Workshop Proceedings

Biometrics: Challenges arising from Theory to Practice

Inaugural BCTP Workshop

Cambridge, United Kingdom, August 22nd 2004

(Satellite Workshop to IEEE - ICPR 2004, August 22nd - 27th 2004)

Editors:

Claus Vielhauer
Simon Lucey
Jana Dittmann
Tsuhan Chen

Workshop Proceedings

Biometrics: Challenges arising from Theory to Practice

ISBN: 3-929757-3

1st Edition, 2004

Editors: Claus Vielhauer, Simon Lucey, Jana Dittmann, Tsuhan Chen

Publisher: Univ.Magdeburg

Otto-von-Guericke University Magdeburg
School of Computer Science
Department of Technical and Business Information Systems (ITI)
Advanced Multimedia and Security Lab (AMSL)

Universitaetsplatz 2
D-39016 Magdeburg
Germany

Printed in Germany.

Introduction

The process of a biometric authentication is one of today's most interesting challenges in pattern recognition. Biometric systems based on a plethora of modalities have been introduced throughout the last couple of decades. By providing a security mechanism based on physical traits or behavior, biometric security systems appear superior to other mechanisms based on ownership or knowledge. Today, it can be stated that a significant number of these systems have reached a high degree of maturity, which is reflected in their broad commercial availability.

With technology migrating from theory to practice, new scientific challenges beyond technical implementation arise. The workshop "Biometric Challenges from Theory to Practice", for the first time held in conjunction with the 17th IEEE International Conference on Pattern Recognition (ICPR) 2004 in Cambridge, United Kingdom, has attracted a great number of researchers from scientific and industrial institutions, as well as academia. From a great number of proposals, the program committee had the honor to put together a program consisting of invited talks and papers, as well as reviewed publications. There will be three invited talks from renowned scientists (Daugman: "Large-scale deployment of biometric devices for iris recognition", Huang: "Automatic Audio-Visual Person Identification in Practice: Some Challenging Issues" and Phillips: "Challenges Face Recognition and Multi-biometrics"). Two well known collegues have agreed to contribute invited papers (Zhang: "Low-Resolution Palmprints for Personal Identification" and Pavesic: "Online personal authentication using hand-geometry and palm-print features: the state of the art"). Together with a total of 13 additionally contributed papers, we are looking forward to an extraordinary comprehensive event.

The topics of papers published in these proceedings reflect the diversity of biometric challenges. For **single biometric modalities**, the workshop has received contributions for hand (palmprint, geometry and grip patterns), fingerprint, face and voice biometrics. In their paper, Zhang et al. discuss palmprint classification based on low-resolution images in the context of aging. In another paper addressing hand biometrics, a new technique for hand authentication based on projection invariant cross ratio hand descriptors is introduced by Zheng et al. Veldhuis and Bazen exemplify a new technique for feature dimensionality reduction by use of the recent modality of biometric grip patterns. Wolf discusses the problem of aging for the modality of speaker recognition, and presents experimental results from a test of 12-month duration. The work of Uludag and Jain addresses the new challenges of using biometric in cryptographic systems by exploring the possibilities to use fingerprint minutiae for a novel cryptographic approach, called Fuzzy Vault. In another fingerprint related work, Bhanu et al. introduce a binomial model to predict performance of fingerprint recognition in large populations. While the vast majority of scientific work in the area of biometrics address user authentication, a number of additional challenging goals exist. Suresh et al., for example, contribute towards automated anchorperson indexing of video sequences, i.e. generation of visual table of contents, based on facial features.

Multimodality, or the synergetic combination of biometric devices is another bias of the workshop. Incorporating the external knowledge privy to a distributed network of practical biometric devices (e.g. time of day, emergency status, construction work, recency of last visit, etc.), one can make more informed decisions and combine decisions from different devices in a more effective manner. This aspect of biometrics has been addressed by a number of authors of these proceedings. Beattie et al. suggest path fusion as a mechanism for combining verification decision across space and time. Fusion approaches for two particular biometric modalities, hand geometry and palmprint features, are provided as a comprehensive state-of-the-art in the paper of Pavesic et al. Yanikoglu and Kholmatov investigate the fusion of two verification instances of fingerprint matching in order to increase privacy. A system for combining acoustic and visual biometric features for person authentication based on an autoassociative neural network (AANN), which has been evaluated by TV sequences is presented by Palanivel et al.

Besides aspects of increasing recognition accuracy of mono- or multimodal biometric authentication systems, four papers of the workshop address the dimensions of **security for biometric references**,

cultural impacts and **evaluation** of experimental data. Vatsa et al. elaborate on the effectiveness of reference protection by watermarking algorithms for embedding iris codes in facial images. Schimke et al. discuss a methodology for cross-cultural evaluation of biometrics with respect to recognition accuracy, social acceptance and legal considerations. The statistical dependency among False Acceptance rates and False Rejection Rates from evaluation data is studied by Bolle et al. Evaluation in attack scenarios to face recognition systems, based on changes in hair style and modification of appearance by adding and removing mustaches and eye glasses are experimentally analyzed in a paper by Aksan et al.

The workshop chairs would like to thank all participants, authors, organizers and the IEEE ICPR management for supporting our efforts to realize this workshop. Our special thanks go to Marten Wenzel for his help in preparing these proceedings. We look forward to a successful meeting and hope for a continuation of this kind of event in the near future.

Claus Vielhauer

Otto-von-Guericke-University of Magdeburg, Germany
School of Computer Science
Advanced Multimedia and Security Lab (AMSL)

Simon Lucey

Carnegie Mellon University of Pittsburgh, USA
Department of ECE
Advanced Multimedia Processing Lab

Jana Dittmann

Otto-von-Guericke-University of Magdeburg, Germany
School of Computer Science
Advanced Multimedia and Security Lab (AMSL)

Tsuhan Chen

Carnegie Mellon University of Pittsburgh, USA
Department of ECE
Advanced Multimedia Processing Lab

Table of Contents

Low-Resolution Palmprints for Personal Identification

[1]David Zhang, [1,2]Adams Wai-Kin Kong, [1,3]Guangming Lu, [3]Xiangqian Wu and [1]Michael Wong

[1]Biometrics Research Centre
Department of Computing,
The Hong Kong Polytechnic University
Kowloon, Hong Kong
{csdzhang, cswkkong, csglu, ,csmkwong}@comp.polyu.edu.hk

[2]Pattern Analysis and Machine Intelligence Lab
University of Waterloo, 200 University Avenue West, Waterloo Ontario, N2L 3G1, Canada

[3]School of Computer Science and Technology
Harbin Institute of Technology, Harbin, 150001, China

Abstract

Recognizing people based on their biological or behavioral characteristics, called biometric authentication, has been applied over hundred years. Two important issues, 1) distinctiveness and 2) permanence should be considered for biometric authentication. Distinctiveness of a biometric refers to that any two persons should be sufficiently different in terms of the features in the biometrics. Permanence of a biometric refers to that the features should be sufficiently invariant over a period of time. In this paper, we will provide evidences for discussing the distinctiveness and permanence of low-resolution palmprints, which has high potential for commercial security systems. In addition to distinctiveness and permanence, we will discuss another fundamental issue, palmprint classification. We define six classes of palmprints based on number of principal lines and their intersection points.

1. Introduction

Using human body or human behavior for personal authentication called biometric authentication has a long history. In fact, we have used it day to day. We commonly recognize people based on their face, voice and gait for social communication. Signatures are recognized as an official verification method for legal and commercial transactions. Fingerprints and DNA have been considered effective methods for forensic applications including criminal investigation, corpse identification and parenthood determination. Recently, more and more effort has been put on developing effective automatic personal identification systems for various security demands. No matter what biometric is used for commercial applications or forensic applications, we have to face two fundamental problems, distinctiveness and permanence. Distinctiveness of a biometric refers to that any two persons should be sufficiently different in terms of the features in the biometric. Permanence of a biometric refers to that the features should be sufficiently invariant over a period of time. Undoubtedly, these two issues control the accuracy and reliability of the biometric security systems.

Low-resolution palmprints have drawn our attention over several years because of theirs rich features including principal lines, wrinkles, and texture, and effective computation. From inked to inkless palmprints, we only have one goal: using a low-resolution palmprint for personal identification. In this paper, low-resolution palmprint is referred to palmprint images with 75 dpi (dot per inch), which is totally different from high resolution palmprint images (500 dpi), where a lot of detailed features such as core and minutiae points can be extracted for forensic applications [1]. All previous papers only concentrated feature representation, matching and system development [3-6]. So far, only some authors give their objective comments for palmprint permanence and distinctiveness [7]. However, none of pervious papers systematically discusses about these issues. Without the deep investigation of these two issues, how can we know that using low-resolution palmprint for personal identification is reliable? In this paper, we will provide various evidences to show that low resolution palmprints are distinctive and stable over a long period of time. To discuss the distinctiveness, two experiments should be conduced. First, we should compare large amounts of palmprints from different persons to investigate whether low-resolution palmprints contain enough distinctive information or not. In the second test, identical twins palmprints are compared to investigate whether they can be separable or not. Some biometrics such face and DNA cannot pass this test. To discuss the permanence, palmprints from the same palms are collected from different time to show the invariance of palmprints.

Conducting the above experiments by human vision is very time-consuming and subjective. We have to find out

an effective feature representation method for low-resolution palmprints to objectively evaluate the similarity between two palmprints. Recently, we modify our previous approach [3] to use orientation of palm lines for personal identification [8]. This method gives a matching score for objectively evaluating the similarity between two palmprints. The matching score range between zero and one. The smaller matching score indicates more similarity between two palmprints. Zero indicates the perfect matching.

In addition to distinctiveness and permanence, classification is another important issue for biometric recognition. Fingerprints have well defined classes, called Henry classes, including, left loop, right loop, whorl, arch and tented arch. However, so far, palmprint classification is not well defined. In this paper, we define classes of palmprint based on the number of principal lines and their intersection points.

The rest of this paper is organized as the following. The databases used in this paper are described in Section 2. The distinctiveness and permanence of palmprints are discussed in Section 3 and Section 4, respectively. Six classes of palmprints are defined in Section 5. A brief conclusion is provided in Section 6.

2. Databases

Two palmprint databases are collected for this paper. All the palmprint images are collected by our self-designed palmprint capture devices. The basic principal of the device has been published in [3]. Originally, the images have two kinds of sizes such as 384×284 and 768×568. In the experiments, all the images are normalized to 384×284. The corresponding resolution is about 75 dpi.

2.1. General Database

The first database contains 7,337 palmprints images from the left and right hands of 364 persons. Each person provides about 10 images each of the left palm and the right palm. In this dataset, 241 people are male, and the age distribution of the subjects is about 82% are younger than 30, about 2% are older than 50 and about 16% is aged between 30 and 40.

2.2. Twin Database

Twin database contains 590 palmprint images from 30 pairs of identical twins' palms. Each of them provides around 10 images for their left palms and 10 images for their right palms. Their age range is between 6 and 45.

3. Distinctiveness

In this test, we interest in the information in the palmprint, whether it is sufficiently enough for identifying a person from large population, or not. In other words, can we find out some palmprints from different palms but they are very similar? To investigate the distinctiveness of palmprints, two experiments should be conducted.

In the first experiment, each image in the general dataset compares all the others in this dataset. The matching score is considered as a genuine matching score if two palmprint images come from the same palm; otherwise, it is considered as an imposter matching score. Each palm in this dataset has several images and therefore, when two palms are compared, many imposter matching scores are generated. Similarly, we have many genuine matching scores for each palm. We take the mean of all the imposter matching scores generated by matching palmprint images belonging to one palm with all the other palmprint images from the other palms to represent the dissimilarity of this palm to the other palms. Similarly, we take the mean of all genuine matching scores of each palm. Therefore, each palm has a mean of genuine matching scores and a mean of imposter matching scores. Fig. 1 plots the means of genuine and imposter matching scores against user identity. We can see the means of genuine and imposter matching scores of each palm are completely separable. It means that palms from different persons contain enough distinctive information for identifying a person from large population.

Test of identical twins is considered as an important test for biometric authentication but not all biometrics including face and DNA can pass this test. To conduct this test, we match a palmprint in the twin database with his/her identical twin sibling to produce imposter matching scores. Since number of images in this database is relatively few, we directly use the matching scores for plotting the imposter distribution; otherwise, we cannot get the smooth imposter distribution. The twins' imposter distribution is given in Fig. 2, which the genuine distribution is estimated by the means of genuine matching scores as shown in Fig. 1. From Fig. 2, identical twins' palmprint can easily be separated, just like twins' fingerprints [2]. Fig. 3 shows that three pairs of palmprint images from three pairs of identical twins. We can observe that identical twins' palmprints still contain a lot of distinctive information.

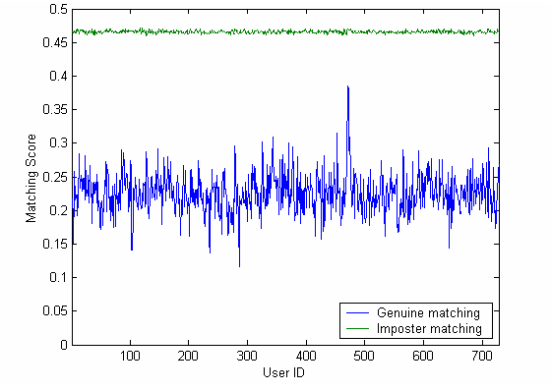


Figure 1. The means of genuine and imposter matching scores for distinctiveness test.

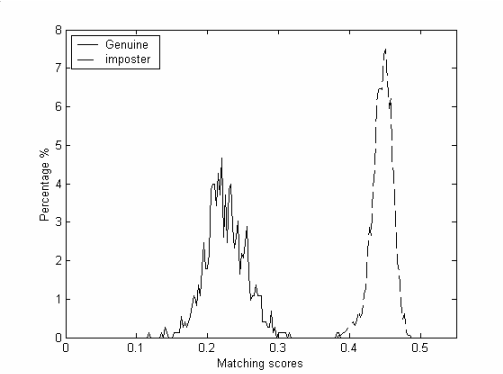


Figure. 2 The genuine and imposter distributions for measuring the similarity of identical twins' palmprints.

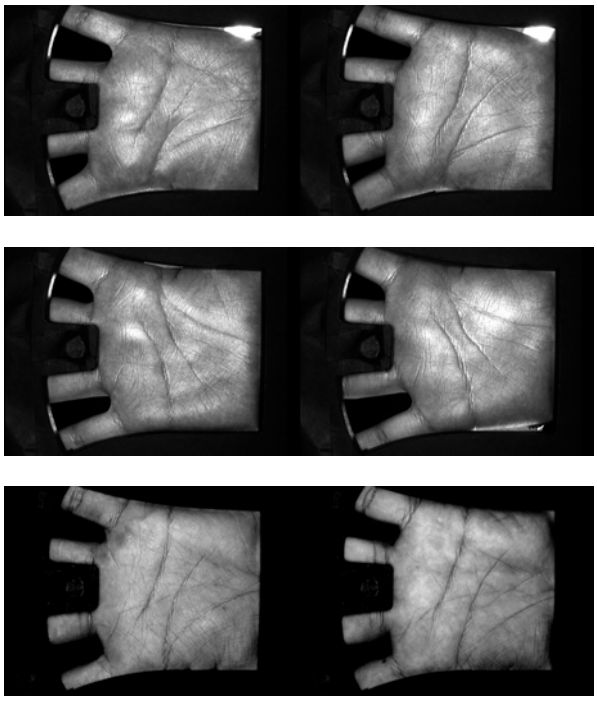
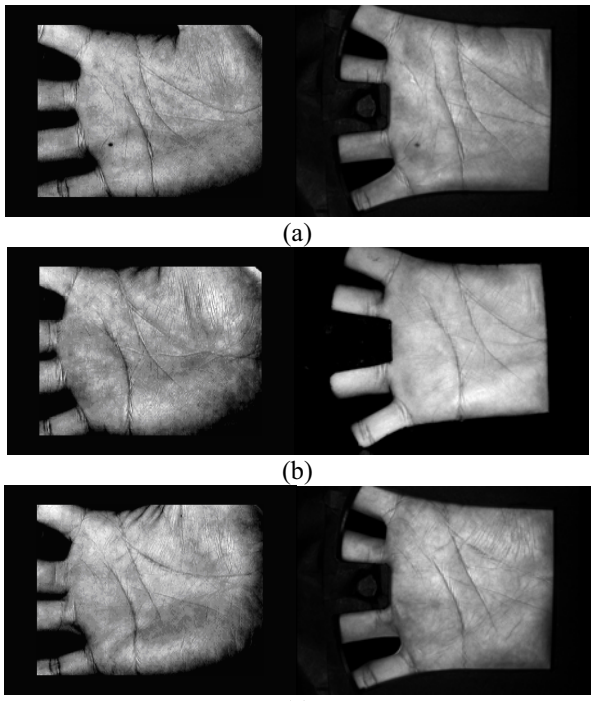


Figure 3. Three pairs of palmprints from three pairs of identical twins.

4. Permanence

Permanence is one of important issues for biometric identification. Each biometric has some variations. Even for DNA, mutation is one of the means to change it. Face change depends on our weight, age and living styles. Undoubtedly, palmprints have similar situation. Our hands are growing from childhood to adulthood, which implies that palmprints is changing at that period of time. Our previous study shows that our method can recognize palmprints collected with a period of time over several months [3]. Figs. 4a, 4b and 4c show three pairs of palmprint images collected with periods of 1,288 days, 1,340 days and 1,166 days, respectively. In term of image intensity, they have some difference since they are collected by different capture devices. In term of the features, principal lines and wrinkles, they are completely stable. We do not discover any observable change from those features.



(a) (b) (c)

Figure 4. Three pairs of palmprint images with long intervals, a) 1,288 days, b) 1,340 days and c) 1,166 days

5. Palmprint Classification

Classification is also an important issue for palmprint identification, especially for large databases. However, so far, a well-defined palmprint classification method does not exist. In this section, based on the number of principal lines and their intersection points, we define a palmprint classification algorithm with six categories.

To classify a palmprint, we first extract its principal lines and then classify the palmprint by the number of the principal lines including heart, life and head lines [4] and the intersections of these principal lines. As the number of each type of principal line is less than or equal to 1, there are at most three principal lines. Two principal lines are said to intersect only if some of their points overlap or some points of one line are the neighbors of some points of another line. If any two principal lines intersect, the number of intersections increases by 1. Therefore, the number of intersections of three principal lines is less than or equal to 3.

Regarding the number of principal lines and the number of the intersections of these lines, palmprints can be classified into following six categories:

Category 1: Palmprints composed of no more than one principal line (Fig. 5 (a));

Category 2: Palmprints composed of two principal lines and no intersection (Fig. 5 (b));

Category 3: Palmprints composed of two principal lines and one intersection (Fig. 5 (c));

Category 4: Palmprints composed of three principal lines and no intersection (Fig. 5 (d));

Category 5: Palmprints composed of three principal lines and one intersection (Fig. 5 (e));

Category 6: Palmprints composed of three principal lines and more than one intersection (Fig. 5 (f)).

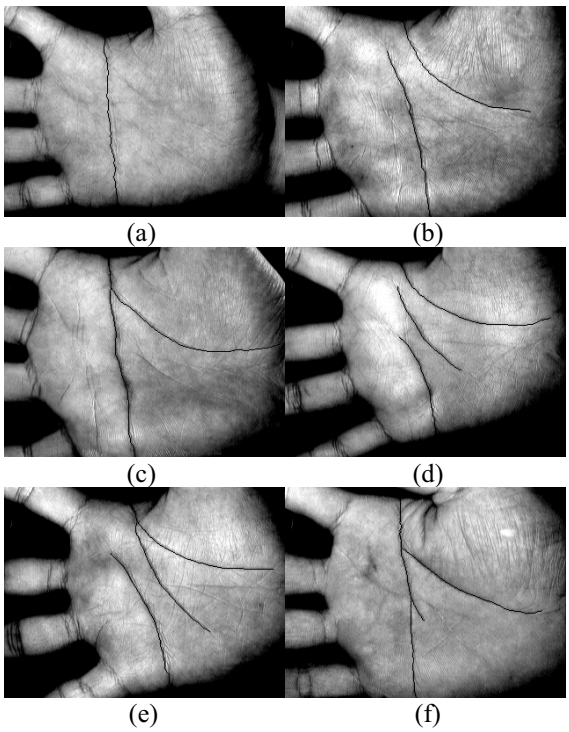


Figure 5. Six typical palmprint images from different palms. (a)-(f), categories 1-6, respectively

6. Conclusion

In this paper, we touch three important problems of using low-resolution palmprint for personal identification - distinctiveness, permanence and classification. We use a modified method of [3] to represent a palmprint, which bases on the orientation of palm lines [8]. For distinctiveness, 7,337 palmprint images from 728 different palms and 590 palmprint images from 60 identical twins' palms are tested. The experimental results show that palmprints contain rich distinctive information for personal identification, just like iris and fingerprints. For permanence, we show three pairs of palmprint images collected with periods of three years. They illustrate that palmprints are stable over a long period of time. According to the experimental results and the images, we believe that palmprint is highly distinctive and stable for personal identification. In addition, based on the number of principal lines and their intersection points, we define six palmprint categories. Based on the palmprint database we collected, we summarize that 0.36% samples belong to Category 1, 1.23% to Category 2, 2.83% to Category 3, 11.81% to Category 4, 78.12% to Category 5 and 5.65% to Category 6.

References

[1] http://www.nectech.com/afis/download/PalmprintDtsht.q.pdf.

[2] A.K. Jain, S. Prabhakar and S. Pankanti, "On the similarity of identical twin fingerprints," *Pattern Recognition*, vol. 35, no. 11, pp. 2653-2662, 2002.

[3] D. Zhang, W.K. Kong, J. You and M. Wong, "On-line palmprint identification", *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1041-1050, 2003.

[4] D. Zhang and W. Shu, "Two novel characteristics in palmprint verification: datum point invariance and line feature matching," *Pattern Recognition*, vol. 32, no. 4, pp. 691-702, 1999.

[5] C.C. Han, H.L. Cheng, K.C. Fan and C.L. Lin, "Personal authentication using palmprint features," *Pattern Recognition*, vol. 36, no 2, pp. 371-381, 2003.

[6] N. Duta, A.K. Jain, and K.V. Mardia, "Matching of palmprint," *Pattern Recognition Letters*, vol. 23, no. 4, pp. 477-485, 2001.

[7] A. K. Jain, A. Ross and S. Prabhakar, "An introduction of biometric recognition," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 14, no 1, pp. 4-20, 2004.

[8] W. K. Kong and D. Zhang, "Competitive coding scheme for palmprint verification", To appear in *ICPR,* 2004.

Comparing Robustness of Watermarking Algorithms on Biometrics Data

Mayank Vatsa[1], Richa Singh[1], Pabitra Mitra[1] and Afzel Noore[2]

1- Department of Computer Science & Engineering,
Indian Institute of Technology, Kanpur, India

2 – Lane Department of Computer Science & Electrical Engineering,
College of Engineering and Mineral Resources
West Virginia University, Morgantown, USA
Email: {mayankv, richas, pmitra}@cse.iitk.ac.in, Afzel.Noore@mail.wvu.edu

Abstract

Attacks on biometric templates are becoming common. Watermarking helps in template protection as well as facilitating multi-biometric verification. Several combinations of watermarks and cover images are possible. In this paper, we present the case where a face image is the cover and an iris code is used as a watermark. The effectiveness of selected watermarking techniques is evaluated by comparing the matching performance using face recognition and iris recognition algorithms. The robustness to various techniques is studied when the watermarked image is subjected to several attacks.

1. Introduction

There are many critical issues in designing a practical biometrics system. They are characterized into major categories based on accuracy, computation speed, cost, security, scalability and real time performance. The security of biometric data and templates is of paramount importance and must be protected from external attacks.

The common attacks in biometric systems are coercive attack, impersonation attack, replay attack, and others. There are some attacks in which the hacker can manipulate the feature extractor, extract specific pre-selected features, and alter the contents of the database where biometric templates are stored. For the development of linked databases as an international commercial highway for information exchange, it is required to protect the template database to keep the information of users secure at all times.

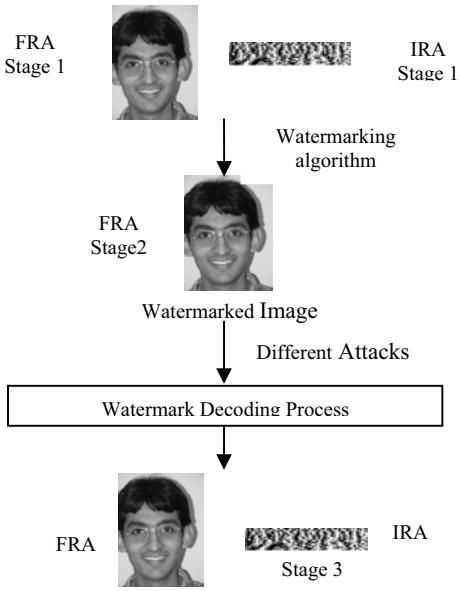
Biometrics template can be secured using encryption and watermarking techniques. Encryption does not provide security once the data is decrypted. On the other hand, watermarking involves embedding information into the host data imperceptibly, to provide additional security. However, embedding a watermark may change the inherent characteristics of the host image. Therefore, the verification performance of the decoded watermarked images should not be inferior compared to the performance of non-watermarked images. Previous work [4, 5, 6, 7] has shown that watermarking can be a good technique to protect biometrics templates. Both the watermark and the cover image should be the biometrics template of the same user to achieve high level security for template protection and multilevel cross-authentication.

In this paper we compare four watermarking algorithms (LSB substitution [1], Modified Correlation (MCBA) [2], Modified 2D Discrete Cosine Transform (M2DCT) [3] and Discrete Wavelet Transform (DWT) [4]) on biometrics template based on their robustness to different attacks. Different choice of watermark data and cover object is found to be the major issue in this robustness comparison. Ideally any biometrics template can be treated as watermark data and another as cover object. But relative information content restricted the admissible combinations. In this comparative study we choose the iris code generated from the iris image as the watermark and the face image as the cover object. The iris code is unique for every individual. In its binary form it is similar to a noisy image and qualifies as a good watermark. The size of face image is sufficiently large to allow the embedding of iris code as watermark without causing any major distortion. We have tested the watermarking algorithm on seven attacks and have summarized our results in this paper.

2. Comparing Robustness

To compare the robustness of watermarking algorithms on biometrics data, a prototype system has been designed (Figure 1). In this system, a face image and an iris image is given as input to the watermarking algorithm where an iris code (in binary form) is embedded into the face image. The resulting watermarked face image is matched using a Face Recognition Algorithm (FRA) and a matching score is calculated. Next, the watermarked face

image is used as input to the watermark decoding process. The output is the extracted face image and the decoded iris code. These outputs are then matched using FRA and Iris Recognition Algorithm (IRA), and the matching scores are calculated for both the images. We next study the effect of attacks on the watermarked face image. The performance of the watermarking algorithm is compared based on the matching scores/ percentage accuracy of the face image and iris code calculated at different stages. In this section algorithms of face recognition, iris recognition, watermarking and the criterion for comparing robustness of watermarking algorithms are discussed in detail.



"Figure 1. Diagram for Robustness Comparison"

2.1 *Face and Iris Recognition*

We have designed a face verification algorithm based on Line Based Face Recognition (LBA) [9]. The matching score obtained by this algorithm is used for verification purposes.

For generating the iris code from the iris image, 1D log Gabor based iris template generation algorithm is used. Iris detection is performed using the algorithm described in [12]. From the output of iris detection i.e., texture of the iris, features are extracted using the algorithm based on 1D log Gabor [10]. These features are encoded into bit patterns called the Iriscode. For generating the iris template, 2D normalized pattern is transformed into a number of 1D signals and convolved with the 1D log Gabor wavelets. This iris code is the textural representation of features of the iris in binary form of size 10x100. An example of an iris code is shown in Figure 2. Bit shifting based Hamming distance matching algorithm [11] is used for iris code matching and obtaining the matching score of the iris recognition algorithm.

"Figure 2. Iris Code"

2.2 Watermarking Algorithms

Four watermarking algorithms are selected as representative approaches covering the spatial, frequency and wavelet domains. These are LSB substitution [1], MCBA [2], M2DCT [3] and DWT [4].

2.3 Criterion for Comparing Robustness

With the large number of watermarking algorithms recently developed, it is challenging to compare the performance of different algorithms. In this paper, we compare the robustness of watermarking algorithms on biometrics data. The algorithms are tested in a standard form using similar operations to make the comparisons meaningful. The dimensions of the cover images (database of Face images) and watermark images (Iriscodes) are fixed. The robustness is tested using grayscale manipulations and geometric transformations. Attacks on watermarked image is tested using JPEG Compression (based on quality factor), blurring (based on kernel size), Gaussian Noise addition (based on mean and variance), Median filtering (based on kernel size), Gamma Correction (based on gamma exponent), converting the image into "eps" format with 72x72 dpi and then converting it back to bmp format and printing the watermarked image on paper at different resolutions and scanning it at different resolutions. The last attack is performed to simulate the case where a credit card is scanned and the biometrics data is used for authentication. Since the synchronization pattern can be combined with different watermarking algorithms, the test of robustness with respect to geometric deformations has not been included here. The guidelines for comparing watermarking algorithm are as follows:

1. Size of cover image (Face) is fixed (1024 x 768).
2. Size of watermark (Iriscodes) is fixed (10x100).
3. Cover images and watermark images are in grayscale and in "bmp" format.
4. Constants/ Kernel of the four algorithms and the attacks are fixed for all the images.
5. One common key is used for all algorithms.
6. Thresholds of matching scores of FRA and IRA are fixed for all cases (3 stages, 4 algorithms and 7 attacks).
7. Perform JPEG compression, Gaussian noise addition, median filter, blurring, gamma correction, conversion of image from 'bmp' format to 'eps' and then again to bmp format, and printing - scanning attacks on the watermarked images of the four algorithms.

8. Test for the verification after every stage (for both face and iris) [Figure 1].

3. Experimental Results

Experiments have been carried out on the database consisting of face and iris images from 50 individuals (5 face and 5 iris images from each individual). Three face images and three iris images from each individual have been used for training and the rest of the images have been used for testing. The percentage accuracy is being used as the criterion for comparing the robustness of the watermarking algorithms. We have verified the user's biometrics data at every stage (Figure 1) and calculated the accuracy. The watermarking algorithm is said to be robust if it shows good verification accuracy when subjected to an attack. The verification accuracy of face recognition and iris recognition before introducing watermarking algorithms (at Stage 1) are 91.2% and 98.48% respectively (including FAR and FRR).

This section is divided into eight subsections. The first subsection shows the accuracy when the watermark is embedded and decoded. The remaining seven subsections evaluate the robustness of various attacks.

3.1. *Watermark Embedding and Decoding*

At Stage 2 only the FRA has been tested as the output is only a face image. At this stage the iris code has been inserted into the face image and the face verification is performed. Table 1 shows the accuracy of the FRA on four watermarking algorithms.

Algorithm	LSB	MCBA	M2DCT	DWT
Accuracy (%)	91.2	91.2	91.2	91.2

Table 1: Accuracy of Watermarked Face Image

This table indicates that the watermark embedding using the four algorithms does not have any effect on face verification. At Stage 3 (without attack), both face verification and iriscode verification are performed. The performance of the watermarking algorithms is shown in Table 2. It also indicates that there is no significant change in the accuracy of either FRA or IRA for the four algorithms.

Algorithm	LSB	MCBA	M2DCT	DWT
FRA Accuracy (%)	91.2	91.2	91.2	91.2
IRA Accuracy (%)	98.48	98.48	98.0	98.48

Table 2: Accuracy after decoding the watermark

3.2. *JPEG Compression*

The first attack tested on the watermarked face image is JPEG compression. After compressing the watermarked face image the decoding process is carried out. Table 3 shows the verification accuracy obtained after decoding the face image and decoded watermark iriscode.

Algorithm	LSB	MCBA	M2DCT	DWT
FRA Accuracy (%)	91.2	91.2	91.2	91.2
IRA Accuracy (%)	0.0	95.42	98.48	98.48

Table 3: Accuracy on watermark recovered from JPEG compressed image

This table indicates that the LSB based watermarking algorithm fails on JPEG compression. In MCBA, the accuracy dropped due to increase in FRR. DWT and M2DCT did not change accuracy on IRA. The face verification showed that it was able to withstand JPEG compression well.

3.3. *Gaussian Noise Addition*

When adding Gaussian noise, LSB fails for Iriscode verification even when the standard deviation is 5 gray scale levels. MCBA gives the accuracy of 82.44% and 54.72% on IRA when the standard deviation is 5 gray levels and 40 gray levels respectively. M2DCT performs better than these two and the accuracy is maintained until the standard deviation of 35 gray levels is reached. Beyond this, the accuracy of the IRA decreases. Of all the four algorithms, the performance of DWT was the best. It maintains an accuracy of 98.48% until the standard deviation of 150 gray levels and outperforms when Gaussian noise is added. There is no change in the accuracy of face verification up to the standard deviation of 150 gray values.

3.4. *Median Filter*

Neither of the algorithms survived the median filtering at any kernel size. At kernel size of 2 and 3, MCBA performs best with the IRA accuracy of 71.56% and 64.88% respectively. When the kernel size is 4, M2DCT performs best with the IRA accuracy of 62.01%. For kernel size of 5, DWT performs best with an IRA accuracy of 55.46%. For FRA the accuracy remains at 91.2% when a maximum kernel size of 5 is used.

3.5. Blurring

All watermarking algorithms, except the LBA, exhibit good resistance to multiple blurring till 5 successive applications of a blurring filter with the kernel of size 3x3.

Table 4 shows the results obtained in the blurring process. For FRA, the accuracy remains almost the same; where as for IRA, the M2DCT performs the best of all four algorithms.

Algorithm	LSB	MCBA	M2DCT	DWT
FRA Accuracy (%)	90.9	90.6	90.9	90.6
IRA Accuracy (%)	46.55	94.89	97.21	97.16

Table 4: Accuracy on recovering the watermark from a Blurred Image

3.6. Gamma Correction

DWT and MCBA have no effect on the results when performing Gamma correction in the range of 1.0 to 9.9; but M2DCT and LSB algorithms have slight change in the accuracy of iris, i.e. 97.32% and 96.78% respectively.

3.7. Conversion into "eps" format and again to "bmp".

This attack is performed to check for the differences on file format conversion. In this experiment, MCBA performs the best with the unchanged accuracy of 98.48%. Table 5 shows the results of this experiment.

Algorithm	LSB	MCBA	M2DCT	DWT
FRA Accuracy (%)	91.2	91.2	91.2	91.2
IRA Accuracy (%)	51.56	98.48	66.71	64.86

Table 5: Accuracy on File Format Conversion

3.8. Robustness on Printing – Scanning

This experiment simulates the case of credit card tampering. The biometric template on a credit card is protected by watermarking it with another biometrics template of the same person to use it for cross verification. To test for the possible attack on such an application, we have printed the watermarked face images at 300, 600 and 1200 dpi, scanned these images at 600 dpi and then extracted the watermark. In this process at 300 dpi and 600 dpi printing, M2DCT performs best with an IRA accuracy of 69.22% and 58.85% respectively. DWT performs best at 1200 dpi printing with an IRA accuracy of 51.44%. The accuracy drops as the printing resolution is increased because of the damage to the watermark.

4. Conclusion

The security and integrity of the biometric data is an important issue for the implementation of biometric systems. Encryption, watermarking, and steganography are possible techniques to secure biometrics data. In this paper a comparative study of watermarking algorithms on biometrics data is performed. We have used face as a cover image and an iris code as watermark. Watermarking helps in template protection as well as facilitating multi-biometric verification. Performance of several watermarking strategies with respect to their verification performance under different algorithms is compared. The results show that the DWT based watermarking algorithm performs best on most of the attacks followed by M2DCT and MCBA.

We notice that in the spatial domain, the LSB watermarking technique is relatively easy to embed the biometric iris code. However, it is highly sensitive to small changes or modifications in the watermarked face image. The loss of the embedded iris code is clearly noticeable when subjected to basic compression for transmission of the watermarked image. The frequency domain techniques are more robust and less susceptible to attacks such as compression, filtering, and image processing operations. This paper shows that the multi resolution property of the discrete wavelet transformation distributes the iris code in the face cover image such that it maintains a high level of robustness and imperceptibility.

5. References

[1]. I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 1673-1687, 1997.

[2]. X. Xia, C. Boncelet, and G. Arce, "A Multiresolution Watermark for Digital Images," *Proceedings of ICIP*, 1997, Vol. I, pp. 548-551.

[3]. M. Barni, F. Bartolini, V. Cappellini and A. Piva, "A DCT domain system for robust image watermarking", *Signal Processing*, Vol. 66, No. 3, 1998, pp. 357-372.

[4]. N. K. Ratha, J. H. Connell and R. M. Bolle, "Secure data hiding in wavelet compressed fingerprint images", *ACM Multimedia 2000,* pp. 127-130.

[5]. U. Uludag, B. Gunsel and M. Ballan, "A spatial method for watermarking of fingerprint images", *Proceedings of WPRIS*, 2001, pp. 26-33.

[6]. A. Jain and U. Uludag, "Hiding Fingerprint Minutiae in Images", *Proceedings of Workshop on Automatic Identification Advanced Technologies* (*AutoID*), 2002, pp. 97-102.

[7]. B. Schneier, "The uses and abuses of biometrics", *Communications of ACM*, Vol. 42, No. 8, 1999, pp. 136.

[8]. J. Dugelay and S. Roche, "A Survey of Current Watermarking Techniques" *Information Techniques for Steganography and Digital Watermarking*, 1999, pp. 121-145.

[9]. O. de Vel and S. Aeberhard, "Line-Based Face Recognition under Varying Pose", *IEEE PAMI*, 1999, Vol. 21, No. 10, pp. 1081-1088.

[10]. J. Bigun and J. M. du Buf, "N-folded symmetries by complex moments in Gabor space and their applications to unsupervised texture segmentation", *IEEE PAMI*, Vol. 16, No. 1, 1994, pp. 80-87.

[11]. J. Daugman, "High confidence visual recognition of persons by a test of statistical independence. *IEEE PAMI*, Vol. 15, No. 11, 1993, pp. 1148–1161.

[12]. P. Richard Wildes, "Iris Recognition: An Emerging Biometric Technology, *Proceedings of IEEE*, Vol. 85, No. 9, 1999, pp. 1348-1363.

Building Access Control using Coordinated Biometric Verification

Michael Beattie, B.V.K. Vijaya Kumar, Simon Lucey, and Ozan Tonguz
Carnegie Mellon University
Electrical and Computer Engineering
Pittsburgh, PA 15213

Abstract

Access control, or granting access to only those identities with an appropriate clearance level is a fundamental component of any building security system. By enforcing access control decisions with electronically controlled locks, such systems limit building access to a limited set of identities. Current building security systems rely on key cards and make use of a building network to distribute valid identities and log accesses. In this paper, we propose the use of biometrics to verify that each key card is presented by the subject to whom it was assigned. To combat errors in biometric verification, we propose further use of the building network to coordinate authenticity decisions from many biometric verifiers spread throughout the building. While other work has investigated the combination of biometric verifiers, we propose an algorithm for combining decisions made across different locations in space and time.

1 Introduction

In an increasingly security-conscious society, access control for "high-value" buildings represents an important tool for protecting both building occupants and the structure itself. This class of buildings might include private or government offices, research laboratories, or even segments of airports that must be closed to the public. Access control for these installations is not a new idea. Many systems have been deployed leveraging key cards and RFID tags to attach an identity to each person requesting access. Because they rely only on possession of a token, existing systems provide only limited security. These systems make no attempt to guarantee the authenticity of a person claiming a particular identity (the *claimant*). A stolen access card would compromise building security in a possibly devastating way. In those scenarios where the highest level of security is required, we propose using a network of coordinated biometric verifiers. Each biometric verifier measures a physiological or behavioral characteristic of the claimant and compares that measurement to a stored template. This enables the verifier to make a strong claim concerning the authenticity of the claimant.

Current biometric verification technologies have a non-negligible error rate. Errors occur due to poor biometric measurements, changes in subject appearance and behavior, and similarities among different claimants. The combination of decisions from multiple biometric verifiers can reduce the error rate of a verification system by introducing decisions made from additional, hopefully diverse, measurements. Such diversity might extend from the use of different types of biometrics (e.g., face and fingerprint) or different verification algorithms. Combining multiple biometrics is not a new idea [1, 2], but most existing work seeks to combine a small number of verifiers working together to cast a single decision. In this document, we discuss the combination of decisions from verifiers spread throughout a building. This is fundamentally different from existing techniques in that the decisions are separated in space and time. We present an efficient methodology for the intelligent combination of decisions from biometric verifiers of differing performance.

2 Biometric Verification: Preliminaries

In order to combine multiple biometric verifiers, we must establish a model for the behavior of a single verifier. Regardless of the biometric traits or algorithms employed, the operation of a biometric verifier proceeds in three fundamental stages: 1) Measure a particular biometric from a presenting claimant; 2) Generate a score by comparing the measurement with a stored template; and 3) Generate a decision by comparing the score to a threshold.

Using nomenclature from [3], we label scores x, thresholds γ, and decisions u. We also identify two distinct classes that might be observed: an authentic claimant (H_1) and an unauthentic claimant (H_0). Scores greater than the

This work has been supported in part by the National Institute of Standards and Technology (NIST).

threshold are accepted as authentic, while scores less than the threshold are rejected as being unauthentic.

Despite general agreement with this model, different verifiers might provide different types of output information. Some may expose the score value, others may provide only a binary decision but allow for the configuration of a threshold, others still may provide a decision without any ability to change the threshold. This heterogeneity may appear across different verifier vendors or even different models from the same vendor. Any attempt to coordinate the combination of decisions from a series of verifiers must attempt to cope with heterogeneity.

There are four possible verification outcomes generated from the product of two possible observation events (authentic and unauthentic claimant) and two possible decisions (accepted and rejected). Of these outcomes, an error occurs when either an authentic claimant is rejected or an unauthentic claimant is accepted. These two events are labeled false accept (FA) and false reject (FR). Minimizing the rate of their occurrence is the key objective in improving verification accuracy. One typically employs the measure of False Accept Rate (FAR) and False Reject Rate (FRR) to ascertain the overall performance of a verifier. The Weighted Error Rate (WER) from [4] is a convenient metric for representing performance with a single value based on R, the *security ratio* or ratio of the importance of FAR to FRR. The results in this paper set R to 1 for equal weights.

$$\mathrm{WER} = \frac{\mathrm{FAR} + \mathrm{R} \cdot \mathrm{FRR}}{1 + \mathrm{R}} \tag{1}$$

3 Optimally Combining Decisions

The optimal combination of verifier decisions falls under the guise of the well-researched topic of *data fusion* [3]. One can envision a scenario in which each verifier casts a local decision u_i and the complete vector of N local decision $\mathbf{u}$ is combined according to some Boolean function. That function is called the *fusion rule* and it can take the form of AND, OR, or any number of other such rules. Again using the nomenclature of [3], we define a fusion rule $F(\cdot)$ that generates a single binary decision u_0. The selection of a fusion rule and local verification thresholds are both important to the problem of optimally combining biometric verifiers.

$$u_0 = F(u_1, \ldots, u_N) = F(\mathbf{u}) \tag{2}$$

In Figure 1, we compare the behavior[1] of applying the OR rule using simple combination, in which thresholds are not changed from their single-sensor configurations, to optimal combination, in which the thresholds have been jointly optimized. It is clear that simple combination provides

[1]The IDIAP2 and EPFL score sets used in this comparison are fully introduced in the evaluation section.

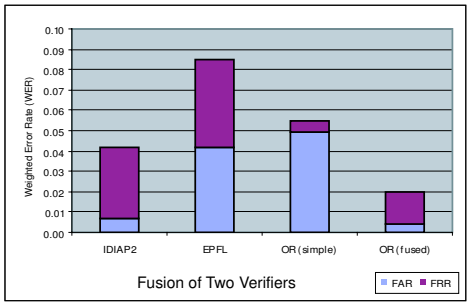


Figure 1. Performance improvement from adjusting thresholds according to the fusion rule

a minimal performance gain over each individual verifier. This is a result of what we call "rule tendencies" for a given fusion rule. The OR rule has a strong tendency toward FA. If any single verifier casts an FA decision, then an unauthentic claimant is accepted as authentic. However, all verifiers must cast an FR decision to cause an authentic claimant to be falsely rejected. As a result, the error rate for verifiers combined using the OR rule is dominated by the FAR. The AND rule reverses this effect.

To improve performance, verification thresholds must be updated according to the fusion rule being employed. In the example of the OR rule in Figure 1, the thresholds of each individual verifier should be increased. As a result, an FA is made to be a rare event at each individual verifier. When all low-FAR verifiers are combined, the FA tendencies of the OR rule are overcome. Once the updated thresholds have been applied, a better weighted error rate can be achieved from the combined verifiers.

One strategy for finding these optimal rules (suggested in [3]) is through Bayesian risk analysis. Applying the concept of a security ratio from the calculation of WER (Eq. 1), we define $R = C_{FA}/C_{FR}$. Optimization is achieved by minimizing the resulting risk function in Eq. 3. Using this method to fuse decisions from a known set of sensors has been treated extensively in [3].

$$\Re = P_{\mathrm{FR}} + \mathrm{R} \cdot P_{\mathrm{FR}} \tag{3}$$

4 Fusion Across Space and Time

The preceding discussion presumed that the set of verifiers being jointly optimized is fixed and that each verifier casts a decision before a combined decision is released. We refer to this scenario as *cluster fusion*. Given a probabilistic score model, the jointly optimal thresholds for verifiers in such a set can be calculated *a priori*. Coordinating verification decisions being made throughout a building represents

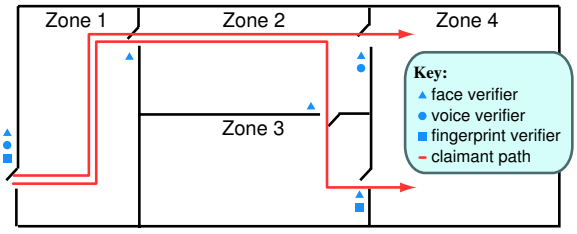


Figure 2. Example paths to Zone 4

a different problem. Within a building, the set of verifiers being combined includes all verifiers to which a claimant has previously presented his biometric. We call an ordered collection of such verifiers a *path*. Paths vary dynamically based on the motion of the claimant, so they cannot be optimized *a priori*. Two examples of possible paths in a hypothetical floor plan are highlighted in Figure 2. As a path grows, the number of available fusion rules becomes extremely large [3]. Some of these available fusion rules lack the rule tendencies described previously. As a result, the penalty associated with ignoring threshold optimization is reduced. This argument permits the combination of decisions over a long path to avoid the complexity associated with determining optimal thresholds.

We describe *path fusion* as a mechanism for coordinating verification decisions across space and time using a network of biometric verifiers. This algorithm provides the service of verification rather than identification, so we must assume that each biometric measurement can be associated with a claimed identity. This might be implemented by requiring a user to carry a key card or RFID tag as suggested in [5]. The claimed identity need not be guaranteed as authentic—it merely provides a hint for linking measurements made at distant verifiers.

To combine these verifiers, we can apply the same risk function given in Eq. 3. In the previous analysis, we fixed a fusion rule and identified the optimal threshold for each verifier in a closed set. Here, we assume that all thresholds are held fixed and identify an optimal fusion rule. By our previous path length argument, rule tendencies will have a smaller effect on paths than small clusters.

Rather than construct a complete fusion rule, we need only find the optimal output for the vector $\mathbf{u}$ of decisions from verifiers along the current path. To determine the best output, we evaluate the risk associated with both FA and FR, then determine the lowest risk decision.

In the spirit of [3], we calculate P_{FA} as follows. Using Bayes rule, we separate $\mathbf{u}$ from u_0. Assuming conditional independence among individual verifiers, we can express the overall $\mathrm{P_{FA}}$ as a product of local decision likelihoods conditioned on H_0. At this point, there is no uncertainty in the first term. It is simply a function of the selected fusion rule.

$$\begin{aligned} P_{FA} &= P\{u_0 = 1|H_0\} && (4)\\ &= P\{u_0 = 1|H_0, \mathbf{u}\}P\{\mathbf{u}|H_0\} && (5)\\ &= P\{u_0 = 1|\mathbf{u}\}\prod_{i=1}^{N} P\{u_i|H_0\} && (6) \end{aligned}$$

Using Eq. 6 and a similar expression for P_{FR}, we can obtain a risk expression for each of the two fusion outputs (Equation 7). In these simple equations, several terms have been dropped because u_0 is known from $\mathbf{u}$ once $F(\cdot)$ is selected. As stated previously, path fusion releases the decision associated with the smaller of the two risks given below.

$$\Re = \begin{cases} R\prod_{i=1}^{N} P(u_i|H_0), & F(\mathbf{u}) = 1 \\ \prod_{i=1}^{N} P(u_i|H_1), & F(\mathbf{u}) = 0 \end{cases} \quad (7)$$

Note that path fusion does not require the evaluation of these $\mathrm{P_{FA}}$ and $\mathrm{P_{FR}}$ expressions at every step, nor does it require a vector of decisions to be passed between verifiers. The algorithm only requires that two running products be maintained to describe the path performance of each claimant. Furthermore, there is no need for a score distribution model beyond an estimated $\mathrm{P_{FA}}$ and $\mathrm{P_{FR}}$ for each verifier. Such estimates can be constructed using relative frequency measurements (in particular, FAR and FRR). This simplicity enables extremely efficient combination of verifiers and also permits the use of heterogeneous device outputs.

5 Evaluation

To evaluate the performance of the path fusion algorithm presented in the previous section, we require a collection of scores from several verifiers. Given these scores and an implementation of the path fusion algorithm, the number of error events can be counted and translated into FAR and FRR values and then compared with the performance of individual verifiers.

As score sets, we have used a set of verification scores collected as part of a biometric verification contest associated with the International Conference on Pattern Recognition in 2000 (ICPR'00). Each of the score sets was generated by applying a different verification algorithm to identities in the XM2VTS database and follows the format specified by the Lausanne Protocol [6].

The score set from each algorithm can be used to represent a different biometric verifier along a client's path. We have constructed a simulator that passes each claimant over a specified path, reading verification decisions associated with comparing a claimant with a particular template along

Path	Step 1	Step 2	Step 3	Step 4
1	EPFL	USYD1	SURREY2	AUT1
2	SURREY2	EPFL	USYD1	IDIAP2
3	IDIAP2	USYD1	IDIAP3	SURREY2

Figure 3. Three paths using XM2VTS scores

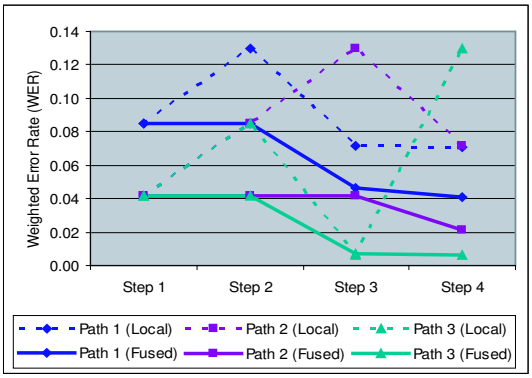


Figure 4. Individual verifier versus path-fused performance.

the way. The previously described path fusion algorithm is used to construct a path-fused decision at each step in the path.

Our evaluation of path fusion isolates three different paths that have been enumerated in Figure 3. Path 1 contains scores from four different face verification algorithms—each with similar performance. Path 2 places a speech verification algorithm from IDIAP at the end of the path. This particular speech algorithm performs better than any of the face verification algorithms, so this path demonstrates the effect of mismatched verifiers in a single path. The third path alternates between high performance speech verification algorithms and typical face verification algorithms to show the effect of this alternation. The performance of both local decisions and path-fused decisions appear in Figure 4.

From the results in Figure 4, it is clear that path fusion provides a strong performance benefit over verifiers operating independently. No performance benefit is seen at the first verifier, because no previous path information is available. The second verifier appears to track the performance of the better of the first two verifiers. More interesting results occur once the third verifier is introduced. At this point, the fused decision performs at least as well as the path's best verifier and tends to outperform all individual verifiers.

As previously discussed, we assume that each claimant presents both his identity and a biometric to each verifier. Our algorithm assumes that the person claiming each identity is fixed as subjects move through the building. If an intruder were to steal one of these identifying devices at an interior node, then he could probably pass the next verifier with little difficulty. Even if the verifier correctly identifies the unauthentic claimant, the decision will likely be overturned by previous verifiers. While this presents a potential security risk to path fusion, multiple local rejections would cause the fused decision to be overturned and the imposter identified. The benefits of path fusion stem from the frequency with which users are verified and the method by which those decisions are combined.

6 Conclusion

In this paper, we have presented a method for providing building access control by combining decisions from multiple biometric verifiers. In particular, we have presented path fusion as a mechanism for combining verification decisions across space and time. Using an existing set of verification scores, we are able to simulate a large number of claimants moving through a building and evaluate the performance of path fusion over each verifier's decisions. The results demonstrate a clear benefit to path fusion—especially considering the computational simplicity of the algorithm.

References

[1] J. Kittler and K. Messer. Fusion of multiple experts in multimodal biometric personal identity verification systems. In *Proc. 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 3–12, September 2002.

[2] L. Hong and A. Jain. Integrating faces and fingerprints for personal identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(12):1295–1307, December 1998.

[3] P.K. Varshney. *Distributed Detection and Data Fusion*. Springer-Verlag, New York, 1997.

[4] E. Bailly-Bailliere et. al. The BANCA database and evaluation protocol. In *Proc. 4th International Conference in Audio- and Video-Based Biometric Person Authentication (AVBPA)*, June 2003.

[5] P. Varshney L. Osadciw and K. Veeramachaneni. Improving personal identification accuracy using multisensor fusion for building access control applications. In *Proc. Fifth Internation Conference on Information Fusion*, volume 2, pages 1176–1183, July 2002.

[6] J. Luettin and G. Maitre. Evaluation protocol for the extended M2VTS database (XM2VTSDB). Technical Report IDIAP-COM 05, IDIAP, 1998.

Fuzzy Fingerprint Vault

Umut Uludag and Anil K. Jain
Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824, USA
{uludagum, jain}@cse.msu.edu

Abstract

Biometrics-based authentication has the potential to eliminate illegal key exchange problem associated with traditional cryptosystems. In this paper, we explore the utilization of a fingerprint minutiae line based representation scheme in a new cryptographic construct called fuzzy vault. Minutiae variability is quantified for a fingerprint database marked by a human expert.

1. Introduction

In traditional cryptography, one or more keys are used to convert the plain text (data to be encrypted) to cipher text (encrypted data): the encrypting key(s) maps the plain text to essentially a sequence of random bits, that can be mapped back to the plain text using the decrypting key(s). Without the knowledge of the correct decrypting keys, the conversion of cipher text to the plain text is *infeasible* (considering time and cost limitations) [1].

Current cryptographic algorithms (e.g., Advanced Encryption Standard (AES) [2], RSA [1]) have a very high proven security but they suffer from the key management problem. All these algorithms fully depend on the assumption that the keys will be kept in absolute secrecy. If the secret key is compromised, the security provided by them immediately falls apart. Another limitation of the these algorithms is that they require the keys to be long and random for higher security, e.g., 128 bits for AES [2], which makes it impossible for users to memorize the keys. As a result, the cryptographic keys are stored somewhere (e.g., in a computer or on a smart card) and released based on some alternative authentication mechanism. The most popular authentication mechanism used for this purpose is based on passwords, which are again cryptographic key-like strings but simple enough for users to remember. Hence, plain text (e.g., multimedia content, email records, financial records, and private encryption keys) protected by a cryptographic algorithm is only as secure as the passwords (weakest link) used for authentication that release the correct decrypting key(s). Simple passwords compromise security; complex passwords are difficult to remember and expensive to maintain. Also, passwords are unable to provide non-repudiation.

Many of these limitations can be eliminated by incorporation of better methods of user authentication. Biometric authentication [3], [4] refers to verifying individuals based on their physiological and behavioral traits such as face, fingerprint, voice, etc. It is inherently more reliable than password-based authentication as biometric characteristics cannot be lost or forgotten. Further, biometric characteristics are difficult to copy, share, and distribute, and require the person being authenticated to be present at the time and point of authentication. Thus, biometrics-based authentication is a potential candidate to replace password-based authentication, either by providing the complete authentication mechanism or by securing the traditional cryptographic keys that contain the plain text.

An interesting cryptographic construct, called *fuzzy vault,* was proposed by Juels and Sudan [5]. This construct, as explained in later sections, has the characteristics that make it suitable for applications that combine biometric authentication and cryptography. In this paper, we explore the use of a fingerprint minutiae representation scheme in this construct (that we call *fuzzy fingerprint vault*). In Section 2, we summarize the related literature. In Section 3, we give specifications about the line-based fingerprint minutiae representation scheme that can be used in securing the fuzzy fingerprint vault. In Section 4, we objectively characterize the variations in fingerprint data using a database that has been marked by a human expert. This helps in quantifying the amount of tolerance that should be introduced into the vault construction. Section 5 concludes the paper.

2. Previous Work

Juels and Sudan's fuzzy vault scheme [5] is an improvement upon the previous work by Juels and Wattenberg [6]. In [5], Alice can place a secret value κ (e.g., private encryption key) in a vault and lock (secure) it using an unordered set A. Bob, using an unordered set B, can unlock the vault (access κ) only if B overlaps with A to a great extent. The procedure for constructing the fuzzy vault is as follows: First, Alice selects a polynomial p of variable x that encodes κ (e.g., by fixing the coefficients of p according to κ). She computes the polynomial projections, $p(A)$, for the elements of A. She adds some randomly generated chaff points that do not lie on p, to arrive at the final point set R. When Bob tries to learn κ (i.e., finding p), he uses his own unordered set B. If B overlaps with A

substantially, he will be able to locate many points in R that lie on p. Using error-correction coding (e.g., Reed-Solomon [7]), it is assumed that he can reconstruct p (and hence κ). The security of the scheme is based on the infeasibility of the polynomial reconstruction problem (i.e., if Bob does not know many points that lie on p, he can not feasibly find the parameters of p, hence he cannot access κ). Note that since this fuzzy vault can work with unordered sets (common in biometric templates, including fingerprint minutiae data), it is a promising candidate for biometric cryptosystems.

Clancy et al. [8] proposed a *fingerprint vault* based on the fuzzy vault of Juels and Sudan [5]. Using multiple minutiae location sets (typically 5), they first find the canonical positions of minutia, and use these as the elements of set A. They added the maximum number of chaff points to find R that locks κ. Note that their system inherently assumes that fingerprints (the one that locks the vault and the one that tries to unlock it) are pre-aligned. This is not a realistic assumption for fingerprint-based authentication schemes (even for iris biometric, this is not true), and limits the applicability of their scheme.

The registration of fingerprints is one of the biggest barriers in the implementation of any fingerprint vault (or any biometrics-based vault). In addition to the possible translational and rotational transformations (see Section 4) and non-linear deformation between two impressions of the same finger, it is possible to have different number of feature points (e.g., missing or spurious minutiae).

In the next section, we propose a *fuzzy fingerprint vault* that uses minutiae lines to lock a secret, using Juels and Sudan's fuzzy vault scheme [5] as the basis.

3. Line-based Minutiae Features

We propose to use a variant of the line-based minutiae representation scheme proposed by Malickas and Vitkus [9] in securing the fuzzy fingerprint vault. Fig. 1 shows the block diagram of the proposed system.

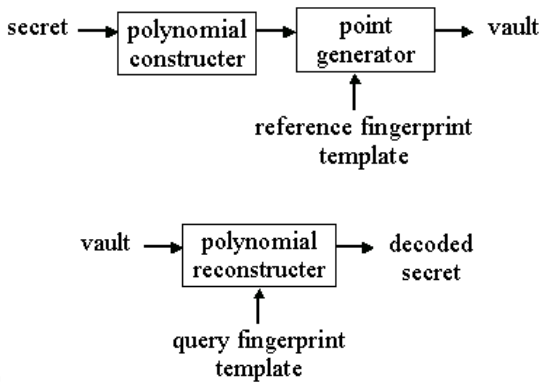


Fig. 1. System block diagram: (a) locking the secret, (b) unlocking the secret.

As explained above, Clancy et al. [8] used only the location of individual minutiae as the locking and unlocking sets for their fingerprint vault. Whereas, in [9], both location and angle of minutiae are used to extract lines for forming the templates. Malickas and Vitkus' method is based on an earlier paper [10] on generic image registration. The main idea is to decompose the registration process into elementary stages and to eliminate only a single transformation parameter (e.g., scaling, translation, or rotation) at each stage [9]. Let I and I' denote the two images to be registered. Assume the current stage of transformation is T_θ. Consider a pair of features f (from I) and f' (from I') of the same type (e.g., point, line). If f and f' have the attributes α and α' (e.g., length, angle) such that $\alpha' = g_\theta(\alpha)$, where g is a bijective function, the parameter θ is called *observable* with respect to the associated feature class and attribute class. The function g allows the current parameter to be estimated as $\theta = h(\alpha, \alpha')$. Each feature pair ($f$, f') votes for one estimate of the parameter. The final transformation parameter is estimated by locating the maximum of the *consensus function* $H(\theta)$ that accumulates the votes.

Malickas and Vitkus [9] assume that minutiae locations (x, y) and angles (φ) are given for reference and query fingerprints, respectively, as:

$$Q = \{(x_1^K, y_1^K, \varphi_1^K), ..., (x_N^K, y_N^K, \varphi_N^K)\} \text{ and}$$
$$P = \{(x_1^L, y_1^L, \varphi_1^L), ..., (x_M^L, y_M^L, \varphi_M^L)\}.$$

Then, the line K_{ij} between minutiae i and j of reference fingerprint is defined as

$$K_{ij} = (x_i^K, y_i^K, \varphi_i^K, x_j^K, y_j^K, \varphi_j^K, d_{ij}^K, \Phi_{ij}^K, \omega_i^K, \omega_j^K)$$

where the first three fields code minutia i, the second three fields code minutia j, d_{ij}^K is the distance between minutiae i and j, Φ_{ij}^K is the line direction and the last two fields code the angles between the line directions and minutiae directions.

Considering that the same sensor is typically used for capturing reference and query fingerprints, estimating the scaling parameter is not necessary. The rotation angle θ is observable for the line direction Φ_{ij}^K via

$$\Delta\Phi_{K_{ij}L_{kl}} = (\Phi_{ij}^K - \Phi_{kl}^L) \bmod 360 .$$

Using the consensus function, it is possible to obtain an estimate for θ. By rotating the lines from reference fingerprint according to this estimate, K_{ij}^R line set is found. Finally, the translation (Δx , Δy) is observable for minutiae locations via

$$\Delta x_{K_{ij}L_{kl}} = \frac{(x_i^R - x_k^L) + (x_j^R - x_l^L)}{2}$$
$$\Delta y_{K_{ij}L_{kl}} = \frac{(y_i^R - y_k^L) + (y_j^R - y_l^L)}{2}.$$

Using this representation for the proposed fuzzy fingerprint vault, we plan to use quantized location and angles, to account for non-linear distortion and eliminate the necessity to use two line feature sets.

4. Experimental Results

In this section, we assess the requirements of a typical fingerprint-based vault application, in terms of the variability of the fingerprint minutiae data. We used a moderate sized database (denoted as GT henceforth) consisting of 450 mated fingerprint pairs collected in a realistic setting and acquired in multiple sessions. The images were obtained with a DBI optical sensor with 500 dpi resolution. The minutiae in all of the images were identified (their location and angle) by a human expert. Further, the expert determined the minutiae correspondence information between minutiae of every mated pair. We decided to use a database where the features were extracted by a human expert since we did not want the characteristics of an automatic minutiae extractor to affect the statistics we wanted to compute. As a result, the minutiae information that we use is the *ground truth* (hence, the database is named GT).

In the following, we present several statistics that we calculated from this database. Note that these statistics are useful in assessing the applicability of fingerprint minutiae features for any fingerprint-based vault.

Fig. 2 shows the minutiae distributions for three sets: total number of minutiae in the images, number of matching minutiae in the images, and the number of minutiae added-to/missed from the originals. We see that in this database, the average number of minutiae is 40. Note that, the missing and added minutiae may eliminate some possibilities for using minutiae representation as locking keys, since even if all of the translational, rotational, and non-linear distortions in the prints are eliminated, the representations for reference and query will not be the same.

We measured the translational and angular differences (measured in pixels and degrees, respectively) between mated minutiae in all of the fingerprint pairs. Note that no preprocessing (e.g., aligning) of images was done here. We wanted to analyze the difference between minutiae pairs originating from the same finger before carrying out such operations. Note that, inherently, such preprocessing is not applicable to fuzzy vault, as the construct inputs just one feature set, not two that can be compared, aligned, etc. We found that the translational difference can be quite large, with a mean difference of nearly 20 pixels. The maximum difference can be as much as 45 pixels, with a relatively high probability of approximately 0.09. For assessing the magnitude of the necessary alignment, the rigid transformation (optimal in the Least Mean Square sense) between mated fingerprint pairs is estimated using the ground truth information. This yields 2D translation and rotation components of the transform. We found that a translation of nearly 20 pixels and a rotation of nearly 3 degrees are needed, on the average, for aligning (hence effectively eliminating) the cited rigid transform. Then, we measured the translational and angular differences between the minutiae of mated pairs after this alignment. As expected, the translational differences decrease considerably, but there is still a mean difference of nearly 4 pixels. This *residual* difference may create correspondence problems for fingerprint-based vaults, since even the alignment is not able to completely eliminate the variability in minutiae features.

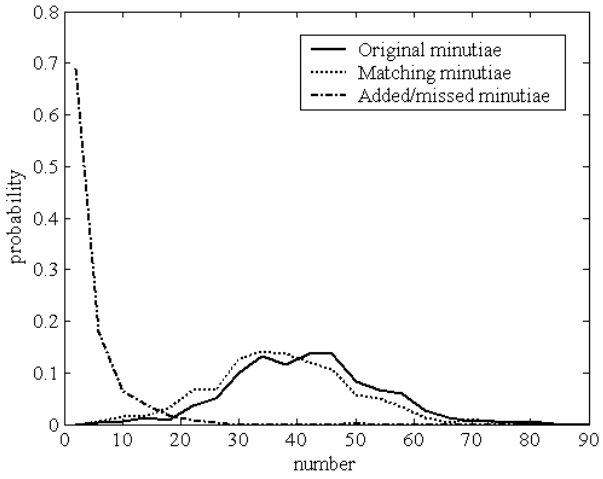


Fig. 2. Minutiae distributions: the curves show the distributions for the number of original minutiae, matching minutiae and added/missed minutiae.

Fig. 3 shows two fingerprint images from the GT database, along with the overlaid lines, obtained via the method given in Section 3. Using the calculated rotation (see Fig. 4) and translation consensus functions, the rotation angle is found to be 2.8 degrees, horizontal translation is found to be 10 pixels and vertical translation is found to be 31 pixels. These values agree closely with the actual parameters. We are currently working on

obtaining performance estimates for the proposed fuzzy fingerprint vault system.

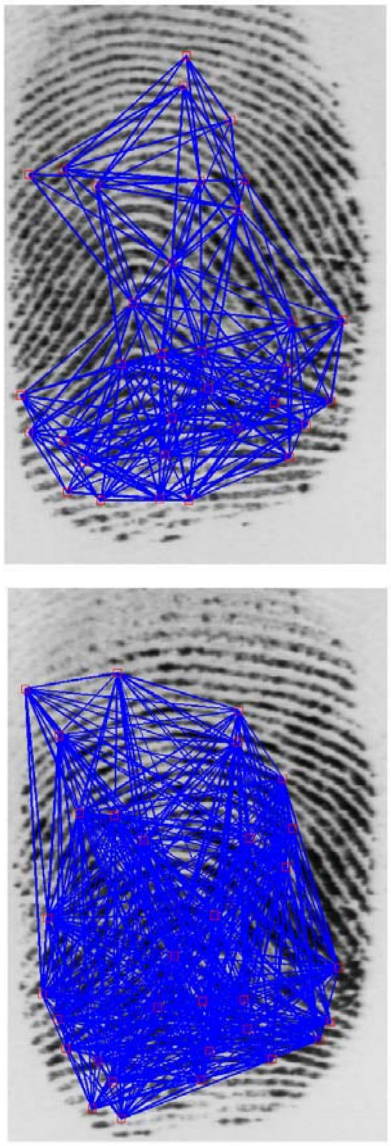


Fig. 3. Fingerprint images with overlaid minutiae lines (top: reference, bottom: query).

5. Conclusions

Based on a new cryptographic construct called fuzzy vault, a fuzzy fingerprint vault system using fingerprint minutiae based lines is proposed. This construct has several characteristics (such as order invariance) that increase its applicability for use with biometric data. Using a fingerprint database marked by a human expert, the variability of minutiae data (before and after alignment) and the alignment parameters are quantified.

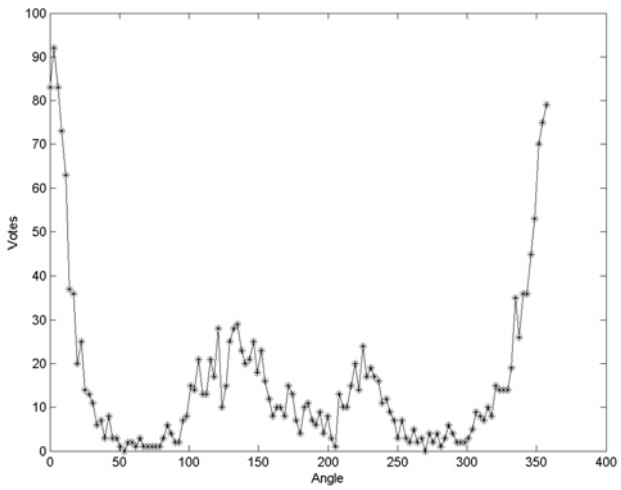


Fig. 4. Rotation consensus function for the fingerprint pair in Fig. 3.

References

[1] W. Stallings, *Cryptography and Network Security: Principles and Practices*, 3. Ed., Prentice Hall, 2003.

[2] NIST, Advanced Encryption Standard (AES), 2001. http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf

[3] A. Jain, R. Bolle, and S. Pankanti, Eds., *Biometrics: Personal Identification in Networked Society*, Kluwer, 1999.

[4] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, 2003.

[5] A. Juels and M. Sudan, "A Fuzzy Vault Scheme", *Proc. IEEE Int'l. Symp. Information Theory*, A. Lapidoth and E. Teletar, Eds., pp. 408, 2002.

[6] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme", In G. Tsudik, Ed., *Sixth ACM Conf. Computer and Comm. Security*, pp. 28-36, 1999.

[7] S. Lin, *An Introduction to Error-Correcting Codes*, Prentice-Hall, 1970.

[8] T. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure Smartcard-Based Fingerprint Authentication", *ACM SIGMM 2003 Multimedia, Biometrics Methods and Applications Workshop*, pp. 45-52, 2003.

[9] A. Malickas and R. Vitkus, "Fingerprint Registration Using Composite Features Consensus", *Informatica, Institute of Mathematics and Informatics (Vilnius)*, vol. 10, no. 4, pp. 389-402, 1999.

[10] C. Shekhar, V. Govindu, and R. Chellappa, "Multisensor Image Registration by Feature Consensus", *Pattern Recognition*, vol. 32, pp. 39-52.

Personal authentication using hand-geometry and palmprint features – the state of the art

N. Pavešić[1)], S. Ribarić[2)], D. Ribarić[3)]
[1)] *Faculty of Electrical Engineering, University of Ljubljana, Slovenia*
[2)] *Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*
[3)] *Intelligent Informatics System, Zagreb, Croatia*
nikola.pavesic@fe.uni-lj.si, slobodan.ribaric@fer.hr, damir.ribaric@zg.hinet.hr

Abstract

In this paper we present an overview of the fundamentals of personal authentication based on hand-geometry measurements and palmprint features. Unimodal and multimodal hand-based systems and technologies are presented, and various levels of fusion for hand-based biometric systems are discussed. Finally, a description of the design and development of a multimodal personal authentication system based on the fusion of hand-geometry and palmprint features at the matching-score level is given.

1. Introduction

Biometrics is a scientific discipline that involves methods of identifying people by their physical and/or behavioural characteristics. The most common physical and behavioural characteristics of an individual used for automatic biometric authentication are as follows: fingerprint, hand-geometry, palmprint, face, iris, retina, DNA, ear, signature, speech, keystroke dynamics, gesture and gait [1], [2] and [3].

A biometric authentication (identification or verification) system uses pattern recognition to automatically recognize an individual on the basis of a measurement of a specific physiological or behavioural characteristic that the individual possesses. An authentication system consists of the following basic modules: sensor, feature generator, matcher, decision and template database; see Fig. 1.

In the feature-generator module the set of discriminatory features is extracted from the raw measurements of an individual biometric characteristic. The matcher module compares the templates (mathematical representations of the features set) against the templates stored in the template database in order to generate matching scores, while the final decision concerning the user's identity (identification) or the user's claimed identity (verification) is taken in the decision module.

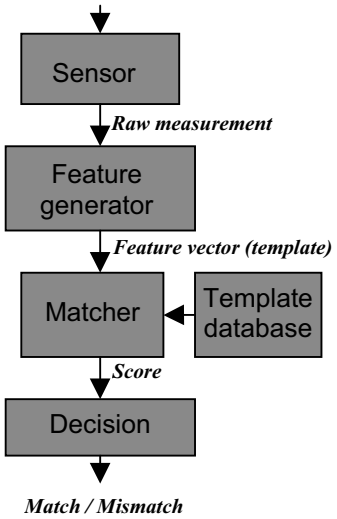


Figure 1. The basic modules of a biometric system.

The most frequently used measures to rate the accuracy of a biometric authentication system are as follows: *false-accept-rate* (FAR), the frequency with which an impostor is falsely accepted; and *false-reject-rate* (FRR), the frequency with which a genuine user is rejected. The error rate at which the FAR equals the FRR, the *equal-error-rate* (EER), is (normally) used as a comparison metric for different biometric systems.

Biometric systems based on a single biometric characteristic are referred to as *unimodal* systems. There are several human and technical factors that influence the performance and operation of a unimodal system, among the most important are the following [2], [3]: *universality*, each person should have the biometric characteristic being acquired; *uniqueness*, there are no two persons that are the same in terms of the biometric characteristic;

permanence, the biometric characteristic should be time-invariant; *collectability of the biometric characteristic*, the biometric characteristic can be measured quantitatively; *performance*, the achievable recognition accuracy, speed and robustness of the biometric system; *acceptability,* to what extent people are willing to accept the biometric system; *circumvention,* how easy it is to fool the biometric system by fraudulent techniques; *scalability*, the feasibility of authenticating people in a large population without unacceptable error rates or throughput times; *maturity of the technology*, the stage of development of the biometric system's technology; and *cost*, an estimation of the total cost to deploy a biometric system. Table 1. presents a comparison of the most common unimodal biometric systems in terms of the above factors.

Table 1. Comparison of the human and technical factors of seven popular unimodal biometric systems. H, M, and L denote high, medium, and low, respectively.

	Fingerprint	Face	Hand geometry	Palmprint	Iris	Voice	Signature
Universality	M	H	M	M	H	M	L
Uniqueness	H	L	M	H	H	L	L
Permanence	H	M	M	M	H	L	L
Collectability	M	H	H	H	M	M	H
Performance	H	L	M	H	H	L	L
Acceptability	M	H	M	M	L	H	H
Circumvention	M	H	M	L	L	H	H
Scalability	H	M	L	H	H	L	H
Maturity	H	M	H	L	M	M	M
Cost	M	L	H	M	H	L	M

Unimodal biometric systems are usually more cost-efficient than multimodal systems. However, a single physical or behavioural characteristic of an individual can sometimes fail to be sufficient for identification. For this reason, multimodal biometric systems, i.e., systems that integrate two or more different biometric characteristics, are being developed to increase the accuracy of decisions and to decrease the possibility of circumventing the system [4]. In general, multimodal biometric systems require integration schemes to fuse the information obtained from the individual biometric modalities. This fusion process can be performed at four different levels: sensor, feature-generation, matching and decision; see Fig. 1.

Generally, a biometric system can be classified according to the method used for capturing and processing the biometric characteristic, i.e., an *on-line* or an *off-line* system. An on-line system captures the biometric characteristics of a person who is physically present at the point of authentication by means of a sensor that is directly connected to a computer for real-time processing, while an off-line system processes previously captured biometric characteristics and the authentication is not performed in real-time.

The biometric system that uses hand-based features is one of the seven leading biometric technologies, and had 11% of the world market in 2004 [4]; see Fig 2. Since the shape of the human hand is not a highly distinctive characteristic, hand-geometry-based systems are used for physical access control, and time and attendance applications. On the other hand, palmprint-based systems, due to the uniqueness and permanence of palmprint features, are typically used in criminal forensic applications.

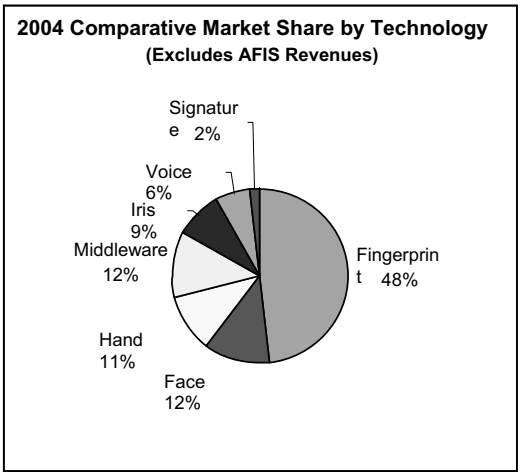
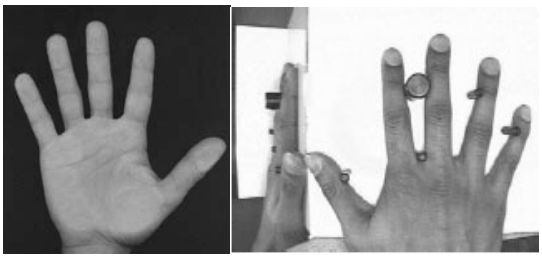


Figure 2. Biometric Market Report (International Biometric Group©) estimates the revenues of various biometrics in 2004 in terms of market share. Note that AFIS are used in forensic applications.

The human hand contains a wide variety of measurable characteristics, e.g., shape, palmprint, fingerprints on the palmar surface of the hand, and veins on the dorsum of the hand, that can be used by biometric systems. Fig. 3 shows typical images of a) the palmar, b) the lateral and c) the dorsal surfaces of the hand.

From these images three classes of biometric features can be extracted: *hand-geometry features,* (e.g., width, thickness and area of the palm, lengths, widths and thickness of fingers), *palmprint features* (e.g., principle

lines, wrinkles, ridges, texture), and *fingerprints features* (e.g., minutiae locations, types, number). These characteristics of the human hand are relatively stable and the hand image from which they are extracted can be acquired relatively easily.



(a) (b) (c)

Figure 3. Typical images of: a) the palmar [25], b) the lateral [10] and c) the dorsal [10] surfaces of the hand.

The image of the palmar surface of the hand is usually acquired by a scanner rated at 180 dots per inch (dpi)/256 grey levels (see Fig. 3 a)) or by a low/medium-resolution CCD camera, located under the transparent platform where the hand is placed. In contrast, the lateral and dorsum surfaces of the hand (see Fig. 3 b) and 3 c)) are captured with a CCD camera placed above the platform with a side-mounted mirror inclined at 45º to the platform. There are usually 4–6 pegs on the platform to guide the placement of the user's hands.

The hand biometrics *shape, palmprint* and *fingerprint* are particularly convenient for fusion because they can be extracted from a single-shot measurement; see Fig. 3 a).

2. Systems based on images of the lateral and dorsal surfaces of the hand

In the literature, several prototypal biometric authentication systems based on extracting a set of hand-geometry features from images of the lateral and dorsal hand surfaces have been proposed.

2.1 Hand-geometry-based systems

Hand-geometry-based authentication systems have been available for more than thirty years. Several companies launched such systems during the 1980s [6], [7] and [8]. With the exception of available information in the form of patents there is no accessible literature referring to research in this area during that period. Some results of recent research and developed prototypes are described below.

Golfarelli et al. [9] described an *on-line* biometric system based on 17 hand-geometrical features, extracted from the image by means of an ad-hoc feature-extraction algorithm. Fig. 4 shows the characteristic points and the geometrical features used in the system.

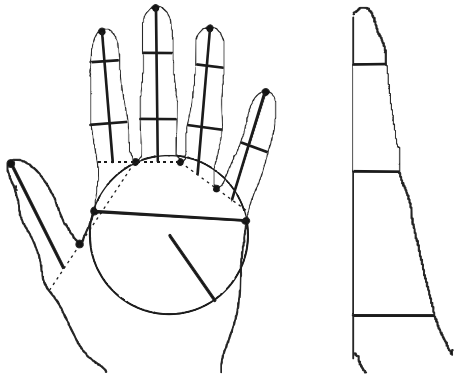


Figure 4. The characteristic points and the 17 geometrical features [9].

100 people took part in a test session where 8 different images of their right hand were taken. With the Bayes classification rule in the matcher module an EER equal to 0.12% was obtained.

Jain et al. [10] described the prototype of an *on-line* verification system based on 16 hand-geometrical features: the length, the width and the thickness of fingers and the widths of the palm (see Fig. 5).

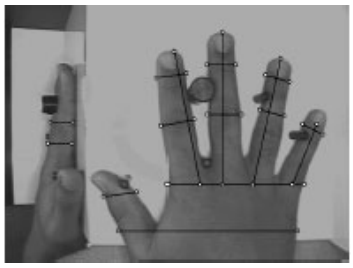


Figure 5. The 16 geometrical features [10].

In the verification phase a 16-dimensional feature vector (stored template) is associated with the claimed identity. This feature vector is then compared with a feature vector of the hand whose identity has to be verified (live template). The system was trained and tested using a database of 50 users. Ten images of each user were captured. Out of 500 images only 360 were used for testing the system (the remaining 140 images were discarded due to incorrect placement of the hand). The performance of the system for 4 different operating points is displayed in Table 2.

Table 2. The results of the system performance testing [2].

FRR	FAR
1 in 33 (3%)	1 in 7 (15%)
1 in 20 (5%)	1 in 10 (10%)
1 in 10 (10%)	1 in 20 (5%)
1 in 3 (30%)	- (0%)

Sanches-Reillo et al. [11] defined and implemented an *on-line* biometric system based on an optimal set of hand-geometry features. After the capturing and pre-processing of the hand images the measurement algorithms are applied. The main distances and angles of the hand are divided into four different categories: width, lengths, deviations, and angles between the inter-finger points. Thirty-one features are extracted, and after applying a discriminatory analysis a feature vector consisting of 25 components is obtained. The feature vectors are the inputs for a comparison process used to determine the identity of the user whose hand has been captured. The nearest-neighbour (1-NN) classifier based on the Euclidean distance d_E and the Hamming distance d_H, the Gaussian Mixture Models (GMM) and the Radial Basis Function (RBF) Neural Networks are used for the classification and verification. The system was trained and tested using a database of 200 images of 20 users. Table 3 summarizes the results of the biometric recognition testing. In the biometric verification testing an EER < 5% is obtained independently of the classification technique and feature vector size used.

Table 3. Percentage of biometric recognition (classification) success compared to the enrolment set size and the feature vector size.

		d_E	d_H	GMM	RBF
No. of enrolment vectors	3	86%	75%	88%	90%
	4	85%	82%	93%	91%
	5	86%	87%	96%	91%
Feature vector size (5 enrol. vectors)	25	86%	87%	96%	91%
	21	84%	86%	97%	95%
	15	86%	88%	96%	89%
	9	77%	75%	91%	82%

Jain et al. [12] presented an authentication method based on the deformable matching of hand shapes. The proposed authentication method is performed in 5 steps: peg removal from the image, contour extraction, finger extraction and alignment, pairwise distance computation and verification (comparison of the Mean Alignment Error (MAE) with a decision threshold T). The system was tested on a database consisting of 353 (2 to 15 images per person) grey-scale hand images (resolution 480 x 485) of 53 people. The best results, i.e., 2% FAR and 3.5% FRR, were obtained for a decision threshold equal to 1.80.

2.2 Finger-geometry-based systems

There are commercial verification systems available that are based on measurements of only 1 or 2 fingers [13]. The single-finger geometry-based biometric system uses only the index finger. This finger pushes a plunger/button, which goes into the device. The rollers, which scan the finger, take measurements of 12 cross-sections of 1½ phalanx of finger.

The two-finger geometry-based biometric system uses a camera-based sensor system to take 3-dimensional measurements of the index and middle finger. From the image a set of the fingers' geometrical features (length, width and thickness of fingers measured on different finger sections) is extracted.

3. System based on the image of the palmar surface of the hand

From an image of the palmar surface of the hand the hand-geometry, the palmprint and the fingerprint features can be extracted. In this paper we will confine ourselves to a description of biometric systems based on hand-geometry and palmprint features.

3.1 Palmprint-based systems

The palm is the inner surface of the hand between the wrist and the fingers. The palmprint is a rich source of information that is useful for personal authentication. The most important features are the three principal lines (the heart line, the head line, and the life line), wrinkles, and ridges. Fig. 6 shows the image of a palmprint and 6 straight black lines that point out the principal lines, the wrinkles and the ridges.

Palmprint-based biometric (*on-line* or *off-line*) systems can be classified according to the applied feature-generation method into systems that extract features in the *original image space* or in the *transformed image space*.

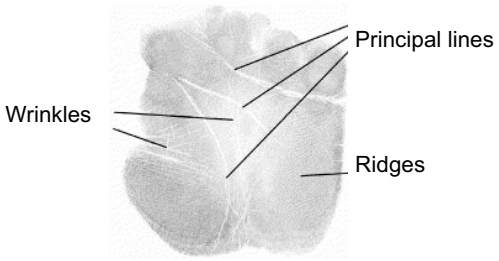


Figure 6. The palmprint image.

3.1.1 Systems based on features extracted in the original image space.

Zhang et al. [14] described an *off-line* palmprint-based verification system based on the end-points and the middle points of the principal lines (datum points) and line-feature matching. For each line segment three features are computed: slope, intercept and angle of inclination. These features, obtained from the lines of two palmprints, are inputs in an invariant line-segment feature-matching process. Two line segments are declared to be the same ***if*** the Euclidian distances between the end-points are less than some threshold D, ***and*** the difference in the angle of inclination is less than some threshold E ***and*** the difference in the intercepts is less than the threshold B, where the thresholds are experimentally determined. The palmprint verification system was tested with 20 pairs of palmprint images from 20 right palms. The experimental results of the identity verification showed the effectiveness of the palmprint verification by obtaining an EER = 0.0% at a decision threshold value of 0.2. The experiments were conducted on 400 x 400 grey-scale inked palmprint images at a resolution of 100 dpi, 256 grey levels.

Duta et al. [15] investigated the feasibility of person identification based on feature points extracted from palmprint images. Their approach is based on a set of feature points extracted from along the principal lines and the associated line orientation. For each palmprint a set of approximately 300 feature points is extracted according to an original algorithm. The decision as to whether two palmprints belong to the same hand is based on computing the matching score between the corresponding sets of feature points of the two palmprints. The matching technique is based on the non-linear deformations of the two sets. The paper palmprints were scanned at a resolution of 200 dpi (image-size 400 x 300 with 256 grey levels). A data set consisting of 30 (15 of each of the two hands) palmprint images of three people was used for experimental purposes. The overlap between the user (genuine) and the impostor distributions is reported to be approximately 5%.

The paper of *P. Ying-Han et al.* [16] introduces an experimental evaluation of the effectiveness of utilizing three well-known orthogonal moments, i.e., Zernike moments, pseudo Zernike moments and Legendre moments, in the application of palmprint verification. These orthogonal moments are able to define statistical and geometrical features containing line-structure information about a palmprint. Experimental results have shown that the performance of the system depends on the moment order as well as on the type of moments. Pseudo Zernike moments of the order of 15 showed the best performance from among all the moments. Its verification rate is 95.75% with FAR = 4.25% and FRR = 4.47% at a decision threshold value of 0.495, which also represents the overall performance of this palmprint verification system. Experiments were conducted using a database consisting of 50 different palmprint classes, with 6 samples for each class.

You et al. [17] proposed a palmprint identification system based on a dynamic selection scheme to facilitate a fast search for the best matching of a palmprint template in the database in a hierarchical fashion. The global texture energy is introduced to guide this dynamic selection of a small set of similar candidates from the database at a coarse level for further matching. At the fine-level identification the same procedure is used as in [14].

C. C. Han et al. [18] described an *on-line* scanner-based personal verification system based on palmprint features. These palmprint features are extracted from the region-of-interest (ROI): the square region in the palm; see Fig. 7.

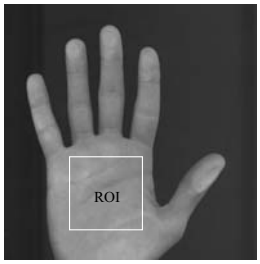


Figure 7. The palmprint region-of-interest.

The multi-resolution feature vectors are derived from the ROI using three different grid sizes (32 x 32, 16 x 16 and 8 x 8). Each component of the feature vector is represented by the mean value of the pixels in the grid element. Two techniques were designed for the identity verification: the multiple-template matching method and the back-propagation neural network method. The hand images of size 845 x 829 in grey-scale format at a resolution of 100 dpi are captured by a scanner. For experimental purposes the 30 hand images of 49

individuals were obtained to construct the database. The best-obtained accuracy rate was 98%.

3.1.2 Systems based on features extracted in the transformed image space.

G. Lu et al. [19] used eigenpalms for palmprint recognition. The images of the palmprints are captured at a resolution of 484 x 384 pixels. They are then aligned and their size is normalized. From these images the palm sub-images with a fixed size (128 x 128 pixels) are extracted and transformed using the Karhunen–Loeve transformation. The extraction of the features involves the proposed eigenspace method with feature-vector lengths of 50, 100, 150 and 200. Classification was performed using a nearest-neighbour classifier based on the weighted Euclidean distance. The test database consists of 191 people, each of whom provided 8 images of their left palm and 8 images of their right palm. Experiments using different numbers of training samples of each person and different lengths of the feature vector are described. The best recognition rate of 99.149% was achieved for 4 training samples and 100 features. Error rates of FAR = 1% and FRR = 0.03% at a decision threshold value of 0.71 were also reported.

W. Li et al. [20] describe a feature-extraction method based on converting a palmprint image from a spatial domain to a frequency domain using a Fourier transform. The features extracted in the frequency domain are used as indexes to the palmprint templates in the database, and the searching process for the best match is conducted in a layered fashion. The experimental results show that palmprint identification based on feature extraction in the frequency domain is effective in terms of accuracy and efficiency. Table 4 shows the results of testing.

Table 4. The results of the palmprint performance testing.

Palmprint images in the database	500
Attempts in testing	2500
Correct answers	2387
Identification rate	95.48%
Average response time (s)	2

The papers of *W. K. Kong et al.* and *D. Zhang et al.* [21], [22] describe an *on-line* palmprint-based system based on the use of an adjusted 2-D Gabor filter to obtain texture information from the palmprint, and a comparison of two 2049-dimensional texture feature vectors using the normalized Hamming distance. In terms of accuracy the best results are obtained using the following filter parameters: orientation equal to $\pi/4$, frequency of the sinusoidal wave equal to 0.0916, and the standard deviation of the Gaussian envelope equal to 5.6179. On a palmprint-images database containing 7752 images collected from 193 individuals, a system EER of 0.6% was reported.

In *X. Wu et al.* [23] an *off-line* palmprint-based verification system based on the use of Fisher's linear discriminant (FLD) was described. The palmprint image is projected from the high-dimensional original palm space to the low-dimensional Fisherpalms space. A database with 3000 palmprint images from 300 different palms was used for testing purposes. For the palmprint region-of-interest with resolution 128 x 128, 64 x 64 and 32 x 32, the feature vector of each testing palmprint is matched against each stored template at each resolution. A nearest-neighbour classifier based on the Euclidean distance is used. The EERs at the 128 x 128, 64 x 64 and 32 x 32 resolutions are 1.00%, 0.95% and 0.82%, respectively.

3.2 Systems based on the fusion of palmprint and hand-geometry features

Shu et al. [24] presented a prototype of an *off-line* system based on the following palmprint features: geometrical features (width, length and area of the palm), principal-line, wrinkle and delta-point features, and minutiae. Principal-line and wrinkle features obtained from a low-resolution image (100 dpi), and delta-point features and minutiae features extracted from a high-resolution image (400 dpi), are fused at the sensor level. The authors evaluated the FAR and FRR for different combinations of palmprint features. Their experiments showed that the combination of eight points on the principal-lines and palm-geometry features gives an acceptable identification accuracy, i.e., FAR = 0.2% and FRR = 0.0%, at the decision threshold T_0. All experiments were carried out on a database containing 48 pairs of prints of the same palm and 844 pairs of prints of different palms.

The paper "A Biometric Identification System based on the Fusion of Hand and Palm Features" by *S. Ribarić et al.* [25] describes the design and development of a prototype system for the on-line identification of an individual based on the fusion of palm features, finger- and palm-geometry parameters. Information fusion at the matching-score level, where the three matchers are combined, is discussed. After training with the template files of 50 people, the system was tested with the template files of 61 people not "seen" during the training phase. The test performance of the system based on the fusion of palmprint features, finger geometry and palm geometry was reported to be FAR = 0.0% and FRR = 1.7%. The FAR and FRR, where the *total-error-rate* TER (TER = FAR + FRR) achieves a minimum, are displayed

in Table 5 for different unimodal and multimodal systems.

Table 5. Performance scores for minimum total error rate for unimodal systems: FG finger-geometry features; PG palm-geometry features; P palmprint features, and on the systems based on the fusion of these features (e.g., FG-P denotes the fusion of finger-geometry (FG) and palmprint (P) features).

	FAR	FRR
FG	0%	5.2%
PG	32.6%	27.7%
P	8.1%	6.1%
FG-PG	0%	4.6%
PG-P	2.6%	2.3%
FG-P	0%	1.8%
FG-PG-P	0%	1.7%

The design and implementation details, as well as the experimental testing of an improved version of the system are given in Section 4.

The paper by *A. Kumar et al.* [26] describes improvements to the performance of an *on-line* palmprint-based verification system by integrating hand-geometry features. A digital camera (1280 x 960 pixels) is used to acquire the hand images. The acquisition setup is inherently simple and does not use any pegs, but users were requested to make sure that their fingers do not touch each other and that most of their hand's back side touches the table. After image pre-processing, extraction of the palmprint image and its normalization, the palmprint feature vector is represented by standard deviations in the *n* overlapping blocks within the palmprint ROI. Additionally, a total of 16 hand-geometry features are used: 4 finger lengths, 8 finger widths, palm width, palm length, hand area, and hand length. These two feature vectors are used for information fusion at the feature-generation level, as well as at the decision level. Experimental results on the image dataset of 100 users are displayed in Table 6.

Table 6. Performance scores for minimum total error rate on 472 test images. FGL and FDL denote fusion at feature generation level and fusion at decision level, respectively.

	FAR	FRR	Threshold
Palmprint	4.49%	2.04%	0.9830
Hand-geometry	5.29%	8.34%	0.9314
FGL	5.08%	2.25%	0.9869
FDL	0%	1.41%	0.9840

Jain et al. [27] describes realization of a multimodal biometric verification system based on face, fingerprint and hand-geometry features that uses fusion at the matching-score level based on learning user-specific matching thresholds as well as the weights of an individual biometric characteristic.

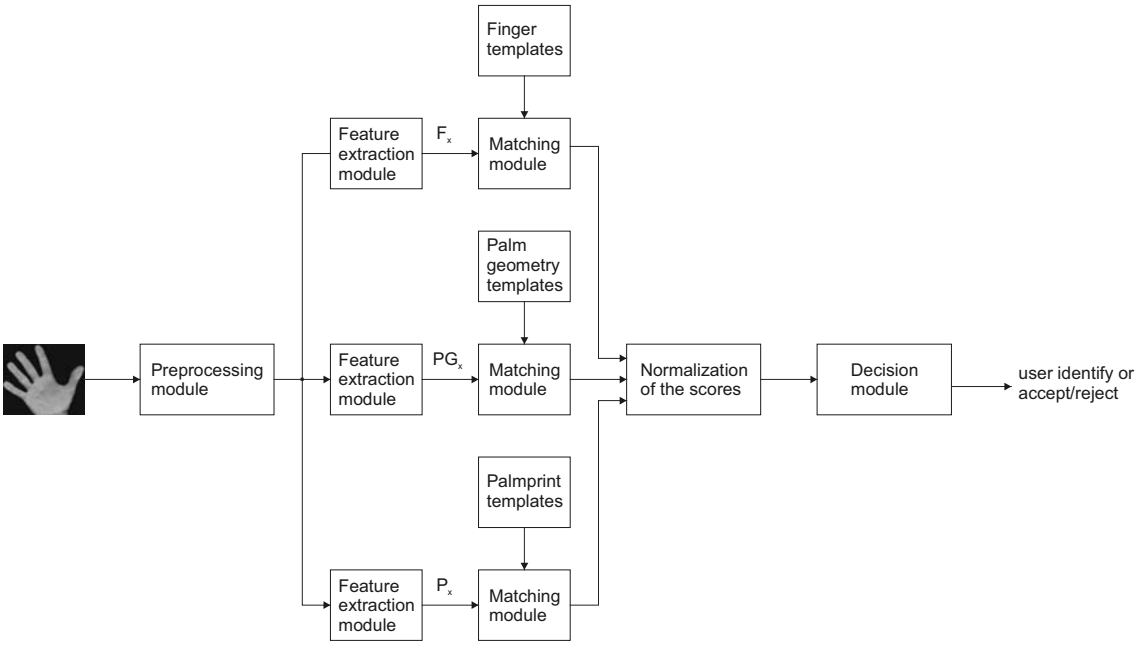


Figure 8. A block-diagram of the multimodal biometric identification system based on the fusion of finger-, palm-geometry and palmprint features at the matching-score level.

4. Design of a three-modal hand-based system

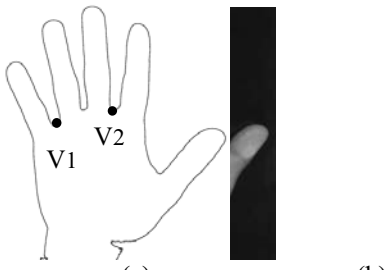
In this section the prototype of an on-line hand-based multimodal authentication system [28] is described. The system is based on the fusion of finger- and palm-geometry features, as well as palmprint features at the matching-score level.

4.1 System description

The block diagram of the proposed system is shown in Fig. 8.

A desktop scanner (180 dpi, 256 grey levels) is used to acquire the hand images. The user is asked to put his/her hand on the flat glass surface of the scanner, with the fingers spread naturally. There are no pegs for controlling the placement of the hand, and there are no requirements for any additional illumination. All the biometric features, i.e., finger-, palm-geometry and palmprint features, are obtained from the hand image in a single shot. Fig. 9 a) shows a typical image obtained by the input device.

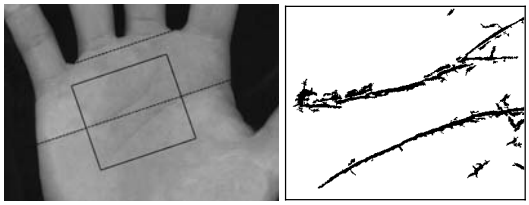
During the pre-processing phase, by applying thresholding the hand is extracted from the image background. Due to the regular and light-controllable conditions of the image-capturing, global thresholding provides satisfactory results. By using a modified contour-tracking algorithm [29], the contour of the hand is extracted (see Fig. 9 b)).



(a) (b)

Figure 9. a) Example of an image of the right hand obtained with a desktop scanner; b) Extracted contour of the hand showing the two reference points (V1 and V2).

From the hand contour two stable points are determined (V1 and V2); see Fig 9 b). Point V1 is used to determine the sub-region (120 x 60 pixels) of the palmprint where a segment of the heart line can be detected. The third stable point is defined as a point on the heart line. It is detected by subsequently applied operations as follows: convolution with a Gaussian mask, Sobel operator, hysteresis thresholding and horizontal projection. These points (V1, V2 and the point on the heart line) are used to define the palmprint's region-of-interest (ROI) with dimensions 315 x 285 pixels. Applying the same sequence of operations the principal lines in the ROI are detected (see Figs. 10 a) and 10 b)).

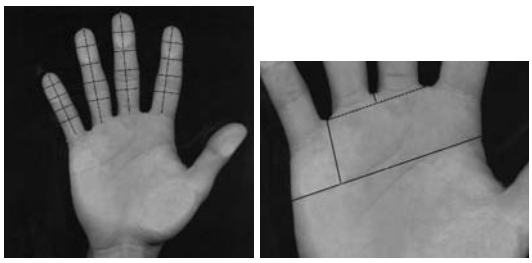


(a) (b)

Figure 10. a) The palm region-of-interest (ROI); b) Pre-processed ROI.

In the feature-extraction modules the three feature vectors F_x, PG_x and P_x are obtained. The 20-component vector F_x contains the features of the four fingers (the lengths and four widths of each finger measured at different heights); see Fig. 11 a).

The 4-component vector PG_x carries information about simple palm-geometry (the width of the palm, the distance between points V1 and V2, and the two distances between the line segments; see Fig. 11 b)).



(a) (b)

Figure 11. a) Finger-geometry features; b) Palm-geometry features.

The 399-component vector P_x relies on palmprint attributes, i.e., on the palmprint's principal lines and texture. The matching between the live-template (represented by 3 feature vectors: F_x, PG_x and P_x) and the user template (F_u, PG_u and P_u) stored in the database is based on a computation of the Euclidian distances: $d(F_x, F_u)$, $d(PG_x, PG_u)$ and $d(P_x, P_u)$. These distances are normalized and transformed into similarities S^F_{xu}, S^{PG}_{xu} and S^P_{xu}. These transformations are performed by three transition functions that are determined experimentally during the learning phase of the system.

The fusion is performed in the decision module, where the total similarity measure $TSM_{xu} = w_1 S^F_{xu} + w_2 S^{PG}_{xu} + w_3 S^P_{xu}$ is computed. The values of the weights are proportional to the three-unimodal system performances. The final decision as to whether the live template matches with the user template is based on an additional rule. This rule requires that the similarity between the live and the user templates has to exceed some decision threshold value. This value is determined during the system validation phase.

In the case that the system is used for identification purposes, then the additional rule (k, l)-NN; (k = l = 3) has to be applied in the decision module.

4.2 Experiment and results

The verification experiments were done on a database consisting of two parts: a user database and an impostor database. The user database consisted of the images of 110 people, with 7 hand images per person. Three of these 7 images were used in the enrolment stage, to create the user database, and the remaining 4 were used for testing. The impostor database consisted of 399 images of 57 people (7 images per person). This setup makes possible 440 (110 x 4) user experiments and 399 impostor experiments. The impostor database was also used as a training database in the template-generation process.

The verification experiments were done as follows: After entering the user PIN code and capturing the user hand image, the system calculates the total similarity measure (TSM) between the live-template and all the corresponding templates in the user database (there are 3 user templates). In the next step, the final decision (the user is accepted or rejected) is based on thresholding with the decision threshold T:

if

$$\min_i \{TSM_i\} \geq T; i = 1,2,3$$

then the user represented by the live template is accepted as a user registered in the database with the corresponding PIN code.

otherwise, the person represented by the live template is rejected as an impostor.

The results, expressed in terms of FRR and FAR, vary depending on the decision threshold value T. Fig. 12 presents the verification test results and shows the dependency of the FAR and the FRR on the threshold value.

From Fig. 12 it can be seen that the described verification system achieves an EER equal to 0.41% at the decision threshold T = 0.814, and a minimum total error rate equal to 0.75% is achieved at T = 0.82. It achieves FAR = 0%, FRR = 0.68% at T = 0.86 and FRR = 0%, FAR = 1.18% at T = 0.78.

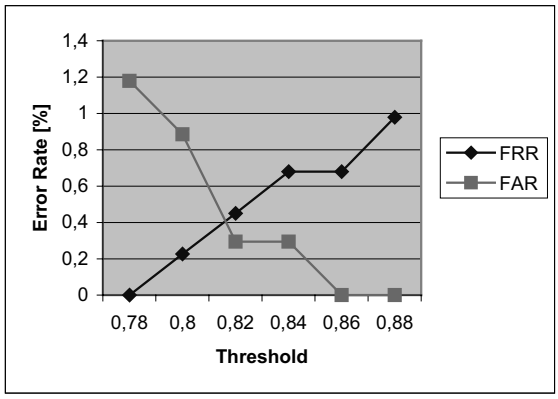


Figure 12. Verification test results, the dependence of the FAR and the FRR on the threshold value.

5. Summary

This paper presents an overview of biometric hand-based systems. Systems based on images of the lateral and dorsal surfaces of the hand, as well as systems based on the image of the palmar surface of the hand are described.

In spite of the fact that hand-geometry features (finger length, width, thickness, curvatures and the relative location of these features) are not unique, hand-geometry-based biometric systems are attractive for a number of reasons. Hand-geometry measurements are easily collectible and non-intrusive, template generation is fairly simple, template size does not exceed 100 bytes, and a stand-alone system is easy to build [2].

The finger-geometry-based biometric systems are smaller than those based on complete hand-geometry parameters, they also tend to be low cost and user friendly (no “criminality” feeling). The main drawbacks of such systems are the low uniqueness and accuracy, as well as the low scalability.

Palmprint authentication is one of the relatively new biometric technologies. Palmprint-based biometric systems, due to the uniqueness and permanence of the palmprint features, are considered as highly accurate and scalable systems.

Multimodal biometric systems increase the performance as well as the scalability, and they are generally more robust to fraudulent technologies. Multimodal authentication systems based on the

integration of hand-geometry and palmprint features at the different fusion levels, are described in the paper. The implementation details of a system based on the fusion of finger-geometry, palm-geometry and palmprint features at the matching-score level are given.

For still higher accuracy and scalability, we propose the development of a multimodal system that integrates features extractable from the palmar surface images, such as finger-geometry, palm-geometry, palmprint and fingerprint features. Such a system could also be interesting because all these features can be obtained from just one high-resolution single-shot image.

6. References

[1] D. Zhang, Automated Biometrics: Technologies & Systems, Kluwer Academic Publishers, USA, 2000.

[2] R.M. Bolle, J.H. Connell, S. Pankati, N.K. Ratha and A.W. Senior, Guide to Biometrics, Springer-Verlag, 2003.

[3] A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition", IEEE Tr. on CSVT, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, 2004, pp. 4-20.

[4]International Biometric Group, http://www.biometricgroup.com/reports/

[5] S. Prabhakar, S. Pankanti and A.K. Jain, "Biometric Recognition: Security and Privacy Concerns", IEEE Security and Privacy Magazine, Vol. 1, No. 2, 2003, pp. 33-42.

[6] Recognition Systems Inc, http://www.recogsys.com.

[7] B. Spencer, "Biometrics in Physical Access Control Issues", Status and Trends, http://www.recogsys.com

[8] R. Zunkel, "Hand Geometry-based Verification", in [1], pp. 87-101.

[9] M. Golfarelli, D. Maio and D. Maltoni, "On the Error-Reject Trade-Off in Biometric Verification Systems", IEEE Tr. on PAMI, Vol. 19, No. 7, 1997, pp.786-796.

[10] A. K. Jain, A. Ross and S. Pankanti, "A Prototype Hand Geometry-based Verification System", 2nd Int. Conference on Audio- and Video-based Personal Authentication (AVBPA), Washington, March 1999, pp. 166-171.

[11] R. Sanchez-Reillo, C. Sanchez-Avila and A. Gonzalez-Marcos, "Biometric Identification Through Hand Geometry Measurements", IEEE Tr. on PAMI, Vol. 22, No. 10, 2000, pp. 1168-1171.

[12] A. K. Jain and N. Duta, "Deformable Matching of Hand Shapes for Verification", Proceedings of IEEE International Conference on Image Processing, Kobe, October 1999, 5 pages.

[13] BioMet Partners Inc., http://www.biomet.ch/.

[14] D. Zhang and W. Shu, Two Novel Characteristics in Palmprint Verification: Datum Point Invariance and Line Feature Matching, Pattern Recognition 32, 1999, pp.691-702.

[15] N. Duta, A. K. Jain and K. V. Mardia, Matching of Palmprints, Pattern Recognition Letters, Vol. 23, No. 4, 2002, pp. 477-485.

[16] P. Ying-Han, T.B.J. Andrew, N.C.L. David and F.S. Hiew, "Palmprint Verification with Moments", Journal of WSCG, Vol.12, No.1-3, WSCG'2004, Feb 2-6, 2003.

[17] J. You, W. Li and D. Zhang, "Hierarchical Palmprint Identification via Multiple Feature Extraction", Pattern Recognition 35, 2002, pp. 847-859.

[18] C-C. Han, H-L. Cheng, C-L. Lin and K-C. Fan, "Personal Authentication Using Palm-print Features", Pattern Recognition 36, 2003, pp. 371-381.

[19] G. Lu, D. Zhang and K. Wang, "Palmprint Recognition using Eigenpalms features", Pattern Recognition Letters 24, 2003, pp. 1463-1467.

[20] W. Li, D. Zhang and Z. Xu, "Palmprint Identification by Fourier Transform", IJPRAI, Vol. 16, No. 4 (2002), pp. 417-432.

[21] W. K. Kong, D. Zhang and W. Li, "Palmprint Feature Extraction using 2-D Gabor Filters", Pattern Recognition 36, 2003, pp. 2339-2347.

[22] D. Zhang, W. K. Kong, J. You and M. Wong, "Online Palmprint Identification", IEEE Tr. on PAMI, Vol. 25, No. 9, 2003, pp. 1041-1050.

[23] X. Wu, D. Zhang and K. Wang, Fisherpalms-based Palmprint Recognition, Pattern Recognition Letters 24, 2003, pp. 2829-2838.

[24] W. Shu and D. Zhang, "Automated Personal Identification by Palmprint", Opt.Eng. 37(8), 1998, pp. 2359-2362.

[25] S.Ribarić, D. Ribarić and N. Pavešić, "A biometric identification system based on the fusion of hand and palm features", Proceedings of the advent of biometrics on the Internet, Rome, Italy, 7-8 Nov. 2002, pp. 79-82.

[26] A. Kumar, D. C. M. Wong, H. C. Shen and A. K. Jain, "Personal Verification Using Palmprint and Hand Geometry Biometric", *Proc. of 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Guildford, UK, June 9-11, 2003, pp. 668-678.

[27] A. K. Jain and A. Ross, " Multibiometric Systems", Communications of the ACM, Special Issue on Multimodal Interfaces , Vol. 47, No. 1, 2004, pp. 34-40.

[28] S.Ribarić, D. Ribarić, N. Pavešić, "Multimodal biometric user-identification system for network-based applications", IEE Proc. Vis. Image Signal Process., Vol. 150, No. 6, 2003, pp. 409-416.

[29] T. Pavlidis, Algorithms for Graphics and Image Processing, Springer Verlag, 1982

Cross Cultural Aspects of Biometrics*

S. Schimke[(1)], C. Vielhauer[(1)], P. K. Dutta[(2)], T. K. Basu[(2)], A. De Rosa[(3)], J. Hansen[(4)], J. Dittmann[(1)], B. Yegnanarayana[(5)]

[(1)] Otto-von-Guericke-University of Magdeburg Germany

[(2)] Indian Institute of Technology Kangpur India

[(3)] University Florence Italy

[(4)] HTTC Darmstadt Germany

[(5)] Indian Institute of Technology Chennai India

Abstract

In this paper we summarize our first research results in the field of Cross Culture user authentication. We will investigate intercultural aspects of biometrics, both of technical and legal nature. Besides biometric based user authentication, Human-to-Computer interfaces are an important part of our work.

We present a methodology for intercultural and multimodal data recording and testing of different hypotheses.

1. Motivation

The goal of our work is to analyze multicultural aspects of biometric speech and writing data input. We will analyze data input for either natural human-to-computer interfaces or biometric authentication purposes. As [1] shows, it is possible to estimate some meta-data like script language, origin, gender and age by statistically analyzing human handwriting. By knowing this meta-data, it appears to be possible to adapt the recognition or authentication algorithms in order to enhance their performance/quality (i.e. False-Match/False-Non-Match Rates, FMR/FNMR).

One goal of our research is to show, that speech or handwritten input is of different suitability for biometric authentication and recognition in **different countries and/or in different languages**. It is also interesting to perform the task of user authentication in bilingual or multilingual environment, which may have special relevance to tracking a particular target user under changing situations. Also, using speech input in addition to handwriting opens the potential to build multimodal environments for a more natural and intuitive handling of computer systems.

Another important aspect of our work is the analysis of **user acceptance** of speech and handwriting modalities for interface or authentication usages. For example handwritten signature verification appears to have some advantages over other biometric modalities in European countries, where it is a traditionally well-established method for manual user authentication. However, the **social or legal perception** of the signature might be different in other cultural or linguistic groups. Our idea is to accomplish an initial survey on the user perception in three different countries (India, Italy and Germany) together with the technical evaluation of speech and handwriting biometrics. In our paper we present our first results with respect to evaluation aspects with focus on privacy and cross cultural issues (section 2). Furthermore we introduce our test methodology and evaluations strategies (section 3). The paper finalizes with a conclusion (section 4).

2. Evaluation Aspects

In this section we briefly discuss privacy issues in the context of collection of biometric and personal data. Other points of discussion are cross cultural issues of user interfaces.

2.1. Privacy Issues

User privacy awareness, i.e. to know, when personal data are taken and for which purpose they are used, is a crucial component for trust in the information society. Without clarity and trust in this area, members of the information society could be scared to be at the mercy of unsearchable surveillance technology. This consideration leads us to improve data control.

To know, that data can move freely and can be a permanent part of scientific progress could also be a crucial component for trust in the information society. If personal data are necessary for scientific progress, they should be explicitly protected.

* This publication has been produced with the assistance of the European Union (project CultureTech, see http://amsl-smb.cs.uni-magdeburg.de/culturetech/). The content of this publication is the sole responsibility of the University Magdeburg and their co-authors and can in no way be taken to reflect the views of the European Union.

Since there are concerns about privacy with antagonizing aspects with respect to gaining and processing biometric data, we will have to take care of that issue. We will analyse traffic restrictions, which are designed to cover privacy and we will measure the range of exemptions for scientific purposes as well.

Some countries have established privacy laws to regulate the handling of personal data. Other countries are within an ongoing legislation process to establish data protection rules. In the project, one area of research is to summarize information about that kind of law in the different countries. This information has to be kept in mind while gaining and processing biometric data in our project. We have the goal to use law more for building a bridge over the gap than for widening the gap.

This information is interesting for itself, since it can be of use in other research and commercial projects, which also have implications to processing of personal data.

2.2. Cross Cultural Issues

Additionally to legal issues, there can be varying user perceptions about handling of biometric data. Therefore we will survey, accompanying the technical evaluations, cultural aspects in order to get information about such perceptions. The goal is to develop a mapping of legal and social concerns in the different regions. Is there an interplay of the social, ethical, and existential orientation on the one hand and specific codes of perception on the other hand if biometric data are part of an interaction? An approach to evaluate social and ethnical perception can be based on an online survey and subsequent statistical analysis of the poll data. For the legal issues we analyse the prevailing case law of the three countries.

Also, we want to analyse, if there are differences in power of authentication with multimodal biometric data. For example, we will evaluate the hypothesis, that handwritten scripts or spoken text can lead to different security levels, depending on the language and script.

3. Methodology

In this section we present the technical concept and the metadata, which we will acquire (3.1), present the test plan (3.2) and discuss our evaluation strategy (3.3).

3.1. Technical Concept

Our software system for recording and evaluating speech and handwriting data is based on a generic system design introduced in [2], extended by audio capability and additional metadata models. Fig.1 presents our design architecture, which consists of the following components:

- **Data Recorder** module: implements the A/D conversion from the audio and handwriting sampling devices. For the sampling, we use tablet PC hardware, equipped with active pen-based (WinTab compatible) digitizer hardware and on-board audio device.
- **Evaluation Database**: stores the complete audio and handwriting signals along with synchronized **metadata**
- **Test Controller** may reproduce user inputs in batch mode process. The operational sequence of batch runs is defined by **Test Profiles**, which feed reproduced signals from the Evaluation Database to plug-in **Algorithms** to be evaluated and protocol their results to the test log.

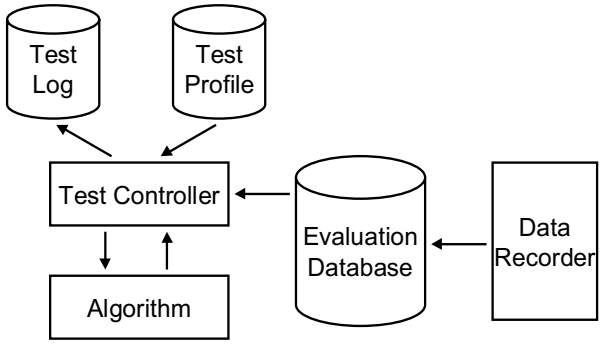


Figure 1 – Model of our evaluation system. The Data recorder collects digitized representations of handwriting signals x(t), y(t) (horizontal and vertical movement signal), p(t) (pen pressure signal), θ(t) and Φ(t) (pen incline signal) during the online handwriting process using a tablet digitizer and audio data during a speech session using a microphone. Signals resulting from the sampling processes are stored to the evaluation database. Based on these samples, the Test Controller may execute user verification and other Algorithms, using predefined, stored Test Profiles. Test results are protocolled to the Test Log.

The following metadata categories are requested and stored within the system. For the sake of standardization, we use ISO norms to describe names of countries, languages and scripts.

- **Person related meta data, acquired to the test subjects** [1][3]:
  - Gender (female or male),
  - Age,
  - Handedness (right or left),
  - Ethnicity (white, black, hispanic, asian, …),
  - Religion,
  - Highest level of education,
  - Place of birth (ISO-3166 [5]),
  - Place of birth of parents (ISO-3166),
  - Place of schooling (ISO-3166)
  - Native language (ISO-639 [6]),

- Known other languages (ISO-639),
- Native script (ISO-15924 [7]),
- Known other scripts (ISO-15924).

- **Process related meta data**:
  - Digitizer device (what kind of handwriting device, microphone, soundcard, other audio hardware, e.g. telephone [4]),
  - Environment (silent audio cabin, noisy laboratory, open air w/o traffic noises),
  - Semantic/type of input (see table 1) and content of input, if not predefined,
  - Used language/script,
  - Block letters or cursive script,
  - Date and time of day.

Input	Style	Sp	Wr
Decimal numbers. (0 – 9)	B	x	x
Latin alphabet	B	x	x
Answer: "What is your good name?"	B/C	x	x
Answer: "Where are you from?"	B/C	x	x
Answer: "How old are you?"	B/C	x	x
Say/write: "Minimum"	B/C	x	x
Say/write: "Maximum"	B/C	x	x
Say/write: "Pay the man first please."	B/C	x	x
Your signature	C		x
A pseudonym	B/C		x
The PIN number "8710"	B/C	x	x
A free chosen pass phrase	B/C	x	x
A free chosen symbol			x

Table 1 – Example types of inputs for English speech and handwriting modality. *Style* is the writing style: *B* for block letters and *C* for cursive script. *Sp* stands for speech input and *Wr* for handwriting input. Find complete list of input types for all languages (English, German, Italian, Indian dialects) in [8].

3.2 Test Plan

We define a **test module** as a set of recordings (speech or handwritten) of one person at one date in one language. The set of recordings consists of different types of input: a) simple questions to answer, words to say or write and phrases to repeat (see table 1 for some English examples), b) continuous text, as shown in extracts in figure 2. The detailed number of recordings in a test module is given in the test list in [8]; typically, for basic input types such as those shown in table 1, we request ten sample instances and for more extensive texts like in figure 2, we ask for one instance within one module. A **test session** is a set of test modules of one person at one date. In our scenario, a test session of one person consists of at least four test modules; handwritten as well as speech input for the native language(s) and for English language. A **test series** is a sequence of test sessions of one person on five days while a duration not longer than a month.

Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. […]

Figure 2 – Excerpt of a test sample text with 330 words overall. The complete sample text can be found in [8].

In each location in Germany, Italy and India, at least ten persons (if possible half of them female) will perform a test series. Each recorded sample of the test modules gets annotated with metadata as described in 3.1.

3.3. Evaluation Concept

A goal of our work is to test different hypotheses, regarding multi cultural aspects of biometric authentication and user interfaces. One hypothesis is that there are differences in speech and handwriting recognition and biometric user authentication results, depending on used language and script, as well as depending on origin of English (as a foreign language for majority of the test subjects) speaking or writing person.

Apart from cultural aspects, we will investigate influence of other person related metadata (see list in 3.1), such as gender or age, on results of authentication and recognition. In [1] Tomai et al state a power of handwritten characters to discriminate persons, belonging to groups of such metadata. For example they correctly recognize a person to be female or male with a probability of 70%. We will try to verify these results and hopefully find other discriminatory features of speech and handwriting. A hypothesis is that it is possible to recognize the origin of an English speaking and/or writing individual on the basis of their manner to speak and/or write.

Beyond this aspect, we will investigate possibilities of fusion of handwriting and speech modalities for estimation of metadata.

4. Conclusions

We have introduced a new approach to include metadata into user authentication systems to evaluate cross cultural impact on biometric authentication processes as well as textual recognition quality.

Creating a database of handwriting and speech test samples from persons with different cultural backgrounds, annotated with valuable metadata, opens the possibility to investigate differences between these dif-

ferent cultures and to fine-tune recognition and authentication algorithms and enhance them, that way.

The novelty of our work is to capture multimodal sample data from persons of different culture groups and to annotate them at the same time with a substantial set of metadata. This opens the possibility for further research activities in the area of inter cultural and multimodal user interfaces and biometric authentication.

References

[1] C. I. Tomai, D. M. Kshirsagar, and S. N. Srihari, "Group Discriminatory Power of Handwritten Characters", *Proceedings of SPIE-IS&T Electronic Imaging 2004*, pp. 116-123.

[2] C. Vielhauer, "Handwriting Biometrics for User Authentication: Security Advances in Context of Digitizer Characteristics", PhD Thesis, submitted to Technical University Darmstadt, Germany, 2004

[3] F. Zöbisch, C. Vielhauer, "A test tool to support brut-force online and offline signature forgery tests on mobile devices", *Proceedings of IEEE ICME 2003*, pp. 60–64.

[4] Richard Norton, "The evolving biometric marketplace to 2006", *Biometric Technology Today*, Oct. 2002, pp 7-8.

[5] ISO 3166 – English country names and code elements. http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html

[6] ISO 639 – Code for the representation of names of languages. http://www.iso.org/iso/en/prods-services/popstds/languagecodes.html

[7] ISO 15924 – Codes for the representation of names of scripts. http://www.unicode.org/iso15924/

[8] Technical Worksheet: List of speech and handwriting input types. http://amsl-smb.cs.uni-magdeburg.de/culturetech/inputs.pdf

Maximum Discrimination Analysis (MDA) as a Means for Dimension Reduction in Biometric Verification

Raymond Veldhuis, Asker Bazen
University of Twente
Faculty EEMCS
P.O. Box 217, 7500 AE Enschede, The Netherlands
R.N.J.Veldhuis@utwente.nl

Abstract

We propose the discrimination distance between the probability densities of genuine feature vector and the entire observation space as an objective function for dimension reduction. This leads to a new method for dimension reduction, called maximum discrimination analysis. *It is demonstrated with synthetic and real data that it has a better verification performance than linear discriminant analysis.*

1 Introduction

In biometric verification systems, it is often desirable to reduce the dimension of the feature vector prior to verification. One reason is that with the reduction of the dimension of the feature vector, also the complexity of the verification system is reduced. The other reason is that, if the parameters of the verification system are estimated from training data, a dimension reduction may improve the verification performance. Often the verification performance measured on test data increases first with the dimensionality of the feature vector, but after a certain point it decreases [4]. This is known as the Hughes phenomenon, [5] or as overtraining or overfitting. In an extreme case, it may happen that, given the dimensionality of the feature vector, there are not enough examples in the training data to estimate the parameters of verifier. This is known as the small-sample-size problem. For a likelihood-ratio-based verifier for Gaussian data, for instance, the number of examples must be greater than the dimension of the feature vector in order to be able to estimate non-singular covariance matrices.

Various methods of dimension reduction have been described in the literature. An overview is presented in [9]. The most basic methods are based on principal-component analysis (PCA) and linear-discriminant analysis (LDA). Both PCA and LDA are realized by linear orthogonal transforms, which project the feature vector on a subspace. PCA determines an orthonormal basis for the entire observation space. Each basis vector is called a *mode*. The first mode is found as the one-dimensional subspace with the highest variance. The second mode is the found as the one-dimensional subspace orthogonal to the first, with the highest remaining variance, and so on. Dimension reduction based on PCA comes down to retaining only those modes that contribute significantly to the variance. LDA also determines an orthonormal basis for the observation space, but now the subsequent modes are found as one-dimensional subspaces with the highest ratio of within-class to between-class variances. It is assumed that such a high ratio is equivalent to a good discrimination between classes. See [12] and [2] for examples of an applications of, respectively, PCA and LDA in face recognition. Variations of LDA have been proposed in e.g. [7, 11].

Dimension reduction based on LDA is optimal, if a single transform is used for all classes and if the performance is averaged over all classes. Because LDA ignores the means of the classes, a better verification performance can be obtained with a class-dependent transform. This is also recognized in [4], where it is suggested to use a transform that optimizes a performance criterion such as the divergence or the Bhattacharyya distance. Examples of such approaches can be found in [9, 11, 10, 8].

We consider likelihood-ratio-based verification for Gaussian data. For this case we present a new method for dimension reduction, called *Maximum Discrimination Analysis* (MDA), which is based on the discrimination [3], or Kullback-Leibler [6], distance. The discrimination distance quantifies the difference between two probability densities. In biometric verification these are the probability density of the genuine feature vector and that of the entire observation space. Note that the entire observation space is also the impostor feature-vector space. I.e. ‘anyone can

be an impostor'. We illustrate that the relation between the logarithm of the Equal-Error Rate (EER) and the discrimination distance is by, good approximation, linearly decreasing. This demonstrates that the discrimination distance is a good measure of performance for a verification system and, therefore, a reasonable objective function for dimension reduction. MDA is a method to find an orthogonal transformation that is capable of reducing the dimension of the feature space while keeping the discrimination distance maximal. A numerical procedure to obtain the MDA transform from the parameters of the probability densities is presented. This procedure is simple and converges rapidly. We show that the verification performance obtained with MDA is superior to that obtained with LDA and demonstrate the suitability of MDA on real-life biometric data, obtained from a grip-pattern recognition experiment [14].

2 Discrimination as a figure of merit

We assume that the elements $x_i, i = 1, \ldots, d$ of a feature vector $\mathbf{x}$, randomly drawn from the observation space, are independent and identically distributed and have Gaussian probability densities with zero mean and unit variance, denoted by $p(\mathbf{x})$. Furthermore, we assume that the elements of a genuine feature vector are uncorrelated. In practice, these conditions can always be met by applying a linear transform that simultaneously whitens the observation space and uncorrelates the genuine feature vector [4]. The probability density of a feature vector belonging to a class with mean $\mathbf{m}$ is denoted by $p(\mathbf{x}|\mathrm{m})$, with $\mathcal{E}\{x_i|\mathrm{m}\} = m_i$ and $\mathcal{E}\{x_i^2|\mathrm{m}\} = \sigma_i^2$. The advantage of these assumptions is that the expressions for the likelihood ratio and discrimination distance become simple. The likelihood ratio given class mean $\mathbf{m}$ is denoted by $L(\mathbf{x};\mathrm{m}) \stackrel{\mathrm{def}}{=} \frac{p(\mathbf{x}|\mathrm{m})}{p(\mathbf{x})}$ and the log-likelihood ratio by $l(\mathbf{x};\mathrm{m}) \stackrel{\mathrm{def}}{=} \log(L(\mathbf{x};\mathrm{m}))$. The discrimination [3] (or Kullback-Leibler distance [6]) is given by

$$D_{\mathrm{dis}} \stackrel{\mathrm{def}}{=} \int \log(\frac{p(\mathbf{x}|\mathrm{m})}{p(\mathbf{x})}) p(\mathbf{x}|\mathrm{m}) \mathrm{d}\mathbf{x} = \mathcal{E}\{l(\mathbf{x};\mathbf{m})|\mathbf{m}\} \quad (1)$$

and the divergence [4] (or symmetrical discrimination) by $D_{\mathrm{div}} = \mathcal{E}\{l(\mathbf{x};\mathbf{m})|\mathbf{m}\} - \mathcal{E}\{l(\mathbf{x};\mathrm{m})\}$.

Next we demonstrate that there is an approximately linear relation between the discrimination and the log equal-error rate, which is an accepted figure of merit for biometric verification. Since there is no closed expression for the equal-error rate, this is done experimentally. Figure 1 shows combinations of the log equal-error rate and discrimination ($\cdot$) and combinations of log equal-error rate and the divergence ($+$) of 100 randomly drawn parameter sets $\{(m_i, \sigma_i^2)\}_{i=1}^d$, with $d = 25$. The divergence is included because in [4] it was recommended as an objective function for dimension reduction. A comparison with other figures of merit is discussed in [13]. The log equal-error rates were computed by means of Monte-Carlo simulations. The σ_i^2 were drawn from a uniform probability density on the interval $[0, 1]$. The m_i were drawn from a Gaussian probability density with zero mean and variance $1 - \sigma_i^2$. This dependency of m_i on σ_i^2 ensures that the sum of the within-class and the between-class variances equals the total variance, which must hold for verification. Figure 1 illustrates that the relation between the log equal-error rate and the discrimination can be approximated well by a straight line. This means that the discrimination can be used to predict the equal-error rate and can, therefore, be accepted as a figure of merit for biometric verification. Figure 1 also shows that the relation between log-equal-error rate and the divergence is less consistent.

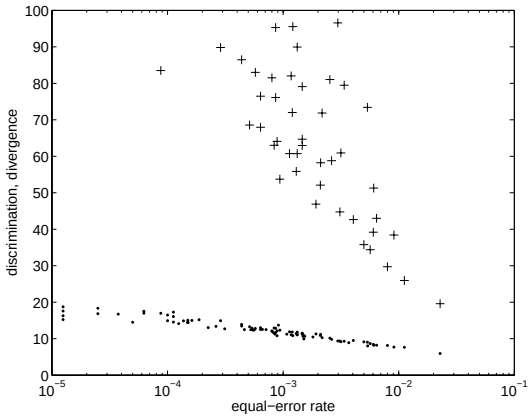


Figure 1. Relation between equal-error rate and discrimination ($\cdot$), and equal-error rate and divergence ($+$).

3 Maximum Divergence Analysis

We rewrite the discrimination (1) as

$$D_{\mathrm{dis}}(\mathbf{m}, \mathbf{\Lambda}) = \frac{1}{2}\left(\mathbf{m}^{\mathrm{T}}\mathbf{m} + \mathrm{trace}(\mathbf{\Lambda}) - \log(|\mathbf{\Lambda}|) - d\right), \quad (2)$$

with $\mathbf{\Lambda}$ an $d \times d$ diagonal matrix with $\lambda_{ii} = \sigma_i^2$. The aim of MDA is to determine a subspace of the observation space with an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $n < d$, such that the discrimination after projection onto this subspace is maximum.

Let $\mathbf{V}_n = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. The discrimination after projection is $D_{\mathrm{dis}}(\mathbf{V}_n^{\mathrm{T}}\mathbf{m}, \mathbf{V}_n^{\mathrm{T}}\mathbf{\Lambda}\mathbf{V}_n)$. For $\mathbf{m} = 0$, it can be shown that the basis $\mathbf{V}_n$ that maximizes this expression is

identical to the one resulting from LDA. If this is not the case, direct maximization is cumbersome. Therefore, we follow an iterative approach in which we first determine the optimal projection onto a one-dimensional subspace. That is, we look for the $\mathbf{v}_1$, with $\|\mathbf{v}_1\| = 1$, which maximizes

$$\mathbf{v}_1^{\mathrm{T}}\mathbf{m}\mathbf{m}^{\mathrm{T}}\mathbf{v}_1 + \mathbf{v}_1^{\mathrm{T}}\mathbf{\Lambda}\mathbf{v}_1 - \log(\mathbf{v}_1^{\mathrm{T}}\mathbf{\Lambda}\mathbf{v}_1). \qquad (3)$$

Note that the order in the first term has changed and that the trace operator and the determinant have disappeared. Each subsequent $\mathbf{v}_i$, $i = 2, \ldots, d$ is then found by maximizing

$$\mathbf{v}_i^{\mathrm{T}}\mathbf{m}\mathbf{m}^{\mathrm{T}}\mathbf{v}_i + \mathbf{v}_i^{\mathrm{T}}\mathbf{\Lambda}\mathbf{v}_i - \log(\mathbf{v}_i^{\mathrm{T}}\mathbf{\Lambda}\mathbf{v}_i) \qquad (4)$$

under the constraints that $\|\mathbf{v}_i\| = 1$ and $\mathbf{v}_i \perp \mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. We write $\mathbf{v}_i = \mathbf{V}_i^{\perp}\mathbf{w}_i$, in which the columns of $\mathbf{V}_i^{\perp}$ are orthonormal and span the subspace orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. The constrained maximization of (4), or (3), is then equivalent to the unconstrained maximization as a function of $\mathbf{w}_i$ and λ of

$$\mathbf{w}_i^{\mathrm{T}}\mathbf{D}\mathbf{w}_i + \mathbf{w}_i^{\mathrm{T}}\mathbf{C}\mathbf{w}_i - \log(\mathbf{w}_i^{\mathrm{T}}\mathbf{C}\mathbf{w}_i) - \lambda(\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_i - 1). \quad (5)$$

with $\mathbf{D} = \mathbf{V}_i^{\perp\mathrm{T}}\mathbf{m}\mathbf{m}^{\mathrm{T}}\mathbf{V}_i^{\perp}$ and $\mathbf{C} = \mathbf{V}_i^{\perp\mathrm{T}}\mathbf{\Lambda}\mathbf{V}_i^{\perp}$. Setting the derivative w.r.t. $\mathbf{w}_i$ equal to zero results in

$$\left(\mathbf{D} + \mathbf{C}(1 - \frac{1}{\mathbf{w}_i^{\mathrm{T}}\mathbf{C}\mathbf{w}_i})\right)\mathbf{w}_i = \lambda\mathbf{w}_i. \qquad (6)$$

This nonlinear equation is solved iteratively by taking $\hat{\mathbf{w}}_i^{(0)} = \mathbf{V}_i^{\perp\mathrm{T}}\mathbf{m}/\|\mathbf{V}_i^{\perp\mathrm{T}}\mathbf{m}\|$ as an initial estimate and selecting $\hat{\mathbf{w}}_i^{(j)}$ as the eigenvector of $\mathbf{D} + \mathbf{C}(1 - \frac{1}{\hat{\mathbf{w}}_i^{(j-1)\mathrm{T}}\mathbf{C}\hat{\mathbf{w}}_i^{(j-1)}})$ for which (5) is maximum. We found that this procedure converges rapidly, e.g. in less than 5 iterations for $d = 50$.

The performance of MDA depends on the parameters $\{(m_i, \sigma_i^2)\}_{i=1}^{d}$. If $m_i = 0$, $i = 1, \ldots, d$, MDA will have the same results as LDA. Otherwise, MDA will result in a higher discrimination after dimension reduction than LDA. How much higher depends on the parameters. In order to illustrate the effectiveness of MDA for dimension reduction, we plot the results of an example of dimension reduction by MDA and LDA with randomly drawn $\{(m_i, \sigma_i^2)\}_{i=1}^{d}$, $d = 25$. The σ_i^2 were drawn from a uniform probability density on the interval $[0, 1]$ and the m_i from a Gaussian probability density with zero mean and variance $1 - \sigma_i^2$. Figure 2 shows the discrimination and the equal-error rate as functions of the reduced number of dimensions for MDA (∇) and LDA ($*$). The figure illustrates that MDA outperforms LDA in terms of both discrimination and equal-error rate. For instance, an equal-error rate of 10^{-2} requires 4 dimensions with MDA, but 7 with LDA. The reason for MDA's better performance is that it makes use of the discriminative power of the mean m, which is ignored by LDA. An approximate and simpler version of MDA, called AMDA, with a performance that is very close to that of MDA is described in [1].

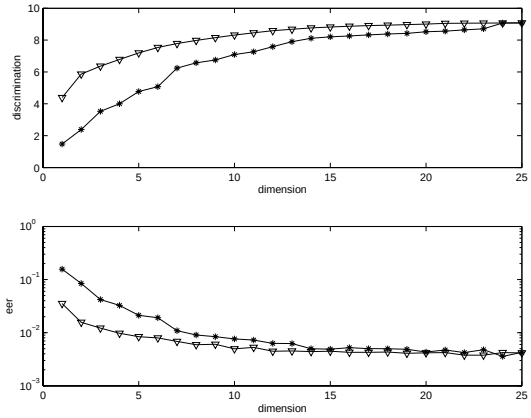


Figure 2. Discrimination (top panel) and equal-error rate (bottom panel) as functions of the number of dimensions for MDA (∇) and LDA ($*$).

4 MDA applied to grip-pattern recognition

We describe the effect on the verification performance of an application of grip-pattern recognition. This biometric is based on the pressure pattern exerted while holding an object. Its application for securing police guns against unauthorized use is described in [14][1]. A pressure sensor mounted on the grip of the gun measures grip patterns such as shown in Figure 3. These are images of 44×44 pixels, with values in the range $0 \ldots 255$. Hence, a feature vector has as many as 1936 elements. In [14] an experiment is described in which 855 grip patterns were gathered from 26 mostly untrained subjects. From each subject 30 to 100 right-hand grip patterns were taken. The data were randomly split into equally sized training and a test sets. A first dimension reduction to $d = 77$ by means of PCA was performed to obtain non-singular covariance matrices. The within-class covariance matrix and the within-class means were estimated from the resulting training set. A second dimension reduction to $d = 25$, i.e. the number of users minus 1, by means of LDA, which principally cannot affect the verification performance, was then performed. The observation space was whitened and the within-class covariance matrix diagonalized. Log-likelihood-ratio classifiers were derived for all users. A global threshold was used for the log-likelihood ratio and the EER for the test set was 4.8%.

We applied MDA and LDA to the resulting feature vectors and compared the EERs. The feature vector's dimension was reduced from 25 to 1 in steps of 1. Figure 4 shows

[1]The continuation of this research will be supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economy Affairs.

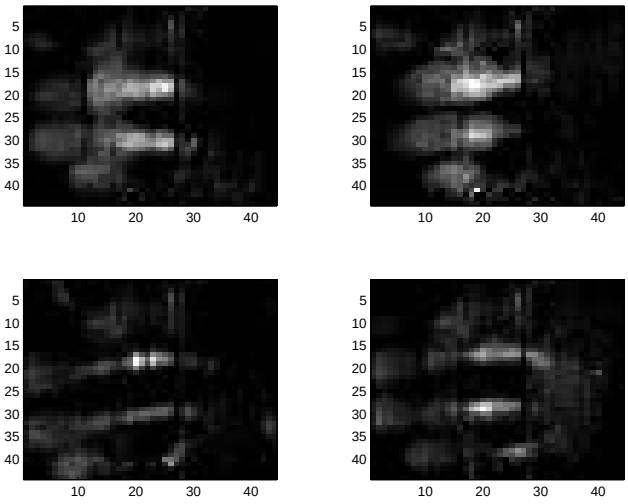


Figure 3. Gray-scale images of the average grip patterns of 4 users.

the EERs of MDA (solid line) and LDA (dashed line) as a function of the dimension. The dotted line is the initial EER at $d = 25$. The curves show that the loss of performance

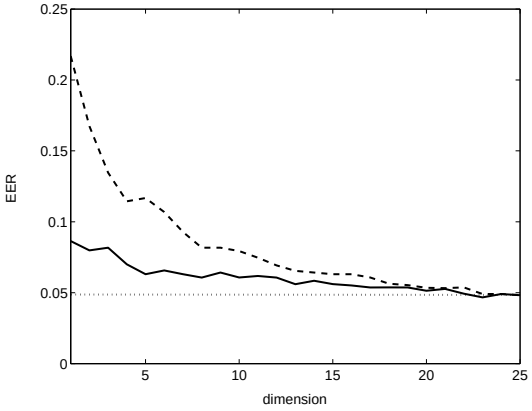


Figure 4. EERs of MDA (solid line) and LDA (dashed line) as a function of the dimension.

is smallest for MDA. In fact, a reduction of the dimension from 25 to 5 leads to an increase of the EER from 4.8 to 6% with MDA and to an increase of 4.8 to 12% with LDA.

5 Conclusions

A new method, MDA, for the reduction of the dimension of the feature vector prior to (biometric) verification has been proposed. It uses the discrimination distance between the probability densities of genuine feature vector and the entire observation space as an objective function. Experiments with synthetic and with grip-pattern data have shown that MDA outperforms LDA in terms of discrimination and equal-error rate. The reason for MDA's better performance is that it makes use of the discriminative power contained in the class mean, which is ignored by LDA.

References

[1] A. Bazen and R. Veldhuis. Detection of cores in fingerprints with improved dimension reduction. In *Proceedings of the* 4^{th} *IEEE Benelux Signal Processing Symposium (SPS-2004)*, pages 41–44, Hilvarenbeek, The Netherlands, 2004.

[2] P. N. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition*, 17(7):711–720, 1997.

[3] R. Blahut. *Principle and Practice of Information Theory*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1987.

[4] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, San Diego, second edition, 1990.

[5] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, 1968.

[6] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[7] N. Kumar and A. G. Andreou. Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition. *Speech Communication*, 26(4):283–297, 1998.

[8] M. Loog and R. Duin. Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):732–739, 2004.

[9] M. Ordowski and G. Meyer. Geometric linear discriminant analysis for pattern recognition. *Pattern Recognition*, 37(3):421–428, 2004.

[10] G. Saon and M. Padmanabhan. Minimum bayes error feature selection for continuous speech recognition. In *Advances in Neural Information Processing Systems*, volume 13, pages 800–806, Cambridge, MA, 2001. MIT Press.

[11] G. Saon, M. Padmanabhan, R. Gopinath, and S. Chen. Maximum likelihood discriminant feature spaces. In *Proceedings ICASSP2000*, Istanbul, 2000.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[13] R. Veldhuis and A. Bazen. Figures of merit for biometric verification and a means for dimension reduction. In *Proceedings of the 25th Symposium on Information Theory in the Benelux*.

[14] R. Veldhuis, A. Bazen, J. Kauffman, and P. Hartel. Biometric verification based on grip-pattern recognition. In *Proceedings of the IS&T/SPIE 16th Annual Symposium Electronic Imaging*, San Jose, CA, USA, January 2004.

How to deceive a face recognizer?

B. Gökberk, L. Akarun, B. Aksan
Computer Engineering Dept., Boğaziçi University
{gokberk, akarun, aksan}@boun.edu.tr

Abstract

Many security systems depend upon face recognizers to identify a person. Many of these systems are passive and are deployed at places such as airline terminals. However, face recognizers are sensitive to deception attacks. Previous studies suggest that hair regions are very crucial in face recognition and the success of a recognizer depends on the success of a pre-segmentation stage which extracts the face region from the hair and the background. Deception attacks which would change the hairstyle, apply make-up or occluding objects to the face would cause many systems to fail. In this study, we study the effects of deception attacks on two basic face recognition systems: a PCA-based system and a Gabor wavelet-based recognizer. We study the performance of the recognizers under different attacks and focus on the selection of features so as to maximize performance under attacks.

1. Introduction

Over the past two decades significant progress has been made in the automatic human face recognition research. Although many successful face recognition systems have been proposed in the literature, the problem is still not considered to be fully solved, especially in real-life applications. The main obstacle can be simply stated as follows: intra-personal variations between human faces is large when compared to inter-personal variations. These variations can be broadly classified into two groups: *external variations* and *internal variations*. Variations due to illumination, head pose, scale and translation are considered to be external variations. However, variations due to hair color, hair style, moustache, beard and eyeglasses as well as facial variations which stem from the subject itself are considered to be internal variations.

One of the studies dealing with internal variations such as expression changes and occlusion is [1], where the AR face database is used to illustrate the superior performance of a local probabilistic approach. The local component based approach has also been studied to deal with external variations in face recognition [2, 3]. When occlusions such as beards and glasses are present, a different approach is to try to remove them [4, 5].

In this paper, our aim is to examine how an impostor can deceive a face recognizer by taking the advantage of internal variations; specifically hair color change, occlusions, and expression variations. After analyzing the effects of such variations aiming to deceive a recognizer, we propose a robust technique that increases the performance of PCA and Gabor-based face recognizers.

2. Face Representation

2.1. PCA-based Method

In PCA, faces are expressed as linear combinations of the eigenvectors of faces. Then, for recognition, the PCA coefficients can be used to denote a face. In its original form, PCA is found to be rather sensitive to image intensity variations, local perturbations, and needs almost perfect correspondence. Image variations which are not present in the training phase generally cause a poor recognition performance. A possible solution to improve the PCA method is to divide the whole face region into subregions and do modular PCA analysis. In a modular PCA analysis, each subregion is handled in isolation, and for each subregion, a different subspace is found. Then local features are extracted and merged to represent a face. An important advantage of modular PCA analysis is that local perturbations can only affect the local coefficients, not the whole face. Figure 1.c shows subregions that we have used in our experiments.

2.2. 2D Gabor Wavelet-based Method

A biologically motivated representation of face images is to code them using convolutions with multi-frequency multi-orientation 2D Gabor–like filters. In order to represent face images using Gabor filters, the intensity image is convolved by Gabor kernels. The set of convolution coefficients for kernels of different orientations and frequencies

at one image pixel constitutes local feature vectors. Local feature vectors are then merged to represent whole face. In this work, we employ Gabor filters as in [6], and use uniform grid-based sparse representation (see Figure 1).

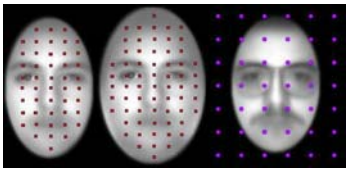


Figure 1. Gabor sampling grid points: (a) Small ellipse, (b) Large ellipse (c) Local PCA regions

3. Similarity Measure and Classifier

In both original PCA-based and Gabor wavelet-based recognition methods, L_2-norm similarity measure is used where classification is done via 1-nearest neighbor algorithm. Although comparing two faces in modular PCA-based approach can be simply done by $||I_i - I_j||$ where $||\cdot||$ denotes $L_2 - norm$, a more robust distance measure can be used by taking advantage of the locality principle. Let $I_i = \{V_1^i, V_2^i, \ldots V_p^i\}$, and $I_j = \{V_1^j, V_2^j, \ldots V_p^j\}$ be two global feature vectors for two different images, and let $d = \{||V_1^i - V_1^j||, ||V_2^i - V_2^j||, \ldots ||V_p^i - V_p^j||\}$ be local Euclidean distance vector between corresponding local feature vectors. Here, each component in the d denotes how similar the local regions are in two images. For robustness against outlier regions, one can simply discard some subregions having smallest similarities, and select t subregions having greatest similarities, and then calculate the overall L_2-norm of this selected subregions. In this paper, we call this technique as asymmetric trimmed distance measure (ATDM). ATDM can also be applied to the Gabor wavelet-based representation technique by selecting most similar t local features among the global feature vectors.

4. Experimental Results

4.1. The Effect of Hair Color Change

In our experiments, we have used a subset of AR face database. For hair color experiments, we have selected four neutral images from 20 males and 20 females where the first two images are from the first session and the other two are from the second session. For each image, we have automatically generated six synthetic face images in increasing order of hair color change from darker color to lighter color. Faces are normalized and rotated according to eye coordinates. After normalization, faces are cropped by an ellipse mask. In hair color experiments, two different ellipse masks are employed. The small mask covers the face outline whereas the large mask covers a wider region that includes the chin and the ears. Samples of original images and six synthetic images for three subjects are shown in Figure 2. Small and large ellipse masks are shown in Figure 1.

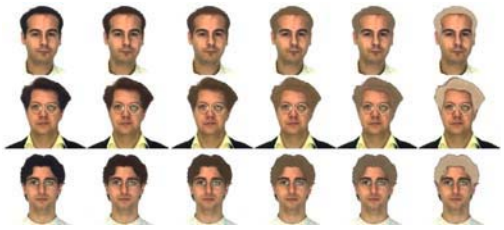


Figure 2. Sample images from hair dataset

In order to see the effect of hair color change, we have designed seven experimental setups. We have four original images per person, and six differently colored synthetic sets where each set contains four images that are modified from original images. Let $S_0 = \{I_0^1, I_0^2, I_0^3, I_0^4\}$ be the original set of images of a person and $S_i = \{I_i^1, I_i^2, I_i^3, I_i^4\}, i = 1 \ldots 6$ be the synthetic image sets of a person. In the first experiment, denoted by H_1, we put two images from S_0 to the training set, and the remaining two images of S_0 to the test set. Since there are six possible configurations for training set image selection, H_1 has six training-test set configurations. Similarly in H_2, six different training sets are formed from S_0 (original images) containing two images per person, and test sets are formed from S_1. So in each one of the six different configurations in H_2, we ask our algorithm to recognize synthetically modified two test images where the training set contains two original images. In such a setup, the first experiment, H_1 determines the baseline classification accuracy of our recognizers, whereas experiments $H_2 \ldots H_7$ determine how much the increasingly modified hair color affects the recognition performance.

In the PCA method, we represent each face image using the first $k = 40$ PCA coefficients. In the Gabor method, we identify uniform grid-like regions in ellipse masks (see Figure 1). The number of grid points is $g = 47$ for the small ellipse mask, and $g = 65$ for the large ellipse mask. At each grid point, local features of dimensionality $5 \times 8 = 40$ are extracted, and then global feature vector is formed by concatenating these. Table 1 displays the classification performances of PCA and Gabor methods on seven experiments. Figure 3 plots the results on Table 1. It is clear from Figure 3 that as the hair color change increases the recognition performances of both PCA and Gabor methods decreases.

However, Gabor method outperforms PCA in all experiments, and its performance is more resistant than PCA. This behavior is especially clear when going from H_4 to H_5. We also see that larger ellipse is better when hair color change is minimal.

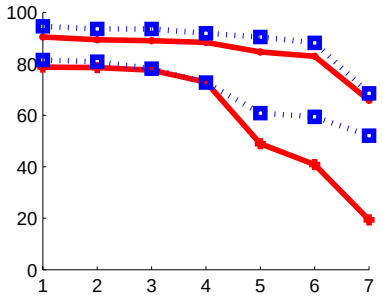


Figure 3. PCA and Gabor performances on small ellipse hair dataset for Euclidean (red lines) and robust ATDM (blue dotted lines)

In the previous section, we saw that intensity variations around the hair region cause the deterioration of both PCA and Gabor-based classifier accuracies. This is an expected result especially for the PCA method because it is known that PCA runs into problems when it is required to code an unknown test image having variations which are not adequately present in the training set. PCA can not generalize well if the learning set (training images) does not cover all possible variations. In our experiments, training images are selected from original faces, and hair color variations are not present in training sets. This explains the poor generalization. For the Gabor method, the local features extracted from hair regions will be different, and this explains the performance drop. The reason why Gabor method performs better than PCA is that Gabor filters are less sensitive to intensity changes, and respond edge-like features in these regions.

After training the PCA and Gabor techniques against internal variations, we applied the robust similarity measure defined in Section 3. Since the PCA technique has shown inferior results, we use modular PCA in combination with the robust distance measure. The recognition accuracies of modular PCA and Gabor wavelet-based robust ATDM are shown in Table 1. Figure 3 depicts these results. In Figure 3, we see that ATDM outperforms the baseline PCA and Gabor methods especially in the difficult experiments, e.g. H_5, H_6, H_7. This is an important observation: ATDM improves the performance when there are highly variable subregions.

We have also performed additional experiments where some synthetically modified images are put into the training set in order to better reflect the variations. The results have shown that adding the synthetic images to the training set, especially the ones with more variation from the original hair color, generally increased the recognition accuracy of the PCA-based representation, as expected. However, the Gabor-based classifier and PCA-based classifier with ATDM did not have a considerable amount of increase in performance.

4.2. Eyeglasses and Moustache

Among internal variations besides hair color change, eyeglasses and moustache differences may be used for deception attacks. To analyze the effects of such variations, a different subset of the AR face database is used. 40 males are selected from the dataset where each male has two neutral images, two images having slight expression variations, and two images with dark eyeglasses. Let S_n, S_e, and S_g denote these image sets respectively. Each set contains two images. We automatically generate synthetic sets from these images by adding moustache to each individual set, and obtain synthetically generated sets: S_{nm}, S_{em}, S_{gm} respectively. Sample images from these sets are shown in Figure 4.

Figure 4. Sample images from the eyeglasses/moustache dataset.

We have designed five experiments to analyze how face recognizers behave under expression, eyeglass, and moustache variations. The training and test set configurations are as follows: E_1 = (Tr:$\{S_n\}$, Ts: $\{S_e\}$), E_2 = (Tr:$\{S_n$,$S_e\}$ Ts: $\{S_g\}$), E_3 = Tr:$\{S_n\}$, Ts:$\{S_{em}\}$), E_4 = (Tr:$\{S_e\}$, Ts:$\{S_{nm}\}$), E_5 = (Tr:$\{S_n$,$S_e\}$, Ts: $\{S_{gm}\}$)

Table 2 shows the recognition accuracies for both standard PCA and Gabor methods and their robust versions on small and large ellipse masks. PCA results show that adding moustache does not cause significant performance degradation. This can be seen by comparing the accuracies in E_1 experiments to the accuracies for experiments E_3 and E_4. Although robust version has an improved accuracy in eyeglass experiments (E_2, E_5), both PCA versions perform poorly. Note that ellipse size does not effect performance in PCA.

In Gabor results, the robust ATDM generally outperforms the standard version. Also, large ellipse mask has performed better than the small one. Moustache experiments

Table 1. Classification accuracies of PCA and Gabor methods for hair dataset.

	PCA				Gabor			
	Small Ellipse		Large Ellipse		Small Ellipse		Large Ellipse	
	Euc.	ATDM	Euc.	ATDM	Euc.	ATDM	Euc.	ATDM
H_1	78,75	81,46	78,54	86,25	90,42	94,58	88,13	93,13
H_2	78,54	80,83	77,92	85,83	89,38	93,54	86,25	91,67
H_3	77,50	78,13	77,71	82,71	88,96	93,54	85,83	91,46
H_4	72,92	72,71	72,29	76,04	88,33	91,88	84,58	90,83
H_5	48,96	60,83	44,79	63,33	84,58	90,42	82,92	90,00
H_6	40,83	59,38	36,04	56,88	82,92	88,13	82,71	90,00
H_7	19,38	52,08	17,92	54,38	65,83	68,54	76,04	83,54

Table 2. Eyeglasses/Moustache Experiments

	E_1	E_2	E_3	E_4	E_5
	PCA				
Euc. (S)	78.75	15.00	80.00	70.00	11.25
ATDM (S)	82.50	35.00	83.75	68.75	28.75
Euc. (L)	78.75	15.00	80.00	70.00	11.25
ATDM (L)	82.50	35.00	83.75	67.50	28.75
	Gabor				
Euc. (S)	71.25	21.25	78.75	76.25	16.25
ATDM (S)	78.75	28.75	81.25	85.00	20.00
Euc. (L)	85.00	48.75	87.50	83.75	50.00
ATDM (L)	85.00	70.00	86.25	91.25	60.00

(E_3, E_4) show that both standard and improved robust Gabor method is superior to PCA. This situation is especially visible in eyeglasses experiments where robust ATDM improves from 35.00 percent (PCA) to 70.00 percent in E_2, and from 28.75 percent (PCA) to 60 percent in E_5. However, since dark eyeglasses cover most of the discriminative regions in the human face, small ellipse masks can not provide useful information enough for recognition. This explains poor accuracies in the Gabor method, i.e., 28.75 percent recognition accuracy in E_2 using the robust ATDM method.

5. Conclusion

In this paper, we analyze several deception attacks which use internal facial variations such as hair color change, expression variations, and occlusions by moustache and eyeglasses. Results show that both PCA-based and Gabor wavelet-based face recognizers are sensitive to these variations, although the latter generally outperforms the first. In hair color experiments, we see that PCA performance deteriorates drastically, while Gabor-based classifier is more robust to color changes. In eyeglasses experiments, since a large portion of a face is occluded, both approaches perform poorly. It is also shown that adding moustache does not effect the recognition rate significantly. After these observations, we propose a robust classifier which uses asymmetric trimmed distance measure. This distance measure is suitable for modular representations. Therefore, a modular PCA algorithm is used to represent local facial regions. Our experiments show that using asymmetric trimmed distance measure with modular PCA and Gabor methods significantly improves the recognition performance when test images have considerable variations such as hair color change and eyeglasses.

References

[1] A. M. Martinex. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Tran. on PAMI*, 24(6):748–763, 2002.

[2] R. Gottumukkal and V. L. Asari. An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25(4):429–436, 2004.

[3] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: Component-based versus global approaches. *Computer Vision and Image Understanding*, 91(1/2):6–21, 2003.

[4] Bon-Woo Hwang and Seong-Whan Lee. Reconstruction of partially damaged face images based on a morphable face model. *IEEE Tran. on PAMI*, 25(3):365–372, 2003.

[5] C. Wu, C. Liu, H.-Y. Shum, Y.-Q. Xu, and Z. Zhang. Automatic eyeglasses removal from face images. *IEEE Tran. on PAMI*, 26(3):322–336, 2004.

[6] L. Wiskott, J. M. Fellous, N. Kruger, and C. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Tran. on PAMI*, 19(7):775–779, 1997.

Template Aging in Speech Biometrics

Andreas Wolf
VOICE.TRUST AG*
Landshuter Allee 12-14, 80637 Munich, Germany
awo@voicetrust.de

Abstract

Biometric systems combine security with user-friendliness. The most important advantage of biometrics over other authentication schemas is the inseparable connection between an individual and his physical attributes. It is assumed, that these attributes like fingertip, iris pattern, face geometry, or speech are more or less unique. But are they constants? Obviously, not exactly. They vary within a certain range. But, does this range itself vary in the course of time? This paper reports on a yearlong experimental evaluation of the template aging effect in different speaker verification systems.

1. Introduction

Biometric systems are secure and user-friendly. Nothing can be forgotten, lost, or stolen, and systems with tests on liveliness are barriers for replay or simulation attacks. You simply have to be *only yourself*. This is the theory. In practice, the recorded biometric attributes vary. They may depend on the disposition of a user, on environmental conditions like lighting, background noise, or temperature. As long as the changes depend on the scanner device, one can try to guarantee fixed, standardized conditions. For speaker verification this can be done, e. g., by utilization of ISDN telephone handsets. But, what happens, if the user itself changes his *attributes*? In the speech example, we know that a cold may have some influence. After getting well the voice is as before. Is it really as before? Does it change in the course of time? If yes, how significantly, and how quickly? If one looks for scientific publications or asks the vendors of biometric solutions, you do not find that much.

In *Best Practices* [6] the following statement has been made: *For scenario evaluations, test data must be separated in time . . . For most systems, this interval may not be known. In such cases, a rule of thumb would be to separate the samples at least by the general time of healing of that body part.* This is not very helpful in the voice case. The same document addresses template aging as an important factor for biometric systems: *Template ageing, . . . will vary in accordance with the delay between creation of the enrolment template, and the verification or identification attempt. Generally, performance a short time after enrolment, when the user appearance and behavior has changed very little, is far better than that obtained weeks or months later.* But what delay exactly is the appropriate one for speaker verification tests? The NIST report [4] gives no hints on how to deal with template aging. The IBG [8] suggests to collect data for an assessment of the template aging six weeks after enrollment, but they give no motivation.

In Section 2, a brief introduction to the principles of speech biometrics and speaker verification systems is given, followed by a formulation of the problem and a description of setup and realization of the experiment in Section 3. Section 4 presents the results and an assessment.

2. Speech Biometrics

Biometrics. Biometrics are automated methods of recognizing a person based on physiological or behavioral characteristics. Among the features measured are face, fingerprints, hand geometry, handwriting, iris, and voice.

Biometrics were applied even in the 19th century. In the 1890s, Alphonse Bertillon sought to identify convicted criminals by a method of multiple body measurements (Bertillonage). Possibly the first known biometrics in practice, as reported from China in the 14th century, were ink stampings of children's palm prints and foot prints on pa-

* Setting the standard, VOICE.TRUST is the worldwide leader in highly secure voice authentication solutions. With over 100,000 licenses sold we are European voice authentication market leader. Since 2000, VOICE.TRUST's solutions for secure authentication have lead to a dramatic reduction in operating costs over conventional authentication solutions of up to 80% at leading companies of all industries. Simple, safe and secure, VOICE.TRUST brings easy-to-use solutions for PIN and Password Reset, Single Sign-On, Remote access, PKI-Support, Caller-Identifi cation and Two-Factor Authentication to the network security, voice-portal, call center and helpdesk markets.

per. Sometimes it seems that speech biometrics is a brand new technology. But this is not true. Over the past 40 years, speech scientists, linguists, and computer scientists have studied the human voice. The concept of a *voiceprint* occurs in the literature even in the early 1960s [5]. One can find papers like [3] on Voice Verification in the 1970s. Texas Instruments was the pioneer of speech systems in the 1960s.

Biometric technologies are about to become the foundation of an extensive array of secure identification and personal verification solutions. As the amount of security breaches and transaction fraud increases, the need for highly secure identification and personal verification technologies becomes apparent [1]. Biometric verification may be preferred over traditional methods using passwords or PIN numbers for various reasons: The person to be identified is required to be physically present at the point-of-identification, and an identification based on biometric techniques renders the need to remember a password or carry a token unnecessary. PINs and passwords may be forgotten or compromised, and token-based methods of identification may be forged, stolen, or lost. By replacing or adding to a PIN-based solution, biometric techniques can potentially prevent unauthorized access to or fraudulent use of ATMs, smart cards, computers, etc. An important issue in the design of a biometric system is to determine how an individual is identified. Depending on the context, a biometric system can be either a verification (authentication) system or an identification system. Verification involves confirming or denying a person's claimed identity. In identification, one has to establish a person's identity.

Speaker verification. In a speaker verification system, a person's identity can be confirmed by analyzing unique speech characteristics such as word pronunciation or spectral energy distribution. Speech verification provides a means to confirm that someone is the one he claims to be by comparing a spoken utterance against a previously recorded biometric template. *Text dependent* speaker verification is based on the comparison of the pronunciation of a requested utterance chosen by the proposed system with a pre-recorded *voice template*. This procedure can be enriched with life test functionality: The system chooses a random utterance from a pool of possible utterances previously recorded under secure conditions and stored as a reference. The verification system prompts the user for an utterance and compares it to the stored template to verify the user's identity. This so-called *Challenge & Response* procedure complicates an unauthorized access. So it is much more difficult possible to use pre-recorded utterances. Another type of speaker verification is the *text independent* one. Here, only general characteristics of a person's voice are evaluated. Currently, one can find the text dependent version to be dominating in available applications.

3. The Template Aging Experiment

The problem. Biometric systems use specific individual characteristics of human beings for identification and verification purposes, like fingerprints, iris, hand geometry, face image, or speech. Obviously, these characteristics change over time. This can be noticed, e.g., observing the growth of children. But even adults change, they age. This aging may have impact on the quality of the biometric data a person may provide to a biometric system. But does aging really have impact on the recognition power of biometric systems? In our case, we are interested in speech. Can an aging be observed with respect to the templates used for speaker recognition systems? In the literature, there are only sparse hints. In the community, several opinions can be heard: Some colleagues expect that after two weeks the ability to verify against a given voice template gets stable, at least for adults having no vocal tract disease, others assume that after 3 months this stability is reached, and a third group says that such a stability will never be reached.

Everybody who has planned experiments with biometric systems will agree that the data collection is very difficult. In the speech biometrics case, test persons have to be motivated to call a telephony application and to repeat several phrases several times. To be able to compute data on long-term characteristics, these calls have to be repeated regularly. Even for a small group of six people, all employed at the same company, this was a non-trivial task. So I want to thank my colleagues Andreas B., Bettina, Harald, Johannes, and Stephan for their patience with my boring data collection application and me. We planned to collect data from a simulated speaker verification system over a period of one year. Once having the data, we wanted to find answers for the following questions:

- Is there a deterioration of the recognition performance of a speaker verification system?
- If yes, which quantity it has?
- If a level of stability can be observed, when this level is reached?
- Are there differences between speaker recognition systems available in the market?
- Which suggestions can be given to operators of speech verification systems with respect to the template aging phenomenon?

Setup of the experiment. Beginning in April 2003, six members of the VOICE.TRUST staff started a series of (almost) weekly speech data collection experiments. To collect the data, an application was used, that was developed originally for the speech data collection for VOICE.TRUST's ongoing Common Criteria evaluation process. This application is telephony based and prompts the caller to repeat

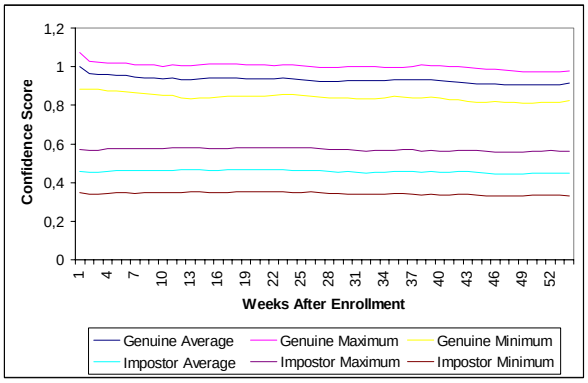


Figure 1. Changes of the speaker recognition performance of system 1 within one year.

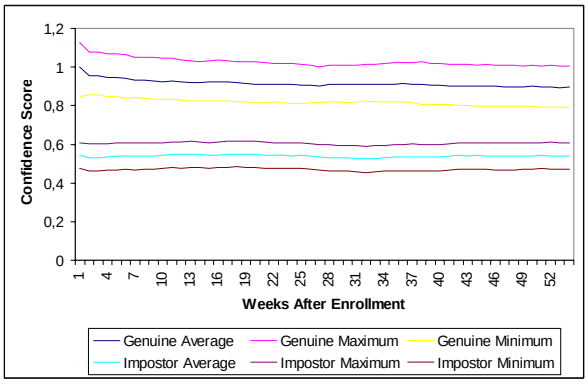


Figure 2. Changes of the speaker recognition performance of system 2 within one year.

several phrases, among them a generic alpha-numeric user login name and five pairs of German given names. Each caller is asked to repeat the same set of phrases. In each call, a speaker is asked to repeat each phrase twice. That way, within one call two variants of six phrases are recorded in WAV files. The callers are asked to use their usual office phone, and to use the same phone for all calls during the year. That way, the normal environment of the test persons has been used, and the data collection was performed in an environment that is very similar to that of real voice authentication applications.

To generate the biometric template in the course of the first call, the callers are required to repeat all phrases four times. The templates have been generated for three different speaker verification engines from different vendors available in the market. For each recorded sound file the confidence value was computed with each appropriate template, that is, genuine access as well as impostor attempts has been simulated. That way, a complete three-dimensional matrix was computed. The first dimension describes the owners of the templates (6), the second the owners of the voice records to be tested (6), and the third the number of weeks since enrollment time (52). Each cell of the matrix then contains 12 confidence values, two for the two records of the generic ID and ten for the two records of the five name pairs. Such a matrix was computed for each of the three participating voice verification engines.

In many voice biometric applications a user gets a second chance if his first one leads to bad confidence values. To model this property, only the better one of the two confidence values for the two repeated records was used, the other one was discarded. Then, the average (arithmetic means), the minimal and the maximal confidence value have been computed for each cell. To eliminate the influence of the daily form of the test persons, for each attempt these min/max/average values have been smoothed by averaging with the appropriate values from the five preceding and succeeding weeks. This procedure was performed on the data of genuine users as well as of impostors. To allow a comparison of the results for different verification engines, the range of possible confidence values was re-scaled to the interval between zero and one. After this, for each set of templates the average confidence value reached by the genuine user in a data collection call done immediately after enrollment was used to re-scale all confidence values computed on that set of templates; so the average confidence value of a genuine user immediately after enrollment is scaled to 1.

Realization. The experiment was started in April 2003, and was finished in April 2004. Six persons called the data collection application described above in a weekly cycle. In periods with pubic holidays some dates have been skipped, but the calls have been made more or less regularly. It has to be mentioned that the motivation of the test persons has been the most difficult task in the entire experiment. Even if one would desire to have more participants to get a higher statistical significance, for such a long period this is not easy. The data has been collected, and the computation of the results has been started as described in the previous Section. The Figures 1, 2, and 3 describe the observed changes of the ”smoothed” speaker recognition performance of the used systems. To see some more details, we additionally give average confidence scores for genuine users as well as for impostors for some time ranges for the three considered systems in Table 1.

4. Results, Assessment, and Conclusions

That last Section contains an assessment of the numerical results of the experiment. Furthermore, it will be discussed, how the template-aging phenomenon can be addressed in real life environments. The BEM [2] suggests *periodic updating of the user's reference template*. How that

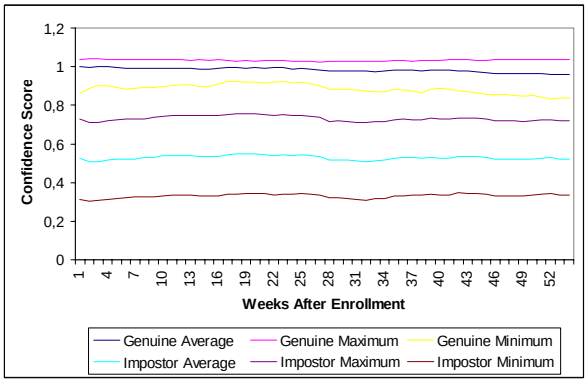


Figure 3. Changes of the speaker recognition performance of system 3 within one year.

Average confi dence in the weeks …	Sys 1		Sys 2		Sys 3	
	Gen.	Imp.	Gen.	Imp.	Gen.	Imp.
0	1.000	0.459	1.000	0.543	1.000	0.525
1-6	0.966	0.451	0.957	0.532	0.998	0.506
1-11	0.946	0.460	0.933	0.538	0.992	0.522
12-22	0.942	0.467	0.917	0.549	0.994	0.548
23-33	0.926	0.455	0.909	0.530	0.980	0.516
34-44	0.929	0.454	0.905	0.534	0.983	0.526
45-53	0.905	0.448	0.898	0.542	0.963	0.526

Table 1. Average confidence scores of genuine users in some time intervals after enrollment.

can be made? And what means *periodic*?

One can see, that all systems have an almost constant performance. In the first six weeks after enrollment, a significant decrease of approx. 4% of the reached confidence values can be observed for two of the three systems, followed by a smaller reduction within the following time. After this period, the confidence decreases by 3.5% to 4.1% in the following 45 weeks. This is less than 0.1% per week. In the long term, all tree systems from different vendors have shown the same behavior.

As one would expect, the average confidence value of impostor attempts does not change significantly over time.

As it is known, speaker recognition systems often apply neural networks, and automated training updates of the networks may lead to over-training. Over-training may have dramatic impact on the performance of any neural network, so it should be avoided to train nets automatically. Consequently, one can assume that there is no strong need for updates of speech-biometric templates. Even for applications with large time intervals between the verification calls, aging effects should not influence the recognition performance significantly, except for high security applications. That is, speech biometrics are suitable for long term authentication.

On the other hand, the results of the experiment presented in this paper allow an estimate on the usability period for templates. Knowing the currently reached confidence score, the required threshold, and the aging rate, it is possible to compute when the confidence probably will fall below the threshold. This additionally allows an estimate on the quality of enrollments. Whenever the estimated usability period of an enrollment computed immediately after the enrollment in a verification session is too short following the intended use of the system, the user can be asked to re-enroll. And, last, but not least, it can be computed automatically, when it makes sense to re-use recorded speech data for re-generation of templates. These automatically generated templates can be reviewed automatically, too, also using recorded and stored reference data. That way, over-training by template extension can be avoided.

For text-dependent speaker recognition one probably would like to increase the pool of phrases available for verification purposes. On the other hand, no user will be willing to enroll, say, 20 phrases at once. So the enrollment should be separated into several sessions, e.g., following each successful authentication. That way, and combined with the deletion of older templates, the set of templates can be rolled. In applications with longer intervals between authentication calls, from the recognition performance point of view, it makes not that much sense, except for intervals of more than a year, which can not be assessed using the results of this paper. For application with almost daily use, it can be recommended to apply template rolling to ensure that whenever possible no template reaches the stable (but lower) confidence level after some weeks.

References

[1] Biometric Consortium, http://www.biometrics.org

[2] Common Criteria Biometric Evaluation Methodology Working Group Biometric Evaluation Methodology, Release 1.0, August 2002, p. 19–21, 23, table 12.

[3] Doddington, G. R.: Personal Identity Verifi cation Using Voice. Proc. ELECTRO-76, 1976.

[4] Doddington, G. R. et al.: The NIST speaker recognition evaluation: Overview, Methodology, Systems, Results, Perspective. Speech Communication, 2000.

[5] Kersta, L. J.: Voiceprint Identifi cation. Nature. Vol. 196, pp. 1253-1257, 1962.

[6] Mansfield, A. J., Wayman, J. L.: Best Practice in Testing and Reporting, Performance of Biometric Devices; Biometric Working Group, Version 2.01, 2002.

[7] McGehee, F.: The Reliability of the Identifi cation of Human Voice. J Gen. Psychol. Vol. 17, pp. 249–271, 1937.

[8] International Biometric Group: Comparative Biometric Testing; IBG, New York, 2003.

Combining Multiple Biometrics to Protect Privacy

Berrin Yanikoglu and Alisher Kholmatov
Sabanci University
Tuzla, Istanbul, 34956, Turkey
berrin@sabanciuniv.edu, alisher@su.sabanciuniv.edu

Abstract

As biometrics are gaining popularity, there is increased concern over the loss of privacy and potential misuse of biometric data held in central repositories. The association of fingerprints with criminals raises further concerns. On the other hand, the alternative suggestion of keeping biometric data in smart cards does not solve the problem, since forgers can always claim that their card is broken to avoid biometric verification altogether.

We propose a biometric authentication framework which uses two separate biometric features combined to obtain a non-unique *identifier of the individual, in order to address privacy concerns. As a particular example, we demonstrate a fingerprint verification system that uses two separate fingerprints of the same individual. A combined biometric ID composed of two fingerprints is stored in the central database and imprints from both fingers are required in the verification process, lowering the risk of misuse and privacy loss. We show that the system is successful in verifying a person's identity given both fingerprints, while searching the combined fingerprint database using a single fingerprint, is impractical.*

1. Introduction

Biometric data is increasingly used in authentication and identification of individuals, replacing password-based security systems. Identification and authentication refers to two different tasks: finding the identity of a person given the biometric versus verifying the identity given the biometric data and the claimed identity.

There are two approaches to a biometric authentication system. In one alternative, enrolled users' biometric data is kept at a central repository and authentication is done by verifying the test data against the reference at the central repository. In the second alternative, a user carries a smart card containing his/her biometric data, and verification is done against the sample in the smart card. There are disadvantages associated with both of these two approaches. In particular there is increased concern over the loss of privacy and potential misuse of biometric data held in central repositories. Biometric data which can *uniquely* identify a person (e.g. fingerprints, iris patterns) can be used to track individuals, linking many separate databases (where the person has been, what he has purchased etc.). There is also fear that the central databases can be used for unintended purposes [5]. For instance, latent fingerprints can be used to search for information about a person in a central database, if such databases are compromised. The association of fingerprints with criminals raise further concerns for fingerprint databases in particular. Similarly, biometric data may reveal certain rare health problems [2], which raises concern about possible discriminatory uses of central databases.

On the other hand, keeping biometric data in smart cards has its own disadvantages. In particular, forgers can claim that their card is broken and avoid biometric verification altogether. Since a smart card may become damaged legitimately, such a situation would need to be solved by non-biometric authentication or by resorting to a central database.

In this paper we propose a biometric authentication framework to address these privacy concerns. In particular, two biometric features (e.g. fingerprints) are combined to obtain a *non-unique* identifier of the individual and stored as such in a central database. While the combined biometric ID is not a unique identifier, relieving concerns of privacy, we show that it can still be used in authenticating a person's identity. As a particular example, we demonstrate a fingerprint verification system that uses two separate fingerprints of the same individual to form a combined biometric ID.

With the proposed method, a person can give two fingerprints for one application (e.g., passport application), and two other fingerprints for another one (e.g., bank), creating two separate biometric IDs. While the person can still be authenticated for either application, it is impossible to link the two databases. Similarly, searching for a person using latent fingerprints is difficult, as one would need to try many such combinations of latent fingerprint pairs.

2. Previous Work

Unlike passwords which can be modified and re-issued if stolen, biometric data is permanent and non-renewable, thus pose great concerns when compromised. Although there is growing concern about the loss of privacy and theft of biometric data, there are very few published research articles on the topic [1, 3, 6, 8]. On a related topic, there are several studies which have shown the vulnerability of biometric authentications systems to spoofing attacks [3, 4, 7].

Tomko proposes the use of biometric data as an encryption key that would be used to encrypt/decrypt his/her PIN number (of which there can be many) [5, 6]. In this way, the fingerprint which uniquely identifies the person is not stored in the database, eliminating any privacy concerns. Indeed, this would be a good solution, however obtaining a unique encryption key from a biometric data, such as a fingerprint, is a challenge. Each impression of a fingerprint for instance is slightly different from another, due to many factors, cut marks, moisture, finger being pressed differently etc., making the task of key generation less than straightforward.

Ratha et al. [3] suggest a framework of cancelable biometrics, where a biometric data undergoes a predefined non-invertible distortion during both enrollment and verification phases; if the transformed biometric is compromised, the user is reenrolled to the system using a new transformation. Likewise, different applications are also expected to use different transformations for the same user. Although this framework hides original (undistorted) biometric and enables revocation of a (transformed) biometric, it introduces the management of transform databases.

In the subsequent section, we propose a biometric authentication framework to alleviate these privacy issues. Although we demonstrate an application of the framework using fingerprints, it can be also generalized to other biometrics.

3. Proposed Method

Each person who enrolls into the system gives two fingerprints, A and B. The minutiae points of these two fingerprints are found and superimposed so that their center of masses are aligned. The obtained combined minutiae list becomes the biometric ID of the person and is stored in the central database. Note that the combined ID can be generated by many different fingerprint pairs, as such, it is *not* a unique identifier of the person.

The enrollment process is shown in Fig. 1, with the combined minutiae image being on the right. Note that in this figure we show the two parts of the combined fingerprint with separate markers for clarity; in fact they should all be marked the same, since they are indistinguishable in the combined list.

Figure 1. Two fingerprints A and B are combined to give the combined fingerprint minutiae on the right. The minutiae points are marked so as to indicate the source finger, but this information is not stored in the database.

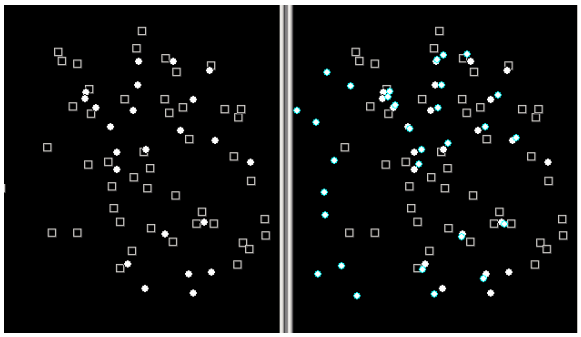


Figure 2. The combined minutiae set is on the left and the registration of the first fingerprint A', shown in blue circles, against the combined fingerprint is shown on the right. The corresponding fingerprints are shown in Fig. 3.

When a person is to be authenticated, s/he gives two fingerprint impressions (A' and B'), *both* of which is used to verify his/her identity. First, one of the fingerprints, say A', is matched against the combined biometric ID, as shown in Fig. 2. Note that even though the minutiae points are marked in the figures with circles and squares so as to indicate their source finger, they are not kept in the combined ID! The matching is done by finding the correspondence between the minutiae of the two fingerprints and the combined fingerprint. Both the minutiae extraction and the point correspondence algorithm are non-essential to the proposed method and any previously developed minutiae detection or correspondence algorithms can be used.

After this first match step, the matched minutiae points are removed from the combined minutiae list, giving

$$A_M + B_M - A'_M$$

where A_M indicates the minutiae list of the fingerprint A, $+$ indicates concatenation and $-$ indicates deletion of *matched*

points. Then, the second fingerprint B' is matched against these remaining minutiae points. The person is authenticated if the ratio of matched minutiae points to the remaining minutiae points left from the combined list plus those from B' is above a certain threshold:

$$score = 2 \times \frac{\left|(A_M + B_M - A'_M) \cap B'_M\right|}{\left|(A_M + B_M - A'_M) + B'_M\right|} \times 100 \quad (1)$$

In case A' matches A perfectly and B' matches B perfectly, the resulting score with this metric is 100. If A' was not successfully matched, it would be reflected in the final score since many minutiae points would be left unmatched, making the denominator large. If B' was not successfully matched, the numerator would be small.

Note that the match rate obtained in the first step is significantly higher than if we just matched the corresponding fingerprints A and A', since the combined ID contains about twice as many minutiae points. In particular, fingerprints with few minutiae points match to several combined fingerprints with large sets of minutiae points. This makes it very difficult to search the combined database using a single fingerprint to find matching records (identification); which is the intended result. On the other hand, it does not reduce the effectiveness of the system: if any minutiae from B are matched by A', it will show in the final score if it matters (if A's and B's minutiae are nearby, it does not matter whose minutiae are matched).

4. Experiments

Four fingerprints (two from one finger and two from another finger) are collected from each of the 100 people contributing to the database. Two of these fingerprints, one from each finger, are added to the reference set: they are used to form the combined ID for the person. The remaining two fingerprints, the second impressions of each finger, are added to the test set: they are used to authenticate the person. Figure 3 shows a quadruple from the database: the top row is the reference set (A and B) and the bottom row is the test set (A' and B'), from left to right. Notice that the fingerprints have many missed minutiae, either due to labeling mistakes, or due to the shifts and deformations in the taking of the imprints.

Once the data is collected, the minutiae points are found and the fingerprint pairs are matched against the stored combined fingerprint, as explained in the previous section. Currently, the minutiae points are marked manually, but the matching is done automatically. However, note that manual labelling of the minutiae points is not essential: any reasonably successful minutiae detection and matching algorithm can be used. The automatic matching is done via a simple

Figure 3. Sample quadruple fingerprints from the database. Top row shows fingerprints A and B; bottom row shows fingerprints A' and B', left to right.

matching algorithm that aligned two point sets by finding the best alignment over all translations and rotations, allowing for some elastic deformation of the fingerprint (accepting two points as matching if they are within a small threshold in this alignment). Since the aim of this paper is to introduce the idea of a combined biometric ID, we only show that the resulting combined ID is non-unique, but that it can still be used to authenticate a person. Hence, minutiae detection and matching were assumed to be given or were simply implemented.

Using the proposed method explained in Section 3, there was a 2% false reject rate (FRR) in the collected database. In other words, 2 out of 100 people in the database were not authorized using their second set of fingerprints (A' and B'). On the other hand, when each of these fingerprint pairs were used as a forgery for all other people (for a total of 9900=100*99 data points), only 1.8% were falsely accepted (FAR). The equal error rate (EER) where FAR and FRR are equal was approximately at 1.9%.

In order to test how much error is introduced with the new authentication scheme (using two fingerprints instead of one), we have calculated the error rate of matching the fingerprints one by one, using the same minutiae detection and matching algorithms. The matching score used was the ratio of the number of matching points over the total number

of points in the matched and the reference fingerprints. For instance, for the A set, it was:

$$score = 2 \times \frac{|A_M \cap A'_M|}{|A_M + A'_M|} \quad (2)$$

In this task, the FRR was found to be 3%: in other words, only 6 fingerprints were falsely rejected out of 200 fingerprints (100x2). When each fingerprint was used as forgery for all the others, the FAR for this test was 2%. Hence, the combined biometric scheme introduced no additional errors, in fact, it reduced the error rate. This should in fact be the case, since we are given more identifying information about the person, however the test have shown that the proposed combination scheme did not hinder verification.

A final test was done to see whether a single fingerprint was sufficient to search the combined fingerprint database (i.e. given only one fingerprint, what are the chances to correctly identify a person?). The scoring method used was based on the proportion of the minutiae points of the presented fingerprint (say A') that matched the template set (A+B). Using this score, the template fingerprint belonging to the correct person gave the highest match in only 24% of the identification tests. When we looked at top-5 results (accepting the person correctly identified if the correct template was in the top-5 alternatives), the identification rate rose to 39% . In large fingerprint databases, a large number of fingerprints would match the template, indicating that both fingerprints are necessary to retrieve someone's records, as intended.

Most of the errors were due to fingerprints that had significant stretching between two instance, as these are not well matched using our simple matching algorithm. Other biggest source of error is fingerprints that have missing left or right parts, due to pressure being applied to one side of the finger while taking the imprint.

4.1. Summary and Conclusions

We have introduced the idea of combining biometrics such that the combined biometric would not be a unique identifier of the person, yet it could still be successfully used for authentication purposes.

We have demonstrated such a system using fingerprints and showed that the authentication error rate is very small (1.9% EER), even with very simple underlying algorithms for minutiae detection and matching. Given that there was actually a decrease in the verification error using the combined biometric, compared to our simple fingerprint verification system (labelled minutiae and simple alignment), we can say that the proposed scheme can be used to increase privacy without hindering the verification process.

We have not actually proven that the combined biometric cannot be used to track a person: it may be possible that a certain pattern of minutiae distribution appears only for a given person. However, the addition of minutiae points from the second fingerprint hides these patterns to the largest extent. In the future, one can further look into how to best combine two biometrics, (e.g. to disperse the minutiae points as much as possible etc.) so as to hide most unique features of a fingerprint. Three separate combined fingerprint minutiae are shown in Fig. 4 to give some idea.

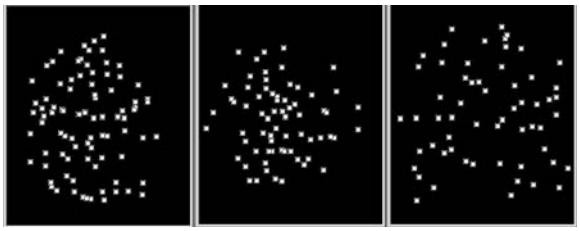


Figure 4. The combined minutiae from 3 different people.

We have collected our own data because we wanted to make sure to have four fingerprints from each user and to have relatively good fingerprints. In future, we will try the same tests with public fingerprint databases, as well as more sophisticated minutiae detection and matching algorithms.

References

[1] M. Crompton. Biometrics and privacy: The end of the world as we know it or the white knight of privacy? In *1st Biometrics Institute Conference*, 2003.

[2] W. H. I. McLean. Genetic disorders of palm skin and nail. *Journal of Anatomy*, 202(1):133–133, 2003.

[3] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[4] S. Schuckers. Spoofing and anti-spoofing measures. *Information Security Technical Report*, 7(4):56–62, 2002.

[5] G. Tomko. Biometrics as a privacy-enhancing technology: Friend or foe of privacy? In *Privacy Laws & Business 9th Privacy Commissioners/Data Protection Authorities Workshop*, 1998.

[6] G. Tomko. Privacy implications of biometrics – a solution in biometric encryption. In *Eighth Annual Conference on Computers, Freedom and Privacy*, 1998.

[7] U. Uludag and A. Jain. Attacks on biometric systems: a case study in fingerprints. In *SPIE-EI 2004, Security, Seganography and Watermarking of Multimedia Contents VI*, 2004.

[8] J. Woodward. Biometrics: Privacy's foe or privacy's friend? In *Proceedings of the IEEE*, volume 85, 1997.

Predicting Fingerprint Recognition Performance from a Small Gallery

Bir Bhanu, Rong Wang and Xuejun Tan
Center for Research in Intelligent Systems
University of California, Riverside
Riverside, California 92521, USA
{bhanu, rwang, xjtan}@vislab.ucr.edu

Abstract

Predicting performance of biometrics is an important problem in a real world application. In this paper we present a binomial model to predict fingerprint recognition performance. We use a fingerprint identification algorithm to find the number of corresponding triangles as the match and non-match scores. Then we use these similarity scores in a binomial prediction model, which uses small gallery to predict performance on a large population. The results on the entire NIST-4 database show that our model can reasonably predict large population performance.

1. Introduction

In order to ensure the high confidence in security biometrics such as ear, face, gait, fingerprint, palm, signature and speech are commonly used. Fingerprint has been used for a long time because of its uniqueness and immutability. Depending on an application there are two kinds of fingerprint recognition systems: verification system and identification system [5]. A verification system will store users' fingerprints as sets of minutiae in the database. Then compare a person's fingerprint with her/his own minutiae set to verify if this person is who she/he claims to be. This is a one to one matching problem. The system can accept or reject this person according to the verification result. An identification system is more complex. For a query fingerprint the system searches the whole database to find out if there are any fingerprint minutiae sets saved in the database that can match it. It conducts one to many matching [5].

How does the fingerprint recognition technique work for large population is often asked in a practical application. In this paper we develop a binomial model to predict large population performance based on small gallery. Firstly we calculate the corresponding triangles between each fingerprint in a probe set with every fingerprint in a gallery. Then we use these corresponding values as similarity scores to estimate the distribution of match and non-match scores. After this we use the Cumulative Match Characteristic (CMC) curve to rank all these scores. CMC curve can show different probabilities of recognizing a fingerprint depending on how similar this query fingerprint to its minutiae set compared with other fingerprints in the gallery [6]. Finally we use a binomial distribution to compute the probability that the match score is within rank r. In this paper we only concern about the performance when the rank is 1. Using this model we can predict fingerprint recognition performance when the database size is increased.

In section 2 the related work is presented, details of fingerprint identification technique and prediction model are given in section 3. In section 4, prediction performance based on NIST-4 is described. Finally in section 5 conclusions are provided.

2. Related work

Fingerprint identification problem can be regarded as the verification performed for the probe image with every gallery image in the database. Additionally indexing followed by verification can solve this problem. In Germain et al. [2], they combine indexing and verification together. Their identification approach is based on triangles. For any three noncolinear minutiae they get a triangle. They use length of each side, ridge count and angles as their features. These features are not robust to distortion. So they undermine the performance [8]. Tan and Bhanu [7] propose another approach to solve identification problem, which is also based on triangles. Their approach has two main differences with Germain's. First one is that they use indexing and verification separately. In the indexing step they get top T hypotheses, then use the verification process to verify these hypotheses. Secondly the features they use are: angles, triangle handedness, triangle direction, maximum side, minutiae density and ridge counts. These features are more robust to distortion than Germain's [8].

Binomial model is very suitable for estimating recognition performance when the database size is large. Until now the prediction models are mostly based on feature space or similarity scores. Johnson et al. [4] build a CMC model that is based on the feature space to predict the gait identification performance. L_2 norm and Mahalanobis distance are used to compute similarity within the feature space. They make an assumption about the density that the population variation is much bigger than the individual variation. Sometimes this assumption is invalid. Wayman [9] and Daugman [1] develop a binomial model that uses the non-match distribution. This model underestimates recognition performance for large galleries. Phillips et al. [6] create a moment model, which uses both the match and non-match distributions. Since all the similarity scores are sampled independently, their results underestimate the identification performance. Johnson et al [3] improve this model by using a multiple non-match scores set. They average match scores on the whole gallery. For each match score they count the number of non-match scores that is larger than this match score, which leads to an error. In reality the distribution of match score is not always uniform.

In this paper we use a binomial model to estimate fingerprint recognition performance for large population. We first estimate the similarity scores distributions and then integrate the non-match distribution according to the match score which can find the probability that the non-match score is larger than the match score. This is different from Phillips' moment model. It can efficiently solve the problem of underestimate recognition performance.

3. Technical approach

We are given two sets of data: gallery and probe. Gallery is a set of fingerprint minutiae saved in the database. For each fingerprint there is one set of minutiae saved in the gallery. Probe is a set of query fingerprints. One finger can have more than one print in the probe set. The fingerprint identification algorithm we used is based on the representation of triangles. For every fingerprint we first extract minutiae. Then randomly choose any three noncolinear minutiae to form a triangle. Thus, one fingerprint can get hundreds of triangles. There are two steps in the identification process: indexing and verification.

3.1. Fingerprint indexing

During the indexing, the features we used to find potential triangles are: minimum angle $\alpha_{\min}$, median angle α_{med}, triangle handedness ϕ, triangle direction η, maximum side λ, minutiae density χ and ridges counts ξ. We compute these features for each fingerprint in the gallery and set up an indexing space $H(\alpha_{\min}, \alpha_{med}, \phi, \eta, \lambda, \chi, \xi)$, the detail explanation of these features can be found in [7].

We compute these features for each query fingerprint and compare them with indexing space H. If the error between them is small enough then we know they are probably the same fingerprint. The output of this process is a list of hypotheses, which are sorted in the decreasing order of the number of potential corresponding triangles. Top T hypotheses are input to the verification process.

3.2. Fingerprint verification

Suppose there are Q and M minutiae in the query and gallery fingerprints respectively. Δ_q and Δ_m are potential corresponding triangles. We assume $F(s, \theta, t_x, t_y)$ is the transformation between query and gallery fingerprints, where s is a scale parameter, θ is a rotation parameter, t_x and t_y are translation parameters. The details of how to estimate the transformation parameters can be found in [7]. If these parameters are less than a threshold then we apply this transformation to the potential corresponding triangles. We compute the distance:

$$d = \arg\min_i \left\{ \left| F\left(\begin{bmatrix} x_{j,1} \\ x_{j,2} \end{bmatrix} \right) - \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix} \right| \right\}$$

where $\{(x_{j,1}, x_{j,2})\}$ and $\{(y_{i,1}, y_{i,2})\}$ are two sets of minutiae in the gallery and query fingerprints, $j = 1,2,...M$ and $i = 1,2,...Q$. If d is smaller than a threshold then we can say that $\{(x_{j,1}, x_{j,2})\}$ and $\{(y_{i,1}, y_{i,2})\}$ are corresponding point. If the number of corresponding points is larger than a threshold then we define Δ_q and Δ_m are corresponding triangles.

3.3. Prediction model

Assume that the size of probe set and gallery are all N. For each fingerprint in the probe set we compute the number of corresponding triangles with every fingerprint in the gallery. The number of corresponding triangles can be used as similarity scores. If we have enough match and non-match scores then we can estimate the Probability Density Function (PDF) of these two distributions. Assume $ms(x)$ and $ns(t)$ represent the distribution of match scores and non-match scores respectively. If the

similarity score is higher then the fingerprints are more similar. The error occurs when any given match score is smaller than the non-match scores. The probability that the non-match score is larger than the match score x is $NS(x)$, where

$$NS(x) = \int_{x}^{\infty} ns(t)dt \tag{1}$$

We rank all the similarity scores in decreasing order. The probability that the match score rank r is given by the binomial probability distribution:

$$C_{r-1}^{N-1}(1-NS(x))^{N-r}\,NS(x)^{r-1} \tag{2}$$

N is the gallery size. Integrating over all the match scores, we get the probability that all the match scores rank r is:

$$\int_{-\infty}^{\infty} C_{r-1}^{N-1}(1-NS(x))^{N-r}\,NS(x)^{r-1}\,ms(x)dx \tag{3}$$

In theory the match scores can be any value within $(-\infty,\infty)$. Finally the probability that all the match scores are within rank r is:

$$P(N,r) = \sum_{i=1}^{r}\int_{-\infty}^{\infty} C_{i-1}^{N-1}(1-NS(x))^{N-i}\,NS(x)^{i-1}\,ms(x)dx \tag{4}$$

Here we assume that the match scores and non-match scores are independent and their distributions are the same for all the fingerprints in the gallery. For the identification problem we only consider the situation where rank $r=1$ because this can evaluate the performance of identification technique. Then this model becomes:

$$P(N,1) = \int_{-\infty}^{\infty}(1-NS(x))^{N-1}ms(x)dx \tag{5}$$

In this model N is the size of large population whose performance needs to be estimated. Small size gallery is used to estimate the distribution of $ms(x)$ and $ns(t)$.

4. Experimental results

All the fingerprints we use in the experiments are from NIST Special Database 4 (NIST-4). There are 2000 pairs of fingerprints, each of them is labeled 'f' and 's' that represent different impressions of a fingerprint followed by an ID number. Since the fingerprints in NIST-4 are collected by an ink-based method, many fingerprints are of poor quality and some of them even contain characters and handwritten lines. The size of fingerprint image is 480×512 pixels and resolution is 500 DPI.

We choose all these 2000 fingerprints. 'f' images are the gallery and 's' images are the probe set respectively. Matching all these fingerprints pairs we get 2000 match scores. Then we randomly select 20 fingerprints from the gallery and another 20 different fingerprints from the probe set. We match them and obtain 20 non-match scores. Repeat this process for 100 times then we get 2000 non-match scores. Distributions of these 2000 similarity scores are showed in Figure 1 and Figure 2. If the match score is less than a threshold T_m then we believe the fingerprints pair does not match. Since 99.95% non-match scores are less than 12 we choose $T_m = 12$. Using this threshold we can compute the probability of correct verification.

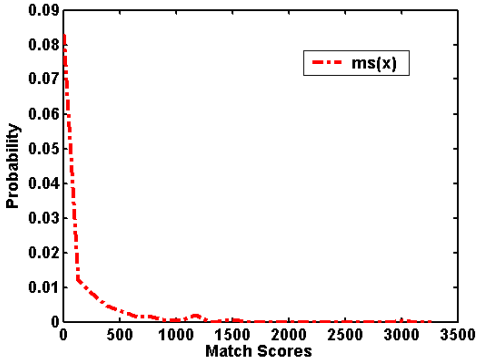


Figure 1. Match scores distribution

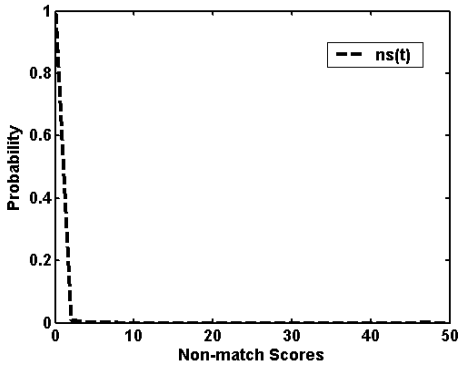


Figure 2. Non-match scores distribution

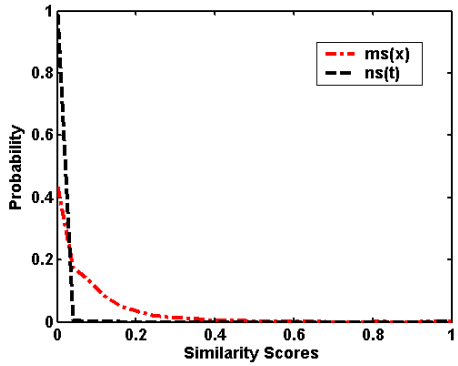


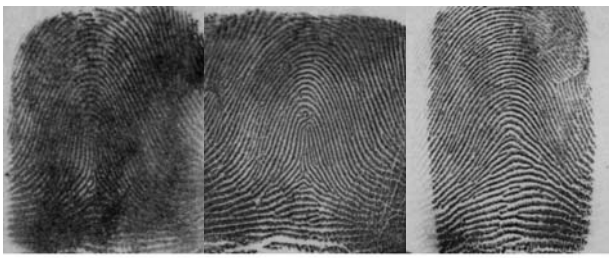
Figure 3. Similarity scores distributions

We randomly choose 40 and 50 fingerprints separately from NIST-4 to be our small gallery to predict the

fingerprint recognition performance for the large population. So the sizes of small gallery are $n = 40$ and $n = 50$, the size of the probe set is the same as small gallery size. We use the verification technique to compute the similarity scores. Figure 3 shows the distributions of match and non-match scores when $n = 50$. Sample results are shown in Table 1. The values on the diagonal are match scores, off diagonal values are non-match scores. Usually match scores should be larger than non-match scores. For fingerprint s0026_03 the match score is 0, while the non-match score between s0026_03 and f0006_09 is 3, obviously this is not correct. Figure 4 shows these three fingerprints from NIST-4. The quality of s0026_03 is not good. It could not find any corresponding triangle with f0026_03, while it has 3 corresponding triangles with f0006_09.

Table 1. Similarity scores for sample image pairs

	s0031_02	s0006_09	s0015_01	s0026_03
f0031_02	810	4	0	0
f0006_09	0	719	0	3
f0015_01	0	0	106	0
f0026_03	0	0	0	0



s0026_03 f0026_03 f0006_09

Figure 4. Three fingerprints from NIST-4

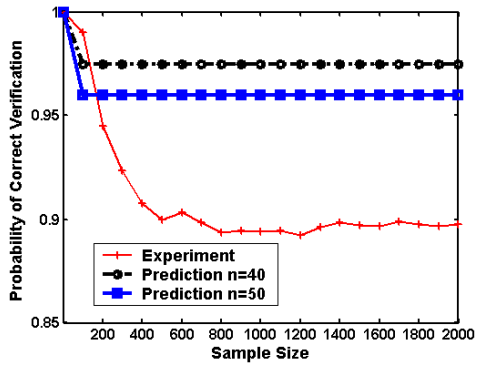


Figure 5. Experimental and prediction performance

Figure 5 shows the experimental and prediction performance results. We use different size of small galleries to estimate fingerprints verification performance on large sample images. We can see that the size of small gallery has effect on the prediction performance. The error reduces with the increase in sample size. So this model can use to predict large population performance.

5. Conclusions

In this paper we use a fingerprint identification algorithm to find the match and non-match scores. We use these scores in a binomial prediction model. The assumption we make for this model is that the match and non-match scores are independent and their distributions are the same for all the fingerprints in the gallery. Based on the results shown in this paper we find that our model can be used to predict large population performance.

References

[1] J. Daugman, Biometric decision landscapes. University of Cambridge computer laboratory, *Technical Report No. TR482*, 2000.

[2] R. S. Germain, A. Califano, and S. Colville, Fingerprint matching using transformation parameter clustering. *IEEE Computational Science and Engineering*, 4(4), pp. 42-49, 1997.

[3] A. Y. Johnson, J. Sun and A. F. Boick, Using similarity scores from a small gallery to estimate recognition performance for large galleries. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 100-103, 2003.

[4] A. Y. Johnson, J. Sun, and A. F. Bobick, Predicting large population data cumulative match characteristic performance from small population data, *the 4th International Conference on AVBPA*, pp. 821-829, June 2003.

[5] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition, *Springer (New York),* 2003.

[6] P. J. Phillips, P.Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, Face Recognition Vendor Test 2002, *Evaluation Report*, March 2003.

[7] X. Tan, and B. Bhanu, Robust Fingerprint Identification. *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 1, pp. 277-280, 2002.

[8] B. Bhanu, X. Tan, Fingerprint indexing based on novel features of minutiae triplets, *IEEE Trans. Pattern Analysis and Machine Intelligence* ***(PAMI)***, 25(5), pp. 616-622, 2002.

[9] J. L. Wayman, Error-rate equations for the general biometric system. *IEEE Robotics & Automation Magazine*, Vol. 6, issue 1, 35-48, 1999.

Personal Identification by Cross Ratios of Finger Features

Gang Zheng[1*] Chia-Jiu Wang[1] Terrance E. Boult[2]
[1]Department of Electrical and Computer Engineering
[2]Department of Computer Science
University of Colorado at Colorado Springs
Colorado Springs, CO 80933-7150 USA

Abstract

In this paper, a touch-free non-intrusive technique of extracting personal hand geometry biometric data is proposed and analyzed. This technique is based on the invariant property of cross ratios of hand features. The descriptor of each hand consists of two-dimensional and one-dimensional cross ratios of the locations of the finger creases. Preliminary results, using a 16 dimensional feature space, show a 100% hit rate on 465 pairwise comparisons, from a database of 14 subjects.

1. Introduction

Effective personal identification technology has increase in importance over the past decade. The most advanced systems require three independent factors: something you have, something you know, and something you are. Many biometric-based identification techniques are available today [2]. Many are still the subjects of research. As application variations results, no known biometric system is ideal. Non-contact non-invasive but accurate system is the general goal. In many cases, the effectiveness is limited by pose variations during imaging or artificial constraints use to reduce the possible poses. In this paper we explore, for the first time, the use of formal projective image invariants to define biometrics.

Hand-based identification techniques have been investigated for decades, but received less attention than other techniques, mostly due to its low reliability, imaging difficulties, and somewhat intrusion. In this paper we introduced the possibility of a robust biometric identification technique using well-known projective invariants of image features of a hand. The hand images are taken by a standard digital camera. Thus, it is non-contact and non-aggressive, while subject to no major pose constraints. The projectively invariant hand features make it more sophisticated and more reliable. Although the experimental results are very preliminary, they may open the door to a new exploration of the projectively invariant biometric-based identification area. And the concept of the projectively invariant biometric may be applied to other biometric identification technologies.

In Section 2, a review of existing technologies is given. The fundamental concepts of the cross ratio are presented in Section 3. The application of cross ratios in hand geometry for personal identification is described in Section 4. Section 5 summarizes the preliminary experimental results. Discussions and conclusions are presented in Section 6.

2. Review of existing work

Hand geometry identification techniques are suitable for applications of moderate security level [2,3,4,5,6]. It has been employed at over 8,000 locations, including the Columbian Legislature, San Francisco International Airport, day care centers, a sperm bank, etc [7].

A. Jain et al [4] developed a hand geometry based verification system. Various features including widths of the fingers, lengths of the fingers, widths of the palm, as well as heights of the fingers and the palm were measured. In order to obtain consistent positions of a hand to be measured, they used five pegs to guide the placement of user's hand on a flat surface of the imaging device. They reported a false acceptance rate (FAR) of 2% and a false rejection rate (FRR) of 15% in a hand geometry based web access system. A. Jain et al [8] developed an algorithm that aligns hand contours and measures the mean alignment error to determine the distinction between two hands. They used the similar five-peg device as in [4] to restrict the motion of user's hand.

R. Sanchez-Reillo et al [3,5] developed a biometric verification system based on hand geometry. Different pattern recognition techniques were used in their work. The input features of the system were geometrical sizes of a hand similar to those in [4]. Pegs were also used on the

*This work is funded in part by Air Force Research Lab under agreement number F49620-03-1-0207 through Network Information and Space Security Center (NISSC), U. Colorado at Colorado Springs. It is also funded in part by DARPA HID program under contract number N00014-00-1-0929, and by the Colorado Institute for Technology Transfer and Implementation

image acquisition device. Using Gaussian mixture model, an error rate of 4.9% was reported.

Although most of these research works achieved valuable success, a common shortcoming exists: the aforementioned work requires the user to place his/her hand on a pegged flat surface. This introduced several problems. First, pegs can deform the shape of the hand. Second, improper placement of a hand can still happen due to the relatively complicated instruction, which will reduce the reliability of the system [9]. Finally it is intrusive and some people don't like to place hands on where others just put their hands [2]. Research has continued to solve these problems.

A. Wong et al [9] presented a peg-free hand geometry system. The peg-free improvement was achieved by using a flatbed optical scanner for hand image capturing. Significant geometrical landmarks of a hand were extracted as the recognition features. They achieved 89% hit rate and 2.2% FAR. Y. Bulatov et al [6] used optical scanner in their work too. Besides the typical sizes of hands, they also measured the areas of some particular regions of hands. Their work obtained an FAR of 1% and FRR of 6%. However, both works still required contact-based imaging devices that are objectionable and somewhat inconvenience.

A. Kumar et al [10] didn't use the regular scanners. They took the pictures of the palm side of a hand using a digital camera. But the hand geometry identification had to be combined with palmprint recognition techniques. And users still had to touch the device. If the hand pose is not constrained, however, the viewpoint change produces perspective projective distortion. Therefore, geometrical measurements cannot be used.

3. Projective invariants and cross ratios

Projective geometry does not preserve Euclidean distances, angles or ratios of distances/angles. Our research seeks features that are projective invariant yet discriminate individuals. The cross ratio, which is the ratio of ratios of some particular properties, is invariant to the projective distortion. We briefly recall the definitions.

The one-dimensional cross ratio τ, is defined in terms of the distances between four collinear points [11, 12, 13, 14]. Let four collinear points A, B, C, and D on line l. Points A', B', C', and D' are their projections on line l'. Then, the 1-D cross ratio can be defined as follows:

$$\tau = \frac{\|AC\|\|BD\|}{\|AD\|\|BC\|} = \frac{\|A'C'\|\|B'D'\|}{\|A'D'\|\|B'C'\|} \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean distances between two points.

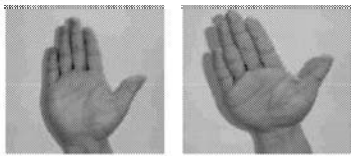
Under projective transformations neither area of planar object nor angle established by two intersecting lines is preserved. But, the cross ratios created by five coplanar points or four coplanar concurrent lines are invariant [11, 12]. On a projective plane π, points A, B, C, and D are on the lines a, b, c, and d, respectively. Point O is the intersecting point of lines a, b, c, and d. No three of these five points, including O, should be collinear. On the other hand, plane Π is the projective transformation of the plane π. Thus, points set A', B', C', D', and O' are the projections of points A, B, C, D, and O. The 2-D cross ratios of two five-point sets in Figure 2 can be computed by Equation (2) [12].

$$\tau = \frac{\sin(\angle AOC)}{\sin(\angle AOD)} \cdot \frac{\sin(\angle BOD)}{\sin(\angle BOC)} = \frac{\sin(\angle A'O'C')}{\sin(\angle A'O'D')} \cdot \frac{\sin(\angle B'O'D')}{\sin(\angle B'O'C')} \quad (2)$$

where $\angle\cdot$ denotes the angle created by three points.

4. Hand identification by cross ratios

In our work, we use a digital camera as the imaging device. We request a user join his/her fingers as closely when taking the hand pictures and keep their hand stretched flat. Thus the hand surface composed of the fingers and the palm will become a rigid two-dimensional plane. We do grant users the flexibility to pose their hands as they like, as long as major part of the hand plane is exposed to the camera. Therefore, the pictures of this hand plane taken from different viewpoints can be regarded as the projections of the original hand plane onto arbitrary projective planes. Figure 1 shows two examples of such pictures

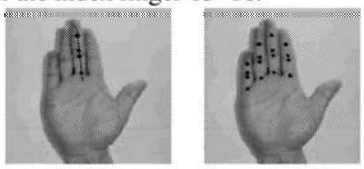


(a) (b)

Figure 1. Hand pictures from different viewpoints

4.1. Extraction of the feature point/sets

In order to acquire the 1-D cross ratios from a hand, four collinear points have to be extracted. One can always find a straight line going through a finger. Then the intersecting points of this line and the crease lines on the finger give the collinear points. This is shown in Figure 2(a). We extracted four points from each finger except the thumb. The four fingers give us 16 feature points in total. This is shown in Figure 2(b). We denote the four points from top to bottom on the pinky finger as 1~4, and those on the ring finger 5~8, those on the middle finger 9~12, and those on the index finger 13~16.



(a) (b)

Figure 2. Feature points on the fingers

The way generating the four feature points on each finger qualifies them for computing the 1-D cross ratios. Thus we get four 1-D cross ratios, one from each finger,

as shown in Figure 3 with the advantage that these features are invariant to individual finger movement.

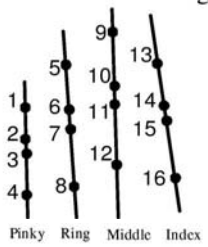


Figure 3. Collinear points layout of 1-D cross ratios

Besides four 1-D cross ratios, we are going to use 2-D cross ratios as well since our object is a 2-D hand plane. Thus, the size of the feature vector will be increased, and the robustness of the technology will be improved. The 2-D cross ratio is computed based on two requirements. The first is a coplanar five-point set. And the second is that no three points which include the reference point *O* should be collinear. We will use the same 16 feature points and find subsets satisfying the requirements. The first requirement is automatically satisfied because all the feature points are extracted from the hand plane. There are plenty of qualified five-point sets within the 4096 possible combinations of 16 points. At this stage of our research, we have selected only 12 five-point layouts as shown in Figure 4. We obtain one 2-D cross ratio from each layout. In Figure 4, each five-point layout is corresponding to the five points *O*, *A*, *B*, *C*, and *D*. The points used in each layout are: (a) 6, 5, 9, 13, 14; (b) 12, 7, 5, 13, 15; (c) 9, 13, 14, 6, 5; (d) 15, 6, 5, 9, 13; (e) 10, 9, 13, 14, 15; (f) 6, 5, 13,14, 15; (g) 9, 13, 10, 6, 5; (h) 10, 7, 6, 5, 9; (i) 12, 11, 13, 14, 15; (j) 7, 5, 9, 10, 11; (k) 14, 12, 10, 9, 13; (l) 15, 7, 6, 5, 13

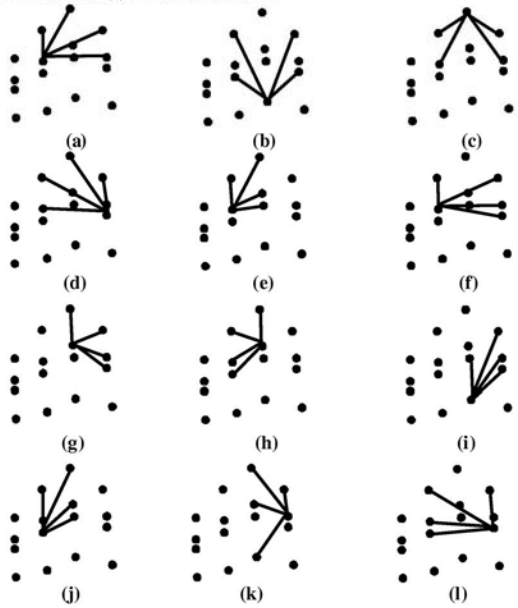


Figure 4. Twelve five-point layouts for 2-D cross ratios

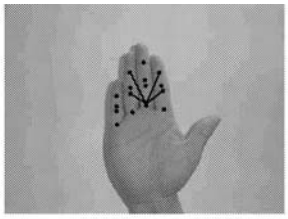


Figure 5. A qualified five-point set on the hand plane

4.2. Feature vector computation

The hand descriptor is represented by a feature vector which is composed of both 1-D and 2-D cross ratios. Thus the feature vector V has a size of sixteen since sixteen cross ratios were computed.

$$V = \{\Psi_1, \Psi_2, \ldots \Psi_N\}, \qquad N = 16 \tag{3}$$

where Ψ_i represents the 2-D cross ratios when i=1~12; Ψ_i represents the 1-D cross ratios when i=13~16. The distance, D, between two feature vectors of two hand images is calculated using Equation (4).

$$D = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\frac{\Psi_i^j - \Psi_i^k}{\mu_i}\right)^2}, \quad N = 16 \tag{4}$$

where Ψ_i^j and Ψ_i^k are the ith cross ratios of the feature vectors of hand j and hand k, and

$$\mu_i = \frac{\Psi_i^j + \Psi_i^k}{2}$$

5. Experimental results

At the early stage of our research, we collected a total of 31 hand pictures from 14 persons. We gave each person an ID number, from 1 to 14, respectively. We also index the pictures. In the table, Pi,j represents the jth picture of Person#i. We tested 465 different pairs, one against another, with fairly promising results. Table 1~4 shows part of our experimental results. The table entries indicate the distances, scaled by 1000, between the feature vectors of two hand images.

Table 1. Pairwise comparison of 10 pictures of 4 persons.

Picture ID / Distance (×1000)		Person #1		Person #2				Person #3		Person #4	
Picture ID		P1,1	P1,2	P2,1	P2,2	P2,3	P2,4	P3,1	P3,2	P4,1	P4,2
Person#1	P1,1	0	9.01	23.56	26.10	25.58	20.85	36.52	38.28	13.59	15.99
	P1,2	9.01	0	24.42	27.45	26.04	21.26	33.10	35.18	14.05	16.44
Person#2	P2,1	23.56	24.42	0	7.95	7.22	9.26	25.43	27.61	15.71	13.69
	P2,2	26.10	27.45	7.95	0	5.33	8.03	23.88	25.29	16.94	14.60
	P2,3	25.58	26.04	7.22	5.33	0	6.09	22.85	24.57	16.50	14.19
	P2,4	20.85	21.26	9.26	8.03	6.09	0	22.93	24.46	12.01	10.12
Person#3	P3,1	36.52	33.10	25.43	23.88	22.85	22.93	0	5.24	26.38	25.10
	P3,2	38.28	35.18	27.61	25.29	24.57	24.46	5.24	0	27.86	26.48
Person#4	P4,1	13.59	14.05	15.71	16.94	16.50	12.01	26.38	27.86	0	3.06
	P4,2	15.99	16.44	13.69	14.60	14.19	10.12	25.10	26.48	3.06	0

Table 1 shows that the pictures from the same person have small distance, and conversely, the pictures from different persons have large distances.

Table 2. Pairwise comparison of different pictures

Picture ID / Distance (×1000) / Picture ID		Person #5			Person #7		Person #9		Person #11	
		P5,1	P5,2	P5,3	P7,1	P7,2	P9,1	P9,2	P11,1	P11,2
Person#5	P5,1	0	8.63	6.16	30.94	32.60	19.39	20.65	21.42	23.84
	P5,2	8.63	0	5.02	35.14	37.35	19.93	21.26	23.13	23.89
	P5,3	6.16	5.02	0	32.27	34.59	18.10	20.12	20.59	21.99
Person#7	P7,1	30.94	35.14	32.27	0	11.69	26.36	27.60	27.03	32.90
	P7,2	32.60	37.35	34.59	11.69	0	27.84	28.38	29.40	35.12
Person#9	P9,1	19.39	19.93	18.10	26.36	27.84	0	4.71	15.38	17.87
	P9,2	20.65	21.26	20.12	27.60	28.38	4.71	0	18.05	20.24
Person#11	P11,1	21.42	23.13	20.59	27.03	29.40	15.38	18.05	0	7.65
	P11,2	23.84	23.89	21.99	32.90	35.12	17.87	20.24	7.65	0

Table 3. Pairwise comparison of different pictures

Picture ID / Distance (×1000) / Picture ID		Person #6		Person #8		Person #10		Person #12	
		P6,1	P6,2	P8,1	P8,2	P10,1	P10,2	P12,1	P12,2
Person#6	P6,1	0	6.16	31.79	33.50	35.88	33.93	53.28	53.31
	P6,2	6.16	0	31.94	33.28	37.25	35.45	52.68	52.61
Person#8	P8,1	31.79	31.94	0	5.74	15.25	17.18	28.04	28.52
	P8,2	33.50	33.28	5.74	0	17.00	18.76	27.53	28.21
Person#10	P10,1	35.88	37.25	15.25	17.00	0	4.68	35.40	36.56
	P10,2	33.93	35.45	17.81	18.76	4.68	0	37.83	39.16
Person#12	P12,1	53.28	52.68	28.04	27.53	35.40	37.83	0	3.54
	P12,2	53.31	52.61	28.52	28.21	36.56	39.16	3.54	0

Table 4. Pairwise comparison of different pictures

Picture ID / Distance (×1000) / Picture ID		Person #3		Person #6		Person #10		Person #14	
		P3,1	P3,2	P6,1	P6,2	P10,1	P10,2	P14,1	P14,2
Person#3	P3,1	0	5.24	35.33	33.83	27.12	25.80	38.00	35.01
	P3,2	5.24	0	35.97	34.65	28.36	27.33	38.80	35.89
Person#6	P6,1	35.33	35.97	0	6.16	35.88	33.93	32.62	33.40
	P6,2	33.83	34.65	6.16	0	37.25	35.45	32.43	32.68
Person#10	P10,1	27.12	28.36	35.88	37.25	0	4.68	27.18	27.65
	P10,2	25.80	27.33	33.93	35.45	4.68	0	29.29	29.68
Person#14	P14,1	38.00	38.80	32.62	32.43	27.18	29.29	0	7.16
	P14,2	35.01	35.89	33.40	32.68	27.65	29.68	7.16	0

Different tests were done based on various thresholds. When the threshold drops down to 0.007, FAR becomes zero; at the same time, FRR equals 2.8%. The Equal Error Rate (ERR) happens when the threshold is set around 0.012.

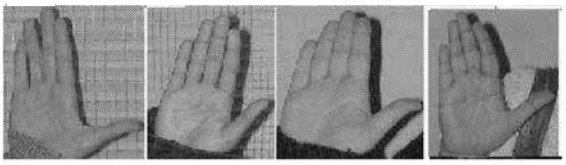


Figure 6: More sample hand images

6. Conclusion

This paper introduced the approach of using projective invariant biometrics and investigated a new approach to hand geometry identification using hand descriptors based on cross ratios. The hand descriptor built by this technique is invariant under projective transformations. The method is peg-free, touch-free, and less intrusive than existing hand-geometry approaches. Preliminary results show a high potential of success using this technique in applications of high security level. While only a 16 dimensional space was used, thousands of projective invariant features are describe and could be used to further discriminate individuals.

7. References

[1] D. Zhang, W. Kong, and J. You, "Online Palmprint Identification", *IEEE trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, 2003, pp. 1041-1050.

[2] Z. Riha, V. Matyas, "Biometric Authentication Systems", *Faculty of Informatics Masaryk University (FIMU) Report Series*, FIMU-RS-2000-08, 2000.

[3] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric Identification through Hand Geometry Measurements", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, 2000, pp. 1168-1171.

[4] A.K. Jain, A. Ross, and S. Pankanti, "A Prototype Hand Geometry-based Verification System", *Proc. Of 2nd Int'l Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, 1999, pp.166-171.

[5] R. Sanchez-Reillo, "Hand Geometry Pattern Recognition Through Gaussian Mixture Modelling", *Proc. of the Int'l. Conf. on Pattern Recognition*, 2000, Vol. II, pp. 941-944.

[6] Y. Bulatov, S. Jambawalikar, P. Kumar, S. Sethia, "Hand Recognition Using Geometric Classifiers", *DIMACS Workshop on Computational Geometry*, 2002.

[7] E. Bowman, "Everything You Need to Know About Biometrics", from http://www.ibia.org, 2000.

[8] A.K. Jain, and N. Duta, "Deformable Matching of Hand Shapes for Verification", *Proc. of IEEE Int'l. Conf. on Image Processing*, 1999, pp. 857-861.

[9] A. Wong, and P. Shi, "Peg-free Hand Geometry Recognition Using Hierarchical Geometry and Shape Matching", *IAPR Workshop on Machine Vision Applications*, 2002, pp. 281-284.

[10] A. Kumar, D. Wong, H.C. Shen, A.K. Jain, "Personal Verification using Palmprint and Hand Geometry Biometric", *Proc. of 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 2003, pp. 668-678.

[11] C. E. Springer, "*Geometry and Analysis of Projective Spaces*", W.H. Freeman and Company, 1964.

[12] J.L. Mundy, A. Zisserman, "*Geometric Invariance in Computer Vision*", The MIT Press, London, England, 1992.

[13] C.A. Rothwell, "*Object Recognition Through Invariant Indexing*", Oxford University Press, 1995.

[14] Olivier Faugeras, "*Three-Dimensional Computer Vision*", The MIT Press, London, England, 1993.

Efficacy of joint person subset bootstrap estimation of confidence interval

Ruud M. Bolle, Nalini K. Ratha and Sharath Pankanti
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
{bolle, ratha, sharat}@us.ibm.com

Abstract

It is widely known that robust and practical estimation methods for non i.i.d. distributions followed by real data is a challenging problem [7]. While, generally, nonparametric methods (e.g., bootstrap [2]) are better suited for estimation for such distributions, recent studies have demonstrated that the subset bootstrap methods [4] better model the dependencies among the data. While many innovative ways construction of subsets for modelling the dependencies among data have been proposed in the literature, subset constructions to study dependencies among the different dependent estimates (e.g., FAR and FRR) have not been yet explored. The objective of this study to propose a method of subset construction to jointly model dependency among FAR and FRR estimates from the data and examine the magnitude of this dependency. We demonstrate our approach using a real dataset of the fingerprints and discuss the implications of our experimental results.

1. Introduction

Realistic performance evaluation is a challenging problem [1, 3, 6, 5]. It is conventionally assumed that the biometric data being sampled is i.i.d. and therefore, any violation of such assumption would result in inaccurate estimation results (e.g., error confidence intervals). In realistic biometric datasets, there is always significant dependence among the data[7]. For example, the match scores generated from fingerprint impressions of a finger are not independent. Similarly, the match scores of involving different fingers of a person may be dependent. Elsewhere in the literature [7], it has been demonstrated that typical biometric test data exhibits significant dependencies among its component entities and error (e.g., FAR, FRR) confidence interval estimation needs to take into account the dependencies among the sample data. A *subset* bootstrap method is typically used to model the dependency among the data nonparametrically to arrive at more realistic confidence intervals.

While the results of subset bootstrap have effectively demonstrated that there is statistical dependence among the match scores associated with a biometric (e.g., left index finger) of a subject and with different biometrics (e.g., left index and left middle finger) of a single subject, there are no studies examining the dependence among the match (e.g., mated) scores of a subject and the non-match scores associated with a biometric of a subject. The objective of this paper is to explore this dependency and characterize the extent of this dependency.

The rest of this paper is organized as follows. Section 2 introduces terminology and confidence intervals. Sections 3 and 4 presents the methodologies for estimating confidence intervals using person subset and joint person subset bootstrap methods. Section 5 presents the experimental methodology used to test the accuracies of the confidence interval estimates. We also present the data used for the experiments and the experimental results in Section 5. In Section 6, we discuss the implications of our results and conclusions.

2. Confidence Intervals for Error Estimates

Suppose we have a database DB of biometric samples acquired from $\mathcal{D}$ biometrics (meaning, these are real-world biometrics, $\mathcal{B}_1, ..., \mathcal{B}_\mathcal{D}$) from which d samples are acquired per biometric. The number $\mathcal{D}$ of biometrics $\mathcal{B}_i$, $i = 1, ..., \mathcal{D}$ may be larger than the number of subjects $\mathcal{P}$ that are used to collect the samples, since people may have more than one of the particular biometric (e.g., finger). In any case, the database contains $d\,\mathcal{D}$ biometric samples, and given a *biometric match engine*, one can compute the test score sets: a set of genuine (match) scores $\mathbf{X} = \{X_1, X_2, ..., X_M\}$ and a set of imposter (mismatch) scores $\mathbf{Y} = \{Y_1, Y_2, ..., Y_N\}$.

Matching mated pairs in DB, i.e., matching samples from the same biometric, gives the sample match score (genuine score) set $\mathbf{X}$; matching samples in DB from different identities (or biometrics) gives the mismatch (imposter) score set $\mathbf{Y}$. In this work, for concreteness sake, we focus on fingerprint databases and fingerprint matchers.

A biometric match engine is in theory completely specified by its $F(s)$, the genuine score distribution, and its $G(s)$,

the imposter score distribution. Equivalently, the biometric matcher is completely specified by $\mathrm{FRR}(T)$ and $\mathrm{FAR}(T)$.

These error rates and probability distributions are related: The false reject rate $\mathrm{FRR}(T)$ is the probability of falsely rejecting a genuine subject, it is the probability $Prob(s \leq T|H_o = true)$. The match score distribution $F(s)$ is defined as $F(s) = Prob(X \leq s|H_o = true)$. Hence, it is seen that $\mathrm{FRR}(x) = F(x)$. The false accept rate $\mathrm{FAR}(T)$ is the probability of falsely accepting a subject, i.e., $Prob(s > T|H_a = true)$. The probability distribution of mismatch scores $G(s) = Prob(Y \leq s|H_a = true)$; i.e., $1 - G(s)$ is $Prob(Y > s|H_a = true)$. We have $\mathrm{FAR}(y) = 1 - G(y)$ and the false accept rate FAR is *not* a probability distribution.

We will only be able to estimate these FAR and FRR error rates within a certain $(1-\alpha)100\%$ range, or confidence interval. Here α is the probability that the true value of the FAR or the FRR are outside the respective confidence intervals. Let us first concentrate on estimating characteristics of the match score distributions F. The mean is one such characteristic of F that can be estimated from $\mathbf{X}$; another characteristic of F that can be estimated from $\mathbf{X}$ is the value of the distribution at x_o, $\hat{F}(x_o)$, this gives the estimate of $\mathrm{FRR}(T)$ at $T = x_o$. For example, the point estimate of F at x_o is given by

$$\begin{aligned} \hat{F}(x_o) = &\quad \mathrm{FRR}'(x_o) = \tfrac{1}{M}\textstyle\sum_{i=1}^{M} \mathbf{1}\,(X_i \leq x_o) \\ = &\quad \tfrac{1}{M}\#(\,X_i \leq x_o). \end{aligned} \tag{1}$$

3. Person Subset Bootstrap Method of Estimation

The bootstrap sampling implicitly assumes that the data being sampled is i.i.d. and therefore, any violation of such assumption would result in inaccurate confidence intervals. In realistic (biometric) datasets, there is always significant dependence among the data. For example, the match scores generated from fingerprint impressions of a finger are not independent. Similarly, the match scores of involving different fingers of a person may be dependent. The subset bootstrap technique recommends that instead of resampling (with replacement) the individual datum samples in the dataset, the data be divided into *subsets* and the resampling process should sample the *entire* subsets. The number and constitution of subsets plays an important role in the estimation of confidence intervals. Depending upon the magnitude of independence of each sample subset (w.r.t. other sample subsets), subset bootstrap resampling will be able to propagate the dependence in the data; consequently the confidence intervals estimated using subset bootstrap will be more realistic than the conventional bootstrap.

Taking match (e.g., mated) scores of fingerprints as a concrete example, the literature [] describes three different types of subset bootstrap sampling. First, each match score constitutes a (singleton) subset in itself. This is conventional bootstrap. In second case, one divide the match scores into $\mathcal{PD}$ subsets such that each subset contains match scores resulting from a single finger. This is referred to as finger subset bootstrap. Finally, $\mathcal{P}$ subsets are constructed such that each subset consists of match scores involved with a single person only. This method of bootstrap is referred to as person subset bootstrap. Since the subsets in person bootstrap are relatively more independent than those in finger subset bootstrap, one expects that person susbset bootstrap should be able to better estimate the FRR confidence intervals. Similarly, finger and person subsets should be able to estimate confidence intervals better than the conventional bootstrap. The subset method can be extended to non-match scores and FAR confidence intervals as well. For both FAR and FRR confidence interval estimation, person subset bootstrap is most effective estimation method; we will therefore focus our discussion to person subset bootstrap for the rest of the section.

Let the biometric dataset consist of scores from $\mathcal{P}$ identities (e.g., persons). In person subset bootstrap, the sets $\mathbf{X}$ and $\mathbf{Y}$ are subdivided as

$$\begin{aligned} \mathbf{X} &= \{\mathcal{X}_i,\ i = 1, ..., \mathcal{P}\} \ \text{ and} \\ \mathbf{Y} &= \{\mathcal{Y}_j,\ j = 1, ..., \mathcal{P}\}, \end{aligned} \tag{2}$$

respectively, where the subsets $\mathcal{X}_i,\ i = 1, ..., \mathcal{P}$ and subsets $\mathcal{Y}_j,\ j = 1, ..., \mathcal{P}$ are more or less independent. Here $\mathcal{P}$ is the number of volunteers and, therefore, the number of score subsets, since each volunteer is associated with a subset.

Let us assume the set $\mathbf{X}$ can be divided into $\mathcal{K}$ subsets $\mathbf{X} = \{\mathcal{X}_1, ..., \mathcal{X}_\mathcal{K}\}$.

A person subset bootstrap estimate (see [2]) of a $(1-\alpha)100\%$ confidence interval for the estimate $\hat{F}(x_o)$ is obtained as follows:

1. Divide the set of match scores $\mathbf{X}$ into $\mathcal{K}$ subsets $\mathcal{X}_1, ..., \mathcal{X}_\mathcal{P}$.
2. Many (B) times do:
   (a) Generate a bootstrap set $\mathbf{X}^\star$ by sampling $\mathcal{P}$ subsets with replacement from $\mathbf{X} = \{\mathcal{X}_1, ..., \mathcal{X}_\mathcal{P}\}$.
   (b) Compute the bootstrap estimate $\hat{F}^\star$ as
   $$\hat{F}^\star(x_o) = \frac{1}{M} \sum_{X_i \in \mathbf{X}^\star} \mathbf{1}\,(X_i \leq x_o).$$
   This gives the set $\mathbf{F}^\star(x_o) = \{\hat{F}^\star_k(x_o), k = 1, ..., B\}$ of B bootstrap estimates.
3. Rank the estimates in $\mathbf{F}^\star(x_o)$:
   $$\begin{aligned} \mathbf{F}^\star_O(x_o) = &\quad \{\hat{F}^\star_{(1)}(x_o) \leq \hat{F}^\star_{(2)}(x_o) \leq ... \\ &\quad \leq \hat{F}^\star_{(B)}(x_o)\}. \end{aligned}$$

4. Eliminate the bottom $(\alpha/2)100\%$ and the top $(\alpha/2)100\%$ of estimates $\hat{F}^{\star}_{(k)}(x_o)$. The leftover set of estimates $\mathbf{F}^{\star\star}(x_o)$ with $B' = (1-\alpha)B$ elements gives the $(1-\alpha)100\%$ confidence interval for $\hat{F}(x_o)$.

The subset bootstrap confidence interval estimation concepts can be extended to the non-match scores as well in a straightforward fashion with one exception. Since the non-match scores involve two different fingers, it turns out that completely independent datasets cannot constructed without sacrificing portions of non-match scores. So, there is an option of either using all of the non-match score data and tolerating some amount of dependence among finger and person subsets or using very little fraction of the non-match score data while ascertaining subset data independence. Typically, the former option is considered more desirable.

4. Joint Person Bootstrap Method of Estimation

Until now B bootstrap sets like $\{\mathbf{X}^{\star}_1, ..., \mathbf{X}^{\star}_B\}$ have been generated in isolation. The set $\mathbf{X}$ is divided up into $\mathcal{M}$ subsets $\mathcal{X}_i$ and the bootstrap sets $\mathbf{X}^{\star}$ are obtained by sampling the subsets with replacement $\mathcal{M}$ times. The bootstrap sets $\mathbf{Y}^{\star}$ are obtained in a similar fashion. Here $\mathcal{M}$ is the number of biometric "entities" (e.g., *either* the number of fingers or the number of volunteers presenting fingers); we continue here with $\mathcal{M} = \mathcal{P}$, the number of volunteers (subjects), since this subdivision gives the more accurate confidence intervals.

We have to rethink the bootstrap sampling a little. The above bootstrap sets are obtained from the samples pretty much in an independent fashion. However, there may be interdependence between sets $\mathbf{X}$ and $\mathbf{Y}$ because the sets are obtained from the same subjects. There may be interdependence between individual match scores X_i and non-match scores Y_j and this dependence has to be somehow replicated into the bootstrap sets $\mathbf{X}^{\star}$ and $\mathbf{Y}^{\star}$ when sampling with replacement.

Note that there may also be much dependence between the match score set $\mathbf{X}$ and the mismatch scores $\mathbf{Y}$. After all, the score sets ($\mathcal{X}_i$ and $\mathcal{Y}_j$) are associated with the same volunteer for $i = j$. Therefore, we could combine the sets of (2) into a set of pairs of subsets:

$$(\mathbf{X}, \mathbf{Y}) = \{(\mathcal{X}_i, \mathcal{Y}_i),\ \ i = 1, ..., \mathcal{P}\}, \tag{3}$$

and sample with replacement from $(\mathbf{X}, \mathbf{Y})$ instead. A joint person subset bootstrap confidence interval for FAR and FRR is now fairly straightforward to construct. That is, B times do—

1. Generate two bootstrap sets $\mathbf{X}^{\star}$ and $\mathbf{Y}^{\star}$ by simultaneously and identically sampling with replacement $\mathcal{P}$ set pairs from $(\mathbf{X}, \mathbf{Y})$ of (3). The two bootstrap sets are
$$(\mathbf{X}^{\star}, \mathbf{Y}^{\star}) = \{(\mathcal{X}^{\star}_1, \mathcal{Y}^{\star}_1), ..., (\mathcal{X}^{\star}_{\mathcal{P}}, \mathcal{Y}^{\star}_{\mathcal{P}})\}.$$
2. Compute the bootstrap estimate $\hat{F}^{\star}$ at threshold x_o as
$$\hat{F}^{\star}(x_o) = \frac{1}{M} \sum_{X_i \in \mathbf{X}^{\star}} \mathbf{1}\,(X_i \le x_o).$$
3. Compute the bootstrap estimate $\hat{G}^{\star}$ at threshold y_o as
$$\hat{G}^{\star}(y_o) = \frac{1}{N} \sum_{Y_i \in \mathbf{Y}^{\star}} \mathbf{1}\,(Y_i \le y_o).$$

   This gives the sets $\mathbf{F}^{\star}(x_o) = \{\hat{F}^{\star}_k(x_o), k = 1, ..., B\}$ and $\mathbf{G}^{\star}(y_o) = \{\hat{G}^{\star}_k(y_o), k = 1, ..., B\}$ of B bootstrap estimates.
4. Rank the estimates in $\mathbf{F}^{\star}(x_o)$, $\mathbf{G}^{\star}(y_o)$ as in person subset bootstraps to obtain $\mathbf{F}^{\star}_O(x_o)$, $\mathbf{G}^{\star}_O(y_o)$, respectively.
5. Eliminate the bottom $(\alpha/2)100\%$ and the top $(\alpha/2)100\%$ from $\hat{F}^{\star}_O(x_o)$, $\hat{G}^{\star}_O(y_o)$ as in person subset bootstrap to obtain FRR and FAR confidence intervals at x_o, y_o, respectively.

5. Experiments

In this work, we attempt to compare the efficacy of different error estimation methods by sequestering a random portion of the biometric data. The non-sequestered data is first used to arrive at false positive and false negative error rate estimates and their respective confidence intervals using (i) person subset bootstrap (method P) and (ii) joint person bootstrap (method Q) methods for arriving at estimation of FRR and FAR confidence intervals. The accuracies of these confidence interval estimates is ascertained using the error rates estimated from the sequestered data. Because of the limited amount of data, the procedure of (randomly) splitting the data into two independent (e.g., train and test) datasets is repeated (rather than generating two new datasets for each confidence interval verification experiment). The experiment is summarized as follows:

1. Randomly split the number of IDs into two sets, A and B, each set containing identical number of IDs.
2. For each method $R \in \{P, Q\}$, use set A to compute the $FAR_R(A)$, $FRR_R(A)$ confidence interval (CI) estimates using CI estimation method R at different thresholds.
3. From set B, estimate $FRR(B)$, $FAR(B)$ at different thresholds.
4. For each method $R \in \{P, Q\}$, check whether $FRR(B)$ $FAR(B)$ estimates are within the corresponding confidence intervals $FRR_R(A)$, $FAR_R(A)$ at different thresholds.

5. By repeating steps 1-4 n number of times, obtain average estimates of probabilities $Prob(FAR(B) \in CI\ of\ FAR_R(A))$, $Prob(FRR(B) \in CI\ of\ FRR_R(A))$ at different thresholds for each confidence interval estimation method R.

We use a private data set. The data are acquired from $C = 114$ different fingers in 2 sessions 5 weeks apart. The subjects are approximately half adult males and half adult females in the age group 22-65. In each session, for each subject, 5 prints of the left and right index finger are acquired. Hence, the database contains a total of $1,140$ impressions, i.e., 10 prints of 114 fingers. The number of match scores m per finger is 90 and the number of non-match scores n per finger is $5,650$. ($M = 10,260$ and $N = 644,100$.)

Figures 1, 2 compare the confidence interval estimates obtained by the two estimation methods for a typical random splitting of the dataset. The confidence interval accuracy verification experiments are summarized in Table 1.

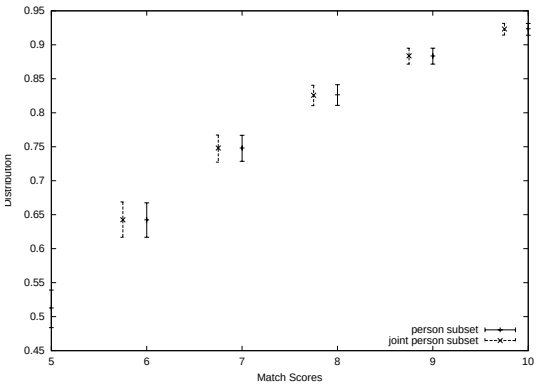


Figure 1. Typical FAR Confidence intervals of the training data using two estimation methods for private data set.

5. Discussion

From Figures 1 and 2, it can be seen that the confidence intervals estimated from the two methods are not significantly different. The confidence interval accuracy verifications results are also almost identical (see Table 1). Thus, it can be readily seen from the illustrations that there is no significant *additional* dependency among the dataset owing to the dependency among the FAR and FRR random variables. Note that this is *not* to infer that there no dependency in the data due to to the dependency among the FAR and FRR random variables. It appears that the person subset bootstrap may have captured the dependency. We are in the process of confirming in this finding for other datasets and design of additional experiments.

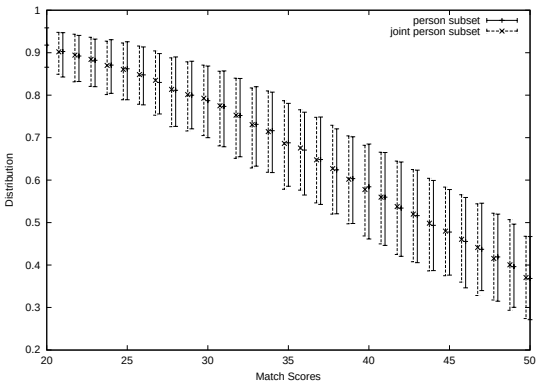


Figure 2. Typical FRR Confidence intervals of the training data using two estimation methods for private data set.

Error / Estimate	FRR (%)	FAR (%)
Person Subset	36.77	25.49
Joint Person Subset	36.93	25.56

Table 1. On the average what percentage of times the 90% training confidence intervals failed to capture the test data for different methods of estimates based on a private dataset used in our experiments?

References

[1] K. Cho, P. Meer, and J. Cabrera. Performance assessment through bootstrap. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(11):1185–1198, Nov. 1997.

[2] B. Efron. Bootstrap methods: Another look at the Jackknife. *Ann. Statistics*, 7:1–26, 1979.

[3] A. K. Jain, R. C. Dubes, and C.-C. Chen. Bootstrap techniques for error estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(5):628–633, Sept. 1987.

[4] R. Liu and K. Singh. Moving blocks Jackknife and Bootstrap capture weak dependence. In R. LePage and L. Billard, editors, *Exploring the Limits of the Bootstrap*, pages 225–248, New York, NY, 1992. John Wiley & Sons, Inc.

[5] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, 33(2):56–63, 2000.

[6] S. M. Weiss. Small sample error rate estimation for k-NN classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(3):285–289, Mar. 1991.

[7] J. L. Wayman. Confidence interval and test size estimation for biometric data. In *Proc. IEEE AutoID'99*, pages 177–184, Oct. 1999.

Anchorperson Indexing and Visual Table of Contents Generation for TV News

Vakkalanka Suresh, S. Palanivel, C. Chandra Sekhar, B. Yegnanarayana
Speech and Vision Laboratory
Department of Computer Science and Engineering
Indian Institute of Technology Madras, Chennai-600 036, India.
Email:{suresh,spal,chandra,yegna}@cs.iitm.ernet.in

Abstract

This paper presents an approach to construct quickly and automatically a compact yet meaningful abstraction of news video contents in a structured format, allowing users to randomly browse large amounts of video data. The two important components in visual table of contents generation are structure analysis *and* video abstraction*. Structure analysis is the process of parsing the news video into story units, and video abstraction is the process of extracting representative frames from each story unit which can serve as entries in the table of contents. For parsing the news video into story units, anchorperson indexing/detection is an important task. Since an anchorperson hosts a news program, locations of anchorperson segments in the news video provides landmarks for detecting story units. In the proposed method, a fast and computationally effective algorithm is employed to automatically detect a face, and a* 73 *dimensional feature vector is derived from the face region. Autoassociative neural network (AANN) model is used to capture the distribution of the extracted facial features. In the video abstraction process a fast key frame selection method based on the motion activity in the video has been proposed. Experimental results show the effectiveness of the proposed method.*

1 Introduction

With the never ending advances in digital technology, more and more video data is generated every day. As the accessible video collections grow, efficient schemes for navigating, browsing, and retrieving video data are required. A good survey of techniques for automatic indexing and retrieval of video data can be found in [1], [2]. Among the various video categories, news programs are important storing objects, due to the fact that they concisely cover large number of topics related to society, politics, business, sports, weather, etc.

In recent years, several news video indexing techniques have been proposed [3], [4], [5]. These systems employ a two stage classification scheme. First, the video is segmented into shots and each shot is then classified into anchor and non-anchor categories. Most of the work in shot classification can be categorized into two classes of approaches. One is model based and the other is unsupervised clustering based.

This paper presents an approach to automatically identify anchor segments using visual clues. First, the complete video is segmented into low and high motion regions, by using a simple motion metric. Low motion segments are the probable anchorperson segments, but may also correspond to non-anchorperson segments like graphic objects and interviews. These low motion segments are further classified into anchorperson and non-anchorperson segments, by using visual anchorperson models. The model will be built on-fly without any user supervision for the first appearance of an anchorperson in the news database, and the model once built for an anchorperson can be used as off-line model to detect the appearance of the same anchorperson on different dates in the news database. The temporal location of anchorperson segments are then used to construct the story units. A motion based algorithm is then employed to extract key frames from each story unit.

This paper is organized as follows: Section 2 describes the video structure analysis process and Section 3 describes the video abstraction process. Experimental results are discussed in Section 4. Finally conclusions and the scope for future work are discussed in Section 5.

2 Structure Analysis

Video structure analysis is the process of extracting temporal and structural information of news video programs. It involves detecting the story boundaries and parsing the complete video into story units. Each news story can be further segmented into an introduction by the anchorperson followed by a detailed report. In general, it can be observed

that during the news story introduction by an anchorperson the background around the anchorperson is almost constant, accompanied by a small motion of the anchorperson in the foreground. Whereas during detailed reporting, motion in the background as well as in the objects of interest is high most of the times. A binary pattern matching method is employed on the motion based binary feature vector derived for the complete video, to segment the complete video into low and high motion segments.

In order to segment the video into low and high motion segments we use a motion metric measuring the foreground object motion, computed from the thresholded difference image between two frames. To reduce the sensitivity of the motion metric to noise, 1D-smoothing technique is applied on individual frames as described in [6].

2.1 Definition of the motion metric

Let f_i and f_j represent the i^{th} and j^{th} 1D-smoothed frames respectively. The thresholded difference image D between f_i and f_j is given by

$$D(x,y) = \begin{cases} 1, & \text{if } |f_i(x,y) - f_j(x,y)| > \tau \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where τ is the threshold. Then the amount of motion M between frames f_i and f_j is obtained by

$$M = \frac{1}{W \times H} \sum_{x=0}^{W} \sum_{y=0}^{H} D(x,y) \quad (2)$$

where W and H represents the width and height of each frame respectively.

2.2 Video segmentation

To reduce the computational complexity as well as to enhance the sensitivity to motion, the metric is computed between frame pairs that are at a fixed interval $\triangle t$ (20 frames) apart. The motion metric based binary feature vector for the complete video is obtained as

$$B_t = \begin{cases} 1, & \text{if } M_t > \lambda \\ 0, & \text{otherwise} \end{cases} \qquad 0 \le t < (T/\triangle t) \quad (3)$$

where λ is the threshold empirically chosen from the data and T is the total number of frames in the video. A binary pattern matching method is employed on this feature vector to segment the complete video into low and high motion segments. Segments of low-motion correspond to the sequence pattern "00...0". These low motion segments correspond to anchorperson segments, and some non-anchor person segments like interviews and graphic objects. By using visual based anchorperson models, these low motion segments are classified into anchorperson and non-anchorperson segments.

2.3 Visual feature extraction

The visual features extracted should be insensitive to the location and size of the anchor and color and visual content in the studio background. The proposed visual feature extraction has three important modules: Face localization, eye location estimation and feature vector extraction.

2.3.1 Face localization

Recent methods for face detection use neural networks [7], skin color segmentation [8] and motion information [9], [10] for tracking faces in video. Our approach of face localization contains two major modules 1) skin regions detection, and 2) face region approximation. A Gaussian Mixture Model (GMM) is used to model the skin color in YC_bC_r color space. Skin color patches extracted from various still images and video frames, covering a large range of skin color appearance have been used for training the model. Given an image, we can classify the regions of the image into two classes by finding the likelihood of each pixel to be a skin pixel. Figure 1(a) shows the result of skin color regions detection applied on the original image Figure 1(b).

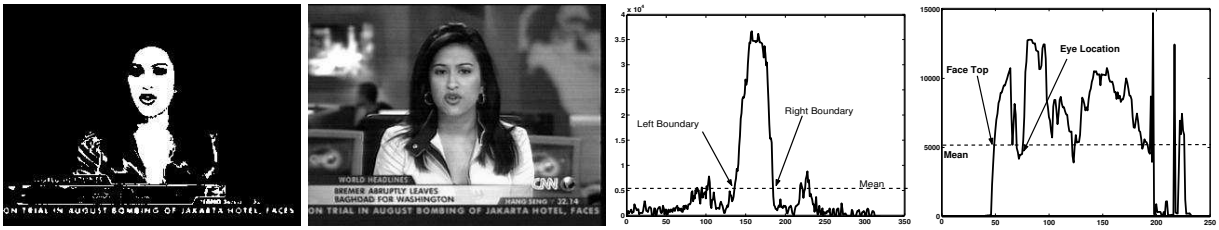
The bounding box for the face region is found using the horizontal and vertical projections of the binary image obtained from the skin region detection as shown in Figures 1(c) and 1(d). The first mean crossing points as we move away from peak location on both sides of the peak value in the horizontal projection are taken as left and right boundaries of the face region. The top part of the face is obtained from the vertical projection of the sub image between left and right boundaries. The height of the face is estimated from the width of the face ($(4/3)\times$width), and a rectangular bounding box for the face can be obtained as shown in Fig. 2(a)

2.3.2 Eye location estimation

Once the approximate face region is found, the next step is to extract the location of eyes. The eye location algorithm is based on our earlier work in [10]. Face region within the bounding box is thresholded to obtain the thresholded face image U, given by

$$U(x,y) = \begin{cases} 255, & \text{if } Y(x,y) < \lambda_1 \text{ and} \\ & C_r(x,y) < \lambda_2 \text{ and} \\ & C_b(x,y) > \lambda_3 \\ f(x,y), & \text{otherwise} \end{cases} \quad (4)$$

where λ_1, λ_2 and λ_3 are the average Y, C_r and C_b values of the pixels in the forehead region, respectively. Morphological closing operation is applied to the thresholded face image, and the centroid of all the blobs are estimated.



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

Figure 1. (a) Binary image obtained from the skin color region detection, (b) corresponding original image, (c) horizontal projection of the binary skin color region detected image and (d) vertical projection of the region within left and right boundaries of binary image

(a) (b)

Figure 2. (a) Rectangular bounding box for the face region and (b) feature vector extraction.

The eyebrow (E) pixels are estimated using

$$E(x,y) = \begin{cases} 1, & \text{if } Y(x,y) \geq \lambda_1 \text{and} \\ & Y(x,y+1) \geq \lambda_1 \text{and} \\ & Y(x,y+2) < \lambda_1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The centroid of the blobs which are nearer to the center of the eyebrow pixels are taken as the location of the eyes.

2.3.3 Feature extraction

A 73 dimensional feature vector is extracted from the face region as shown in Fig. 2(b). The position of the 73 facial regions are estimated relative to the location of the eyes. The distance between the eyes is used to estimate the size of each region. The region can be of size 2×2, 3×3 or 4×4 pixels. The average gray value in each region is used as an element in the 73 dimensional feature vector.

2.4 AANN based anchorperson indexing

Autoassociative neural network (AANN) models are the feedforward neural networks performing an identity mapping of the input space, and are used to capture the distribution of the input data [11]. The structure of the AANN model used in our study is $73L\ 90N\ 30N\ 90N\ 73L$, where L denotes a linear unit, and N denotes a nonlinear unit. The AANN based anchorperson indexing has the following steps;

1. Consider the low motion segments in the descending order of duration.
2. Apply face detection algorithm for the first few frames of the current low motion segment. If a face is detected extract visual features from the entire segment, else declare the low motion segment as non-anchorperson segment and goto Step 6.
3. Test against the existing off-line anchorperson models. If any model gives confidence score above certain threshold α, declare the low motion segment as anchorperson segment and goto Step 6.
4. Test against the models if any, created from the current video as described in Step 5. If any model gives a confidence score above the threshold α then goto Step 6.
5. Train the visual model using the features extracted from the current segment.
6. Repeat steps 2 to 5 for the next low motion segment.
7. For each visual model created using segments of the current video, find the temporal distance between the first and last occurrence of the low motion segments corresponding to that model. If the distance is greater than $(3/4)^{th}$ of the total duration of the news video, the model is declared as an anchorperson model and all the low motion segments corresponding to this model are declared as anchor segments.

3 Video Abstraction

Video abstraction is the process of creating a presentation of visual information about the structure of a video, which should be much smaller than the original video. Key

frames play an important role in the video abstraction process. Key frames are still images, extracted from original video data, that best represent the content of a story in an abstract manner. Since motion is the major indicator for content change, dominant motion components resulting from camera operations and large moving objects are the most important source of information. So, in an effective approach to key frame extraction, the number of key frames needed to represent a segment of video should be based on temporal variations of video content in the segment.

In our approach to key frame extraction, the binary motion vector derived during the structure analysis process, as defined in Section 2.1 is reused to extract the key frames. As described earlier, the low activity video segments correspond to binary pattern "00..0" and the high activity video segments correspond to the binary pattern "11...1" or "01" or "10". From each of the low motion regions, the middle frame is taken as the key frame and from each of the high activity regions, key frames are extracted at the interval $\triangle t$ (20 frames).

4 Experimental Results

The proposed method has been evaluated on more than 6 hours of news video data recorded at 25 frames per second and frame size 320 × 240 from 4 news channels: BBC World, CNN, NDTV 24 × 7 and ETV. The collection includes 10 subjects. The details of the experimental results are given in Table 1. To evaluate the performance of the proposed method of anchorperson indexing, we use the standard *precision* and *recall* criteria, shown in the following:

$$\text{precision} = \frac{\text{number of hits}}{\text{number of hits + number of false alarms}} \quad (6)$$

$$\text{recall} = \frac{\text{number of hits}}{\text{number of hits + number of misses}} \quad (7)$$

A precision of 99.35% and a recall of 98.7% for story segmentation is achieved.

Table 1. Details of experimental results

Total number of news stories	155
Hits	153
Misses	2
False alarms	1

5 Conclusions

We have presented an AANN model based approach to automatically detect and index anchorpersons in a news video and to construct visual-table-of-contents for a given news video. The proposed method is superior to the methods where off-line trained audio-visual models of anchorpersons are used which involve manual collection of training data and provide little flexibility. In our approach the anchorperson models are created on-line without any human supervision and the models once created can be used as off-line models to detect the anchorperson appearance in a different video.

References

[1] R. Brunelli, O. Mich, and C. Modena, "A survey on the automatic indexing of video data," *Journal of Visual Communication and Image Representation*, vol. 10, no. 2, pp. 78–112, 1999.

[2] Y. Wang, Z. Liu, and J.-C. Huang, "Multimedia content analysis using both audio and visual clues," *IEEE Signal Processing Magazine*, vol. 17, pp. 12–36, Nov. 2000.

[3] M. Bertini, A. D. Bimbo, and P. Pala, "Content-based indexing and retrieval of TV news," *Pattern Recognition Letters*, vol. 22, pp. 503–516, Apr 2001.

[4] X. Gao and X. Tang, "Unsupervised video-shot segmentation and model-free anchorperson detection for news video story parsing," *IEEE Trans. Circuits, Systems, Video Technology*, vol. 12, pp. 765–776, Sept. 2002.

[5] A. Albiol, L. Torres, and E. J. Delp, "The indexing of persons in news sequences using audio-visual data," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, (Hong Kong), Apr 6-10, 2003.

[6] P. K. Kumar, S. Das, and B. Yegnanarayana, "One-dimensional processing of images," in *Int. Conf. Multimedia Processing and Systems*, (Chennai, India), pp. 181-185, Aug. 13-15, 2000.

[7] H. A. Rowley, S. Baluj, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 23–38, May 1998.

[8] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavlet packet analysis," *IEEE Trans. Multimedia*, vol. 1, pp. 264–277, Sept. 1999.

[9] B. Li and R. Cehellappa, "A generic approach to simultaneous tracking and verification in video," *IEEE Trans. Image Processing*, vol. 11, pp. 530–544, May 2002.

[10] S. Palanivel, B. S. Venkatesh, and B. Yegnanarayana, "Real time face authentication system using autoassociative neural network models," in *Int. Conf. Multimedia and Expo*, (Baltimore, MD, USA), pp. 257-260, July 6-9, 2003.

[11] B. Yegnanarayana and S. Kishore, "AANN: an alternative to GMM for pattern recognition," *Neural Networks*, vol. 15, pp. 459–469, Jan 2002.

Person Authentication Using Acoustic and Visual Features

S. Palanivel, C. Chandra Sekhar and B. Yegnanarayana
Speech and Vision Laboratory
Department of Computer Science and Engg.
Indian Institute of Technology Madras
Chennai-600 036, India
Email:{spal,chandra,yegna}@cs.iitm.ernet.in

B.V.K. Vijaya Kumar
Department of Electrical and Computer Engg.
Carnegie Mellon University
Pittsburgh
PA 15213, U.S.A
Email:kumar@ece.cmu.edu

Abstract

This paper proposes a method for automatic person authentication using acoustic and visual features. The method uses motion information to estimate the face region, and the face region is processed in the YC_rC_b color space to determine the location of the eyes. The system models the nonlip region of the face using a Gaussian distribution, and it is used to estimate the center of the mouth. Visual features are extracted relative to the location of the eyes and the center of the mouth using multiscale morphological dilation. Acoustic features are derived from the speech signal, and are represented by weighted linear prediction cepstral coefficients (WLPCC). Autoassociative neural network (AANN) models are used to capture the distribution of the extracted acoustic and visual features. The evidence from acoustic and visual models are combined using a sum rule. The performance of the method is evaluated for TV broadcast news data and the system achieves about 5.6% equal error rate for 50 subjects.

1 Introduction

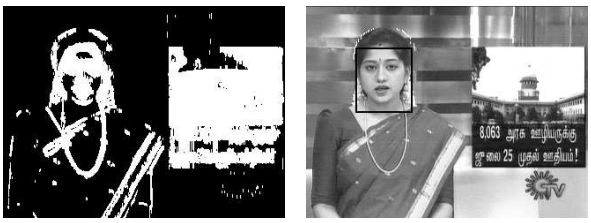
Automatic person authentication by machine appears to be difficult, while it is done effortlessly by human beings. A survey of speech-based bimodal speaker recognizers is given in [1]. The terms acoustic, facial and visual features refer to the features extracted from the speech, face and mouth image, respectively. Audio-video based person authentication methods use either acoustic and facial modalities [2],[3],[4],[5] or acoustic and visual modalities [6],[7],[8]. The mel frequency cepstral coefficients (MFCC) and weighted linear prediction cepstral coefficients (WLPCC) are commonly used as acoustic features. The visual features such as discrete cosine transform (DCT) of the lip region [7], eigenlips [6],[8] are used to represent the mouth image. Our earlier results on person authentication using facial features are given in [9]. The automatic person authentication system proposed in this paper consists of four modules, namely face localization, acoustic feature extraction, visual feature extraction and the model for person authentication. The face localization method is explained in Section 2. The visual and acoustic feature extraction techniques are described in Section 3 and 4, respectively. Section 5 describes the AANN model for person authentication. The experimental results are given in Section 6.

2 Face localization

Detecting faces automatically from the intensity or color image is an essential task for many applications like person authentication, face tracking and video indexing [10]. Face localization is a face detection problem with the assumption that an input image contains a single face. We used a simple method to estimate the face region using only the motion information in order to implement the system in real time. In our method, the face region is determined from the upper head contour points which are extracted from the accumulated difference image. The accumulated difference image is scanned from top to bottom to find out an approximate top center pixel (c_x, c_y) of the moving region. The head contour points are extracted by scanning the accumulated difference image from the pixel (c_x, c_y). The width of the face (w_1) is determined from the head contour points, and the face region is estimated using w_1 and (c_x, c_y). Figure 1(a) shows the accumulated difference and Figure 1(b) shows the extracted head contour points and the face region.

3 Visual feature extraction

This paper proposes a method for extracting visual features from the the mouth image, which are relative to the location of the eyes and the center of the mouth. Among the facial features, eyes and mouth are the most prominent features used for estimating the size and pose of the face [11],[12]. Sections 3.1 and 3.2 describe the methods for locating the eyes and mouth, respectively.



(a) (b)

Figure 1. Face localization. (a) Accumulated difference image. (b) Face region.

(a) (b)

Figure 2. Eye location estimation. (a) Thresholded face. (b) Location of the eyes.

3.1 Eye location estimation

The template-based approach is commonly used for locating the eyes, and the method given in [11],[12] use the gray-scale morphological operations (e.g., dilation and erosion) [13]. For locating the eyes, the face region is converted from RGB to YC_rC_b color space as given by

$$\begin{cases} Y = & 0.299R + 0.587G + 0.114B \\ C_r = & R - Y \\ C_b = & B - Y \end{cases} \quad (1)$$

where R, G and B are the red, green and blue component of the color image, respectively. The forehead and nose regions have high luminance (Y) than the eye regions. Similarly, the eye region has low red chrominance (C_r) and high blue chrominance (C_b) than the forehead and the nose region. Using these facts, the face region is thresholded to obtain the thresholded face image (U), given by

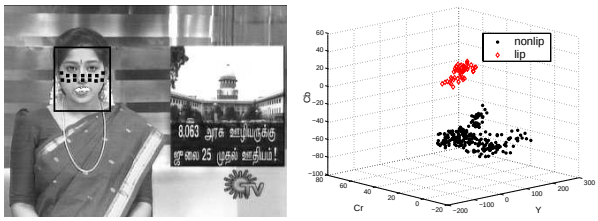
$$U(i,j) = \begin{cases} 255, & \text{if } Y(i,j) < \lambda_1 \text{ and } C_r(i,j) < \lambda_2 \\ & \text{and } C_b(i,j) > \lambda_3 \\ I(i,j), & \text{otherwise} \end{cases} \quad (2)$$

where λ_1, λ_2 and λ_3 are the average Y, C_r and C_b values of the pixels in the forehead region, respectively. The forehead region is estimated from the width of the face (w_1) and the pixel (c_x, c_y). Morphological closing operation is applied to the thresholded face image, and the centroid of all the blobs are estimated.

The relative positions of the centroids with respect to the rectangular bounding box enclosing the face region and the contrast information in the eyebrow region are used to determine the location of the eyes. Figure 2(a) shows the thresholded face image, and Figure 2(b) shows the location of the eyes. The method can detect the location of the eyes in the presence of eye glasses as long as the eyes are visible.

3.2 Mouth center estimation

The mouth or lip image analysis has received considerable attention in the area of speech recognition and person recognition. Mouth image segmentation is a necessary step for visual feature extraction. For estimating the center of the mouth, we model the color distribution of the nonlip regions of the face using a Gaussian distribution and it is used to detect the lip region pixels. The nonlip regions and the detected lip pixels are shown in Figure 3 (a). The distribution of Y, C_r and C_b values of the nonlip and the lip pixels are shown in Figure 3 (b). The center of the mouth is estimated using the pixel coordinates of the detected lip pixels.



(a) (b)

Figure 3. Mouth center estimation. (a) Nonlip regions and the detected lip pixels. (b) Distribution of nonlip and lip pixels.

3.3 Feature extraction

The appearance of mouth image or shape of the lip contour during speaking gives significant information for recognizing humans especially for females. The shape of the lip contour is a dominant visual feature in the mouth region. Visual features are associated with local maxima because the lip and the interior of the mouth has low luminance (Y) than the nonlip region. The local maxima can be extracted using the morphological dilation [13]. A rectangular rigid grid is placed over the mouth region and the multiscale morphological dilation is used for feature extraction. The location of the eyes and its angle are used to determine the size and orientation of the grid, respectively. The grid consists of 25 nodes, and the position of these nodes are determined relative to the location of the eyes and the center of the mouth.

Given an image I: $\mathcal{D} \subseteq \mathcal{Z}^2 \longrightarrow \mathcal{Z}$ and a structuring function G: $\mathcal{G} \subseteq \mathcal{Z}^2 \longrightarrow \mathcal{Z}$, the dilation of the image I by the structuring function G is denoted as $(I \oplus G_\sigma)$, and it is defined

by

$$(I \oplus G_\sigma)(i,j) = \max_{x,y}\{I(i-x, j-y) + G(x,y)\} \quad (3)$$

where $-M_a \leq x, y \leq M_b$, with $1 \leq i \leq w$, $1 \leq j \leq h$. w and h are the width and height of the image, respectively. The size of the structuring function is decided by the parameters M_a and M_b, and is given by $(M_a + M_b + 1) \times (M_a + M_b + 1)$. For a flat structuring function ($G(x,y) = 0$) the dilation can be expressed as

$$(I \oplus G_\sigma)(i,j) = \max_{x,y}\{I(i-x, j-y)\} \quad (4)$$

The dilation operation (4) is applied at each grid node for $\sigma = 1, 2, ..., m$ to obtain m visual feature vectors from the mouth image. The distance between the eyes (d) is used to determine the parameters M_a, M_b and m. These parameters are chosen in such a way that $M_a + M_b + 1$ for $\sigma = m$ is less than or equal to the minimal distance between two nodes of the grid. The value $M_a = \lfloor d/64 + 0.5 \rfloor + \lfloor (\sigma - 1)/2 \rfloor$, $M_b = \lfloor d/64 \rfloor + \lfloor \sigma/2 \rfloor$ and $m = 3$ has been used in our experiments. Figs. 4 (a) shows the visual regions used for extracting the feature vectors for $\sigma = 2$ and 3, respectively. Each visual feature vector is normalized to [-1, 1], the normalized visual feature vector is less sensitive to variation in the image brightness.

(a) (b)

Figure 4. Visual feature extraction. (a) σ=2. (b) σ=3.

4 Acoustic feature extraction

Speaker information can be extracted from the speech signal at subsegmental, segmental and suprasegmental levels. The segmental features are the features extracted from short (10-30ms) segments of the speech signal. In this paper, the differenced speech signal is segmented into frames of 20 ms, with a shift of 5 ms. A 14^{th} order linear prediction (LP) analysis is used to capture the properties of the signal spectrum. The spectrum of speech signal is attributed primarily to the shape of the vocal tract. The recursive relation between the predictor coefficients and cepstral coefficients is used to convert the 14 LP coefficients into 19 LP cepstral coefficients. The LP cepstral coefficients for each frame are linearly weighted to get the WLPCC. A 19 dimension WLPCC for each frame is used as a feature vector.

5 Autoassociative neural network model for person authentication

Autoassociative neural network models are feedforward neural networks performing an identity mapping of the input space, and are used to capture the distribution of the input data [14]. The five layer Autoassociative neural network model as shown in Figure 5, is used to capture the distribution of the feature vectors.

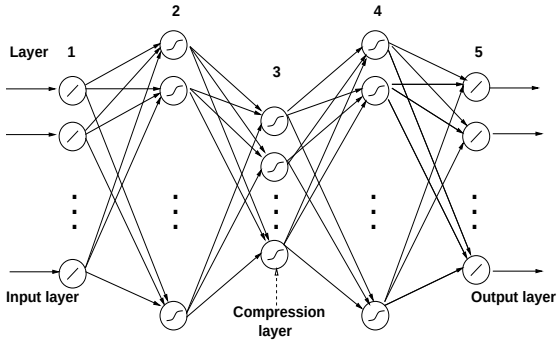


Figure 5. AANN model used for person authentication.

The structures of the AANN models used in our study are $19L38N4N38N19L$ and $25L40N10N40N25L$ for capturing the distributions of the acoustic and visual features of a subject, respectively, where L denotes a linear unit and N denotes a nonlinear unit. The integer value indicates the number of units used in that layer. The nonlinear units use $tanh(s)$ as the activation function, where s is the activation value of the unit. The standard backpropagation learning algorithm is used to adjust the weights of the network to minimize the mean square error for each feature vector.

6 Experimental results

Performance of the person authentication system is evaluated for 50 subjects, 32 females and 18 males. For enrolling a subject, an AVI file of 60 sec duration at 12 fps is recorded with a resolution of 320 $\times$ 240. The speech signal is recorded at 8000 samples per second. Since during news reading, the background around the news reader is almost constant, accompanied by a small motion of the reader in the foreground, the motion information is used to estimate the face region as described in Section 2. If there is a significant head movement during newsreading then the interframe difference image can be used to track the face region [9]. The face localization method is computationally efficient, and it is invariant to size of the face and lighting conditions. The method assumes that there is no moving object in the background, and it is a reasonable assumption for person authentication. The visual features are extracted automatically as described in Section 3 for 300 mouth images

Table 1. Person authentication results

	Acoustic	Visual	Acoustic+visual
Equal error rate	11.2%	8.1%	5.6%

in the video. The distance between the eye location varies between 24 to 33 pixels. The acoustic features are extracted as described in Section 4. The extracted acoustic feature vectors are given as input to the AANN model $19L38N4N38N19L$, and the network is trained for 100 epochs as described in Section 5 for capturing the distribution. Similarly the distribution of the 900 visual feature vectors are captured using an AANN model $25L40N10N40N25L$, and the network is trained for 50 epochs.

For authenticating the identity claim of a subject, an AVI file of 10 sec duration at 12 fps is recorded, one month after collecting the training data. The acoustic feature vectors are extracted from the speech signal. Each feature vector is given as input to the corresponding model. The output of the model is compared with the input to compute the error. The error (ε) is transformed into a confidence score (c) by using the equation $c = exp(-\varepsilon)$. Similarly, the visual feature vectors are extracted from the mouth images in the video. Each visual feature vector is given as input to the corresponding model, and the confidence score is estimated. The confidence scores from the acoustic and visual models are summed, and the result is used to accept or reject the identity claim of the subject.

In the database of 50 subjects, there will be 50 authentic claims and 49×50 impostor claims. The acoustic, visual and combined confidence scores are calculated for all the claims. The performance of person authentication system is measured in terms of equal error rate (EER). In our experiment, the score normalization techniques such as Z-norm, T-norm and ZT-norm [15] are not used, and the EER is obtained by employing subject independent thresholds. The person authentication results are given in Table 1. The method is invariant to size of the face and its position in the image, and only the confidence score from the claimant model is used for authentication. The face localization and feature extraction techniques are computationally inexpensive and the method tests the identity claim of a subject at about 6 frames/s on a PC with 2.3 GHZ CPU.

7 Conclusion

In this paper, we have proposed a method for automatic person authentication using acoustic and visual features. The AANN models are used to capture the distribution of acoustic and visual feature vectors. The method extracts visual features relative to the location of the eyes and the center of the mouth. The face localization and feature extraction techniques are computationally inexpensive, and the method tests the identity claim of a subject within a reasonable time.

References

[1] C.C. Chibelushi Yilmaz, F. deravi, and J.S.D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. MultiMedia*, vol. 4, no. 1, pp. 23–37, March 2002.

[2] T. Choudhury and B. Clarkson and T. Jebara and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proc. Audio-and video-based Biometric Person Authentication*, Washindton, D.C., March 1999, pp. 176–181.

[3] A. Senior, C. Neti, and A. Senior, "On the use of visual information for improving audio-based speaker recognition," in *Audio-Visual speech processing conference*, Santa Cruz, CA, August 1999.

[4] C. Sanderson and K.K. Paliwal, "Noise resistant audio-visual verifi cation via structural constraints," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, Hong Kong, April 2003, pp. 716–719.

[5] T.J. Hazen, E. Weinstein, and B. Heisele, "Multi-modal face and speaker indentifi cation on a handheld device," in *Workshop on multimodal user authentication*, Santa Barbara, California, December 2003, pp. 113–120.

[6] P. Jourlin, J. Luettin, D. Genoud, and H. Wassner, "Acoustic-labial speaker verifi cation," *Pattern Recognition Letters*, vol. 18, pp. 853–858, 1997.

[7] U.V. Chaudhari, G.N. Ramaswamy, G. Potamianos, and C. Neti, "Audio-visual speaker recognition using time-varying stream reliability prediction," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, Hong Kong, April 2003, pp. 712–715.

[8] A. Kanak, E. Erzin, Y. Yemez, and A.M. Tekalp, "Joint audio-video processing for biometric speaker identifi cation," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, Hong Kong, April 2003, pp. 377–380.

[9] S. Palanivel, B.S. Venkatesh, and B. Yegnanarayana, "Real time face authentication system using autoassociative neural network models," in *IEEE Int'l Conf. Multimedia and Expo.*, Baltimore, July 2003, pp. 257–260.

[10] M. Yang, D.J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, January 2002.

[11] A. Nikolaidis and I. Pitas, "Facial feature extraction and pose determination," *Pattern Recognition*, vol. 33, no. 11, pp. 1783–1791, 2000.

[12] R. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002.

[13] P.T. Jackway and M. Deriche, "Scale-space properties of the multiscale morphological dilation-erosion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 38–51, January 1996.

[14] B. Yegnanarayana and S.P. Kishore, "AANN: an alternative to GMM for pattern recognition," *Neural Networks*, vol. 15, pp. 459–469, January 2002.

[15] *National Institute of Standards and Technology (NIST)- Speaker Recognition Workshop Report*, University of Maryland, Baltimore, 2003.

ISBN: 3-929757-3