: Hidden Markov Model (HMM) (Galassi et al., 2007; Poha et al., 2009; Rodrigues et al., 2006) or Gaussian Mixture Models (GMM) (Hosseinzadeh & Krishnan, 2008) ;
- Neural network learning (Bartmann et al., 2007; Clarke & Furnell, 2006; Obaidat & Sadoun, 1997; Rogers & Brown, 1996) ;
- SVM learning (Giot et al., 2009c; Rao, 2005; Sang et al., 2004; Yu & Cho, 2004).

4.2 Verification

The verification consists in verifying if the input of the user corresponds to the claimed identity. The way of capturing these inputs greatly depends on the kind of used keystroke dynamics system (*e.g.*, for static authentication, the user must type its login and password). While the features are extracted from the raw biometric sample (same procedure than during the enrollment), they are compared to the model of the claimed user. Usually, the verification module returns a comparison score. If this score is below than a predefined threshold, the user is authenticated, otherwise, he is rejected. Several verification methods exist and depend on the way the enrollment is done, so they are similar to the present list. *query* represents the query biometric sample (the test capture to compare to the model). $\|\cdot\|_p$ represents the p norm of vector. The main families of computing are (Güven & Sogukpinar, 2003):

- The minimal distance computing.

In (Monrose & Rubin, 1997), the euclidean distance between the query and each of enrolled samples is computed. The comparison score is the min of these distances.

$$score = \min \|query - enrolled_u\|_2, \forall u \in [1, Card(enrolment)] \quad (10)$$

- The statistical methods.

One of the oldest methods is based on bayesian probabilities (Bleha et al., 1990). μ is the mean value of enrolled samples:

$$score = \frac{(query - \mu)^t (query - \mu)}{\|query\|_2 \cdot \|\mu\|_2} \quad (11)$$

A normalized version is also presented in the study. The statistical method presented in (Hocquet et al., 2006) computes the score depending on the mean μ and the standard deviation σ of the enrolled samples:

$$score = 1 - \frac{1}{Card(query)} \left\| \exp \left(-\frac{|query - \mu|}{\sigma} \right) \right\|_1 \quad (12)$$

Filho & Freire (2006) present another method which also computes a distance:

$$score = \|query - \mu\|_2^2 \quad (13)$$

- Application of fuzzy rules de Ru & Eloff (1997).
- Class verification.
For classifiers able to give a label, the verification consists in verifying if the guessed label corresponds to the label of the claimed identity (cf. neural networks, SVM, $k - nn$).
- Some methods are based on the disorder degree of vectors (Bergadano et al., 2002).
- Others are based on timing discretisation (Hocquet et al., 2006).
- Bioinformatic methods based on string motif searching are also used (Revett et al., 2007).

4.3 Improving the performance

Different ways can be used to improve the performance of the recognition. Several studies (Bartmann et al., 2007; Hosseinzadeh & Krishnan, 2008; Killourhy & Maxion, 2010; Revett, 2009) request the user to type the verification text several times (mainly between two and three), when he is rejected, in order to give him more chances of being verified. Such procedure reduces the False Rejection Rate without growing the False Acceptance Rate too much. Other studies try to update the model of a user after being authenticated (Hosseinzadeh & Krishnan, 2008; Revett, 2009). This way, the model tracks the behavior modifications of the user through time, and integrate them in the model. As the keystroke data deviates progressively with time, performance degrades with time when not using such procedure. It is not always clear in the various studies if the template update is done in a supervised way (impostors samples never added), or in a semi-supervised way (samples added if the classifier recognizes them as being genuine). Even if the aim is to improve performance, the result can be totally different: semi-supervised methods may add impostor samples in the model. This way, the model deviates from the real biometric data of the user and attracts more easily impostors samples. Classifier performances greatly depend on the number of used samples to compute them. Chang (2006b) artificially generates new samples from the enrolled samples in order to improve keystroke recognition. The system uses a transformation in frequential domains thanks to wavelets. Another way to improve recognition performance is to fuse two samples together (Bleha & Obaidat, 1991). This way, timing values are smoothed when merging the two samples and light hesitation are suppressed. The fusion (Ross et al., 2006) of several keystroke dynamics methods on the same query is also a good way to improve performances:

- Bleha et al. (1990) associate a bayesian classifier to a minimal distance computing between the query vector and the model.
- Hocquet et al. (2006) apply a fusion between three different keystroke dynamics methods, which greatly improves the performance.
- Different kinds of weighted sums score fusion functions are proposed in Giot, El-Abed & Rosenberger (2010); Teh et al. (2007).

Keystroke dynamics has also been successfully fused with other modalities, like face (Giot, Hemery & Rosenberger, 2010) or speaker recognition (Montalvao Filho & Freire, 2006). Hwang et al. (2006) have defined various measures to get the unicity, consistency and discriminability. By analysing the behavior of these measures comparing the recognition performance, they find that it is possible to improve performance by asking users to artificially add pauses (helped by cues for being synchronized) when typing the password. Karnan et al. (2011) propose an interesting review of most of the keystroke dynamics recognition methods.

4.4 User identification

The verification consists in verifying if the identity of the claimant is correct, while the identification consists to determine the identity of the user. We may find methods specific to identification, or compare the query to each model, the identity being the owner of the model returning the lowest distance (or a reject if this distance is higher a threshold). Bleha et al. (1990) use a bayesian classifier to identify the user. Identification based on keystroke dynamics has not been much experimented in the literature.

5. Evaluation of keystroke dynamics systems

Despite the obvious advantages of keystroke dynamics systems in enhancing traditional methods based on a *secret*, its proliferation is still not as much as expected. The main drawback is notably the lack of a generic evaluation method for such systems. We need a reliable evaluation methodology in order to put into obviousness the benefit of a new method. Nowadays, several studies exist in the state-of-the-art to evaluate keystroke dynamics systems. It is generally realized within three aspects: performance, satisfaction and security.

5.1 Performance

The goal of this evaluation aspect is to quantify and to compare keystroke dynamics systems. In order to compare these systems, we need generally to compute their performance using a predefined protocol (acquisition conditions, test database, performance metrics, ...). According to the International Organization for Standardization ISO/IEC 19795-1 (2006), the performance metrics are divided into three sets:

- Acquisition performance metrics such as the Failure-To-Enroll rate (FTE).
- Verification system performance metrics such as the Equal Error Rate (EER).
- Identification system performance metrics such as the False-Negative and the False-Positive Identification Rates (FNIR and FPIR, respectively).

Several benchmark databases exist in order to compare keystroke dynamics systems. A benchmark database can contain real samples from individuals, which reflect the best the real use cases. Nevertheless, it is costly in terms of efforts and time to create such a database. As argued by Cherifi et al. (2009), a good benchmark database must satisfy various requirements:

1. As keystroke dynamics is a behavioral modality, the database must be captured among different sessions, with a reasonable time interval between sessions, in order to take into account the variation of individuals behavior.
2. The database must also contain fake biometric templates to test the robustness of the system. It seems that there is no other reference to this kind of experiment in the literature.
3. The benchmark must embed a large diversity of users (culture, age, ...). This point is essential for any biometrics, but, it is really difficult to attain.

We present an overview of the existing benchmark databases:

DB 1 Chaves

Montalvão *et al.* have used the same keystroke databases in several papers (Filho & Freire, 2006). The databases are available at <http://itabi.infonet.com.br/biochaves/br/download.htm>. The databases do not seem to be yet available on their website. The maximum number of users in a database is 15, and, the number of provided samples per user is 10. Each database contains the raw data. The database is composed of couples of ASCII code of the pressed key and the elapsed time since the last key down event. Release of a key is not tracked. Four different databases have been created. Most databases were built under two different sessions spaced of one week or one month (depending on the database). Each database is stored in raw text files.

DB 2 DSN2009

Killourhy & Maxion (2009) propose a database of 51 users providing four hundred samples captured in height sessions (there are fifty inputs per session). The delay between each session is one one day at minimum, but the mean value is not stated. This is the dataset having the most number of samples per user, but, a lot of them are typed on a short period (50 at the same time). Each biometric data has been captured when typing the following password: “.tie5Roanl”. The database contains some extracted features: hold time, interval between two pressures, interval between the release of a key, and the pressure of the next one. The database is available at <http://www.cs.cmu.edu/~keystroke/>. It is stored in raw text, csv or Excel files.

DB 3 Greyc alpha

Giot *et al.* (2009a) propose the most important public dataset in term of users. It contains 133 users and, 100 of them provided samples of, at least, five distinct sessions. Each user typed the password “greyc laboratory” twelve times, on two distinct keyboards, during each session (which give 60 samples for the 100 users having participated to each session). Both extracted features (hold time and latencies) and raw data are available (which allow to build other extracted features). The database is available at <http://www.ecole.ensicaen.fr/~rosenber/keystroke.html>. It is stored in an sqlite database file.

DB 4 Pressure-Sensitive Keystroke Dynamics Dataset

Allen (2010) has created a public keystroke dynamics database using a pressure sensitive keyboard. The database is available at <http://jddesign.net/2010/04/pressure-sensitive-keystroke-dynamics-dataset/> in a csv or sql file. It embeds the following raw data: key code, time when pressed, time when release, pressure force. 104 users are present on the database, but, only 7 of them provided a significant amount of data (between 89 and 504), whereas the 97 other have only provided between 3 and 15 samples. Three different passwords have been typed: “pr7q1z”, “jeffrey allen” and “drizzle”.

DB 5 Fixed Text

The most recent database has been released in 2010 Bello et al. (2010). 58 volunteers participated to the experiment. Each session consists in typing 14 phrases extracted from books and 15 common UNIX commands. It seems that almost all the users have done only one session. The database is available at <http://www.citefa.gov.ar/si6/k-profiler/dataset/> in a raw text file. Press and release times for each key are saved, as well as the user agent of the browser from which the session has been done, the age, gender and handedness of the user and other information.

We can see that some databases are available. Each of them has been created for keystroke dynamics on computer (*i.e.* no public dataset available for smartphones). Despite this, these databases do not always fit the previous requirements, which may explain why none of them have been used by researchers different than their creators. Although, it would be the best kind of dataset, no public dataset has been built with one login/password different for each user. Table 3 presents a summary of these public datasets.

Dataset	Type	Information	Users	Samples /users	Sessions
Filho & Freire (2006)	Various	Press events	< 15	< 10	2
Killourhy & Maxion (2009)	1 fixed string	Duration and 2 latencies	51	400	8
Giot et al. (2009a)	1 fixed String	Press and release events. Duration and 3 latencies	> 100	60	5
Allen (2010)	3 fixed strings	Press and release events and pressure	7/97	(89-504)/(3-15)	few months
Bello et al. (2010)	14 phrases and 15 unix commands	Press and release time	58	1	1

Table 3. Summary of keystroke dynamics datasets

Most of the proposed keystroke dynamics methods in the literature have quantified their methods using different protocols for their data acquisition (Giot et al., 2009c; Killourhy & Maxion, 2009). Table 4 illustrates the differences of the used protocols in this research area for some major studies. The performance comparison of these methods is quite impossible, as stated in (Crawford, n.d.; Giot et al., 2009a; Karnan et al., 2011; Killourhy & Maxion, 2009), due to several reasons. First, most of these studies have used different protocols for their data acquisition, which is totally understandable due to the existence of different kinds of keystroke dynamics systems (static, continuous, dynamic) that require different acquisition protocols. Second, they differ on the used database (number of individuals, separation between sessions . . .), the *acknowledgement* of the password (if it is an imposed password, a high *FTA* is expected), the used keyboards (which may deeply influences the way of typing), and the use of different or identical passwords (which impacts on the quality of impostors' data). In order to resolve such problematic, Giot et al. (2011) presents a comparative study of seven methods (1 contribution against 6 methods existing in the literature) using a predefined protocol, and GREYC alpha database (Giot et al., 2009a). The results from this study show a promising EER value equal to to 6.95%. To our knowledge, this is the only work that compares

keystroke methods within the same protocol, and using a publicly available database. The performance of keystroke dynamics systems (more general speaking, of behavioral systems) provides a lower quality than the morphological and biological ones, because they depend a lot on user's feelings at the moment of the data acquisition: user may change his way of performing tasks due to its stress, tiredness, concentration or illness. Previous works presented by Cho & Hwang (2006); Hwang et al. (2006) focus on improving the quality of the captured keystroke features as a mean to enhance system overall performance. Hwang et al. (2006) have employed pauses and cues to improve the uniqueness and consistency of keystroke features. We believe that it is relevant to more investigate the quality of the captured keystroke features, in order to enhance the performance of keystroke dynamics systems.

Paper	A	B	C	D	E	FAR	FRR
Obaidat & Sadoun (1997)	8 weeks	15	112	no	no	0%	0%
Bleha et al. (1990)	8 weeks	36	30	yes	yes	2.8%	8.1%
Rodrigues et al. (2006)	4 sessions	20	30	/	no	3.6%	3.6%
Hocquet et al. (2007)	/	38	/	/	no	1.7%	2.1%
Revett et al. (2007)	14 days	30	10	/	no	0.15%	0.2%
Hosseinzadeh & Krishnan (2008)	/	41	30	no	no	4.3%	4.8%
Monrose & Rubin (1997)	7 weeks	42	/	no	no	/	20%
Revett et al. (2006)	4 weeks	8	12	/	/	5.58%	5.58%
Killourhy & Maxion (2009)	8 sessions	51	200	yes	no	9.6%	9.6%
Giot et al. (2009c)	5 sessions	100	5	yes	no	6.96%	6.96%

Table 4. Summary of the protocols used for different studies in the state-of-the-art (A: Duration of the database acquisition, B: Number of individuals in the database, C: Number of samples required to create the template, D: Is the acquisition procedure controlled?, E: Is the threshold global?). "/" indicates that no information is provided in the article.

5.2 Satisfaction

This evaluation aspect focuses on measuring users' acceptance and satisfaction regarding the system (Theofanos et al., 2008). It is generally measured by studying several properties such as easiness to use, trust in the system, *etc.* The works done by El-Abed et al. (2010); Giot et al. (2009b) focusing on studying users' acceptance and satisfaction of a keystroke dynamics system (Giot et al., 2009a), show that the system is well perceived and accepted by the users. Figure 7 summarizes users' acceptance and satisfaction while using the tested system. Satisfaction factors are rated between 0 and 10 (0 : not satisfied · · · 10 : quite satisfied). These results show that the tested system is well perceived among the five acceptance and satisfaction properties. Moreover, there were no concerns about privacy issues during its use. In biometrics, there is a potential concern about the misuse of personal data (*i.e.*, templates) which is seen as violating users' privacy and civil liberties. Hence, biometric systems respecting this satisfaction factor are considered as useful.

5.3 Security

Biometric authentication systems present several drawbacks which may considerably decrease their security. Schneier (1999) compares traditional security systems with biometric systems. The study presents several drawbacks of biometric systems including:

- The lack of secrecy: everybody knows our biometric traits such as iris,
- and, the fact that a biometric trait cannot be replaced if it is compromised.

El-Abed et al. (2011) propose an extension of the Ratha *et al.* model (Ratha et al., 2001) to categorize the common threats and vulnerabilities of a generic biometric system. Their proposed model is divided into two sets as depicted in figure 6: architecture threats and system overall vulnerabilities.

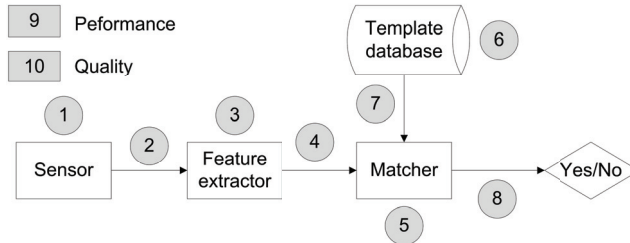


Fig. 6. Vulnerability points in a general biometric system.

5.3.1 Set I architecture threats

- 1) Involves presenting a fake biometric data to the sensor. An example of such attack is the zero-effort attempts. Usually, attackers try to impersonate legitimate users having weak templates;
- 2) and 4) In a replay attack, an intercepted biometric data is submitted to the feature extractor or the matcher bypassing the sensor. Attackers may collect then inject previous keystroke events features using a keylogger;
- 3) and 5) The system components are replaced with a Trojan horse program that functions according to its designer specifications;
- 6) Involves attacks on the template database such as modifying or suppressing keystroke templates;
- 7) The keystroke templates can be altered or stolen during the transmission between the template database and the matcher;
- 8) The matcher result (accept or reject) can be overridden by the attacker.

5.3.2 Set II system overall vulnerabilities

9) Performance limitations

By contrast to traditional authentication methods based on “what we know” or “what we own” (0% comparison error), biometric systems is subject to errors such as False Acceptance Rate (FAR) and False Rejection Rate (FRR). This inaccuracy illustrated by statistical rates would have potential implications regarding the level of security provided by a biometric system. Doddington et al. (1998) assign users into four categories:

- Sheep: users who are recognized easily (contribute to a low FRR),
- Lambs: users who are easy to imitate (contribute to a high FAR),
- Goats: users who are difficult to recognize (contribute to a high FRR), and
- Wolves: users who have the capability to spoof the biometric characteristics of other users (contribute to a high FAR).

A poor biometric in term of performance, may be easily attacked by lambs, goats and wolves users. There is no reference to this user classification in the keystroke dynamics literature. Therefore, it is important to take into consideration system performance within the security evaluation process. The Half Total Error Rate (HTER) may be used as an illustration of system overall performance. It is defined as the mean of both error rates FAR and FRR:

$$HTER = \frac{FAR + FRR}{2} \quad (14)$$

10) Quality limitations during enrollment

The quality of the acquired biometric samples is considered as an important factor during the enrollment process. The absence of a quality test increases the possibility of enrolling authorized users with weak templates. Such templates increase the probability of success of zero-effort impostor, hill-climbing and brute force (Martinez-Diaz et al., 2006) attempts. Therefore, it is important to integrate such information within the security evaluation process. In order to integrate such information, a set of rules is presented in (El-Abed et al., 2011).

According to the International Organization for Standardization ISO/IEC FCD 19792 (2008), the security evaluation of biometric systems is generally divided into two complementary assessments:

1. Assessment of the biometric system (devices and algorithms), and
2. Assessment of the environmental (for example, is the system is used indoor or outdoor?) and operational conditions (for example, tasks done by system administrators to ensure that the claimed identities during enrollment of the users are valid).

A type-1 security assessment of a keystroke dynamics system (Giot et al., 2009a) is presented in El-Abed et al. (2011). The presented method is based on the use of a database of common threats and vulnerabilities of biometric systems, and the notion of risk factor. A risk factor, for each identified threat and vulnerability, is considered as an indicator of its importance. It is calculated using three predefined criteria (effectiveness, easiness and cheapness) and is defined between 0 and 1000. More the risk factor is near 0, better is the robustness of the Target of Evaluation (ToE). Figure 7 summarizes the security assessment of the TOE, which illustrates the risk factors of the identified threats and system overall vulnerabilities among the ten assessment points (the maximal risk factor is retained from each point).

5.4 Discussion

The evaluation of keystroke dynamics modality are very few in comparison to other types of modalities (such as fingerprint modality). As shown in section 5.1, there is only a few public databases that could be used to evaluate keystroke dynamics authentication systems. There is none competition neither existing platform to compare such behavioral modality. The results presented in the previous section show that the existing keystroke dynamics methods provide promising recognition rates, and such systems are well perceived and accepted by users. In our opinion, we believe that keystroke dynamics systems belong to the possible candidates that may be implemented in an Automated Teller Machine (ATM), and can be widely used for e-commerce applications.

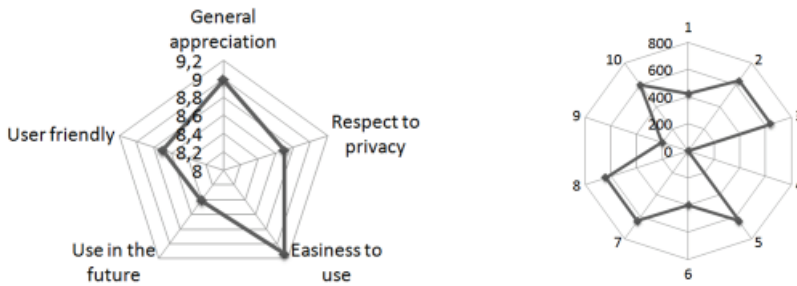


Fig. 7. Satisfaction (on the left) and security (on the right) assessment of a keystroke dynamics based system.

6. Conclusion and future trends

We have presented in this chapter an overview of keystroke dynamics literature. More information on the subject can be found in various overviews: Revett (2008, chapter 4) deeply presents some studies. We believe that the future of the keystroke dynamics is no more on desktop application, whereas it is the most studied in the literature, but in the *mobile* and *internet* worlds, because mobile phones are more popular than computers and its use is very democratized. They are more and more powerful every year (in terms of calculation and memory) and embeds interesting sensors (pressure information with tactile phones). Mobile phone owners are used to use various applications on their mobile and they will probably agree to lock them with a keystroke dynamics biometric method. Nowadays, more applications are available in a web browser. These applications use the classical couple of login and password to verify the identity of a user. Integrating them a keystroke dynamics verification would harden the authentication process. In order to spread the keystroke modality, it is necessary to solve various problems related to:

- The cross devices problem. We daily use several computers which can have different keyboards on timing resolution. These variability must not have an impact on the recognition performances. Users tend to change often their mobile phone. In an online authentication scheme (were the template is stored on a server), it could be useful to not re-enroll the user on its new mobile phone.
- The aging of the biometric data. Keystroke dynamics, is subject to a lot of intra class variability. One of the main reasons is related to the problem of template aging: performances degrade with time because user (or impostors) type differently with time.

7. Acknowledgment

The authors would like to thank the Lower Normandy Region and the French Research Ministry for their financial support of this work.

8. References

Ahmed, A. & Traore, I. (2008). *Handbook of Research on Social and Organizational Liabilities in Information Security*, Idea Group Publishing, chapter Employee Surveillance based on Free Text Detection of Keystroke Dynamics, pp. 47–63.

- Allen, J. D. (2010). *An analysis of pressure-based keystroke dynamics algorithms*, Master's thesis, Southern Methodist University, Dallas, TX.
- Araujo, L., Sucupira, L.H.R., J., Lizarraga, M., Ling, L. & Yabu-Uti, J. (2005). User authentication through typing biometrics features, *IEEE Transactions on Signal Processing* 53(2 Part 2): 851–855.
- Azevedo, G., Cavalcanti, G., Carvalho Filho, E. & Recife-PE, B. (2007). An approach to feature selection for keystroke dynamics systems based on pso and feature weighting, *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*.
- Balagani, K. S., Phoha, V. V., Ray, A. & Phoha, S. (2011). On the discriminability of keystroke feature vectors used in fixed text keystroke authentication, *Pattern Recognition Letters* 32(7): 1070 – 1080.
- Bartmann, D., Bakdi, I. & Achatz, M. (2007). On the design of an authentication system based on keystroke dynamics using a predefined input text, *Techniques and Applications for Advanced Information Privacy and Security: Emerging Organizational, Ethical, and Human Issues* 1(2): 149.
- Bello, L., Bertacchini, M., Benitez, C., Carlos, J., Pizzoni & Cipriano, M. (2010). Collection and publication of a fixed text keystroke dynamics dataset, *XVI Congreso Argentino de Ciencias de la Computacion (CACIC 2010)*.
- Bergadano, F., Gunetti, D. & Picardi, C. (2002). User authentication through keystroke dynamics, *ACM Transactions on Information and System Security (TISSEC)* 5(4): 367–397.
- Bleha, S. & Obaidat, M. (1991). Dimensionality reduction and feature extraction applications in identifying computer users, *IEEE transactions on systems, man and cybernetics* 21(2): 452–456.
- Bleha, S., Slivinsky, C. & Hussien, B. (1990). Computer-access security systems using keystroke dynamics, *IEEE Transactions On Pattern Analysis And Machine Intelligence* 12 (12): 1216–1222.
- Boechat, G., Ferreira, J. & Carvalho, E. (2006). Using the keystrokes dynamic for systems of personal security, *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 18, pp. 200–205.
- Campisi, P., Maiorana, E., Lo Bosco, M. & Neri, A. (2009). User authentication using keystroke dynamics for cellular phones, *Signal Processing, IET* 3(4): 333–341.
- Chang, W. (2006a). Keystroke biometric system using wavelets, *ICB 2006*, Springer, pp. 647–653.
- Chang, W. (2006b). Reliable keystroke biometric system based on a small number of keystroke samples, *Lecture Notes in Computer Science* 3995: 312.
- Chen, Y.-W. & Lin, C.-J. (2005). Combining svms with various feature selection strategies, *Technical report*, Department of Computer Science, National Taiwan University, Taipei 106, Taiwan.
- Cherifi, F., Hemery, B., Giot, R., Pasquet, M. & Rosenberger, C. (2009). *Behavioral Biometrics for Human Identification: Intelligent Applications*, IGI Global, chapter Performance Evaluation Of Behavioral Biometric Systems, pp. 57–74.
- Cho, S. & Hwang, S. (2006). Artificial rhythms and cues for keystroke dynamics based authentication, *In International Conference on Biometrics (ICB)*, pp. 626–632.
- Clarke, N. & Furnell, S. (2006). Advanced user authentication for mobile devices, *computers & security* 27: 109–119.

- Clarke, N. L. & Furnell, S. M. (2007). Authenticating mobile phone users using keystroke analysis, *International Journal of Information Security* 6: 1–14.
- Conklin, A., Dietrich, G. & Walz, D. (2004). Password-based authentication: A system perspective, *Proceedings of the 37th Hawaii International Conference on System Sciences, Hawaii*.
- Crawford, H. (n.d.). Keystroke dynamics: Characteristics and opportunities, *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on, IEEE*, pp. 205–212.
- de Magalhaes, T., Revett, K. & Santos, H. (2005). Password secured sites: stepping forward with keystroke dynamics, *International Conference on Next Generation Web Services Practices*.
- de Ru, W. G. & Eloff, J. H. P. (1997). Enhanced password authentication through fuzzy logic, *IEEE Expert: Intelligent Systems and Their Applications* 12: 38–45.
- Doddington, G., Liggett, W., Martin, A., Przybocki, M. & Reynolds, D. (1998). Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation, *ICSLP98*.
- Dozono, H., Itou, S. & Nakakuni, M. (2007). Comparison of the adaptive authentication systems for behavior biometrics using the variations of self organizing maps, *International Journal of Computers and Communications* 1(4): 108–116.
- El-Abed, M., Giot, R., Hemery, B. & Rosenberger, C. (2010). A study of users' acceptance and satisfaction of biometric systems, *44th IEEE International Carnahan Conference on Security Technology (ICCST)*.
- El-Abed, M., Giot, R., Hemery, B., Schwartzmann, J.-J. & Rosenberger, C. (2011). Towards the security evaluation of biometric authentication systems, *IEEE International Conference on Security Science and Technology (ICSST)*.
- Eltahir, W., Salami, M., Ismail, A. & Lai, W. (2008). Design and Evaluation of a Pressure-Based Typing Biometric Authentication System, *EURASIP Journal on Information Security, Article ID 345047(2008)*: 14.
- Epp, C. (2010). *Identifying emotional states through keystroke dynamics*, Master's thesis, University of Saskatchewan, Saskatoon, CANADA.
- Filho, J. R. M. & Freire, E. O. (2006). On the equalization of keystroke timing histograms, *Pattern Recognition Letters* 27: 1440–1446.
- Gaines, R., Lisowski, W., Press, S. & Shapiro, N. (1980). Authentication by keystroke timing: some preliminary results, *Technical report*, Rand Corporation.
- Galassi, U., Giordana, A., Julien, C. & Saitta, L. (2007). Modeling temporal behavior via structured hidden markov models: An application to keystroking dynamics, *Proceedings 3rd Indian International Conference on Artificial Intelligence (Pune, India)*.
- Giot, R., El-Abed, M. & Chri (2011). Unconstrained keystroke dynamics authentication with shared secret, *Computers & Security* pp. 1–20. [in print].
- Giot, R., El-Abed, M. & Rosenberger, C. (2009a). Greyc keystroke: a benchmark for keystroke dynamics biometric systems, *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, IEEE Computer Society, Washington, District of Columbia, USA, pp. 1–6.
- Giot, R., El-Abed, M. & Rosenberger, C. (2009b). Keystroke dynamics authentication for collaborative systems, *International Symposium on Collaborative Technologies and Systems*, pp. 172–179.
- Giot, R., El-Abed, M. & Rosenberger, C. (2009c). Keystroke dynamics with low constraints svm based passphrase enrollment, *IEEE International Conference on Biometrics: Theory,*

- Applications and Systems (BTAS 2009)*, IEEE Computer Society, Washington, District of Columbia, USA, pp. 1–6.
- Giot, R., El-Abed, M. & Rosenberger, C. (2010). Fast learning for multibiometrics systems using genetic algorithms, *The International Conference on High Performance Computing & Simulation (HPCS 2010)*, IEEE Computer Society, Caen, France, pp. 1–8.
- Giot, R., Hemery, B. & Rosenberger, C. (2010). Low cost and usable multimodal biometric system based on keystroke dynamics and 2d face recognition, *IAPR International Conference on Pattern Recognition (ICPR)*, IAPR, Istanbul, Turkey, pp. 1128–1131. Acceptance rate: 54/100.
- Giot, R. & Rosenberger, C. (2011). A new soft biometric approach for keystroke dynamics based on gender recognition, *Int. J. of Information Technology and Management (IJITM), Special Issue on: "Advances and Trends in Biometric"* pp. 1–17. [in print].
- Grabham, N. & White, N. (2008). Use of a novel keypad biometric for enhanced user identity verification, *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*, pp. 12–16.
- Guven, A. & Sogukpinar, I. (2003). Understanding users' keystroke patterns for computer access security, *Computers & Security* 22(8): 695–706.
- Hocquet, S., Ramel, J.-Y. & Cardot, H. (2006). Estimation of user specific parameters in one-class problems, *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, pp. 449–452.
- Hocquet, S., Ramel, J.-Y. & Cardot, H. (2007). User classification for keystroke dynamics authentication, *The Sixth International Conference on Biometrics (ICB2007)*, pp. 531–539.
- Hosseinzadeh, D. & Krishnan, S. (2008). Gaussian mixture modeling of keystroke patterns for biometric applications, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 38(6): 816–826.
- Hwang, S.-s., Lee, H.-j. & Cho, S. (2006). Improving authentication accuracy of unfamiliar passwords with pauses and cues for keystroke dynamics-based authentication, *Intelligence and Security Informatics* 3917: 73–78.
- Ilonen, J. (2003). Keystroke dynamics, *Advanced Topics in Information Processing–Lecture* .
- ISO/IEC 19795-1 (2006). Information technology biometric performance testing and reporting, *Technical report*, International Organization for Standardization ISO/IEC 19795-1.
- ISO/IEC FCD 19792 (2008). Information technology – security techniques – security evaluation of biometrics, *Technical report*, International Organization for Standardization ISO/IEC FCD 19792.
- Janakiraman, R. & Sim, T. (2007). Keystroke dynamics in a general setting, *Lecture notes in computer science* 4642: 584.
- Kang, P. & Cho, S. (2009). A hybrid novelty score and its use in keystroke dynamics-based user authentication, *Pattern Recognition* p. 30.
- Karnan, M., Akila, M. & Krishnaraj, N. (2011). Biometric personal authentication using keystroke dynamics: A review, *Applied Soft Computing* 11(2): 1565 – 1573. The Impact of Soft Computing for the Progress of Artificial Intelligence.
- Khanna, P. & Sasikumar, M. (2010). Recognising Emotions from Keyboard Stroke Pattern, *International Journal of Computer Applications IJCA* 11(9): 24–28.
- Killourhy, K. & Maxion, R. (2008). The effect of clock resolution on keystroke dynamics, *Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection*, Springer, pp. 331–350.

- Killourhy, K. & Maxion, R. (2009). Comparing anomaly-detection algorithms for keystroke dynamics, *IEEE/IFIP International Conference on Dependable Systems & Networks, 2009. DSN'09*, pp. 125–134.
- Killourhy, K. & Maxion, R. (2010). Keystroke biometrics with number-pad input, *IEEE/IFIP International Conference on Dependable Systems & Networks, 2010. DSN'10*.
- Kohonen, T. (1995). Self-organising maps, *Springer Series in Information Sciences* 30.
- Lopatka, M. & Peetz, M. (2009). Vibration sensitive keystroke analysis, *Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning*, pp. 75–80.
- Marsters, J.-D. (2009). *Keystroke Dynamics as a Biometric*, PhD thesis, University of Southampton.
- Martinez-Diaz, M., Fierrez-Aguilar, J., Alonso-Fernandez, F., Ortega-Garcia, J. & Siguenza, J. (2006). Hill-climbing and brute force attacks on biometric systems: a case study in match-on-card fingerprint verification, *Proceedings of the IEEE of International Carnahan Conference on Security Technology (ICCST)*.
- Modi, S. K. & Elliott, S. J. (2006). Kesyroke dynamics verification using spontaneously generated password, *IEEE International Carnahan Conferences Security Technology*.
- Monrose, F., Reiter, M. & Wetzel, S. (2002). Password hardening based on keystroke dynamics, *International Journal of Information Security* 1(2): 69–83.
- Monrose, F. & Rubin (1997). Authentication via keystroke dynamics, *Proceedings of the 4th ACM conference on Computer and communications security*, ACM Press New York, NY, USA, pp. 48–56.
- Monrose, F. & Rubin, A. (2000). Keystroke dynamics as a biometric for authentication, *Future Generation Computer Syststems* 16(4): 351–359.
- Montalvao Filho, J. & Freire, E. (2006). Multimodal biometric fusion–joint typist (keystroke) and speaker verification, *Telecommunications Symposium, 2006 International*, pp. 609–614.
- Nguyen, T., Le, T. & Le, B. (2010). Keystroke dynamics extraction by independent component analysis and bio-matrix for user authentication, in B.-T. Zhang & M. Orgun (eds), *PRICAI 2010: Trends in Artificial Intelligence*, Vol. 6230 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 477–486.
- Obaidat, M. & Sadoun, B. (1997). Verification of computer users using keystroke dynamics, *Systems, Man and Cybernetics, Part B, IEEE Transactions on* 27(2): 261–269.
- Pavaday, N., ., S. S. & Nugessur, S. (2010). Investigating & improving the reliability and repeatability of keystroke dynamics timers, *International Journal of Network Security & Its Applications (IJNSA)*, 2(3): 70–85.
- Pohoa, V. v., Pohoa, S., Ray, A. & Joshi, S. S. (2009). Hidden markov model (hmm)-based user authentication using keystroke dynamics, patent.
- Rao, B. (2005). *Continuous keystroke biometric system*, Master's thesis, University of California.
- Ratha, N. K., Connell, J. H. & Bolle, R. M. (2001). An analysis of minutiae matching strength, *Audio- and Video-Based Biometric Person Authentication*.
- Revett, K. (2008). *Behavioral biometrics: a remote access approach*, Wiley Publishing.
- Revett, K. (2009). A bioinformatics based approach to user authentication via keystroke dynamics, *International Journal of Control, Automation and Systems* 7(1): 7–15.
- Revett, K., de Magalhães, S. & Santos, H. (2006). Enhancing login security through the use of keystroke input dynamics, *Lecture notes in computer science* 3832.
- Revett, K., de Magalhaes, S. & Santos, H. (2007). On the use of rough sets for user authentication via keystroke dynamics, *Lecture notes in computer science* 4874: 145.

- Rodrigues, R., Yared, G., do NCosta, C., Yabu-Uti, J., Violaro, F. & Ling, L. (2006). Biometric access control through numerical keyboards based on keystroke dynamics, *Lecture notes in computer science* 3832: 640.
- Rogers, S. J. & Brown, M. (1996). Method and apparatus for verification of a computer user's identification, based on keystroke characteristics. US Patent 5,557,686.
- Ross, A. & Jain, A. (2004). Biometric sensor interoperability: A case study in fingerprints, *Proc. of International ECCV Workshop on Biometric Authentication (BioAW)*, Springer, pp. 134–145.
- Ross, A., Nandakumar, K. & Jain, A. (2006). *Handbook of Multibiometrics*, Springer.
- Sang, Y., Shen, H. & Fan, P. (2004). Novel impostors detection in keystroke dynamics by support vector machine, *Proc. of the 5th international conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT 2004)*.
- Schneier, B. (1999). Inside risks: the uses and abuses of biometrics, *Commun. ACM*.
- Song, D., Venable, P. & Perrig, A. (1997). User recognition by keystroke latency pattern analysis, *Retrieved on 19*.
- Spillane, R. (1975). Keyboard apparatus for personal identification.
- Stefan, D. & Yao, D. (2008). Keystroke dynamics authentication and human-behavior driven bot detection, *Technical report*, Technical report, Rutgers University.
- Teh, P., Teoh, A., Ong, T. & Neo, H. (2007). Statistical fusion approach on keystroke dynamics, *Proceedings of the 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System-Volume 00*, IEEE Computer Society, pp. 918–923.
- Theofanos, M., Stanton, B. & Wolfson, C. A. (2008). Usability & biometrics: Ensuring successful biometric systems, *Technical report*, The National Institute of Standards and Technology (NIST).
- Umphress, D. & Williams, G. (1985). Identity verification through keyboard characteristics, *Internat. J. Man-Machine Studies* 23: 263–273.
- Yu, E. & Cho, S. (2004). Keystroke dynamics identity verification – its problems and practical solutions, *Computers & Security* 23(5): 428–440.

DWT Domain On-Line Signature Verification

Isao Nakanishi, Shouta Koike, Yoshio Itoh and Shigang Li
Tottori University
Japan

1. Introduction

Biometrics attracts attention since person authentication becomes very important in networked society. As the biometrics, the fingerprint, iris, face, ear, vein, gate, voice and signature are well known and are used in various applications (Jain et al., 1999; James et al., 2005). Especially, assuming mobile access using a portable terminal such as a personal digital assistant (PDA), a camera, microphone, and pen-tablet are normally equipped; therefore, authentication using the face, voice and/or signature can be realized with no additional sensor.



Fig. 1. A PDA with a pen-tablet

On the other hand, the safety of biometric data is discussed actively. Every human being has limited biometrics, for example, only ten fingerprints and one face. If the biometric data are leaked out and it is known whose they are, they are never used for authentication again.

To deal with this problem, cancelable biometric techniques have been proposed, which use not biometric data directly but one-to-one transformed data from the biometric data. However, such a technique is unnecessary if the biometrics itself is cancelable.

Among various biometric modalities, only the signature is cancelable from a viewpoint of spoofing. Even if a signature shape is known by others, it is possible to cope with the problem by changing the shape. Especially, in on-line signatures, the habit during writing is biometrics and it is not remained in the signature shape; therefore, to imitate it is quite difficult even if the signature shape is copied. As a result, the on-line signature verification is actively researched

(Dimauro et al., 2004; Fierrez & Ortega-Garcia, 2007; Jain et al., 2002; Plamondon & Srihari, 2000). However, the verification performance tends to be degraded since the on-line signature is a dynamic trait.

We have proposed a new on-line signature verification method in which a pen-position parameter is decomposed into sub-band signals using the discrete wavelet transform (DWT) and total decision is done by fusing verification results in sub-bands (Nakanishi et al., 2003; 2004; 2005). The reason why we use only the pen-position parameter is that detecting functions of other parameters such as pen-pressure, pen-altitude, and/or pen-direction are not equipped in the PDA.

However, since the signature shape is visible, it is relatively easy to forge the pen-position parameter by tracing genuine signatures by others. In the proposed method, individual features of a signature are enhanced and extracted in the sub-band signals, so that such well-forged signatures can be distinguished from genuine ones. Additionally, in the verification process of the proposed method, dynamic programming (DP) matching is adopted to make it possible to verify two data series with different number of sampled points. The purpose of the DP matching is to find the best combination between such two data series. Concretely, a DP distance is calculated in every possible combination of the two data series and as a result the combination which has the smallest DP distance is regarded as the best.

But there are problems in use of the DP matching. The DP distance is obtained as dissimilarity; therefore, signatures with large DP distances are rejected even if they are of genuine. For instance, in a pen-tablet system, a pen-up while writing causes large differences in coordinate values of pen-position and so increases false rejection. On the other hand, signatures with small DP distances are accepted even if they are forgery. The DP matching forces to match two signatures even if either is forgery. It increases false acceptance.

Consequently, we propose simply-partitioned DP matching. Two data series compared are divided into several partitions and the DP distance is calculated every partition. The DP distance is initialized at the start of a next partition, so that it reduces excessively large DP distances, that is, the false rejection. On the other hand, limitation of combination in matching is effective for rejecting forgeries; therefore, it reduces the false acceptance.

There is another important problem when we use the DP matching. The DP distance is proportional to the number of signature's sampled data, that is, signature complexity (shape), so that if it is used as a criterion in verification, each signature (user) has a different optimal threshold. But, it is general to use a single threshold commonly in an authentication system. If the common threshold is used for all signatures, it results in degradation of verification performance. Therefore, we have studied threshold equalization in the on-line signature verification (Nakanishi et al., 2008). We propose new equalizing methods based on linear and nonlinear approximation between the number of sampled data and optimal thresholds.

2. DWT domain on-line signature verification

In this section, we briefly explain the proposed on-line signature verification in the DWT domain.

2.1 System overview

A signal flow diagram is shown in Fig. 2. An on-line signature is captured as x and y coordinate (pen-position) data in a digital pen-tablet system and their sampled data are

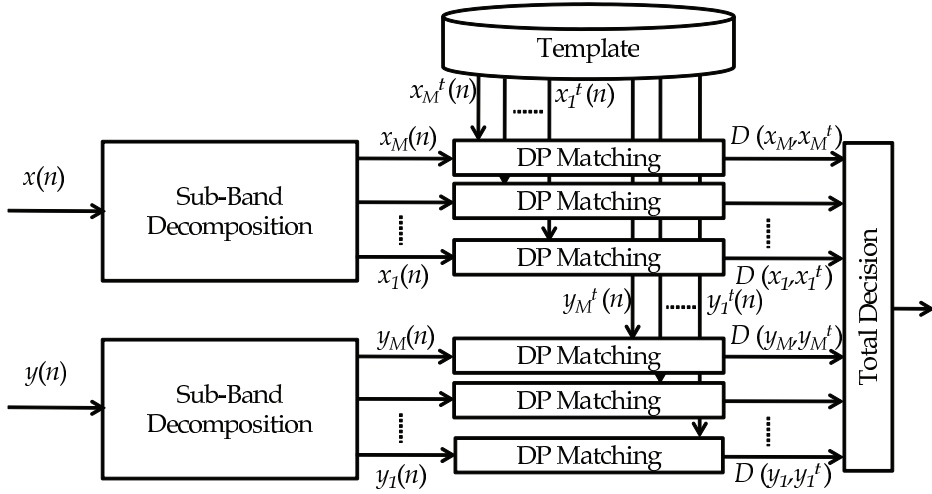


Fig. 2. DWT domain on-line signature verification

given by $x(n)$ and $y(n)$ where $n = 0, 1, \dots, S_n - 1$ and S_n is the number of sampled data. They are respectively normalized in both time and amplitude domains and then decomposed into sub-band signals using sub-band filters by the DWT (Nakanishi et al., 2003; 2004; 2005). In advance of verification, sub-band signals are enrolled as a template for each user. Templates are generated by ensemble-averaging several genuine signatures. Please refer to Ref. (Nakanishi et al., 2003) for the details. At the verification stage, each decomposed signal is compared with its template based on the DP matching and a DP distance is obtained at each decomposed level. Final score is calculated by combing the DP distances at appropriate sub-bands in both coordinates. Total decision is done by comparing the final score with a threshold and it is verified whether the signature data are of genuine or not.

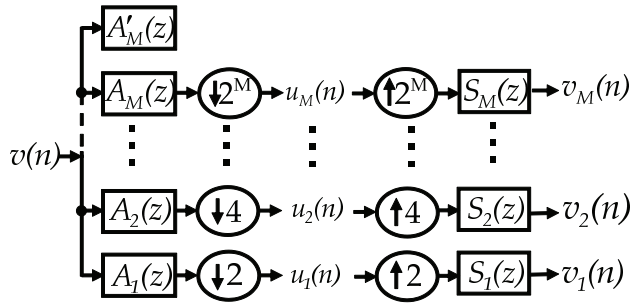
2.2 Feature extraction by DWT

In the following, the $x(n)$ and $y(n)$ are represented as $v(n)$ together for convenience. The DWT of the pen-position data: $v(n)$ is defined as

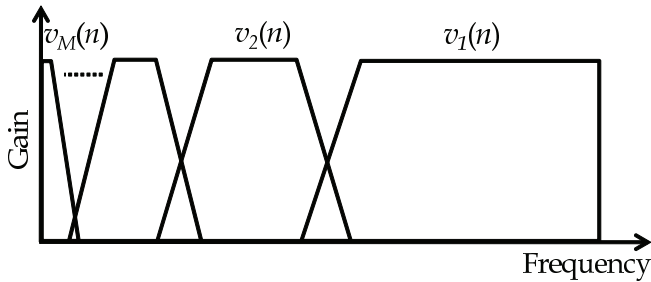
$$u_k(n) = \sum_m v(m) \overline{\Psi_{k,n}(m)} \tag{1}$$

where $\Psi_{k,n}(m)$ is a wavelet function and $\bar{\cdot}$ denotes the conjugate. k is a frequency (level) index. It is well known that the DWT corresponds to an octave-band filter bank (Strang & Nguyen, 1997) of which parallel structure and frequency characteristics are shown in Fig. 3, where $(\downarrow 2^k)$ and $(\uparrow 2^k)$ are down-sampling and up-sampling, respectively. M is the maximum level of the sub-band, that is, the decomposition level. $A_k(z)$ and $S_k(z)$ ($k = 1, \dots, M$) are analysis filters and synthesis ones, respectively.

The synthesized signal: $v_k(n)$ in each sub-band is the signal in higher frequency band and called *Detail* which corresponds to the difference between signals. Therefore, we adopt the *Detail* as an enhanced individual feature which can be extracted with no specialized function: pen-pressure, pen-altitude, and/or pen-direction which are not equipped in the PDA.



(a) Parallel structure



(b) Frequency characteristics

Fig. 3. Sub-band decomposition by DWT

Let us get another perspective on the effect of the sub-band decomposition using Fig. 4. Each signature is digitized at equal (common) sampling period using a pen-tablet system. In the proposed system, writing time of all signatures is normalized in order to suppress intra-class variation. Concretely, the sampling period of each signature is divided by the number of sampled data and so becomes real-valued.

Even genuine signatures have different number of sampled data; therefore, all signatures have different normalized sampling periods, that is, different sampling frequencies.

In general, variation of writing time in the genuine signatures is small, so that their sampling periods (frequencies) are comparable as shown in Fig. 4 (a). On the other hand, in the case of forged signatures, the variation of writing time is relatively large since it is not easy for forgers to imitate writing speed and rhythm of genuine signatures. Thus, sampling periods (frequencies) of the forged signatures become greatly different from those of the genuine signatures as in Fig. 4 (b).

The maximum frequency: f_m of the octave-band filter bank is determined by the sampling frequency based on the "sampling theory". If the sampling frequencies are greatly different, each octave band (decomposition level) includes greatly different frequency elements as illustrated in (b). In other words, even if levels compared are the same, frequency elements included in one level are different from the other, so that the differences between genuine signatures and forged ones are accentuated.

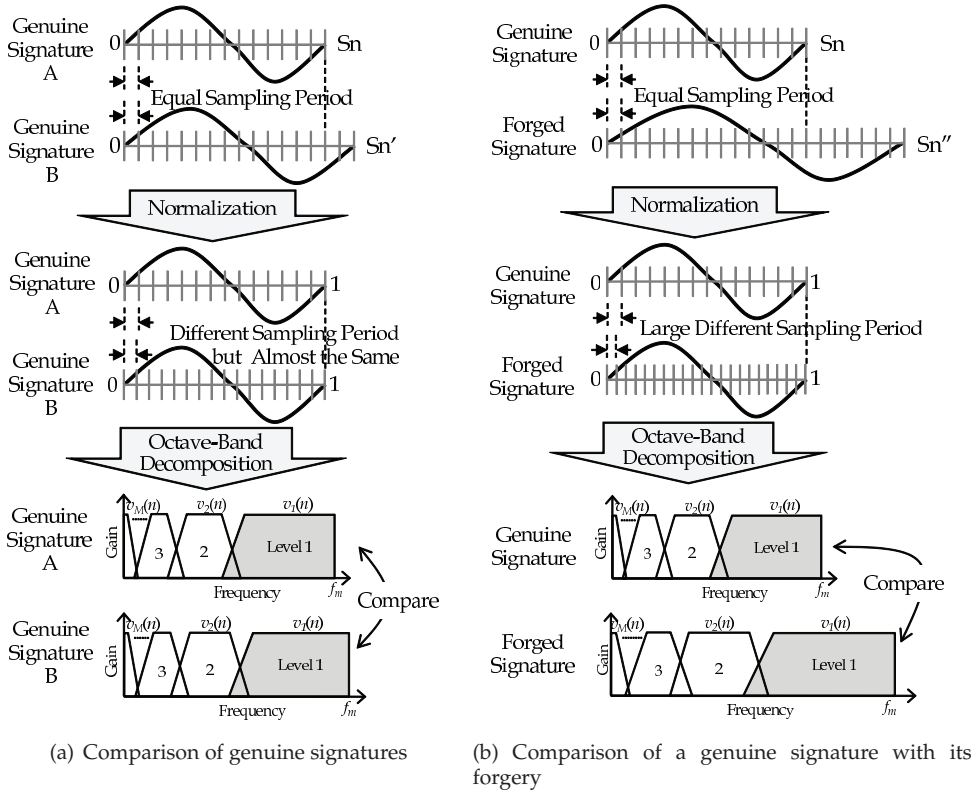


Fig. 4. Effect of sub-band decomposition

The advantage was confirmed comparing with a time-domain method (Nakanishi et al., 2005). Of course, if forgers imitate writing speed and rhythm of genuine signatures, it is impossible for the proposed method to distinguish forged signatures from genuine ones.

2.3 Verification by DP matching

Since on-line signatures have large intra-class variation, one-to-one matching cannot be applied in verification. In order to deal with the problem, the verification was performed every stroke (intra-stroke or inter-stroke) in the conventional system (Nakanishi et al., 2003; 2004; 2005; 2008). However, a part of signature databases eliminates the data in inter-strokes and so we could not apply the conventional system to such a database.

Therefore, we introduce DP matching into the verification process. The DP matching is effective in finding the best combination between two data series even if they have different number as illustrated in Fig. 5.

Letting the two data series be $a(i)$ ($i = 0, 1, \dots, I - 1$) and $b(j)$ ($j = 0, 1, \dots, J - 1$), the local distance at k th is defined as

$$d(k) = |a(i)_k - b(j)_k| \quad (k = 0, 1, \dots, K - 1) \tag{2}$$

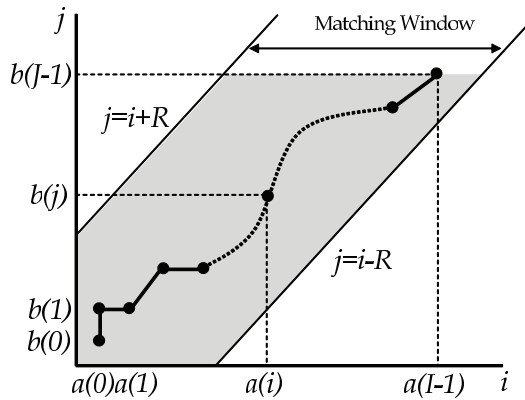


Fig. 5. DP matching

where instead of i and j , k is used as another time index since these data are permitted to be referred redundantly.

By accumulating the local distances in one possible combination between the two series, a DP distance is given by

$$D(a, b) = \sum_{k=0}^{K-1} w(k)d(k) \tag{3}$$

where $w(k)$ is a weighting factor. After calculating the DP distance in all possible combinations, we can find the best combination by searching the combination with the smallest DP distance.

Moreover, since the DP distance depends on the number of sampled data, the normalized DP distance is used in general.

$$nD(a, b) = D(a, b) / \sum_{k=0}^{K-1} w(k) \tag{4}$$

Assuming the weight is symmetric: (1-2-1) and the initial value is zero,

$$\sum_{k=0}^{K-1} w(k) = I + J. \tag{5}$$

In the proposed method, the DP distance is obtained at each sub-band level. Let the DP distance at l th level be $D(v, v^t)_l$ where v is sampled data series of a signature for verification and v^t is that of a template, the normalized DP distance is given by

$$nD(v, v^t)_l = D(v, v^t)_l / (V_n + T_n) \tag{6}$$

where V_n is the number of sampled data in the verification signature and T_n is that of the template.

A total distance (TD) is obtained by accumulating the normalized DP distances in sub-bands.

$$TD = c_x \cdot \frac{1}{L} \sum_{l=M-L+1}^M nD(x, x^t)_l + c_y \cdot \frac{1}{L} \sum_{l=M-L+1}^M nD(y, y^t)_l \quad (7)$$

where c_x and c_y are weights for combining the DP distances in x and y coordinates and $c_x + c_y = 1, c_x > 0, c_y > 0$. L is the number of levels used in the total decision.

2.4 Verification experiments

In order to confirm verification performance of the proposed system, we carried out experiments in the following conditions. The wavelet function was Daubechies8. The maximum level of the sub-band: M was 8 and the number of levels used in the total decision: L was 4. The combination weights were $c_x = c_y = 0.5$, which mean to take the average. For generating templates, data of five genuine signatures were ensemble-averaged.

We used a part of the on-line signature database: SVC2004 in which the data in inter-strokes were eliminated. The number of subjects was 40 and 17 subjects signed their names in Chinese characters and the rest in alphabetical ones. For collecting skilled forgeries, imposters could see how genuine signatures were being written. The total number of signatures was 1600. Please refer to Ref. (SVC2004, 2004) for more information.

The verification performance was evaluated by using an equal error rate (EER) where a false rejection rate (FRR) was equal to a false acceptance rate (FAR). The EER of the proposed system was 20.0 %. For reference, the EER of the conventional system was 28.3 %, so that it is confirmed that introducing the DP matching is effective for not only applying to the standard database but also improving verification performance.

On the other hand, assuming to use individually-optimal thresholds for all subjects, we averaged EERs of all subjects and then so obtained EER of 15.3 %. This is a rough evaluation but suggests that if a single common threshold is optimal for all subjects, the verification performance could be improved further. This issue is examined in Sect. 4.

3. Simply partitioned DP matching

There is another issue to be overcome in order to improve the verification performance. For instance, in a pen-tablet system, when the pen tip is released from the surface of the tablet, it is not guaranteed to get precise coordinate values of pen-position. It sometimes brings large differences from template data and then leads to a large DP distance. The signature with such a large DP distance is rejected even if it is genuine. This increases false rejection.

Conversely, signatures with small DP distances are accepted even if they are forgery. In particular, skilled forgeries (well-forged signatures) could make the DP distance smaller. The DP matching forces to match two signatures even if either is forged one. This increases false acceptance.

Consequently, we propose simply-partitioned DP (spDP) matching. The concept is illustrated in Fig. 6, where the number of partitions is four. Both data series: $a(i)$ and $b(j)$ are divided into several partitions of the same integer number and a sub DP distance is calculated every partition. If the division leaves remainders, they are singly distributed to partitions. The sub DP distances are initialized at the start of next partitions and a total DP distance is obtained by summing the sub DP distances in all partitions.

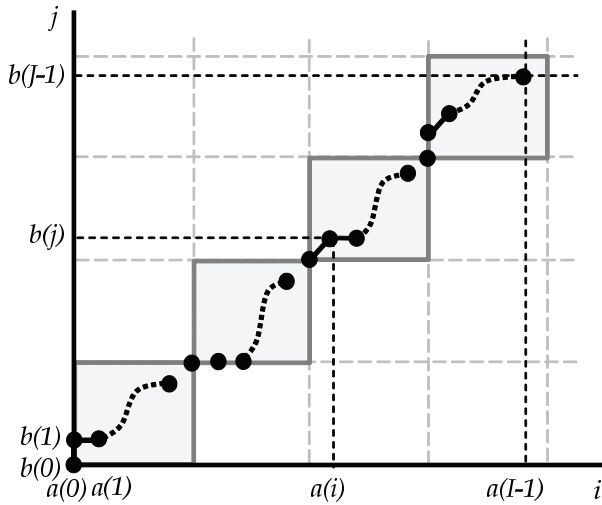


Fig. 6. Simply-partitioned DP matching (Q=4)

Assuming that genuine signatures have equivalent rhythms in writing, even if their data are partitioned, rhythms in corresponding partitions compared are still equivalent. Therefore, when a verification signature is genuine, appropriate matching pairs tend to exist in diagonal direction in Fig. 6. As a result, the spDP matching has no ill effect for false rejection.

Furthermore, even if excessively a large sub DP distance is caused by the irregular pen-up mentioned above in a partition, it is initialized at the start of the next partition, so that the spDP matching prevents the total DP distance from becoming excessively large and has an effect on reducing false rejection.

On the other hand, it is difficult for forgers to copy rhythms in writing of genuine signatures, so that the rhythm in each partition of forged signatures becomes different from that of genuine ones. Resultingly, matching pairs between the genuine signature and its forged one are not in the diagonal direction and so are excluded even if they have small DP distances. The spDP matching is also effective in reducing false acceptance.

Such a concept that inappropriate pairs are excluded by partitioning the DP distance has been already proposed (Sano et al., 2007; Yoshimura & Yoshimura, 1998) but they assume to write Chinese (Kanji) characters in standard style and the partitioning is done every character or stroke. Therefore, they could not be directly applied to the case of a cursive style (connected characters). Of course, they need additional processing for character or stroke detection.

Let the number of partitions and the sub DP distance be Q and $D(v, v^t)^q$, respectively, the normalized DP distance at the sub-band level: l is obtained by summing sub DP distances in all partitions.

$$nD(v, v^t)_l = \left(\sum_{q=1}^Q D(v, v^t)^q \right) / (V_n + T_n) \tag{8}$$

A total distance (TD) is given by Eq. (7).

By the way, the matching window is generally adopted in the DP matching as shown in Fig. 5 in order to reduce calculation amount by excluding unlikely pairs. Comparing between Figs. 5 and 6, it is clear that the spDP matching is more effective for excluding inappropriate pairs than the matching window.

3.1 Evaluation of spDP matching

We evaluated verification performance using the spDP matching. Conditions are similar with those in Sect. 2.4. EERs in various numbers of partitions are summarized in Table 1 where the case of 0 partitions corresponds to the conventional normalized DP matching.

Number of Partitions	0	2	3	4	5	6
EER (%)	20.0	17.8	16.4	16.6	17.0	16.4

Table 1. EERs in various numbers of partitions

From these results, it is confirmed that the spDP matching decreased the EER by 2-3%. In the following, we set the number of partitions at 4.

4. Threshold equalizing

There is an important issue to be overcome as mentioned in Sect. 2.4 in order to improve verification performance. In not only on-line signature verification but also all biometric authentication systems, final scores are compared with a threshold which is preliminary determined. In addition, the threshold should be common to all users. Therefore, when the final score (the DP distance) of each user is greatly different from those of others, the verification performance tends to be degraded by using the common threshold.

In general, the normalized DP distance given by Eq. (4) is used for dealing with this problem. However, the normalization also makes the DP distances of forged signatures small and thereby might increase false acceptance.

We have studied to equalize the threshold instead of using the normalization (Nakanishi et al., 2008). A total distance (TD) is rewritten as

$$TD = c_x \cdot \frac{1}{L} \sum_{l=M-L+1}^M D(x, x^t)_l + c_y \cdot \frac{1}{L} \sum_{l=M-L+1}^M D(y, y^t)_l \quad (9)$$

where please be aware that unnormalized DP distance $D(v, v^t)_l$ is used.

Generally, complex signatures have large number of sampled data since they consume relatively long time for writing. The larger the number of sampled data of a signature becomes, the larger intra-class variation becomes and as a result, it makes a DP distance large. Final decision is achieved by comparing the DP distance with a threshold; therefore, to make the DP distance inversely proportional to the number of sampled data suppresses the variation range of the DP distance and then it leads to equalization of thresholds.

Based on this concept, the conventional equalization is defined as

$$TD_{eq}^p = \frac{\gamma}{T_n^p} TD^p \quad (10)$$

where p is user number and TD^p , TD_{eq}^p and T_n^p are the total distance, the equalized total (final) distance and the number of sampled data of the template of the user, respectively. γ is a

constant for adjusting the final distance to an appropriate value. When the number of sampled data in a signature is too small, the final distance of the signature is enlarged. Conversely, large number of sampled data in a signature reduces the final distance. The effect of the threshold equalizing was already confirmed (Nakanishi et al., 2008).

4.1 New threshold equalizing methods

Figure 7 shows the relation between the number of sampled data in signatures (templates) and their optimal thresholds (total DP distances) using the spDP matching ($Q = 4$) in SVC2004, where the thresholds which bring EERs are regarded as optimal.

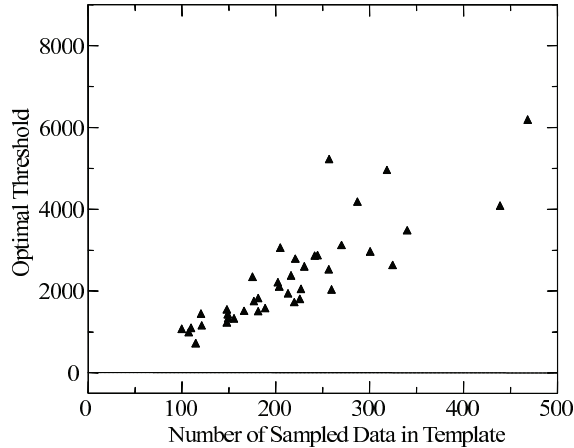


Fig. 7. Relation between the number of sampled data and optimal thresholds

The optimal thresholds are widely distributed; therefore, it is easy to guess that common use of a single threshold is not good for verification performance. In addition, the relation between the number of sampled data and the optimal threshold is not simple differently from that assumed in the conventional equalization.

4.1.1 Equalization using linear approximation

Assuming that the relation between the number of sampled data and the optimal threshold is approximated by a linear function, the total DP distance is equalized as

$$TD_{eq}^p = \frac{\gamma}{\alpha \cdot T_n^p + \beta} TD^p \quad (11)$$

where γ is the adjustment constant as well as the conventional method. α and β are the gradient and intercept of the linear function.

4.1.2 Equalization using nonlinear approximation

On the other hand, the relation between the number of sampled data and the optimal threshold could be fitted by a nonlinear function. The total DP distance is adjusted by using an exponential function as

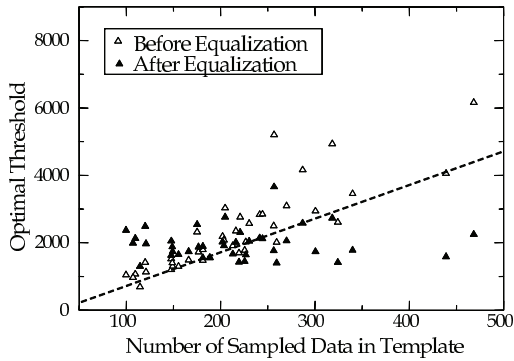
$$TD_{eq}^p = \frac{\gamma}{\exp(\alpha \cdot T_n^p + \beta)} TD^p \quad (12)$$

where α and β are constants for fitting the nonlinear function to the relation between the number of sampled data and the optimal threshold.

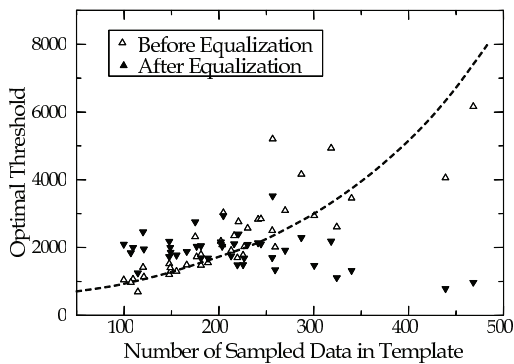
4.2 Evaluation of threshold equalizing

In order to verify effectiveness of the threshold equalizing methods, we evaluated verification performance using the SVC2004, again. Conditions are the same as those in Sect. 2.4. The number of partitions in DP matching was 4.

The distribution of optimal thresholds after equalization is compared with that before equalization in Fig. 8 where the uncolored triangles are before equalization and the black ones are after equalization. The broken lines indicate approximation functions where $\alpha = 10$ and $\beta = -293$ in the linear case and $\alpha = 0.0069$ and $\beta = 3.3$ in the nonlinear case.



(a) Linear case



(b) Nonlinear case

Fig. 8. Distribution of optimal thresholds before and after equalization in the linear and nonlinear approximation cases

From a viewpoint of their universality, it is better to determine them using a training data set, which is independent of a test data set. However, the proposed equalizing methods are based on rough approximation of the relation between the number of sampled data and the optimal threshold in the SVC2004. If the relation in the training data set is equivalent with that in the

test data set, the proposed methods does not depend on the data used. The approximation depends on not training data sets but databases. The larger the number of data becomes, the more universal the constants. In both cases, γ was set to a value which adjusts the thresholds to around 2000.

It is confirmed that the optimal thresholds, that is, the DP distances were adjusted to around 2000 and the variation range of the DP distances was narrowed.

For achieving quantitative evaluation, we analyzed statistical variance of optimal threshold values before and after the equalization. The variance before the equalization was 0.27 but after the equalization it was reduced to 0.05 in the linear case and 0.07 in the nonlinear case.

Method	EER(%)	Variance
Unnormalized DP	25.4	0.54
Normalized DP	20.0	0.17
4-partitioned DP	16.6	0.05
Conventional equalization	19.9	0.24
Linear equalization	19.0	0.22
Nonlinear equalization	19.5	0.23
4-partitioned DP + Linear equalization	14.6	0.05
4-partitioned DP + Nonlinear equalization	14.9	0.07

Table 2. EERs and variances in various methods

Finally, EERs and variances in various methods are summarized in Table 2. Comparing the EER and variance in the 4-partitioned DP matching to those in the normalized DP one, it is confirmed that the proposed spDP matching is more effective in improving verification performance than the general-used DP matching. Similarly, the proposed new threshold equalization methods are confirmed to be more efficient than the normalized DP matching and the conventional method. Moreover, combining the spDP matching with the new threshold equalization is much more effective. Especially, the smallest EER of 14.6% and variance of 0.05 were obtained when the threshold equalization using the linear approximation was applied. As confirmed in Fig. 8 (b), the adjustment in the nonlinear case might be excessive when the number of sampled data was large. It is a future problem to adopt other functions for approximating the relation between the number of sampled data and the optimal threshold. On the other hand, the EER of about 15% may not be absolutely superior to those of other on-line signature verification methods. However, it is possible to introduce the spDP matching and/or the threshold equalizing into the methods based on the DP matching and it might also improve their performance.

5. Conclusions

We have studied on-line signature verification in the DWT domain. In order to improve the verification performance, we introduced spDP matching and threshold equalizing into the verification process.

In the spDP matching, two data series compared were divided into partitions, a sub DP distance was calculated every partition, and then a total DP distance was obtained by summing the sub DP distances. The sub DP distances were initialized at the start of next partitions; therefore, accumulative distances were also initialized and the total DP distance

was prevented from becoming excessively large. It was effective in reducing false rejection. Also, the spDP matching reduced false acceptance since limitation of combination in matching excluded inappropriate matching pairs.

In the threshold equalizing, by approximating the relation between the number of sampled data in a signature and its optimal threshold by linear or nonlinear functions, the variation range of optimal thresholds of all signatures were suppressed and as a result, it prevented the verification performance from being degraded by using a single common threshold for all signatures.

In experiments using a part of the signature database: SVC2004, it was confirmed that each proposed method was efficient in improving the verification performance. Moreover, combining the spDP matching with the threshold equalizing was more effective and reduced the error rate by about 5% comparing with the general-used DP matching.

We have an issue that there might be more effective approximate functions for threshold equalization. Also, we evaluated signature's complexity by using the number of sampled data but it is expected to use sub-band signals for evaluating the complexity.

6. References

- Dimauro, G.; Impedovo, S.; Lucchese, M. G.; Modugno, R. & Pirlo, G. (2004). Recent Advancements in Automatic Signature Verification, *Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition*, Oct. 2004, pp. 179-184
- Fierrez, J. & Ortega-Garcia, J. (2007). On-Line Signature Verification, In: *Handbook of Biometrics*, A. K. Jain, P. Flynn, and A. A. Ross, (Eds.), Chapter 10, Springer, New York
- Jain, A. K.; Griess, F. D. & Connell, S. D. (2002). On-Line Signature Verification, *Pattern Recognition*, Vol. 35, No. 12, Dec. 2002, pp. 2963-2972
- Jain, A.; Bolle, R. & Pankanti, S. (1999). *BIOMETRICS*, Kluwer Academic Publishers, Massachusetts
- Wayman, J.; Jain, A.; Maltoni, D. & Maio, D. (2005). *Biometric Systems*, Springer, London
- Nakanishi, I.; Nishiguchi, N.; Itoh, Y. & Fukui, Y. (2003). On-Line Signature Verification Method Based on Discrete Wavelet Transform and Adaptive Signal Processing, *Proceedings of Workshop on Multimodal User Authentication*, Santa Barbara, USA, Dec. 2003, pp.207-214
- Nakanishi, I.; Nishiguchi, N.; Itoh, Y. & Fukui, Y. (2004). On-Line Signature Verification Based on Discrete Wavelet Domain Adaptive Signal Processing, *Proceedings of International Conference on Biometric Authentication*, Hongkong, Jul. 2004, pp. 584-591
- Nakanishi, I.; Nishiguchi, N.; Itoh, Y. & Fukui, Y. (2005). On-Line Signature Verification Based on Subband Decomposition by DWT and Adaptive Signal Processing, *Electronics and Communications in Japan, Part 3*, Vol. 88, No.6, Jun. 2005, pp.1-11
- Nakanishi, I.; Sakamoto, H.; Itoh, Y. & Fukui, Y. (2008). Threshold Equalization for On-Line Signature Verification, *IEICE Trans. Fundamentals*, Vol.E91-A, No.8, Aug. 2008, pp. 2244-2247
- Plamondon, R. & Srihari, S. N. (2000). On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, Jan. 2000, pp. 63-84

- Sano, T.; Wada, N.; Yoshida, T. & Hangai, S. (2007). A Study on Segmentation Scheme for DP Matching in Japanese Signature Verification (in Japanese), *Proceedings of 2007 IEICE General Conference*, Mar. 2007, p. B-18-4
- Strang, G. & Nguyen, T. (1997). *Wavelet and Filter Banks*, Wellesley-Cambridge Press, Massachusetts.
- SVC2004. (2004). URL: <http://www.cse.ust.hk/svc2004/index.html>
- Yoshimura, M. & Yoshimura, I. (1998). An Off-Line Verification Method for Japanese Signatures Based on a Sequential Application of Dynamic Programming Matching Method (in Japanese), *IEICE Trans. Information and Systems*, Vol. J81-D-II, No. 10, Oct. 1998, pp. 2259-2266

Part 3

Medical Biometrics

Heart Biometrics: Theory, Methods and Applications

Foteini Agrafioti, Jiexin Gao and Dimitrios Hatzinakos

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering University of Toronto 10 Kings College Road, Toronto Canada

1. Introduction

Automatic and accurate identity validation is becoming increasingly critical in several aspects of our every day lives such as in financial transactions, access control, traveling, healthcare and other. Traditional strategies to automatic identity recognition include items such as PIN numbers, tokens, passwords and ID cards. Despite the wide deployment of such tactics, the authenticating means is either entity or knowledge-based which rises concerns with regard to their ease of acquisition and use from unauthorized third parties.

According to the latest *The US Federal Commission Report, February 2010* (n.d.), in 2009 identity theft was the number one complaint category (a total of 721,418 cases of consumer complaints). As identity theft can take different forms, credit card fraud was the most prominent (17%), followed by falsification of government documents (16%), utilities fraud (15%), employment fraud (13%) and other. Among these cases, true-identity theft constitutes only a small portion of the complaints, while ID falsification appears to be the greatest threat. Unfortunately, the technology for forgery advances without analogous counterfeit improvements.

Biometric recognition was introduced as a more secure means of identity establishment. Biometric features are characteristics of the human body that are unique for every individual and that can be used to establish his/her identity in a population. These characteristics can be either physiological or behavioral. For instance, the face, the iris and the fingerprints are physiological features of the body with identifying information. Examples of behavioral features include the keystroke dynamics, the gait and the voice. The fact that biometric features are directly linked with the users presents an extraordinary opportunity to bridge the security gaps caused by traditional recognition strategies. Biometric features are difficult to steal or counterfeit when compared to PIN numbers or passwords. In addition, the convenience by which a biometric feature can be presented makes the respective systems more accessible and easy to use.

However, biometric recognition has a drawback that rises from the nature of the authenticating modality. As opposed to static PIN numbers or passwords, biometric recognition may present *false rejection* since usually no two readings of the same biometric feature are identical. Anatomical, psychological or even environmental factors affect the appearance of the biometric feature at a particular instance. For instance, faces may be presented to the recognizers under various expressions, different lighting settings or with

occlusion (glasses, hats etc). This may introduce significant variability (commonly referred to as *intra-subject* or *intra-class* variability) and the challenge is to design algorithms that can find robust biometric patterns.

Although the intra-subject variability is universal for all biometric modalities, every feature has unique characteristics. For instance, face pictures may be acquired from distance which makes them suitable for surveillance. On the opposite, fingerprints need direct contact with the sensing device, and despite the robust biometric signature that there exists, most of the error arises from inefficient processing of the image. Therefore, given a set of standards it is difficult, if not impossible, to choose one feature that satisfies all criteria. Every biometric feature has its own strengths and weaknesses and deployment choices are based on the characteristics of the envisioned application environment.

On the assumption that intra-subject variability can be sufficiently addressed with appropriate feature extraction, another consideration with this technology is the robustness to circumvention and replay attacks. Circumvention is a form of biometric feature forgery, for example the use of falsified fingerprint credentials that were copied from a print of the original finger. A replay attack is the presentation to the system of the original biometric feature from an illegitimate subject, for example voice playbacks in speaker recognition systems. Biometric obfuscation is another prominent risk with this technology. There are cases where biometric features are intentionally removed to avoid establishment of the true identity (for example asylum-seekers in Europe *Peter Allen, Calais migrants mutilate fingerprints to hide true identity, Daily Mail* (n.d.)). With the wide deployment of biometrics, these attacks are becoming frequent and concerns are once again rising on the security levels that this technology can offer.

Concentrated efforts have been made for the development the next generation of biometric characteristics that are inherently robust to the above mentioned attacks. Characteristics that are internal to the human body have been investigated such as vein patterns, the odor and cognitive biometrics. Similarly, the *medical biometrics* constitutes another category of new biometric recognition modalities that encompasses signals which are typically used in clinical diagnostics. Examples of medical biometric signals are the electrocardiogram (ECG), phonocardiogram (PPG), electroencephalogram (EEG), blood volume pressure (BVP), electromyogram (EMG) and other.

Medical biometrics have been actively investigated only within the last decade. Although the specificity to the individuals had been observed before, the complicated acquisition process and the waiting times were restrictive for application in access control. However, with the development of dry recoding sensors that are easy to attach even by non-trained personnel, the medical biometrics field flourished. The rapid advancement between 2001-2010 was supported by the fact that signal processing of physiological signals (or *biosignals*) had already progressed for diagnostic purposes and a plethora of tools were available for biometric pattern recognition.

The most prominent strength of medical biometrics is the robustness to circumvention, replay and obfuscation attacks. If established as biometrics, then the respective systems are empowered with an inherent shield to such threats. Another advantage of medical biometrics is the possibility of utilizing them for continuous authentication, since they can provide a fresh biometric reading every couple of seconds. This work is interested in the ECG signal, however the concepts presented herein may be extended to all medical biometric modalities.

2. ECG biometrics: Motivation and challenges

The ECG signal describes the electrical activity of the heart over time. It is recorded non-invasively with electrodes attached at the surface of the body. Traditionally, physicians use the ECG to gain insight on heart conditions, usually with complementary tests required in order to finalize a diagnosis. However, from a biometrics perspective, it has been demonstrated that the ECG has sufficient detail for identification.

The advantages of using the ECG for biometric recognition can be summarized as universality, permanence, uniqueness, robustness to attacks, liveness detection, continuous authentication and data minimization. More precisely,

1. *Universality* refers to the ability of collecting the biometric sample from the general population. Since the ECG is a vital signal, this property is satisfied naturally.
2. *Permanence* refers to the ability of performing biometric matches against templates that have been designed earlier in time. This essentially requires that the signal is stable over time. As it will be discussed later, the ECG is affected by both physical and psychological activity, however even though the specific local characteristics of the pulses may change, the overall diacritical waves and morphologies are still observable.
3. *Uniqueness* is guaranteed in the ECG signal because of its physiological origin. While ECG signals of different individuals conform to approximately the same pattern, there is large inter-individual variability due to the various electrophysiological parameters that control the generation of this waveform. Uniqueness will be further discussed in Section 3.1.
4. *Robustness to attacks*. The particular appearance of the ECG waveform is the outcome of several sympathetic and parasympathetic factors of the human body. Controlling the waveform or attempting to mimic somebody else's ECG signal is extremely difficult, if not impossible. To the best of our knowledge there is currently no means of falsifying an ECG waveform and presenting it to a biometric recognition system. Obfuscation is also addressed naturally.
5. *Liveness detection*. ECG offers natural liveness detection, being only present in a *living* subject. With this modality the recognizer can trivially ensure sensor liveness. Other biometric modalities, such the iris or the fingerprint require additional processing to establish the liveness of the reading.
6. *Continuous authentication*. As opposed to static iris or fingerprint images, the ECG is a dynamic biometric feature that evolves with time. When deployed for security in welfare monitoring environments, a fresh reading can be obtained every couple of second to re-authenticate an identity. This property is unique to medical biometrics and can be vital in avoiding threats such as field officer impersonation.
7. *Data minimization*. Privacy intrusion is becoming increasingly critical in environments of airtight security. One way to address this problem is to utilize as less identifying credentials as possible. Data minimization is a great possibility with ECG biometrics because there are environments where the collection of the signal is performed irrespective of the identification task. Examples of such environments are tele-medicine, patient monitoring in hospitals, field agent monitoring (fire-fighters, policemen, soldiers etc).

Despite the advantages, notable challenges are presented with this technology when envisioning large-scale deployment:

1. *Time dependency.* With time-varying biosignals there is high risk of instantaneous changes which may endanger biometric security. Recordings of the cardiac potential at the surface of the body are very prone to noise due to movements. However, even in the absence of noise, the ECG signal may destabilize with respect to a biometric template that was constructed some time earlier. The reason for this is the direct effect that the body's physiology and psychology have on the cardiac function. Therefore, a central aspect of the ECG biometrics research is the investigation of the sources of intra-subject variability.
2. *Collection periods.* As opposed to biometrics such as the face, the iris or the fingerprint, where the biometric information is available for capturing at any time instance, this is not the case with the ECG signal. Every heart beat is formed within approximately a second, which essentially means that longer waiting times are expected with this technology especially when long ECG segments are required for feature extraction. The challenge however is to minimize the number of pulses that the algorithm uses for recognition, as well as the processing time.
3. *Privacy implications.* When collecting ECG signals a large amount of sensitive information is collected inevitably. The ECG signal may reveal current and past medical conditions as well as hints on the instantaneous emotional activity of the monitored individual. Traditionally, the ECG is available to physicians only. Thus, the possibility of linking ECG samples to identities can imply catastrophic privacy issues.
4. *Cardiac Conditions.* Although cardiac disorders are not as a frequent damaging factor as injuries for more conventional biometrics (fingerprint, face), they can limit ECG biometric methods. Disorders can range from an isolated irregularity (Atria and ventricle premature contractions) to severe conditions which require immediate medical assistance. The biometric challenge is therefore to design algorithms that are invariant to tolerable everyday ECG irregularities Agrafioti & Hatzinakos (2008a).

3. Electrocardiogram fundamentals

The electrocardiogram (ECG) is one of the most widely used signals in healthcare. Recorded at the surface of the body, with electrodes attached in various configurations, the ECG signal is studied for diagnostics even at the very early stage of a disease. In essence, this signal describes the electrical activity of the heart over time, and pictures the sequential depolarization and repolarization of the different muscles that form the myocardium.

The first recording device was developed by the physiologist Williem Einthoven in the early 20th century, and for this discovery he was rewarded with the Nobel Prize in Medicine in 1924 Sornmo & Laguna (2005). Since then, ECG became an indispensable tool in clinical cardiology. The deployment of this signal in biometric recognition and affective computing is relatively young.

Figure 1 shows the salient components of an ECG signal i.e., the *P* wave, the *QRS* complex and the *T* wave. The *P* wave describes the depolarization of the right and left atria. The amplitude of this wave is relatively small, because the atrial muscle mass is limited. The absence of a *P* wave typically indicates ventricular ectopic focus. This wave usually has a positive polarity, with a duration of approximately 120 ms, while its spectral content is limited to 10-15 Hz, i.e., low frequencies.

The *QRS* complex corresponds to the largest wave, since it represents the depolarization of the right and left ventricles, being the heart chambers with substantial mass. The duration of this complex is approximately 70-110 ms in a normal heartbeat. The anatomic characteristics

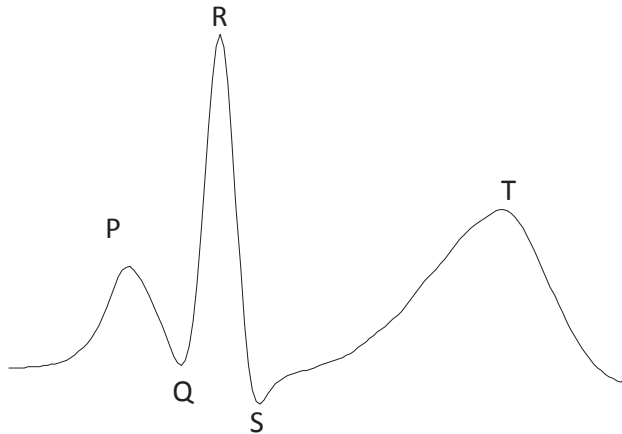


Fig. 1. Main components of an ECG heart beat.

of the *QRS* complex depend on the origin of the pulse. Due to its steep slopes, the spectrum of a *QRS* wave is higher compared to that of other ECG waves, and is mostly concentrated in the interval of 10-40 Hz.

Finally, the *T* wave depicts the ventricular repolarization. It has a smaller amplitude, compared to the *QRS* complex, and is usually observed 300 ms after this larger complex. However, its precise position depends on the heart rate, e.g., appearing closer to the *QRS* waves at rapid heart rates.

3.1 Inter-individual variability

This section will briefly discuss the physiological rationale for the use of ECG in biometric recognition. Overall, healthy ECG signals from different people conform to roughly the same repetitive pulse pattern. However, further investigation of a person's ECG signal can reveal notably unique trends which are not present in recordings from other individuals. The inter-individual variability of ECG has been extensively reported in the literature Draper et al. (1964); Green et al. (1985); Hoekema et al. (2001); Kozmann et al. (1989; 2000); Larkin & Hunyor (1980); Pilkington et al. (2006).

More specific, the ECG depicts various electrophysiological properties of the cardiac muscle. Model studies have shown that physiological factors such as the heart mass orientation, the conductivity of various areas and the activation order of the heart, are sources of significant variability among individuals Hoekema et al. (2001); Kozmann et al. (2000).

Furthermore, geometrical attributes such as the exact position and orientation of the myocardium, and torso shape designate ECG signals with particularly distinct and personalized characteristics. Other factors affecting the ECG signal are the timing of depolarization and repolarization and lead placement. In addition, except for the anatomic idiosyncrasy of the heart, unique patterns have been associated to physical characteristics such as the body habitus and gender Green et al. (1985); Hoekema et al. (2001); Kozmann et al. (1989; 2000); Simon & Eswaran (1997). The electrical map of the area surrounding the heart may also be affected by variations of other organs in the thorax Hoekema et al. (2001).

In fact, various methodologies have been proposed to eliminate the differences among ECG recordings. The idea of clearing off the inter-individual variability is typical when seeking to establish healthy diagnostic standards Draper et al. (1964). Automatic diagnosis

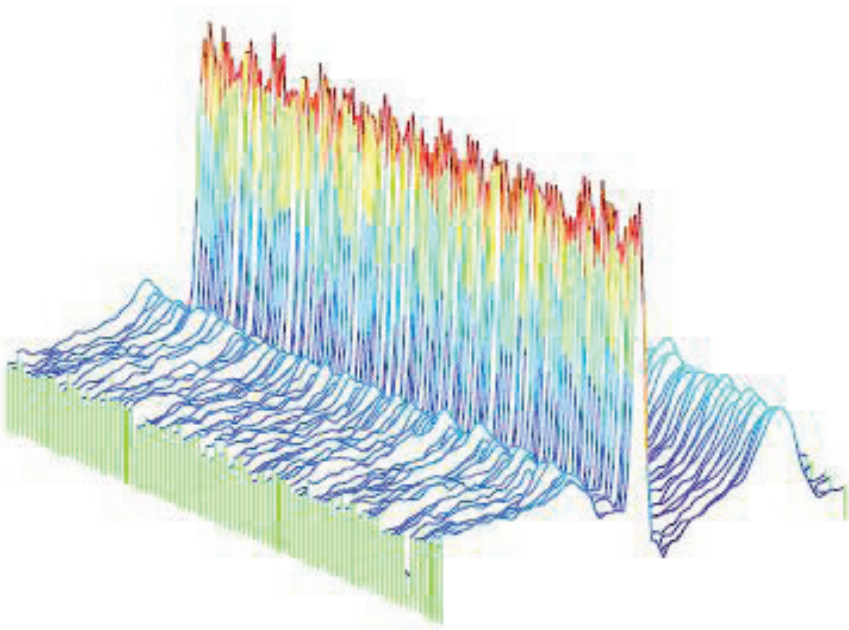


Fig. 2. Variability surrounding the *QRS* complex among heart beats of the same individual.

of pathologies using the ECG is infeasible if the level of variability among healthy people is high Kozmann et al. (2000). In such algorithms, personalized parameters of every subject are treated as random variables and a number of criteria have been defined to quantify the degree of subjects' similarities on a specific feature basis.

4. Related research

Prior works in the ECG biometric recognition field can be categorized as either fiducial points dependent or independent. Fiducials are specific points of interest on an ECG heart beat such as the ones shown in Figure 1. Therefore, fiducial based approaches rely on local features of the heart beats for biometric template design, such as the temporal or amplitude difference between consecutive fiducial points. On the other hand, fiducial points independent approaches treat the ECG signal or isolated heart beats holistically and extract features statistically based on the overall morphology of the waveform. This distinction has a direct analogy to face biometric systems, where one can operate locally and extract biometric features such the distance between the eyes or the size of the mouth. A holistic approach in this case would then be to analyze the facial image globally.

Both approaches have advantages and disadvantages. While fiducial oriented features risk to miss identifying information hidden behind the overall morphology of the biometric, holistic approaches deal with a large amount of redundant information that needs to be eliminated. The challenge in the later case, is remove this information in a way that the intra-subject variability is minimized and the inter-subject is maximized. For the ECG case, detecting fiducial points is a very obscure process due to the high variability of the signal. Figure 2 shows an example of aligned ECG heart beats which belong to the same individual. Even

though the *QRS* complex is perfectly aligned there is significant variability surrounding the *P* and the *T* wave, rendering the localization of these waves' onsets and offsets very difficult. In fact, there is no universally acknowledged rule that can guide this detection Hoekema et al. (2001).

This section provides an overview of fiducial dependent and independent approaches to ECG biometrics that are currently found in the literature. A comparison is also provided in Tables 1 and 2.

5. Fiducial based approaches

Among the earliest works in the area is Biel et al. (2001) proposal, in 2001, for a fiducial feature extraction algorithm, which demonstrated the feasibility of using ECG signals for human identification. The standard 12 lead system was used to record signals from 20 subjects of various ages. Special experimentation was carried out to test variations due to lead placement in terms of the exact location and the operators who place the electrodes.

Out of 30 clinical diagnosis features that were estimated for each of the 12 leads, only 12 features were retained for matching by inspection of the correlation matrix. These features pictured local characteristics of the pulses, such as the *QRS* complex and *T* wave amplitudes, *P* wave duration and other. This feature set was subsequently fed to SIMCA for training and classification. Results of combining different features were compared to demonstrate that the best case classification rate was 100% with just 10 features.

Kyoso & Uchiyama (2001), also proposed fiducial based features for ECG biometric recognition. Overall, four feature parameters were selected i.e., the *P* wave duration, *PQ* interval, *QRS* complex and *QT* durations. These features were identified on the pulses by applying a threshold to the second order derivative. The subject with the smallest Mahalanobis distance between each two of the four feature parameters was selected as output. The highest reported performance was 94.2% for using just the *QRS* and *QT* intervals.

In 2002, Shen et al. (2002) reported an ECG based recognition method with seven fiducial based features defined based on the *QRS* complex. The underlying idea was that this wave is less affected by varying heart rates, and thus is appropriate for ECG biometric recognition.

The proposed methodology encompassed two steps: During a first step, template matching was used to compute the correlation coefficient among *QRS* complexes in the gallery set in order to find possible candidates and prune the search space. A decision based neural network (DBNN) was then formed to strengthen the validation of the identity resulting from the first step. While the first step managed to correctly identify only 85% of the cases, the neural network resulted in 100% recognition.

More complete biometric recognition tests were reported in 2004, by Israel et al. (2005). This work presented the three clear stages of ECG biometric recognition i.e., preprocessing, feature extraction and classification. In addition, a variety of experimental settings are described in Israel et al. (2005) such as, electrode placement and physical stress.

The proposed system employed only temporal features. A filter was first applied to retain signal information in the band 1.1- 40 Hz and discard the rest of the spectral components which were attributed to noise. By targeting to keep discriminative information while applying a stable filter over the gallery set, different filtering techniques were examined to conclude to a local averaging, spectral differencing and Fourier band-pass filter. The highest identification rate achieved was close to 100% which generally established the ECG signal as a biometric feature that is robust to heart rate variability.

A similar approach was reported in the same year by Palaniappan & Krishnan (2004). In addition to commonly used features within QRS complex, a form factor, which is a measure of signal complexity, was proposed and tested as input to a neural network classifier. An identification rate of 97.6% was achieved over recordings of 10 individuals, by training a MLP-BP with 30 hidden units.

Kim et al. (2005), proposed a method to normalize time domain features by Fourier synthesizing an up-sampled ECG heart beat. In addition, the *P* wave was excluded when calculating the features since it disappears when heart rate increases. With this strategy, the performance improved significantly when the testing subjects were performing physical activities.

Another work that addressed the heart rate variations was by Saechia et al. (2005) in 2005. The heart beats were normalized to a the healthy durations and then divided into three sub-sequences: *P* wave, *QRS* complex and *T* wave. The Fourier transform was applied on a heart beat itself and all three sub-sequences. The spectrum were then passed to a neural network for classification. It was shown that false rate was significantly lower (17.14% to 2.85%) by using the three sub-sequences instead of the original heart beat.

Zhang & Wei (2006), suggested 14 commonly used features from ECG heart beats on which a PCA was applied to reduce dimensionality. A classification method based on Bayes' Theorem was proposed to maximize the posterior probability given prior probabilities and class-conditional densities. The proposed method outperformed Mahalanobis' distance by 3.5% to 13% depending on the particular lead that was used.

Singh & Gupta (2008), proposed a way to delineate the *P* and *T* waveforms for accurate feature extraction. By examining the ECG signal within a preset window before *Q* wave onset and apply a threshold to its first derivative, the precise position of *P* was revealed. In addition the onset, peak and offset of the *P* wave were detected by tracing the signal and examining the zero crossings in its first derivative. The system's accuracy was 99% as tested over 25 subjects. In 2009, Boumbarov et al. (2009), investigated different models, such as HMM-GMM (Hidden markov model with Gaussian mixture model), HMM-SGM (Hidden markov model with single Gaussian model) and CRF (Conditional Random Field), to determine different fiducial points in an ECG segment, followed by PCA and LDA for dimensionality reduction. A neural network with radial basis function was realized as the classifier and the recognition rate was between 62% to 94% for different subjects.

Ting & Salleh (2010), described in 2010 a nonlinear dynamical model to represent the ECG in a state space form with the posterior states inferred by an Extended Kalman Filter. The Log-likelihood score was used to compare the estimated model of a testing ECG to that of the enrolled templates. The reported identification rate was 87.5% on the normal beats of 13 subjects from the MIT Arrhythmia database. It was also reported that the method was robust to noise for SNR above 20 dB.

The Dynamic time warping or FLDA were used in Venkatesh & Jayaraman (2010), together with the nearest neighbor classifier. The proposed system was comprised of two steps as follows: First the FLDA and nearest neighbor operated on the features and then a DTW classifier was applied to additionally boost the performance (100% over a 12-subject database). For verification, only features related to *QRS* complex were selected due to their robustness to heart rate variability. Same two-stage setting was applied together with a threshold and the reported performance was 96% for 12 legitimates and 3 intruders.

Another fiducial based method was proposed by Tawfik et al. (2010). In this work, the ECG segment between the *QRS* complex and the *T* wave was first extracted and normalized in the

time domain by using Framingham correction formula or assuming constant QT interval. The DCT was then applied and the coefficients were fed into a neural network for classification. The identification rate was 97.727% for that of Framingham correction formula and 98.18% for that with constant QT interval. Furthermore, using only the QRS complex without any time domain manipulation yielded a performance of 99.09%.

6. Fiducial independent approaches

On the non-fiducial methodologies side the majority of the works were reported after 2006. Among the earliest is Plataniotis et al. (2006) proposal for an autocorrelation (AC) based feature extractor. With the objective of capturing the repetitive pattern of ECG, the authors suggested the AC of an ECG segment as a way to avoid fiducial points detection. It was demonstrated that the autocorrelation of windowed ECG signals, embeds highly discriminative information in a population. However, depending on the original sampling frequency of the signal, the dimensionality of a segment from the autocorrelation was considerably high for cost efficient applications. To reduce the dimensionality and retain only useful for recognition characteristics, the discrete cosine transform (DCT) was applied. The method was tested on 14 subjects, for which multiple ECG recordings were available, acquired a few years apart. The identification performance was 100%.

Wübbeler et al. (2007), have also reported an ECG based human recognizer by extracting biometric features from a combination of Leads I, II and III i.e., a two dimensional heart vector also known as the characteristic of the electrocardiogram. To locate and extract pulses a thresholding procedure was applied. For classification, the distance between two heart vectors as well as their first and second temporal derivatives were calculated. A verification functionality was also designed by setting a threshold on the distances. Authenticated pairs were considered those which were adequately related, while in any other case, input signals were rejected. The reported false acceptance and rejection rates were 0.2% and 2.5% corresponding to a 2.8% equal error rate (EER). The overall recognition rate of the system was 99% for 74 subjects.

A methodology for ECG synthesis was proposed in 2007 by Molina et al. (2007). An ECG heartbeat was normalized and compared with its estimate which was constructed from itself and the templates from the claimed identity. The estimated version was produced by a morphological synthesis algorithm involving a modified dynamic time warping procedure. The Euclidean distance was used as a similarity measure and a threshold was applied to decide the authenticity. The highest reported performance was 98% with a 2

In 2008, Chan et al. (2008), reported ECG signal collection from the fingers by asking the participants to hold two electrode pads with their thumb and index finger. The Wavelet distance (WDIST) was used as the similarity measure with a classification accuracy of 89.1%, which outperformed other methods such as the percent residual distance (PRD) and the correlation coefficient (CCORR). Furthermore, a new recording session was conducted on several misclassified subjects which improved the system performance to 95%.

In the same year, Chiu et al. (2008), proposed the use of DWT on heuristically isolated pulses. More precisely, every heart beat was determined on the ECG signal, as 43 samples backward and 84 samples forward from every R peak. The DWT was used for feature extraction and the Euclidean distance as the similarity measure. When the proposed method was applied to a database of 35 normal subjects, a 100% verification rate was reported. The author also pointed

out that false rate would increase if 10 subjects with cardiac arrhythmia were included in the database.

Fatemian & Hatzinakos (2009), also suggested the Wavelet transform to denoise and delineate the ECG signals, followed by a process wherein every heart beat was resampled, normalized, aligned and averaged to create one strong template per subject. A correlation analysis was directly applied to test heart beats and the template since the gallery size was greatly reduced. The reported recognition rate was 99.6% for a setting where every subject has 2 templates in the gallery.

The Spectrogram was employed in Odinaka et al. (2010) to transform the ECG into a set of time-frequency bins which were modeled by independent normal distributions. Dimensionality reduction was based on Kullback-Leibler divergence where a feature is selected only if the relative entropy between itself and the nominal model (which is the spectrogram of all subjects in database) is larger than a certain threshold. Log-likelihood ratio was used as a similarity measure for classification and different scenarios were examined. For enrollment and testings over the same day, a 0.37% ERR was achieved for verification and a 99% identification rate. For different days, the performance was 5.58% ERR and 76.9% respectively.

Ye et al. (2010), applied the discrete wavelet transform (DWT) and independent component analysis (ICA) on ECG heart beat segments to obtain 118 and 18 features respectively (the feature vectors were concatenated). The dimensionality of the feature space was reduced from 136 to 26 using a PCA which retained 99% of the data's variance. An SVM with Gaussian radial basis function was used for classification with a decision level fusion of the results from the two leads. Rank-1 classification rate of 99.6% was achieved for normal heart beats. Another observation was that even though dynamic features such as the *R-R* interval proved to be beneficial for arrhythmia classification, they were not as good descriptors for biometric recognition.

Coutinho et al. (2010), isolated the heart beats and performed 8-bit uniform quantization to map the ECG samples to strings from a 256-symbol alphabet. Classification was based on finding the template in the gallery set that results in the shortest description length of the test input (given the strings in the particular template) which was calculated by the Ziv-Merhav cross parsing algorithm. The reported classification accuracy was 100% on a 19-subject database in presence of emotional state variation.

Autoregressive modeling was used in Ghofrani & Bostani (2010). The ECG signal was segmented with 50% overlap and an AR model of order 4 was estimated so that its coefficients are used for classification. Furthermore, the mean PSD of each segment was concatenated as add-on features which increased the system performance to 100% using a KNN classifier. The proposed method outperformed the state-of-art fractal-based approach such as the Lyapunov exponent, ApEn, Shannon Entropy, and the Higuchi chaotic dimension.

Li & Narayanan (2010), proposed a method to model the ECG signal in both the temporal and cepstral domain. The Hermite polynomial expansion was employed to transform heart beats into Hermite polynomial coefficients which were then modeled by an SVM with a linear kernel. Cepstral features were extracted by simple linear filtering and modeled by GMM/GSV(GMM supervector). The highest reported performance was 98.26% with a 5% ERR corresponding to a score level fusion of both temporal and cepstral classifiers.

Tables 1 and 2 provide a high level comparison of the works that can currently be found in the literature.

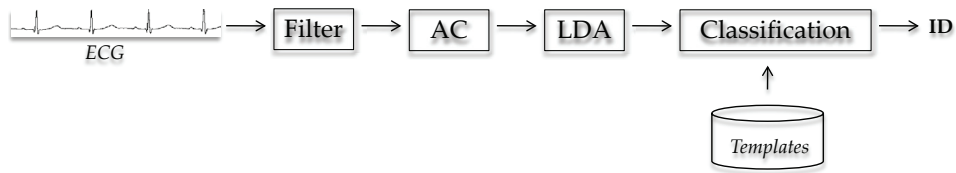


Fig. 3. Block diagram of the AC/LDA algorithm.

7. The AC/LDA algorithm

The AC/ LDA, is a fiducial points independent methodology originally proposed in Agrafioti & Hatzinakos (2008b) and later expanded to cardiac irregularities in Agrafioti & Hatzinakos (2008a). This method, relies on a small segment of the autocorrelation of 5 sec ECG signals. The 5 sec duration has been chosen experimentally, as it is fast enough for real life applications, and also allows to capture the ECG's repetitive property. The reader should note that this ECG window is allowed to cut the signal even in the middle of pulse, and it does not require any prior knowledge on the heart beat locations. The autocorrelation (AC) is computed for every 5 sec ECG using:

$$\hat{R}_{xx}(m) = \sum_{i=0}^{N-|m|-1} x(i)x(i+m) \quad (1)$$

where $x(i)$ is an ECG sample for $i = 0, 1, \dots, (N - |m| - 1)$, $x(i + m)$ is the time shifted version of the ECG with a time lag m , and N is the length of the signal.

Out of \hat{R}_{xx} only a segment $\phi(m)$, $m = 0, 1, \dots, M$, defined by the zero lag instance and extending to approximately the length of a QRS complex¹ is used for further processing. This is because this wave is the least affected by heart rate changes, thus utilizing only this segment for discriminant analysis, makes the system robust to the heart rate variability. Figure 3, shows a block diagram of the AC/LDA algorithm.

In a distributed system, such as human verification on smart phones, every user can use his/her phone in one of the two modes of operation i.e., *enrollment* or *verification*. During enrollment, ECG sensors located on the smart phone, first acquire an ECG sample from a subject's fingers and then a biometric signature is designed and saved. Similarly, during verification a newly acquired sample is matched against the store template. Essentially, the verification decision is performed with a threshold on the distance of the two feature vectors. Although verification performs one-to-one matches, false acceptance can be controlled by learning the patterns of possible attackers. Therefore for smart phone based verification the objective is to *optimally reduce the within class variability while learning patterns of the general population*. This can be done with LDA training over a large generic dataset, against which a new enrollee will be learned.

More specific, given a generic dataset (which can be anonymous to ensure privacy) the autocorrelation of every 5 sec ECG segment is computed using Eq. 1. This results into a number of AC segments $\Phi(m)$ against which an input AC feature vector $\phi_{input}(m)$ is learned.

¹ A QRS complex lasts for approximately 100 msec

Method	Principle	Performance	Number of Subjects
Kyoso & Uchiyama (2001)	Analyzed four fiducial based features from heart beats, to determine those with greater impact on the identification performance	94.2%	9
Biel et al. (2001)	Use a SIEMENS ECG apparatus to record and select appropriate medical diagnostic features for classification	100%	20
Shen et al. (2002)	Use template matching and neural networks to classify QRS complex related characteristics	100%	20
Israel et al. (2005)	Analyze fiducial based temporal features under various stress conditions	100%	29
Palaniappan & Krishnan (2004)	Use two different neural network architectures for classification of six QRS wave related features	97.7%	10
Kim et al. (2005)	By normalizing ECG heartbeat using Fourier synthesis the performance under physical activities was improved	N/A	10
Saechia et al. (2005)	Examined the effectiveness of segmenting ECG heartbeat into three subsequences	97.15%	20
Zhang & Wei (2006)	Bayes' classifier based on conditional probability was used for identification and was found superior to Mahalanobis' distance.	97.4%	502
Plataniofts et al. (2006)	Analyze the autocorrelation of ECGs for feature extraction and apply DCT for dimensionality reduction	100%	14
Wübbeler et al. (2007)	Utilize the characteristic vector of the electrocardiogram for fiducial based feature extraction out of the QRS complex	99%	74
Molina et al. (2007)	Morphological synthesis technique was proposed to produced a synthesized ECG heartbeat between the test sample and template	98%	10
Chan et al. (2008)	Wavelet distance measure was introduced to test the similarity between ECGs	95%	50

Table 1. Summary of related to ECG based recognition works.

Method	Principle	Performance	Number of Subjects
Singh & Gupta (2008)	A new method to delineate P and T waves	99%	25
Fatemian & Hatzinakos (2009)	Less templates per subject in gallery set to speed up computation and reduce memory requirement	99.6%	13
Boumbarov et al. (2009)	Neural network with radial basis function was employed as the classifier	86.1%	9
Ting & Salleh (2010)	Use extended Kalman filter as inference engine to estimate ECG in state space	87.5%	13
Odinaka et al. (2010)	Time frequency analysis and relative entropy to classify ECGs	76.9%	269
Venkatesh & Jayaraman (2010)	Apply dynamic time warping and Fisher's discriminant analysis on ECG features	100%	15
Tawfik et al. (2010)	Examined the system performance when using normalized QT and QRS and using raw QRS	99.09%	22
Ye et al. (2010)	Applied Wavelet transform and Independent component analysis, together with support vector machine as classifier to fuse information from two leads	99.6%	36 normal and 112 arrhythmic
Coutinho et al. (2010)	Treat heartbeats as a strings and using Ziv-Merhav parsing to measure the cross complexity	100%	19
Ghofrani & Bostani (2010)	Autoregressive coefficient and mean of power spectral density were proposed to model the system for classification	100%	12
Li & Narayanan (2010)	Fusion of temporal and cepstral features	98.3%	18

Table 2. (Continued) Summary of related to ECG based recognition works

Let the number of classes in the generic dataset be C . The training set will then involve $C + 1$ classes as follows:

$$\Phi(m) = [\Phi_1(m), \Phi_2(m) \dots \Phi_C(m), \Phi_{input}(m)] \quad (2)$$

and for every subject i in $C + 1$, a number of C_i AC vectors are available. This is because during enrollment, longer ECG recordings can be acquired, so that multiple segments of the user's biometric participate in training. The longer the training ECG signal the lower the chances of false rejection. Furthermore, since this is only required in the enrollment mode of operation, it does not affect the overall waiting time of the verification.

Given $\Phi(m)$, LDA will find a set of k feature basis vectors $\{\psi_v\}_{v=1}^k$ by maximizing the ratio of between-class and within-class scatter matrix. Given the transformation matrix Ψ , a feature vector is projected using:

$$Y_i(k) = \Psi^T \Phi_i(m) \quad (3)$$

where eventually $k \ll m$ and at most C .

An advantage of distributed verification is that smart phone can be optimized experimentally for the intra-class variability of a particular user. This can be done during enrollment, by choosing the smallest distance threshold at which an individual is authenticated. Essentially, rather than imposing universal distance thresholds for all enrollees, every device can be "tuned" to the expected variability of the user.

8. ECG signal collection

The performance of the framework discussed in Section 7 was evaluated over ECG recordings collected at the BioSec.Lab², at the University of Toronto. Overall, two recording sessions took place, scheduled a couple of weeks apart, in order to investigate the permanence of the signal in terms of verification performance. During the first session, 52 healthy volunteers were recorded for approximately 3 min each. The experiment was repeated a month later for 16 of the volunteers.

The signals were collected from the subject's wrists, with the Vernier ECG sensor. The wrists were selected for this recording so that the morphology of the acquired signal can resemble the one collected by a smart phone from the subject's fingers. The sampling frequency was 200Hz. During the collection, the subjects were given no special instructions, in order to allow for mental state variability to be captured in the data. The recordings of the 36 volunteers who participated to the experiment only once were used for generic training. For the 16 volunteers that two recordings were available, the earliest ones were used for enrollment and the latter for testing.

9. Experimental performance

Preprocessing of the signals is a very important because ECG is affected by both high and low frequency noise. For this reason, a butterworth bandpass filter of order 4 was used, centered between 0.5 Hz and 40Hz based on empirical results. After filtering, the autocorrelation was computed according to Eq. 1, for the generic dataset, the enrollment records and the testing ones.

Each of the enrollees' recordings were appended to the generic dataset individually, and a new LDA was trained for every enrollee. The performance each recognizer was then tested with matches the respective subject's recordings in the test set. To estimate the False Acceptance

² <http://www.comm.utoronto.ca/~biometrics/>

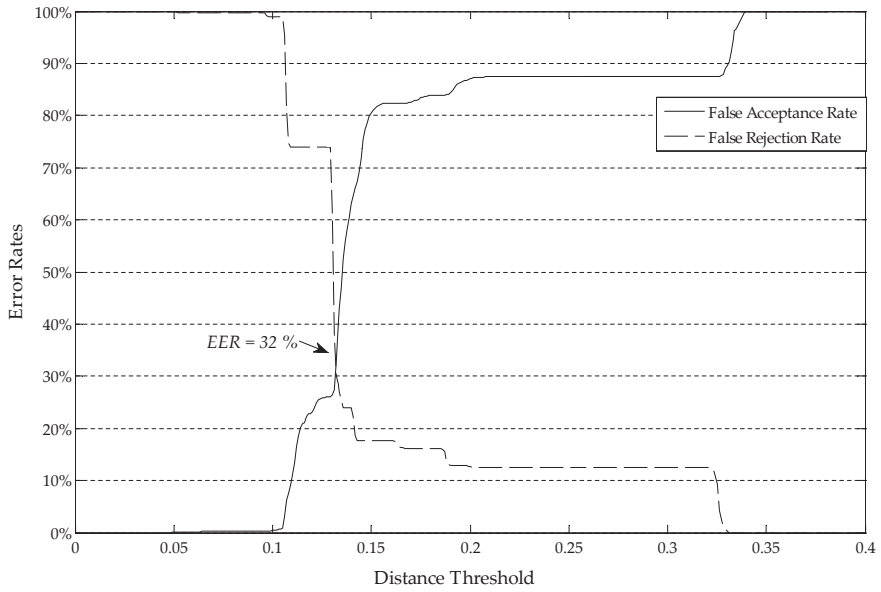


Fig. 4. False Acceptance and Rejection Rates when imposing universal recognition thresholds.

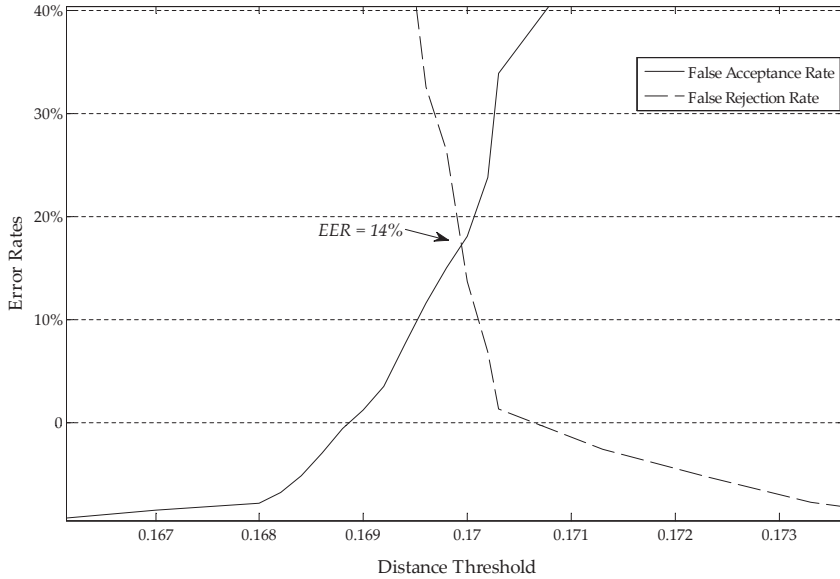


Fig. 5. False Acceptance and Rejection Rates after personalization of the thresholds.

Rate, the remaining enrolles (i.e., subjects who did not participate in the generic pool), acted as intruders to the system. This subset of recordings is unseen to the current LDA, and thus constitutes, the unknown population.

Figure 4 demonstrates the tradeoffs between false acceptance (FA) and rejection (FR) when the same threshold values are imposed for all users. The Equal Error Rate (EER) i.e., the rate at which false acceptance and rejection rates are equal is 32%. This performance is generally unacceptable for a viable security solution.

When verification is performed locally on a smart device, one can take advantage of the fact that every card can be optimized to a particular individual. This treatment controls the false rejection since the matching threshold is "tuned" to the biometric variability of the individual. Figure 5 demonstrates the error distribution for the case of personalized thresholds, when aligning the individual ROC plots of the enrolles. The EER drops on average to 14%, with a significant number of subjects exhibiting EER between 0%- 5%. More details pertaining to this result can be found in Gao et al. (2011).

10. Conclusion

This chapter discussed on one of the most extensively studied medical biometric feature, the ECG. ECG reflects the cardiac electrical activity over time, and presents significant intra subject variability due to electrophysiological variations of the myocardium. However, as it was argued there are significant advantages of using ECG for human identification such as universality, permanence and uniqueness. Therefore, a number of approaches have been developed to address the challenges of designing templates that are robust to the heart rate variability.

The reported performance of the AC/LDA approach as well as of other methodologies in the literature, establish the ECG in the biometrics world and render its use in human recognition very promising. It is crucial however to perform large scale tests that can generalize the current performance as well as to address the privacy implications that arise with this technology.

11. References

- Agrafioti, F. & Hatzinakos, D. (2008a). ECG biometric analysis in cardiac irregularity conditions, *Signal, Image and Video Processing* pp. 1863–1703.
- Agrafioti, F. & Hatzinakos, D. (2008b). Fusion of ECG sources for human identification, *3rd Int. Symp. on Communications Control and Signal Processing*, Malta, pp. 1542–1547.
- Biel, L., Pettersson, O., Philipson, L. & Wide, P. (2001). ECG analysis: a new approach in human identification, *IEEE Trans. on Instrumentation and Measurement* 50(3): 808–812.
- Boumbarov, O., Velchev, Y. & Sokolov, S. (2009). ECG personal identification in subspaces using radial basis neural networks, *IEEE Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems*, pp. 446–451.
- Chan, A., Hamdy, M., Badre, A. & Badee, V. (2008). Wavelet distance measure for person identification using electrocardiograms, *Instrumentation and Measurement, IEEE Transactions on* 57(2): 248–253.
- Chiu, C. C., Chuang, C. & Hsu, C. (2008). A novel personal identity verification approach using a discrete wavelet transform of the ECG signal, *International Conference on Multimedia and Ubiquitous Engineering*, pp. 201–206.

- Coutinho, D., Fred, A. & Figueiredo, M. (2010). One-lead ECG-based personal identification using Ziv-Merhav cross parsing, *20th Int. Conf. on Pattern Recognition*, pp. 3858–3861.
- Draper, H., Peffer, C., Stallmann, F., Littmann, D. & Pipberger, H. (1964). The corrected orthogonal electrocardiogram and vectorcardiogram in 510 normal men (frank lead system), *Circulation* 30: 853–864.
- Fatemian, S. & Hatzinakos, D. (2009). A new ECG feature extractor for biometric recognition, *16th International Conference on Digital Signal Processing*, pp. 1–6.
- Gao, J., Agrafioti, F., Mohammadzade, H. & Hatzinakos, D. (2011). ECG for blind identity verification in distributed systems, *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*.
- Ghofrani, N. & Bostani, R. (2010). Reliable features for an ECG-based biometric system, *17th Iranian Conference of Biomedical Engineering*, pp. 1–5.
- Green, L., Lux, R., Williams, C. H. R., Hunt, S. & Burgess, M. (1985). Effects of age, sex, and body habitus on qrs and st-t potential maps of 1100 normal subjects, *Circulation* 85: 244–253.
- Hoekema, R., G.Uijen & van Oosterom, A. (2001). Geometrical aspect of the interindividual variability of multilead ECG recordings, *IEEE Trans. Biomed. Eng.* 48: 551–559.
- Israel, S. A., Irvine, J. M., Cheng, A., Wiederhold, M. D. & Wiederhold, B. K. (2005). ECG to identify individuals, *Pattern Recognition* 38(1): 133–142.
- Kim, K. S., Yoon, T. H., L., J., Kim, D. J. & Koo, H. S. (2005). A robust human identification by normalized time-domain features of Electrocardiogram, *27th Annual Int. Conf on Eng. in Medicine and Biology Society*, pp. 1114–1117.
- Kozmann, G., Lux, R. & Green, L. (1989). Sources of variability in normal body surface potential maps, *Circulation* 17: 1077–1083.
- Kozmann, G., Lux, R. & Green, L. (2000). Geometrical factors affecting the interindividual variability of the ECG and the VCG, *J. Electrocardiology* 33: 219–227.
- Kyoso, M. & Uchiyama, A. (2001). Development of an ECG identification system, *23rd Annual International Conference of the Engin.g in Medicine and Biology Society*.
- Larkin, H. & Hunyor, S. (1980). Precordial voltage variation in the normal electrocardiogram, *J. Electrocardiology* 13: 347–352.
- Li, M. & Narayanan, S. (2010). Robust ECG biometrics by fusing temporal and cepstral information, *20th International Conference on Pattern Recognition*, pp. 1326–1329.
- Molina, G. G., Bruekers, F., Presura, C., Damstra, M. & van der Veen, M. (2007). Morphological synthesis of ECG signals for person authentication, *15th European Signal Proc. Conf., Poland*.
- Odinaka, I., Lai, P.-H., Kaplan, A., O'Sullivan, J., Sirevaag, E., Kristjansson, S., Sheffield, A. & Rohrbaugh, J. (2010). Ecg biometrics: A robust short-time frequency analysis, *IEEE International Workshop on Information Forensics and Security*, pp. 1–6.
- Palaniappan, R. & Krishnan, S. (2004). Identifying individuals using ECG beats, *International Conference on Signal Processing and Communications*, pp. 569–572.
- Peter Allen, *Calais migrants mutilate fingerprints to hide true identity*, *Daily Mail* (n.d.). <http://www.dailymail.co.uk/news/worldnews/article-1201126/Calais-migrants-mutilate-fingertips-hide-true-identity.html>.
- Pilkington, T., Barr, R. & Rogers, C. L. (2006). Effect of conductivity interfaces in electrocardiography, *Springer New York*. 30: 637–643.
- Plataniotis, K., Hatzinakos, D. & Lee, J. (2006). ECG biometric recognition without fiducial detection, *Proc. of Biometrics Symposiums (BSYM)*, Baltimore, Maryland, USA.

- Saechia, S., Koseeyaporn, J. & Wardkein, P. (2005). Human identification system based ECG signal, *TENCON 2005*, pp. 1–4.
- Shen, T. W., Tompkins, W. J. & Hu, Y. H. (2002). One-lead ECG for identity verification, *Proc. of the 2nd Conf. of the IEEE Eng. in Med. and Bio. Society and the Biomed. Eng. Society*, Vol. 1, pp. 62–63.
- Simon, B. P. & Eswaran, C. (1997). An ECG classifier designed using modified decision based neural network, *Comput. Biomed. Res.* 30: 257–272.
- Singh, Y. & Gupta, P. (2008). ECG to individual identification, *nd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems 2*.
- Sornmo, L. & Laguna, P. (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier.
- Tawfik, M., Selim, H. & Kamal, T. (2010). Human identification using time normalized QT signal and the QRS complex of the ECG, *7th International Symposium on Communication Systems Networks and Digital Signal Processing*, pp. 755–759.
- The US Federal Commission Report, February 2010 (n.d.). <http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2009.pdf>.
- Ting, C. M. & Salleh, S. H. (2010). ECG based personal identification using extended kalman filter, *10th International Conference on Information Sciences Signal Processing and their Applications*, pp. 774–777.
- Venkatesh, N. & Jayaraman, S. (2010). Human electrocardiogram for biometrics using DTW and FLDA, *20th International Conference on Pattern Recognition (ICPR)*, pp. 3838–3841.
- Wübbeler, G., Stavridis, M., Kreiseler, D., Boussejot, R. & Elster, C. (2007). Verification of humans using the electrocardiogram, *Pattern Recogn. Lett.* 28(10): 1172–1175.
- Ye, C., Coimbra, M. & Kumar, B. (2010). Investigation of human identification using two-lead electrocardiogram (ECG) signals, *4th Int. Conf. on Biometrics: Theory Applications and Systems*, pp. 1–8.
- Zhang, Z. & Wei, D. (2006). A new ECG identification method using bayes' teorem, *TENCON 2006*, pp. 1–4.

Human Identity Verification Based on Heart Sounds: Recent Advances and Future Directions

Francesco Beritelli and Andrea Spadaccini
Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica (DIEEI)
University of Catania
Italy

1. Introduction

Identity verification is an increasingly important process in our daily lives. Whether we need to use our own equipment or to prove our identity to third parties in order to use services or gain access to physical places, we are constantly required to declare our identity and prove our claim.

Traditional authentication methods fall into two categories: proving that you know something (i.e., password-based authentication) and proving that you own something (i.e., token-based authentication).

These methods connect the identity with an alternate and less rich representation, for instance a password, that can be lost, stolen, or shared.

A solution to these problems comes from biometric recognition systems. Biometrics offers a natural solution to the authentication problem, as it contributes to the construction of systems that can recognize people by the analysis of their anatomical and/or behavioral characteristics. With biometric systems, the representation of the identity is something that is directly derived from the subject, therefore it has properties that a surrogate representation, like a password or a token, simply cannot have (Jain et al. (2006; 2004); Prabhakar et al. (2003)). The strength of a biometric system is determined mainly by the trait that is used to verify the identity. Plenty of biometric traits have been studied and some of them, like fingerprint, iris and face, are nowadays used in widely deployed systems.

Today, one of the most important research directions in the field of biometrics is the characterization of novel biometric traits that can be used in conjunction with other traits, to limit their shortcomings or to enhance their performance.

The aim of this chapter is to introduce the reader to the usage of heart sounds for biometric recognition, describing the strengths and the weaknesses of this novel trait and analyzing in detail the methods developed so far and their performance.

The usage of heart sounds as physiological biometric traits was first introduced in Beritelli & Serrano (2007), in which the authors proposed and started exploring this idea. Their system is based on the frequency analysis, by means of the Chirp z -Transform (CZT), of the sounds produced by the heart during the closure of the mitral tricuspid valve and during the closure of the aortic pulmonary valve. These sounds, called S1 and S2, are extracted from the input

signal using a segmentation algorithm. The authors build the identity templates using feature vectors and test if the identity claim is true by computing the Euclidean distance between the stored template and the features extracted during the identity verification phase.

In Phua et al. (2008), the authors describe a different approach to heart-sounds biometry. Instead of doing a structural analysis of the input signal, they use the whole sequences, feeding them to two recognizers built using Vector Quantization and Gaussian Mixture Models; the latter proves to be the most performant system.

In Beritelli & Spadaccini (2009a;b), the authors further develop the system described in Beritelli & Serrano (2007), evaluating its performance on a larger database, choosing a more suitable feature set (Linear Frequency Cepstrum Coefficients, LFCC), adding a time-domain feature specific for heart sounds, called First-to-Second Ratio (FSR) and adding a quality-based data selection algorithm.

In Beritelli & Spadaccini (2010a;b), the authors take an alternative approach to the problem, building a system that leverages statistical modelling using Gaussian Mixture Models. This technique is different from Phua et al. (2008) in many ways, most notably the segmentation of the heart sounds, the database, the usage of features specific to heart sounds and the statistical engine. This system proved to yield good performance in spite of a larger database, and the final Equal Error Rate (EER) obtained using this technique is 13.70 % over a database of 165 people, containing two heart sequences per person, each lasting from 20 to 70 seconds.

This chapter is structured as follows: in Section 2, we describe in detail the usage of heart sounds for biometric identification, comparing them to other biometric traits, briefly explaining how the human cardio-circulatory system works and produces heart sounds and how they can be processed; in Section 3 we present a survey of recent works on heart-sounds biometry by other research groups; in Section 4 we describe in detail the structural approach; in Section 5 we describe the statistical approach; in Section 6 we compare the performance of the two methods on a common database, describing both the performance metrics and the heart sounds database used for the evaluation; finally, in Section 7 we present our conclusions, and highlight current issues of this method and suggest the directions for the future research.

2. Biometric recognition using heart sounds

Biometric recognition is the process of inferring the identity of a person via quantitative analysis of one or more traits, that can be derived either directly from a person's body (physiological traits) or from one's behaviour (behavioural traits).

Speaking of physiological traits, almost all the parts of the body can already be used for the identification process (Jain et al. (2008)): eyes (iris and retina), face, hand (shape, veins, palmprint, fingerprints), ears, teeth etc.

In this chapter, we will focus on an organ that is of fundamental importance for our life: the heart.

The heart is involved in the production of two biological signals, the Electrocardiograph (ECG) and the Phonocardiogram (PCG). The first is a signal derived from the electrical activity of the organ, while the latter is a recording of the sounds that are produced during its activity (heart sounds).

While both signals have been used as biometric traits (see Biel et al. (2001) for ECG-based biometry), this chapter will focus on hearts-sounds biometry.

2.1 Comparison to other biometric traits

The paper Jain et al. (2004) presents a classification of available biometric traits with respect to 7 qualities that, according to the authors, a trait should possess:

- **Universality:** each person should possess it;
- **Distinctiveness:** it should be helpful in the distinction between any two people;
- **Permanence:** it should not change over time;
- **Collectability:** it should be quantitatively measurable;
- **Performance:** biometric systems that use it should be reasonably performant, with respect to speed, accuracy and computational requirements;
- **Acceptability:** the users of the biometric system should see the usage of the trait as a natural and trustable thing to do in order to authenticate;
- **Circumvention:** the system should be robust to malicious identification attempts.

Each trait is evaluated with respect to each of these qualities using 3 possible qualifiers: H (high), M (medium), L (low).

We added to the original table a row with our subjective evaluation of heart-sounds biometry with respect to the qualities described above, in order to compare this new technique with other more established traits. The updated table is reproduced in Table 1.

The reasoning behind each of our subjective evaluations of the qualities of heart sounds is as follows:

- **High Universality:** a working heart is a *conditio sine qua non* for human life;
- **Medium Distinctiveness:** the actual systems' performance is still far from the most discriminating traits, and the tests are conducted using small databases; the discriminative power of heart sounds still must be demonstrated;
- **Low Permanence:** although to the best of our knowledge no studies have been conducted in this field, we perceive that heart sounds can change their properties over time, so their accuracy over extended time spans must be evaluated;
- **Low Collectability:** the collection of heart sounds is not an immediate process, and electronic stethoscopes must be placed in well-defined positions on the chest to get a high-quality signal;
- **Low Performance:** most of the techniques used for heart-sounds biometry are computationally intensive and, as said before, the accuracy still needs to be improved;
- **Medium Acceptability:** heart sounds are probably identified as unique and trustable by people, but they might be unwilling to use them in daily authentication tasks;
- **Low Circumvention:** it is very difficult to reproduce the heart sound of another person, and it is also difficult to record it covertly in order to reproduce it later.

Of course, heart-sounds biometry is a new technique, and some of its drawbacks probably will be addressed and resolved in future research work.

Biometric identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	M	L	L	M	H
Heart sounds	H	M	L	L	L	M	L

Table 1. Comparison between biometric traits as in Jain et al. (2004) and heart sounds

2.2 Physiology and structure of heart sounds

The heart sound signal is a complex, non-stationary and quasi-periodic signal that is produced by the heart during its continuous pumping work (Sabarimalai Manikandan & Soman (2010)). It is composed by several smaller sounds, each associated with a specific event in the working cycle of the heart.

Heart sounds fall in two categories:

- **primary sounds**, produced by the closure of the heart valves;
- **other sounds**, produced by the blood flowing in the heart or by pathologies;

The primary sounds are S1 and S2. The first sound, S1, is caused by the closure of the tricuspid and mitral valves, while the second sound, S2, is caused by the closure of the aortic and pulmonary valves.

Among the other sounds, there are the S3 and S4 sounds, that are quieter and rarer than S1 and S2, and murmurs, that are high-frequency noises.

In our systems, we only use the primary sounds because they are the two loudest sounds and they are the only ones that a heart always produces, even in pathological conditions. We separate them from the rest of the heart sound signal using the algorithm described in Section 2.3.1.

2.3 Processing heart sounds

Heart sounds are monodimensional signals, and can be processed, to some extent, with techniques known to work on other monodimensional signals, like audio signals. Those

techniques then need to be refined taking into account the peculiarities of the signal, its structure and components.

In this section we will describe an algorithm used to separate the S1 and S2 sounds from the rest of the heart sound signal (2.3.1) and three algorithms used for feature extraction (2.3.2, 2.3.3, 2.3.4), that is the process of transforming the original heart sound signal into a more compact, and possibly more meaningful, representation. We will briefly discuss two algorithms that work in the frequency domain, and one in the time domain.

2.3.1 Segmentation

In this section we describe a variation of the algorithm that was employed in (Beritelli & Serrano (2007)) to separate the S1 and S2 tones from the rest of the heart sound signal, improved to deal with long heart sounds.

Such a separation is done because we believe that the S1 and S2 tones are as important to heart sounds as the vowels are to the voice signal. They are stationary in the short term and they convey significant biometric information, that is then processed by feature extraction algorithms.

A simple energy-based approach can not be used because the signal can contain impulsive noise that could be mistaken for a significant sound.

The first step of the algorithm is searching the frame with the highest energy, that is called SX1. At this stage, we do not know if we found an S1 or an S2 sound.

Then, in order to estimate the frequency of the heart beat, and therefore the period P of the signal, the maximum value of the autocorrelation function is computed. Low-frequency components are ignored by searching only over the portion of autocorrelation after the first minimum.

The algorithm then searches other maxima to the left and to the right of SX1, moving by a number P of frames in each direction and searching for local maxima in a window of the energy signal in order to take into account small fluctuations of the heart rate. After each maximum is selected, a constant-width window is applied to select a portion of the signal.

After having completed the search that starts from SX1, all the corresponding frames in the original signal are zeroed out, and the procedure is repeated to find a new maximum-energy frame, called SX2, and the other peaks are found in the same way.

Finally, the positions of SX1 and SX2 are compared, and the algorithm then decides if SX1, and all the frames found starting from it, must be classified as S1 or S2; the remaining identified frames are classified accordingly.

The nature of this algorithm requires that it work on short sequences, 4 to 6 seconds long, because as the sequence gets longer the periodicity of the sequence fades away due to noise and variations of the heart rate.

To overcome this problem, the signal is split into 4-seconds wide windows and the algorithm is applied to each window. The resulting sets of heart sounds endpoint are then joined into a single set.

2.3.2 The chirp z -transform

The Chirp z -Transform (CZT) is an algorithm for the computation of the z -Transform of sampled signals that offers some additional flexibility to the Fast Fourier Transform (FFT) algorithm.

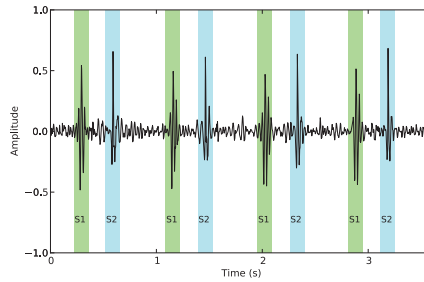


Fig. 1. Example of S1 and S2 detection

The main advantage of the CZT exploited in the analysis of heart sounds is the fact that it allows high-resolution analysis of narrow frequency bands, offering higher resolution than the FFT.

For more details on the CZT, please refer to Rabiner et al. (1969)

2.3.3 Cepstral analysis

Mel-Frequency Cepstrum Coefficients (MFCC) are one of the most widespread parametric representation of audio signals (Davis & Mermelstein (1980)).

The basic idea of MFCC is the extraction of cepstrum coefficients using a non-linearly spaced filterbank; the filterbank is instead spaced according to the Mel Scale: filters are linearly spaced up to 1 kHz, and then are logarithmically spaced, decreasing detail as the frequency increases.

This scale is useful because it takes into account the way we perceive sounds.

The relation between the Mel frequency \hat{f}_{mel} and the linear frequency f_{lin} is the following:

$$\hat{f}_{mel} = 2595 \cdot \log_{10} \left(\frac{1 + f_{lin}}{700} \right) \quad (1)$$

Some heart-sound biometry systems use MFCC, while others use a linearly-spaced filterbank. The first step of the algorithm is to compute the FFT of the input signal; the spectrum is then feeded to the filterbank, and the i -th cepstrum coefficient is computed using the following formula:

$$C_i = \sum_{k=1}^K X_k \cdot \cos \left(i \cdot \left(k - \frac{1}{2} \right) \cdot \frac{\pi}{K} \right) \quad i = 0, \dots, M \quad (2)$$

where K is the number of filters in the filterbank, X_k is the log-energy output of the k -th filter and M is the number of coefficients that must be computed.

Many parameters have to be chosen when computing cepstrum coefficients. Among them: the bandwidth and the scale of the filterbank (Mel vs. linear), the number and spectral width of filters, the number of coefficients.

In addition to this, differential cepstrum coefficients, typically denoted using a Δ (first order) or $\Delta\Delta$ (second order), can be computed and used.

Figure 2 shows an example of three S1 sounds and the relative MFCC spectrograms; the first two (a, b) belong to the same person, while the third (c) belongs to a different person.

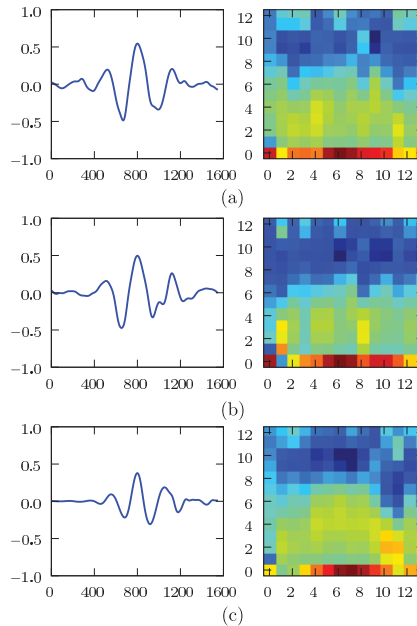


Fig. 2. Example of waveforms and MFCC spectrograms of S1 sounds

2.3.4 The First-to-Second Ratio (FSR)

In addition to standard feature extraction techniques, it would be desirable to develop ad-hoc features for the heart sound, as it is not a simple audio sequence but has specific properties that could be exploited to develop features with additional discriminative power.

This is why we propose a time-domain feature called First-to-Second Ratio (FSR). Intuitively, the FSR represents the power ratio of the first heart sound (S1) to the second heart sound (S2). During our work, we observed that some people tend to have an S1 sound that is louder than S2, while in others this balance is inverted. We try to represent this diversity using our new feature.

The implementation of the feature is different in the two biometric systems that we described in this chapter, and a discussion of the two algorithms can be found in 4.4 and 5.4.

3. Review of related works

In the last years, different research groups have been studying the possibility of using heart sounds for biometric recognition. In this section, we will briefly describe their methods.

In Table 2 we summarized the main characteristics of the works that will be analyzed in this section, using the following criteria:

- **Database** - the number of people involved in the study and the amount of heart sounds recorded from each of them;

- **Features** - which features were extracted from the signal, at frame level or from the whole sequence;
- **Classification** - how features were used to make a decision.

We chose not to represent performance in this table for two reasons: first, most papers do not adopt the same performance metric, so it would be difficult to compare them; second, the database and the approach used are quite different one from another, so it would not be a fair comparison.

Paper	Database	Features	Classification
Phua et al. (2008)	10 people 100 HS each	MFCC LBFC	GMM VQ
Tran et al. (2010)	52 people 100m each	Multiple	SVM
Jasper & Othman (2010)	10 people 20 HS each	Energy peaks	Euclidean distance
Fatemian et al. (2010)	21 people 6 HS each 8 seconds per HS	MFCC, LDA, energy peaks	Euclidean distance
El-Bendary et al. (2010)	40 people 10 HS 10 seconds per HS	autocorrelation cross-correlation complex cepstrum	MSE kNN

Table 2. Comparison of recent works about heart-sound biometrics

In the rest of the section, we will briefly review each of these papers.

Phua et al. (2008) was one of the first works in the field of heart-sounds biometry. In this paper, the authors first do a quick exploration of the feasibility of using heart sounds as a biometric trait, by recording a test database composed of 128 people, using 1-minute heart sounds and splitting the same signal into a train and a testing sequence. Having obtained good recognition performance using the HTK Speech Recognition toolkit, they do a deeper test using a database recorded from 10 people and containing 100 sounds for each person, investigating the performance of the system using different feature extraction algorithms (MFCC, Linear Frequency Band Cepstra (LFBC)), different classification schemes (Vector Quantization (VQ) and Gaussian Mixture Models (GMM)) and investigating the impact of the frame size and of the training/test length. After testing many combinations of those parameters, they conclude that, on their database, the most performing system is composed of LFBC features (60 cepstra + log-energy + 256ms frames with no overlap), GMM-4 classification, 30s of training/test length.

The authors of Tran et al. (2010), one of which worked on Phua et al. (2008), take the idea of finding a good and representative feature set for heart sounds even further, exploring 7 sets of features: temporal shape, spectral shape, cepstral coefficients, harmonic features, rhythmic features, cardiac features and the GMM supervector. They then feed all those features to a feature selection method called RFE-SVM and use two feature selection strategies (optimal and sub-optimal) to find the best set of features among the ones they considered. The tests

were conducted on a database of 52 people and the results, expressed in terms of Equal Error Rate (EER), are better for the automatically selected feature sets with respect to the EERs computed over each individual feature set.

In Jasper & Othman (2010), the authors describe an experimental system where the signal is first downsampled from 11025 Hz to 2205 Hz; then it is processed using the Discrete Wavelet Transform, using the Daubechies-6 wavelet, and the D4 and D5 subbands (34 to 138 Hz) are then selected for further processing. After a normalization and framing step, the authors then extract from the signal some energy parameters, and they find that, among the ones considered, the Shannon energy envelopogram is the feature that gives the best performance on their database of 10 people.

The authors of Fatemian et al. (2010) do not propose a pure-PCG approach, but they rather investigate the usage of both the ECG and PCG for biometric recognition. In this short summary, we will focus only on the part of their work that is related to PCG. The heart sounds are processed using the Daubechies-5 wavelet, up to the 5th scale, and retaining only coefficients from the 3rd, 4th and 5th scales. They then use two energy thresholds (low and high), to select which coefficients should be used for further stages. The remaining frames are then processed using the Short-Term Fourier Transform (STFT), the Mel-Frequency filterbank and Linear Discriminant Analysis (LDA) for dimensionality reduction. The decision is made using the Euclidean distance from the feature vector obtained in this way and the template stored in the database. They test the PCG-based system on a database of 21 people, and their combined PCG-ECG systems has better performance.

The authors of El-Bendary et al. (2010) filter the signal using the DWT; then they extract different kinds of features: auto-correlation, cross-correlation and cepstra. They then test the identities of people in their database, that is composed by 40 people, using two classifiers: Mean Square Error (MSE) and k-Nearest Neighbor (kNN). On their database, the kNN classifier performs better than the MSE one.

4. The structural approach to heart-sounds biometry

The first system that we describe in depth was introduced in Beritelli & Serrano (2007); it was designed to work with short heart sounds, 4 to 6 seconds long and thus containing at least four cardiac cycles (S1-S2).

The restriction on the length of the heart sound was removed in Beritelli & Spadaccini (2009a), that introduced the quality-based best subsequence selection algorithm, described in 4.1.

We call this system “structural” because the identity templates are stored as feature vectors, in opposition to the “statistical” approach, that does not directly keep the feature vectors but instead it represents identities via statistical parameters inferred in the learning phase.

Figure 3 contains the block diagram of the system. Each of the steps will be described in the following sections.

4.1 The best subsequence selection algorithm

The fact that the segmentation and matching algorithms of the original system were designed to work on short sequences was a strong constraint for the system. It was required that a human operator selected a portion of the input signal based on some subjective assumptions. It was clearly a flaw that needed to be addressed in further versions of the system.

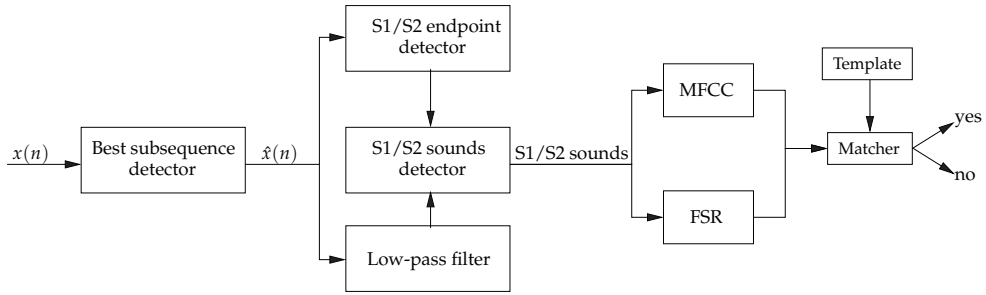


Fig. 3. Block diagram of the proposed cardiac biometry system

To resolve this issue, the authors developed a quality-based subsequence selection algorithm, based on the definition of a quality index $DHS_{QI}(i)$ for each contiguous subsequence i of the input signal.

The quality index is based on a cepstral similarity criterion: the selected subsequence is the one for which the cepstral distance of the tones is the lowest possible. So, for a given subsequence i , the quality index is defined as:

$$DHS_{QI}(i) = \frac{1}{\sum_{k=1}^4 \sum_{\substack{j=1 \\ j \neq k}}^4 d_{S1}(j, k) + \sum_{k=1}^4 \sum_{\substack{j=1 \\ j \neq k}}^4 d_{S2}(j, k)} \tag{3}$$

Where d_{S1} and d_{S2} are the cepstral distances defined in 4.5.

The subsequence \bar{i} with the maximum value of $DHS_{QI}(\bar{i})$ is then selected as the best one and retained for further processing, while the rest of the input signal is discarded.

4.2 Filtering and segmentation

After the best subsequence selection, the signal is then given in input to the heart sound endpoint detection algorithm described in 2.3.1.

The endpoints that it finds are then used to extract the relevant portions of the signal over a version of the heart sound signal that was previously filtered using a low-pass filter, which removed the high-frequency extraneous components.

4.3 Feature extraction

The heart sounds are then passed to the feature extraction module, that computes the cepstral features according to the algorithm described in 2.3.

This system uses $M = 12$ MFCC coefficients, with the addition of a 13-th coefficient computed using an $i = 0$ value in Equation 2, that is the log-energy of the analyzed sound.

4.4 Computation of the First-to-Second Ratio

For each input signal, the system computes the FSR according to the following algorithm.

Let N be the number of complete S1-S2 cardiac cycles in the signal. Let P_{S1_i} (resp. P_{S2_i}) be the power of the i -th S1 (resp. S2) sound.

We can then define \overline{P}_{S1} and \overline{P}_{S2} , the average powers of S1 and S2 heart sounds:

$$\overline{P_{S1}} = \frac{1}{N} \sum_{i=1}^N P_{S1_i} \quad (4)$$

$$\overline{P_{S2}} = \frac{1}{N} \sum_{i=1}^N P_{S2_i} \quad (5)$$

Using these definitions, we can then define the First-to-Second Ratio of a given heart sound signal as:

$$FSR = \frac{\overline{P_{S1}}}{\overline{P_{S2}}} \quad (6)$$

For two given DHS sequences x_1 and x_2 , we define the FSR distance as:

$$d_{FSR}(x_1, x_2) = |FSR_{dB}(x_1) - FSR_{dB}(x_2)| \quad (7)$$

4.5 Matching and identity verification

The crucial point of identity verification is the computation of the distance between the feature set that represents the input signal and the template associated with the identity claimed in the acquisition phase by the person that is trying to be authenticated by the system.

This system employs two kinds of distance: the first in the cepstral domain and the second using the FSR.

MFCC are compared using the Euclidean metric (d_2). Given two heart sound signals X and Y , let $X_{S1}(i)$ (resp. $X_{S2}(i)$) be the feature vector for the i -th S1 (resp. S2) sound of the X signal and Y_{S1} and Y_{S2} the analogous vectors for the Y signal. Then the cepstral distances between X and Y can be defined as follows:

$$d_{S1}(X, Y) = \frac{1}{N^2} \sum_{i,j=1}^N d_2(X_{S1}(i), Y_{S1}(j)) \quad (8)$$

$$d_{S2}(X, Y) = \frac{1}{N^2} \sum_{i,j=1}^N d_2(X_{S2}(i), Y_{S2}(j)) \quad (9)$$

Now let us take into account the FSR. Starting from the d_{FSR} as defined in Equation 7, we wanted this distance to act like an amplifying factor for the cepstral distance, making the distance bigger when it has a high value while not changing the distance for low values.

We then normalized the values of d_{FSR} between 0 and 1 ($d_{FSR_{norm}}$), we chose a threshold of activation of the FSR (th_{FSR}) and we defined k_{FSR} , an amplifying factor used in the matching phase, as follows:

$$k_{FSR} = \max\left(1, \frac{d_{FSR_{norm}}}{th_{FSR}}\right) \quad (10)$$

In this way, if the normalized FSR distance is lower than th_{FSR} it has no effect on the final score, but if it is larger, it will increase the cepstral distance.

Finally, the distance between X and Y can be computed as follows:

$$d(X, Y) = k_{FSR} \cdot \sqrt{d_{S1}(X, Y)^2 + d_{S2}(X, Y)^2} \quad (11)$$

5. The statistical approach to heart-sounds biometry

In opposition to the system analyzed in Section 4, the one that will be described in this section is based on a learning process that does not directly take advantage of the features extracted from the heart sounds, but instead uses them to infer a statistical model of the identity and makes a decision computing the probability that the input signal belongs to the person whose identity was claimed in the identity verification process.

5.1 Gaussian Mixture Models

Gaussian Mixture Models (GMM) are a powerful statistical tool used for the estimation of multidimensional probability density representation and estimation (Reynolds & Rose (1995)). A GMM λ is a weighted sum of N Gaussian probability densities:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^N w_i p_i(\mathbf{x}) \quad (12)$$

where \mathbf{x} is a D -dimensional data vector, whose probability is being estimated, and w_i is the weight of the i -th probability density, that is defined as:

$$p_i(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)'\Sigma_i(\mathbf{x}-\mu_i)}$$

The parameters of p_i are $\mu_i \in \mathbb{R}^D$ and $\Sigma_i \in \mathbb{R}^{D \times D}$, that together with $w_i \in \mathbb{R}^N$ form the set of values that represent the GMM:

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \quad (13)$$

Those parameters of the model are learned in the training phase using the Expectation-Maximization algorithm (McLachlan & Krishnan (1997)), using as input data the feature vectors extracted from the heart sounds.

5.2 The GMM/UBM method

The problem of verifying whether an input heart sound signal s belongs to a stated identity I is equivalent to a hypothesis test between two hypotheses:

$$H_0 : s \text{ belongs to } I$$

$$H_1 : s \text{ does not belong to } I$$

This decision can be taken using a likelihood test:

$$S(s, I) = \frac{p(s|H_0)}{p(s|H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ < \theta & \text{reject } H_0 \end{cases} \quad (14)$$

where θ is the decision threshold, a fundamental system parameter that is chosen in the design phase.

The probability $p(s|H_0)$, in our system, computed using Gaussian Mixture Models.

The input signal is converted by the front-end algorithms to a set of K feature vectors, each of dimension D , so:

$$p(s|H_0) = \prod_{j=1}^K p(x_j|\lambda_I) \quad (15)$$

In Equation 14, the $p(s|H_1)$ is still missing. In the GMM/UBM framework (Reynolds et al. (2000)), this probability is modelled by building a model trained with a set of identities that represent the demographic variability of the people that might use the system. This model is called Universal Background Model (UBM).

The UBM is created during the system design, and is subsequently used every time the system must compute a matching score.

The final score of the identity verification process, expressed in terms of log-likelihood ratio, is

$$\Lambda(s) = \log S(s, I) = \log p(s|\lambda_I) - \log p(s|\lambda_W) \quad (16)$$

5.3 Front-end processing

Each time the system gets an input file, whether for training a model or for identity verification, it goes through some common steps.

First, heart sounds segmentation is carried on, using the algorithm described in Section 2.3.1. Then, cepstral features are extracted using a tool called *sfbccep*, part of the SPro suite (Gravier (2003)). Finally, the FSR, computed as described in Section 5.4, is appended to each feature vector.

5.4 Application of the First-to-Second Ratio

The FSR, as first defined in Section 4.4, is a sequence-wise feature, i.e., it is defined for the whole input signal. It is then used in the matching phase to modify the resulting score.

In the context of the statistical approach, it seemed more appropriate to just append the FSR to the feature vector computed from each frame in the feature extraction phase, and then let the GMM algorithms generalize this knowledge.

To do this, we split the input heart sound signal in 5-second windows and we compute an average FSR (\overline{FSR}) for each signal window. It is then appended to each feature vector computed from frames inside the window.

5.5 The experimental framework

The experimental set-up created for the evaluation of this technique was implemented using some tools provided by ALIZE/SpkDet, an open source toolkit for speaker recognition developed by the ELISA consortium between 2004 and 2008 (Bonastre et al. (n.d.)).

The adaptation of parts of a system designed for speaker recognition to a different problem was possible because the toolkit is sufficiently general and flexible, and because the features used for heart-sounds biometry are similar to the ones used for speaker recognition, as outlined in Section 2.3.

During the world training phase, the system estimates the parameters of the world model λ_W using a randomly selected subset of the input signals.

The identity models λ_i are then derived from the world model W using the Maximum A-Posteriori (MAP) algorithm.

During identity verification, the matching score is computed using Equation 16, and the final decision is taken comparing the score to a threshold (θ), as described in Equation 14

5.6 Optimization of the method

During the development of the system, some parameters have been tuned in order to get the best performance. Namely, three different cepstral feature sets have been considered in (Beritelli & Spadaccini (2010b)):

- $16 + 16 \Delta + E + \Delta E$
- $16 + 16 \Delta + 16 \Delta \Delta$
- $19 + 19 \Delta + E + \Delta E$

However, the first of these sets proved to be the most effective

In (Beritelli & Spadaccini (2010a)) the impact of the FSR and of the number of Gaussian densities in the mixtures was studied. Four different model sizes (128, 256, 512, 1024) were tested, with and without FSR, and the best combination of those parameters, on our database, is 256 Gaussians with FSR.

6. Performance evaluation

In this section, we will compare the performance of the two systems described in Section 4 and 5 using a common heart sounds database, that will be further described in Section 6.1.

6.1 Heart sounds database

One of the drawbacks of this biometric trait is the absence of large enough heart sound databases, that are needed for the validation of biometric systems. To overcome this problem, we are building a heart sounds database suitable for identity verification performance evaluation.

Currently, there are 206 people in the database, 157 male and 49 female; for each person, there are two separate recordings, each lasting from 20 to 70 seconds; the average length of the recordings is 45 seconds. The heart sounds have been acquired using a Thinklabs Rhythm Digital Electronic Stethoscope, connected to a computer via an audio card. The sounds have been converted to the Wave audio format, using 16 bit per second and at a rate of 11025 Hz. One of the two recordings available for each person used to build the models, while the other is used for the computation of matching scores.

6.2 Metrics for performance evaluation

A biometric identity verification system can be seen as a binary classifier.

Binary classification systems work by comparing matching scores to a threshold; their accuracy is closely linked with the choice of the threshold, which must be selected according to the context of the system.

There are two possible errors that a binary classifier can make:

- **False Match (Type I Error):** accept an identity claim even if the template does not match with the model;
- **False Non-Match (Type II Error):** reject an identity claim even if the template matches with the model

The importance of errors depends on the context in which the biometric system operates; for instance, in a high-security environment, a Type I error can be critical, while Type II errors could be tolerated.

When evaluating the performance of a biometric system, however, we need to take a threshold-independent approach, because we cannot know its applications in advance. A common performance measure is the Equal Error Rate (EER) (Jain et al. (2008)), defined as the error rate at which the False Match Rate (FMR) is equal to the False Non-Match Rate (FNMR). A finer evaluation of biometric systems can be done by plotting the Detection Error Tradeoff (DET) curve, that is the plot of FMR against FNMR. This allows to study their performance when a low FNMR or FMR is imposed to the system.

The DET curve represents the trade-off between security and usability. A system with low FMR is a highly secure one but will lead to more non-matches, and can require the user to try the authentication step more times; a system with low FNMR will be more tolerant and permissive, but will make more false match errors, thus letting more unauthorized users to get a positive match. The choice between the two setups, and between all the intermediate security levels, is strictly application-dependent.

6.3 Results

The performance of our two systems has been computed over the heart sounds database, and the results are reported in Table 3.

System	EER (%)
Structural	36.86
Statistical	13.66

Table 3. Performance evaluation of the two heart-sounds biometry systems

The huge difference in the performance of the two systems reflects the fact that the first one is not being actively developed since 2009, and it was designed to work on small databases, while the second has already proved to work well on larger databases.

It is important to highlight that, in spite of a 25% increment of the size of the database, the error rate remained almost constant with respect to the last evaluation of the system, in which a test over a 165 people database yielded a 13.70% EER.

Figure 4 shows the Detection Error Trade-off (DET) curves of the two systems. As stated before, a DET curve shows how the analyzed system performs in terms of false matches/false non-matches as the system threshold is changed.

In both cases, fixing a false match (resp. false non-match) rate, the system that performs better is the one with the lowest false non-match (resp. false match) rate.

Looking at Figure 4, it is easy to understand that the statistical system performs better in both high-security (e.g., FMR = 1-2%) and low-security (e.g., FNMR = 1-2%) setups.

We can therefore conclude that the statistical approach is definitely more promising than the structural one, at least with the current algorithms and using the database described in 6.1..

7. Conclusions

In this chapter, we presented a novel biometric identification technique that is based on heart sounds.

After introducing the advantages and shortcomings of this biometric trait with respect to other traits, we explained how our body produces heart sounds, and the algorithms used to process them.

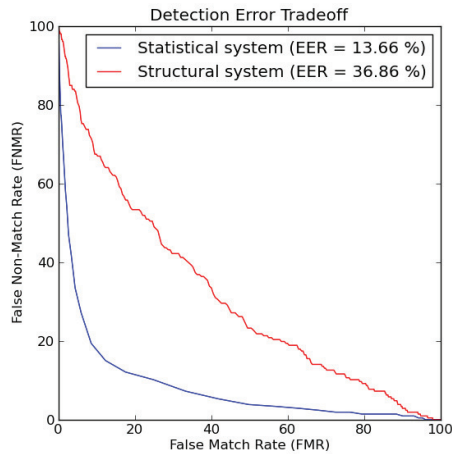


Fig. 4. Detection Error Tradeoff (DET) curves of the two systems

A survey of recent works on this field written by other research groups has been presented, showing that there has been a recent increase of interest of the research community in this novel trait.

Then, we described the two systems that we built for biometric identification based on heart sounds, one using a structural approach and another leveraging Gaussian Mixture Models. We compared their performance over a database containing more than 200 people, concluding that the statistical system performs better.

7.1 Future directions

As this chapter has shown, heart sounds biometry is a promising research topic in the field of novel biometric traits.

So far, the academic community has produced several works on this topic, but most of them share the problem that the evaluation is carried on over small databases, making the results obtained difficult to generalize.

We feel that the community should start a joint effort for the development of systems and algorithms for heart-sounds biometry, at least creating a common database to be used for the evaluation of different research systems over a shared dataset that will make possible to compare their performance in order to refine them and, over time, develop techniques that might be deployed in real-world scenarios.

As larger databases of heart sounds become available to the scientific community, there are some issues that need to be addressed in future research.

First of all, the identification performance should be kept low even for larger databases. This means that the matching algorithms will be fine-tuned and a suitable feature set will be identified, probably containing both elements from the frequency domain and the time domain.

Next, the mid-term and long-term reliability of heart sounds will be assessed, analyzing how their biometric properties change as time goes by. Additionally, the impact of cardiac diseases on the identification performance will be assessed.

Finally, when the algorithms will be more mature and several independent scientific evaluations will have given positive feedback on the idea, some practical issues like computational efficiency will be tackled, and possibly ad-hoc sensors with embedded matching algorithms will be developed, thus making heart-sounds biometry a suitable alternative to the mainstream biometric traits.

8. References

- Beritelli, F. & Serrano, S. (2007). Biometric Identification based on Frequency Analysis of Cardiac Sounds, *IEEE Transactions on Information Forensics and Security* 2(3): 596–604.
- Beritelli, F. & Spadaccini, A. (2009a). Heart sounds quality analysis for automatic cardiac biometry applications, *Proceedings of the 1st IEEE International Workshop on Information Forensics and Security*.
- Beritelli, F. & Spadaccini, A. (2009b). Human Identity Verification based on Mel Frequency Analysis of Digital Heart Sounds, *Proceedings of the 16th International Conference on Digital Signal Processing*.
- Beritelli, F. & Spadaccini, A. (2010a). An improved biometric identification system based on heart sounds and gaussian mixture models, *Proceedings of the 2010 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, IEEE, pp. 31–35.
- Beritelli, F. & Spadaccini, A. (2010b). A statistical approach to biometric identity verification based on heart sounds, *Proceedings of the Fourth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE2010)*, IEEE, pp. 93–96. URL: <http://dx.medra.org/10.1109/SECURWARE.2010.23>
- Biel, L. & Pettersson, O. & Philipson, L. & Wide, P. (2001). ECG Analysis: A New Approach in Human Identification, *IEEE Transactions on Instrumentation and Measurement* 50(3): 808–812.
- Bonastre, J.-F., Scheffer, N., Matrouf, D., Fredouille, C., Larcher, A., Preti, R., Pouchoulin, G., Evans, N., Fauve, B. & Mason, J. (n.d.). Alize/Spkdet: a state-of-the-art open source software for speaker recognition.
- Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech and Signal Processing* 28(4): 357–366.
- El-Bendary, N., Al-Qaheri, H., Zawbaa, H. M., Hamed, M., Hassanien, A. E., Zhao, Q. & Abraham, A. (2010). Hsas: Heart sound authentication system, *Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress on*, pp. 351–356.
- Fatemian, S., Agrafioti, F. & Hatzinakos, D. (2010). Heartid: Cardiac biometric recognition, *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pp. 1–5.
- Gravier, G. (2003). SPro: speech signal processing toolkit. URL: <http://gforge.inria.fr/projects/spro>
- Jain, A. K., Flynn, P. & Ross, A. A. (2008). *Handbook of Biometrics*, Springer.
- Jain, A. K., Ross, A. A. & Pankanti, S. (2006). Biometrics: A tool for information security, *IEEE Transactions on Information Forensics and Security* 1(2): 125–143.
- Jain, A. K., Ross, A. A. & Prabhakar, S. (2004). An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video Technology* 14(2): 4–20.

- Jasper, J. & Othman, K. (2010). Feature extraction for human identification based on envelopegram signal analysis of cardiac sounds in time-frequency domain, *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, Vol. 2, pp. V2-228 –V2-233.
- McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*, Wiley.
- Phua, K., Chen, J., Dat, T. H. & Shue, L. (2008). Heart sound as a biometric, *Pattern Recognition* 41(3): 906–919.
- Prabhakar, S., Pankanti, S. & Jain, A. K. (2003). Biometric recognition: Security & privacy concerns, *IEEE Security and Privacy Magazine* 1(2): 33–42.
- Rabiner, L., Schafer, R. & Rader, C. (1969). The chirp z-transform algorithm, *Audio and Electroacoustics, IEEE Transactions on* 17(2): 86 – 92.
- Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models, *Digital Signal Processing*, p. 2000.
- Reynolds, D. A. & Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models, *IEEE Transactions on Speech and Audio Processing* 3: 72–83.
- Sabarimalai Manikandan, M. & Soman, K. (2010). Robust heart sound activity detection in noisy environments, *Electronics Letters* 46(16): 1100 –1102.
- Tran, D. H., Leng, Y. R. & Li, H. (2010). Feature integration for heart sound biometrics, *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 1714 –1717.

Investigation of Temporal Change in Heartbeat in Transition of Sound and Music Stimuli

Makoto Fukumoto and Hiroki Hasegawa

*Fukuoka Institute of Technology,
Mukogawa Women's University,
Japan*

1. Introduction

Music is widely believed as one of the most effective media forms that effect on human psycho-physiologically. Many people expect the effects of music and uses music pieces in various situations. For example, relaxation effect of music is used in change in personal mind in daily life, creation of sedative atmosphere in a home, and therapeutic purposes. Oppositely, parts of music pieces excite people in disco and party. These various effects of music have been investigated in many previous studies. Especially, relaxation effects of music were investigated from various view points with psycho-physiological experiments. Although many previous studies investigated the effects, the effects have not been clarified at all. One of the reasons of that is existence of many music factors; melody, tempo, harmony, rhythm, etc. Anyway, investigations of music and its effects will contribute to theoretical use of music for therapy and so on.

This chapter aims to investigate the temporal change in heartbeat intervals in a transition between different sound stimuli. Heartbeat is one of the most important physiological indices and is often used as physiological index to investigate the effect of music and sound stimuli, because the heartbeat reflects autonomic nervous activity [Pappano, 2008] of a listener and is easy to measure. Furthermore, a device measuring electrocardiogram is generally cheaper than devices measuring other physiological indices such as electroencephalogram. Although many previous studies have investigated the effect of music and sound with heartbeat interval as physiological index, very few previous studies have investigated the change in heartbeat in the transition of different stimulus; most of the previous studies have observed average of heartbeat intervals a certain range in pre- and post-listening sound stimulus. Observing temporal change in heartbeat is important and contributes to improvement of exposure method of music and sound to the listeners. For example, time intervals eliciting effect of music and sound and decreasing the effect are important information to determine the time length to exposure them to the listeners.

The experimental method in the present study is set by referring to our previous study [Fukumoto et al., 2009]. As further investigations, we newly add No-sound stimulus to relaxation music piece (*Air* by Bach), white noise, which are music and sound stimuli employed as sound stimuli in the previous study. In the listening experiment of our previous study, two min relaxing music piece and white noise were employed as the different sound stimuli, and these sound stimuli were played twice alternately after five min rest. The alternate exposure of different sound stimuli is also employed in the present study.

By employing No-sound stimulus as sound stimulus, we can observe change in heartbeat in start and end of noise and music stimuli. It means observing the change in heartbeat caused by listening sound stimuli.

Objectives of this study are fundamental investigations of temporal change in heartbeat in a transition between sound stimuli; music, noise, and No-sound. Figure 1 illustrates the objectives and explains that listening sounds elicit deceleration or acceleration of heartbeat. Time cost is also investigated. The results of this study will contribute to develop the studies in music therapy and the studies on musical system reflecting a user's *KANSEI* information automatically.

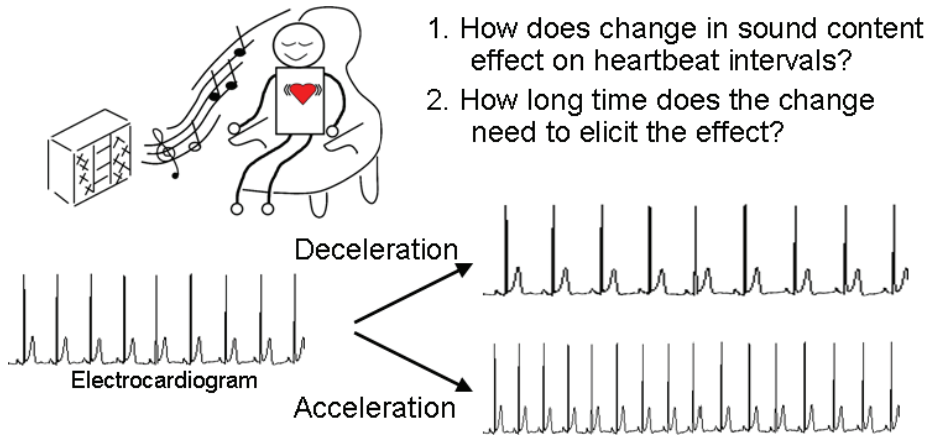


Fig. 1. Illustration of objectives of this study.

As mentioned above, we generally believe relaxation and excitation effects of sounds and utilize the effect in music therapy, reducing patient's anxiety during a surgical operation and personal use to quickly change our mind and so on. Especially for the effects of music, many researchers have investigated the effect with psycho-physiological indices [Dainow, 1977], however, the effects have not been clarified completely yet.

Heartbeat has been often used to investigate the effects of music, because heartbeat is non-invasive physiological index being measured with relatively convenient and low-cost device. Moreover, as mentioned above, heartbeat reflects autonomic nervous activity, and the decrease of the heart rate is caused by a combination of two factors; increase in parasympathetic activity and decrease in sympathetic activity [Pappano, 2008]. Most of the previous studies using heartbeat as physiological index have tried to investigate the effect of music by comparing heartbeats before and after listening to the musical piece. However, heartbeat information has not been used well, because we do not know how long time the sound stimuli would cost to elicit the change of heartbeat. According to a previous study, impression for the listening to music piece on the listener is determined with 1 s [Bigand et al., 2005]. Referring to this finding, response of heartbeat needs longer than 1 s, because general physiological changes come after psychological changes.

Some previous studies have investigated the temporal cost that sound elicits heartbeat approximately with different sound stimuli and different conditions. Etzel et al. have investigated the physiological effects of musical pieces inducing different moods on listener [Etzel et al., 2005]. They have not mentioned about the concrete time cost, however, the

results of heart rate development seem to elicit the change of heart rate from 30 s to 40 s. Gomez et al. have investigated the physiological effects of 30 s noise and music [Gomez & Danuser, 2004]. The results of their study showed that part of noises and musical pieces elicited physiological change including heart rate within 30 s. Moreover, Hazama et al. have associated listener's preference (with 2-point scale) and heartbeat interval for 40 s musical pieces obtained from their evolutionary computation method composing musical piece [Hazama & Fukumoto, 2009]. Their result showed different distribution of heartbeat intervals for preferred and not preferred musical pieces. These previous studies give us useful and interesting information about the time cost, however, precise time that sound stimulus need to elicit the change of heartbeat has not been clarified.

In our previous study [Fukumoto et al., 2009], average of heartbeat intervals in music section was larger than that in noise section significantly ($P < 0.05$). Furthermore, after the transition of sound stimuli, it took about 30 s to change heartbeat interval from previous section. In the investigation of the change in heartbeat interval, 20 s sliding window was employed. Additionally, a questionnaire asking relaxation feeling was used as psychological index. After the exposure of the sound stimuli, a questionnaire asks the subjects relaxation feeling for noise and music sections, respectively. Psychological result showed that music section induced the higher relaxation feeling than noise section significantly ($P < 0.001$). The significant difference in subjective relaxation feeling meant that the relaxing music piece and the noise were greatly different. These results in our previous study support the findings by other previous studies that investigated the effects of sound stimuli on heartbeat.

What kind areas does this study contribute for? First, as described above, theoretical use of music is first candidate of the application of this study. For example, in Guided Imagery Method, patients listen to several music pieces for a long time. If a therapist know a time length that music pieces need to elicit change in heartbeat, the knowledge must be useful to make a selection of music pieces.

From engineering point of view, with a mind to develop a musical system that reflects user's *KANSEI* and psychological condition of listeners automatically, several previous studies have investigated the psycho-physiological response for sound stimulus with various physiological indices [Aoto & Ookura, 2007; Chung & Vercoe, 2006; Hazama & Fukumoto, 2009; Healy et al., 1998; Kim & André, 2004; Sugimoto et al., 2008; Yoshida et al., 2006]. Heartbeat is included the indices, and revealing time cost eliciting the change of heartbeat by sounds will contribute to musical information techniques such as automatic musical composition based on physiological index [Fukumoto & Imai, 2008].

This section has explained the background, the previous studies, and the applications as introductions of this study. The remains of this chapter are constructed as below. The section 2 describes experimental method and sound materials, and the section 3 shows results of the experiment. The section 4 discusses the effects of the sound stimuli on temporal change in heartbeat based on the experimental results. Finally, the section 5 concludes this chapter.

2. Procedure and materials

This section describes experimental method used in this study. Basically, the experimental method used in this study is referring to our previous study [Fukumoto et al., 2009]: different sound stimuli are included in one experimental set, and electrocardiogram is measured in listening to the sound stimuli. Music and noise were used in our previous

study. To investigate the effects of various transitions of sound stimuli, we add mute sound as a sound stimulus. Therefore, three kinds of transitions and their inverted sequences are used in the listening experiment.

2.1 Procedure

Time length of one experimental set was thirteen min. The set of the listening experiment was basically constructed from two parts; five min rest and eight min sound stimuli. The part of eight min sound stimuli was composed of two min different sound stimuli. These sounds were played two times respectively with change places their sequence. Figure 2 shows example of one experimental set. Especially, the change in heartbeat intervals in second presentations of Sound A and B were used in analysis, because the subjects already realized the sound stimuli in second presentations; remove of surprise for first time listening to the sound stimulus. Based on this experimental set, three experiments were performed, and each of the experiments included two different sound stimuli; Noise and Music, Noise and No-sound, Music and No-sound.

Sixteen males and females (mean age: 21.9 ± 0.7 years) participated in the listening experiments as subjects. None of the subjects had professional or college-level music experience. Beforehand for the experiments, the subjects were instructed not to eat, drink and smoke anything from 30 min before the listening experiment. As described in the next section, each of three experiments included two conditions, and all of the subjects participated in the one condition in all of three experiments. Therefore, there were eight kind of combination of experimental conditions: Each of the subjects participated in three experimental sets. Sixteen subjects were randomly and counter-balanced assigned to the eight combinatins.

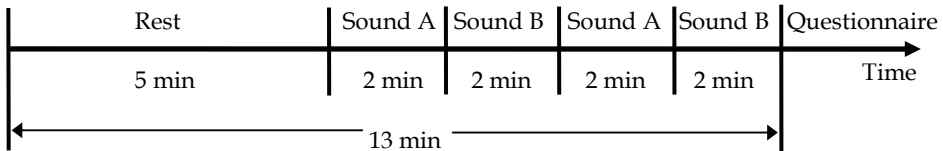


Fig. 2. Experimental procedure of one experimental set.

2.2 Sound stimuli

As sound stimuli, a music piece, white noise, and no sound were employed. In the listening experiment, *Air in G* composed by Bach was used as relaxing musical piece in music sections. This music piece is well known as its relaxing mood and was used as sedative musical piece in a previous study [Yamada et al., 2000]. White noise was used in noise sections. Format of these sound stimuli was WAVE format, and these sound stimuli were played by notebook. Both of the music piece and white noise were stereo recorded sound. In the no-sound section, sound was not played. The subjects listened to these sound stimuli through a headphone.

With three experiments, the change in heart beat in the transition between different sounds was investigated. Each of the three experiments was mainly composed of two different sound stimuli. Figures from 3 to 5 show outline of waveforms of sound stimuli including first 5 min rest. Horizontal axis means time, and vertical axis means sound amplitude. As shown in the outline of waveforms, to prevent to induce strange feeling to the subjects, music stimuli and noise were introduced after 10 s of fade-in and finished with 10 s of fade-out. This control of volume enabled us to construct the experiment composed of continuous

different sound stimuli. In our previous study [Fukumoto et al., 2009], volume of sound stimuli was adjusted initially by the subjects themselves by listening to white noise before the listening experiment. In the present study, to adjust the experimental condition between the subjects further, the volume of the sound stimuli were fixed. The volume of noise and musical piece was around 66.0 to 70.0 dB(A).

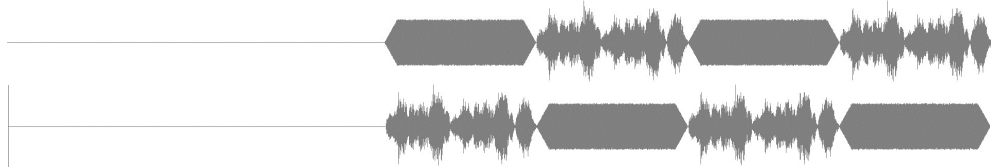


Fig. 3. Volumes of sound stimuli used in the experiment 1: (upper: Noise to Music condition, lower: Music to Noise condition).



Fig. 4. Volumes of sound stimuli used in the experiment 2: (upper: Noise to No-sound condition, lower: No-sound to Noise condition).

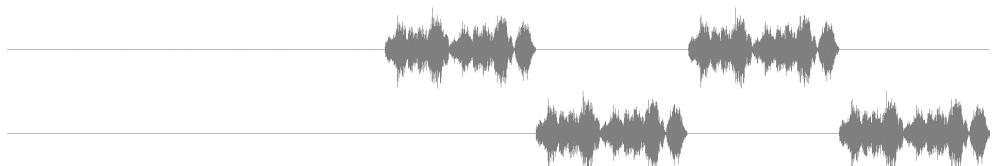


Fig. 5. Volumes of sound stimuli used in the experiment 3: (upper: Music to No-sound condition, lower: No-sound to Music condition).

2.3 Psychological index

In the questionnaire after listening all of the sound stimuli, semantic differential method [Osgood et al., 1957] was used to ask the subjects relaxation feeling. The subjects estimated their relaxation feelings with 7-point scale of “relaxed - stressful” for the afforded two sound stimuli, respectively. Additionally, in the experiment 1, the questionnaire also asked the subjects that the subject had an experience listening to the musical piece played in music sections. These questions were written on a paper in Japanese and were explained by an experimenter, and the subjects answered with writing on the paper by themselves. Additionally, in statistical analysis for relaxation feeling, sign test was used.

2.4 Physiological index

In the physiological analysis, R-waves were detected from the subjects' electrocardiogram. R-wave approximately represents time of heartbeat, and R-R intervals represent temporal development of heartbeat intervals. Development of heartbeat interval is represented from $(t_n, t_{n+1}-t_n)$, where t_n means time of n-th R-wave. In the analyses, 2 min and 20 s windows

were used to observe the change in heartbeat intervals. First, all heartbeat intervals were detected from electrocardiogram. Then, average of heartbeat intervals included in each window (t_n is in the temporal range of each window) was calculated.

Two kinds of windows were used for different observation. 2 min window was same as its temporal length of each section and was used for broad and rough observation for change in heartbeat intervals in each section. On the other hand, 20 s was used for detail observation, and temporal length of 20 s window was determined referring to a method of a previous study [Yoshida et al., 2006]. Heartbeat intervals in general condition contain two kinds of period of heartbeat fluctuations reflecting autonomic nervous activity and respiration, and the temporal lengths are 4 s and 10 s (in 60 heartbeats per 1-min). Length of 20 s window was available for omitting these heartbeat fluctuations. In a part of the analyses, the 20 s window slid as queue processing, and time of the window was defined as central time obtained from average of first and last time of the window. The analyses with 20 s window were mainly applied for latter two listening sections.

3. Experimental results

This section shows the psychological and physiological results of the listening experiment. In the statistical analyses, pairwise comparison was used, because individual variation of subjective evaluation and physiological indices were considered large.

3.1 Results of psychological index

First, result of questionnaire only for the experiment 1 showed that all of the subjects knew the music piece played in music sections.

Figure 6 shows subjective relaxation feeling in the experiment 1. Higher point means higher subjective relaxation feeling. As shown in Fig. 6, large difference between relaxation feelings of Noise and Music conditions was observed. Statistical analysis for this result showed that there was significant difference between Noise and Music conditions ($P < 0.001$).

Figure 7 shows subjective relaxation feelings of Noise and No-music Stimuli in the experiment 2. Average point of No-sound was larger than that of Noise a little bit, however, there was no significant difference.

Figure 8 shows subjective relaxation feelings of Music and No-sound in the experiment 3. Relaxation point of Music was also larger than that of No-sound ($P < 0.001$).

As summary of the results of relaxation feelings, Music stimulus elicited the largest relaxation feeling. Noise elicited the smallest. No-sound was intermediate, however, Noise and No-sound were almost same level.

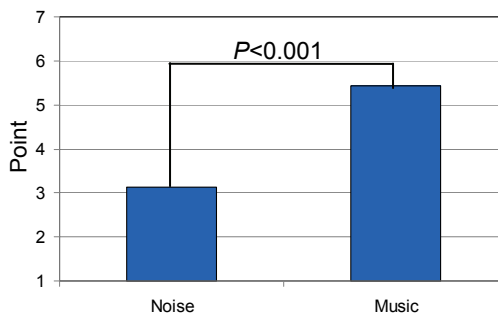


Fig. 6. Subjective relaxation feeling to Noise and Music stimuli in the experiment 1 ($N=16$).

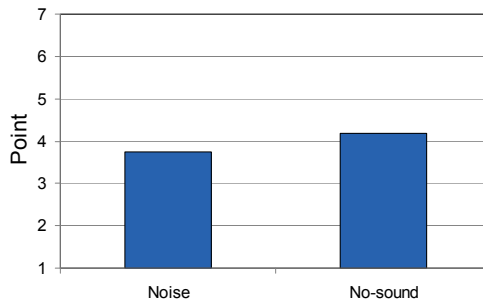


Fig. 7. Subjective relaxation feeling to Noise and No-sound stimuli in the experiment 2 ($N=16$).

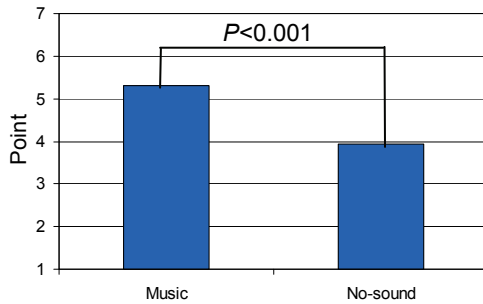


Fig. 8. Subjective relaxation feeling to Music and No-sound stimuli in the experiment 3 ($N=16$).

3.2 Results of physiological index

This subsection shows the results of physiological index. First, rough change in heartbeat intervals in each section is investigated. Then, detail change in heartbeat is observed with 20 s window.

3.2.1 Change in heartbeat in each 2 min section

Figures from 9 to 11 show whole changes in heartbeat in each three experiment, respectively. In the analysis, average heartbeat intervals in 2 min sections were used. For rest section prior to listening to sound stimuli, last 2 min in 5 min rest was utilized as analyzed section. Each result was composed of average and standard deviation between subjects and was obtained after analysis of each subject.

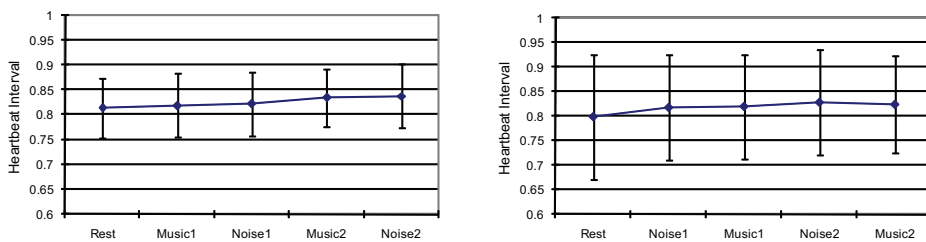


Fig. 9. Average heartbeat interval in each 2 min section in the experiment 1 (Left: Music to Noise condition ($N=8$), Right: Noise to Music condition ($N=8$)).

A common tendency between the results shown in Figures from 9 to 11 was gradual extend of heartbeat intervals in correspondence with time development. It means that lowest average heartbeat was observed in the rest section except Music to No-sound condition in the experiment 3. From these results, large differences between different sound stimuli were not observed.

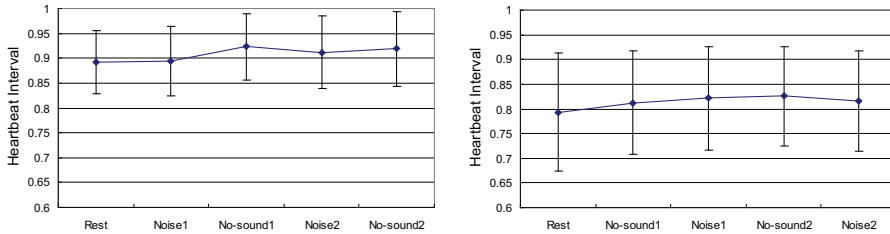


Fig. 10. Average heartbeat interval in each 2 min section in the experiment 2 (Left: Noise to No-sound condition ($N=8$), Right: No-sound to Noise condition ($N=8$)).

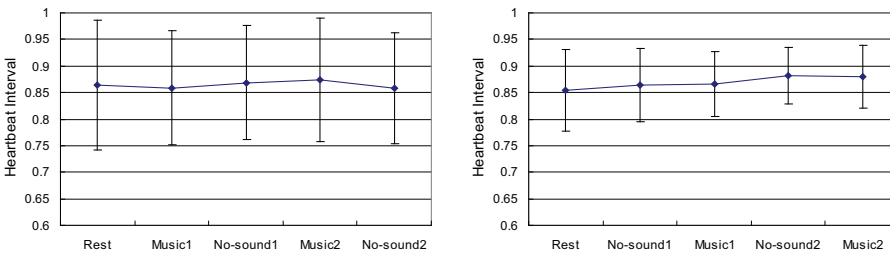


Fig. 11. Average heartbeat interval in each 2 min section in the experiment 3 (Left: Music to No-sound condition ($N=8$), Right: No-sound to Music condition ($N=8$)).

3.2.2 Detail change in heartbeat in a transition of sound stimuli

Figures 12 and 13 show the results of the experiment 1. Figure 12 shows change in heartbeat in latter 2 sections from 540 s to 780 s with 20 s window. The sound changed its content at 660 s. We can observe gradual change in heartbeat and tendencies that shorten of heartbeat in noise sections and extension of heartbeat in music sections. To investigate time cost eliciting change in heartbeat by listening to each sound stimulus, 20 s sliding window was used. As previous analysis section, 650 s window (640 s to 660 s) was used. From 650 s and latter windows are target sections and compared with 650 s statistically. Figure 13 shows

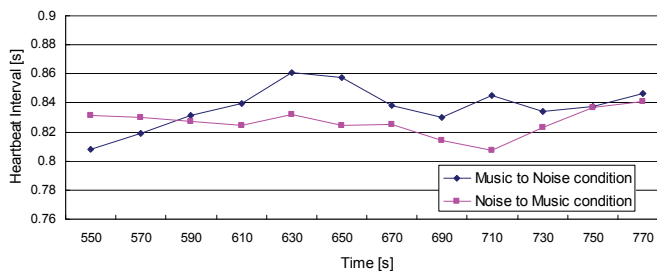


Fig. 12. Average heartbeat interval in 20 s sections from 540 to 780 s in the experiment 1.

the results of analysis with sliding window. Gradual change with sliding window was observed, and plotted line means P -value obtained from statistical analysis. In the result of Noise to Music condition, heartbeat interval tended to be extended from previous section in 761 s window. It means that 111 s was needed to change heartbeat interval by listening to music from listening to noise. In the result of Music to Noise, heartbeat interval was shortened in 678 s window significantly: Time cost to elicit the change of heartbeat from Music to Noise was 28 s.

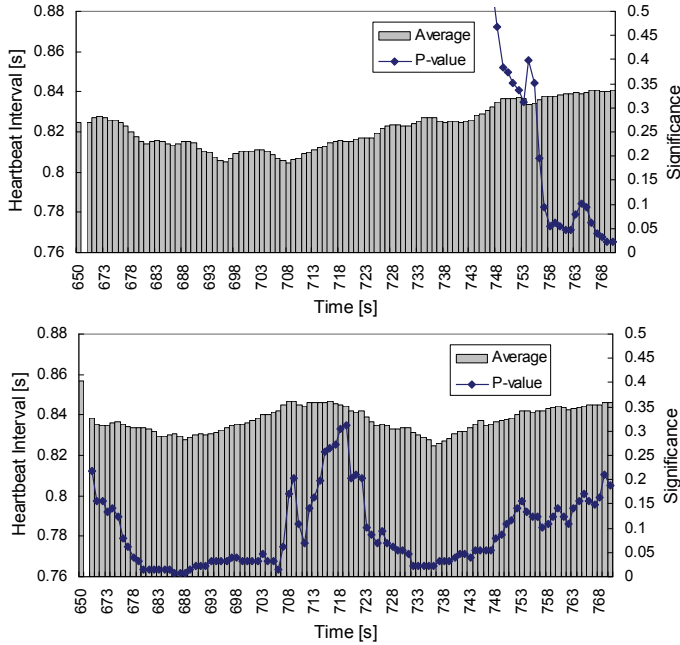


Fig. 13. Detail observation in the transition using 20 s sliding window in the experiment 1 (upper: Noise to Music condition, lower: Music to Noise condition).

Figure 14 shows change in heartbeat in latter 2 sections in the experiment 2. Gradual change in heartbeat was also observed in this result.

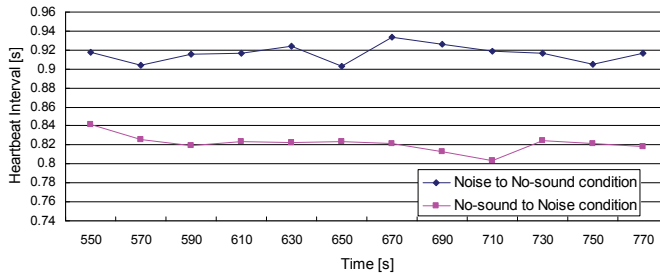


Fig. 14. Average heartbeat interval in 20 s sections from 540 to 780 s in the experiment 2.

Figure 15 shows the results of analysis with sliding window. In the result of Noise to No-sound condition, heartbeat interval tended to be extended from previous section in 671 s

window: Smallest P -value was observed in the section ($P = 0.0547$). After that, heartbeat was gradually extended. 21 s was needed to change heartbeat interval from listening to noise. In the result of No-sound to Noise, heartbeat interval was shortened around 690 s window, furthermore, it was shortened in 756 s window significantly.

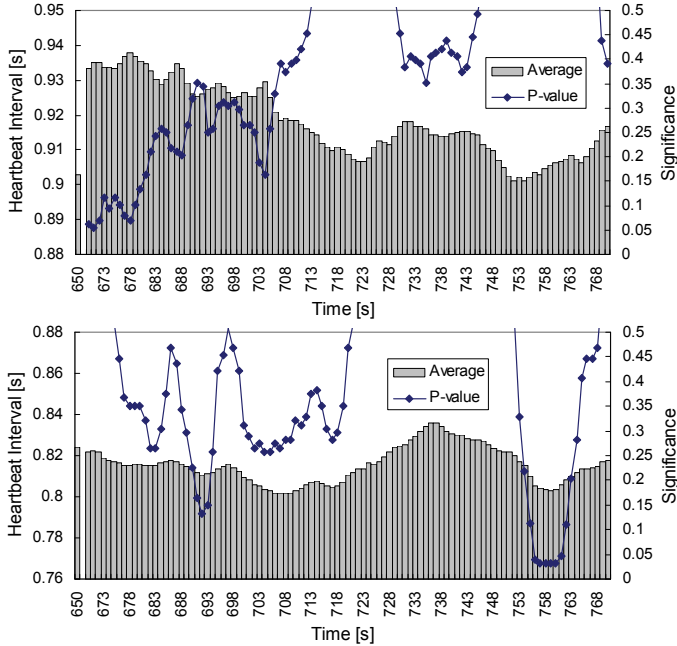


Fig. 15. Detail observation in the transition using 20 s sliding window in the experiment 2 (upper: Noise to No-sound condition, lower: No-sound to Noise condition).

Figure 16 shows change in heartbeat in latter 2 sections in the experiment 3. Gradual and rapid changes in heartbeat were observed in this result.

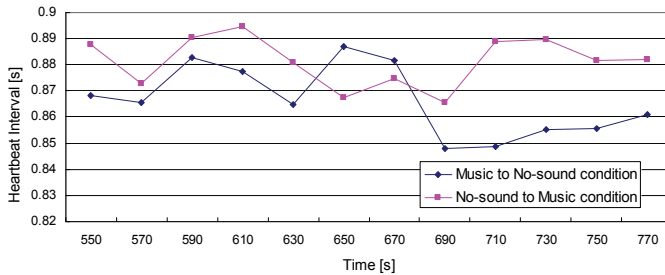


Fig. 16. Average heartbeat interval in 20 s sections from 540 to 780 s in the experiment 3.

Figure 17 shows the results of analysis with sliding window. In the result of Music to No-sound condition, heartbeat interval tended to be shortened from previous section in 686 s window. After that, heartbeat interval kept same level till the end of the listening experiment. In the result of No-sound to Music, heartbeat interval was extended around 752 s window.

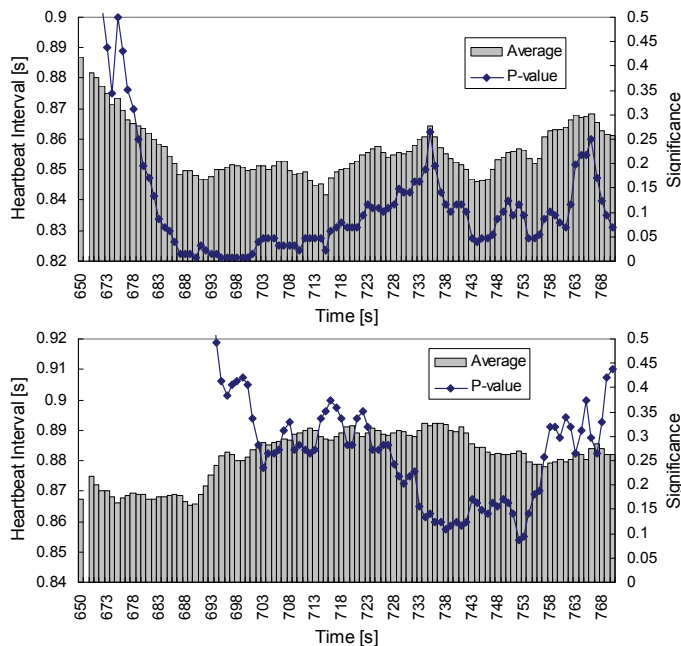


Fig. 17. Detail observation in the transition using 20 s sliding window in the experiment 3 (upper: Music to No-sound condition, lower: No-sound to Music condition).

4. Discussion

Result of relaxation feelings of sound stimuli showed that music elicited highest relaxation and noise elicited lowest one. These results support our previous study with more concrete investigations. No-sound was intermediate. The sequence of them is reasonable, however, difference between Noise and No-sound was little (no significant). Generally, noise is believed as sound everyone dislikes. One of the reasons for that is the subjects were in the listening experiment without any task. If they have tasks to do, noise deny doing the tasks, therefore, they might feel the noise as more negative. In addition, adjustment of volume of sound stimuli between the subjects might reduce the volume for some of the subjects (The adjustment was not applied for the sound stimuli in our previous study). The reduction of volume might elicit little negative impression for noise.

Entire change in 2 min average heartbeat interval did not show large difference between sections, while a significant difference was observed between music and noise conditions in our previous study [Fukumoto et al., 2009]. Adjustment of volume might effect on change in heartbeat interval as small change. Same tendency between the present and the previous study as gradual extend of heartbeat interval was observed. The tendency is considered as caused of sitting on a chair for a long time.

Detail observation in a transition between different stimuli showed obvious change in heartbeat intervals. In the all experiments, tendency and significant change in heartbeat were observed. Time cost eliciting the change and direction of the change (extension or shorten) was quite different between experimental conditions: The direction of change obeyed subjective relaxation feelings. In our previous study that employed music and noise

as sound stimuli, about 30 s was needed to observe obvious change in heartbeat interval from previous condition to post transition. In the present study, time cost of noise to elicit the change from music piece was 28 s. While the time cost of noise was almost same time length as our previous study, the time cost of music was longer than our previous study: 111 s was needed. As mentioned above, the difference was considered as caused of adjustment of sounds' volume.

For Noise condition, the time costs were 28 s from Music. From No-sound to Noise, to observe significant change, it costs 106 s. However, around 690 s window, shorten of heartbeat was observed. With these results, Noise elicited the change in heartbeat than Music. The reason why noise affects on heartbeat earlier than music is considered that noise is a continuous sound content: from start to end, noise sound contents unpleasant sound. On the other hand, music effects on psycho-physiologically by its development. The different of sound contents might be cause of difference of time cost. Additionally, physiological change is considered faster for unpleasant and dangerous stimuli, because the human have to defend or run away from dangerous things quickly.

No-sound was a new additional condition from our previous study, and it played a role of release from music and noise and obvious start of the stimuli from no-sound. From noise to no-sound, it was needed almost 20 s to elicit the extension heartbeat intervals: With the analysis used in the present study, shorter reaction was not measured. The time cost was the fastest among the experimental conditions. Furthermore, from Music to No-sound, 36 s was needed to shorten the heartbeat intervals. The time cost was shorter than that of music. Generally, impressions and feelings to sound stimuli are believed to remain for a long time, however, the results with no-sound condition suggest that change in heartbeat interval occurred by sound stimuli disappear from 20 s to 36 s. Strength of the impressions and feelings would relate to the time cost.

In some results, although temporal changes in heartbeat intervals just after the transition were observed, higher *P*-values around the end of post section were also observed. This tendency was also shown in temporal development of heartbeat intervals. Homeostasis is an important function assisting the body in maintaining a constant internal environment [Rubinson & Lang, 2008], and this function keeps heartbeat interval in certain range. In the physiological evaluation processes, it should be noted that heartbeat interval has its limits to change.

Furthermore, the change in heartbeat was mainly caused of psychological change as discussed above, however, some previous studies indicated the possibility that tempo of the sound stimulus entrain listener's heartbeat. Previous studies have investigated the physiological effect of tempo of sound stimuli on heartbeat with simple tone [Bason & Celler, 1976] and musical piece [Kusunoki et al., 1972] and have observed the entrainment and synchronization of heartbeat by the sound stimulus. According to the results of these previous studies, the change of heartbeat interval might be partly affected physiologically from tempo of sound stimulus. On the other hand, there was no possibility that heartbeat interval in noise section was entrained by white noise because white noise does not have any tempo and cycle. The difference of physical property of sound stimuli used in the listening experiment might affect the change of heartbeat intervals in each section. To clarify the effect of the tempo of sound stimuli, further investigation comparing tempo of musical piece and heartbeat interval is needed.

5. Conclusions

In this study, as fundamental investigation of temporal development of heartbeat interval in listening to sound stimuli, we investigated the effects of relaxing musical piece and white

noise, and no-sound on heartbeat through three listening experiments. These sound stimuli induced the subjects' different relaxation feelings between the sound stimuli. Averages of heartbeat intervals were almost same level. As precise observation with statistical analysis, detail temporal development of heartbeat interval in transitions between two different sound stimuli were observed using 20-s sliding window. Some of the results, especially for noise, supports previous studies. However, time cost of change in heartbeat by listening to music were longer than the result in our previous study.

Some previous studies have investigated the relationship user's *KANSEI* and physiological response in listening sound and have tried to apply the relation between them for developing musical system; selection, arrangement and creation of musical piece. These approach aims to reflect user's *KANSEI* to the system and to absorb individual difference. The results of this study will contribute to these trials, especially if the system uses heartbeat interval as physiological index.

As future study, we will investigate the psycho-physiological effects of the sound stimuli inducing different relaxation feeling with spectral analysis [Akselrod et al., 1981]. With the analysis, we can separately evaluates autonomic nervous activities based on heartbeat intervals; sympathetic and parasympathetic nervous activity. The results with the analysis would show us more precise physiological change in the change in the sound stimuli, and it will contribute to more effective applications based on user's physiological information.

6. Acknowledgment

This work was supported partly by Grant from Computer Science Laboratory, Fukuoka Institute of Technology.

7. References

- Akselrod, S., Gordon, D., Ubel, F. A., Shanon, D. C., Barger, A. C., & Cohen, R. J. (1981). Power Spectrum Analysis of Heart Rate Fluctuation: A Quantitative Probe of Beat-to-Beat Cardiovascular Control, *Science*, Vol. 213, No. 10, pp. 220-222, doi:10.1126/science.6166045
- Aoto, T. & Ohkura, M. (2007). Study on Usage of Biological Signal to Evaluate Kansei of a System, *Proceedings of the 1st International Conference on Kansei Engineering and Emotion Research 2007*, Sapporo, L-9
- Bason, P. T. & Celler, B. G. (1972). Control of the Heart Rate by External Stimuli, *Nature*, Vol. 238, pp. 279-280
- Bigand, E., Filipic, S., & Lalitte P. (2005). The time course of emotional responses to music, *Annals of the New York Academy of Sciences*, pp. 429-437
- Chung, J. & Vercoe, G. S. (2006). The affective remixer: personalized music arranging, *Conference on Human Factors in Computing Systems*, pp. 393-398
- Dainow, E. (1977). Physical effects and motor responses to music, *Journal of Research in Music Education*, pp. 211-221
- Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D., & Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction, *International Journal of Psychophysiology*, Vol. 61, pp. 57-69, doi:10.1016/j.ijpsycho.2005.10.025
- Fukumoto, M. & Imai, J. (2008). Evolutionary Computation System for Musical Composition using Listener's Heartbeat Information, *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 3, No. 6, pp. 629-631, doi:10.1002/tee.20324

- Fukumoto, M., Hasegawa, H., Hazama, T., & Nagashima, T. (2009). Temporal Development of Heartbeat Intervals in Transition of Sound Stimuli Inducing Different Relaxation Feelings, *Proceedings of Biometrics and Kansei Engineering*, pp. 84-89, ISBN: 978-0-7695-3692-7, Cieszyn, Poland
- Gomez, P. & Danuser, B. (2004). Affective and physiological responses to environmental noises and music, *International Journal of Psychophysiology*, Vol. 53, No. 2, pp. 91-103, doi:10.1016/j.ijpsycho.2004.02.002
- Hazama T. & Fukumoto, M. (2009). Relationship between Evaluation Values in GA using Physiological Index and Subjective Evaluation, *Proceedings of the 2009 IEICE General Conference*, 2009, p.105, (in Japanese)
- Healey, J., Picard, R., Dabek, F. (1998). A New Affect-Perceiving Interface and Its Application to Personalized Music Selection, *Proceedings of the 1998 Workshop on Perceptual User Interfaces*
- Kim, S. & André, E. (2004). A Generate and Sense Approach to Automated Music Composition, *Proceedings of the 9th international conference on Intelligent user interfaces*, Funchal, Madeira, Portugal, pp. 268-270
- Kusunoki, Y., Fukumoto, M., & Nagashima, T. (2003). A Statistical Method of Detecting Synchronization for Cardio-Music Synchrogram, *IEICE Transactions, Fundamentals*, Vol. E86-A, No. 9, 2003, pp. 2241-2247
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. (1957). *The measurement of meaning*, University of Illinois Press, IL, USA
- Pappano, A. J. (2008). 'Section IV: The Cardiovascular System', in Berne and Levy Physiology (6th ed.), Koepfen, B. M. & Stanton, B. A. (Eds.), Mosby Elsevier, PA, USA, pp. 287-414
- Rubinson, K. & Lang, E. J. (2008) 'Section II: The Nervous System', in Berne and Levy Physiology (6th ed.), Koepfen, B. M. & Stanton, B. A. (Eds.), Mosby Elsevier, PA, USA, pp. 51-230
- Sugimoto, T., Legaspi, R., Ota, A., Moriyama, K., Kurihara S., & Numao, M. (2008). Modelling affective-based music compositional intelligence with the aid of ANS analyses, *Knowledge-Based Systems*, Vol. 21, Issue 3, pp. 200-208, doi:10.1016/j.knosys.2007.11.010
- Yamada, T., Yamazaki, I., Misaki, K., & Sawada, Y. (2000). A Basic Study on Psychophysiological Changes in Listening Music, *Japanese Bulletin of Arts Therapy*, vol. 31, no. 2, 2000, pp. 33-41, (in Japanese)
- Yoshida, Y., Yokoyama, K. & Ishii, N. (2006). Real-time Continuous Assessment Method for Mental and Physiological Condition using Heart Rate Variability, *IEEE Transactions on Electronics, Information and Systems*, vol. 126, no. 12, 2006, pp. 1441-1446 (in Japanese)

The Use of Saliva Protein Profiling as a Biometric Tool to Determine the Presence of Carcinoma among Women

Charles F. Streckfus and Cynthia Guajardo-Edwards
*University of Texas Dental Branch at Houston,
United States of America*

1. Introduction

1.1 Background information

Biometrics is the science and technology of measuring and analyzing biological data. It also refers to technologies that measures and analyzes human body characteristics for identification purposes. In the context of this book chapter, identification will refer to the recognition of those individuals in a disease state *i.e.*, carcinoma of the breast. Using “start-of-the-art” mass spectrometry protein analysis, the author will demonstrate the use of salivary protein profiles to recognize individuals at risk for carcinoma of the breast.

Proteomic analyses of varying body fluids are propelling the field of medical research forward at unprecedented rates due to its consistent ability to identify proteins that are at the femtomole level in concentration. These advancements have also benefited biometric research to the point where saliva is currently recognized as an excellent diagnostic medium for biometric authentication of human body characteristics. The saliva microbiome, for example, is reputed to be biometrically as accurate as a fingerprint. Collectively, these efforts are in the area of biological verification; however, biometric can be applied to identify the biological characteristics of a diseased individual.

1.2 Why Saliva as a diagnostic media?

1.2.1 Analytical advantages of Saliva

Saliva as a diagnostic fluid has significant biochemical and logistical advantages when compared to blood. Bio-chemically, saliva is a clear liquid with an average protein concentration of 1.5 to 2.0 mg/ml. As a consequence of this low protein concentration, it was once assumed that this was a major drawback for using saliva as a diagnostic fluid; however, current ultra sensitive analyte detection techniques have eliminated this barrier. Saliva specimen preparation is simple involving centrifugation prior to storage and the addition of a cocktail of protease inhibitors to reduce protein degradation for long-term storage.

Blood is a far more complex medium. A decision has to be made as to whether to use serum or plasma. Serum has a total protein concentration of approximately 60-80 mg/ml. Since serum possesses more proteins than saliva, assaying trace amounts of “factors” (*e.g.*, oncogenes, etc.), may result in a greater risk of non-specific interference and a greater chance

for hydrostatic (and other) interactions between the factors and the abundant serum proteins. Serum also possesses numerous carrier proteins, *e.g.*, albumin, which must either be removed or treated prior to being assayed for protein content. Additionally, it has been demonstrated that clotting removes many background proteins, which may be altered in the presence of disease. It has been demonstrated that enzymatic activity continues during this process, which may cleave proteins from many relevant pathways (Koomen et al., 2005). It would be ideal if all enzymatic activity in serum would cease at the time of collection; however, proteomic analyses of serum has shown that this is not the case. As a consequence, plasma is also being explored as a diagnostic fluid. The main consideration in using plasma is the selection of a proper anticoagulant (Koomen et al., 2005; Teisner et al., 1983). Heparin for example can be used as an anti-clotting agent; however, current research has found that heparin has a relatively short half life (3 to 4 hours) and can produce products of coagulation which are abundantly comparable to those assayed in serum. Based on these observations, it is recommended that blood specimens be collected with ethylenediamine tetraacetic acid (EDTA).

1.2.2 Collection advantages of Saliva

From a logistical perspective, the collection of saliva is safe (*e.g.*, no needle punctures), non-invasive and relatively simple, and may be collected repeatedly without discomfort to the patient [4]. Consequently it may be possible to develop a simplified method for “home-testing”, testing in a “health fair” setting or in dental clinics where individuals are available for periodic oral examinations. This diagnostic potential could reach many individuals who for personal, logistical or economical reasons lack access to preventive care.

Blood is a more complicated medium to collect. It requires highly trained personnel to collect it and if collected incorrectly, can lead to misinterpretations which can result in patient mismanagement (Ernest & Balance, 2006). Blood specimens need to be collected in a specific sequence and under-filling tubes with additives may possibly alter protein analyses. Additionally, if specimens are collected during hospital or clinical settings, there may be a lapse of time before being processed.

1.2.3 Saliva collection

The oral cavity receives secretions from three pairs of major salivary glands and numerous minor salivary glands that are located on the oral buccal mucosa, palate, and tongue each producing a unique type of secretion with varying protein constituents (Birkhed & Heintze, 1989). For example the parotid and Von Ebner glands (located on the tongue) produce serous secretions while the minor salivary glands produce mucinous secretions. The submandibular and sublingual glands, however, produce mixed secretions which are both serous and mucinous. As a consequence, composite or “whole” saliva is preferred as it enhances the chances of finding a biomarker due to the variety of sources from which it derives and because of its simplicity to collect.

There are basically two types of saliva to collect. One type is “resting” or unstimulated whole saliva and the other is stimulated whole saliva. There are several methods for collecting unstimulated whole saliva. These include the draining or drool method, spitting method, suction method and the swab method. These methods will yield 0.47, 0.47, 0.54 and 0.52 ml/minute of saliva respectively. Of the four methods, the most reliable is the suction method with a reliability coefficient of $r = 0.93$. It also revealed a within subject variance of

0.14. This is a very reliable method; however, a vacuum pump is required to collect the specimens (Birkhed & Heintze, 1989).

There are several drawbacks when using unstimulated saliva. The major problem is the small amount of saliva derived from collection. The 0.47-0.54 ml/minute is the range for healthy individuals (0.25–0.35 ml/minute normal range) under ideal conditions using those aforementioned collection methodologies. If the subject is taking medications that decrease flow rates (e.g., anti-hypertensive medications) the amount collected will be significantly reduced. Additionally, if the subject has autoimmune disorders (e.g., Sjögren's syndrome), has under gone head and neck radiation, or is very elderly, it will be difficult to obtain 0.5 ml over a five minute period. Unstimulated saliva flow rates are also influenced by circadian and circannual rhythms. Therefore, for consistency, individuals will need to be serially assessed at approximately the same time of day that the baseline specimen was collected. All other participants will need to be collected at approximately the same time in order to reduce inter-variability among the participating subjects. In conclusion, due to the small quantity of specimen obtained from these techniques and the large within subject variance, one can conclude that using unstimulated saliva is not the ideal medium for cancer biomarker discover.

The alternative to using unstimulated whole saliva is obviously to use stimulated whole saliva. Stimulated secretions produce about three times the volume of unstimulated secretions and are not subjected to the effects of circadian rhythm. Additionally, you will be able to collect sufficient quantities of saliva despite health status and medication usage. The flow rate range is 1 - 3 ml/minute for healthy individuals (Birkhed & Heintze, 1989; Gu et al., 2004).

There are two methods for collecting stimulated whole saliva. One method of collection is the gustatory method and the other is the reflexive or "masticatory" technique. The gustatory technique requires the use of an oral based secretory stimulant. Citric acid is the most widely used stimulant. Five drops of a 1-6% citric acid solution is applied to the dorsum of the tongue every 30 seconds. The saliva accumulates in the mouth and is expectorated intermittently for a period of five minutes. This technique produces copious amounts of saliva; however, the reliability is only $r = 0.76$ and has a within subject variance of 0.49.

The reflexive method is based on the reflex response occurring during the mastication of a bolus of food. Usually, a standardized bolus (1 gram) of paraffin or a gum base (Wrigley Co., Peoria, IL) is given to the test subject and they chew the substance at a regular rate. The subject expectorates intermittently during the collection period for duration of five minutes. This is an accurate technique as it has a reliability coefficient of $r = 0.95$ and a within subject variability of 0.11. The authors recommend this salivary collection method for biomarker discovery.

The procedure for collecting Stimulated Whole Salivary Gland Secretions is as follows: A standard piece of unflavored gum base (1.0 - 1.5 g.) is placed in the subject's mouth. The armamentarium used for this procedure is illustrated in Figure 1. The patient is asked to swallow any accumulated saliva and then instructed to chew the gum at a regular rate (using a metronome). The subject, upon sufficient accumulation of saliva in the oral cavity, expectorates periodically into a preweighed disposable plastic cup. This procedure is continued for a period of five minutes. The cup with the saliva specimen is reweighed and the flow rate determined gravimetrically. The volume and flow rate is then recorded along with a brief description of the specimen's physical appearance (Gu et al., 2004).

1.2.4. Long-term Saliva specimen banking

Roughly 2 - 5 ml of whole saliva will be obtained from the individual. In order to minimize the degradation of the proteins, protease inhibitor cocktail (Sigma, 1 mg/ml whole saliva) and 1 mM of sodium orthovanadate are added immediately after sample collection (Shevchenko et al., 2002). All samples are kept on ice during the process. The specimen is next divided into 0.5 ml aliquots, placed into bar code labeled cryotubes, and frozen (-80°C). To assess specimen degradation, ten healthy subjects were serially sampled for saliva over a five-year period. We used *c-erbB-2* to test for specimen stability as this is a large 185-kDa protein, which would be susceptible to degradation by proteases and other biochemical activity. The results are shown in Figure 1 and illustrate protein stability when frozen at -80°C. These results are consistent with Wu et al, 1993 where they assayed serially sampled salivary specimens which were collected over a ten year period for total protein, lactoferrin (77 kDa) and histidine rich proteins concentrations. In their study, they found no concentration differences due to specimen aging.



Fig. 1. Armamentarium for the collection and storage of stimulated whole saliva

1.3 Studies using Saliva protein profiling for disease state detection

The majority of the literature concerning human saliva biometrics is associated with the oral cavity and its associated maladies. An example of this statement is demonstrated in a manuscript assessing salivary proteins associated with burning mouth syndrome (Moura et al., 2007). The principle objective the present study was to analyze the characteristics of salivary production and its composition in individuals with burning mouth syndrome. The investigators compared salivary flow rates, potassium, iron, chloride, thiocyanate, magnesium, calcium, phosphorus, glucose, total protein and urea concentrations, as well as the expression profile of salivary proteins by SDS-PAGE among healthy individuals and those diagnosed with burning mouth syndrome. The results of the study showed that mean salivary flow rates among control patients were lower than that of burning mouth syndrome patients. Chloride, phosphorus and potassium levels were elevated in patients with burning

mouth syndrome ($p = 0.041$, 0.001 and 0.034 , respectively). Total salivary protein concentrations were reduced in individuals with burning mouth syndrome ($p = 0.223$). Additionally, the analysis of the expression of salivary proteins by Coomassie blue SDS-PAGE revealed a lower expression of low molecular weight proteins in individuals with burning mouth syndrome compared to healthy controls. The results suggested that the identification and characterization of low molecular weight salivary proteins in burning mouth syndrome may be important in understanding BMS pathogenesis, thus contributing to its diagnosis and treatment.

Another study using salivary protein profiles investigated the modification of the salivary proteome occurring in type 1 diabetes and to highlight potential biomarkers of the disorder. High-resolution two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry was combined to perform a large scale analysis of the salivary specimens. The proteomic comparison of saliva samples from healthy subjects and poorly controlled type-1 diabetes patients revealed a modulation of 23 proteins. Fourteen isoforms of α -amylase, one prolactin inducible protein, three isoforms of salivary acidic protein-1, and three isoforms of salivary cystatins SA-1 were detected as under expressed proteins, whereas two isoforms of serotransferrin were over expressed secondary to type-1 diabetes. The proteins under expressed were all known to be implicated in the oral anti-inflammatory process, suggesting that the pathology induced a decrease of non-immunological defense of oral cavity. As only particular isoforms of proteins were modulated, type-1 diabetes seemed to differentially affect posttranslational modification of the proteins (Hirtz et al., 2006).

An additional study (Delaleu et al., 2008) investigated the involvement of 87 proteins measured in serum and 75 proteins analyzed in saliva in spontaneous experimental Sjögren's syndrome. In addition, they intended to compute a model of the immunological situation representing the overt disease stage of Sjögren's syndrome. In this animal study, they used non-diabetic, non-obese diabetic mice for salivary gland dysfunction. The mice aged 21 weeks and were evaluated for salivary gland function, salivary gland inflammation and extra-glandular disease manifestations. The analytes, comprising chemokines, cytokines, growth factors, autoantibodies and other biomarkers, were quantified using multi-analyte profile technology and fluorescence-activated cell sorting. Age-matched and sex-matched Balb/c mice served as a reference. The investigators found non-diabetic, non-obese diabetic mice tended to exhibit impaired salivary flow, glandular inflammation and increased secretory SSB (anti-La) levels. Thirty-eight biomarkers in serum and 34 in saliva obtained from non-diabetic, non-obese diabetic mice were significantly different from those in Balb/c mice. Eighteen biomarkers in serum and three chemokines measured in saliva could predict strain membership with 80% to 100% accuracy. Factor analyses identified principal components mostly correlating with one clinical aspect of Sjögren's syndrome and having distinct associations with components extracted from other families of proteins. They concluded that the autoimmune manifestations of Sjögren's syndrome are greatly independent and associated with various immunological processes; however, CD40, CD40 ligand, IL-18, granulocyte chemotactic protein-2 and anti-muscarinic M3 receptor IgG3 may connect the different aspects of Sjögren's syndrome. Processes related to the adaptive immune system appear to promote Sjögren's syndrome with a strong involvement of T-helper-2 related proteins in hyposalivation. This approach further established saliva as an attractive biofluid for biomarker analyses in Sjögren's syndrome and provides a basis for the comparison and selection of potential drug targets and diagnostic markers (Delaleu et al., 2008).

2. Current research in Salivary protein profiling for cancer detection

2.1 Methods

2.1.1 Study design

The purpose of this study was to determine if individuals could be protein profiled and potentially classified as having cancer. The investigator also wanted to ascertain if there were alterations of the protein profiles due to the primary tissue site and the varying degree of tumor staging. In order to achieve this objective, the investigators collected saliva from women that were healthy and from those diagnosed with carcinoma breast in the following stages: Stage 0, Stage I, Stage IIa and Stage IIb. All the tumors were adenocarcinomas. Additionally, specimens were collected from women diagnosed with varying gynecological carcinomas. These included women diagnosed with moderate cervical dysplasia, severe cervical dysplasia and cervical carcinoma in situ. These tumors were all squamous cell carcinomas. Women diagnosed with ovarian and endometrial carcinomas were also included in the study. These malignancies were identified as adenocarcinomas. The final group consisted of women diagnosed with head and neck squamous cell carcinomas ten women with varying stages of development. Due to the difficulty in obtaining early stage tumors for ovarian, endometrial and head/neck carcinomas, a composite of varying staged patients formed these saliva pools.

This study was performed under the UTHSC IRB approved protocol number HSC-DB-05-0394. All procedures were in accordance with the ethical standards of the UTHSC IRB and with the Helsinki Declaration of 1975, as revised in 1983. The specimens were banked at the University of Texas Dental Branch Saliva repository, which stores the specimens at -80°C . Ten saliva specimens were pooled for each type of carcinoma. The saliva samples were pooled by combining equal volumes of cleared stimulated whole saliva from a set of archived healthy and cancer subjects. The subjects were matched for age and race and were non-tobacco users. Previous studies by the investigator have demonstrated that properly prepared specimens can remain in storage for a long period of time.

2.1.2 Saliva collection and sample preparation

Stimulated whole salivary gland secretion is based on the reflex response occurring during the mastication of a bolus of food. Usually, a standardized bolus (1 gram) of paraffin or a gum base (generously provided by the Wrigley Co., Peoria, IL) is given to the subject to chew at a regular rate. The individual, upon sufficient accumulation of saliva in the oral cavity, expectorates periodically into a preweighed disposable plastic cup. This procedure is continued for a period of five minutes. The volume and flow rate is then recorded along with a brief description of the specimen's physical appearance (Navazesh & Christensen, 1982). The cup with the saliva specimen is reweighed and the flow rate determined gravimetrically. The authors recommend this salivary collection method with the following modifications for consistent protein analyses. A protease inhibitor from Sigma Co (St. Louis, MI, USA) is added along with enough orthovanadate from a 100mM stock solution to bring its concentration to 1mM. The treated samples were centrifuged for 10 minutes at top speed in a table top centrifuge. The supernatant was divided into 1 ml aliquots and frozen at -80°C .

2.1.3 LC-MS/MS mass spectroscopy with isotopic labeling

Recent advances in mass spectrometry, liquid chromatography, analytical software and bioinformatics have enabled the researchers to analyze complex peptide mixtures with the

ability to detect proteins differing in abundance by over 8 orders of magnitude (Wilmarth et al., 2004). One current method is isotopic labeling coupled with liquid chromatography tandem mass spectrometry (IL-LC-MS/MS) to characterize the salivary proteome (Gu et al., 2004). The main approach for discovery is a mass spectroscopy based method that uses isotope coding of complex protein mixtures such as tissue extracts, blood, urine or saliva to identify differentially expressed proteins (18). The approach readily identifies changes in the level of expression, thus permitting the analysis of putative regulatory pathways providing information regarding the pathological disturbances in addition to potential biomarkers of disease. The analysis was performed on a tandem QqTOF QStar XL mass spectrometer (Applied Biosystems, Foster City, CA, USA) equipped with a LC Packings (Sunnyvale, CA, USA) HPLC for capillary chromatography. The HPLC is coupled to the mass spectrometer by a nanospray ESI head (Protana, Odense, Denmark) for maximal sensitivity (Shevchenko et al., 2002). The advantage of tandem mass spectrometry combined with LC is enhanced sensitivity and the peptide separations afforded by chromatography. Thus even in complex protein mixtures MS/MS data can be used to sequence and identify peptides by sequence analysis with a high degree of confidence (Birkhed et al., 1989; Gu et al., 2004; Shevchenko et al., 2002; Wilmarth et al., 2004).

Isotopic labeling of protein mixtures has proven to be a useful technique for the analysis of relative expression levels of proteins in complex protein mixtures such as plasma, saliva urine or cell extracts. There are numerous methods that are based on isotopically labeled protein modifying reagents to label or tag proteins to determine relative or absolute concentrations in complex mixtures. The higher resolution offered by the tandem Qq-TOF mass spectrometer is ideally suited to isotopically labeled applications (Gu et al., 2004; Koomen et al 2004; Ward et al., 1990).

Applied Biosystems recently introduced iTRAQ reagents (Gu et al., 2004; Koomen et al 2004; Ward et al., 1990), which are amino reactive compounds that are used to label peptides in a total protein digest of a fluid such as saliva. The real advantage is that the tag remains intact through TOF-MS analysis; however, it is revealed during collision induced dissociation by MSMS analysis. Thus in the MSMS spectrum for each peptide there is a fingerprint indicating the amount of that peptide from each of the different protein pools. Since virtually all of the peptides in a mixture are labeled by the reaction, numerous proteins in complex mixtures are identified and can be compared for their relative concentrations in each mixture. Thus even in complex mixtures there is a high degree of confidence in the identification.

2.1.4 Salivary protein analyses with iTRAQ

Briefly, the saliva samples were thawed and immediately centrifuged to remove insoluble materials. The supernatant was assayed for protein using the Bio-Rad protein assay (Hercules, CA, USA) and an aliquot containing 100 µg of each specimen was precipitated with 6 volumes of -20°C acetone. The precipitate was resuspended and treated according to the manufacturers instructions. Protein digestion and reaction with iTRAQ labels was carried out as previously described and according to the manufacturer's instructions (Applied Biosystems, Foster City, CA). Briefly, the acetone precipitable protein was centrifuged in a table top centrifuge at 15,000 × g for 20 minutes. The acetone supernatant was removed and the pellet resuspended in 20 µl dissolution buffer. The soluble fraction was denatured and disulfides reduced by incubation in the presence of 0.1% SDS and 5 mM TCEP (tris-(2-carboxyethyl)phosphine)) at 60°C for one hour. Cysteine residues were

blocked by incubation at room temperature for 10 minutes with MMTS (methyl methanethiosulfonate). Trypsin was added to the mixture to a protein:trypsin ratio of 10:1. The mixture was incubated overnight at 37°C.

The protein digests were labeled by mixing with the appropriate iTRAQ reagent and incubating at room temperature for one hour. On completion of the labeling reaction, the four separate iTRAQ reaction mixtures were combined. Since there are a number of components that can interfere with the LC-MS/MS analysis, the labeled peptides are partially purified by a combination of strong cation exchange followed by reverse phase chromatography on preparative columns. The combined peptide mixture is diluted 10 fold with loading buffer (10 mM KH₂PO₄ in 25% acetonitrile at pH 3.0) and applied by syringe to an ICAT Cartridge-Cation Exchange column (Applied Biosystems, Foster City, CA) column that has been equilibrated with the same buffer. The column is washed with 1 ml loading buffer to remove contaminants.

To improve the resolution of peptides during LCMSMS analysis, the peptide mixture is partially purified by elution from the cation exchange column in 3 fractions. Stepwise elution from the column is achieved with sequential 0.5 ml aliquots of 10 mM KH₂PO₄ at pH 3.0 in 25% acetonitrile containing 116 mM, 233 mM and 350 mM KCl respectively. The fractions are evaporated by Speed Vacuum to about 30% of their volume to remove the acetonitrile and then slowly applied to an Opti-Lynx Trap C18 100 ul reverse phase column (Alltech, Deerfield, IL) with a syringe. The column was washed with 1 ml of 2% acetonitrile in 0.1% formic acid and eluted in one fraction with 0.3 ml of 30% acetonitrile in 0.1% formic acid. The fractions were dried by lyophilization and resuspended in 10 ul 0.1% formic acid in 20% acetonitrile. Each of the three fractions was analyzed by reverse phase LCMSMS. The analytical strategy is illustrated in Figure 2.

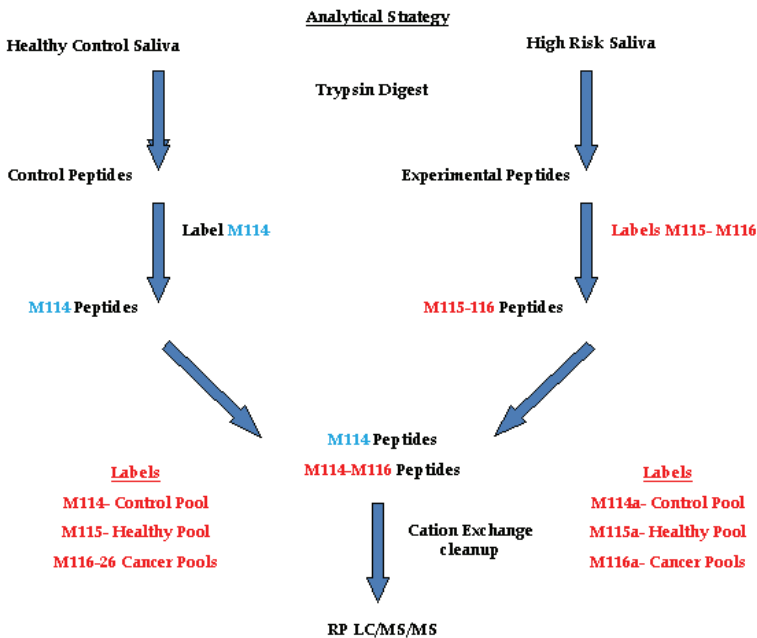


Fig. 2. Analytical strategy for quantifying peptides using iTRAQ tagging

2.1.5 Reverse phase LCMSMS

The desalted and concentrated peptide mixtures were quantified and identified by nano-LC-MS/MS on an API QSTAR XL mass spectrometer (AB Sciex Instruments) operating in positive ion mode. The chromatographic system consists of an UltiMate nano-HPLC and FAMOS auto-sampler (Dionex LC Packings). Peptides were loaded on a 75cm x 10 cm, 3cm fused silica C18 capillary column, followed by mobile phase elution: buffer (A) 0.1% formic acid in 2% acetonitrile/98% Milli-Q water and buffer (B): 0.1% formic acid in 98% acetonitrile/2% Milli-Q water. The peptides were eluted from 2% buffer B to 30% buffer B over 180 minutes at a flow rate 220 nL/min. The LC eluent was directed to a NanoES source for ESI/MS/MS analysis. Using information-dependent acquisition, peptides were selected for collision induced dissociation (CID) by alternating between an MS (1 sec) survey scan and MS/MS (3 sec) scans. The mass spectrometer automatically chooses the top two ions for fragmentation with a 60 second dynamic exclusion time. The IDA collision energy parameters were optimized based upon the charge state and mass value of the precursor ions. Each saliva sample set there are three separate LCMSMS analyses.

The accumulated MSMS spectra are analyzed by ProQuant and ProGroup software packages (Applied Biosystems) using the SwissProt database for protein identification. The ProQuant analysis was carried out with a 75% confidence cutoff with a mass deviation of 0.15 Da for the precursor and 0.1 Da for the fragment ions. The ProGroup reports were generated with a 95% confidence level for protein identification.

2.1.6 Bioinformatics

The Swiss-Prot database was employed for protein identification while the PathwayStudio® bioinformatics software package was used to determine Venn diagrams were also constructed using the NIH software program (<http://ncrr.pnl.gov>). Graphic comparisons with log conversions and error bars for protein expression were produced using the ProQuant® software.

2.1.7 Western blot analysis for marker validation

2.1.7.1 Preparation of samples

We selected the protein profilin-1 for validating the presence of these proteins in saliva. The profilin-1 antibody was a rabbit polyclonal from the Abcam Co. #Ab10608 diluted 1:1000. The saliva samples from a healthy individuals, benign tumor subjects and individuals diagnosed with Stage IIa her2/neu receptor positive breast cancer subjects and Stage IIa her2/neu receptor negative breast cancer subjects were pooled by combining equal volumes of cleared stimulated whole saliva from a set of archived specimens.

The pooled saliva was mixed with loading buffer (Laemmli buffer containing BME) in 1:1 ratio. The sample was then incubated at 95°C for 5 minutes and was then loaded onto the 4-15% Tris-HCl polyacrylamide gel. Four-fifteen percent Tris-HCl polyacrylamide gels were loaded with molecular weight markers, controls, and the pooled saliva samples. Electrophoresis was run at 200 Volts, 30 minutes in 1X TGS buffer. The gels were equilibrated and extra thick blot paper and PVDF membranes were soaked in 1X TGS buffer for 15 minutes prior to running Western Transfer. Semi-dry transfer apparatus was used. Transfer conditions were 0.52mA constant, 17 volts, 19 minutes. Polyvinylidene fluoride (PVDF) membranes were air dried for minimum of 1 hour. Dry PVDF membranes were activated in methanol for about 10 seconds then transferred to soak in 1X PBS-T for 3 washes

of 5 minutes each. The membranes were then incubated for 1 hour in 5% NFDM in PBS-T. Afterwards the membranes were washed 3 times for 5 minutes in PBS-T. The membranes were incubated overnight with a primary antibody in PBS-T. The membranes were washed 3 times for 5 minutes in PBS-T and were incubated for 4 hours with secondary antibody (HRP conjugate) in PBS-T. Again, the membranes were washed 3 times for 5 minutes in PBS-T. Finally, the membranes were treated with ECL plus detection reagents and photographed with exposure of 800 seconds.

3. Experimental results in salivary protein profiling for cancer detection

3.1 Mass spectrometry analysis

Tables 1-5 summarize the results of the proteomic analysis and illustrates protein comparisons between breast (Stage 0 - IIb), cervical (moderate, severe dysplasia and in situ)

Proteins		Breast				Ovarian	Endo.	Cervical			H & N
		Staging									
Gene ID	Accession	Stage 0	Stage I	Stage IIa	Stage IIb	Variable	Variable	Mod.	Severe	Stage 0	Variable
A1AT	P01009										4.32
ANXA3	P12429										0.74
CO3	P01024										2.16
HPTR	P00739										3.57
K1C16	P08779									5.37	
COBA1	P12107							0.57			
LUZP1	Q86V48							0.62			
ZN248	Q8NDW4							0.84			
CYTC	P01034							0.77			
KAC	P01834					1.42					
K2C6C	P48666	3.40									
K1CJ	P13645		0.47		0.26						
SCOT2	Q9BYC2			0.83	0.50						
VEGP	P31025	1.36		0.47							
PROF1	P07737	0.74									
NGAL	P80188	0.90									
NUCB2	P80303		1.28								
HEMO	P02790	0.74									
CYTD	P28325		0.82								
CRIS3	P54108	0.65									
ACBP	P07108	1.42									
KLK	P06870	0.80	1.46	0.86							
PIP	P12273	0.88	0.80								
PERL	P22079		1.31	0.88							
PPIB	P23284		1.32								

Table 1. The table shows which proteins are unique to each type and stage of carcinoma

Protein Status	Breast				Ovarian	Endo.	Cervical			H & N
	Staging									
	Stage 0	Stage I	Stage IIa	Stage IIb	Variable	Variable	Mod.	Severe	Stage 0	Variable
Unique Proteins										
Up Regulated	3	4	0	0	1	0	0	0	1	2
Down Regulated	6	3	4	2	0	0	4	0	0	2
Total Proteins	9	7	4	2	1	0	4	0	1	4
Common Proteins										
Up Regulated	16	7	2	7	17	21	20	13	13	15
Down Regulated	8	11	11	7	5	5	7	4	8	7
Total Proteins	24	18	13	14	22	26	27	17	21	22
Total Number Proteins										
Up Regulated	19	11	2	7	18	21	20	13	14	17
Down Regulated	14	14	15	9	5	5	11	4	8	9
Grand Total	33	25	17	16	23	26	31	17	22	26

Table 2. The table demonstrates the unique, common and total protein counts among the varying carcinomas and their stages

ovarian, endometrial and head and neck cancer subjects. The values exhibited in table 1 and table 2 represent the ratio of disease state to healthy state. Simply stated the ratio for each protein is a result of the log sum integrated area under each peak formed when plotting the intensity versus mass-to-charge ratios for each peptide (Boehm et al., 2007). The value is divided by the corresponding value yielded for that particular protein in the healthy cohort. Significance between ratios is determined by t-test analysis. Those with values greater than 1.000 were considered up-regulated proteins while those with values less than 1.000 were considered down-regulated proteins.

In total, 166 proteins were identified at a confidence level at >95. Of these there were 76 proteins that were determined to be expressed significantly different ($p < 0.05$) in the saliva from cancer subjects as compared to the healthy individuals. Table 1 lists the 25 proteins that are unique to each type of carcinoma with their corresponding ratios. Unique being defined as not associated with the other types of carcinoma. Eleven of the unique proteins were up-regulated and fourteen were down-regulated. There were no unique proteins associated with endometrial carcinoma or severe dysplasia of the cervix. Carcinoma of the breast and its associated stages accounted for nearly 58% of the total unique proteins.

Table 3 indicates the function of the associated proteins. Generally, the proteins appear to be diverse with the exception of the cytoskeletal associated proteins which comprised 20% of the protein panel. Additionally, over 50% of the proteins in table 3 are referenced in the literature for the presence of carcinoma in blood and cell supernatants from cancer cell lines (Polanski & Anderson, 2006).

The following table represents the up and down-regulated proteins that over-lapped the various types of carcinomas. Numerous proteins, 21 percent, were common to both squamous cell carcinoma and the adenocarcinoma cancer types. This was particularly noticed among the S100 proteins. There were 13 proteins that were exclusive to the gynecological carcinomas despite the fact that some were squamous cell carcinoma or adenocarcinoma cancer types. These proteins comprised 25% of the protein panel. Table 5 illustrates the function of the proteins listed in table 4.

Gene ID	Accession	Protein Name	Reported Function
A1AT	P01009	Alpha 1 protease inhibitor	Protease inhibitor
ACBP	P07108	AcylCoA binding protein	Transport protein
ANXA3	P12429	Annexin A3	Inhibits phospholipase activity
CO3	P01024	Complement 3 precursor	Activates complement system
COBA1	P12107	Collagen alpha-1 chain	Fibrillogenesis
CRISP3	P54108	Cysteine rich secretory protein	Immune response
CYTC	P01034	Cystatin C	Inhibitor of cysteine proteases
CYTD	P28325	Cystatin D precursor	Protein degradation & inhibitor
HEMO	P02790	Hemopexin precursor	Heme transporter protein
HPTR	P00739	Haptoglobin - 1 related protein	proteolysis
K1C16	P08779	Cytoskeleton 16	Cytoskeleton associated
K1CJ	P13645	Cytoskeleton 10	Cytoskeleton associated
K2C6C	P48666	Cytoskeleton 6C	Cytoskeleton associated
KAC	P01834	Immunoglobulin kappa chain C region	Immune response
KLK	P06870	Kallikrein-1	Cleaves kininogen
LUZP1	Q86V48	Leucine zipper protein-1	Nucleus functioning protein
NGAL	P80188	Neutrophil Gelatinase associated lipocalin	Iron trafficking protein
NUCB2	P80303	Nucleobindin-2 precursor	Calcium binding protein
PERL	P22079	Lactoperoxidase	Transport, antimicrobial
PIP	P12273	Prolactin inducible protein precursor	Secretory actin binding protein
PPIB	P23284	Peptidyl-prolyl cis trans isomerase B	Accelerates protein folding
PROF1	P07737	Profilin-1	Cytoskeleton associated
SCOT2	Q9BYC2	Succinyl CoA	Ketone body catabolism
VEGP	P31025	Lipocalin - 1	Ligand binding protein
ZN248	Q8NDW4	Zinc finger protein 248	Transcriptional regulation

Table 3. The table reveals the protein functions associated with the varying carcinomas and their stages

Proteins		Breast				Ovarian	Endo.	Cervical		H & N	
		Stage 0	Stage I	Stage IIa	Stage IIb	Variable	Variable	Mod.	Severe	Stage 0	Variable
Gene ID	Accession										
1433S	O70456					1.50	2.14	2.09	2.12		
ADA32	Q8TC27					0.52	0.69				
ALBU	P02768	0.86		0.91	1.46	1.15	1.23			1.94	2.50
ALK1	P03973					2.94		1.33			
AMYS	P04745		0.84	1.11	0.75	0.84	0.47	0.58	0.55	0.78	0.65
ANXA1	P04083	1.54	1.43				3.79	2.04	2.06		0.63
BPIL1	Q8N4F0	1.15				1.94	0.86				1.78
CAH6	P23280	1.46	0.85				1.57				
CYTA	P01040	1.63						1.16			
CYTN	P01037	0.77						0.53	1.16	0.67	0.73
CYTS	P01036	0.80			0.64						0.59
CYTT	P09228		1.14		0.59			0.49		0.62	0.50
DEF3	P59666					3.86	2.61	1.17			

Table 4. Represents the up and down-regulated proteins that over-lapped the various types of carcinomas.

Proteins		Breast				Ovarian	Endo.	Cervical		H & N	
		Staging									
Gene ID	Accession	Stage 0	Stage I	Stage IIa	Stage IIb	Variable	Variable	Mod.	Severe	Stage 0	Variable
DMBT1	Q9UGM3					1.83	1.40	1.22			
ENOA	P06733		0.74	0.81			1.71			1.43	1.55
FABPE	Q01469	1.56							0.55		
FGRL1	Q8N441					0.21	0.44	0.64			
HPT	P00738	0.65			2.22	1.94				2.64	
IGHA1	P01876					1.23	1.55	1.55	1.23		1.34
IGHA2	P01877						1.81	1.65			2.18
IGHG1	P01857	1.26	0.83					1.41		1.48	1.50
IGHG2	P01859		0.82			0.80				1.58	1.89
IGJ	P01591		1.23				1.41			0.77	1.30
K1C13	P13646						3.50	6.21	5.91		0.73
K1C14	P02533					5.34	6.55				
K1CM	P13646	3.92				2.95					
K1CP	P08779			0.13	0.13	2.61					
K22O	Q01546					7.47	5.40				
K2C1	P04264	0.46	0.63	0.33	0.32			2.07	2.22		
K2C4	P19013	4.42				4.47	4.17	5.46	5.91	1.11	0.54
K2C5	P13647					3.08	5.39	2.65	3.49	0.75	
K2C6A	P02538					4.34	5.78				
K2C6E	P48668			0.10	0.09						
LAC	P01842			0.84			1.43	1.51			1.62
LCN1	P31025						0.72	1.33			
LV3B	P80748					1.33		1.26			
MUC	P01871	1.22	0.72								1.45
MUC5B	Q9HC84	1.25	0.71	0.54	1.26	1.39		2.04	1.75		2.15
PERM	P05164		0.72				1.88				
PIGR	P01833	0.82								0.88	1.29
PRDX1	Q06830						1.67	1.22			
S10A7	P31151	2.05				0.67	1.26		0.74	0.76	
S10A8	P05109	1.46	1.36	0.76	1.33		1.49	3.46	3.35	1.42	1.73
S10A9	P06702		1.53	0.67	1.15			3.86	3.10	1.38	1.73
SPLC2	Q96DR5	1.22	1.18	0.87	0.68			0.74	1.22	2.11	
SPRR3	Q9UBC9	1.10	0.87		1.42					1.92	
THIO	P10599	1.51								1.43	
TRFE	P02787	0.80	0.80	0.83						2.00	2.36
TRFL	P02788	1.36	1.19							0.64	
TRY	P07477			1.46	1.66			0.84	0.88		
ZA2G	P25311	0.90						0.91	1.13	1.29	

Table 4. Represents the up and down-regulated proteins that over-lapped the various types of carcinomas.(continuation)

Gene ID	Accession	Protein Name	Reported Function
1433S	O70456	1433 sigma	Signaling
ADA32	Q8TC27	Disintegrin	Role in sperm development
ALBU	P02768	Albumin	Protein transport
ALK1	P03973	Antileukoproteinase	Acid-stable proteinase inhibitor
AMYS	P04745	Amylase	Enzyme
ANXA1	P04083	Annexin A1	Involved in exocytosis
BPIL1	Q8N4F0	Bactericidal/permeability-increasing protein-like 1	Lipid binding
CAH6	P23280	Carbonic anhydrase 6	Hydration of carbon dioxide.
CYTA	P01040	Cystatin-A	Intracellular thiol proteinase inhibitor.
CYTN	P01037	Cystatin-SN	Cysteine proteinase inhibitors
CYTS	P01036	Cystatin-S	Inhibits papain and ficin
CYTT	P09228	Cystatin-SA	Thiol protease inhibitor
DEF3	P59666	Neutrophil defensin 3	Antimicrobial activity
DMBT1	Q9UGM3	Deleted in malignant brain tumors 1	Candidate tumor suppressor gene
ENOA	P06733	Alpha-enolase	Multifunctional enzyme
FABPE	Q01469	Fatty acid-binding protein, epidermal	Keratinocyte differentiation
FGRL1	Q8N441	Fibroblast growth factor receptor 1	Negative effect on cell proliferation
HPT	P00738	Haptoglobin	Combines with free hemoglobin
IGHA1	P01876	Ig alpha-1 chain C region	Immune Response
IGHA2	P01877	Ig alpha-2 chain C region	Immune Response
IGHG1	P01857	Ig gamma-1 chain C region	Immune Response
IGHG2	P01859	Ig gamma-2 chain C region	Immune Response
IGJ	P01591	Immunoglobulin J chain	Immune Response
K1C13	P13646	Cytokeratin 13	Cytoskeleton
K1C14	P02533	Cytokeratin 14	Cytoskeleton
K1CM	P13646	Keratin, type I cytoskeletal 13	Cytoskeleton
K1CP	P08779	Keratin, type I cytoskeletal 16	Cytoskeleton
K22O	Q01546	Keratin, type II cytoskeletal 2 oral	Contributes to terminal cornification
K2C1	P04264	Keratin, type II cytoskeletal 1	Regulate the activity of kinases
K2C4	P19013	Keratin, type II cytoskeletal 4	Cytoskeleton
K2C5	P13647	Keratin, type II cytoskeletal 5	Protein binding
K2C6A	P02538	Keratin, type II cytoskeletal 6A	Protein binding
K2C6C	P48666	Keratin, type II cytoskeletal 6C	Structural molecule activity
K2C6E	P48668	Keratin, type II cytoskeletal 6C	Cytoskeleton organization
LAC	P01842	Lactoperoxidase	Antimicrobial
LCN1	P31025	Lipocalin-1 precursor	Plays a role in taste reception.
LV3B	P80748	Ig lambda chain V-III region LOI	Activates complement pathway
MUC	P01871	Ig Mu Chain	Immune Response
MUC5B	Q9HC84	Mucin-5B	Contribute to the lubricating
PERM	P05164	Myeloperoxidase	Microbiocidal activity
PIGR	P01833	Polymeric immunoglobulin receptor	Binds IgA and IgM at cell surface
PRDX1	Q06830	Peroxiredoxin-1	Involved in redox regulation

Table 5. The table represents the function of the common proteins associated with the various types of carcinomas.

Gene ID	Accession	Protein Name	Reported Function
S10A7	P31151	S100 A7	Interacts with RANBP9
S10A8	P05109	S100 A8	Calcium-binding protein
S10A9	P06702	S100 A9	Calcium-binding protein
SPLC2	Q96DR5	Epithelial carcinoma associated protein-2	Lipid binding protein
SPRR3	Q9UBC9	Small proline-rich protein 3	Protein of keratinocytes.
THIO	P10599	Thioredoxin	Redox activity
TRFE	P02787	Serotransferrin	Iron binding transport proteins
TRFL	P02788	Lactotransferrin precursor	Iron binding transport proteins
TRY	P07477	Trypsin-1	Activity against synthetic substrates
ZA2G	P25311	Zinc alpha-2 glycoprotein	Signaling

Table 5. The table represents the function of the common proteins associated with the various types of carcinomas.(continuation)

3.2 Western blot analyses

The results of the western blot suggest the presence of profilin-1 in saliva. Figure 3 illustrates the presence of profilin-1 protein in both the human submandibular gland cell lysates and in whole saliva.

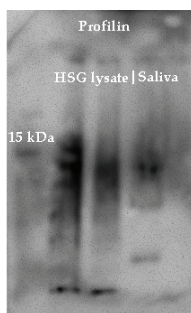


Fig. 3. The western blot indicates the presence of the 15 kDa profilin-1 protein in human submandibular gland cell lysates and stimulated whole saliva

The western blot analyses also revealed the presence of profilin-1 In SKBR3 cell lysates.

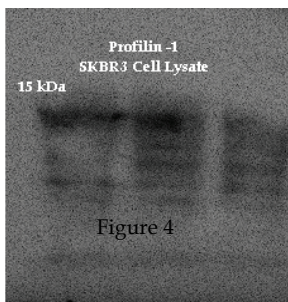


Fig. 4. Also revealed the presence of profilin-1 In SKBR3 HER2/neu receptor positive breast cancer cell lysates.

Figure 5 illustrates the presence of profilin-1 in the saliva sampled from healthy, benign and malignant tumor patients. Profilin is a down-regulated protein in the presence of malignancy and it is visualized by the lighter bands associated with malignancy. It is also worth noting that the Her2/neu receptor negative band is darker than the Her2/neu receptor positive counterpart suggesting further down-regulation of the profilin-1 protein.

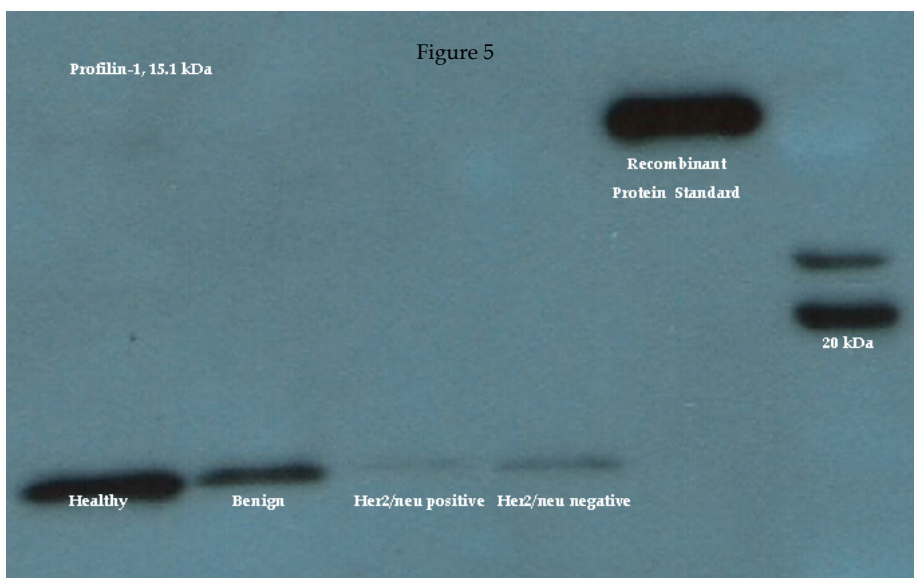


Fig. 5. This figure illustrates the presence and modulation of salivary profilin-1 protein secondary to HER2/neu receptor status.

3.3 Conclusions

It is interesting to note that the salivary protein profile mimics the findings of protein alterations secondary to cancer in serum, tissues and cell lysates that are numerous cited in the literature (Polanski et al, 2006). For example, there are salivary protein alterations in the cytoskeleton phenotype, which are strongly implicated in tumor growth and cancer metastases. Additionally, there are alterations in the metabolic, growth and signaling pathways all of which provide support for the concept that salivary protein profiles are altered secondary to the presence of cancer.

The investigator has spent 15 years investigating the phenomena and postulates, that in the presence of disease, *i.d.*, carcinoma, that there is an over abundance of protein resulting from the rapid growth of the malignancy which in turn, produces a humoral response in the salivary glands. This response results in altered salivary protein concentrations. Another possible explanation is active transport of the proteins of interest. It is plausible that these proteins are secreted into saliva as consequence of localized regulatory function in the oral cavity via signal transduction similar to the proposed explanation of HER-2/*neu* protein in nipple aspirates. These "loop" mechanisms, in health, appear to be in equilibrium both intercellularly and extracellularly with each pathway fulfilling the resultant phenotypic process of growth, proliferation, and differentiation.

The results of this preliminary study suggest that there are two panels of biomarkers that need to be assessed. There are those biomarkers that are unique to each tumor site *i.e.*, breast, cervix, etc. and those that are common to various types of malignancies. The finding may prove to be useful. It may be possible to develop a “global” profile for the overall detection of cancer while using the unique protein identifiers to determine the site of the tumor. Theoretically, by using the biomarkers S100A7, S100A8 and S100A9 you may predict the presence of a malignancy and by using the biomarkers listed in table 1 you, potentially could determine the site of the tumor. This is merely speculation at this time; however, advances in mass spectrometry and microarray technology may make the concept cost effective and logistically feasible.

These results are very preliminary and investigated only the abundant proteins found in saliva. Further research is required to determine the low abundant protein profiles associated with the various carcinomas. Additionally, the peptidome and the fragmetome need to be evaluated to further enhance the profile. Other carcinomas such as lung, colon, pancreas, bladder, skin, brain and kidney at varying stages of development also need to be studied and their resultant profiles added to the catalogue of cancer associated protein profiles. One major question that is essential to answer is, “Are we identifying the primary site of the tumor’s origin?” This information is essential for rendering “tailored” treatment regimens.

The results of this study suggest that it is feasible to employ biometric principles for cancer detection. It is obvious from the resultant data that cancer is a very complex disease process involving numerous gene alterations in varying numbers of molecular pathways. This in turn renders the concept of the single biomarker as passé with respect to cancer detection. The multi-marker approach is evident as it avoids the pitfalls of a single marker concept as demonstrated by the current shortcomings of the prostate specific antigen marker for prostate cancer. Due to the overwhelming complexity of the malady high-throughput protein detection coupled with protein modeling in health and disease may yield the tools necessary for life saving early cancer detection.

4. Acknowledgement

The research presented in this chapter was supported by the Avon Breast Cancer Foundation (#07-2007-071), Komen Foundation (KG080928), Gillson-Longenbaugh Foundation and the Texas Ignition Fund. The authors would also like to thank Dr. William Dubinsky for the LC-MS/MS salivary mass spectrometry analyses and Dr. Karen Storthz for her technical expertise concerning the western blot analyses.

5. References

- Birkhed D & Heintze U (1989). Salivary secretion rate, buffer capacity, and pH. In: *Human Saliva: Clinical Chemistry and Microbiology* - Volume 1. Tenovuo JO, editor, pp. (25-30), CRC press, Boca Raton, ISBN 0-8493-6391-8.
- Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J & Wieand HS (1999): Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* , 91. 18, (Sept, 1999), pp. (1541-1548), ISSN 0027-8874.
- Delaleu N, Immervoll H, Cornelius J & Jonsson R. Biomarker profiles in serum and saliva of experimental Sjögren's syndrome: associations with specific autoimmune

- manifestations. *Arthritis Research & Therapy* 10. 1,(Feb 2008), (R22-R36), ISSN (electronic) 1478-6362.
- Ernst DJ & Balance LO (2006). Quality collection: the phlebotomist's role in pre-analytical errors. *Med Lab Obs*, 83. 9, (Sept 2006), pp. (30-38), ISSN 0580-7247.
- Gu S, Liu Z, Pan S, Jiang Z, Lu H, Amit O, Bradbury E.M, Hu CA & Chen X (2004). Global investigation of p53-induced apoptosis through quantitative proteomic profiling using comparative amino acid-coded tagging. *Mol Cell Proteomics*, 3.10, (Oct 2004), pp. (998-1008), ISSN 1535-9476.
- Hirtz C, Chevalier F, Sommerer N, Raingard I, Bringer J, Rossignol M & and Deville de Périere D (2006). Salivary protein profiling in type I diabetes using two-dimensional electrophoresis and mass spectrometry. *Clinical Proteomics* 2. 2, (Sept 2006), pp. (117-127), ISSN 1542-6416.
- Kallenberg CG, Vissink A, Kroese FG, Abdulahad WH, Bootsma H. What have we learned from clinical trials in primary Sjögren's syndrome about pathogenesis? *Arthritis Res Ther*. 2011 Feb 28;13(1):205.
- Koomen JM, Li D, Xiao L, Liu TC, Coombes KR, Abbruzzese J & Kobayashi R (2005). Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker recovery. *J Prot Res*, 4. 3, (2005), pp. (972-981), ISSN 1535-3893.
- Koomen JM, Zhao H, Li D, Abbruzzese J, Baggerly K & Kobayashi R (2004). Diagnostic protein discovery using proteolytic peptide targeting and identification. *Rapid Com Mass Spec*, 18. 21, (Nov 2004), pp. (2537-2548) ISSN 1097-0231.
- Moura S, Sousa J, Lima D, Negreiros A, Silva F & Costa L. (2007). Burning mouth syndrome (BMS): sialometric and sialochemical analysis and salivary protein profile. *Gerodontology*, 24, (Sept., 2007), pp. (173-176), ISSN 0734-0664.
- Navazesh M, Christensen C (1982). A comparison of whole mouth resting and stimulated salivary measurements. *J Dent Res*, 61. 10, (Oct 1982), pp. 1158-1162, ISSN 0022-0345.
- Polanski M & Anderson NL. A list of candidate cancer biomarkers for targeted proteomics (2006). *Biomarker Insights*, 7. 1, (Feb 2006), pp. (1-48), ISSN 1177-2719.
- Shevchenko AV, Chernushevic I, Shevchenko A, Wilm M & Mann M (2002). "De novo" sequencing of peptides recovered from in-gel digested proteins by nanoelectrospray tandem mass spectrometry. *Mol Biotech*, 20. 1, (Jan 2002), pp. (107-118) ISSN 1073-6083.
- Teisner B, Davey MW & Grudzinskas JG (1983). Interaction between heparin and plasma proteins analyzed by crossed immunoelectrophoresis and affinity chromatography. *Clin Chem Acta*, 127. 3, (Feb 1983) pp. 413-417, ISSN 0009-8981.
- Ward LD, Reid, GE, Moritz RL & Simpson RJ (1990). Strategies for internal amino acid sequence analysis of proteins separated by polyacrylamide gel electrophoresis. *J Chroma*, 519. 1, (Oct 1990), pp. (199-216), ISSN 0021-9673.
- Wilmarth PA, Riviere MA, Rustvold DL, Lauten JD, Madden TE & David LL (2004). Two dimensional liquid chromatography study of the human whole saliva proteome. *J Prot Res*, 3. 5, (Sept-Oct, 2004), pp. (1017-1023), ISSN 1535-3893.
- Wu AJ, Atkinson JC, Fox PC, Baum BJ & Ship JA (1993). Cross-sectional and longitudinal analyses of stimulated parotid salivary constituents in healthy, different-aged subjects. *J Geron*, 41. 5, (Sept,1993); pp. (M219-224), ISSN 1079-5006.