EURASIP Journal on Advances in Signal Processing

Recent Advances in Biometric Systems: A Signal Processing Perspective

Guest Editors: Natalia A. Schmid, Stephanie Schuckers, Jonathon Phillips, and Kevin Bowyer

Recent Advances in Biometric Systems: A Signal Processing Perspective

EURASIP Journal on Advances in Signal Processing

Recent Advances in Biometric Systems: A Signal Processing Perspective

Guest Editors: Natalia A. Schmid, Stephanie Schuckers, Jonathon Phillips, and Kevin Bowyer

Contents

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 128752, 2 pages
doi:10.1155/2009/128752

Editorial

Recent Advances in Biometric Systems: A Signal Processing Perspective

Natalia A. Schmid,[1] Stephanie Schuckers,[2] Jonathon Phillips,[3] and Kevin Bowyer[4]

[1] *West Virginia University, Morgantown, WV 26506, USA*
[2] *Clarkson University, Potsdam, NY 13699, USA*
[3] *National Institute of Standard and Technology, Gaithersburg, MD 20899, USA*
[4] *University of Notre Dame, Notre Dame, IN 46556, USA*

Correspondence should be addressed to Natalia A. Schmid, natalias@csee.wvu.edu

Received 31 December 2009; Accepted 31 December 2009

We were pleased to receive a total of thirty-nine submissions to the special issue on "Recent advances in biometric systems: a signal processing perspective." The Guest Editors divided up the responsibility for the submissions, and each submission was reviewed by a minimum of two experts in the relevant area of biometrics. Following the first round of reviews, some of the submissions were revised by the authors and then underwent a second round of review. The final result of the process is the set of fifteen papers that appear in this special issue.

The first six papers all deal with face recognition in some respect. Then we have one paper dealing with iris biometrics and one dealing with recognition by gait. The topic of the next two papers is fingerprint image analysis and the following paper addresses the related topic of palmprint analysis. The next two papers cover issues in signature verification. Lastly, there is one paper on retinal verification and one on using electrocardiogram signals as a biometric. The broad variety of topics in this special issue represents the dynamism and breadth of biometrics.

In "Recognition of faces in unconstrained environments: a comparative study," Ruiz-del-Solar, Verschae and Correa present the results of a comparative study of existing face recognition methods in the context of unconstrained environments. The recognition approaches considered include two local-matching methods, histograms of LBP features and Gabor Jet descriptors, one holistic method, generalized PCA, and two novel image-matching methods, SIFT-based and ERCF-based. The FERET, LFW, UCHFaceHRI, and FRGC face databases are used in the evaluation. Two conclusions are that there is a large dependence of the methods on the amount of face and background information in the image, and that outdoor illumination results in a large decrease in the performance of all of the methods.

In "Facial expression biometrics using statistical shape models," Shark et al. perform face recognition by combining 3D range images and expression. The authors' method is based on a shape space vector derived from a statistical shape model for 3D range data. Experimental results are reported on the SUNY Binghamton BU-3DFE dataset of the 3D face images. Results are reported for both recognition and expression classification.

In "Evolutionary discriminant feature extraction with application to face recognition," Lu et al. present a technique that searches for subspaces to represent faces. The search technique is based on evolutionary computing and is designed to be efficient. One reason the algorithm is efficient is because the search space is confined to discriminatory subspaces.

In "Comparison of spectral-only and spectral/spatial face recognition for personal identity verification," Pan et al. compare the performance of single-band, multiband, and combined spectral/spatial approaches to face recognition. They use the eigenface algorithm from the CSU Face Identification Evaluation System for the basic recognition engine. Multiband eigenface methods in which the multiple bands are processed independently are shown to improve face recognition performance relative to single-band results.

The new spectral-face approach is proposed to preserve both spectral and spatial properties and shown to provide even better performance.

In "Talking-face identity verification, audiovisual forgery and robustness issues," Karam et al. develop an interesting multimodal approach involving face appearance and speaker recognition. They emphasize the aspect of robustness to imposter attacks. Using audio conversion and an MPEG-4 compliant face animation system, they also demonstrate the production of audio-visual forgeries that substantially increase the equal-error rate of an identify verification system.

Protecting the privacy of biometric samples has become an active area of research in biometrics. In "Sorted index numbers for privacy preserving face recognition," Wang and Hatzinakos introduce the concept of sorted index numbers to protect privacy. The sorted index numbers technique converts a feature vector into an ordered set of indices. The authors show the effectiveness of this technique for protecting the privacy of biometric samples.

In "A new user-dependent iris recognition system based on an area preserving pointwise level set segmentation approach", Barzegar and Moin compare their new approach to iris recognition with other methods. A level set approach is used in finding the papillary and limbic boundaries. The approach is claimed to have advantages in cases where the iris is partly occluded. Results of the comparison to five other approaches using three different iris image datasets indicate an improvement in accuracy and speed of processing.

In "Gait recognition using wearable motion recording sensors," Gafurov and Snekkenes investigate the use of wearable motion recording sensors for gait-based person recognition. Such wearable sensors record motion of the body parts during walking. This paper analyzes acceleration signals from the foot, hip, pocket, and arm. The authors also analyze the robustness of the proposed recognition method under three distinct security attacks including a minimal effort-mimicry, knowing the closest person in the database in terms of gait similarity, and knowing the gender of the user in the database.

In fingerprint recognition, there has been debate about the existence of biometric "goats." A biometric goat is a person who consistently has an unusually high false nonmatch rate. In "Inter-subject differences in false non-match rates for a fingerprint-based authentication system," Kelkboom et al. look for the existence of goats. In their study of fingerprint performance, the authors find that 10% of the subjects account for a large portion of the false nonmatches and are classified as biometric goats.

Many applications would benefit from the implementation of biometric authentication with smart cards. In "Integrating the fingerprint verification into the smart card-based healthcare information system" by Pan et al. a typical fingerprint verification algorithm is integrated in a smart card and smart card reader where various designs are compared in terms of real-time execution and security and privacy tradeoffs.

In "Development of a new cryptographic construct using palmprint based fuzzy vault" by Amioy Kumar and Ajay Kumar, this research focuses on a combination of biometrics and cryptography to create a "fuzzy vault" for secure authentication. Asymmetric approaches typically have high security but require high computation. This paper uses the combination of symmetric and asymmetric cryptography in a palmprint authentication system to alleviate the drawbacks of a symmetric-only system.

This special issue presents two novel methods for recognition of individuals based on sample signatures. In "A novel criterion for writer enrolment based on a time-normalized signature sample entropy measure," Garcia-Salicetti et al. promote time-normalized sample entropy as a novel criterion for writer enrollment. They also propose a novel criterion for writer enrollment targeting enhanced signature verification. In "On-line signature verification using fourier descriptors," Yanikoglu and Khomatov involve Fourier descriptors in the process of feature extraction and template formation. The application of Fast Fourier Transform results in a compact representation with a fixed number of coefficients. The main challenge that the authors address in their paper is the design of matching algorithms. The improved performance is achieved through a fusion of the proposed system with a state-of-the-art Dynamic Time Warping (DTW) system.

In their work "Retinal verification using a feature points based biometric pattern," Rouco et al. present a novel approach to the selection of landmark points in the retinal vessel tree. The approach is based on extracting a set of landmarks (bifurcations and crossovers of retinal vessel tree). However, the use of reference structures is avoided, which allows the system to cope with a wider range of images and users. Together with new set of features a new similarity metric is introduced, and a careful analysis of the proposed method is performed using a large and diverse database of retinal images.

In "A sequential procedure for individual identity verification using ECG," Irvine and Israel tackle an intriguing biometrics modality that has received relatively little attention to date. They are interested in identity verification that uses the minimum number of heartbeats of electrocardiogram data for verification. Initial experiments on datasets representing twenty-nine and seventy-five persons indicate that fifteen or fewer heartbeats of data are sufficient in nearly all instances.

We hope that you enjoy reading this selection of papers that samples the variety of modalities and themes in current biometrics research.

Natalia A. Schmid
Stephanie Schuckers
Jonathon Phillips
Kevin Bowyer

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 184617, 19 pages
doi:10.1155/2009/184617

Research Article

Recognition of Faces in Unconstrained Environments: A Comparative Study

Javier Ruiz-del-Solar, Rodrigo Verschae, and Mauricio Correa

Department of Electrical Engineering, Universidad de Chile, Avenida Tupper 2007, 837-0451 Santiago, Chile

Correspondence should be addressed to Javier Ruiz-del-Solar, jruizd@cec.uchile.cl

Received 10 October 2008; Revised 31 January 2009; Accepted 13 March 2009

Recommended by Kevin Bowyer

The aim of this work is to carry out a comparative study of face recognition methods that are suitable to work in unconstrained environments. The analyzed methods are selected by considering their performance in former comparative studies, in addition to be real-time, to require just one image per person, and to be fully online. In the study two local-matching methods, histograms of LBP features and Gabor Jet descriptors, one holistic method, generalized PCA, and two image-matching methods, SIFT-based and ERCF-based, are analyzed. The methods are compared using the FERET, LFW, UCHFaceHRI, and FRGC databases, which allows evaluating them in real-world conditions that include variations in scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background and photographic quality. Main conclusions of this study are: there is a large dependence of the methods on the amount of face and background information that is included in the face's images, and the performance of all methods decreases largely with outdoor-illumination. The analyzed methods are robust to inaccurate alignment, face occlusions, and variations in expressions, to a large degree. LBP-based methods are an excellent election if we need real-time operation as well as high recognition rates.

1. Introduction

Many different face-recognition approaches have been developed in the last few years [1–4], ranging from classical Eigenspace-based methods (see, e.g., eigenfaces [5]), to sophisticated systems based on thermal's information, high-resolution images, or 3D models (see, e.g., [4, 6, 7]). However, the recognition of faces in unconstrained environments has not been completely solved [8]. In addition, some time-demanding applications, such as searching faces in nonannotated or partially annotated databases (i.e., news databases, the Internet, etc.) and HRI (Human-Robot Interaction), impose extra requirements of real-time operation, just one image per person and fully on-line operation (no off-line enrollment), which are difficult to achieve.

In this general context, the aim of this article is to carry out a comparative study of face-recognition methods by considering these requirements. The main motivation is the lack of direct and detailed comparisons of this kind of methods under the same conditions. The results of this comparative study are a guide for developers of face-recognition systems. As aforementioned, we concentrate ourselves on methods that fulfill the following requirements: (i) *full on-line operation*: no off-line enrollment stages. All processes must run on-line. The system has to be able to build the face database incrementally from scratch; (ii) *real-time operation*: the recognition process should be fast enough to allow real-time interaction in case of HRI or to search large databases in reasonable time (a few seconds or a couple of minutes depending on the application and the size of the database); (iii) *one single image per person problem*: one two-dimensional face image of an individual should be enough for his/her later identification. Databases containing just one face image per person should be considered. The main reasons are savings in storage and computational costs and the impossibility of obtaining more than one face image from a given individual in certain situations. In addition, we want to consider standard 2D images, and not high-resolution, 3D or thermal images that are not always available and that can slow down the recognition process; (iv) *unconstrained*

environments: no restrictions over environmental conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background, and photographic quality are required.

Thus, in this study two local-matching, one holistic, and two novel image-matching methods are selected by considering their fulfillment of the aforementioned requirements and their performance in former comparative studies of face-recognition methods [2, 9–12]. The two local-matching methods, namely, histograms of LBP (Local Binary Patterns) features [13] and Gabor-Jet features with Borda count classifiers [10] are selected considering their performance in the studies reported in [2, 10]. Among the holistic methods, a member of the eigenspace-based family of face-recognition methods is included, generalized PCA (Principal Component Analysis) with Euclidian distance and modified LBP features to achieve illumination invariance [11] (the restriction of *one single image per person* does not allow to include easily other members of the family). In addition, two novel face-recognition methods based on advanced image-matching methods are also considered: SIFT (Scale-Invariant Feature Transform) descriptors with local and global matching methods [12] and ERCF (Extremely Randomized Clustering Forest) of SIFT Descriptors used together with linear classifiers [14]. This last method, although not being real-time, is included for comparison purposes, because of the excellent results it has obtained in the LFW database [15].

The comparative study is carried out using the FERET [10], LFW (Labeled Faces in the Wild) [8], UCHFaceHRI [12], and FRGC (Face Recognition Grand Challenge) databases [16, 17]. We choose to use the very well-known FERET database, because it is one of the most employed face databases, and therefore it allows comparing results to other studies. In addition, we think that robustness when using a large database is also important and FERET contains more than 1,000 individuals. We include the LFW database because it is specially designed to study the problem of unconstrained face recognition. It corresponds to a set of more than 13,000 images of faces collected from the web, images which exhibit natural variability in pose, lighting, focus, resolution, facial expression, age, gender, race, accessories, make-up, occlusions, background, and photographic quality. The only constraint on these faces is that they were detected using the Viola-Jones face detector [18]; therefore, they correspond to frontal and quasifrontal faces. We also include in this study the new UCHFaceHRI, which is especially designed to compare face analysis methods for HRI. This database contains 30 individuals and includes images with natural variations in illumination (indoor and outdoor), scale, pose, and expressions. Finally, we consider experiments using the FRGC dataset, whose data corpus consists of 50,000 recordings, divided into training and validation partitions. We used FRGC's experiments 1 and 4, designed to measure progress on recognition from controlled and uncontrolled frontal face images. Thus, the comparative study includes 4 stages. (1) In the first stage all methods (except ERCF) are compared using the FERET database. Aspects such as variable illumination, alignment's accuracy, occlusions, and dependence on the database's size are measured, and the results are analyzed in terms of recognition rate and computational costs. (2) Some selected methods are further analyzed using the more challenging conditions defined by LFW. In addition to all the variability expressed in the LFW images, we analyze the dependence of the methods on the alignment's accuracy as well as on the amount of background and face's information considered in the analysis of the images. (3) The best variants of each of these methods (including selected distance's metrics and cropping-size for each case) are further analyzed using the natural requirements defined in the UCHFaceHRI database. (4) Finally, the best performing methods in all tests are analyzed and compared to state-of-the-art methods using the FRGC database. This study corresponds to an extended version of the one presented in [19].

This paper is structured as follows. The methods under analysis are described in Section 2. In Sections 3–6 the comparative analysis of these methods is presented. Finally, in Section 7 results are discussed, and conclusions are given.

2. Methods under Comparison

As mentioned above, the algorithms' selection criteria are their fulfillment of the defined requirements, and their performance in former comparative studies of face-recognition methods [2, 9–12]. In the comparison we decided to consider local-matching, holistic, and advanced image-matching methods.

Local-matching methods behave well when just one image per person is available [2], and some of them have presented very good results in standard databases such as FERET [10]. Thus, taking into account the results of [10], and our requirements of high-speed operation, we selected two methods to be analyzed. The first one is based on the use of histograms of LBP features, and the second one is based on the use of Gabor filters and Borda count classifiers.

When analyzing which holistic methods to include, the first idea was to consider methods based on eigenspace-decompositions (see a basic categorization in [9]). However, these methods normally fail when just one image per person is available, mainly because they have difficulties to build the required representation models. This difficulty can be overcome if a generalized face representation is built. Such representation can be built using a generalized PCA model. Thus, we incorporated to the study a face-recognition method based on a generalized PCA model.

We also decided to consider in this study advanced image-matching methods, which are not very popular in the face-recognition community, but which have been successfully applied in other computer vision contexts. Thus, taking into account that local interest points and descriptors (see, e.g., SIFT [20]) have been already used to solve successfully some other biometric problems (see, e.g., fingerprint verification [21] and off-line signature verification [22]), and as a first stage of complex face-recognition systems [6], we decided to test the suitability of a SIFT-based face-recognition system in this study. Finally, we also included

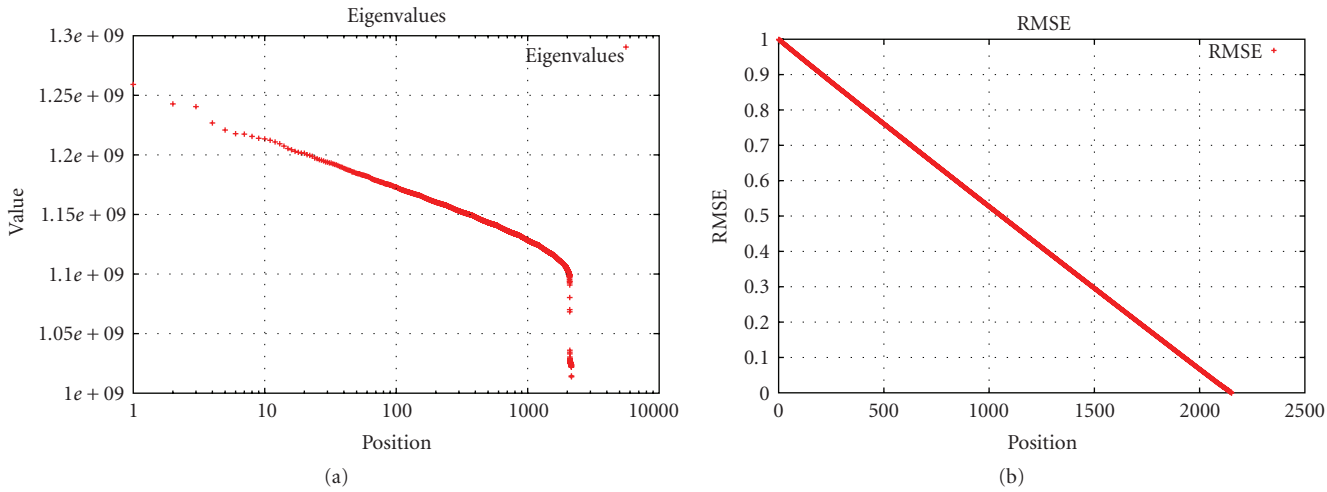


Figure 1: (a) Spectrum of the eigenvalues in the employed generalized PCA representation. Training set of size 2,152. (b) RMSE of the employed representation.

the recently proposed ERCF [14], a tree-based classification method designed to verify if a pair of images corresponds to the same object or not. The reason to include this last method in the comparison is the excellent results that have been obtained in recognizing faces in the LFW database [15]. The use of SIFT features for face authentication was investigated in [23] for the first time; however, no comparisons with other methods were presented.

The aforementioned selected methods are described in the next sections.

2.1. Generalized PCA. We implemented a face-recognition method that uses generalized PCA as projection algorithm, the Euclidian distance as similarity measure, and modified LBP features [24]. We used a generalized PCA approach, which consists on building a PCA representation in which the model does not depend on the individuals to be included in the final database, that is, on their face's images, because the PCA projection model is built using face's images that belong to a different set of persons. This allows applying this method in the case when just one single image per person is available. Our PCA model was built using 2,152 face images obtained from different face databases and the Internet. For compatibility with the results presented in [9], the model was built using face images scaled and cropped to 100×185 pixels and was aligned using eye's information. Using a similar approach to the one described in [16], we analyzed the validity of this generalized PCA representation by verifying that the main part of the eigenspectrum, that is, the spectrum of the ordered eigenvalues, is approximately linear between the 10th and 1,500th components, using a logarithmic scale for the components (see Figure 1(a)). The RMSE [9] was used as a criterion to select the appropriate number of components to be used. To achieve a RMSE between 0.9 and 0.5, the number of employed PCA components has to be in the range of 200 to 1,050 (see Figure 1(b)). Taking into account these results as well as the tradeoff between number of components and speed, we choose to implement two flavors of our system, one with 200 components and one with 500. Modified LBP features were used because according to the study presented in [11], this feature-space transformation (together with SQI) is one of the most suitable algorithms to achieve illumination compensation and normalization in eigenspace-based face-recognition systems.

2.2. LBP Histograms. Face recognition using histograms of LBP features was originally proposed in [13] and used by many groups since then. In the original approach, three different levels of locality are defined: pixel level, regional level, and holistic level. The first two levels of locality are realized by dividing the face image into small regions from which LBP features are extracted and histograms are used for efficient texture information representation. The holistic level of locality, that is, the global description of the face, is obtained by concatenating the regional LBP extracted features. The recognition is performed using a nearest neighbor classifier in the computed feature space using one of the three following similarity measures: histogram intersection, log-likelihood statistic, and Chi square. We implemented this recognition system, without considering preprocessing (cropping using an elliptical mask and histogram equalization are used in [13]), and by choosing the following parameters: (i) images divided in 10 (2×5), 40 (4×10), or 80 (4×20) regions, instead of using the original divisions which range from 16 (4×4) to 256 (16×16), and (ii) the mean square error as similarity measure, instead of the log-likelihood statistic. We also carried out preliminary experiments for replacing the LBP features by modified LBP features, but better results were always obtained by using the original LBP features. Thus, considering the 3 different image divisions and the 3 different similarity measures, we get 9 flavors of this face-recognition method.

2.3. Gabor Jets Descriptors. Local-matching approaches for face recognition are compared in [10]. The study analyzes

several local feature representations, classification methods, and combinations of classifier alternatives. Taking into account the results of their study, the authors implemented a system that integrates the best possible choice at each step. That system uses Gabor jets descriptors as local features, which are uniformly distributed over the images, one wavelength apart. In each grid position of the test and gallery image and at each scale (multiscale analysis), the Gabor jets are compared using normalized inner products, and these results are combined using the Borda count method. In the Gabor feature representation, only Gabor magnitudes are used, and 5 scales and 8 orientations of the Gabor filters are adopted. We implemented this system using all parameters described in [10] (filter frequencies and orientations, grid positions, face image size).

2.4. SIFT Descriptors. Wide-baseline matching approaches based on local interest points and descriptors have become increasingly popular and have experienced an impressive development in recent years. Typically, local interest points are extracted independently from both a test and a reference image and then characterized by invariant descriptors, and finally the descriptors are matched until a given transformation between the two images is obtained. Lowe's system [20] using SIFT descriptors and a probabilistic hypothesis rejection-stage is a popular choice for implementing object-recognition systems, given its recognition capabilities, and near real-time operation. However, Lowe's system's main drawback is the large number of false positive detections. This drawback can be overcome by the use of several hypothesis rejection stages as, for example, in the L&R system [21]. This system has already been used in the construction of robust fingerprint verification systems [21] and for off-line signature verification [22]. Here, we use the L&R system to build a face-recognition system, with three different flavors. In the first one, *Full*, all verification stages defined in [21] are used, while in the second one, *Simple*, just the probabilistic hypothesis rejection stages are employed. In the third one, *Matches*, the number of matching key points without using any rejection stages is considered.

2.5. ERCF: Extremely Randomized Clustering Forest. In [14] a robust method to learn a similarity measure is proposed, which allows to discriminate whether a pair of object's images corresponds to the same object or not (the objects could be faces). The method is especially designed to be used in object recognition problems and makes use of ERCF and SIFT descriptors. The learning is done for specific object classes, such as frontal faces or specific views of cars. The method basically consists of three stages. In the first stage, pairs of similar patches, measured in terms of a normalized cross-correlation, are selected. In the second stage, each pair of patches is coded (quantized) by means of an ERCF of SIFT descriptors. ERCF is a sparse representation of the image that is built using classification trees. Each classification tree is generated using SIFT descriptors and used for vector quantization. In the third stage, the quantized pairs of patches are used to build a feature vector, which is finally used to evaluate the similarity of the image pair using a linear classifier. In this study we use the author's implementation of the method, available on http://lear.inrialpes.fr/people/nowak/similarity/index.html.

2.6. Notation: Methods and Variants. We use the following notation to refer to the methods and their variations: A, B, and C. (i) A describes the name of the face-recognition algorithm: H is Histogram of LBP features, PCA is generalized PCA with modified LBP features, GJD is Gabor Jets Descriptors, SD is L&R system with SIFT descriptors, and ERCF is Extremely Randomized Clustering Forest; (ii) B denotes the similarity measure: HI is Histogram Intersection, MSE is Mean square error, XS is Chi square, BC is Borda Count, and EU is Euclidian Distance, except for the case of SD and ERCF, which do not use any explicit distance's measure; (iii) C describes additional parameters: number of divisions in the case of the LBP-based method, number of principal components in the case of PCA, size of the reference-set for the GJD case (see explanation in Section 4), and flavor (*Full*, *Simple,* or *Matches*) in the case of SD.

3. Comparative Study Using the FERET Database

Face images are scaled and cropped to 100×185 pixels and 203×251 (for compatibility with former studies [9, 10]), except for the case of the PCA method in which, for simplicity, just one image size (100×185) was employed (the generalized PCA model depends on the image cropping). In all cases, faces are aligned by centering the eyes in the same relative positions, at a fixed distance between the eyes, which was 62 pixels for the 100×185 size images and 68 pixels for the 203×251 size images. The amount of face information and background contained in the cropped images can be measured using the normalized width (*nw*) and height (*nh*), defined as the image width/height divided by the distance between eyes. This means that the *nw*/*nh* of the analyzed images are 1.6/3.0 for images of 100×185 pixels and 3.0/3.7 for images of 201×253 pixels. To compare the methods we used the FERET evaluation procedure [25], which established a common data set and a common testing protocol for evaluating semiautomated and automated face-recognition algorithms. We used the following sets: (i) *fa* set (1,196 images), used as gallery set (contains frontal faces of 1,196 people); (ii) *fb* set (1,195 images), used as test set 1 (in *fb* subjects were asked for a different facial expression than in *fa*); (iii) *fc* set (194 images,) used as test set 2 (in *fc* pictures were taken under different lighting conditions). In all cases the information about the eyes' position provided by FERET was used for the face alignment.

In addition, we carried out extra experiments by adding noise to the position of the eyes in the *fb* set, and also by adding artificial occlusions in these images. The goal was to test the robustness of the different methods. Finally, we also compared the computational performance of the methods. ERCF was not considered in this first comparison, neither in the FRGC experiments, because the method is not real-time and, to carry out all the experiments, it takes a very

Table 1: FERET fa-fb and fa-fc tests. Top-1 recognition rate. Noise in eye positions and face occlusion is tested in the fa-fb test. OR: Original. OC: Original plus Occlusion. The best results for each condition are presented in bold. Methods that have differenc

Method	100×185						203×251					
	fa-fb					*fa-fc*	*fa-fb*					*fa-fc*
	OR	Noise in eye positions			OC		OR	Noise in eye positions			OC	
		2.5%	5%	10%				2.5%	5%	10%		
H-HI-10	95.6	**95.0**	**91.3**	**81.8**	93.6	12.9	95.1	23.7	22.4	16.4	93.4	50.0
H-MSE-10	95.6	**95.0**	**91.3**	**81.8**	93.6	12.9	95.1	23.7	22.4	16.4	93.4	50.0
H-XS-10	95.7	94.7	**92.3**	**82.2**	78.4	14.9	95.1	41.3	39.4	31.0	86.1	60.8
H-HI-40	**96.5**	**96.0**	89.7	70.9	95.1	57.2	96.5	41.0	39.7	27.5	95.2	85.1
H-MSE-40	**96.5**	**96.0**	89.7	70.9	95.1	57.2	96.5	41.0	39.7	27.5	95.2	85.1
H-XS-40	95.5	93.6	87.0	67.4	92.1	47.4	97.4	76.6	71.4	53.8	95.0	88.1
H-HI-80	**97.2**	**95.6**	90.1	71.5	**96.7**	71.1	96.9	61.1	55.7	40.6	96.6	91.8
H-MSE-H-MSE-80	**97.2**	**95.6**	90.1	71.5	**96.7**	71.1	96.9	61.1	55.7	40.6	96.6	91.8
H-XS-80	96.3	94.1	88.3	68.0	94.4	62.9	97.4	87.8	83.9	64.9	**96.7**	92.8
PCA-MSE-200	73.1	55.9	40.7	16.2	63.6	52.1	—	—	—	—	—	—
PCA-MSE-500	76.1	60.3	42.9	16.0	64.9	57.2	—	—	—	—	—	—
GJD-BC	91.4	89.6	85.0	63.1	74.5	**79.9**	**98.5**	95.0	93.6	73.9	**97.7**	**99.0**
SD-FULL	74.3	75.7	73.5	71.5	67.3	7.7	97.1	**96.2**	**95.7**	**95.3**	95.6	67.5
SD-SIMPLE	73.1	75.3	73.1	71.0	68.6	5.7	**97.5**	**96.7**	**96.4**	**96.2**	95.3	63.9
SD-MATCHES	70.3	70.3	67.6	66.7	58.6	4.7	93.9	93.7	94.6	92.3	90.1	44.0

long time. However, the method is considered in the LFW and UCHFaceHRI experiments.

Original fa-fb Test. Table 1 shows *top-1* RR (Recognition Rate) achieved by the different methods under comparison in the original *fa-fb* test, which corresponds to a test with few variations in the acquisition process (uniform illumination, no occlusions). We use the information of the annotated eyes, without adding any noise. From the experiments the following can be observed.

(i) The results obtained with our own implementation of the methods are consistent with those of other studies. The best H-X-X flavors achieved in the 203×251 face images a similar performance (97.4% versus 97%) than the one reported in the original work [13]. GJD-BC achieved a slightly lower performance (98.5% versus 99.5%) than in the original work [10]. When comparing these results to the ones obtained by other authors using more complex systems based on hybrid Gabor-LBP [26], Gabor-Fisher [27], or Fisher-Gabor-LBP [28]—98%, 99% and 99.6%, respectively, we observe that those results are similar or slightly better than ours; however, our systems are much simpler. There are no reports of the use of the generalized PCA or SIFT methods in these datasets.

(ii) The best results ($\sim$ 98.5%) are obtained by GJD-BC, followed by the SD and H-X-80 variants, all using 203×251 images. Nevertheless, other H-X-X variants also get very good results. Interestingly, some H-X-X variants get $\sim$ 97% even using 100×185 size images. The results obtained by the PCA methods are the lowest.

(iii) The performance of the GJD-X-X and SD-X methods depends largely on the normalized size of the cropped images, probably because the methods use information about face shape and contour, which does not appear in the 100×185 images.

Eye Detection Accuracy. Most of the face-recognition methods are very sensitive to face alignment, which depends directly on the accuracy of the eye detection process; eye position is usually the primary, and sometimes the only, source of information for face alignment. For analyzing the sensitivity of the different methods on the eye position's accuracy, we added white noise to the position of the annotated eyes in the *fb* images (see example in Figure 2(a)). The noise was added independently to the x and y eye positions. Table 1 shows the *top-1* RR achieved by the different methods. Our main conclusions are the following.

(i) SD-X methods are almost invariant to the position of the eyes in the case of using 203×251 face images. With 10% error in the position of the eyes, the top-1 RR decreases in just $\sim$2%. The invariance is due to the fact that this method aligns test and gallery images by itself.

(ii) In all other cases the performance of the methods decreases largely with the error in the eye position, probably because they are based on the matching between holistic or feature-based representations of the images. However, if the eye position error is bounded to 5%, the results obtained by some H-X-X variants using 100×185 face images ($\sim$90%) are still acceptable.

Partial Face Occlusions. To analyze the behavior of the different methods in response to partial occlusions on the face area, *fb* face images were divided into 10 different areas (2 columns and 5 rows). One of these areas was randomly selected and its pixels set to 0 (black). See example in Figure 2(b). Thus, in this test each face image of *fb* has one tenth of its area occluded. Table 1 shows the *top-1* RR achieved by the different methods. The main conclusions are as follows.

(i) GJD-BC and H-XS-80 achieve the highest top-1 RR in the 203×251 case, 97.7% and 96.7%, respectively.

(ii) Some H-X-X variants are very robust to face occlusions (e.g., H-HI-10, H-MSE-10, H-X-80) independently of using face images of 100×185 or 203×251 pixels.

(iii) SD-X variants are also robust to occlusions in the 203×251 case.

(iv) PCA is not robust to occlusions; its performance decreases in about 10% compared to the nonoccluded case.

Variable Illumination. Variable illumination is one of the factors with strong influence in the performance of face-recognition methods. Although there are some specialized face databases for testing algorithm invariance against variable illumination (e.g., PIE, YaleB), we choose to use the *fa-fc* test set, because (i) it considers a large number of individuals (394 versus 10 in Yale B and 68 in PIE), and (ii) the illumination conditions are more natural in the *fc* images. Table 1 shows the *top-1* RR achieved by the different methods in this test. The main conclusions are as follows.

(i) The results obtained with our own implementation of the methods are consistent with those of other studies. The best H-X-X flavors achieve in the 203×251 case a higher performance (92.8% versus 79%) than the one reported in the original work [13], probably due to the different image's partitions that we use in our implementation. The best GJD-BC flavors achieve a slightly lower performance (99% versus 99.5%) than the original implementation [10]. When comparing these results to the ones obtained by other authors using more complex systems based on hybrid Gabor-LBP [26], Gabor-Fisher [27], or Fisher-Gabor-LBP [28]—98%, 97% and 99%, respectively—we observe that those results are similar to ours; however, our systems are much simpler. There are no reports of the use of the generalized PCA or SIFT methods in the same database.

(ii) Best performance is achieved by GJD-BC (99%), and second best by H-XS-80 (~93%). In both cases using images cropped to 203×251 pixels.

(iii) In all cases much better results were obtained using larger face images (203×251).

(iv) PCA-X-X and SD-X methods show a low performance in this dataset.

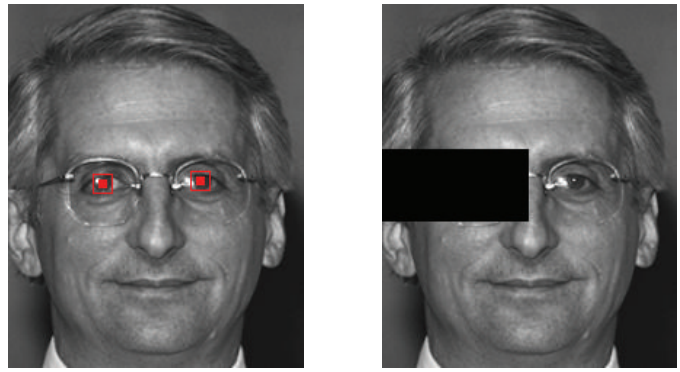
(v) H-X-X methods with a large number of partitions show better performance than variants with a small number of partitions (~93% versus ~50% in the case of using 203×251 images and ~71% versus ~13% in the case of 100×185 images).

Computational Performance. As aforementioned one of the requirements imposed to the methods under comparison is real-time operation. In addition, the memory required by the different methods is very important in some applications where memory could be an expensive resource. Table 2 shows the computational and memory costs of the different methods under comparison, when images of 100×185 are considered. For the case of measuring the computational costs, we considered the feature-extraction time (FET) and the matching time (MT). In the case of measuring memory costs, we considered the database memory (DM), which is the required amount of memory to have the whole database (features) in memory, and the model memory (MM), which is the required amount of memory to have the method model, if any, in memory (PCA matrices for the PCA case and filter bank for the Gabor method). We show the results for databases of 1, 10, 100, and 1,000 individuals (face images). If we consider that in many applications the database size is in the range 10–100 persons, the fastest methods are the H-X-X ones. The second fastest methods are the GJD-BC ones. To achieve real-time operation with a database of 100 or fewer elements, all methods are suitable, except PCA-based methods. In databases of 10–100 individuals, H-X-X and GJD-X-X require less than 8 MBytes of memory (they do not need to keep a model in memory). In the case of H-X-X methods, the required memory increases linearly with the number of partitions.

Summary. As a result of all these experiments we decided to further test these methods in more demanding conditions using the LWF and UCHFaceHRI databases. In this stage we discarded the PCA method, because in all tests it turns to be the weakest one, getting always the lowest scores.

4. Comparative Study Using the LFW Database

The LFW database [8] consists of 13,233 images faces of 5,749 different persons, obtained from news images by means of a face detector (Viola-Jones detector [18]). There are no eyes/fiducial point annotations; the faces were just aligned using the output of the face detector. The faces aligned using the funneling algorithm [29] are also available. The images of the LFW database have a very large degree of variability in the face's expression, age, race, background, and illumination conditions (see Figure 3). Also, unlike other databases, the recognition is only to be done by comparing pairs of images, instead of searching for the most similar face in the database. The idea is that the algorithm being evaluated is given a pair of images, and it has to output whether the two images correspond to the same person or not. There are two evaluation settings already defined by the authors of the LFW: the image restricted setting and the image unrestricted setting. The image restricted setting is the



(a) (b)

Figure 2: Face image of 203 × 251 pixels. (a) Image with eye position (red dot) and square showing a 10% error in the eye position. (b) Image with partial occlusion.

Table 2: Computational and memory costs. FET: Feature Extraction Time. MT: Matching Time. PT: Processing Time. DM: Database Memory. MM: Model Memory. TM: Total Memory. Time measures are in milliseconds; memory measures are in Kbytes. DB sizes of 1, 10, 100, and 1,000 faces are considered. An image size of 100 × 185 pixels is considered.

Method	FET	MT	PT (FET + MT)				DM	MM	TM (DM + MM)			
			1	10	100	1000			1	10	100	1000
H-X-10	15	0.11	15	16	26	120	11	0	11	110	1100	11000
H-X-40	15	0.29	15	18	44	305	41	0	41	410	4100	41000
H-X-80	15	0.42	15	19	57	435	80	0	80	800	8000	80000
PCA-MSE-200	170	0.02	170	170	172	190	0,8	137800	137801	137808	137878	138585
PCA-MSE-500	360	0.02	360	360	362	380	2	137800	137802	137820	137996	139757
GJD-BC	50	0.25	50	53	75	300	33	1240	1273	1572	4559	34427
SD-X	4.7	1.03	6	15	108	1036	428	0	428	4284	42845	428451

most difficult one, and it is the one considered here. Under this setting the only information that the algorithm can use is the image pair; no information of the identity of the faces in the images can be used, that is, the algorithm is restricted to work only using the image pair at hand. The systems are trained (if required) and evaluated using a 10-fold validation procedure, where the folds are symmetric in the sense that the number of matching pairs and nonmatching pairs is the same. See [8] for details.

In the first experiments (Sections 4.1 and 4.2) images were cropped to 100 × 185 pixels (see Figures 4(a) and 4(b)). Given that the mean distance between eyes is 42 pixels, the normalized width and height are $nw = 2.4$ and $nh = 4.4$. We analyze and compare two cases, unaligned and aligned. In the unaligned case, face images have a coarse alignment, which is the one produced by the face detector that was used to obtain the images. In the aligned case, the funnelling algorithm is used to obtain a more accurate alignment. Afterwards, in Section 4.3, all methods are analyzed, considering different region sizes, where the face's images are cropped considering larger and smaller bounding boxes. These experiments analyze the effect of using different amounts of background and face's information in the recognition process (see Figure 4).

Given that the LFW database only requires comparing pairs of faces, and that an important part of the GJD method is the ranking done using Borda count, we had to adapt it to this condition. To accomplish this, we first define a reference set of faces, which is built by randomly selecting face images (e.g., 50) of the same characteristics than the ones under comparison. Then, we take one of the two face images under comparison, and we compare it against the images of the reference set plus the second image under comparison. The relative ranking, computed using Borda count, obtained by the second face image is considered as a measure of the similarity between the pair of images. To obtain a symmetric similarity measure, we repeated the same procedure by switching the roles of the two images, and then averaging the two obtained rankings. The average value was taken as the final similarity measure of the pair of images. We considered three different sizes for the reference set: 10, 50, and 100 faces. To show the importance of using Borda count method, results using the Euclidean distance between the GJD descriptors are also given for comparative purposes.

SD-Full does not work properly in this database, and consequently its results are omitted. In addition, when using the LBP-based methods, HI and MSE always obtained the same recognition results, and therefore the HI case is also omitted. The results corresponding to ERCF consider complete images (250 × 250), and they correspond to those

Figure 3: Examples of faces from the LFW, randomly selected from people with name starting with A.

Figure 4: Examples of faces with different cropping (LFW database). (a) 100×185, unaligned; (b) 100×185, aligned; (c) 81×150, aligned; (d) 122×225, aligned; (e) 125×125, aligned; (f) 250×250, aligned. The last row shows the normalized image's width and height (nw/nh). The images are shown maintaining their relative sizes.

presented in [15]. We use the original results although in our own experiments we got very similar results.

4.1. Experiments Using Unaligned Faces. Table 3 (second and third columns) shows the results for all methods under comparison in the unaligned LFW database. It should be remembered that in the unaligned LFW, all images have a coarse alignment. In all cases (except for ERCF), regions of 100×185 pixels containing the centered face in the 250×250 image were cropped ($nw = 2.4$ and $nh = 4.4$). As it can be observed, the results obtained with our own implementation of the methods are consistent with those of other studies results (in terms of the relative order of the classification accuracy). However the accuracies are low, going form 60% to 72%, values that show the difficulty of the database at hand. In the case of the H-X-X methods,

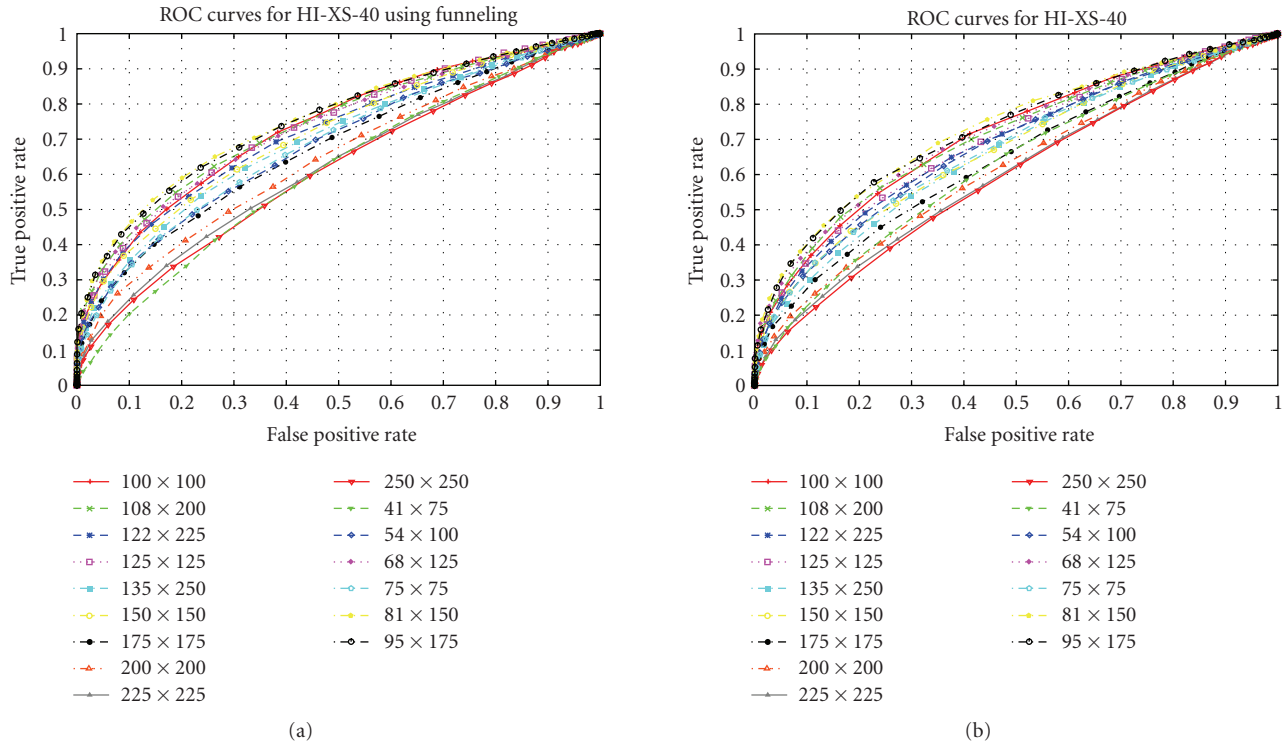


Figure 5: Effect of the image's region size on the performance of the H-XS-40. (a) Faces aligned using funneling; (b) unaligned faces.

best results are obtained with H-X-80, that is, when using the largest number of divisions. The difference between using the Chi-Square and the Mean Square Error is not significant, although the Chi-Square measure gives slightly better results in all cases. For the method based on the GJD, best results are obtained when using the proposed Borda count methodology (it increases the performance in circa 2% over the Euclidean distance); 100 reference images gives slightly better results than 10 or 50. Both methods based on SD got the lowest performance (about 60%–62%). The performance of ERCF is quite good, being ∼ 4% larger than the second best method (GJD-BC-100).

4.2. Experiments Using Aligned Faces. The faces were aligned using the funneling algorithm [29]. Funneling is an unsupervised algorithm for object alignment based on the concept of congealing. Congealing basically consists of searching a sequence of transformations (in this case affine transforms and translations) that are applied to a set of images in order to minimize an entropy measure on the set of images. After having built the congealing model, the transformations can be applied to an unseen image (funneling it) to obtain an aligned image. The main advantage of this method is that it can work in complex objects and that it does not require any labeling during training.

Table 3 (last two columns) shows the results for all methods under comparison using aligned faces. As in the case of unaligned faces (except for ERCF); the face region was cropped considering a region of 100×185 pixels centered in the 250×250 image ($nw = 2.4$ and $nh = 4.4$). Compared to the case of unaligned faces, all methods, but GJD-X-X and SD-Simple, improve or maintain their performance. The H-X-X methods obtain the largest improvement, 2% to 3%, depending on the variant being considered. Again, in the case of LBP based methods, best results are obtained with H-X-80, that is, when using the largest number of divisions, and the Chi-Square distance's measure, with a performance similar to GJD. For the variants based on the GJD, best results are obtained when Borda Count is used (it increases the performance in circa 3% over the Euclidean distance), and 100 reference images gives slightly better results than 10 or 50. However, in this case, the results were slightly worse than the ones obtained for the case of unaligned faces. Again, best results are obtained by ERCF, but this time being about 5% over the second best method.

4.3. Experiments Using Different Windows Sizes. In this section we analyze the effect of using different region sizes in the performance of the analyzed methods. Note that increasing the size of the regions corresponds to adding or removing different amounts of background to the region being analyzed, given that we are not decreasing the scale of the faces. The experiments were performed considering squared image regions, ranging from 50×50 to 250×250, with a step of 25 pixels, and considering regions of ratio $1 : 1.85$ (as in the previous section), ranging from 41×75 to 135×250, with a step of 25 pixels. Results are presented in Figures 5–8 in form of ROC curves. By observing the results, the first thing we can see is the importance of the relative size of the region, that is, the amount of face and

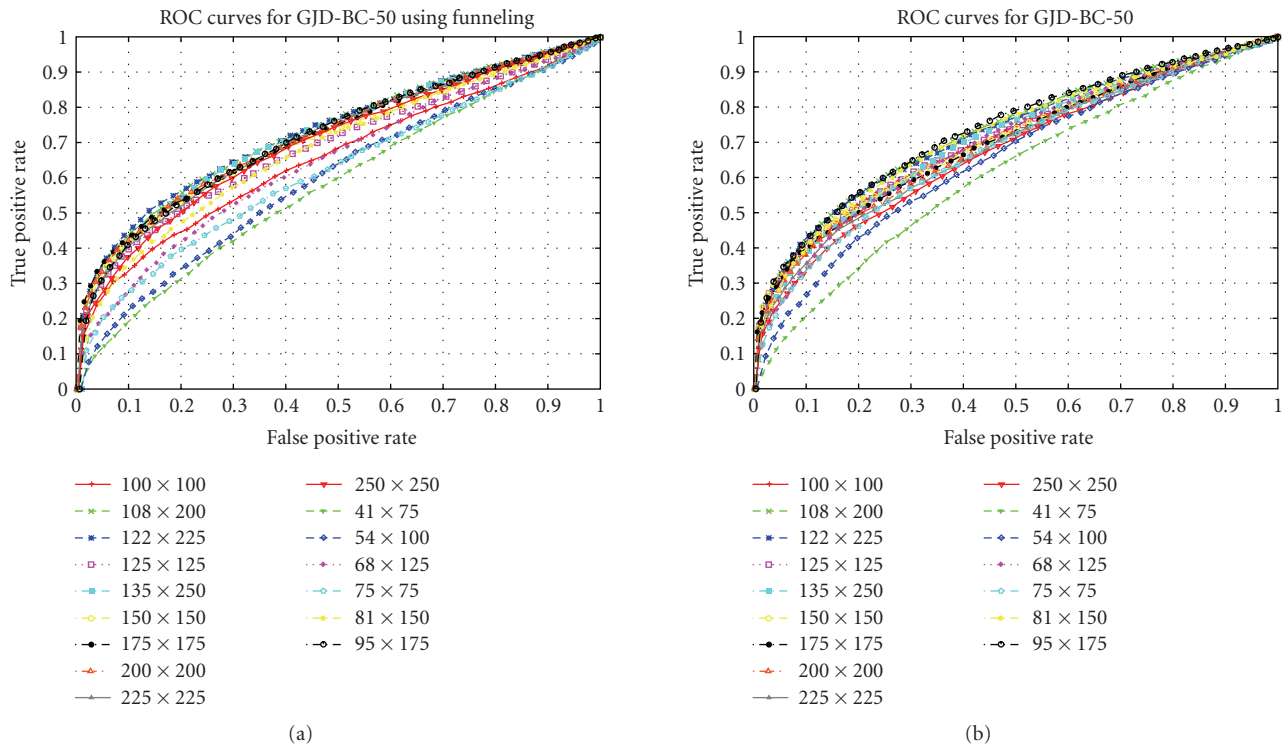


FIGURE 6: Effect of the image's region size on the performance of the GJD-BC-50 method. (a) Faces aligned using funneling; (b) unaligned faces.

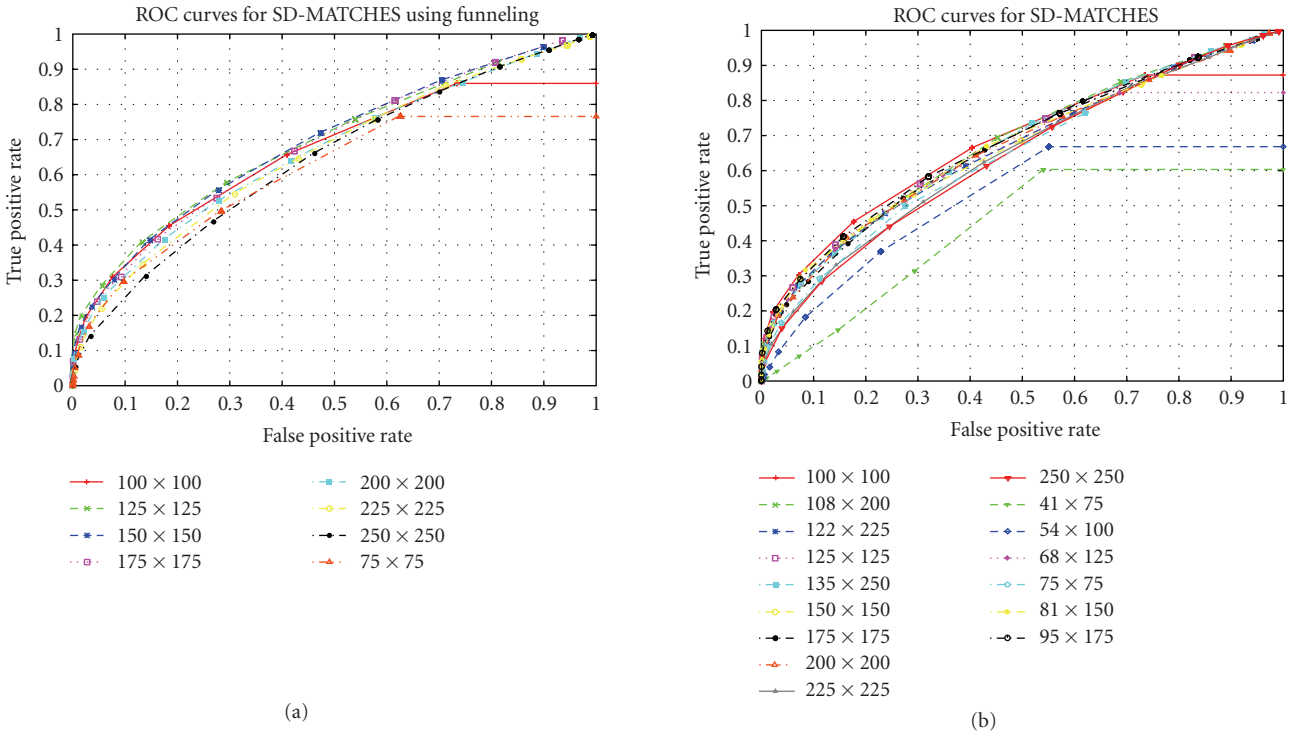


FIGURE 7: Effect of the image's region size on the performance of SD-MATCHES method. (a) Faces aligned using funneling, (b) unaligned faces.

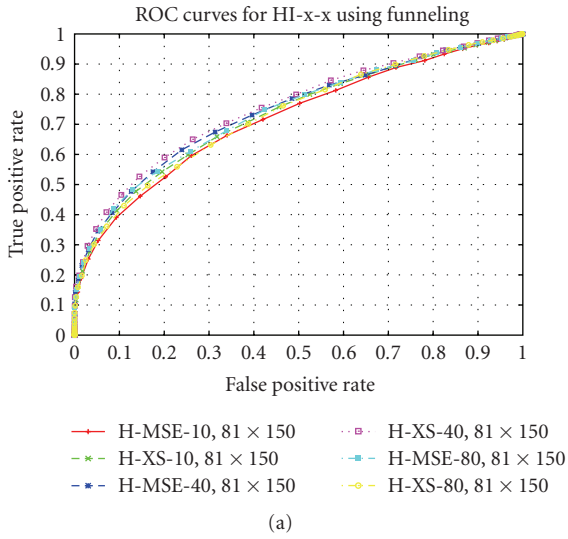


Figure 8: Comparison of the best working flavors of each method when funneling is used: (a) H-X-X, (b) GJD-BC, (c) SD-MATCHES.

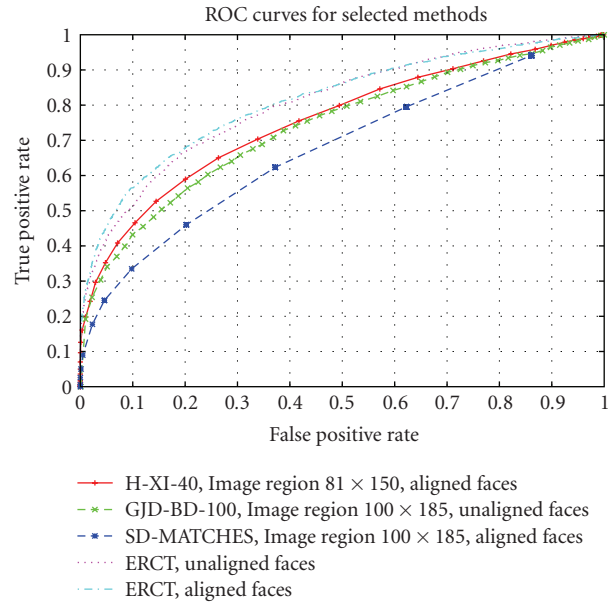


Figure 9: ROC curves of the best working variant of each method. Experiments were performed on faces aligned using funneling.

background information being analyzed on the performance of all algorithms. In Figure 4, the different amounts of face and background information that each image's size includes can be observed. The second thing is that in all cases (independently of the distance's measure and the method's parameters), small region sizes present the worst results, followed by the largest region sizes. Best results are obtained using medium-size regions.

Figure 5 shows the results for HI-XS-40. Best results are obtained for aligned images of size 81 × 150 (see also Figure 8(a)), which contains some background, but not very much (see Figure 4(c)). In the case of unaligned images, best results are obtained for images of size 95×175. Similar results were obtained when using 10 and 80 image divisions. For a fixed number of divisions, the Chi-Square measure works better than the mean square error (results not shown for space reasons).

Figure 6 shows the results for GJD-BC-50. Best results are obtained for aligned images of size 122 × 225 (see Figure 4(d)). In the case of unaligned images, best results are obtained for images of size 95 × 175. The most important thing that must be noticed here (see also Figure 8(b)) is that when the optimal image size is used and aligned faces are considered, using 10, 50, or 100 reference images; very similar results are given (in terms of MCA 0.6838, 0.6838, and 0.6847, resp.). This also holds when unaligned faces are used, but the difference is slightly larger (in terms of MCA 0.6752, 0.6780, and 0.6808, resp.). The experiments with reference sets of 10 and 100 images are not shown for space reasons.

Figures 7 and 8(c) show the results for SD-MATCHES. Best results are obtained again for aligned images; in this case a size of 125 × 125 gives better results (see Figure 4(e)). In the case of unaligned images, best results are obtained for

Table 3: Correct classification rates (LFW database, restricted setting). Experiments were performed on cropped regions of size 100 × 185 (**nw** = 2.4 and **nh** = 4.4), except for ERCF that considers the full image. MCA: *Mean classification accuracy.* SME: *Standard error of the mean.* In bold are the best results of each method.

	Without alignment		With alignment (funneling)	
Method	MCA	SME	MCA	SME
H-MSE-10	0.6375	0.0049	**0.6585**	0.0046
H-XS-10	0.6500	0.0043	**0.6668**	0.0044
H-MSE-40	0.6217	0.0055	**0.6527**	0.0057
H-XS-40	0.6383	0.0064	**0.6650**	0.0059
H-MSE-80	0.6527	0.0047	**0.6725**	0.0032
H-XS-80	0.6532	0.0053	**0.6785**	0.0055
GJD-EU	**0.6410**	0.0084	0.6375	0.0071
GJD-BC-10	**0.6777**	0.0080	0.6753	0.0082
GJD-BC-50	**0.6770**	0.0075	0.6742	0.0061
GJD-BC-100	**0.6798**	0.0065	0.6762	0.0069
SD-MATCHES	0.6015	0.0049	**0.6215**	0.0036
SD-SIMPLE	**0.6295**	0.0071	0.6288	0.0051
ERCF (from [15])	0.7245	0.0040	**0.7333**	0.0060

Table 4: Correct classification rates of the best methods (LFW database, restricted setting). MCA: *Mean Classification Accuracy.* SME: *Standard Error of the Mean.*

Method	Region Size	MCA	SME
SD-MATCHES, aligned faces	125 × 125	0.6410	0.0062
H-XS-40, aligned faces	81 × 150	0.6945	0.0048
GJD-BC-100, aligned faces	122 × 225	0.6847	0.0065
ERCF aligned faces (from [15])	250 × 250	0.7333	0.0060

Table 5: Processing Time. Time measures are in milliseconds. We carried out the experiments on a computer running Linux with an Intel Core 2 Duo E6750 2.66 GHz (2 GB RAM). FET/MT: Feature Extraction/Matching Time.

Method	H	H	H	GJD	GJD	GJD	GJD	SD	ERCF
Parameters	X-10	X-40	X-80	BC-1	BC-10	BC-50	BC-100	X	From [14]
FET (ms)	2.45	2.45	2.45	62	62	62	62	4.7	—
MT (ms)	0.033	0.118	0.230	0.37	2.63	5.59	15.55	64.7	2000
Image size	81 × 150			122 × 225				125 × 125	100 × 185

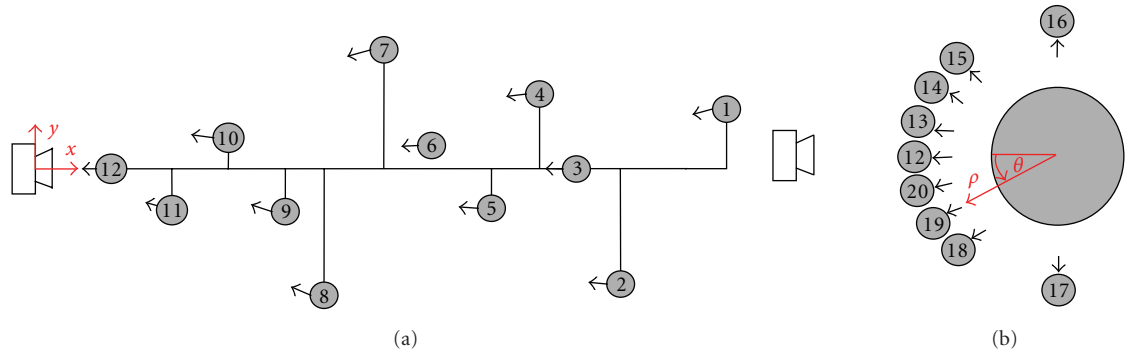


Figure 10: Experimental setup for image acquisition at different (a) distances and (b) angles. Arrows indicate the angular pose of the subjects. (a) Cartesian coordinates of acquisition points (in centimetres) relative to the camera's focus: P1 (1088,90), P2 (906,−180), P3 (785,0), P4 (755,151), P5 (665,−51), P6 (574,30), P7 (514,181), P8 (423,−181), P9 (332,−61), P10 (272,30), P11 (181,−61), and P12 (90,0). (b) Polar coordinates of acquisition points (radius in centimetres and angles in degrees) relative to the camera's focus: P16 (90,90°), P15 (90,45°), P14 (90,30°), P13 (90,15°), P20 (90,−15°), P19 (90,−30°), P18 (90,−45°), and P17 (90,−90°).

images of size 100×100. In all cases, best results are obtained with the SD-Matches variant. Very low performance results are obtained by the SD-Simple variant.

Finally, Figure 9 shows the ROC curves of the best variant of each method.

4.4. Discussion. If one analyzes the performance obtained by the different methods, ERCF obtains clearly the best results (see Table 4). Best LBP-based method (H-XS-40) is almost 3.9% below ERCF, and about 1% over GJD's best method (GJD-BC-100). However, if one now analyzes the processing speed of the methods, the best variant of LBP-based methods (H-XS-40) is at least 400 times faster than ERCF (see Table 5), and 30 times faster than the best Gabor method (GJD-BC-100). The high processing time of ERCF and GJD can be too restrictive for some applications, in particular in the ones that require real-time operation (e.g., HRI) as well in applications where very large amounts of data are being analyzed (e.g., search in a very large multimedia database). The Borda count ranking of each of the features is the slowest operation of GJD, while the slowest part of ERCF corresponds to the computation of the normalized cross-correlation in the selection of pairs of regions to be quantized using ERCF. However, it should be noted that in a face identification scenario, as the one reported for the FERET case, it is not required that GJD use a reference set. In this case (GJD-BC-1 in Table 5), the method needs about 63 milliseconds to analyze a face image.

It is also interesting to analyze which kind of information uses each method by looking at the optimal regions they use. The regions are shown in Figure 4 as well as the normalized width and height of the face images (last row). SD methods, specifically SD-Matches that has an optimal region size of 125×125 (see Figure 4(e)), show better performance when there is the fewest possible background in the image, but without removing any part of the face. The methods need as much as possible face information to obtain a correct matching. However, the background disturbs the matching process (a face keypoint could be matched to a background keypoint). In the case of LBP-based methods, specifically for H-XS-40 that has an optimal region size of 81×150 (see Figure 4(c)), it seems that some background but not much helps. Probably, this additional information about the face's contour helps the recognition process. Finally, in the case of the GJD methods, specifically GJD-BC-100 that has an optimal region size of 122×225 (see Figure 4(d)), the image contains much more background. The reason is twofolds: (i) the Gabor-filters encode information about the contour of the face, and (ii) large regions allow the use of large filters, which encoded large-scale information.

It is important to compare the optimal region sizes of the methods in the LWF, with the sizes used in the FERET experiments. However, it should be noted that the images in both databases have different resolutions. Therefore, instead of comparing region sizes, normalized image's width (nw) and height (nh) need to be used. In our FERET experiments, the nw/nh values are 1.6/3.0 for the 100×185 case, and 3.0/3.7 for the 203×251 case. The nw and nh values of the optimal region sizes, in the LWF case, are shown in Figure 4. By comparing these normalized values we observe that there is a concordance. (i) SD and GJD methods behave much better when normalized sizes of 3.0/3.7 are used in FERET, and in LFW behave better with values of 3.0/3.0 for the case of the SD method, and 2.9/5.4 for the case of the GJD method. (ii) In the case of the H-XS-40 method, similar results are obtained in FERET with 1.6/3.0 or 3.0/3.7, which is concordant with the selected values of 2.4/4.4 in LWF. Naturally, the normalized values in both databases are not the same, because in the FERET case we decided to use just two, fixed image's sizes, while in the LWF we allow the methods to choose the best values.

Finally, it is interesting to analyze how much the methods' performance depends on the alignment's accuracy. By observing Table 3, it can be seen that the methods with the largest dependence on the alignment's accuracy are the H-X-X. These results are consistent with the one obtained in the FERET database. On the opposite site, SD methods are very robust to alignment errors, which is also consistent with the results obtained in the FERET case. As it can be noticed, GJD performs worse when alignment is used. We think this is related with the way in which the used alignment method (funneling) works. Funneling aligns the whole face (shape), and not the eyes. As observed in the results obtained for FERET, GJD seems to be very sensible to good eyes' alignment.

5. Comparative Study Using Real HRI Database

The UCHFaceHRI database was built with the goal of allowing the study of face analysis methods in tasks such as detection, recognition, and relative pose determination of humans using face information, for HRI (Human-Robot Interaction) applications. The database contains images from 30 individuals, which were taken in 20 different relative camera-individual poses (see acquisition points in Figure 10), in outdoor and in indoor settings, at a resolution of 1024×768 pixels. Five different face expressions were considered for the case of the frontal face (P12 acquisition point): neutral expression, surprised, angry, sad, and happy. Thus, the database contains 48 images for each individual. Each of these 48 face images is specified as Fjkl, where j indicates that the image was taken at the acquisition point Pj and k indicates which expression is associated to this image (neutral: $k = a$, surprised: $k = b$, angry: $k = c$, sad: $k = d$, happy: $k = e$). This index is valid only in the case of images taken in the acquisition point P12. Finally, l indicates if the image was taken in an indoor ($l = i$) or an outdoor ($l = o$) environment. Figure 11 shows the 24 indoor images corresponding to a given individual. The database can be downloaded in [30].

In all experiments the F12ai face images composed the gallery set. We define 14 specific and global test sets, to analyze the methods' invariance to the scale, orientation, and expression of the faces, considering indoor and outdoor illumination conditions as follows.

(i) Scale test sets. S-I: Scale Indoor (images F10i-F11i), S-O: Scale Outdoor (images F10o-F12o).

(ii) Expression test sets. E-I: Expression-Indoor (images F12bi-F12ei), E-O: Expression-Indoor (images F12bo-F12eo).

(iii) Rotation test sets. R-I: Rotation-Indoor (images F13i, F14i, F19i, F20i), R-I/15: Rotation-Indoor in 15 degrees (images F13i, F20i), R-I/30: Rotation-Indoor in 30 degrees (images F14i, F19i), R-O: Rotation-Outdoor (images F13o, F14o, F19o, F20o), R-O/15: Rotation-Outdoor in 15 degrees (images F13o, F20o), R-O/30: Rotation-Outdoor in 30 degrees (images F14o, F19o).

(iv) Global test sets. Scale: S = S-I + S-O, Expression: E = E-I + E-O, Rotation: R = R-I + R-O, Global Indoor: G-I = S-I + E-I + R-I, Global Outdoor: G-O = S-O + E-O + R-O, and Global: G = D + E + R = G-I + G-O.

In the experiments we considered the best working variants (distance's measure and region's size) of each method (H, GJD-BC, SD, and ERCF), according to the results obtained in LFW. To have the same conditions than in the LWF experiments, the faces were aligned using the annotated eyes, and the cropping was done without using funnelling, but using the estimated bounding box that would have been obtained if funnelling was used. This estimation was obtained by measuring the eyes positions of a subset of 20 LWF-funnelled images. As in the case of LWF, the distance between eyes was 42 pixels.

For the evaluation of ERCF, we trained a system using the implementation of the author (available on http://lear.inrialpes.fr/people/nowak/similarity/index.html) of ERCF and the same parameters used to obtain the results presented in [15], which were obtained by a direct communication with the authors of the LFW database. For ERCF we are presenting results for four cases, each one corresponding to a different value of C when training the SVM classifier. The results presented in the previous section for ERCF correspond to $C = 1$. Here we used as training set, the complete test set of LFW (6000 pairs of images).

Table 6 shows the top-1 recognition rates obtained in these tests. Main conclusions are as follows.

(i) Comments on indoor/outdoor tests are as follows.

(a) For all methods, much better results are obtained for indoor faces than for outdoor faces. This is a clear indication that the analyzed methods are not robust to outdoor illumination. Some improvement may be achieved if preprocessing stages are added.

(b) H-X-X methods obtain the highest recognition rate with outdoor faces, followed by GJD and ERCF.

(c) SD performance is strongly affected by outdoor illumination.

(ii) Comments about Scale tests are as follows.

(a) The best performing method is H-X-X, followed by GJD, ERCF, and SD, in that order. However, if we consider only indoor images (S-I set), the best performing methods are H-XS-40 and ERCF, followed by the SD-variants. GJD got the lowest top-1 RR.

(b) In the case of outdoor images all methods have a very low performance, with the best ones (H-HI-40 and H-MSE-40) achieving only a 50% top-1 RR.

(iii) Comments about Expression tests:

(a) HI-X-X shows the best performance followed by ERCF. In the third place comes GJD followed by SD. The same holds if we consider only indoor images (E-I).

(b) In the case of outdoor images, all methods have a very low performance, with the best one (H-XS-40) achieving only a 50.7% top-1 RR.

(iv) Comments about rotation tests are as follows.

(a) Which methods is the best depends on the amount of rotation in the images and on the illuminations conditions. In case of low rotations (15 degrees) with indoor or outdoor illumination, HI-X-X got the highest top-1 RR. In case of higher rotations (30 degrees) and indoor illumination, the same happens. However, in case of 30 degrees rotation and outdoor illumination, ERCF got the top-1 RR.

(b) In indoor conditions, SD is more robust to rotations than GJD. Moreover, SD-Matches and SD-Simple present the second best results in some indoor image cases. However, in outdoor conditions their performance is quite low.

(c) In general terms, the performance of some methods in indoor images with 15 degrees rotation is acceptable (~76%). However, no method gives acceptable results for outdoor images with low rotation (15 degrees), or for rotations in 30 degrees.

(v) Comments about global results are as follows.

(a) Overall, best results are obtained in most of the cases by one of the HI-X-X variants (7 out of 8 subset test, S-I, S-O, E-I, E-O, R-I/15, R-I/30, R-O/15). The second best method is ERCR (being the best in R-O/30 and the second best in most of the cases). If we consider only indoor conditions, GJD and SD got a similar performance, with one of the SD variants (SD-Simple) obtaining slightly better results than GJD. However, if both indoor and outdoor images are considered, the third best method is GJD.

Table 6: UCHFaceHRI tests. Top-1 recognition rate. Experiments are performed with detected eyes. In bold are the best results for each condition. Methods that have differences of 1% or less are considered as having the same performance. See main text for a description about the different experiments.

Method	S-I	S-O	S	E-I	E-O	E	R-I/15	R-I/30	RI	R-O/15	R-O/30	RO	R	G-I	G-O	G
H-HI-40	95.0	**50.0**	**68.0**	**92.5**	48.7	**68.1**	**75.9**	**32.8**	**54.3**	**53.4**	24.1	**38.8**	**46.6**	**78.0**	**45.8**	**60.4**
H-MSE-40	95.0	**50.0**	**68.0**	**92.5**	48.7	**68.1**	**75.9**	**32.8**	**54.3**	**53.4**	24.1	**38.8**	**46.6**	**78.0**	**45.8**	**60.4**
H-XS-40	**98.3**	45.6	66.7	89.2	**50.7**	**67.8**	69.0	27.6	48.3	51.7	24.1	**37.9**	43.1	75.0	**45.2**	58.7
GJD-BC-F	85.0	38.9	57.3	73.3	48.0	59.3	43.1	10.3	26.7	36.2	15.5	25.9	26.3	57.4	38.5	47.1
SD-SIMPLE	91.7	11.1	43.3	61.7	9.3	32.6	56.9	15.5	36.2	5.2	3.4	4.3	20.3	57.8	8.1	30.7
SD-FULL	86.7	6.7	38.7	63.3	8.0	32.6	34.5	6.9	20.7	5.2	5.2	5.2	12.9	51.4	6.7	27.0
SD-MATCHES	88.3	12.2	42.7	51.7	8.0	27.4	46.6	27.6	37.1	10.3	6.9	8.6	22.8	53.4	9.3	29.3
ERCF C = 1e-06	96.7	35.6	60.0	76.7	40.0	56.3	36.2	29.3	32.8	46.6	**27.6**	37.1	34.9	63.5	37.9	49.5
ERCF C = 0.0001	96.7	42.2	64.0	80.0	42.0	58.9	46.6	31.0	38.8	43.1	**27.6**	35.3	37.1	67.2	39.9	52.3
ERCF C = 0.1	**98.3**	24.4	54.0	74.2	30.0	49.6	56.9	20.7	38.8	34.5	15.5	25.0	31.9	65.2	27.0	44.3
ERCF C = 1	**98.3**	24.4	54.0	74.2	30.0	49.6	56.9	20.7	38.8	34.5	15.5	25.0	31.9	65.2	27.0	44.3

6. Comparative Study Using FRGC

From the reported experiments it can be observed that the methods that perform better in our experiments are the LBP-based (H-X-X) and Gabor-based (GJD) ones. These methods are further analyzed using the FRGC ver2.0 database [17]. This database consists of 50,000 face images divided into training and validation partitions. In our experiments the training partition was not used, because one of our main requirements is that methods under comparison should be fully on-line. The validation partition consists of data from 4,003 subject sessions. A subject session consists of controlled and uncontrolled images. The controlled images were taken in a studio setting, and they are full frontal facial images taken under two lighting conditions and with two facial expressions (smiling and neutral), while the uncontrolled images were taken in varying illumination conditions [17]. Each set of uncontrolled images contains two expressions, smiling and neutral. In our analysis we will focus on two FRGC tests: *Experiment 1*, which corresponds to a control experiment where the gallery and the probe sets consist of controlled still images, and *Experiment 4*, which measures recognition performance from uncontrolled images (the probe set consist of single uncontrolled still images; the gallery is composed by controlled still images).

Figure 12 shows the ROC curve obtained in experiment 1 by the best methods under comparison. It should be stressed that in our test we have used all possible image pair comparisons that can be carried out in experiment 1 ($16,028 \times 16,028$), and not the image pairs defined by the ROC I–ROC III FRGC subexperiments that some papers report. As it can be observed the obtained results are concordant with the ones of similar reported approaches , for instance in [31, 32]. But, if we compare these methods with recent kernel-based approaches, as the ones proposed by Liu [33] (Gabor-Multiclass-KFDA) or Zhao et al. (LBP KFDA) [31], we observe that kernel approaches obtain much higher results than LBP- or Gabor-based approaches, about 10% higher verification rate for a given FAR. However, it should be remembered that the kernel approaches need to be trained in the database, and they are much slower than the methods under comparison. From Figure 12 it is also interesting to note the dependency of the LBP-based methods' performance on the number of partitions. Methods using a larger number of partitions get better results than methods using a smaller number of partitions. This phenomenon although being logic was not clearly observed in the other databases. Probably with very large database the number of partitions is an important parameter to be considered.

We also analyzed the methods under comparison using the FRGC, experiment 4. By analyzing the results, similar conclusions were obtained: (i) the results are concordant with the ones of similar approaches reported in the literature (see, e.g., [26]), (ii) kernel approaches get much better results, and (iii) the performance of the LBP-based methods depends on the number of partitions.

7. Discussion and Conclusions

In this article, a comparative study among face-recognition methods in unconstrained environments was presented. The analyzed methods were selected by considering their suitability for the defined requirements—real-time operation, just one image per person, fully on-line (no training), robust behavior in unconstrained environments, and their performance in former studies. The comparative study was carried out using three databases: FERET, LFW, and UCHFaceHRI. The well-known FERET database was used as a baseline for comparison, and experiments were carried out in different subsets that include variations in illumination, nonaccurate eye's annotations, and occlusions. The LFW database implicitly includes aspects such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background, and photographic quality, while the UCHFaceHRI explicitly includes aspects such as scale (distance to the camera), expressions (neutral, surprised, angry, sad, and happy), pose (0, ± 15 , and ± 30 degrees of

Figure 11: UCHFaceHRI database. Examples of 24 indoor images corresponding to an individual. The face-image F*jki* corresponds to an image taken acquisition point *j* (see Figure 10). *i* stands for indoor. In the case of the F12*k*i images, the *k* index means: (a) Neutral expression, (b) Surprised, (c) Angry, (d) Sad, and (e) Happy.

out-of-plane rotation), and illumination (indoor/outdoor). The methods under comparison are generalized-PCA, LBP histograms, Gabor Jets descriptors, SIFT descriptors, and ERCF. We will comment about the main results of this study, and we will draw some conclusions of this work.

Comments on the Size of the Face Region. What was very surprising to us is the large dependence of the methods to the amount of face and background information that is included in the face's images. This effect was clearly seen in our FERET and LFW experiments. For instance, in the FERET case, SD increases its recognition rate in more than 20% depending on the size of the face images. In the LWF case where experimental conditions are much harder, LBP-based methods and SD increase their recognition rates in ~4%, depending on the size of the face images. We also observe that the different methods have different requirements. LBP-based methods concentrate themselves mostly in the face area, but it seems that additional information about the face's chin, which is only observed if some background is included in the images, helps the recognition process. On the other hand, GJD methods need much more background. The reason is twofolds: (i) the Gabor filters encode information about the contour of the face, and (ii) large regions allow the use of large filters, which encoded large-scale information. SD methods show better performance when there is the fewest possible background in the image, but without removing any part of the face. The methods needs as much as possible face information to obtain a correct matching, but the background disturbs this process (a face keypoint could be matched to a background keypoint).

Comments on the Illumination Conditions. Most of the methods behave very well in natural, indoor illumination conditions, the exception being SD. This can be clearly seen in the FERET experiments (*fa-fc*). However, this situation changes drastically with outdoor illumination conditions. The performance of all methods decreases largely with outdoor illumination. Clearly, face recognition in outdoor conditions is still a nonsolved problem.

Comments on Pose Variations. Invariance against pose variations is a second main problem in face recognition. In

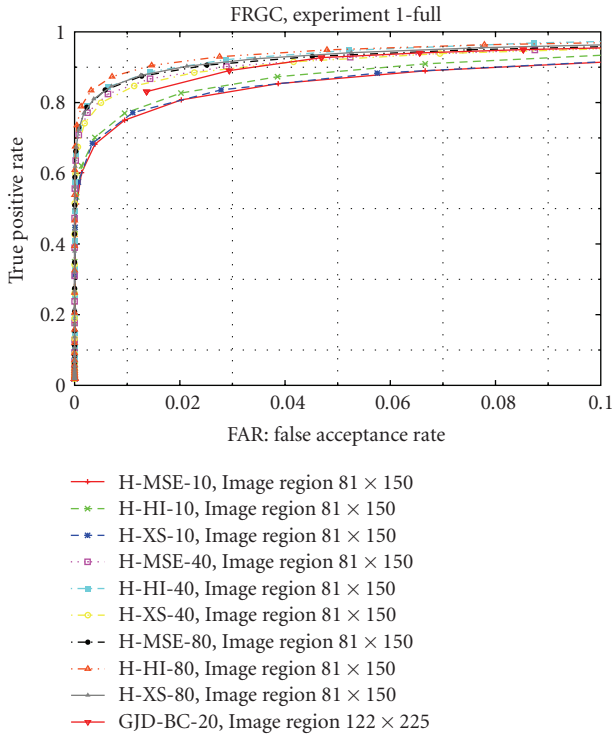


Figure 12: ROC curves of the best methods under comparison in FRGC, experiment 1.

the UCHFaceHRI experiments it can be observed that yaw rotations in 15 degrees affect largely the performance of all methods; the recognition rates decrease in more than 20%. In the 30-degrees case the situation is even worse, the recognition rates fall in more than 60%. In relation, we also believe that the main reason for the low results that are obtained in the LFW database is due to the variations in the faces' pose.

Comments on Alignment, Occlusions, and Expressions. From our experiments we conclude that the analyzed methods are robust to inaccurate alignment, face occlusions, and variations in expressions, to a large degree. Accepting that these factors affect the face-recognition process, their influence in the algorithms' performance is much lower than outdoor illumination or pose's variations.

Conclusions about the Performance of Methods. The question of which method is the best is a very difficult one. However, we could say that LBP-based methods are an excellent election if we need real-time operation as well as high recognition rates. In the UCHFaceHRI experiments some of the LBP variants got the best results, while in the LWF case they got the second best results.

Gabor-based methods are also an adequate election. Although they got a lower performance in UCHFaceHRI than LBP-based methods, they got a similar performance in LFW, and slightly better results in FERET. However, Gabor-based methods are slower than LBP ones. Probably some work can be done to develop strategies that select which filters to use (some research in this direction has been reported in [10]). A last interesting aspect to be mentioned is that the proposed strategy of using a reference set of images in the case of comparing pairs of images was successful and better than using the Euclidian distance.

ERCF is a novel and promising matching method. However, it has some drawbacks, the first one being its low processing speed, which does not allow its application in real-time conditions. Moreover, the method has several parameters, and it seems that its performance depends on the correct selection of them. Thus, although the method achieves the best results in the LFW database, being clearly superior to the others, it got the second place in the UCHFaceHRI experiments. In these experiments LBP-based methods work better than ERCF, in particular in difficult cases such as outdoor images, out-of-plane rotation, and facial expressions. This may be due to the fact that the learning done by ERCF does not generalize as the results reported for LFW seem to indicate. This may be due to the fact that the images from LFW were obtained from news images, which in general are taken by professional photographers, and therefore are obtained under good illumination, and because they are also taken in indoor conditions, which are the cases where ERCF works best.

SD methods performed very well in some of our experiments, achieving similar recognition rates than LBP-based and Gabor-based methods. However, SD methods have a large dependence to illuminations conditions. This is especially true for the case of outdoor illumination, were the methods' performance decrease largely. It is interesting to note that the large dependence of SD methods to illumination conditions is not clearly reported in the SIFT-related literature.

The generalized PCA method got the worse results in the FERET experiments and was not further analyzed in this study. We believe that under the main requirements of this study (real-time operation, just one image per person, and no training stages), eigenspace-based holistic methods are not competitive against the other methods.

When the best methods under analysis are compared against novel kernel-based approaches [31, 33] (e.g., in the FRGC database), they obtain a lower performance. However, it should be noted that kernel-based methods are intended to be used in other kinds of applications, which do not have the requirements of real-time and full on-line operation.

Future Work. We believe that still there are many aspects that can be improved in the recognition of faces in unconstrained environments. However, in the medium term, we will concentrate on: (i) the analysis of pre-processing algorithms and other strategies to achieve invariance against outdoor illumination conditions, (ii) the combined use of methods (e.g., ERCF and LBP-based or kernel-based and LBP) that can allow achieving, at the same time, high recognition rates and processing speed, (iii) the study of the influence of face's resolution in the recognition process, and (iv) a more deep analysis of the facial expression effect in the recognition of faces.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: a survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[3] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[4] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: a survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.

[7] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 880–893, 2008.

[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, Mass, USA, October 2007.

[9] J. Ruiz-del-Solar and P. Navarrete, "Eigenspace-based face recognition: a comparative study of different approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 35, no. 3, pp. 315–325, 2005.

[10] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2617–2628, 2007.

[11] J. Ruiz-del-Solar and J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: a comparative study of different pre-processing approaches," *Pattern Recognition Letters*, vol. 29, no. 14, pp. 1966–1979, 2008.

[12] M. Correa, J. Ruiz-del-Solar, and F. Bernuy, "Face recognition for human-robot interaction applications: a comparative study," in *Proceedings of the RoboCup International Symposium*, Lecture Notes in Computer Science, Suzhou, China, July 2008.

[13] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1632–1646, 2008.

[15] Labeled Faces in the Wild database, "Results," http://vis-www.cs.umass.edu/lfw/results.html.

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, et al., "Overview of the face recognition grand challenge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 947–954, San Diego, Calif, USA, June 2005.

[17] Face Recognition Grand Challenge, http://www.frvt.org/FRGC.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 511–518, Kauai, Hawaii, USA, December 2001.

[19] R. Verschae, J. Ruiz-del-Solar, and M. Correa, "Face recognition in unconstrained environments: a comparative study," in *Proceedings of the Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition (ECCV '08)*, pp. 1–12, Marseille, France, October 2008, CD Proceedings.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] J. Ruiz-del-Solar, P. Loncomilla, and Ch. Devia, "A new approach for fingerprint verification based on wide baseline matching using local interest points and descriptors," in *Proceedings of the 2nd IEEE Pacific Rim Symposium on Image and Video Tecnology (PSIVT '07)*, vol. 4872 of *Lecture Notes in Computer Science*, pp. 586–599, Santiago, Chile, December 2007.

[22] J. Ruiz-Del-Solar, Ch. Devia, P. Loncomilla, and F. Concha, "Offline signature verification using local interest points and descriptors," in *Proceedings of the 13th Iberoamerican Congress on Pattern Recognition (CIARP '08)*, vol. 5197 of *Lecture Notes in Computer Science*, pp. 22–29, Havana, Cuba, September 2008.

[23] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of SIFT features for face authentication," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '06)*, p. 35, New York, NY, USA, June 2006.

[24] B. Fröba and A. Ernst, "Face detection with the modified census transform," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 91–96, Seoul, Korea, May 2004.

[25] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[26] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," in *Proceedings of the 3rd International Workshop on Analysis and Modeling of Faces and Gestures (AMFG '07)*, vol. 4778 of *Lecture Notes in Computer Science*, pp. 235–249, Rio de Janeiro, Brazil, October 2007.

[27] Y. Su, S. Shan, X. Chen, and W. Gao, "Patch-based Gabor fisher classifier for face recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 2, pp. 528–531, Hong Kong, August 2006.

[28] S. Shan, W. Zhang, Y. Su, X. Chen, and W. Gao, "Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 606–609, Hong Kong, August 2006.

[29] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV '07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[30] UCHFaceHRI database, http://vision.die.uchile.cl/2/Databases.htm.

[31] J. Zhao, H. Wang, H. Ren, and S. C. Kee, "LBP discriminant analysis for face verification," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 3, pp. 167–172, San Diego, Calif, USA, June 2005.

[32] H. Yang and Y. Wang, "A LBP-based face recognition method with hamming distance constraint," in *Proceedings of the 4th*

International Conference on Image and Graphics (ICIG '07), pp. 645–649, Chengdu, China, August 2007.

[33] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 261542, 17 pages
doi:10.1155/2009/261542

Research Article

Facial Expression Biometrics Using Statistical Shape Models

Wei Quan, Bogdan J. Matuszewski (EURASIP Member), Lik-Kwan Shark, and Djamel Ait-Boudaoud

Applied Digital Signal and Image Processing Research Centre, University of Central Lancashire, Preston PR1 2HE, UK

Correspondence should be addressed to Bogdan J. Matuszewski, bmatuszewski1@uclan.ac.uk

Received 30 September 2008; Revised 2 April 2009; Accepted 18 August 2009

Recommended by Jonathon Phillips

This paper describes a novel method for representing different facial expressions based on the shape space vector (SSV) of the statistical shape model (SSM) built from 3D facial data. The method relies only on the 3D shape, with texture information not being used in any part of the algorithm, that makes it inherently invariant to changes in the background, illumination, and to some extent viewing angle variations. To evaluate the proposed method, two comprehensive 3D facial data sets have been used for the testing. The experimental results show that the SSV not only controls the shape variations but also captures the expressive characteristic of the faces and can be used as a significant feature for facial expression recognition. Finally the paper suggests improvements of the SSV discriminatory characteristics by using 3D facial sequences rather than 3D stills.

1. Introduction

Facial expressions provide important information in communication between people and can be used to enable communication with computers in a more natural way. Recent advances in imaging technology and ever increasing computing power have opened up a possibility of automatic facial expression recognition. Up till now some research efforts have been exploited in applications such as human-computer interaction (HCI) systems [1], video conferencing [2], and augmented reality [3]. From the biometric perspective, the automatic expression recognition has been investigated in the context of patients' monitoring in the intensive care and neonatal units [4] for signs of pain and anxiety, behavioural research on children's ability to learn emotions by interacting with adults in different social contexts [5], identifying level of concentration [6], that is, for detecting drivers' tiredness, and finally in aiding face recognition. Facial expression representation, which forms one of the most important elements in the facial expression recognition system, is concerned with extraction of facial features for representing variations of expressions. Good features for representing the facial expressions should enable interpretation of various face articulations without any limitation of race, gender, and age. Furthermore, it should also have the capability of reducing the complexity of classification algorithms.

Generally, facial expressions can be represented in two forms, namely, holistic representation and local representation [7]. For the holistic representation, the face is processed as a single entity. Wang and Yin [8] introduced a holistic representation method for representing facial expressions, which is named the topographic context (TC). In this method a grey-scale facial image is treated as a topographic terrain surface in a 3D space with the height of the terrain represented by the image intensity at each pixel. As the result of the topographic analysis, each pixel of the image is described by one of the topographic labels: peak, ridge, saddle, hill, flat, ravine, and pit. The topographic context has been also extended for 3D facial surfaces by Wang et al. [9], where it is referred to as the primitive surface feature method. Huang et al. [10] proposed a method for expression representation based on the local binary pattern, which is originally designed for the texture description. The local binary pattern is calculated by encoding the information of depth difference of a 3D facial surface. Active appearance model (AAM) is a statistical model of shape and grey level of object of interest and mainly used for 2D facial images. For the facial expression representation, the AAM is built on the facial images which are manually selected with a set

of landmarks localised around the facial features such as eyebrows, eyes, mouth, and nose [11]. As an extension of the AAM, the 3D morphable model was developed by Blanz and Vetter [12]. Instead of using manually selected sparse facial landmarks, the 3D morphable model uses all the data points of 3D facial scans to represent the geometrical information. This model has been used to control 3D facial surfaces from a 2D image, across variations in pose, ranging from frontal to profile view, and a wide range of illuminations. B-spline is a parametric model which is often used to describe surfaces. When used with 3D facial data, a large number of data points can be efficiently modelled by a small number of B-spline's control points [13]. When combined with the facial action coding system (FACS) [14], the control points are placed in areas that correspond to action units, and the expression of a face can be generated automatically by adjusting the B-spline's control points.

In contrast to the holistic approaches, the local representation methods focus on the local features or areas that are prone to change with facial expressions. Saxena et al. [15] introduced the localised geometric model to locally represent facial expressions. Their method uses the classical edge detectors with colour analysis for extracting the local appearances of a face such as eyebrows, lips, and nose. Subsequently a feature vector containing measurements of the facial appearances, such as the height of eyebrows, brow distance, mouth height, mouth width, and lip curvature, is created for the facial expression classification. A local parameterised model proposed by Black and Yacoob [16] is developed based on image motion which is calculated using the optical flow of facial image sequences. The image motion not only accurately models a nonrigid facial motion but also provides a concise description that is related to the motion of local facial features to recognise facial expressions. Kobayashi et al. [17] used a point-based geometric model for the facial expression representation. The model contains 30 facial characteristic points in the frontal-view of the face. These facial characteristic points are around the areas that are the most affected by change of facial expressions, such as eyes, nose, brows, and mouth.

In this paper, a novel method for representing facial expressions is proposed based on the authors' previous work [18–20], which postulates that the shape space vectors constitute a significant feature space for the recognition of facial expressions. The proposed method uses only 3D shape information, with the texture not being used at all. The method is therefore inherently invariant to variations in scene illumination conditions, background clutter, and to some extent angle of view. This is in a striking contrast to the methods based on texture where these factors severely limit their practical applicability. Additionally as the texture is not being used, it does not have to be captured; hence fast full frame 3D acquisition techniques based on the time-of-fly principle [21] can be used (3D scanners capturing in excess of 40 frames/sec are commercially available) instead of more computationally intensive, and therefore slower, stereovision scanning systems. The shape space vector (SSV) is the key element in the statistical shape model (SSM), which models the high-dimensional shape variations observed in the training data set using projections on a low-dimensional shape space. In order to obtain the SSV two consecutive stages are necessary, namely, (i) model building stage and (ii) model fitting stage. In the model building stage, the correspondences of points between all faces present in the training data set are established first so that the training data set can be aligned into a common reference face. Subsequently the principal component analysis (PCA) technique is applied to the aligned training data set to obtain the SSM of the shape variations. In the model fitting stage, an iterative algorithm based on a modified iterative closest point (ICP) method is used to gradually adjust the pose parameters and optimise the shape parameters in order to match the model to the newly observed facial data. The pose parameters consist of a translation vector, a rotation matrix, and a scaling factor, whereas the shape parameters are embedded in the SSV. In order to validate the discriminatory ability of the SSV, 3D synthetic faces generated from the FaceGen Modeller [22] and real 3D facial scans from the BU-3DFE database [23] are used for the separability analysis in the SSV domain. The experiments on recognition of facial expressions using a selection of standard classification tools are also presented.

The remainder of this paper is organised as follows. Section 2 introduces the details of construction of the SSM. Section 3 describes the procedure used for fitting the model to the facial data that has not been included in the training data set. Section 4 provides results of qualitative and quantitative separability analysis. Results of facial expression recognition using some popular classification algorithms operating on the SSV feature space are presented in Section 5. Finally, concluding remarks are given in Section 6, and a potential improvement of the expression representation using the SSV constructed for dynamic 3D data is briefly discussed in Section 7.

2. Statistical Shape Model

The statistical shape model (SSM) is developed based on the point distribution model (PDM) which was proposed by Cootes et al. [24], and it is one of the most widely used techniques for the model-based data representation and registration. The model describes shape variations based on the statistic calculated from the position of the corresponding points in the training data set. In order to build an SSM, the correspondence of points between different 3D faces in the training data set must be established first. Subsequently the principal component analysis (PCA) is applied to the mutually aligned training data set.

2.1. Estimating Point Correspondence. The knowledge of the correspondence of points between 3D faces in the training data set is essential, because the incorrect correspondence can either introduce too much variations or lead to illegal instance of the model [24]. In the case of the data used in this paper the correspondence of points for the database generated using the FaceGen Modeller

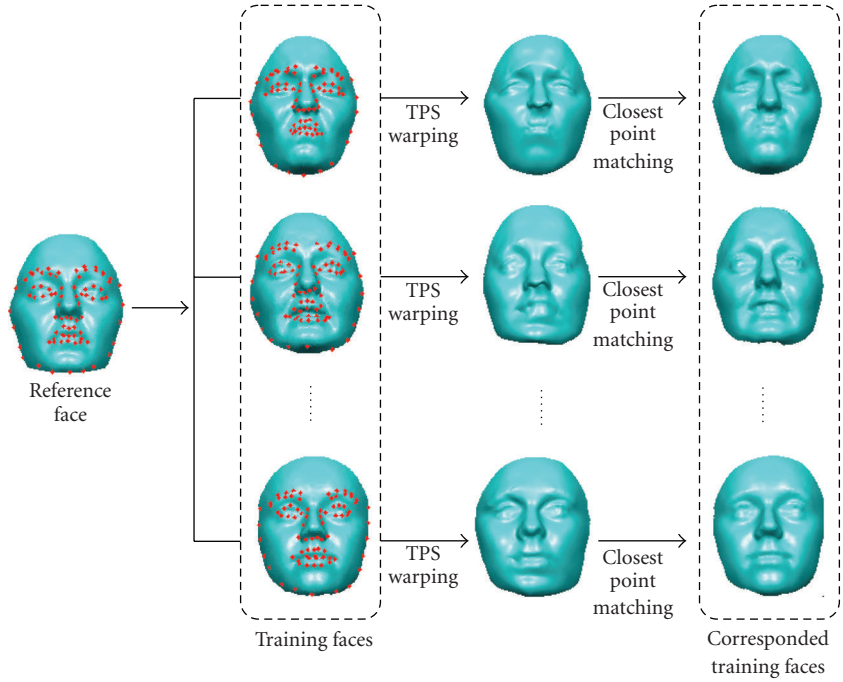


Figure 1: Example of point correspondence estimation in the training data set, with example images from the BU-3DFE.

is explicitly provided by the software, whereas the dense correspondence of points for the faces in the BU-3DFE database is estimated based on a set of facial landmarks included in the database.

In this work, the estimation of the correspondence is achieved in three steps: (i) facial landmark determination, (ii) thin-plate spline (TPS) warping, and (iii) closest point matching. The first step is to identify the corresponding facial landmarks on the reference and training faces. The second step is to warp the reference face to different training face using TPS transformation that is calculated based on the selected facial landmarks as control points [25]. The last step is to estimate the point correspondence between the warped reference face and different training faces based on the closest distance metric. Figure 1 shows the framework of computing the dense point correspondence of different training faces from the BU-3DFE database. The reference face is usually selected as a face containing neutral expression with the mouth closed. Such selection of the reference face helps to avoid wrong correspondences in the case of matching between closed-mouth and open-mouth shapes. If the reference face were selected with the mouth open, after dense correspondence estimation, each point in the open-mouth area of the reference face will find an incorrect corresponding point in the training face within the closed-mouth region even though those corresponding points of the open-mouth area do not exist in the training faces with mouth closed.

2.1.1. Thin-Plate-Spline Warping. The TPS warping technique is a point-based registration method which was first proposed by Bookstein [26]. The TPS warping can be used for interpolation as well as approximation. For the TPS interpolation, the positions of corresponding landmarks are assumed to be known exactly and the corresponding landmarks are forced to match exactly each other after warping [25, 27]. For the TPS approximation, the landmark position errors are taken into account, implying that the corresponding landmarks are not forced to match exactly after warping is applied. It can be shown that the solution of the approximation problem is equivalent to inclusion of a regularisation term in the cost function along with a fidelity term which is exactly the same as used in the definition of the interpolation problem [28]. In this work, the corresponding facial landmarks are manually labeled on the 3D face scans, and their positions are always prone to some errors. Therefore, the TPS approximation model is more suitable for our application.

Given sparse corresponding facial landmarks in the reference face and one of the training faces, represented, respectively, by $\tilde{\mathbf{P}} = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \ldots, \tilde{\mathbf{p}}_L)^T$ and $\tilde{\mathbf{Q}} = (\tilde{\mathbf{q}}_1, \tilde{\mathbf{q}}_2, \ldots, \tilde{\mathbf{q}}_L)^T$, where $\tilde{\mathbf{p}}_k = (\tilde{x}_{pk}, \tilde{y}_{pk}, \tilde{z}_{pk})^T$ and $\tilde{\mathbf{q}}_k = (\tilde{x}_{qk}, \tilde{y}_{qk}, \tilde{z}_{qk})^T$ denote x, y, and z coordinates of the kth corresponding pair and L is the total number of corresponding facial landmarks, the objective is to find the TPS warping function that warps the reference face to the training face. The interpolating warping function, F, has to fulfill the following constraint for all the landmarks in $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$:

$$F(\tilde{\mathbf{p}}_i) = \tilde{\mathbf{q}}_i, \quad i = 1, 2, \ldots, L, \tag{1}$$

where the deformation model is defined in terms of warping function $F(\mathbf{p}_j)$ with

$$F\left(\mathbf{p}_j\right) = \begin{bmatrix} f_x\left(\mathbf{p}_j\right) \\ f_y\left(\mathbf{p}_j\right) \\ f_z\left(\mathbf{p}_j\right) \end{bmatrix}, \tag{2}$$

where $\mathbf{p}_j = (x_{pj}, y_{pj}, z_{pj})^T$ is a point on the reference face and the warping functions for x, y, and z coordinates are defined as follows

$$f_x\left(\mathbf{p}_j\right) = a + a_x x_{pj} + a_y y_{pj} + a_z z_{pj} + \sum_{i=1}^{L} w_{xi} U\left(\left\|\widetilde{\mathbf{p}}_i - \mathbf{p}_j\right\|\right), \tag{3}$$

$$f_y\left(\mathbf{p}_j\right) = b + b_x x_{pj} + b_y y_{pj} + b_z z_{pj} + \sum_{i=1}^{L} w_{yi} U\left(\left\|\widetilde{\mathbf{p}}_i - \mathbf{p}_j\right\|\right), \tag{4}$$

$$f_z\left(\mathbf{p}_j\right) = c + c_x x_{pj} + c_y y_{pj} + c_z z_{pj} + \sum_{i=1}^{L} w_{zi} U\left(\left\|\widetilde{\mathbf{p}}_i - \mathbf{p}_j\right\|\right). \tag{5}$$

Function U is a radial basis function of the form

$$U(r) = r^2 \log r^2, \tag{6}$$

where r is a distance between two points. According to Bookstein [26], the coefficients of the TPS interpolation model can be calculated from

$$\mathbf{K}_c \mathbf{W}_c + \mathbf{P}_c \mathbf{A}_c = \widetilde{\mathbf{Q}}, \tag{7}$$

and

$$\mathbf{P}_c^T \mathbf{W}_c = \mathbf{0}, \tag{8}$$

where $\widetilde{\mathbf{Q}}$ is a $L \times 3$ matrix which contains facial landmarks on the target face and written as

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \widetilde{x}_{q1} & \widetilde{y}_{q1} & \widetilde{z}_{q1} \\ \widetilde{x}_{q2} & \widetilde{y}_{q2} & \widetilde{z}_{q2} \\ \cdots & \cdots & \cdots \\ \widetilde{x}_{qL} & \widetilde{y}_{qL} & \widetilde{z}_{qL} \end{bmatrix}. \tag{9}$$

$\mathbf{W}_c$ and $\mathbf{A}_c$ are the matrices containing coefficients of the TPS interpolation and defined as

$$\mathbf{W}_c = \begin{bmatrix} w_{x1} & w_{y1} & w_{z1} \\ w_{x2} & w_{y2} & w_{z2} \\ \cdots & \cdots & \cdots \\ w_{xL} & w_{yL} & w_{zL} \end{bmatrix}, \qquad \mathbf{A}_c = \begin{bmatrix} a & b & c \\ a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{bmatrix}, \tag{10}$$

whereas matrix $\mathbf{K}_c$ that contains the radial basis functions is defined as

$$\mathbf{K}_c = \begin{bmatrix} U(r_{11}) & U(r_{12}) & \cdots & U(r_{1L}) \\ U(r_{21}) & U(r_{22}) & \cdots & U(r_{2L}) \\ \cdots & \cdots & \cdots & \cdots \\ U(r_{L1}) & U(r_{L2}) & \cdots & U(r_{LL}) \end{bmatrix}, \tag{11}$$

and the radial basis function $U(r_{ij})$ is

$$U\left(r_{ij}\right) = \left\|\widetilde{\mathbf{p}}_i - \widetilde{\mathbf{q}}_j\right\|^2 \log\left(\left\|\widetilde{\mathbf{p}}_i - \widetilde{\mathbf{q}}_j\right\|^2\right). \tag{12}$$

$\mathbf{P}_c$ is the matrix including all corresponding landmarks of the reference face and defined as

$$\mathbf{P}_c = \left[\mathbf{1}\widetilde{\mathbf{P}}\right], \tag{13}$$

and matrix $\widetilde{\mathbf{P}}$ is defined as

$$\widetilde{\mathbf{P}} = \begin{bmatrix} \widetilde{x}_{p1} & \widetilde{y}_{p1} & \widetilde{z}_{p1} \\ \widetilde{x}_{p2} & \widetilde{y}_{p2} & \widetilde{z}_{p2} \\ \cdots & \cdots & \cdots \\ \widetilde{x}_{pL} & \widetilde{y}_{pL} & \widetilde{z}_{pL} \end{bmatrix}. \tag{14}$$

In the TPS approximation model, the interpolation condition has to be weakened since the landmark localisation errors have to be taken into account. The regularisation term needs to be added into the TPS interpolation model in order to control smoothness of the transformation. The coefficients of the TPS approximation model can be calculated as

$$(\mathbf{K}_c + \lambda_c \mathbf{I})\mathbf{W}_c + \mathbf{P}_c \mathbf{A}_c = \widetilde{\mathbf{Q}}, \tag{15}$$

where $\lambda_c > 0$ is a relative weighting factor between the interpolating behavior and the smoothness of the transformation. For small λ_c, the TPS warping maintains a good approximation of the landmarks. For large λ_c, the TPS warping function becomes very smooth and adopts very little to the local structures present in the data.

2.1.2. Closest Point Matching. After the TPS approximation, the shape of the reference face is warped to match the training face. Since the shape of the reference face is close to the shape of the training face, the dense point correspondence of the reference face for the training face can be computed using the closest distance metric. With the Euclidean distance $d(\mathbf{p}, \mathbf{q})$ between two points $p = (x_p, y_p, z_p)^T$ and $q = (x_q, y_q, z_q)^T$ are defined as

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\left(x_p - x_q\right)^2 + \left(y_p - y_q\right)^2 + \left(z_p - z_q\right)^2}. \tag{16}$$

Denoting a set of points of the training face by $\{\mathbf{q}_i, i \in [1, N]\}$, the closest distance between a point $\mathbf{p} = (x_p, y_p, z_p)^T$ of the reference face and the training face is defined as

$$d(\mathbf{p}, \{\mathbf{q}_i, i \in [1, N]\}) = \arg\min_i (d(\mathbf{p}, \mathbf{q}_i)). \tag{17}$$

Using the TPS approximation and closest point matching, the dense point correspondence between the reference face and a training face can be established. This process is applied to all the training faces such that all of them are in correspondence. The training faces from the BU-3DFE database contain between 13 000 and 20 000 mesh polygons with 8711 to 9325 vertices. The reference face used in this paper has 15 687 mesh polygons and 8925 vertices. After performing the TPS approximation and closest point matching, it is likely that there will be multi-to-one correspondences between a training face and the reference face. It is impossible to avoid this completely due to the nature of the closest point matching technique. In order to reduce the number of such correspondences, a subdivision surface method has been used to increase the number of vertices in the training faces [29].

2.2. Principal Component Analysis. Using the standard principal component analysis (PCA), each 3D face in the training data set can be approximately represented in a low-dimensional shape vector space [30] instead of the original high-dimensional data vector space. Given a training data set of M faces, $\mathbf{Q}_i (i = 1, 2, \ldots, M)$, each containing N corresponding data points $\mathbf{Q}_i \in R^{3N}$, where $\mathbf{Q}_i$ contains all the data points of the ith face encoded as a 3N-dimensional vector. The first step of the PCA is to calculate the mean vector $\overline{\mathbf{Q}}$ (representing the mean 3D face):

$$\overline{\mathbf{Q}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{Q}_i. \tag{18}$$

Let $\mathbf{C}$ be defined as the covariance matrix calculated from the training data set:

$$\mathbf{C} = \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{Q}_i - \overline{\mathbf{Q}}\right)\left(\mathbf{Q}_i - \overline{\mathbf{Q}}\right)^T. \tag{19}$$

By building a matrix $\mathbf{X}$ of "centered" data vectors with $\mathbf{Q}_i - \overline{\mathbf{Q}}$ as the ith column of matrix $\mathbf{X}$, covariance matrix $\mathbf{C}$ can be calculated as

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T, \tag{20}$$

where matrix $\mathbf{C}$ has $3N$ rows and columns. Since the number of faces, M, in the training data set is smaller than the number of data points, the eigen decomposition of matrix $\mathbf{C}' = \mathbf{X}^T\mathbf{X}$ is performed first [31]. The first M largest eigenvalues $\lambda_i (i = 1, \ldots, M)$ and eigenvectors $\mathbf{u}_i (i = 1, \ldots, M)$ of the original covariance matrix, $\mathbf{C}$, are then determined, respectively, from

$$\lambda_i = \lambda_i', \tag{21}$$

$$\mathbf{u}_i = \frac{\mathbf{X}\mathbf{u}_i'}{||\mathbf{X}\mathbf{u}_i'||}, \tag{22}$$

where λ_i' and $\mathbf{u}_i'$ are eigenvalues and eigenvectors of matrix $\mathbf{C}'$, respectively. By using these eigenvalues and eigenvectors, the data points on any 3D face in the training data set can be approximately represented using a linear model of the form

$$\hat{\mathbf{Q}} = \mathbf{W}\mathbf{b} + \overline{\mathbf{Q}}, \tag{23}$$

where $\mathbf{W} = [\mathbf{u}_1, \ldots, \mathbf{u}_i, \ldots, \mathbf{u}_K]$ is a $3N \times K$ so-called "Shape Matrix" of K eigenvectors, or "modes of variation", which correspond to the K largest eigenvalues, and $\mathbf{b} = [b_1, \ldots, b_i, \ldots, b_K]$ is the shape space vector (SSV), which controls contribution of each eigenvector, $\mathbf{u}_i$, in the approximated surface $\hat{\mathbf{Q}}$ [12]. The shape matrix $\mathbf{W}$ is database-dependent. In a case when new faces are added to the existing database, this shape matrix needs to be recalculated. Most of the surface variations can usually be modelled by a small number of modes K. Equation (23) can be used to generate new examples of faces by changing the SSV, $\mathbf{b}$, with suitable limits [24]. According to the work proposed by Edwards et al. [11], the suitable limits of the SSM are typically defined as

$$-3\sqrt{\lambda_i} \le b_i \le 3\sqrt{\lambda_i}. \tag{24}$$

Figure 2 shows the effect of varying the first three largest principal components of the two models. These models were built using 450 training faces from the FaceGen and BU-3DFE database, respectively.

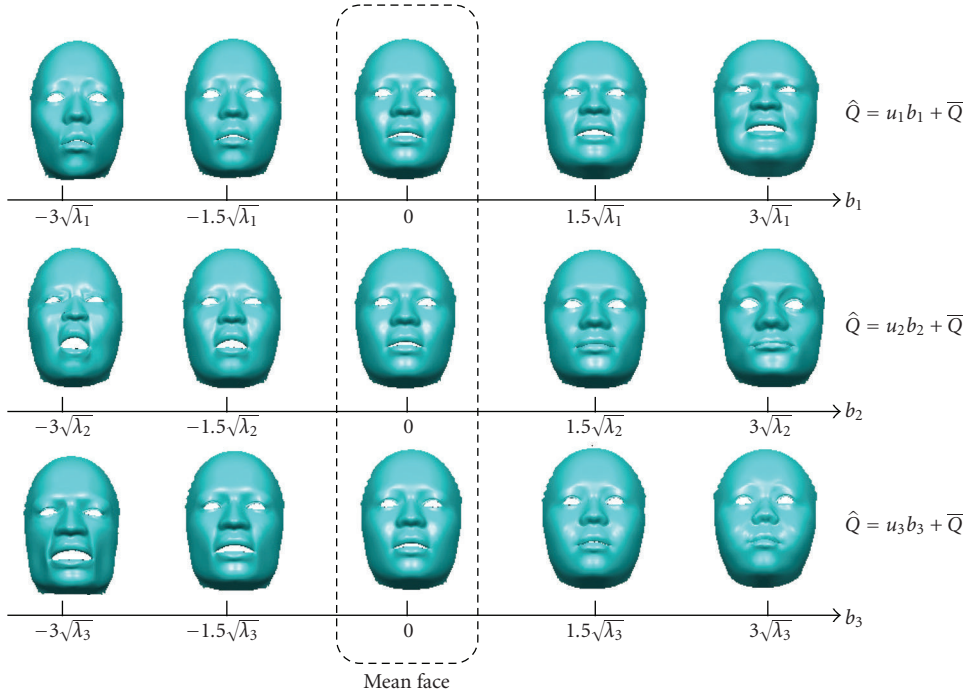
3. Model Fitting

Provided that the faces in the database are representative of the faces in the population, a new face from the same population, which has not been included in the training data, can be represented using the derived SSM. In the proposed method, the model fitting is treated as a surface registration problem, which includes the estimation of the pose parameters and shape parameters of the model. Whilst the pose parameters include a translation vector, a rotation matrix, and a scaling factor, the shape parameters are defined by the SSV. As described in the following subsection, the algorithm starts by aligning a new face with the mean face of the model using similarity transformation. Subsequently the model continues to be refined by iteratively estimating the SSV and pose parameters.

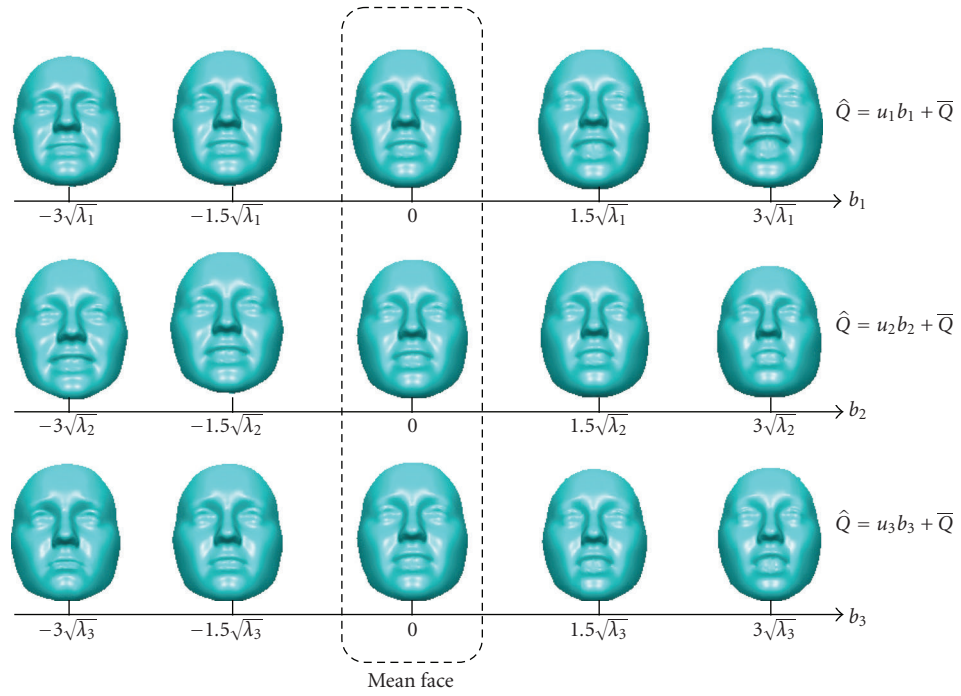
3.1. Similarity Registration. The iterative closest point (ICP) method can be used to achieve similarity registration between the model mean face and a new face. The ICP [32] is a widely used point-based surface matching algorithm. This procedure iteratively refines the alignment by alternately estimating points correspondence and finding the best similarity transformation that minimises a cost function between the corresponding points. In this work the cost function is defined using Euclidean distance:

$$E = \sum_{i=1}^{N}||\mathbf{q}_i' - (s\mathbf{R}\mathbf{q}_i + \mathbf{t})||^2, \tag{25}$$

where $\mathbf{q}_i'$ and $\mathbf{q}_i (i = 1, \ldots, N)$ are, respectively, the corresponding vertices from the model and the data face. $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{t}$ is a 3×1 translation vector, and s is a scaling factor. Following the algorithms in [33, 34], $\mathbf{R}$, $\mathbf{t}$, and s are calculated as follows.

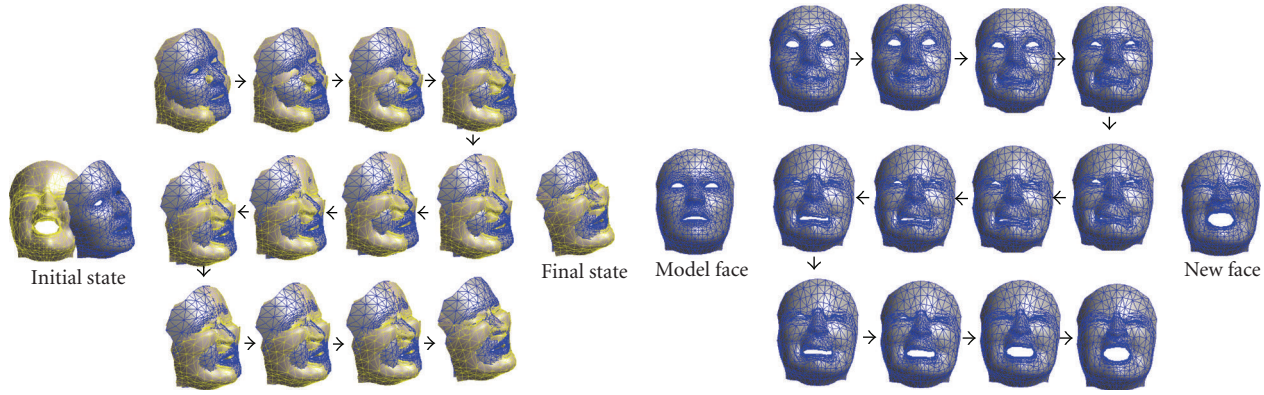


(a) From top to bottom: superposition of the mean face and weighted first three principal components calculated from the FaceGen synthetic database. In each case the principal component weights vary between $\pm 3\sqrt{\lambda_i}$



(b) From top to bottom: superposition of the mean face and weighted first three principal components calculated from the BU-3DFE database. In each case the principal component weights vary between $\pm 3\sqrt{\lambda_i}$

Figure 2: Effects of changing the contribution of the first three principal components of the shape space vector on the models derived from the FaceGen and BU-3DFE data sets.



(a) Example of intermediate results obtained during iterations of the similarity registration

(b) Example of the model deformations during refinement iterations

FIGURE 3: An example of the model fitting.

(1) From the point sets, $\{\mathbf{q}_i\}$ and $\{\mathbf{q}'_i\}(i = 1,\ldots,N)$, compute the mean vectors, $\overline{\mathbf{q}}$ and $\overline{\mathbf{q}}'$:

$$\overline{\mathbf{q}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_i, \tag{26}$$

$$\overline{\mathbf{q}}' = \frac{1}{N}\sum_{i=1}^{N}\mathbf{q}'_i. \tag{27}$$

(2) Calculate $\mathbf{p}_i$ and $\mathbf{p}'_i(i = 1,\ldots,N)$:

$$\mathbf{p}_i = \mathbf{q}_i - \overline{\mathbf{q}}, \tag{28}$$

$$\mathbf{p}'_i = \mathbf{q}'_i - \overline{\mathbf{q}}'. \tag{29}$$

(3) Calculate the matrix $\mathbf{H}$:

$$\mathbf{H} = \sum_{i=1}^{N}\mathbf{p}'_i\mathbf{p}_i^T. \tag{30}$$

(4) Find the SVD of $\mathbf{H}$:

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T. \tag{31}$$

(5) Compute the rotation matrix:

$$\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{32}$$

$$\mathbf{D} = \begin{cases} \mathbf{I}, & \text{if } \det(\mathbf{U}\mathbf{V}^T) = +1, \\ \operatorname{diag}(1,1,-1), & \text{if } \det(\mathbf{U}\mathbf{V}^{\mathbf{T}}) = -1. \end{cases} \tag{33}$$

(6) Find the translation vector and scaling factor:

$$s = \frac{\operatorname{tr}(\mathbf{P}\mathbf{P}'^T\mathbf{R})}{\operatorname{tr}(\mathbf{P}\mathbf{P}^T)}, \tag{34}$$

$$\mathbf{t} = \overline{\mathbf{q}}' - s\mathbf{R}\overline{\mathbf{q}}, \tag{35}$$

where $\mathbf{P} = [\mathbf{p}_1,\ldots,\mathbf{p}_N]$ and $\mathbf{P}' = [\mathbf{p}'_1,\ldots,\mathbf{p}'_N]$ are $3 \times N$ matrices.

In (32), matrix $\mathbf{D}$ is used as a "safeguard" making sure that the calculated matrix $\mathbf{R}$ is a rotation matrix and not a reflection in 3D space. The outline of the similarity registration procedure is given in Algorithm 1. The criterion used to terminate the iteration of the algorithm is based on the variation of the distance between the two surfaces at two successive iterations. According to the experimental results, the iteration of similarity registration is terminated when the variation, τ, is below 0.1 mm. Figure 3(a) shows an example of the results obtained by the similarity registration. The position of the model is fixed and the new face is transformed to align to the model. Although there are noticeable local misalignments, that is, around the mouth and eyes, due to different facial expressions, they are globally well matched.

3.2. Model Refinement. With the data registered to the current model using similarity transformation, the objective of the model refinement is to deform the model so that it is better aligned to the transformed data points. To estimate the optimal pose and shape parameters the whole process has to iterate. This can be seen as a superposition of the ICP method and the least squares projection onto the shape space. The least squares projection onto the shape space provides the SSV, $\hat{\mathbf{b}}$, which controls the deformations of the model. It is also postulated here that at the convergence point this vector can be used as a feature for interpretation of the face articulation. The SSV, $\hat{\mathbf{b}}$, for an observed face is calculated from

$$\hat{\mathbf{b}} = \mathbf{W}^T\left(\mathbf{Q}_c - \overline{\mathbf{Q}}\right), \tag{36}$$

Input: points $\mathbf{Q}$ from the new face and points $\mathbf{Q}'$ from the current model.
Output: transformed points $\mathbf{Q}$ using estimated similarity transformation.
Initialisation:
set threshold $\tau(\tau > 0)$ for terminating the iteration, $k = 0$,
$d_0 = \inf$, $e = \tau$;
while $e \geq \tau$ **do**
$k = k + 1$;
Compute correspondence $q_i \leftrightarrow q'_{j(i)}$ with
$j(i) = \arg\min_{j \in 1:N} \|q_i - q'_j\|$;
Compute pose parameters : $\mathbf{R}$, $\mathbf{T}$, and s using Equations (26)–(35);
Transform the points from set $\mathbf{Q}$ using similarity transformation $q_i = sRq_i + T$ and update set $\mathbf{Q}$ accordingly;
Measure misalignment d_k between corresponding points in the point set of new face $\mathbf{Q}$ and the point set of the model $\mathbf{Q}'$;
$e = d_{k-1} - d_k$;
End

Algorithm 1: Similarity registration.

Input: points $\mathbf{Q}$ of a new face, and the face model: $\mathbf{W}$,$\overline{\mathbf{Q}}$.
Output: Estimation of the SSV, $\hat{\mathbf{b}}$.
Initialization:
set threshold $\sigma(\sigma > 0)$ for terminating the iteration, $\hat{\mathbf{b}}_0 = 0$,
$k = 0$;
while $\|\hat{\mathbf{b}}_k - \hat{\mathbf{b}}_{k-1}\| \geq \sigma$ **do**
Calculate points from the deformed model: $\hat{\mathbf{Q}} = \mathbf{W}\hat{\mathbf{b}}_k + \overline{\mathbf{Q}}$;
Register points sets $\hat{\mathbf{Q}}$ and $\mathbf{Q}$ using Algorithm 1 and obtain the corresponding points $\mathbf{Q}_c$ for the transformed new face;
$k = k + 1$;
Project corresponding points $\mathbf{Q}_c$ onto the shape space
$\hat{\mathbf{b}}_k = \mathbf{W}^T(\mathbf{Q}_c - \overline{\mathbf{Q}})$;
end

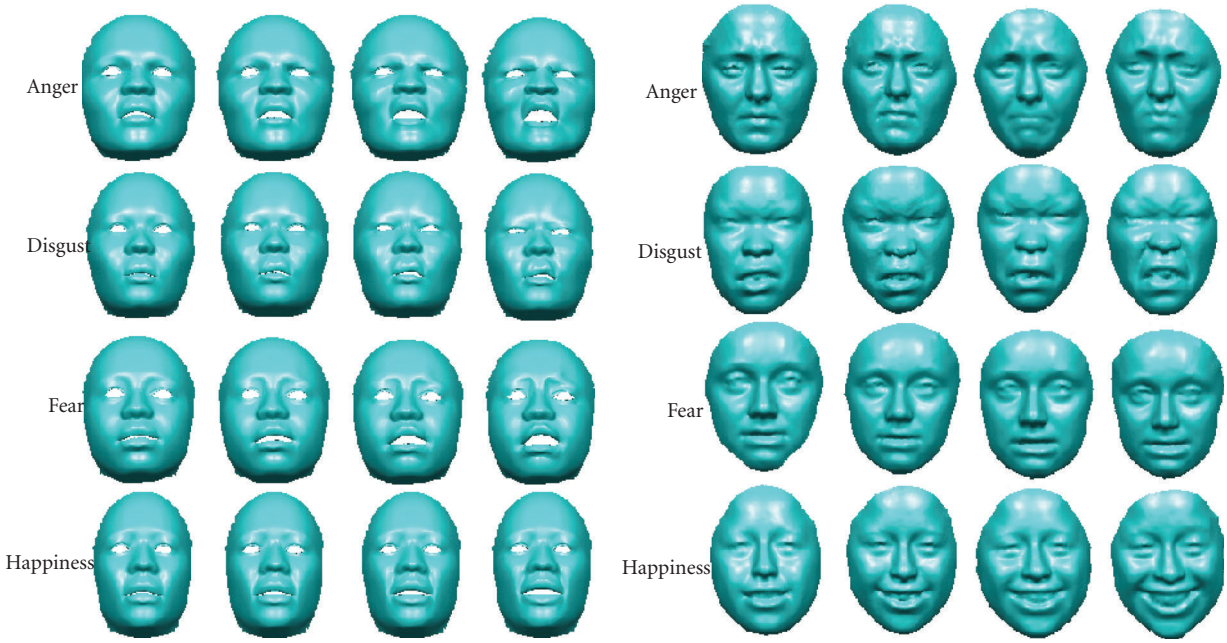
Algorithm 2: Model refinement.

where $\mathbf{Q}_c \in R^{3N}$ is a vector which contains N corresponding data points representing the new face. The mean vector of data points $\overline{\mathbf{Q}}$ and shape matrix $\mathbf{W}$ are obtained from (18) and (22), respectively. The details of the algorithm are explained in Algorithm 2. The criterion used to terminate the iteration of the model refinement is based on the change of the SSVs at two successive iterations. According to the experimental results, the iteration of the algorithm is terminated when the change of the SSVs, σ, is below 5. For most cases, it is seen that the shape variation of the model during the model refinement is negligible when the change of the SSVs is smaller that this preset threshold.

An example of the results obtained from the model refinement is shown in Figure 3(b). In this case the model is matched to a face with a strong fear expression. The intermediate states illustrate how the model is being deformed to match the new face during the refinement iterations.

4. Separability Analysis

To assess if the SSV can be used as a feature space for the facial expression analysis and recognition, the separability of the SSV-based features has been analysed, using qualitative and quantitative methods. In the qualitative analysis, the separability of the SSV-based features is examined visually in a low-dimensional SSV space. The quantitative analysis is carried out using one of the numerical separability criteria. Four types of data sets have been used in the separability analysis; they are 3D synthetic faces generated from the FaceGen Modeller, manually selected 3D facial landmarks from the BU-3DFE database, 3D face scans from the BU-3DFE database, and automatically detected 3D facial landmarks from the BU-3DFE database. All these data sets cover a wide variety of ethnicity, age range, as well as gender. Face samples from the FaceGen and BU-3DFE data sets



(a) Face samples from the FaceGen database

(b) Face samples from the BU-3DFE database

Figure 4: Face samples showing four different subjects and expressions with four levels of expression intensity.

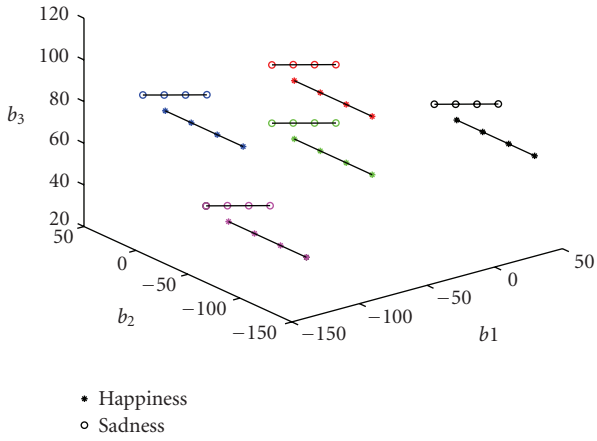
showing different individuals and different expressions are shown in Figure 4. The faces used for testing are not included in the training data sets used for building the SSM.

4.1. Qualitative Evaluation. Since the high-dimensional SSV-based features are hard to visualise, only the first three elements of the SSV are used for qualitative analysis. For different types of data, the first three principal components retain different levels of variability present in the training data set. With the retained variability defined as $\sum_{i=1}^{3} \lambda_i / \sum_{i=1}^{M} \lambda_i$, where λ_i and M are given in Section 2.2, the first three principal components retain around 51% of the total data variability for the model built using 450 synthetic faces. For the model built from the facial landmarks the first three principal components retain around 42% data shape variability, whereas for the model built using dense set of facial points the first three principal components retain 35% of the variability. The last two models were built using the same 450 faces randomly selected from the BU-3DFE database.

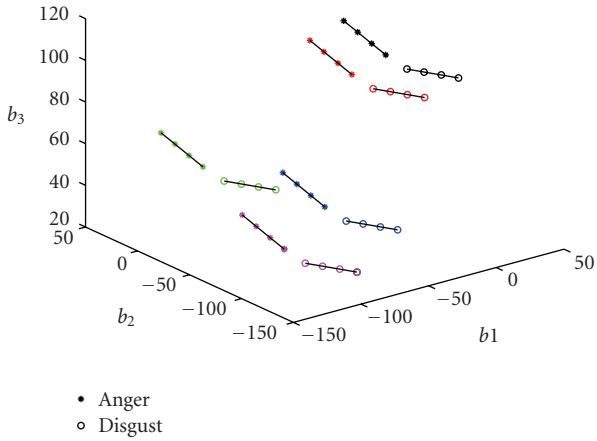
4.1.1. 3-D Synthetic Faces. Firstly, the 3D synthetic faces generated from the FaceGen Modeller are used to show the separability of the SSV-based features. The FaceGen Modeller is a commercial software designed to create realistic faces with controllable type and level of expressions for subjects of any ethnic origin or gender. Since the correspondence information is provided for all the face vertices (3428 vertices are used to represent all the synthetic faces), the SSM can be built directly without correspondence search. However, it needs to be stressed that the priori knowledge about the correspondence, for the faces in the training data set, was only used in the model building stage. In the model fitting stage the information about the data correspondence was ignored and the correspondence search was included in finding the SSV representation of the faces from the test sets.

For the evaluation, a training data set of 450 3D synthetic faces from 18 subjects was used to build the SSM. A sample of faces from the training data set is shown in Figure 4(a). Another 450 synthetic faces of 18 subjects were used for testing. The training and testing faces are mutually exclusive. First, for clarity of the presentation, Figure 5 shows the separability of the synthetic faces' SSVs for selected expression pairs with five different subjects and varying expression's intensity. The SSVs of the same subject and representing the same expression with various expression's intensity are linked together. Considering the expression's intensity as only variable the corresponding SSVs are aligned on the same line segment. It can be observed that the SSV-based features corresponding to different subjects and different facial expressions are well separated; furthermore the orientation of each line seems to define a type of the expression. Figure 6 shows the separability of the synthetic faces' SSVs for all six basic expressions and five subjects shown in different colours. It can be seen that the SSVs representing different expressions for the same subject are clustered together and the SSVs representing the same expression are located on the line segments having the same orientation which is independent of the subject.

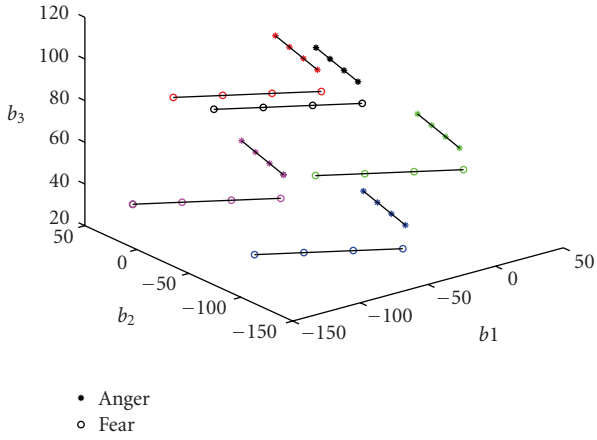
From the obtained results, showing clustered lines in the SSV space, it seems reasonable to postulate that the FaceGen Modeller uses a linear shape space model for face generation, whereby different eigen subspaces represent different face expressions as well as different face types. Such an approach



(a) Happiness versus Sadness



(b) Anger versus Disgust



(c) Anger versus Fear

Figure 5: Visualization of the synthetic faces separability using first three elements of the SSV and five different subjects.

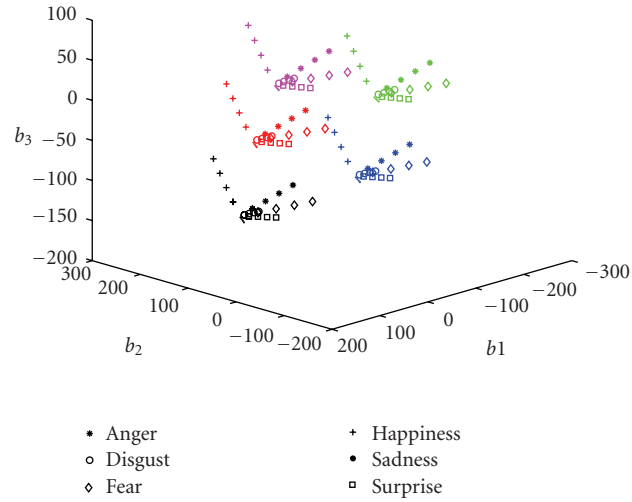
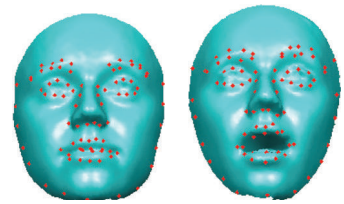


Figure 6: Visualization of the synthetic faces separability for six expressions and five subjects.

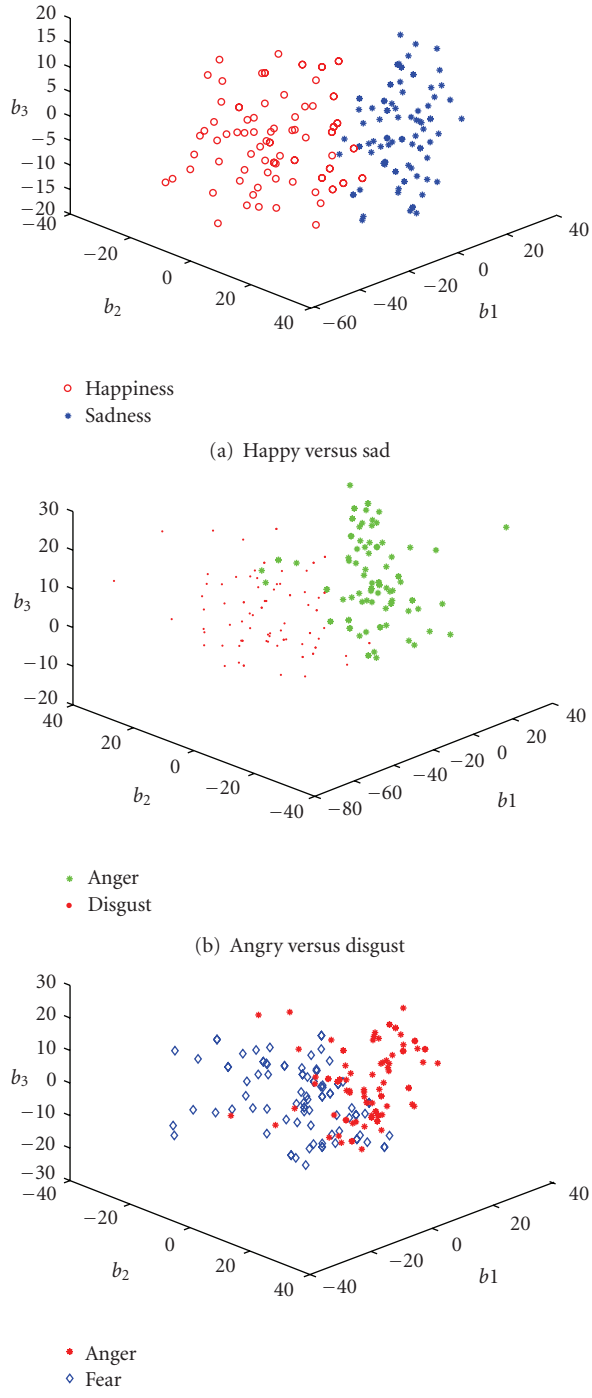


(a) Neutral. (b) Surprise.

Figure 7: Example of manually selected landmarks in two different faces from the BU-3DFE database.

for face generation was previously proposed in computer graphics literature [35]. From the presented results, it can be concluded that the proposed face registration method is able to recover the facial expression and subject control parameters used in the face generation model (e.g., orientations of the clustered lines in the SSV space define eigen faces responsible for generating different expressions in the FaceGen shape space model, whereas positions of the clustered lines define the subject's identity, as shown in Figures 5 and 6).

4.1.2. Manually Selected Facial Landmarks. To test that the SSV feature space can be used for classification of expressions present in real faces and in the same time to circumvent any potential problems caused by wrong data correspondence, tests were carried out on the SSM derived from manually selected landmarks on faces from the BU-3DFE database. Each set of 3D facial landmarks provided in the database contains 83 facial points, which are manually labeled around the areas that are most affected by changes of facial expressions including eyes, nose, brows, and mouth. Figure 7 illustrates positions of the landmarks on two different faces. The BU-3DFE database contains 100 subjects; for each subject, 25 various expressions are included, which can be categorised into neutral, happy, disgust, fear, angry, surprise, and sad [23]. The SSM was built using landmarks from 450 faces belonging to 18 randomly selected subjects. Another set of landmarks from a different set of 450 faces from 18 different subjects was used as a test set.



(a) Happy versus sad

(b) Angry versus disgust

(c) Angry versus fear

Figure 8: Separability analysis for manually selected landmarks using first three principal components.

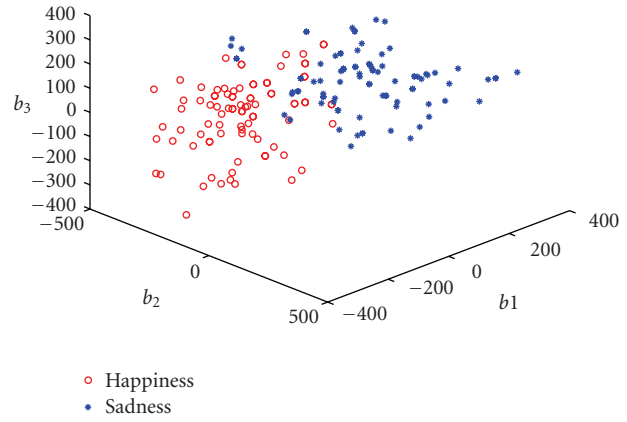
Figure 8 demonstrates the separability of the SSV feature space, derived using manually selected landmarks. The first three elements of the SSV were used with five types of facial expressions. Figure 8(a) shows that facial expressions of happy and sad can be easily separated even in a low-dimensional SSV feature space. This is in agreement with the general consensus that the expressions of sadness and happiness are the most recognisable human expressions as confirmed by a number of psychophysical test. Some of the expressions are not as well separated in the feature space as, for example, "angry" and "fear", as shown in Figure 8(c). Although they are partly "mixed" together in the low-dimensional shape space, it is still possible to separate the majority of these facial expressions. Again this result reflects findings of psychophysical tests, which confirm that expressions such as anger and fear can be easily misclassified by a human observer [36].

4.1.3. Full 3D Face Scans. The results from the previous section show that with the use of the SSV feature space it is possible to discriminate facial expressions on real facial scans. Unfortunately, although the SSM built from manually selected landmarks uses real faces, the correspondence is established manually. This approach would not be a satisfactory solution for most applications as the manual landmark selection is too tedious and time consuming. In this section discriminatory characteristics of the SSV feature space constructed using a dense set of facial points, as described in Section 3, are examined. As explained there, the correspondence is estimated automatically during the pose estimation stage of the model fitting process. It should be noted here that as the dense correspondence is not given in the training data set, the correspondence between points on different training facial scans is also estimated during the model building phase as explained in Section 2.1.

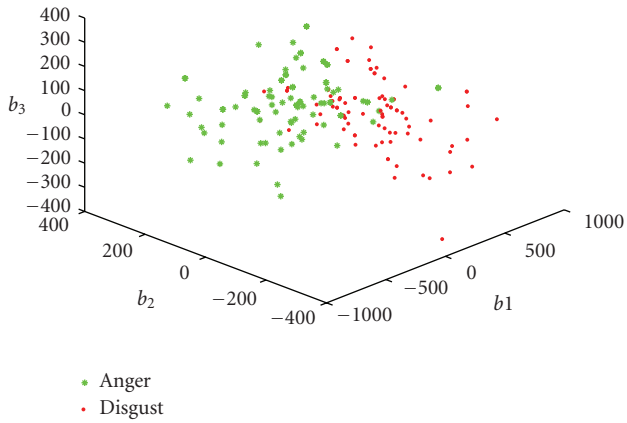
Figure 9 illustrates the separability of the facial expressions in the feature space of the first three principal components of the SSV built from the full facial scans. As in the previous section five different facial expressions were used. Similarly to the results shown for the manually selected facial landmarks the results demonstrate again that the SSV feature space offers a good expression separability.

4.1.4. Automatically Selected Facial Landmarks. As shown in the previous section, the SSV feature space built from full facial scans, using dense facial points, provides good separability of expressions. Additionally this approach is more practical as the correspondence is estimated automatically. Intuitively discriminatory characteristics of the SSV feature space can be further improved by using only information from the facial regions which are articulated the most during different expressions. In the "full facial scan" approach, all the points contribute to the SSM, but some points, that is, on a forehead, carry very little information about face expression. These points would still contribute to the variations of the SSM model as they would represent variability of facial shape for different subjects. Evaluation was therefore carried out to use the "full facial scan" SSM first to establish the correspondence between the model and the data and subsequently used the SSM built from predefined facial landmarks on the model for the facial expression representation.

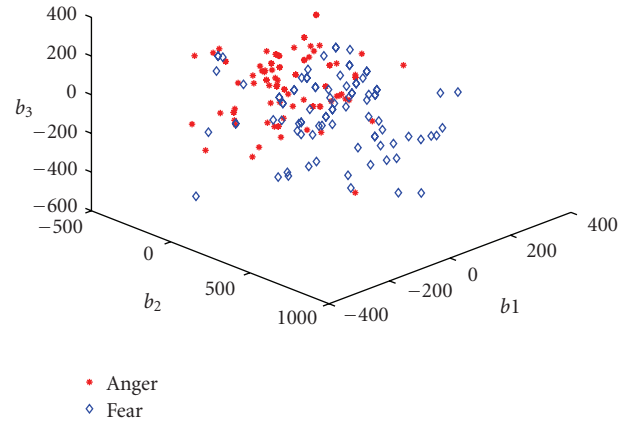
This approach is in principle very similar to using the SSV representing variations of the manually selected landmarks, with the difference that landmark selection



(a) Happy versus sad



(b) Angry versus disgust



(c) Angry versus fear

Figure 9: Separability of the facial expressions in the feature space of first three principal components of the SSV built from the full facial scans.

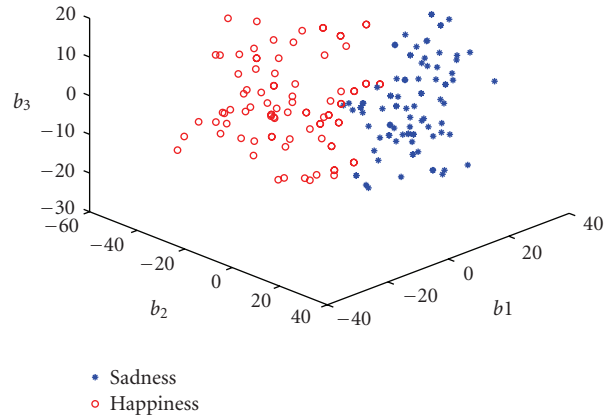
is automated, where the automation is achieved through registration of the "full facial scan" SSM with a new face. Since the corresponding indices of the facial landmarks on the model are already known, the positions of the corresponding landmarks on a new face scan can be directly estimated when the model is matched to the new face scan. In this case the surface registration error may introduce variability in the position of the landmarks which in turn may have negative effects on the classification performance. To examine registration accuracy of the proposed method tests were carried with the synthetic and real faces. In the experiments, for each data type, the model has been matched to 450 faces which were not used for the model building. Subsequently the Euclidean distance between corresponding landmarks on the deformed model and the test faces was calculated. The average distance between corresponding landmarks on the synthetic faces and the model, calculated from all the 450 test faces, was 1.49 mm with maximum error of 3.95 mm, whereas corresponding distances obtained for the real faces were 3.56 mm and 7.64 mm, respectively. The bigger registration errors obtained for the real faces are mainly thought to be due to the errors in the manual selection of the facial landmarks. Indeed it is believed that the errors in the manual landmark selection, used in the model building stage, have more influence on the method performance than the registration error.

Similar to the previous experiments, the model is built using 450 face scans from 18 randomly selected subjects, and another 450 face scans from 18 subjects are used for testing. Figure 10 shows the separability test for the proposed method. As before the first three principal components are used to represent five facial expressions. Compared to the case with the manually selected facial landmarks, the SSV feature space offers a comparable performance on separability of expressions.

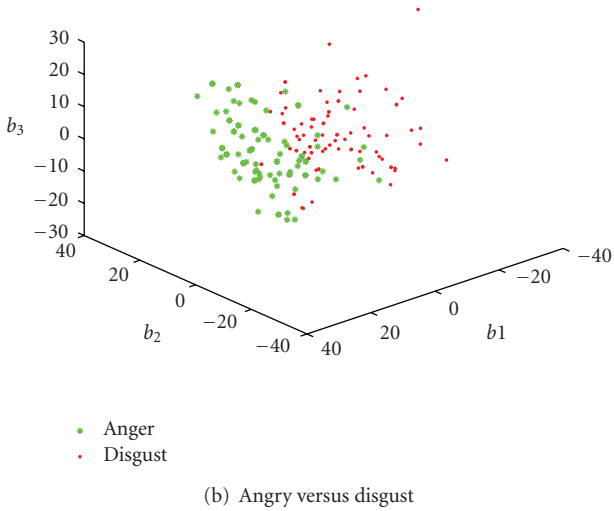
4.2. Quantitative Evaluation. The separability of the SSV-based features has been demonstrated qualitatively in the preceding section. This qualitative analysis shows that the SSV feature space exhibits good facial expressions separability. Due to the way the synthetic data is generated, the SSV-based features in that case were seen to form very distinctive linear patterns with different line directions responsible for different expressions. From experiments with real facial scans from the BU-3DFE database, the best performance is achieved when landmarks are used to build the SSM.

In order to further investigate the separability of the SSV-based features, a quantitative evaluation was carried out. For this analysis, only the SSM which was generated using the data from the real scans was included in the test. The data sets included (i) manually selected facial landmarks, (ii) full face scans, and (iii) automatically selected facial landmarks. In this quantitative evaluation, a computable criterion based on the within-class and between-class distances [37] was used to measure the separability of expressions in the corresponding SSV feature spaces. A similar criterion has been used by Wang and Yin [8] to evaluate the separability of topographic context (TC) and intensity-based features for the facial expression analysis and recognition. The criterion relies on the average between-class distance in the case of multiple categories, which is defined as follows:

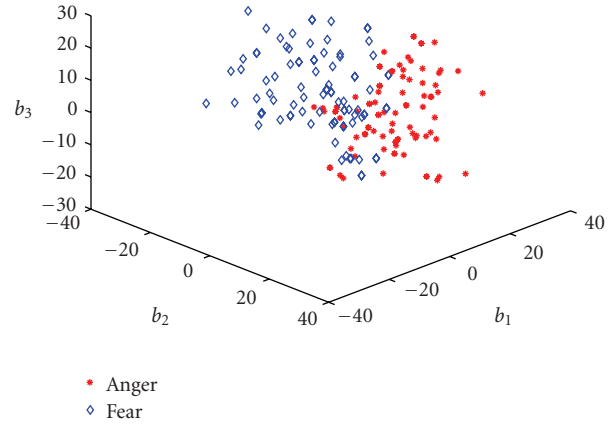
$$J_1(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N_c} P_i \sum_{j=1}^{N_c} P_j \frac{1}{M_i M_j} \sum_{k=1}^{M_i}\sum_{l=1}^{M_j} \delta\left(\mathbf{x}_k^i, \mathbf{x}_l^j\right), \tag{37}$$



(a) Happy versus sad



(b) Angry versus disgust



(c) Angry versus fear

FIGURE 10: Separability analysis for automatically selected landmarks using first three principal components.

where M_i and M_j are the number of samples in classes μ_i, and μ_j, x_k^i, and x_l^j are the K-dimensional feature vectors (SSV) with labels μ_i and μ_j. N_c is the number of distinct classes. P_i and P_j are the class-prior probabilities, and $\delta(x_k^i, x_l^j)$ denotes the distance between two samples, which is usually calculated using Euclidean distance. $J_1(x)$ can be represented in a compact form by using the so-called within-class scatter matrix $\mathbf{S}_W$ and between-class scatter matrix $\mathbf{S}_B$ [38], which are defined as follows:

$$\mathbf{S}_W = \sum_{i=1}^{N_c} P_i \frac{1}{M_i} \sum_{k=1}^{M_i} \left(\mathbf{x}_k^i - \mathbf{m}_i\right)\left(\mathbf{x}_k^i - \mathbf{m}_i\right)^T,$$
$$\mathbf{S}_B = \sum_{i=1}^{N_c} P_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \tag{38}$$

where $\mathbf{m}_i$ is the mean of samples in the ith class:

$$\mathbf{m}_i = \frac{1}{M_i} \sum_{k=1}^{M_i} \mathbf{x}_k^i, \tag{39}$$

and $\mathbf{m}$ is the mean for all of the samples:

$$\mathbf{m} = \sum_{i=1}^{N_c} P_i \mathbf{m}_i. \tag{40}$$

Using (38), $J_1(\mathbf{x})$ can be rewritten in the following form:

$$J_1(\mathbf{x}) = \mathrm{tr}(\mathbf{S}_W + \mathbf{S}_B). \tag{41}$$

Although $J_1(\mathbf{x})$ is an efficient and computable separability criterion for feature selection, it is not appropriate for comparing two or more features since the calculated value of $J_1(\mathbf{x})$ depends on the scale and dimensionality of the feature space. In order to compare two or more features which lie in different spaces with different scales and dimensionalities, a new criterion, $J_2(\mathbf{x})$, similar to $J_1(\mathbf{x})$, is used (as in [8]) based on a natural logarithm of the ratio of the determinant of the within-class scatter matrix and between-class scatter matrix. The new metric is defined as

$$J_2(\mathbf{x}) = \ln \frac{|(\mathbf{S}_W + \mathbf{S}_B)|}{s_{\max}}, \tag{42}$$

where $s_{\max}$ is the entry which contains the maximum value in matrix $\mathbf{\Sigma}$, and matrix $\mathbf{\Sigma}$ is obtained using the singular value decomposition (SVD) of matrix $\mathbf{S}_W$:

$$\mathbf{S}_W = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{43}$$

The larger the value of $J_2(\mathbf{x})$, the better the samples are separated. For comparison, the models using manually selected landmarks, full face scans, and automatically selected landmarks are built using the same 450 face scans as described in the previous sections. As shown in Figure 11, for the same ratio of retained variability in the model training data, $J_2(\mathbf{x})$ calculated for the SSV feature space of manually selected landmarks is always the highest. It is not though significantly different from $J_2(\mathbf{x})$ calculated for automatically selected landmarks when the retained variability is within the most commonly used range of 70% to 90%. As expected the separability based on $J_2(\mathbf{x})$ is the worst for the SSV computed from the full face scans.

Table 1: Recognition rate.

Feature type/classifier	LDA (% ± SD)	QDC (% ± SD)	NNC (% ± SD)
Synthetic faces	98.00 ± 1.33	**100.00** ± 0.00	70.89 ± 2.52
Real faces	**81.89** ± 6.96	80.11 ± 6.87	79.00 ± 7.09
Manually selected landmarks	**84.67** ± 4.12	82.44 ± 5.48	83.22 ± 6.42
Automatically selected landmarks	**82.78** ± 4.64	80.34 ± 5.03	81.78 ± 5.28

SD: Standard Deviation.

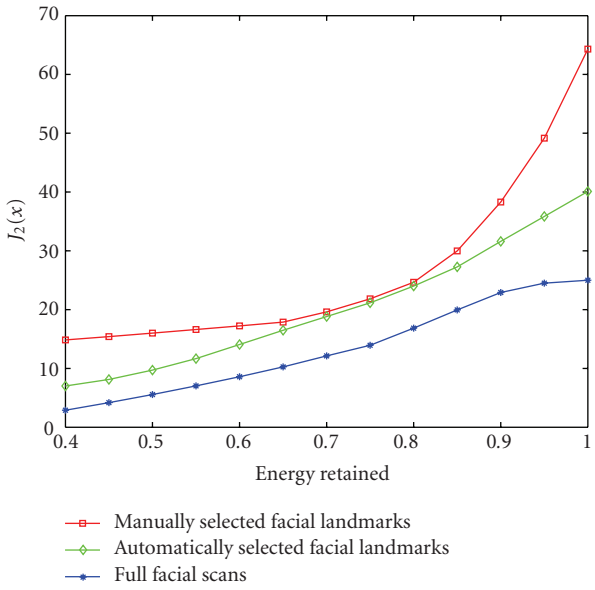


Figure 11: Quantitative evaluation of facial expression separability in the SSV feature spaces.

Table 2: Confusion matrix of the LDA classifier for the synthetic faces.

Input/output	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)
Anger	**94.00**	2.00	2.00	2.00	0.00	0.00
Disgust	0.00	**100.00**	0.00	0.00	0.00	0.00
Fear	0.00	2.00	**98.00**	0.00	0.00	0.00
Happy	0.00	0.00	0.00	**96.00**	2.00	2.00
Sad	0.00	0.00	0.00	0.00	**100.00**	0.00
Surprise	0.00	0.00	0.00	0.00	0.00	**100.00**

5. Experiments on Facial Expression Recognition

The separability analyses performed in the previous section indicate that the SSV feature space can be used in principle for classification of facial expressions. In this section, the person-independent facial expression recognition experiments using the high-dimensional SSV are conducted to further validate discriminatory properties of the SSV feature space. Again, four different types of facial data were used in the experiments. For each type of facial data, 900 faces from

Table 3: Confusion matrix of the LDA classifier the real faces.

Input/output	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)
Anger	**82.64**	3.48	4.17	3.47	4.86	1.39
Disgust	7.64	**78.47**	3.48	5.56	2.08	2.78
Fear	4.17	3.47	**72.59**	12.50	5.56	1.39
Happy	2.78	5.56	8.33	**83.33**	0.00	0.00
Sad	4.17	3.47	11.11	0.00	**81.25**	0.00
Surprise	0.00	0.00	4.17	2.78	0.00	**93.06**

Table 4: Confusion matrix of the LDA classifier for the manually selected landmarks.

Input/output	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)
Anger	**90.97**	4.17	0.00	0.00	4.86	0.00
Disgust	2.08	**89.58**	2.78	3.47	0.69	1.39
Fear	0.00	4.86	**70.14**	4.86	14.58	5.56
Happy	1.38	3.47	6.94	**88.19**	0.00	0.00
Sad	9.72	0.00	11.81	5.56	**72.92**	0.00
Surprise	2.08	0.00	1.39	0.00	0.00	**96.52**

Table 5: Confusion matrix of the LDA classifier for the automatically selected landmarks.

Input/output	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)
Anger	**90.28**	0.00	2.08	3.47	4.17	0.00
Disgust	4.16	**81.94**	4.16	2.78	1.40	5.56
Fear	2.78	4.16	**65.97**	8.18	11.81	5.56
Happy	5.56	0.00	6.94	**87.50**	0.00	0.00
Sad	3.47	5.56	10.42	3.47	**77.08**	0.00
Surprise	2.08	0.00	3.47	0.00	0.00	**94.44**

36 subjects are used containing six basic facial expressions of anger, disgust, fear, happiness, sadness, and surprise. These faces are divided into six subsets. Each subset contains six subjects with 25 faces per subject representing different expressions. During algorithm evaluation one of the subset is selected as the test subset while the remaining sets are used to construct the training database. Such experiment is repeated six times, with the different subsets selected as the test subset each time. As the focus of this paper is on the feature extraction and not on design of the best

Table 6: Confidence confusion matrix for the human observers using 2D video sequences.

Input/output	Anger (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)	Pain (%)
Anger	**55.39**	26.03	5.19	0.00	5.13	5.31	2.94
Disgust	7.70	**68.86**	5.22	0.00	8.47	4.59	5.16
Fear	3.80	9.02	**46.90**	0.00	7.13	23.90	9.26
Happiness	0.27	0.98	0.71	**92.95**	1.15	2.35	1.59
Sadness	4.07	5.87	3.63	0.71	**74.15**	3.22	8.33
Surprise	0.60	7.54	21.84	1.04	2.46	**64.64**	1.88
Pain	4.94	9.45	9.46	2.30	18.96	3.85	**51.04**

possible classification algorithm, three well-know (off-the-shelf) classification methods have been used, namely; linear discriminant analysis (LDA) [39], quadratic discriminant classifier (QDC) [40], and nearest neighbor classifier (NNC) [37]. The detailed description of these methods is beyond the scope of this paper but can be found in most of the textbooks on pattern recognition. The average recognition rates as well as standard deviations, calculated from all the six experiments using different subsets of faces, for the four different types of facial data, are give in Table 1. To have a fair comparison, the size of the SSV for each data type has been selected in such a way that the retained variability in each corresponding SSM is as similar as possible. For the results presented below, SSV for the synthetic data has 27 elements corresponding to 95.31% of retained variability, SSV for the full facial scans has 39 elements corresponding to 94.77%, whereas the SSV for the facial landmarks (both manually and automatically selected landmarks are using the same model) has 18 elements corresponding to 95.12%.

As shown in Table 1, all the classifiers achieve a similar recognition rate for the same data type with the extremely hight rates achieved for the synthetic faces for all the classifiers but the NNC classifier. For the facial data from the BU-3DFE database, the manually selected landmarks' SSVs always reach the highest recognition rate, whereas the real faces' SSVs always achieve the lowest rate. Tables 2 to 5 show LDA classifier confusion matrices for all the different data types used in the experiments.

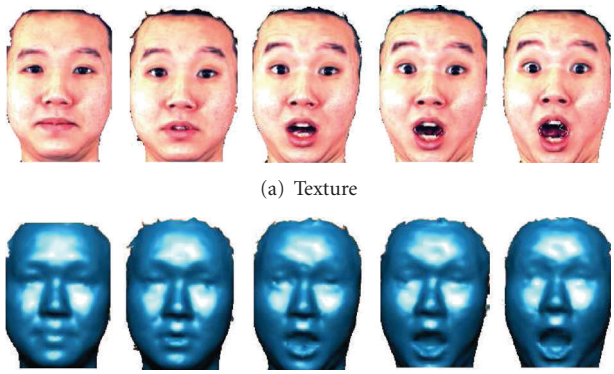
The presented results show that the SSV-based features can be used for recognition of facial expressions. The results for the manually selected landmarks are included only for a reference as using this data type is not practical due to lengthy process of landmarks' selection. From the presented results it can be seen that the best recognition rate of 82.78% obtained for the automatically selected landmarks is comparable with the best recognition rate of 84.67% obtained for the manually selected landmarks. This shows that the deformable surface registration method described in Section 3 is able to recover correct correspondences. An interesting insight into classification performance can be gained by looking at the confusion matrices. From Table 5 showing the confusion matrix of the LDA classifier for the automatically selected landmarks, it can be concluded that the anger and surprise expressions are all classified with above 90% accuracy, whereas the fear expression is only classified correctly in 65%. This can lead to the question about adequacy of the ground truth data. This is a difficult problem as the human expressions are very subjective by their nature. To demonstrate this Table 6 shows the confidence confusion matrix obtained for the human observers. This data has been obtained as a part of the project aiming to build and validate a 3D dynamic human facial expression database [41]. The specific results shown in the table are based on 10 observers asked to rank their confidence about recognising 7 facial expressions represented in 210 video clips and each video clip lasts 3 seconds. As it can be seen in the table the observers were very confident about recognising the happy expression whereas the fear expression was often confused with the surprise expression. This shows a "subjective" nature of the ground truth data. Although recognition rate of 65% for the fear expression in Table 5 seems to be quite low, when taking into account results presented in Table 6, they can be considered as reasonable.

6. Conclusions

A novel method for facial expression representation has been presented in this paper. It uses only 3D shape information, and therefore, in contrast to most of the methods using texture, our method is invariant to changes in the illumination, background, and to some extent viewing angle. The proposed method assumes that the SSV efficiently encodes facial expressions, and this encoding can be separated from the SSV variations caused by observing different faces. The performed tests indeed confirmed this hypothesis showing that the proposed representation is, at least partially, invariant to changes of the face ethnicity, gender, or age. A number of different configurations of the SSM have been tested. These include the SSM built from facial landmarks as well as full facial scans of real as well as simulated data. A fully automatic method has also been proposed for estimation of the SSV, with an iterative procedure which in turn estimates correspondence and shape parameters.

7. Future Work

In the method described in this paper the statistical shape model is built using a single database. In the case of the multiple databases which are subsequently integrated or combined together, a further improvement of the method



(a) Texture

(b) Shape

Figure 12: An example of the dynamic 3D face sequence.

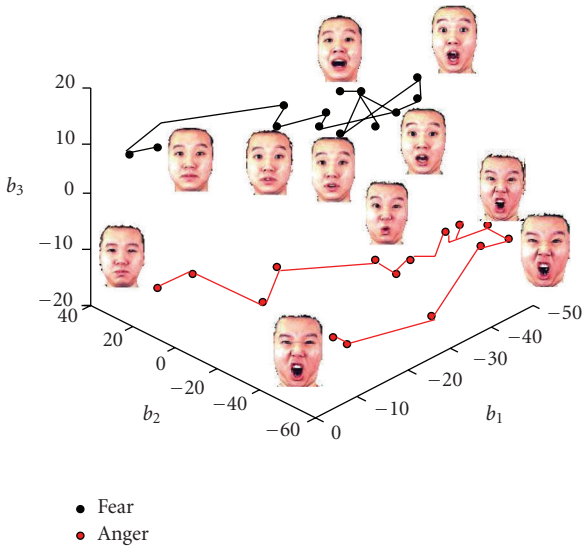


Figure 13: Trajectories of the first three principal components of the SSV-based feature on dynamic face sequences.

would include construction of a hierarchical system, where firstly the face type is decided upon, and subsequently the facial expression is recognized using shape model built from the facial expression database constructed for that specific face type detected in the previous step.

The separability results presented in the paper show that the SSV feature space can offer generally good separation for different expressions. For some expressions though such as angry and fear, the method provides only a limited separation, at least for the data used in the experiments. As a result, these expressions can be easily confused. One way to improve the separation of these "difficult" expressions is to provide more information to the model. From the reported psychophysical test it can be concluded that temporal information of the expression articulation provides important cues for human observers and helps them to correctly read expressions. Following this observation some simple tests were conducted with dynamic 3D facial scans. The dynamic face sequences are captured by the 3dMD scanner [42] in ADSIP research centre, and the facial landmarks set on each face in the sequence were manually labeled subsequently. An example of face sequence is shown in Figure 12. Using the face sequences, the trajectory of each specified facial expression is recorded and displayed in the 3D feature space. Figure 13 shows two trajectories plotted in the SSV domain for sequences representing fear and angry expressions. It can be seen that these trajectories are well separated in the SSV domain, thereby illustrating the potential usefulness of the temporal information of the face articulation for automatic expression classification.

Acknowledgments

The authors would like to acknowledge Dr. Lijun Yin from Binghamtopn University (USA) for making available to them BU-3DFED database. This work has been supported in part by the MEGURATH project (EPSRC grant no. EP/D077540/1).

References

[1] L. Yin, X. Wei, P. Longo, and A. Bhuvanesh, "Analyzing facial expressions using intensity-variant 3D data for human computer interaction," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, pp. 1248–1251, 2006.

[2] P. Eisert and B. Girod, "Analyzing facial expressions for virtual conferencing," *IEEE Computer Graphics and Applications*, vol. 18, no. 5, pp. 70–78, 1998.

[3] C. L. Lisetti and D. J. Schiano, "Automatic facial expression interpretation: where human-computer interaction, artificial intelligence and cognitive science intersect," *Pragmatics and Cognition*, vol. 8, no. 1, pp. 185–235, 2000.

[4] S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack, "Machine recognition and representation of neonatal facial displays of acute pain," *Artificial Intelligence in Medicine*, vol. 36, no. 3, pp. 211–222, 2006.

[5] S. D. Pollak and P. Sinha, "Effects of early experience on children's recognition of facial displays of emotion," *Developmental Psychology*, vol. 38, no. 5, pp. 784–791, 2002.

[6] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proceedings of the IEEE International Workshop on Human Computer Interaction (HCI '07)*, vol. 4796 of *Lecture Notes in Computer Science*, pp. 6–18, October 2007.

[7] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[8] J. Wang and L. Yin, "Static topographic modeling for facial expression recognition and analysis," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 19–34, 2007.

[9] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D facial expression recognition based on primitive surface feature distribution," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 17–26, 2006.

[10] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3D face recognition," in *Proceedings of the British Machine Vision Conference*, pp. 879–888, 2006.

[11] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *Proceedings of the 5th European Conference on Computer Vision*, pp. 581–595, 1998.

[12] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[13] M. Hoch, G. Fleischmann, and B. Girod, "Modeling and animation of facial expressions based on B-splines," *The Visual Computer*, vol. 11, pp. 87–95, 1994.

[14] P. Ekman and W. Friesen, *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, 1978.

[15] A. Saxena, A. Anand, and A. Mukerjee, "Robust facial expression recognition using spatially localized geometric model," in *Proceedings of the International Conference on Systemics, Cybernetics and Informatics*, pp. 124–129, 2004.

[16] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.

[17] H. Kobayashi and F. Hara, "Facial interaction between animated 3D face robot and human beings," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3732–3737, 1997.

[18] W. Quan, B. J. Matuszewski, L.-K. Shark, and D. Ait-Boudaoud, "Low dimensional surface parameterisation with applications in biometrics," in *Proceedings of the 4th International Conference Medical Information Visualisation: BioMedical Visualisation (MediViz '07)*, pp. 15–22, 2007.

[19] W. Quan, B. J. Matuszewski, L.-K. Shark, and D. Ait-Boudaoud, "3-D facial expression representation using B-spline statistical shape model," in *Proceedings of the Vision, Video and Graphics Workshop*, 2007.

[20] W. Quan, B. J. Matuszewski, L.-K. Shark, and D. Ait-Boudaoud, "3-D facial expression representation using statistical shape model," in *Proceedings of the BMVA Symposium on 3D Video—Anaysis, Display and Application*, 2008.

[21] H. B. Jähne and F. Haußecker, *Computer Vision and Applications*, Academic Press, New York, NY, USA, 2000.

[22] FaceGen Modeller, "Singular Inversions," 2003, http://www.facegen.com/.

[23] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR '06)*, vol. 2006, pp. 211–216, 2006.

[24] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[25] X. Lu and A. K. Jain, "Deformation modeling for robust 3D face matching," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 1377–1383, 2006.

[26] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[27] X. Lu and A. K. Jain, "Deformation analysis for 3D face matching," in *Proceedings of the 7th IEEE Workshop on Applications of Computer Vision (WACV/MOTIONS '05)*, vol. 1, pp. 99–104, 2005.

[28] K. Rohr, H. S. Stiehl, R. Sprengel, et al., "Point-based elastic registration of medical image data using approximating thin-plate splines," in *Proceedings of the 4th International Conference on Visualization in Biomedical Computing*, pp. 297–306, 1996.

[29] J. Peters and U. Reif, "The simplest subdivision scheme for smoothing polyhedra," *ACM Transactions on Graphics*, vol. 16, no. 4, pp. 420–431, 1997.

[30] A. Blake and M. Isard, *Active Contours*, Springer, Berlin, Germany, 1998.

[31] T. Vrtovec, D. Tomazevic, B. Likar, L. Travnik, and F. Pernus, "Automated construction of 3D statistical shape models," *Image Analysis and Stereology*, vol. 23, pp. 111–120, 2004.

[32] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[33] K. S. Arun and T. S. Huang, "Least-square fitting of two 3-D point sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 5, pp. 698–700, 1987.

[34] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[35] J. Ahlberg, "CANDIDE-3—an updated parameterized face," Tech. Rep. LiTH-ISY-R-2326, Image Coding Group, Department of EE, Linköping University, Linköping, Sweden, January 2001.

[36] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[37] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2001.

[38] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Proceedings of the 3rd International Conference on Face & Gesture Recognition*, pp. 336–341, 1998.

[39] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, New York, NY, USA, 2nd edition, 1989.

[40] B. D. Ripley, *Pattern Recogntion and Neural Networks*, Cambridge University Press, Cambridge, UK, 1996.

[41] B. J. Matuszewski, C. Frowd, and L. K. Shark, "Dynamic 3D facial database," Faculty of Science and Technology, University of Central Lancashire, 2008.

[42] 3DMD 3D Scanner, 3DMD, 2006, http://www.3dmd.com/.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 465193, 12 pages
doi:10.1155/2009/465193

Research Article

Evolutionary Discriminant Feature Extraction with Application to Face Recognition

Qijun Zhao,[1] David Zhang,[1] Lei Zhang,[1] and Hongtao Lu[2]

[1] *Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Hong Kong*
[2] *Department of Computer Science & Engineering, Shanghai Jiao Tong University, Shanghai 200030, China*

Correspondence should be addressed to Lei Zhang, cslzhang@comp.polyu.edu.hk

Received 27 September 2008; Revised 8 March 2009; Accepted 8 July 2009

Recommended by Jonathon Phillips

Evolutionary computation algorithms have recently been explored to extract features and applied to face recognition. However these methods have high space complexity and thus are not efficient or even impossible to be directly applied to real world applications such as face recognition where the data have very high dimensionality or very large scale. In this paper, we propose a new evolutionary approach to extracting discriminant features with low space complexity and high search efficiency. The proposed approach is further improved by using the bagging technique. Compared with the conventional subspace analysis methods such as PCA and LDA, the proposed methods can automatically select the dimensionality of feature space from the classification viewpoint. We have evaluated the proposed methods in comparison with some state-of-the-art methods using the ORL and AR face databases. The experimental results demonstrated that the proposed approach can successfully reduce the space complexity and enhance the recognition performance. In addition, the proposed approach provides an effective way to investigate the discriminative power of different feature subspaces.

1. Introduction

Biometrics has become a promising technique for personal authentication. It recognizes persons based on various traits, such as face, fingerprint, palmprint, voice, and gait. Most biometric systems use the images of those traits as inputs [1]. For example, 2D face recognition systems capture facial images from persons and then recognize them. However, there are many challenges in implementing a real-world face recognition system [2–4]. A well-known challenge is the "curse of dimensionality," which is also a general problem in pattern recognition [5]. It refers to the fact that as the dimension of data increases, the number of samples required for estimating the accurate representation of the data grows exponentially. Usually, the spatial resolution of a face image is at least hundreds of pixels and usually will be tens of thousands. From the statistical viewpoint, it will require tens of thousands of face samples to deal with the face recognition problem. However, it is often very difficult, even impossible, to collect so many samples. The dimensionality reduction techniques, including feature selection and extraction, are therefore widely used in face recognition systems to solve or alleviate this problem. In this paper, we will present a novel evolutionary computation-based approach to dimensionality reduction.

The necessity of applying feature extraction and selection before classification has been well demonstrated by researchers in the realm of pattern recognition [5, 6]. The original data are often contaminated by noise or contain much redundant information. Direct analysis on them could then be biased and result in unsatisfied classification accuracy. On the other hand, the raw data are usually of very high dimensionality. Not only does this lead to expensive computational cost but also causes the "curse of dimensionality." This may lead to poor performance in applications such as face recognition.

Feature extraction and selection are slightly different. Feature selection seeks for a subset of the original features. It does not transform the features but prunes some of them. Feature extraction, on the other hand, tries to acquire a new feature subset to represent the data by transforming the original data. Mathematically, given an $n \times N$ sample

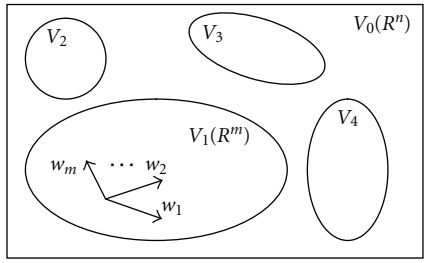


Figure 1: Linear feature extraction: from the subspace viewpoint.

matrix $X = [x_1 x_2 \cdots x_N]$ (n is the original dimension of samples, and N is the number of samples), a linear feature extraction algorithm could use an $n \times m$ transform matrix W to transform the data as $Y = W^T X = [y_1 y_2 \cdots y_N]$, where "$T$" is the transpose operator. Here, $0 < m \ll n$ is the dimension of the transformed feature subspace. Figure 1 illustrates this process. Suppose that the original data lie in the n-dimensional space V_0. Feature extraction is then to find out one of its subspaces which has the best discriminability and is called feature subspace, say the m-dimensional space V_1. In linear cases, an optimal projection basis of the feature subspace, $\{w_1, w_2, \ldots, w_m \in R^n\}$, can be calculated such that certain criterion is optimized. These basis vectors compose the column vectors in the transform matrix W.

Feature extraction is essentially a kind of optimization problem, and several criteria have been proposed to steer the optimization, for example, minimizing reconstruction error, maximizing reserved variance while reducing redundancy, and minimizing the within-class scatterance while maximizing the between-class scatterance, and so forth. Using such criteria, many feature extraction algorithms have been developed. Two well-known examples are Principal Component Analysis (PCA) [7] and Linear Discriminant Analysis (LDA) [5]. They represent two categories of subspace feature extraction methods [8] that are widely used in face recognition [3, 9–17]. In the context of face recognition, various feature subspaces have been studied [16, 17], for example, the range space of S_b and the null space of S_w. Here, S_b and S_w are the between-class and within-class scatter matrixes, defined as $S_b = (1/N)\sum_{j=1}^{L} N_j (M_j - M)(M_j - M)^T$ and $S_w = (1/N)\sum_{j=1}^{L}\sum_{i \in I_j} (x_i - M_j)(x_i - M_j)^T$, where $M = \sum_{i=1}^{N} x_i/N$ is the mean of all the N training samples, and $M_j = \sum_{i \in I_j} x_i/N_j$ is the mean of samples in the jth class ($j = 1, 2, \ldots, L$). A significant issue involved in these methods is how to determine m, that is, the dimension of the feature subspace. Unfortunately, neither PCA nor LDA gives systematic ways to determine the optimal dimension in the sense of classification accuracy. Currently, people usually choose the dimension by experience [9, 10, 18]. For example, the dimensionality of PCA-transformed space is set to 20 or 30 or $(N - 1)$, where "N" is the number of samples, and the dimensionality of LDA-transformed space is set to $(L - 1)$, where "L" is the number of classes. However, such method does not necessarily guarantee the best classification performance as we will show in our experiments. In addition, it is impractical or too expensive to search the whole solution space blindly in real applications such as face recognition because of the very high dimensionality of the original data.

Recently, some researchers [18–32] have explored the use of evolutionary computation (EC) methods [28] for feature selection and extraction. In these methods, the solution space is searched in guided random way, and the dimensionality of the feature subspace is automatically determined. Although these methods successfully avoid the manual selection of feature subspace dimensionality and good results have been reported on both synthetic and real-world datasets, most of them have very high space complexity and are often not applicable for high dimensional or large scale datasets [29]. In this paper, by using genetic algorithms (GA) [30], we will propose an evolutionary approach to extracting discriminant features for classification, namely, evolutionary discriminant feature extraction (EDFE). The EDFE algorithm has low space complexity and high search efficiency. We will further improve it by using the bagging technique. Comprehensive face recognition experiments have been performed on the ORL and AR face databases. The experimental results demonstrated the success of the proposed algorithms in reducing the space complexity and enhancing the recognition performance. In addition, the proposed method provides a way to experimentally compare the discriminability of different subspaces. This will benefit both researchers and engineers in analyzing and determining the best feature subspaces.

The rest of this paper is organized as follows. Sections 2 and 3 introduce in detail the proposed EDFE and bagging EDFE (BEDFE) algorithms. Section 4 shows the face recognition experimental results on the ORL and AR face databases. Section 5 gives some discussion on the relation between the proposed approach and relevant methods. Finally, Section 6 concludes the paper.

2. Evolutionary Discriminant Feature Extraction (EDFE)

This section presents the proposed EDFE algorithm, which is based on GA and subspace analysis. Algorithm 1 shows the procedures of EDFE.

2.1. Data Preprocessing: Centralization and Whitening. All the data are firstly preprocessed by centralization, that is, the total mean is subtracted from them:

$$\overline{x}_i = x_i - M, \quad i = 1, 2, \ldots, N. \tag{1}$$

The centralization applies a translational transformation to the data so that their mean is moved to the origin. This helps to simplify subsequent processing without loss of accuracy.

Generally, the components of data could span various ranges of values and could be of high order of magnitude. If we calculate distance-based measures like scatterance directly on the data, the resulting values can be of various orders of magnitude. As a result, it will be difficult to combine such measures with others. This is particularly serious in defining fitness for GA-based methods. Therefore, we further whiten the centralized data to normalize their variance to unity.

Step 1.
Preprocess the data using whitened principal component analysis (WPCA).
- Centralization
- Whitening

Step 2.
Calculate a search space for the genetic algorithm (GA).
- For example, the null space of S_w and the range space of S_b
- Heuristic knowledge can be used in defining search spaces

Step 3.
Use GA to search for an optimal projection basis in the search space defined in Step 2.
3.1. Randomly generate a population of candidate projection bases.
3.2. Evaluate all individuals in the population using a predefined fitness function.
3.3. Generate a new population using selection, crossover and mutation according to the fitness of current individuals.
3.4. If the stopping criterion is met, retrieve the optimal projection basis from the fittest individual and proceed to Step 4; otherwise, go back to 3.2 to repeat the evolution loop.

Step 4.
Use a classifier to classify new samples in the feature subspace obtained in Step 3.
- For example, Nearest Mean Classifier (NMC)

Algorithm 1: Procedures of the proposed EDFE algorithm.

This is done by the eigenvalue decomposition (EVD) on the covariance matrix of data. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n (\geq 0)$ be the eigenvalues of $S_t = (1/N)\sum_{i=1}^{N}(x_i - M)(x_i - M)^T$ and $\alpha_1, \alpha_2, \ldots, \alpha_n$ the corresponding eigenvectors. The whitening transformation matrix is then

$$W_{\text{WPCA}} = \left[\frac{\alpha_1}{\sqrt{\lambda_1}}\frac{\alpha_2}{\sqrt{\lambda_2}}\cdots\frac{\alpha_{N-1}}{\sqrt{\lambda_{N-1}}}\right]. \tag{2}$$

Here, we set the dimensionality of the whitened space to $(N-1)$, the rank of the covariance matrix. This means that we keep all the directions with nonzero variances, which ensures that no potential discriminative information is discarded from the data in whitening. Let $\overline{X}$ and $\widetilde{X}$ be the centralized and whitened data, then we have

$$\widetilde{X} = W_{\text{WPCA}}^T \overline{X}. \tag{3}$$

It can be easily proven that $\widetilde{S}_t = (1/N)\widetilde{X}\widetilde{X}^T = I_{N-1}$, where I_{N-1} is the $(N-1)$ dimensional identity matrix. In addition to normalizing the data variance, this whitening process also decorrelates the data components. For simplicity, we denote the preprocessed data in the whitened space still by X, omitting the tildes.

2.2. Calculating the Constrained Search Space. Unlike existing GA-based feature extraction algorithms, the proposed EDFE algorithm imposes some constraints on the search space so that the GA can search more efficiently in the constrained space. This idea originates from the fact that guided search, given correct guidance, is always better than blind search. It is widely accepted that heuristic knowledge, if properly used, could significantly improve the performance of systems. Keeping this in mind, we combine the EDFE algorithm with a scheme of constraining the search space as follows.

According to the Fisher criterion

$$W_{\text{LDA}} = \arg\max_{W}\left\{J_{\text{LDA}}(W) = \frac{W^T S_b W}{W^T S_w W}\right\}, \tag{4}$$

the most discriminative directions are most probably lying in the subspaces generated from S_w and S_b. Researchers [16, 17] have investigated the null space of S_w, denoted by null(S_w), and the range space of S_b, denoted by range(S_b), using analytical methods. It can be proved that the solution to (4) lies in these subspaces. We will use the EDFE algorithm to search for discriminant projection directions in null(S_w), range(S_w), and range(S_b), respectively, and compare their discriminability in recognizing faces. In this section, we present a method to calculate these three spaces. If some other subspace is considered, it is only needed to take its basis as the original basis of the search space.

Before proceeding to the detailed method of calculating the basis for null(S_w), range(S_w), and range(S_b), we first give the definitions of these three subspaces as follows

$$\text{null}(S_w) = \{v \mid S_w v = 0,\ S_w \in R^{n\times n},\ v \in R^n\}, \tag{5}$$

$$\text{range}(S_w) = \{v \mid S_w v \neq 0,\ S_w \in R^{n\times n},\ v \in R^n\}, \tag{6}$$

$$\text{range}(S_b) = \{v \mid S_b v \neq 0,\ S_b \in R^{n\times n},\ v \in R^n\}. \tag{7}$$

According to the definitions of S_w and S_b, the ranks of them are, respectively,

$$\text{rank}(S_w) = \min\{n, N-L\}, \qquad \text{rank}(S_b) = \min\{n, L-1\}. \tag{8}$$

These ranks determine the numbers of vectors in the bases of range(S_w), null(S_w), and range(S_b). Next, we introduce an efficient method to calculate the basis.

To get a basis of range(S_w), we use the EVD again. However, in real applications of image recognition, the

dimensionality of data, n, is often very high. This makes it computationally infeasible to conduct EVD directly on $S_w \in R^{n\times n}$. Instead, we calculate the eigenvectors of S_w from another $N \times N$ matrix S'_w [9]. Let

$$H_w = [x_1 x_2 \cdots x_N] \in R^{n\times N}, \tag{9}$$

then

$$S_w = \frac{1}{N} H_w H_w^T. \tag{10}$$

Note that the data have already been centralized and whitened. Let

$$S'_w = \frac{1}{N} H_w^T H_w, \tag{11}$$

and suppose (λ, α') to be an eigenvalue and the associated eigenvector of S'_w, that is,

$$S'_w \alpha' = \lambda \alpha'. \tag{12}$$

Substituting (7) into (8) gives

$$\frac{1}{N} H_w^T H_w \alpha' = \lambda \alpha'. \tag{13}$$

Multiplying both sides of (9) with H_w, we have

$$\frac{1}{N} H_w H_w^T H_w \alpha' = \lambda H_w \alpha'. \tag{14}$$

With (10) and (6), there is

$$S_w \cdot (H_w \alpha') = \lambda \cdot (H_w \alpha'), \tag{15}$$

which proves that $(\lambda, H_w\alpha')$ are the eigenvalue and eigenvector of S_w. Therefore, we first calculate the rank(S_w) dominant eigenvectors of S'_w, $(\alpha'_1, \alpha'_2, \ldots, \alpha'_{\text{rank}(S_w)})$, which have largest positive associated eigenvalues. A basis of range(S_w) is then given by

$$B_{\text{range}}(S_w) = \{\alpha_i = H_w \alpha'_i \mid i = 1, 2, \ldots, \text{rank}(S_w)\}. \tag{16}$$

The basis of range(S_b) can be calculated in a similar way. Suppose that the N column vectors of $H_b \in R^{n\times N}$ consist of M_j $(j = 1, 2, \ldots, L)$ with N_j entries, then $S_b = (1/N) H_b H_b^T$. Let $S'_b = (1/N) H_b^T H_b$ and $\{\beta'_i \mid i = 1, 2, \ldots, \text{rank}(S_b)\}$ be its rank(S_b) dominant eigenvectors. The basis of range(S_b) is then

$$B_{\text{range}}(S_b) = \{\beta_i = H_b \beta'_i \mid i = 1, 2, \ldots, \text{rank}(S_b)\}. \tag{17}$$

Based on the basis of range(S_w), it is easy to get the basis of null(S_w) through calculating the orthogonal complement space of range(S_w).

2.3. Searching: An Evolutionary Approach

2.3.1. Encoding Individuals. Binary individuals are widely used owing to their simplicity; however, the specific definition is problem dependent. As for feature extraction, it depends on how the projection basis is constructed. The construction of projection basis vectors is to generate candidate transformation matrixes for the GA algorithm. Usually, the whole set of candidate projection basis vectors are encoded in an individual. This is the reason why the space complexity of existing GA-based feature extraction algorithms is so high. For example, in EP [18], one individual has $(5n^2 - 4n)$ bits, where n is the dimensionality of the search space. In order to reduce the space complexity and make the algorithm more applicable for high dimensional data, we propose to construct projection basis vectors using the linear combination of the basis of search space and the orthogonal complement technique. As a result, only one vector is needed to encode for each individual.

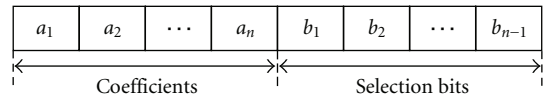


Figure 2: The individual defined in EDFE. Each coefficient is represented by 11 bits.

First, we generate one vector via linearly combining the basis of the search space. Suppose that the search space is R^n and let $\{e_i \in R^n \mid i = 1, 2, \ldots, n\}$ be a basis of it, and let $\{a_i \in R \mid i = 1, 2, \ldots, n\}$ be the linear combination coefficients. Then we can have a vector as follows:

$$v = \sum_{i=1}^{n} a_i e_i. \tag{18}$$

Second, we calculate a basis of the orthogonal complement space in R^n of $V = \text{span}\{v\}$, the space expanded by v. Let $\{u_i \in R^n \mid i = 1, 2, \ldots, n-1\}$ be the basis, and $U = \text{span}\{u_1, u_2, \ldots, u_{n-1}\}$, then

$$R^n = V \oplus U, \tag{19}$$

where "$\oplus$" represents the direct sum of vector spaces, and

$$U = V^{\perp}, \tag{20}$$

where "$\perp$" denotes the orthogonal complement space. Finally, we randomly choose part of this basis as the projection basis vectors.

According to the above method of generating projection basis vectors, the information encoded in an individual includes the n combination coefficients and $(n-1)$ selection bits. Each coefficient is represented by 11 bits with the leftmost bit denoting its sign ("0" means negative and "1" positive) and the remaining 10 bits representing its value as a binary decimal. Figure 2 shows such an individual, in which the selection bits $b_1, b_2, \ldots, b_{n-1}$, taking a value of "0" or "1," indicate whether the corresponding basis vector is chosen as a projection basis vector or not. The individual under such definition has $(12n - 1)$ bits. Apparently, it is much shorter than that by existing GA-based feature extraction algorithms (such as EP), and consequently the proposed EDFE algorithm has a much lower space complexity.

2.3.2. Evaluating Individuals. We evaluate individuals from two perspectives, pattern recognition and machine learning. Our ultimate goal is to accurately classify data. Therefore, from the perspective of pattern recognition, an obvious measure is the classification accuracy in the obtained feature subspace. In fact, almost all existing GA-based feature extraction algorithms use this measure in their fitness functions. They calculate this measure based on the training samples or a subset of them. However, after preprocessing the data using WPCA, the classification accuracy on the training samples is always almost 100%. In [26], Zheng et al. also pointed this out when they used PCA to process the training data. They then simply ignored its role in evaluating individuals. Different from their method, we keep this classification term in the fitness function but use a validation set, instead of the training set. Specifically, we randomly choose from the N_{va} samples to create a validation set Ω_{va} and use the remaining $N_{\text{tr}} = (N - L \times N_{\text{va}})$ samples as the training set Ω_{tr}. Assume that $N_{\text{va}}^c(D)$ samples in the validation set are correctly classified in the feature subspace defined by the individual D on the training set Ω_{tr}; the classification accuracy term for this individual is then defined as

$$\zeta_a(D) = \frac{N_{\text{va}}^c(D)}{N_{\text{va}}}. \tag{21}$$

From the machine learning perspective, the generalization ability is an important index of machine learning systems. In previous methods, the between-class scatter is widely used in fitness functions. However, according to the Fisher criterion, it is better to simultaneously minimize the within-class scatter and maximize the between-class scatter. Thus, we use the following between-class and within-class scatter distances of samples in the feature subspace:

$$\begin{aligned} d_b(D) &= \frac{1}{N}\sum_{j=1}^{L} N_j \left(M_j - M\right)^T \left(M_j - M\right), \\ d_w(D) &= \frac{1}{L}\sum_{j=1}^{L} \frac{1}{N_j} \sum_{i \in I_j} (y_i - M_j)^T \left(y_i - M_j\right) \end{aligned} \tag{22}$$

to measure the generalization ability as

$$\zeta_g(D) = d_b(D) - d_w(D). \tag{23}$$

Here, M and M_j, $j = 1, 2, \ldots, L$, are calculated based on $\{y_i \mid i = 1, 2, \ldots, N\}$ in the feature subspace.

Finally we define the fitness function as the weighted sum of the above two terms:

$$\zeta(D) = \pi_a \zeta_a(D) + (1 - \pi_a)\zeta_g(D), \tag{24}$$

where $\pi_a \in [0, 1]$ is the weight. The accuracy term ζ_a in this fitness function lies in interval $[0, 1]$. Thus, it is reasonable to make the value of the second generalization term ζ_g be of a similar magnitude order to ζ_a. This verifies the motivation of data preprocessing by centralizing and whitening.

2.3.3. Generating New Individuals. To generate new individuals from the current generation, we use three genetic operators, selection, crossover, and mutation. The selection is based on the relative fitness of individuals. Specifically, the ratio of the fitness of an individual to the total fitness of the population determines how many times the individual will be selected as parent individuals. After evaluating all individuals in the current population, we select $(S - 1)$ pairs of parent individuals from them, where S is the size of the GA population. Then the population of the next generation consists of the individual with the highest fitness in the current generation and the $(S - 1)$ new individuals generated from these parent individuals.

The crossover operator is conducted under a given probability. If two parent individuals are not subjected to crossover, the one having higher fitness will be chosen into the next generation. Otherwise, two crossover points are randomly chosen, one of which is within the coefficient bits and the other is within the selection bits. These two points divide both parent individuals into three parts, and the second part is then exchanged between them to form two new individuals, one of which is randomly chosen as an individual in the next generation.

At last, each bit in the $(S - 1)$ new individuals is subjected to mutation from "0" to "1" or reversely under a specific probability. After applying all the three genetic operators, we have a new population for the next GA iteration.

2.3.4. Imposing Constraints on Searching. As discussed before, to further improve the search efficiency and the performance of the obtained projection basis vectors, some constraints are necessary for the search space. Thanks to the linear combination mechanism used by the proposed EDFE algorithm, it is very easy to force the GA to search in a constrained space. Our method is to construct vectors by linearly combining the basis of the constrained search space, instead of the original space. Take null(S_w), the null space of S_w, as an example. Suppose that we want to constrain the GA to search in null(S_w). Let $\{\alpha_i \in R^n \mid i = 1, 2, \ldots, m\}$ be the eigenvectors of S_w associated with zero eigenvalues. They form a basis of null(S_w). After obtaining a vector v via linearly combining the above basis, we have to calculate the basis of the orthogonal complement space of $V = \text{span}\{v\}$ in the constrained search space null(S_w), but not the original space R^n (referring to Section 2.1). For this purpose, we first calculate the isomorphic space of V in R^m, denoted by $\hat{V} = \text{span}\{P^T v\}$, where $P = [\alpha_1 \alpha_2 \cdots \alpha_m]$ is an isomorphic mapping. We then calculate a basis of the orthogonal complement space of $\hat{V}$ in R^m. Let $\{\hat{\beta}_i \in R^m \mid i = 1, 2, \ldots, m - 1\}$ be the obtained basis. Finally, we map this basis back into null(S_w) through $\{\beta_i = P\hat{\beta}_i \in R^n \mid i = 1, 2, \ldots, m - 1\}$.

The following theorem demonstrates that $\{\beta_i \mid i = 1, 2, \ldots, m - 1\}$ comprise a basis of the orthogonal complement space of V in null(S_w).

Theorem 1. *Assume that $A \subset R^n$ is an m-dimensional space, and $P = [\alpha_1 \alpha_2 \cdots \alpha_m]$ is an identity orthogonal basis of A, where $\alpha_i \in R^n$, $i = 1, 2, \ldots, m$. For any $v \in A$, suppose that $\hat{V} = span\{P^T v\} \subset R^m$. Let $\{\hat{\beta}_i \in R^m \mid i = 1, 2, \ldots, m - 1\}$*

be an identity orthogonal basis of the orthogonally complement space of $\hat{V}$ in R^m, then $\{\beta_i = P\widehat{\beta}_i \in R^n \mid i = 1, 2, \ldots, m-1\}$ is a basis of the orthogonally complement space of $V = span\{v\}$ in A.

Proof. See Appendix 6. □

3. Bagging EDFE

The EDFE algorithm proposed above is very applicable to high-dimensional data because of its low space complexity. However, since it is based on the idea of subspace methods like LDA, it could suffer from the outlier and over-fitting problems when the training set is large. Moreover, when there are many training samples, the null(S_w) becomes small, resulting in poor discrimination performance in the space. Wang and Tang [33] proposed to solve this problem using two random sampling techniques, random subspace and bagging. To improve the performance of the EDFE algorithm on large scale datasets, we propose to incorporate the bagging technique into the EDFE algorithm and hence develop the bagging evolutionary discriminant feature extraction (BEDFE) algorithm.

Bagging (acronym for Bootstrap AGGregatING), proposed by Breiman [34], uses resampling to generate several random subsets (called random bootstrap replicates) from the whole training set. From each replicate, one classifier is constructed. The results by these classifiers are integrated using some fusion scheme to give the final result. Since these classifiers are trained from relatively small bootstrap replicates, the outlier and over-fitting problems for them are expected to be alleviated. In addition, the stability of the overall classifier system can be improved by integration of several (weak) classifiers.

Like Wang and Tang's method, we randomly choose some classes from all the classes in the training set. The training samples belonging to these classes compose a bootstrap replicate. Usually, the unchosen samples become useless in the learning process. Instead, we do not overlook these data, but rather use them for validation and calculate the classification accuracy term in the fitness function. Below are the primary steps of the BEDFE algorithm.

(1) Preprocess the data using centralizing and whitening.

(2) Randomly choose some classes, say $\hat{L}$ classes, from all the L classes in the training set. The samples belonging to the $\hat{L}$ classes compose a bootstrap replicate used for training, and those belonging to the other $(L - \hat{L})$ classes are used for validation. Totally, K replicates are created (different replicates could have different classes).

(3) Run the EDFE algorithm on each replicate to learn a feature subspace. In all, K feature subspaces are obtained.

(4) Classify each new sample using a classifier in the K feature subspaces, respectively. The resulting K results are combined by a fusion scheme to give the final result.

There are two key steps in the BEDFE algorithm: how to do validation and classification, and how to fuse the results from different replicates. In the following we present our solutions to these two problems.

3.1. Validation and Classification. As shown above, a training replicate is created from the chosen $\hat{L}$ classes. Based on this training replicate, an individual in the EDFE population generates a candidate projection basis of feature subspace. All the samples in the training replicate are projected into this feature subspace. The generalization term in the fitness function is then calculated from these projections. To obtain the value of the classification accuracy term, we again randomly choose some samples from all the samples of each class in the $(L - \hat{L})$ validation classes to form the validation set. We then project the remaining samples in these classes to the feature subspace and calculate the mean as the prototype for each validation class according to the projections. Finally, the chosen samples are classified based on these prototypes using a classifier. The classification rate is used as the value of the classification accuracy term in the fitness function.

After running the EDFE algorithm on all the replicates, we get K feature subspaces as well as one projection basis for each of them. For each feature subspace, all the training samples (including the samples in training replicates and validation classes) are projected into the feature subspace, and the means of all classes are calculated as the prototypes of them. To classify a new sample, we first classify it in each of the K feature subspaces based on the class prototypes in that space and then fuse the K results to give the final decision, which is introduced in the following part.

3.2. The Majority Voting Fusion Scheme. A number of fusion schemes [35, 36] have been proposed in literature of multiple classifiers and information fusion. In the present paper, we only focus on Majority Voting for its intuitiveness and simplicity. Let $\{M_j^k \in R^{l_k} \mid j = 1, 2, \ldots, L; k = 1, 2, \ldots, K\}$ be the prototype of class j in the kth feature subspace, whose dimensionality is l_k. Given a new sample (represented as a vector), we first preprocess it by centralization and whitening; that is, the mean of all the training samples is subtracted from it, and the resulting vector is projected into the whitened PCA space learned from the training samples. Denote by x_t the preprocessed sample. It is projected into each of the K feature subspaces, resulting in y_t^k in the kth feature subspace, and classified in these feature subspaces, respectively. Finally, the Majority Voting scheme is employed to fuse the classification results obtained in the K feature subspaces.

Majority Voting is one of the simplest and most popular classifier fusion schemes. Take the Nearest Mean Classifier (NMC) and the kth feature subspace as an example. The NMC assigns x_t to the class $c^k \in \{1, 2, \ldots, L\}$ such that

$$\left\| y_t^k - M_{c^k}^k \right\| = \min_{j \in \{1,2,\ldots,L\}} \left\| y_t^k - M_j^k \right\|. \tag{25}$$

In other words, it votes for the class whose prototype is closest to y_t^k. After classifying x_t in all the K feature

Table 1: General information and settings of the used databases.

Database	Sub number	Size number	Image/Sub number	Train number	Validation number	Test number
ORL	40	92×112	10	4(2)	1(3)	5
AR	120	80×100	14	6(3)	1(4)	7

[a] From the first column to the last column: the name of the database, the number of subjects, the size of images, the number of images per subject, the number of training samples per subject, the number of validation samples per subject, and the number of test samples per subject.
[b]The numbers in parentheses are the numbers of samples per validation subject used by BEDFE to calculate the class prototypes and to evaluate the training performance.

subspaces, we get K results $\{c^k \mid k = 1, 2, \ldots, K\}$. Let Votes$(i)$ be the number of votes obtained by class i, that is,

$$\text{Votes}(i) = \#\left\{c^k = i \mid k = 1, 2, \ldots, K\right\}, \tag{26}$$

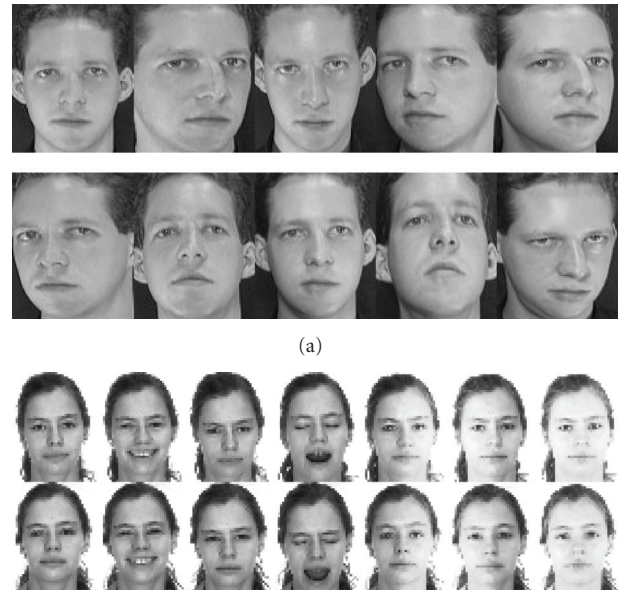
where "#" denotes the cardinality of a set. The final class label of x_t is determined to be c if

$$\text{Votes}(c) = \max_{i \in \{1,2,\ldots,L\}} \text{Votes}(i). \tag{27}$$

4. Face Recognition Experiments

Face recognition experiments have been performed on the ORL and AR face databases. Due to the high dimensionality of the data, conventional GA-based feature extraction methods like EP [18] and EDA [37] cannot be directly applied to these two databases unless reducing the dimensionality in advance. By contraries, the EDFE and BEDFE algorithms proposed in this paper can still work very well with them. As an application of the algorithms, we will use them to investigate the discriminative ability of the three subspaces, null(S_w), range(S_w), and range(S_b). We will experimentally demonstrate the necessity of carefully choosing the dimension of feature subspace. Finally, we will compare the proposed algorithms with some state-of-the-art methods in the literature, that is, Eigenfaces [9], Fisherfaces [10], Null-space LDA [16], EP [18], and EDA+Full-space LDA [32].

4.1. The Face Databases and Parameter Settings. The ORL database of faces [38] contains 400 face images of 40 distinct subjects. Each subject has 10 different images, which were taken at different times. These face images have variant lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). They also display slight pose changes. The size of each image is 92×112 pixels, with 256 gray levels per pixel. The AR face database [39] has much larger scale than the ORL database. It has over 4000 color images of 126 people (70 men and 56 women), which have different facial expressions, illumination conditions, and occlusions (wearing sun-glasses and scarf). In our experiments, we randomly chose some images of 120 subjects and discarded the samples of wearing sun-glasses and scarf. In the resulting dataset, there are 14 face images for each chosen subject, totally 1680 images. All these images were converted to gray scale images, and the face portion on them was manually cropped and resized to 80×100 pixels. For both databases, all images were preprocessed by histogram equalization. Table 1 lists some general information of the two databases, and Figure 3 shows some sample images of them.



(a)

(b)

Figure 3: Some sample images in the (a) ORL and (b) AR face databases.

In the GA algorithm, we set the probability of crossover to 0.8, the probability of mutation to 0.01, the size of population to 50, and the number of generations to 100. For the weight of the classification accuracy term in the fitness function, we considered the following choices for EDFE: $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. After finding the weight which gives the best classification accuracy for a dataset, we adopted it in BEDFE on the dataset. Regarding the number of bagging replicates in BEDFE, we conducted experiments for four cases, that is, using 3, 5, 7, and 9 replicates, and then chose the best one among them. The results will be presented in the following parts.

To create an evaluation face image set, all the sample images in each database were randomly divided into three parts: the training set, the validation set, and the test set. In the experiments on the ORL database, four images were randomly chosen from the samples of each subject for training, one image for validation and the remaining five images for test. In the experiments on the AR database, six images were randomly chosen for training from the samples of each subject, one image for validation, and the rest seven

Table 2: Recognition accuracy of EDFE in different subspaces.

Database	Null(S_w)	Range(S_w)	Range(S_b)
ORL	90.3%	78.1%	79.5%
AR	95.33%	80.95%	81.67%

Table 3: Recognition accuracy of BEDFE in different subspaces.

Database	Null(S_w)	Range(S_w)	Range(S_b)
ORL	91.3%	80.02%	81.14%
AR	96.86%	83.1%	83.38%

Table 4: The mean and standard deviation of recognition accuracy (%) of the proposed EDFE and BEDFE methods and some other state-of-the-art methods on the ORL and AR face databases.

Method	ORL face database	AR face database
Eigenfaces	90.15 ± 3.2	82.68 ± 0.9
Fisherfaces	91.6 ± 1.51	96.99 ± 0.7
Null-space LDA	89.75 ± 1.21	96.71 ± 0.59
EP	80.31 ± 3.5	N/A
EDA+Full-space LDA	92.5 ± 2.1	97.02 ± 0.8
EDFE+Full-space(S_w)	93 ± 1.8	97.9 ± 0.7
BEDFE+Full-space(S_w)	95.5 ± 1.12	98.55 ± 0.46

images for test. For the methods Eigenfaces, Fisherfaces, Null-space LDA, and EP, no validation set is required. Thus we combined the training images and validation images to form the training set for them. The case was a little bit different for the experiments with BEDFE, where the division of samples is on the class level (each subject is a class). Specifically, we first randomly chose five (seven) images from each subject to compose the test set of the ORL (AR) database. Among the remaining images, a subset of classes was randomly chosen. The samples belonging to these classes were used for training whereas those belonging to the other classes composed the validation set. From each class in the validation set, some samples were randomly chosen to calculate the prototype of the class, and the remaining ones were used for evaluation. On the ORL database, two images of each validation class were randomly chosen for class prototype calculation, and the other three images of the class were used to evaluate the training performance. On the AR database, three images were randomly chosen from each validation class for computing the class prototype, and the other four images of the class were used for training performance evaluation. The last three columns in Table 1 summarize these settings.

Totally, we created 10 evaluation sets from each of the two databases and ran algorithms over them one by one. We will use the mean and standard deviation of recognition accuracy over the 10 evaluation sets to evaluate the performance of different methods. When applying EP to the ORL databases, the dimensionality of the data should be reduced in advance due to the high space complexity of EP. We reduced the data to a dimension of the number of training samples minus one using WPCA (note that the role of WPCA here is different from that in the proposed EDFE algorithm). To evaluate the performance of Eigenfaces on the databases, we tested all possible dimensions of PCA-transformed subspace (between 1 and $N-1$) and found out the one with the best classification accuracy. As for Fisherfaces, we set the dimension of PCA-transformed subspace to the rank of S_t and tried all possible dimensions of LDA-transformed subspace (between 1 and $L-1$).

4.2. Investigation on Different Subspaces. Three subspaces, null(S_w), range(S_w), and range(S_b), are thought to contain rich discriminative information within data [16, 17]. As mentioned above, the algorithms proposed in this paper provide a method to constrain the search in a specific subspace. Hence, we can restrict the algorithms to search for a solution within that subspace. Here we report our experimental results in investigating the above three subspaces using the EDFE and BEDFE algorithms on the ORL and AR databases.

Table 2 shows the average recognition accuracy of EDFE in the three different subspaces on the ten evaluation sets of ORL and AR databases. The presented classification accuracies are the best ones among those obtained using different weights. On all the ten evaluation sets of both ORL and AR databases, null(S_w) gives the best results, which are significant better than the other two subspaces. On the other hand, there is no big difference between the performance of range(S_w) and range(S_b). This is not surprising because in the null space of S_w, if exists, samples in the same class will be condensed to one point. Then if a projection basis in it can be found to make the samples of different classes separable from each other, the classification performance on these samples will be surely the best. However, for new samples unseen in the training set, the classification accuracy on them depends on the accuracy of the estimation of S_w. Another problem with null(S_w) is that its dimensionality is bounded by the minimum of the dimensionality of the data and the difference between the number of samples and the number of classes. Consequently, as the number of training samples increases, this null space could become too small to contain sufficient discriminant information. In this case, we propose to incorporate the bagging technique to the EDFE algorithm to enhance its performance. The results of BEDFE are given in Table 3, from which similar conclusion can be drawn.

4.3. Investigation on Dimensionality of Feature Subspaces. In order to show the importance of carefully choosing the dimensionality of feature subspaces, we calculated the average recognition accuracy of Eigenfaces and Fisherfaces on the ten evaluation sets taken from the ORL and AR face databases when different numbers of features were chosen for the feature subspaces. The possible dimension of the feature subspace obtained by Eigenfaces on the ORL evaluation sets is between 1 and 199 (i.e., the number of samples minus one), whereas that on the AR evaluation sets is between 1 and 839. As for Fisherfaces, we set the dimension of PCA-reduced feature subspace to 720 (i.e., the number of samples minus the number of classes) and tested all the possible dimension

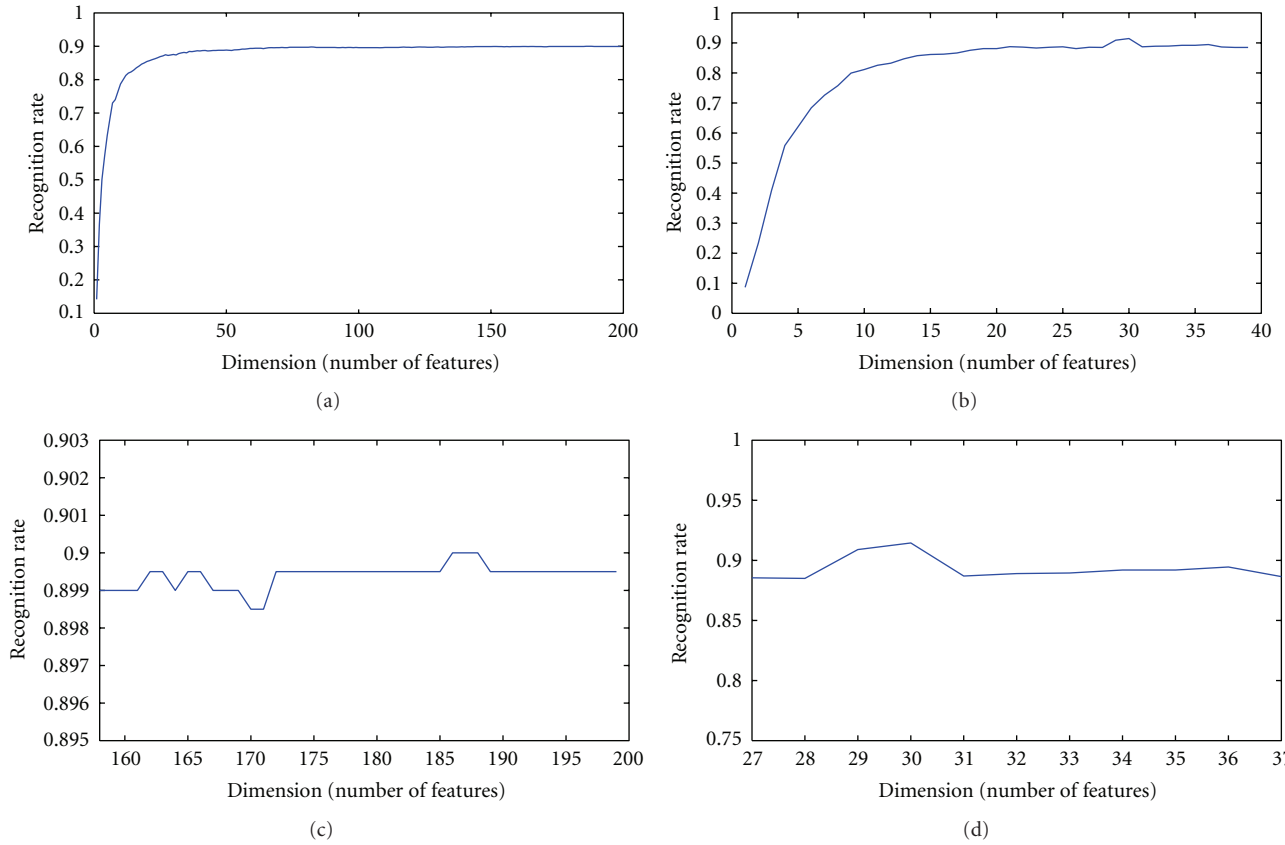


Figure 4: The curves of the average recognition accuracy of (a) Eigenfaces and (b) Fisherfaces on the ORL face database versus the number of features or the dimension of feature subspaces. (c) and (d) are the corresponding enlarged last parts of the curves.

of LDA-reduced feature subspace from 1 to 119 (i.e., the number of classes minus one). According to the experimental results, the overall trend of recognition accuracy is increasing as the number of features (i.e., the dimension of feature subspace) increases. However, the best accuracy is often obtained not at the largest possible dimension (i.e., the number of samples minus one in case of Eigenfaces and the number of classes minus one in case of Fisherfaces). Figures 4 and 5 show the curves of the average recognition accuracy of Eigenfaces and Fisherfaces on ORL and AR face databases versus the dimension of feature subspaces (to clearly show that the best accuracy is achieved not necessarily at the largest possible dimension, we also display the last part of the curves in an enlarged view). From these results, we can see that the dimension at which the best recognition accuracy is achieved varies with respect to the datasets. Therefore, using a systematic method like the ones proposed in this paper to automatically determine the dimension of feature subspaces is very helpful to a subspace-based recognition system.

4.4. Performance Comparison. Finally, we compared the proposed algorithms with some state-of-the-art methods in literature, including Eigenfaces [9], Fisherfaces [10], Null-space LDA [16], EP [18], and EDA+Full-space LDA [32]. Considering that both null(S_w) and range(S_w) have useful discriminative information, we ran our proposed EDFE and BEDFE methods in both null(S_w) and range(S_w) and then employed the same fusion method used by [32] to fuse the results obtained in these two subspaces. We called them EDFE+Full-space(S_w) and BEDFE+Full-space(S_w). We implemented these methods by using Matlab and evaluated their performance on the ten evaluation sets of ORL and AR face databases. But as for the EP method, it is too computationally complex to be applicable (N/A) on the AR face database (in Matlab an error of 'out of memory' will be reported to the EP method). We calculated the mean and standard deviation of the recognition rates for all the methods. The results are listed in Table 4 (the results of Eigenfaces and Fisherfaces are according to the best results obtained in the last subsection).

It can be seen from the results that the proposed EDFE and BEDFE methods overwhelm their counterpart methods in the average recognition accuracy. Moreover, by using the bagging technique, the BEDFE method performs much more stable than EDFE, and it has the smallest deviation of recognition accuracy among all the methods. A possible reason for such improvement on the stability is that by using smaller training sets and multiple feature subspace fusion, the outlier and over-fitting problems of conventional machine learning and pattern recognition

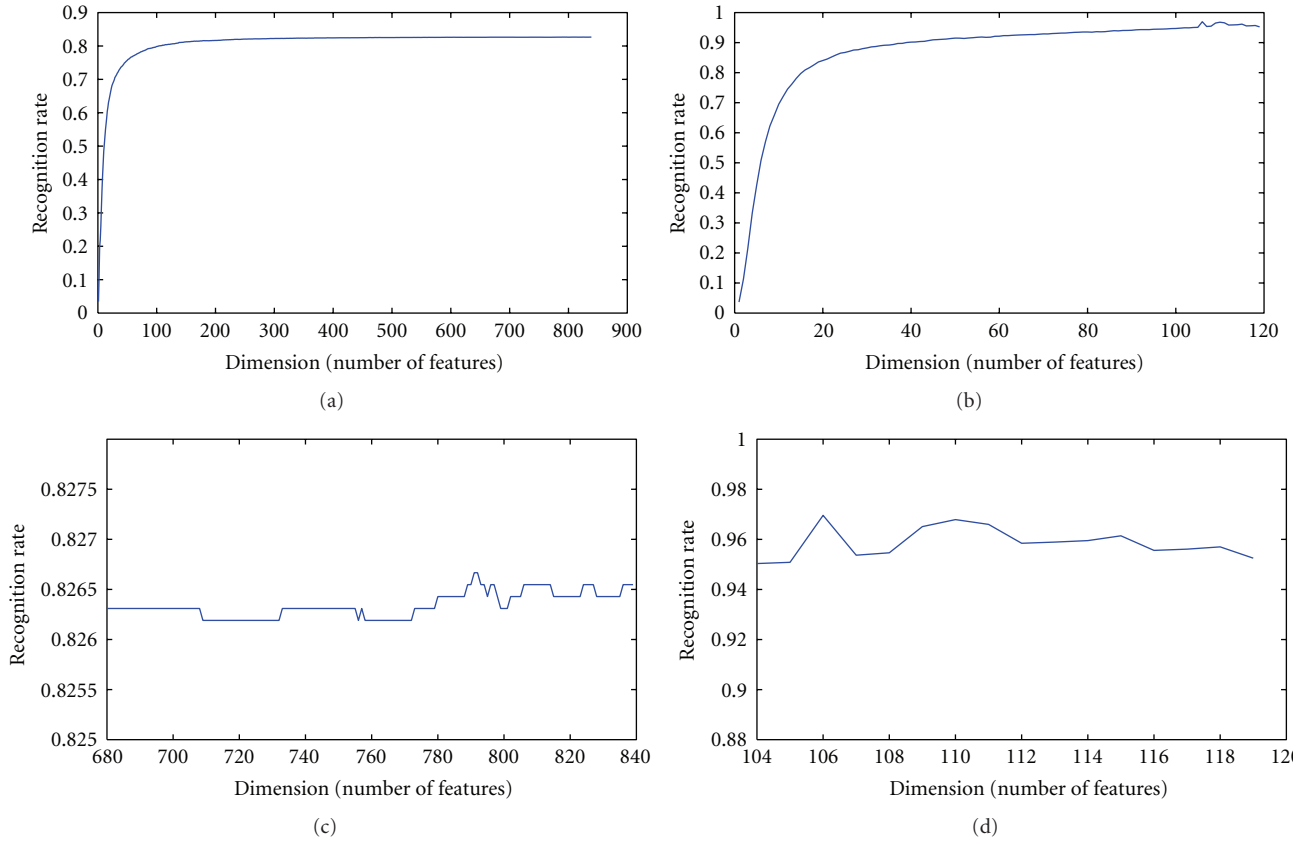


Figure 5: The curves of the average recognition accuracy of (a) Eigenfaces and (b) Fisherfaces on the AR face database versus the number of features or the dimension of feature subspaces. (c) and (d) are the corresponding enlarged last parts of the curves.

systems could be alleviated. Moreover, the improvement on recognition accuracy made by the proposed EDFE and BEDFE compared with the other methods could be due to their better generalization ability. In Eigenfaces, Fisherfaces, Null-space LDA, and EDA+Full-space LDA, the projection basis used for dimension reduction is directly calculated from certain covariance or scatter matrix of the training data. Instead, the methods proposed in this paper begin the search of optimal projection basis from these directly calculated ones and iteratively approach the best one via the linear combination of them. The linear combination not only ensures that the resulting projection basis still lies in the feature subspace but also enhances the generalization ability of the obtained projection basis by adjusting them according to the recognition accuracy on some validation data.

5. Discussion

In the proposed EDFE and BEDFE algorithms, we take the classification accuracy term as a part of the fitness function of the GA. It is then naturally optimized as the GA population evolves. Unlike existing evolutionary computation-based feature extraction methods like EP [18], we define this term on a randomly chosen validation sample set, but not the training set. Since the validation set's role is to simulate new test samples, the performance of the resulting feature subspace is supposed to be more reliable. We also set up a Fisher criterion-like term as another part of the GA's fitness function and optimize it in an iterative way, avoiding the matrix inverse operation required by the conventional LDA method. As a result, the proposed algorithms could alleviate the small sample size (SSS) problem of LDA.

Current PCA- and LDA-based subspace methods such as Eigenfaces and Fisherfaces require setting the dimensionality for the feature subspace in advance. They fail to provide systematic way to automatically determine the dimensionality from the classification viewpoint. Since the optimal dimensionality of feature subspace in terms of recognition rates will vary across datasets, it is desired to select automatically the optimal dimensionality for specific datasets, instead of using a predefined one. The proposed EDFE and BEDFE algorithms provide such a way by employing the stochastic optimization scheme of GA.

Some other GA-based feature selection/extraction methods have been also proposed in literature. Although these GA-based feature selection methods, such as EDA+Full-space LDA [32], GA-PCA and GA-Fisher [26], have the advantage in lower space and time requirement, they are limited in the ability of searching discriminative features. On the other hand, those GA-based feature extraction methods

have high space complexity and are thus not applicable to high dimensional and large scale datasets. For example, in EP [18], an individual has $(5n^2 - 4n)$ bits. In the recently proposed EDA algorithm [37], the individual has to encode $(n \times m)$ weights, which are between -0.5 and 0.5 (here, n is the dimension of the original data space, and m is the dimension of the feature subspace). On the contrary, the individual in the proposed algorithms has only $(12n - 1)$ bits (note that in face recognition applications, m is usually much larger than 12 and a number of bits have to be used to represent a decimal weight used by EDA). As a result, the space complexity is significantly reduced in our proposed methods. After incorporating the bagging technique with the proposed EDFE algorithm, it becomes more stable by eliminating possible outliers and fusing different feature subspaces, and hence more suitable for high-dimensional and large scale datasets.

Another problem with existing GA-based feature extraction methods lies in their blind search strategy. Using the linear combination and orthogonal complement techniques, the EDFE successfully provides a way to impose constraints on the search space of GA. This also enables EDFE to effectively make use of the heuristic information of the discriminative feature subspace to improve its search efficiency and classification performance. In addition, the proposed algorithms make it possible to investigate the discriminative ability of different feature subspaces.

6. Conclusions

In this paper, we proposed an evolutionary approach to extracting discriminative features, namely, evolutionary discriminant feature extraction (EDFE) as well as its bagging version (BEDFE). The basic idea underlying the EDFE algorithm is to use the genetic algorithm (GA) to search for an optimal discriminative projection basis in a constrained subspace with the goal of making the data in different clusters much easier to be separated. The primary contribution of this paper includes (1) reducing the space complexity of GA-based feature extraction algorithms; (2) enhancing the search efficiency, stability, as well as recognition accuracy; (3) providing an effective way to investigate different feature subspaces. Experiments on the ORL and AR databases have been performed to validate the proposed methods.

There are still some issues worthy further study on the proposed approach. Firstly, it is a supervised linear feature extraction method. Therefore, how to extend it to nonlinear cases deserves further study. Secondly, the latest progress in the research on GA, for example, how to set up an initial population and how to choose proper GA parameters, could give us some useful hints on further improving the proposed methods. Thirdly, more promising results could be obtained by exploring other criteria to evaluate the feature subspaces and incorporating them into the evolutionary approach, for instance, those of recently proposed manifold learning algorithms [40, 41]. Finally, it could be very interesting to investigate the discriminability of other subspaces using the proposed EDFE and BEDFE algorithms.

Appendix

Proof of Theorem 1

Proof. First it can be easily proved that for all $i \in \{1, 2, \ldots, m-1\}$, $\beta_i \in A$. Since $\beta_i = P\hat{\beta}_i = \sum_{j=1}^{m} \beta_{ij}\alpha_j$, β_i can be represented by a basis of A. Thus $\beta_i \in A$.

Secondly, let us prove $A = U \oplus V$, where $U = \text{span}\{\beta_1, \beta_2, \ldots, \beta_{m-1}\}$. This is to prove that $\{\beta_1, \beta_2, \ldots, \beta_{m-1}, v\}$ is a linear independent bundle. Since $\beta_i = P\hat{\beta}_i$ and $P^T P = E_m$, which is an m-dimensional identity matrix, we have $\beta_i^T \beta_j = \hat{\beta}_i^T \hat{\beta}_j$. However, $\{\hat{\beta}_i \mid i = 1, 2, \ldots, m-1\}$ is an identity orthogonal basis. Thus, $\beta_1, \beta_2, \ldots, \beta_{m-1}$ are orthogonal to each other.

Furthermore, for all $i \in \{1, 2, \ldots, m-1\}$, $\beta_i^T v = (P\hat{\beta}_i)^T v = \hat{\beta}_i^T P^T v$. Because $\{\hat{\beta}_i \mid i = 1, 2, \ldots, m-1\}$ is an identity orthogonal basis of the orthogonal complement space of $\hat{V} = \text{span}\{P^T v\}$ in R^m, $\hat{\beta}_i^T (P^T v)$ should be zero. Therefore, $\beta_i^T v = 0$, that is, β_i is also orthogonal to v.

To sum up, we get that $\{\beta_1, \beta_2, \ldots, \beta_{m-1}, v\}$ is a linear independent bundle containing m orthogonal vectors in the m dimensional space A. Thus $A = U \oplus V$. □

Acknowledgment

This work is partially supported by the Hong Kong RGC General Research Fund (PolyU 5351/08E).

References

[1] D. Zhang and X. Y. Jing, *Biometric Image Discrimination Technologies*, Idea Group, 2006.

[2] P. J. Phillips, P. J. Flynn, T. Scruggs, et al., "Overview of the face recognition grand challenge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 947–954, June 2005.

[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[4] N. V. Chawla and K. W. Bowyer, "Actively exploring creation of face space(s) for improved face recognition," in *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI '07)*, vol. 1, pp. 809–814, Vancouver, Canada, July 2007.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, Calif, USA, 2nd edition, 1990.

[6] N. A. Schmid and J. A. O'Sullivan, "Thresholding method for dimensionality reduction in recognition systems," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2903–2920, 2001.

[7] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.

[8] E. Oja, *Subspace Methods of Pattern Recognition*, John Wiley & Sons, New York, NY, USA, 1983.

[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71–86, 1991.

[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[11] K. Messer, J. Kittler, M. Sadeghi, et al., "Face authentication test on the BANCA database," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, pp. 523–532, 2004.

[12] B. Moghaddam, A. Pentland, et al., "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[13] B. Moghaddam, T. Jebara, A. Pentland, et al., "Bayesian face recognition," *Pattern Recognition*, vol. 33, no. 11, pp. 1771–1782, 2000.

[14] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1222–1228, 2004.

[15] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.

[16] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "New LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[17] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[18] C. J. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570–582, 2000.

[19] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, no. 5, pp. 335–347, 1989.

[20] H. Vafaie and K. De Jong, "Genetic algorithms as a tool for restructuring feature space representations," in *Proceedings of the 7th International Conference on Tools with Artificial Intelligence*, pp. 8–11, 1995.

[21] H. Vafaie and K. De Jong, "Feature space transformation using genetic algorithms," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 2, pp. 57–65, 1998.

[22] M. G. Smith and L. Bull, "Feature construction and selection using genetic programming and a genetic algorithm," in *Proceedings of the 6th European Conference on Genetic Programming (EuroGP '03)*, C. Ryan, et al., Ed., vol. 2610 of *Lecture Notes in Computer Science*, pp. 229–237, Springer, Berlin, Germany, August 2003.

[23] M. Pei, et al., "Genetic algorithms for classification and feature extraction," in *Proceedings of the Annual Meeting of the Classification Society of North America (CSNA '95)*, pp. 22–25, Denver, Colo, USA, June 1995.

[24] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 2, pp. 164–171, 2000.

[25] Q. Zhao and H. Lu, "GA-driven LDA in KPCA space for facial expression recognition," in *Proceedings of the 1st International Conference on Natural Computation (ICNC '05)*, vol. 3611 of *Lecture Notes in Computer Science*, pp. 28–36, Springer, Changsha, China, August 2005.

[26] W.-S. Zheng, J.-H. Lai, and P. C. Yuen, "GA-Fisher: a new LDA-based face recognition algorithm with selection of principal components," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 5, pp. 1065–1078, 2005.

[27] Q. Zhao, H. Lu, and D. Zhang, "A fast evolutionary pursuit algorithm based on linearly combining vectors," *Pattern Recognition*, vol. 39, no. 2, pp. 310–312, 2006.

[28] D. Dumitrescu, *Evolutionary Computation*, CRC Press, 2000.

[29] Q. Zhao, D. Zhang, and H. Lu, "A direct evolutionary feature extraction algorithm for classifying high dimensional data," in *Proceedings of the National Conference on Artificial Intelligence (AAAI '06)*, vol. 1, pp. 561–566, Boston, Mass, USA, 2006.

[30] D. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, Adison-Wesley, 1989.

[31] Q. Zhao, H. Lu, and D. Zhang, "Parsimonious feature extraction based on genetic algorithms and support vector machines," in *Lecture Notes in Computer Science*, vol. 3971, pp. 1387–1393, Springer, Berlin, Germany, May 2006.

[32] X. Li, B. Li, H. Chen, X. Wang, and Z. Zhuang, "Full-space LDA with evolutionary selection for face recognition," in *Proceedings of the International Conference on Computational Intelligence and Security (ICCIAS '06)*, vol. 1, pp. 696–701, November 2006.

[33] X. Wang and X. Tang, "Random sampling for subspace face recognition," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 91–104, 2006.

[34] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[35] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[36] J. Kittler and F. M. Alkoot, "Sum versus vote fusion in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 110–115, 2003.

[37] A. Sierra and A. Echeverria, "Evolutionary discriminant analysis," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 81–92, 2006.

[38] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138–142, 1994.

[39] A. M. Martinez and R. Benavente, "The AR face database," Tech. Rep. 24, 1998.

[40] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[41] X. He and P. Niyogi, "Locality preserving projections," Tech. Rep., Department of Computer Science, University of Chicago, Chicago, Ill, USA, 2003.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 943602, 6 pages
doi:10.1155/2009/943602

Research Article

Comparison of Spectral-Only and Spectral/Spatial Face Recognition for Personal Identity Verification

Zhihong Pan,[1] Glenn Healey,[2] and Bruce Tromberg[3]

[1] *Galileo Group Inc., 100 Rialto Place Suite 737, Melbourne, FL 32901, USA*
[2] *Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697, USA*
[3] *Beckman Laser Institute, 1002 East Health Sciences Road, Irvine, CA 92612, USA*

Correspondence should be addressed to Zhihong Pan, zpan@galileo-gp.com

Received 29 September 2008; Revised 22 February 2009; Accepted 8 April 2009

Recommended by Kevin Bowyer

Face recognition based on spatial features has been widely used for personal identity verification for security-related applications. Recently, near-infrared spectral reflectance properties of local facial regions have been shown to be sufficient discriminants for accurate face recognition. In this paper, we compare the performance of the spectral method with face recognition using the eigenface method on single-band images extracted from the same hyperspectral image set. We also consider methods that use multiple original and PCA-transformed bands. Lastly, an innovative spectral eigenface method which uses both spatial and spectral features is proposed to improve the quality of the spectral features and to reduce the expense of the computation. The algorithms are compared using a consistent framework.

1. Introduction

Automatic personal identity authentication is an important problem in security and surveillance applications, where physical or logical access to locations, documents, and services must be restricted to authorized persons. Passwords or personal identification numbers (PINs) are often assigned to individuals for authentication. However, the password or PIN is vulnerable to unauthorized exploitation and can be forgotten. Biometrics, on the other hand, use personal intrinsic characteristics which are harder to compromise and more convenient to use. Consequently, the use of biometrics has been gaining acceptance for various applications. Many different sensing modalities have been developed to verify personal identities. Fingerprints are a widely used biometric. Iris recognition is an emerging technique for personal identification which is an active area of research. There are also studies to use voice and gait as primary or auxiliary means to verify personal identities.

Face recognition has been studied for many years for human identification and personal identity authentication and is increasingly used for its convenience and noncontact measurements. Most modern face recognition systems are based on the geometric characteristics of human faces in an image [1–4]. Accurate verification and identification performance has been demonstrated for these algorithms based on mug shot type photographic databases of thousands of human subjects under controlled environments [5, 6]. Various 3D face models [7, 8] and illumination models [9, 10] have been studied for pose and illumination-invariant face recognition. In addition to methods based on gray-scale and color face images over the visible spectrum, thermal infrared face images [11, 12] and hyperspectral face images [13] have also been used for face recognition experiments. An evaluation of different face recognition algorithms using a common dataset has been of general interest. This approach provides a solid basis to draw conclusions on the performance of different methods. The Face Recognition Technology (FERET) program [5] and the Face Recognition Vendor Test (FRVT) [6] are two programs which provided independent government evaluations for various face recognition algorithms and commercially available face recognition systems.

Most biometric methods, including face recognition methods, are subject to possible false acceptance or rejection. Although biometric information is difficult to duplicate, these methods are not immune to forgery, or so-called spoofing. This is a concern for automatic personal identity

authentication since intruders can use artificial materials or objects to gain unauthorized access. There are reports showing that fingerprint sensor devices have been deceived by Gummi fingers in Japan [14] and fake latex fingerprints in Germany [15]. Face and iris recognition systems can also be compromised since they use external observables [16]. To counter this vulnerability, many biometric systems employ a liveness detection function to foil attempts at biometric forgery [17, 18]. To improve system accuracy, there is strong interest in research to combine multiple biometric characteristics for multimodal personal identity authentication [19, 20]. Since hyperspectral sensors capture spectral and spatial information they provide the potential for improved personal identity verification.

Methods that have been developed consider the use of representations for visible wavelength color images for face recognition [21, 22] as well as the combination of color and 3D information [23]. In this work, we examine the use of combined spectral/spatial information for face recognition over the near-infrared (NIR) spectral range. We show that the use of spatial information can be used to improve on the performance of spectral-only methods [13]. We also use a large NIR hyperspectral dataset to show that the choice of spectral band over the NIR does not have a significant effect on the performance of single-band eigenface methods. On the other hand, we show that band selection does have a significant effect on the performance of multiband methods. In this paper we develop a new representation called the spectral-face which preserves both high-spectral and high-spatial resolution. We show that the spectral eigenface representation outperforms single-band eigenface methods and has performance that is comparable to multiband eigenface methods but at a lower computational cost.

2. Face Recognition in Single-Band Images

A hyperspectral image provides spectral information, normally in radiance or reflectance, at each pixel. Thus, there is a vector of values for each pixel corresponding to different wavelengths within the sensor spectral range. The reflectance spectrum of a material remains constant in different images while different materials exhibit distinctive reflectance properties due to different absorbing and scattering characteristics as a function of wavelength. In the spatial domain, there are several gray-scale images that represent the hyperspectral imager responses of all pixels for a single spectral band. In a previous study [24], seven hyperspectral face images were collected for each of 200 human subjects. These images have a spatial resolution of 468×494 and 31 bands with band centers separated by 0.01 μm over the near-infrared (0.7 μm–1.0 μm). Figure 1 shows calibrated hyperspectral face images of two subjects at seven selected bands which are separated by 0.06 μm over 0.7 μm–1.0 μm. We see that the ratios of pixel values on skin or hair between different bands are dissimilar for the two subjects. That is, they have unique hyperspectral signatures for each tissue type. Based on these spectral signatures, a Mahalanobis distance-based method was applied for face recognition tests and accurate face recognition rates were achieved. However, the performance was not compared with classic face recognition methods using the same dataset.

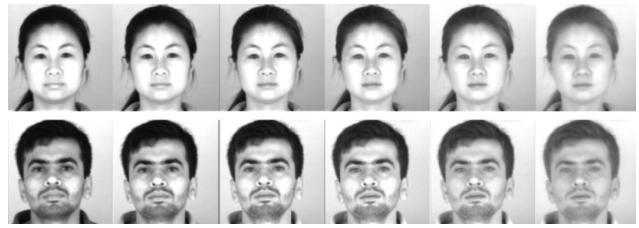


Figure 1: Selected single-band images of two subjects.

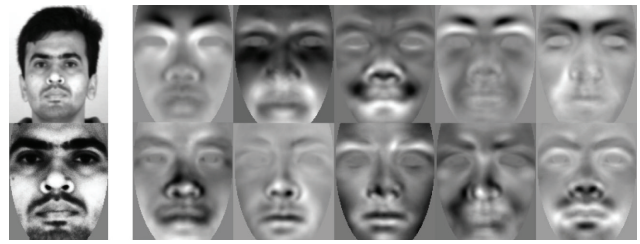


Figure 2: Example of eigenfaces in one single-band.

The CSU Face Identification Evaluation System [25] provides a standard set of well-known algorithms and established experimental protocols for evaluating face recognition algorithms. We selected the Principal Components Analysis (PCA) Eigenfaces [26] algorithm and used cumulative match scores as in the FERET study [5] for performance comparisons. To prepare for the face recognition tests, a gray-scale image was extracted for each of the 31 bands from a hyperspectral image. The coordinates of both eyes were manually positioned before processing by the CSU evaluation programs. In the CSU evaluation system all images were transformed and normalized so that they have a fixed spatial resolution of 130×150 pixels and the eye coordinates are the same. Masks were used to void nonfacial features. Histogram equalization was also performed on all images before the face recognition tests were conducted. For each of the 200 human subjects, there are three front-view images with the first two (fg and fa) having neutral expression and the other (fb) having a smile. All 600 images were used to generate the eigenfaces. Figure 2 shows one single-band image before and after the normalization, and the first 10 eigenfaces for the dataset. The number of eigenfaces used for face recognition was determined by selecting the set of most significant eigenfaces which account for 90% of the total energy.

Given the wth band of hyperspectral images U and V, the Mahalanobis Cosine distance [27] is used to measure the similarity of the two images. Let $u_{w,i}$ be the projection of the wth band of U onto the ith eigenface and let $\sigma_{w,i}$ be the standard deviation of the projections from all of the wth band images onto the ith eigenface. The Mahalanobis projection of U_w is $M_w = (m_{w,1}, m_{w,2}, \ldots, m_{w,I})$ where $m_{w,i} = u_{w,i}/\sigma_{w,i}$. Let N_w be the similarly computed Mahalanobis

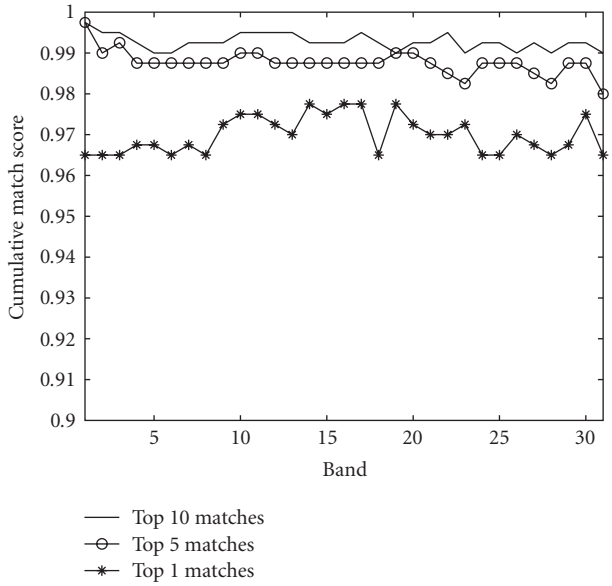


Figure 3: Cumulative match scores of single-band images at different wavelengths.

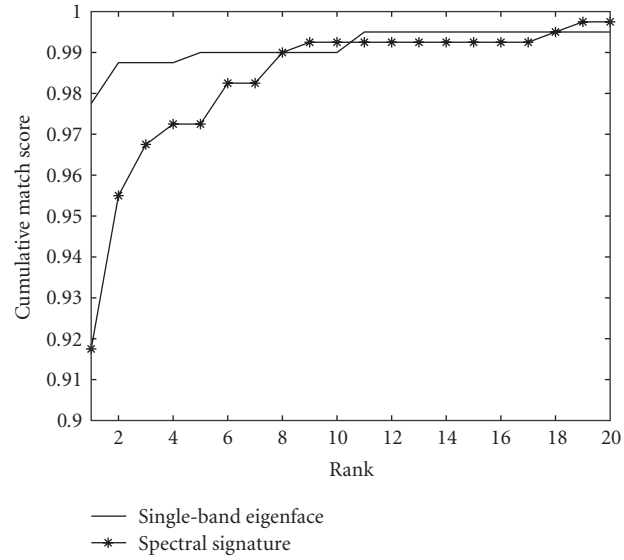


Figure 4: Cumulative match scores of spectral signature method and the best single-band eigenface method.

projection of V_w. The Mahalanobis Cosine distance between U and V for the wth band is defined by

$$D_{U,V}(w) = -\frac{M_w \cdot N_w}{|M_w||N_w|}, \tag{1}$$

which is the negative of the cosine between the two vectors. For the 200 subjects, the fg images were grouped in the gallery set and the fa and fb images were used as *probes* [5]. The experiments follow the *closed universe* model where the subject in every image in the probe set is included in the gallery. For each probe image, the Mahalanobis Cosine distance between the probe and all gallery images is computed. If the correct match is included in the group of gallery images with the N smallest distances, we say that the probe is correctly matched in the top N. The cumulative match score for a given N is defined as the fraction of correct matches in the top N from all probes. The cumulative match score for $N = 1$ is called the recognition rate. Figure 3 plots the cumulative match scores for $N = 1, 5$, and 10 respectively. Band 1 refers to the image acquired at 700 nm and band 31 refers to the image acquired at 1000 nm. We see that all bands provide high recognition rates, with more than 96% of the probes correctly identified for $N = 1$ and over 99% for $N = 10$. It is important to consider the statistical significance of the results. For this purpose, we model the fraction of the probes that are correctly identified by a binomial distribution with a mean given by the measured identification rate p. The variance σ^2 of p is given by $400p(1-p)$ where 400 is the number of probes [28]. For an identification rate of 0.97 we have $\sigma = 3.4$ which corresponds to a standard deviation in the identification rate of 0.009 and for an identification rate of 0.99 we have $\sigma = 1.99$ which corresponds to a standard deviation in the identification rate of 0.005. Thus, for each of the three curves plotted in Figure 3 the variation in performance across bands is not statistically significant. Figure 4 compares the cumulative match scores using the spectral signature method [13] and the single-band eigenface method using the most effective band. We see that the spectral signature method performs well but somewhat worse than the best single-band method for matches with N less than 8. For $N = 1$, a recognition rate of 0.92 corresponds to a standard deviation in the recognition rate of 0.014 which indicates that the difference between the two methods in Figure 4 is statistically significant. The advantage of the spectral methods is pose invariance which was discussed in a previous work [13] but which is not considered in this paper.

3. Face Recognition in Multiband Images

We have shown that both spatial and spectral features in hyperspectral face images provide useful discriminants for recognition. Thus, we can consider the extent of performance improvements when both features are utilized. We define a distance between images U and V using

$$D_{U,V} = \sqrt{\sum_{w=1}^{W} (1 + D_{U,V}(w))^2}, \tag{2}$$

where the index w takes values over a group of W-selected bands that are not necessarily contiguous. Note that the additive 1 is to ensure a nonnegative value before the square.

Redundancy in a hyperspectral image can be reduced by a Principal Component Transformation (PCT) [29]. For a hyperspectral image $U = (U_1, U_2, \ldots, U_W)$, the PCT generates $U' = (U'_1, U'_2, \ldots, U'_W)$, where $U'_i = \sum_j \varepsilon_{ij} U_j$. The principal components $U'_1, U'_2, \ldots, U'_W$ are orthogonal to each other and sorted in order of decreasing modeled variance. Figure 5 shows a single-band image at 700 nm and the first five principal components that are extracted from

Figure 5: Five principal band images of one subject after PCT.

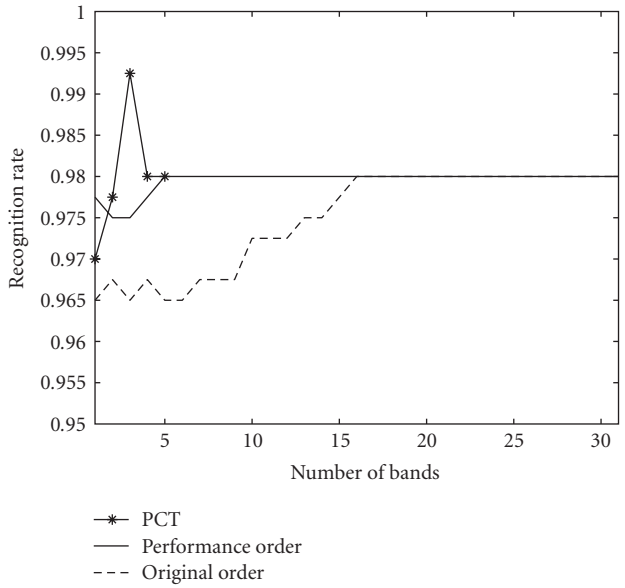


Figure 6: Recognition rate of multiband eigenface methods.

the corresponding hyperspectral image. We see that the first principal component image resembles the single-band image while the second and third component images highlight features of the lips and eyes. We also see that there are few visible features remaining in the fourth and fifth principal components.

Figure 6 plots the recognition rates for different multiband eigenface methods. First we selected the bands in order of increasing center wavelength and performed eigenface recognition tests for the first one band, two bands and up to 31 bands, respectively. We also sorted all 31 bands in descending order of recognition rate and performed the same procedure for the face recognition tests. From Figure 6 we see that both methods reach a maximum recognition rate of 98% when using multiple bands. However, when the number of bands is less than 16, the multiband method performs better if the bands are sorted in advance from the highest recognition rate to the lowest. We also used the leading principal components for multiband recognition. We see in Figure 6 that over 99% of the probes were correctly recognized when using the first three principal bands. Increasing the number of principal bands beyond 3 causes performance degradation. The original-order algorithm in Figure 6 achieves a recognition rate of approximately 0.965 for less than ten bands which corresponds to a standard deviation in recognition rate of 0.009. Thus, the performance difference between this method and the PCT-based method is significant between 3 and 9 bands. Note that the PCT was performed on each hyperspectral image individually with

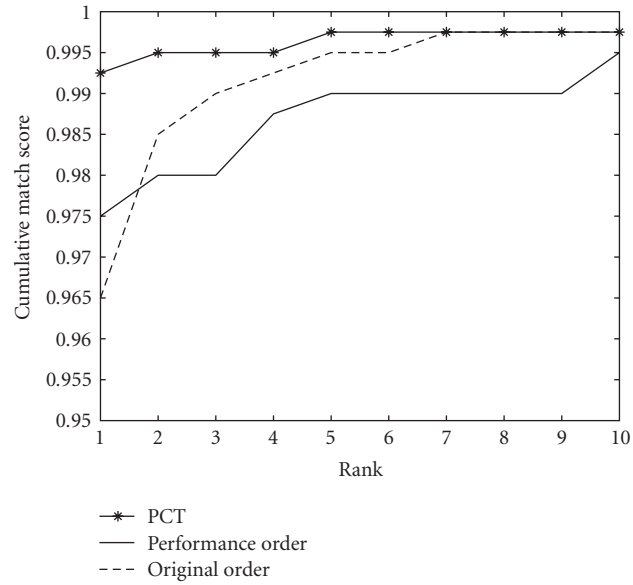


Figure 7: Cumulative match scores of multiband eigenface methods.

different sets of ε_{ij}. The PCT can also be implemented using the same coefficients for faster computation.

Figure 7 also compares the recognition performance of the three multiband methods discussed in the previous paragraph where each algorithm uses only the first three bands. It is interesting that sorting the bands according to performance improves the recognition rate for $N = 1$ but worsens the performance somewhat for larger values of N. In either case, the multiband method based on the PCT has the best performance for $N < 7$ and is equivalent to the original-order method for larger values of N.

4. Face Recognition Using Spectral Eigenfaces

We showed in Section 3 that multiband eigenface methods can improve face recognition rates. In these algorithms, the multiple bands are processed independently. A more general approach is to consider the full spectral/spatial structure of the data. One way to do this is to apply the eigenface method to large composite images that are generated by concatenating the 31 single-band images. This approach, however, will significantly increase the computational cost of the process. An alternative is to subsample each band of the hyperspectral image before concatenation into the large composite image. For example, Figure 8 shows a 31-band image after subsampling so that the total number of pixels is equivalent to the number of pixels in a 130×150 pixel single-band image. We see that significant spatial detail is lost due to the subsampling.

A new representation, called spectral-face, is proposed to preserve both spectral and spatial properties. The spectral-face has the same spatial resolution as a single-band image so the spatial features are largely preserved. In the spectral domain, the pixel values in the spectral-face are extracted sequentially from band 1 to band 31 then from band 1 again. For example, the value of pixel i in spectral-face equals

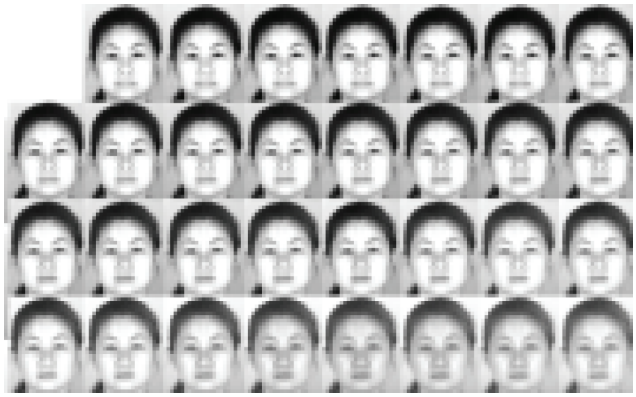


Figure 8: A sample image composed from 31 bands with low-spatial resolution.

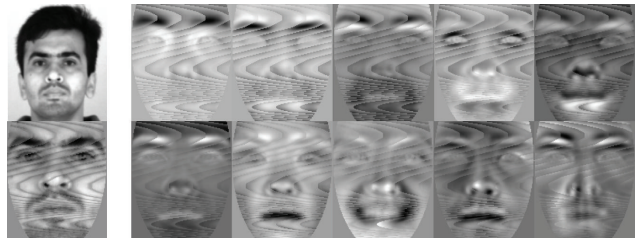


Figure 9: One sample spectral-face and the first 10 spectral eigenfaces.

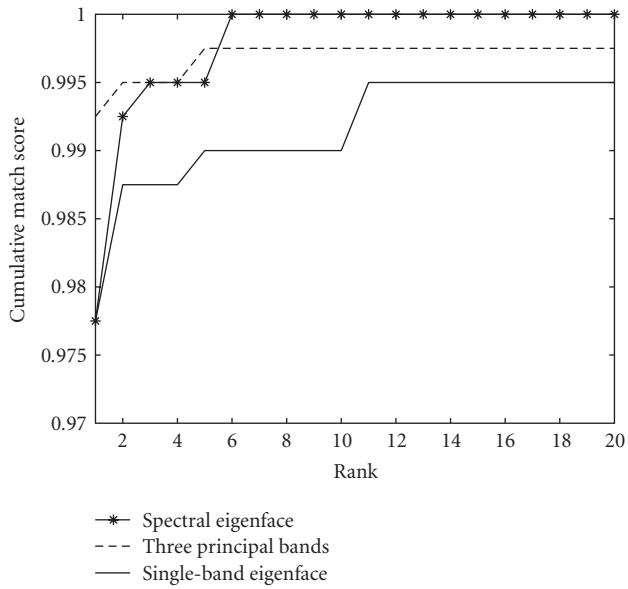


Figure 10: Comparison of spectral eigenface method with single-band and multiband methods.

the value of pixel i in band w where w is the remainder of i divided by 31. Figure 9 shows an original single-band image together with the normalized spectral-face image in the left column. Spectral-face has improved spatial detail as compared with Figure 8. The pattern on the face in Figure 9 demonstrates the variation in the spectral domain. With the spectral-face images, the same eigenface technique is applied for face recognition. The first 10 spectral eigenfaces are shown on the right side of Figure 9. It is interesting to observe that the eighth spectral eigenface highlights the teeth feature in smiling faces.

The spectral eigenface method was applied to the same dataset as the single-band and multiband methods. The cumulative match scores for $N = 1$ to 20 are shown in Figure 10. The best of the single-band methods, which corresponds to band 19 (880 nm), is included for performance comparison with the spectral eigenface method. We see that the spectral eigenface method has better performance for all ranks. The best of the multiband methods, which combines the first three principal bands, is also considered. The multiband method performs better than the spectral eigenface method for small values of the rank, but performs worse for larger values of the rank. For this case, an identification rate of 0.99 corresponds to a standard deviation in identification rate of 0.005. Thus, the two multiple-band methods have a statistically significant advantage over the single-band eigenface method for ranks between 3 and 10. Note that the multiple principal band method requires more computation than the spectral eigenface method.

5. Conclusion

Multimodal personal identity authentication systems have gained popularity. Hyperspectral imaging systems capture both spectral and spatial information. The previous work [24] has shown that spectral signatures are powerful discriminants for face recognition in hyperspectral images. In this work, various methods that utilize spectral and/or spatial features were evaluated using a hyperspectral face image dataset. The single-band eigenface method uses spatial features exclusively and performed better than the pure spectral method. However, the computational requirements increase significantly for eigenface generation and projection. The recognition rate was further improved by using multiband eigenface methods which require more computation. The best performance was achieved with the highest computational complexity by using principal component bands. The spectral eigenface method transforms a multiband hyperspectral image to a spectral-face image which samples from all of the bands while preserving spatial resolution. We showed that this method performs as well as the PCT-based multiband method but with a much lower computational requirement.

Acknowledgments

This work was conducted when the author was with the Computer Vision Laboratory at the University of California, Irvine, USA. This work has been supported by the DARPA Human Identification at a Distance Program through AFOSR Grant F49620-01-1-0058. This work has also been supported by the Laser Microbeam and Medical Program (LAMMP) and NIH Grant RR01192. The data was acquired at the Beckman Laser Institute on the UC Irvine

campus. The authors would like to thank J. Stuart Nelson and Montana Compton for their valuable assistance in the process of IRB approval and human subject recruitment.

References

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[2] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[3] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[4] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[5] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[6] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002: overview and summary," Tech. Rep., Defense Advanced Research Projects Agency, Arlington, Va, USA, March 2003.

[7] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[8] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, 2005.

[9] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, 1997.

[10] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[11] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 72–114, 2003.

[12] J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener, "Comparison of visible and infra-red imagery for face recognition," in *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (AFGR '96)*, pp. 182–187, Killington, Vt, USA, October 1996.

[13] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.

[14] J. Leyden, "Gummi bears defeat fingerprint sensors," *The Register*, May 2002.

[15] A. Harrison, "Hackers claim new fingerprint biometric attack," *Security Focus*, August 2003.

[16] M. Lewis and P. Statham, "CESG biometric security capabilities programme: method, results and research challenges," in *Biometric Consortium Conference*, Crystal City, Va, USA, September 2004.

[17] J. Bigun, H. Fronthaler, and K. Kollreider, "Assuring liveness in biometric identity authentication by real-time face tracking," in *Proceedings of IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS '04)*, pp. 104–111, Venice, Italy, July 2004.

[18] T. Tan and L. Ma, "Iris recognition: recent progress and remaining challenges," in *Biometric Technology for Human Identification*, vol. 5404 of *Proceedings of SPIE*, pp. 183–194, Orlando, Fla, USA, April 2004.

[19] J. Kittler, J. Matas, K. Jonsson, and M. U. Ramos Sánchez, "Combining evidence in personal identity verification systems," *Pattern Recognition Letters*, vol. 18, no. 9, pp. 845–852, 1997.

[20] J. Kittler and K. Messer, "Fusion of multiple experts in multimodal biometric personal identity verification systems," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 3–12, Kauai, Hawaii, USA, December 2002.

[21] J. Yang, D. Zhang, Y. Xu, and J.-Y. Yang, "Recognize color face images using complex eigenfaces," in *Proceedings of International Conference on Advances in Biometrics (ICB '06)*, vol. 3832 of *Lecture Notes in Computer Science*, pp. 64–68, Hong Kong, January 2006.

[22] S. Yoo, R.-H. Park, and D.-G. Sim, "Investigation of color spaces for face recognition," in *Proceedings of IAPR Conference on Machine Vision Applications (MVA '07)*, pp. 106–109, Tokyo, Japan, May 2007.

[23] F. Tsalakanidou, D. Tzovaras, and M. G. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1427–1435, 2003.

[24] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. 334–339, Institute of Electrical and Electronics Engineers, Madison, Wis, USA, June 2003.

[25] D. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, "The CSU face identification evaluation system: its purpose, features and structure," in *Proceedings of the 3rd International Conference Computer Vision Systems (ICVS '03)*, vol. 2626 of *Lecture Notes in Computer Science*, pp. 304–313, Graz, Austria, April 2003.

[26] M. A. Turk and A. P. Pentland, "Face recogntion using eigenfaces," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '91)*, pp. 586–591, Maui, Hawaii, USA, June 1991.

[27] J. R. Beveridge, D. S. Bolme, M. Teixeira, and B. Draper, "The CSU face identification evaluation system user's guide: version 5.0," Tech. Rep., Computer Science Department, Colorado State University, Fort Collins, Colo, USA, May 2003.

[28] A. Papoulis, *Probability and Statistics*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1990.

[29] P. J. Ready and P. A. Wintz, "Information extraction, SNR improvement, and data compression in multispectral imagery," *IEEE Transactions on Communications*, vol. 21, no. 10, pp. 1123–1131, 1973.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 746481, 15 pages
doi:10.1155/2009/746481

Research Article

Talking-Face Identity Verification, Audiovisual Forgery, and Robustness Issues

Walid Karam,[1] Hervé Bredin,[2] Hanna Greige,[3] Gérard Chollet,[4] and Chafic Mokbel[1]

[1] *Computer Science Department, University of Balamand, 100 El-Koura, Lebanon*
[2] *SAMoVA Team, IRIT-UMR 5505, CNRS, 5505 Toulouse, France*
[3] *Mathematics Department, University of Balamand, 100 El-Koura, Lebanon*
[4] *TSI, Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris, France*

Correspondence should be addressed to Walid Karam, walid@balamand.edu.lb

Received 1 October 2008; Accepted 3 April 2009

Recommended by Kevin Bowyer

The robustness of a biometric identity verification (IV) system is best evaluated by monitoring its behavior under impostor attacks. Such attacks may include the transformation of one, many, or all of the biometric modalities. In this paper, we present the transformation of both speech and visual appearance of a speaker and evaluate its effects on the IV system. We propose *MixTrans*, a novel method for voice transformation. *MixTrans* is a mixture-structured bias voice transformation technique in the cepstral domain, which allows a transformed audio signal to be estimated and reconstructed in the temporal domain. We also propose a face transformation technique that allows a frontal face image of a client speaker to be animated. This technique employs principal warps to deform defined MPEG-4 facial feature points based on determined facial animation parameters (FAPs). The robustness of the IV system is evaluated under these attacks.

1. Introduction

With the emergence of smart phones and third and fourth generation mobile and communication devices, and the appearance of a "first generation" type of mobile PC/PDA/phones with biometric identity verification, there has been recently a greater attention to secure communication and to guarantee the robustness of embedded multimodal biometric systems. The robustness of such systems promises the viability of newer technologies that involve e-voice signatures, e-contracts that have legal values, and secure and trusted data transfer regardless of the underlying communication protocol. Realizing such technologies require reliable and error-free biometric identity verification systems.

Biometric identity verification (IV) systems are starting to appear on the market in various commercial applications. However, these systems are still operating with a certain measurable error rate that prevents them from being used in a full automatic mode and still require human intervention and further authentication. This is primarily due to the variability of the biometric traits of humans over time because of growth, aging, injury, appearance, physical state, and so forth. Impostors attempting to be authenticated by an IV system to gain access to privileged resources could take advantage of the non-zero error rate of the system by imitating, as closely as possible, the biometric features of a genuine client.

The purpose of this paper is threefold. (1) It evaluates the performance of IV systems by monitoring their behavior under impostor attacks. Such attacks may include the transformation of one, many, or all of the biometric modalities, such as face or voice. This paper provides a brief review of IV techniques and corresponding evaluations and focuses on a statistical approach (GMM). (2) It also introduces *MixTrans*, a novel mixture-structure bias voice transformation technique in the cepstral domain, which allows a transformed audio signal to be estimated and reconstructed in the temporal domain. (3) It proposes a face transformation technique that allows a 2D face image of the client to be animated. This technique employs principal warps to deform defined MPEG-4 facial feature points based

on determined facial animation parameters (FAPs). The BANCA database is used to test the effects of voice and face transformation on the IV system.

The rest of the paper is organized as follows. Section 2 introduces the performance evaluation, protocols, and the BANCA database. Section 3 is a discussion of audiovisual identity verification techniques based on Gaussian Mixture Models. Section 4 describes the imposture techniques used, including MixTrans, a novel voice transformation technique, and face transformation based on an MPEG-4 face animation with thin-plate spline warping. Section 5 discusses the experimental results on the BANCA audiovisual database. Section 6 wraps up with a conclusion.

2. Evaluation Protocols

Evaluation of audiovisual IV systems and the comparison of their performances require the creation of a reproducible evaluation framework. Several experimental databases have been set up for this purpose. These databases consist of a large collection of biometric samples in different scenarios and quality conditions. Such databases include BANCA [1], XM2VTS [2], BT-DAVID [3], BIOMET [4], and PDAtabase [5].

2.1. The BANCA Database. In this work, audiovisual verification experiments and imposture were primarily conducted on the BANCA Database [1]. BANCA is designed for testing multimodal identity verification systems. It consists of video and speech data for 52 subjects (26 males, 26 females) in four different European languages (English, French, Italian, and Spanish). Each language set and gender was divided into two independent groups of 13 subjects (denoted g1 and g2). Each subject recorded a total of 12 sessions, for a total of 208 recordings. Each session contains two recordings: a true client access and an informed impostor attack (the client proclaims in his own words to be someone else). Each subject was prompted to say 12 random number digits, his or her name, address, and date of birth.

The 12 sessions are divided into three different scenarios.

(i) *Scenario c (controlled)*. Uniform blue background behind the subject with a quiet environment (no background noise). The camera and microphone used are of good quality (sessions 1–4).

(ii) *Scenario d (degraded)*. Low quality camera and microphone in an "adverse" environment (sessions 5–8).

(iii) *Scenario a (adverse)*. Cafeteria-like atmosphere with activities in the background (people walking or talking behind the subject). The camera and microphone used are also of good quality (sessions 9–12).

BANCA has also a world model of 30 other subjects, 15 males and 15 females.

Figure 1 shows example images from the English database for two subjects in all three scenarios.

The BANCA evaluation protocol defines seven distinct training/test configurations, depending on the actual conditions corresponding to training and testing. These experimental configurations are Matched Controlled (MC), Matched Degraded (MD), Matched Adverse (MA), Unmatched Degraded (UD), Unmatched Adverse (UA), Pooled Test (P), and Grand Test (G) (Table 1).

The results reported in this work reflect experiments on the "*Pooled test*," also known as the "P" protocol, which is BANCA's most "difficult" evaluation protocol: world and client models are trained on session 1 only (controlled environment), while tests are performed in all different environments (Table 1).

2.2. Performance Evaluation. The evaluation of a biometric system performance and its robustness to imposture is measured by the rate of errors it makes during the recognition process. Typically, a recognition system is a "comparator" that compares the biometric features of a user with a given biometric reference and gives a "score of likelihood." A decision is then taken based on that score and an adjustable defined acceptance "threshold." Two types of error rates are traditionally used.

(i) *False Acceptance Rate (FAR)*. The *FAR* is the frequency that an impostor is accepted as a genuine client. The *FAR* for a certain enrolled person n is measured as

$$\mathrm{FAR}(n) = \frac{\text{Number of successful haox attempts against a person } n}{\text{Number of all haox attempts against a person } n}, \tag{1}$$

and for a population of N persons, $\mathrm{FAR} = (1/n)\sum_{n=1}^{N}\mathrm{FAR}(n)$.

(ii) *False Rejection Rate (FRR)*. The *FRR* is the frequency that a genuine client is rejected as an impostor:

$$\mathrm{FRR}(n) = \frac{\text{Number of rejected verification attempts a genuine person } n}{\text{Number of all verification attempts a genuine person } n}, \tag{2}$$

$$\mathrm{FRR} = \frac{1}{N}\sum_{n=1}^{N}\mathrm{FRR}(n).$$

Table 1: Summary of the 7 training/testing configurations of BANCA.

	Test Sessions	Train Sessions			
		1	5	9	1, 5, 9
Client	2–4	MC			
Impostor	1–4				
Client	6–8	UD	MD		
Impostor	5–8				
Client	10–12	UA		MA	
Impostor	9–12				
Client	2–4, 6–8, 10–12	**P**			G
Impostor	1–12				

To assess visually the performance of the authentication system, several curves are used: the Receiver Operating Characteristic (ROC) curve [6, 7], the Expected Performance Curve (EPC) [8], and the Detection error trade-off (DET) curve [9]. The ROC curve plots the *sensitivity* (fraction of true positives) of the binary classifier system versus *specificity* (fraction of false positives) as a function of the threshold. The closer the curve to 1 is, the better the performance of the system is.

While ROC curves use a biased measure of performance (EER), the EPC introduced in [8] provides an unbiased estimate of performance at various operating points.

The DET curve is a log-deviate scale graph of *FRR* versus *FAR* as the threshold changes. The *EER* value is normally reported on the DET curve: the closer *EER* to the origin is, the better the performance of the system is. The results reported in this work are in the form of DET curves.

3. Multimodal Identity Verification

3.1. Identification Versus Verification. Identity recognition can be divided into two major areas: authentication and Identification. Authentication, also referred to as verification, attempts to verify a person's identity based on a claim. On the other hand, identification attempts to find the identity of an unknown person in a set of a number of persons. Verification can be though of as being a one-to-one match where the person's biometric traits are matched against one template (or a template of a general "world model") whereas identification is a one-to-many match process where biometric traits are matched against many templates.

Identity verification is normally the target of applications that entail a secure access to a resource. It is managed with the client's knowledge and normally requires his/her cooperation. As an example, a person's access to a bank account at an automatic teller machine (ATM) may be asked to verify his fingerprint or look at a camera for face verification or speak into a microphone for voice authentication. Another example is the fingerprint readers of most modern laptop computers that allow access to the system only after fingerprint verification.

Person identification systems are more likely to operate covertly without the knowledge of the client. This can be used, for example, to identify speakers in a recorded group conversation, or a criminal's fingerprint or voice is cross checked against a database of voices and fingerprints looking for a match.

Recognition systems have typically two phases: enrollment and test. During the enrollment phase, the client deliberately registers on the system one or more biometric traits. The system derives a number of features for these traits to form a client print, template, or model. During the test phase, whether identification or verification, the client is biometrically matched against the model(s).

This paper is solely concerned with the identity verification task. Thus, the two terms verification and recognition referred to herein are used interchangeably to indicate verification.

3.2. Biometric Modalities. Identity verification systems rely on multiple biometric modalities to match clients. These modalities include voice, facial geometry, fingerprint, signature, iris, retina, and hand geometry. Each one of these modalities has been extensively researched in literature. This paper focuses on the voice and the face modalities.

It has been established that multimodal identity verification systems outperform verification systems that rely on a single biometric modality [10, 11]. Such performance gain is more apparent in noisy environments; identity verification systems that rely solely on speech are affected greatly by the microphone type, the level of background noise (street noise, cafeteria atmosphere, ...), and the physical state of the speaker (sickness, mental state, ...). Identity verification systems based on the face modality is dependent on the video camera quality, the face brightness, and the physical appearance of the subject (hair style, beard, makeup, ...).

3.2.1. Voice. Voice verification, also known as speaker recognition, is a biometric modality that relies on features influenced by both the structure of a person's vocal tract and the speech behavioral characteristics. The voice is a widely acceptable modality for person verification and has been a subject for research for decades. There are two forms of speaker verification: text dependent (constrained mode), and text independent (unconstrained mode). Speaker verification is treated in Section 3.3.

3.2.2. Face. The face modality is a widely acceptable modality for person recognition and has been extensively researched. The face recognition process has matured into a science of sophisticated mathematical representations and matching processes. There are two predominant approaches to the face recognition problem: holistic methods and feature-based techniques. Face verification is described in Section 3.4.

3.3. Speaker Verification. The speech signal is an important biometric modality used in the audiovisual verification system. To process this signal a feature extraction module calculates relevant feature vectors from the speech waveform. On a signal window that is shifted at a regular rate a feature vector is calculated. Generally, cepstral-based feature

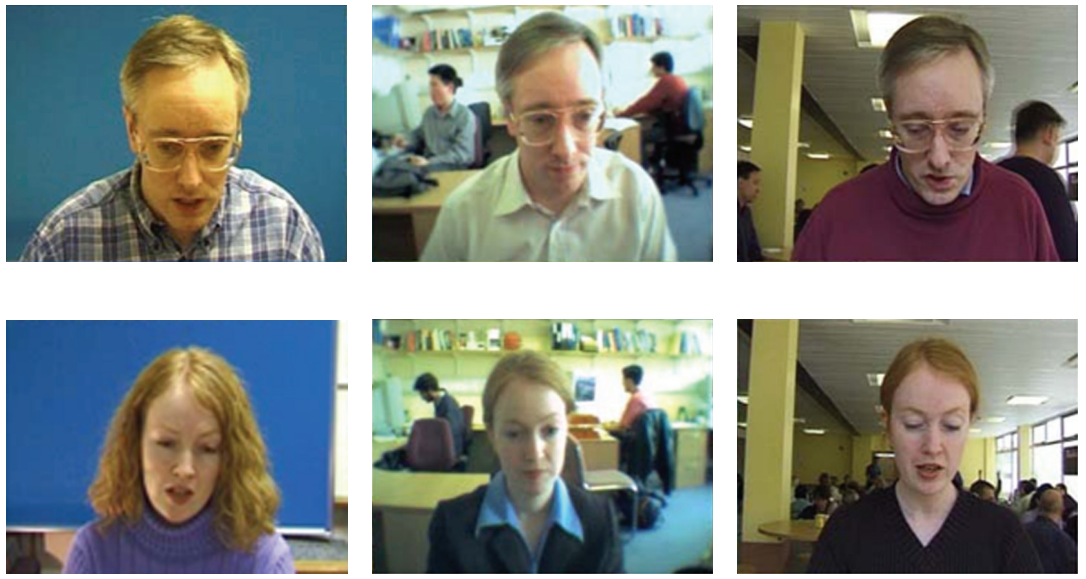


Figure 1: Screenshots from the BANCA database for two subjects in all three scenarios:Controlled (left), degraded (middle), and adverse (right).

vectors are used. A stochastic model is then applied to represent the feature vectors from a given speaker. To verify a claimed identity, new utterance feature vectors are generally matched against the claimed speaker model and against a general model of speech that may be uttered by any speaker, called the world model. The most likely model identifies if the claimed speaker has uttered the signal or not. In text independent speaker verification, the model should not reflect a specific speech structure, that is, a specific sequence of words. State-of-the art systems use Gaussian Mixture Models (GMMs) as stochastic models in text-independent mode. A tutorial on speaker verification is provided in [12].

3.3.1. Feature Extraction. The first part of the speaker verification process is the speech signal analysis. Speech is inherently a nonstationary signal. Consequently, speech analysis is normally performed on short fragments of speech where the signal is presumed stationary. To compensate for the signal truncation, a weighting signal is applied on each window.

Coding the truncated speech windows is achieved through variable resolution spectral analysis [13]. The most common technique employed is filter-bank analysis; it is a conventional spectral analysis technique that represents the signal spectrum with the log-energies using a filter-bank of overlapping band-pass filters.

The next step is cepstral analysis. The cepstrum is the inverse Fourier transform of the logarithm of the Fourier transform of the signal. A determined number of mel frequency cepstral coefficients (MFCCs) are used to represent the spectral envelope of the speech signal. They are derived from the filter-bank energies. To reduce the effects of signals recorded in different conditions, Cepstral mean subtraction and feature variance normalization is used. First- and second-order derivatives of extracted features are appended to the feature vectors to account for the dynamic nature of speech.

3.3.2. Silence Detection. It is well known that the silence part of the signal alters largely the performance of a speaker verification system. Actually, silence does not carry any useful information about the speaker, and its presence introduces a bias in the score calculated, which deteriorates the system performance. Therefore, most of the speaker recognition systems remove the silence parts from the signal before starting the recognition process. Several techniques have been used successfully for silence removal. In our experiments, we suppose that the energy in the signal is a random process that follows a bi-Gaussian model, a first Gaussian modeling the energy of the silence part and the other modeling the energy of the speech part. Given an utterance and more specifically the computed energy coefficients, the bi-Gaussian model parameters are estimated using the EM algorithm. Then, the signal is divided into speech parts and silence parts based on a maximum likelihood criterion. Treatment of silence detection can be found in [14, 15].

3.3.3. Speaker Classification and Modeling. Each speaker possesses a unique vocal signature that provides him with a distinct identity. The purpose of speaker classification is to exploit such distinctions in order to verify the identity of a speaker. Such classification is accomplished by modeling speakers using a Gaussian Mixture Model (GMM).

Gaussian Mixture Models. A mixture of Gaussians is a weighted sum of M Gaussian densities

$$P(x\lambda) = \sum_{i=1:M} \alpha_i f_i(x), \tag{3}$$

where x is an MFCC vector, $f_i(x)$ is a Gaussian density function, and α_i is the corresponding weights. Each Gaussian

is characterized by its mean μ_i and a covariance matrix $\sum_i$. A speaker model λ is characterized by the set of parameters $(\alpha_i, \mu_i, \sum_i)_{i=1:M}$.

For each client, two GMMs are used: the first corresponds to the distribution of the training set of speech feature vectors of that client, and the second represents the distribution of the training vectors of a defined "world model."

To formulate the classification concept, assume that a speaker is presented along with an identity claim C. The feature vectors $V = \{\vec{v}_i\}_{i=1}^{N}$ are extracted. The average log likelihood of the speaker having identity C is calculated as

$$\mathcal{L}(X \mid \lambda_c) = \frac{1}{N}\sum_{i=1}^{N} \log p(\vec{x}_i \mid \lambda_c), \tag{4}$$

where $p(\vec{x}_i \mid \lambda_c) = \sum_{j=1}^{N_G} m_j \mathcal{N}(\vec{x}; \vec{\mu}_j, \sum_j), \lambda = \{m_j \vec{\mu}_j, \sum_j\}_{j=1}^{N_G}$, and $\mathcal{N}(\vec{x}; \vec{\mu}_j, \sum_j) = (1/(2\pi)^{D/2}|\sum_j|^{1/2}) e^{(1/2)(\vec{x}-\vec{\mu}_j)^T \sum_j (\vec{x}-\vec{\mu}_j)}$ is a multivariate Gaussian function with mean $\vec{\mu}_i$ and diagonal covariance matrix $\sum$, and D is the dimension of the feature space, λ_c is the parameter set for person C, N_G is the number of Gaussians, m_j = weight for Gaussian j, and $\sum_{j=1}^{N_j} m_j = 1, m_j \geq 0 \ \forall j$.

With a world model of w persons, the average log likelihood of a speaker being an impostor is found as

$$\mathcal{L}(X \mid \lambda_w) = \frac{1}{N}\sum_{i=1}^{N_W} \log p(\vec{x}_i \mid \lambda_w). \tag{5}$$

An opinion on the claim is then found: $\mathcal{O}(X) = \log \mathcal{L}(X \mid \lambda_c) - \log \mathcal{L}(X \mid \lambda_w)$.

As a final decision to whether the face belongs to the claimed identity, and given a certain threshold t, the claim is accepted when $\mathcal{O}(X) \geq t$ and rejected when $\mathcal{O}(X) < t$.

To estimate the GMM parameters λ of each speaker, the world model is adapted using a Maximum a Posteriori (MAP) adaptation [16]. The world model parameters are estimated using the Expectation Maximization (EM) algorithm [17].

GMM client training and testing is performed on the speaker verification toolkit BECARS [18]. BECARS implements GMMs with several adaptation techniques, for example, Bayesian adaptation, MAP, maximum likelihood linear regression (MLLR), and the unified adaptation technique defined in [19].

3.4. Face Verification. Face verification is a biometric person recognition technique used to verify (confirm or deny) a claimed identity based on a face image or a set of faces (or a video sequence). The process of automatic face recognition can be thought of as being comprised of four stages:

(i) face detection, localization and segmentation;

(ii) normalization;

(iii) facial Feature extraction and tracking;

(iv) classification (identification and/or verification).

These subtasks have been independently researched and surveyed in literature and are briefed next.

3.4.1. Face Detection. Face detection is an essential part of any face recognition technique. Given an image, face detection algorithms try to answer the following questions.

(i) Is there a face in the image?

(ii) If there is a face in the image, where is it located?

(iii) What are the size and the orientation of the face?

Face detection techniques are surveyed in [20, 21].

The face detection algorithm used in this work has been introduced initially by Viola and Jones [22] and later developed further by Lienhart and Maydt [23]. It is a machine learning approach based on a boosted cascade of simple and rotated haar-like features for visual object detection.

3.4.2. Face Tracking in a Video Sequence. Face tracking in a video sequence is a direct extension of still image face detection techniques. However, the coherent use of both spatial and temporal information of faces makes the detection techniques more unique.

The technique used in this work employs the algorithm developed by Lienhart on every frame in the video sequence. However, three types of tracking errors are identified in a talking face video.

(i) More than one face is detected, but only one actually exists in a frame.

(ii) A wrong object is detected as a face.

(iii) No faces are detected.

Figure 2 shows an example detection from the BANCA database [1], where two faces have been detected, one for the actual talking-face subject, and a false alarm.

The correction of these errors is done through the exploitation of spatial and temporal information in the video sequence as the face detection algorithm is run on every subsequent frame. The correction algorithm is summarized as follows.

(a) *More than one face area detected*. The intersections of these areas with the area of the face of the previous frame are calculated. The area that corresponds to the largest calculated intersection is assigned as the face of the current frame. If the video frame in question is the first one in the video sequence, then the decision to select the proper face for that frame is delayed until a single face is detected at a later frame and verified with a series of subsequent face detections.

(b) *No faces detected*. The face area of the previous frame is assigned as the face of the current frame. If the video frame in question is the first one in the video sequence, then the decision is delayed as explained in part (a).

(c) *A wrong object detected as a face*. The intersection area with the previous frame face area is calculated. If this intersection ratio to the area of the previous face is less than a certain threshold, for example, 80%, the previous face is assigned as the face of the current frame.

3.4.3. Face Normalization. Normalizing face images is a required preprocessing step that aims at reducing the variability of different aspects in the face image such as contrast and illumination, scale, translation, rotation, and face masking. Many works in literature [24–26] have normalized face images with respect to translation, scale, and in-plane rotation, while others [27, 28] have also included masking and affine warping to properly align the faces. Craw and Cameron in [28] have used manually annotated points around shapes to warp the images to the mean shape, leading to shape-free representation of images useful in PCA classification.

The preprocessing stage in this work includes four steps.

(i) Scaling the face image to a predetermined size (w_f, h_f).

(ii) Cropping the face image to an inner-face, thus disregarding any background visual data.

(iii) Disregarding color information by converting the face image to grayscale.

(iv) Histogram equalization of the face image to compensate for illumination changes.

Figure 3 shows an example of the four steps.

3.4.4. Feature Extraction. The facial feature extraction technique used in this work uses DCT-*mod2* proposed by Sanderson and Paliwal in [29]. This technique is used in this work for its simplicity and performance in terms of computational speed and robustness to illumination changes.

A face image is divided into overlapping $N \times N$ blocks. Each block is decomposed in terms of orthogonal 2D DCT basis functions and is represented by an ordered vector of DCT coefficients:

$$\left[c_0^{(b,a)} c_1^{(b,a)} \cdots c_{M-1}^{(b,a)}\right]^T, \tag{6}$$

where (b, a) represent the location of the block, and M is the number of the most significant retained coefficients. To minimize the effects of illumination changes, horizontal and vertical delta coefficients for blocks at (b, a) are defined as first-order orthogonal polynomial coefficients, as described in [29].

The first three coefficients c_0, c_1, and c_2 are replaced in (6) by their corresponding deltas to form a feature vector of size $M + 3$ for a block at (b, a):

$$\left[\Delta^h c_0 \Delta^v c_0 \Delta^h c_1 \Delta^v c_1 \Delta^h c_2 \Delta^v c_2 c_3 c_4 \cdots c_{M-1}\right]^T. \tag{7}$$

3.4.5. Face Classification. Face verification can be seen as a two-class classification problem. The first class is the case when a given face corresponds to the claimed identity (client), and the second is the case when a face belongs to an impostor. In a similar way to speaker verification, a GMM is used to model the distribution of face feature vectors for each person.

3.5. Fusion. It has been shown that biometric verification systems that combine different modalities outperform single modality systems [30]. A final decision on the claimed identity of a person relies on both the speech-based and the face-based verification systems. To combine both modalities, a fusion scheme is needed.

Various fusion techniques have been proposed and investigated in literature. Ben-Yacoub et al. [10] evaluated different binary classification approaches for data fusion, namely, Support Vector Machine (SVM), minimum cost Bayesian classifier, Fisher's linear discriminant analysis, C4.5 decision classifier, and multilayer perceptron (MLP) classifier. The use of these techniques is motivated by the fact that biometric verification is merely a binary classification problem. An overview of fusion techniques for audio-visual identity verification is provided in [31].

Other fusion techniques used include the weighted sum rule and the weighted product rule. It has been shown that the sum rule and support vector machines are superior when compared to other fusion schemes [10, 32, 33].

The weighted sum rule fusion technique is used in this study. The sum rule computes the audiovisual score s by weight averaging: $s = w_s s_s + w_f s_f$, where w_s and w_f are speech and face score weights computed so as to optimize the equal error rate (EER) on the training set. The speech and face scores must be in the same range (e.g., $\mu = 0$, $\sigma = 1$) for the fusion to be meaningful. This is achieved by normalizing the scores as follows:

$$s_{\text{norm}(s)} = \frac{s_s - \mu_s}{\sigma_s}, \qquad s_{\text{norm}(f)} = \frac{s_f - \mu_f}{\sigma_f}. \tag{8}$$

4. Audiovisual Imposture

Audiovisual imposture is the deliberate modification of both speech and face of a person so as to make him sound and look like someone else. The goal of such an effort is to analyze the robustness of biometric identity verification systems to forgery attacks. An attempt is made to increase the acceptance rate of an impostor. Transformations of both modalities are treated separately below.

4.1. Speaker Transformation. Speaker transformation, also referred to as voice transformation, voice conversion, or speaker forgery, is the process of altering an utterance from a speaker (impostor) to make it sound as if it was articulated by a target speaker (client). Such transformation can be effectively used to replace the client's voice in a video to impersonate that client and break an identity verification system.

Speaker transformation techniques might involve modifications of different aspects of the speech signal that carries the speaker's identity such as the formant spectra, that is, the coarse spectral structure associated with the different phones in the speech signal [34], the excitation function, that is, the "fine" spectral detail [35], the prosodic features, that is, aspects of the speech that occur over timescales larger than individual phonemes, and the mannerisms such as particular word choice or preferred phrases, or all kinds

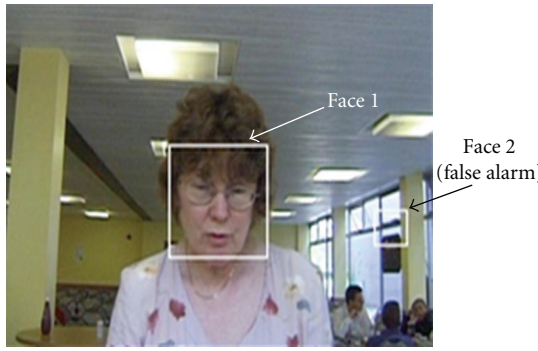


Figure 2: Face detection and tracking.

of other high-level behavioral characteristics. The formant structure and the vocal tract are represented by the overall spectral envelope shape of the signal, and thus they are major features to be considered in voice transformation [36].

Several voice transformation techniques have been proposed in literature. These techniques can be classified as text-dependent methods and text independent methods. In text-dependent methods, training procedures are based on parallel corpora, that is, training data have the source and the target speakers uttering the same text. Such methods include vector quantization [37, 38], linear transformation [36, 39], formant transformation [40], probabilistic transformation [41], vocal tract length normalization (VTLN) [42], and prosodic transformation [38]. In text-independent voice conversion techniques, the system trains on source and target speakers uttering different text. Techniques include text-independent VTLN [42], maximum likelihood constrained adaptation [43], and client memory indexation [44, 45].

The analysis part of a voice conversion algorithm focuses on the extraction of the speaker's identity. Next, a transformation function is estimated. At last, a synthesis step is achieved to replace the source speaker characteristics by those of the target speaker.

Consider a sequence of spectral vectors uttered by the source speaker (impostor) $X_s = [x_1, x_2, \ldots, x_n]$, and a sequence pronounced by the target speaker comprising the same words $Y_t = [y_1, y_2, \ldots, y_n]$. Voice conversion is based on the estimation of a conversion function $\mathcal{F}$ that minimizes the mean square error $\epsilon_{\text{mse}} = E\lfloor \|y - \mathcal{F}(x)\|^2 \rfloor$, where E is the expectation.

Two steps are useful to build a conversion system: training and conversion. In the training phase, speech samples from the source and the target speakers are analyzed to extract the main features. These features are then time aligned, and a conversion function is estimated to map the source and the target characteristics (Figure 4).

The aim of the conversion is to apply the estimated transformation rule to an original speech pronounced by the source speaker. The new utterance sounds like the same speech pronounced by the target speaker, that is, pronounced by replacing the source characteristics by those of the target voice. The last step is the resynthesis of the signal to reconstruct the source speech voice (Figure 5).

Voice conversion can be effectively used by an impostor to impersonate a target person and hide his identity in an attempt to increase the acceptance rate of the impostor by the identity verification system.

In this paper, *MixTrans*, a new mixture-structured bias voice transformation, is proposed and is described next.

4.1.1. MixTrans. A linear time-invariant transformation in the temporal domain is equivalent to a bias in the cepstral domain. However, speaker transformation may not be seen as a simple linear time-invariant transformation. It is more accurate to consider the speaker transformation as several linear time-invariant filters, each of them operating in a part of the acoustical space. This leads to the following form for the transformation:

$$\mathcal{T}_\theta(\mathbf{X}) = \sum_k \prod_k (\mathbf{X} + \mathbf{b}_k) = \sum_k \prod_k \mathbf{X} + \sum_k \prod_k \mathbf{b}_k = \mathbf{X} + \sum_k \prod_k \mathbf{b}_k, \tag{9}$$

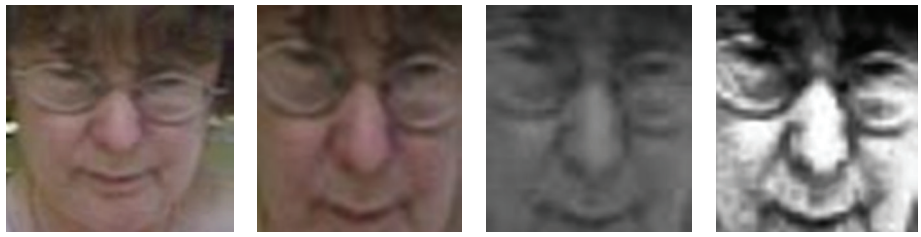
where $\mathbf{b}_k$ represents the kth bias, and $\prod_k$ is the probability of being in the kth part of the acoustical space given the observation vector $\mathbf{X}$. $\prod_k$ is calculated using a universal GMM modeling the acoustic space.

Once the transformation is defined, its parameters have to be estimated. We suppose that speech samples are available for both the source and the target speakers but do not correspond to the same text. Let λ be the stochastic model for a target client. λ is a GMM of the client. Let $\mathbf{X}$ represent the sequence of observation vectors for an impostor (a source client). Our aim is to define a transformation $\mathcal{T}_\theta(\mathbf{X})$ that makes the source client vector resemble the target client. In other words, we would like to have the source vectors be best represented by the target client model λ through the application of the transformation $\mathcal{T}_\theta(\mathbf{X})$. In this context the Maximum likelihood criterion is used to estimate the transformation parameters:

$$\hat{\theta} = \arg\max_\theta \mathcal{L}(\mathcal{T}_\theta(\mathbf{X}) \mid \lambda). \tag{10}$$

Since λ is a GMM, $\mathcal{T}_\theta(\mathbf{X})$ is a transform of the source vectors $\mathbf{X}$, and $\mathcal{T}_\theta(\mathbf{X})$ depends on another model λ_w, then $\mathcal{L}(\mathcal{T}_\theta(\mathbf{X}) \mid \lambda)$ in (10) can be written as

$$\begin{aligned}
&\mathcal{L}(\mathcal{T}_\theta(\mathbf{X} \mid \lambda) \\
&\quad = \prod_{t=1}^{T} \mathcal{L}(\mathcal{T}_\theta(\mathbf{X}_t) \mid \lambda) \\
&\quad = \prod_{t=1}^{T} \sum_{m=1}^{M} \frac{1}{(2\pi)^{D/2} \mid \sum_m \mid^{1/2}} e^{-(1/2)(\mathcal{T}_\theta(\mathbf{X}_t)-\mu_m)^T \sum_m^{-1} (\mathcal{T}_\theta(\mathbf{X}_t)-\mu_m)} \\
&\quad = \prod_{t=1}^{T} \sum_{m=1}^{M} \frac{1}{(2\pi)^{D/2} \mid \sum_m \mid^{1/2}} \\
&\qquad \times e^{-(1/2)(\mathbf{X}_t + \sum_{k=1}^{K} \prod_{kt} b_k - \mu_m)^T \sum_m^{-1} (\mathbf{X}_t + \sum_{k=1}^{K} \prod_{kt} b_k - \mu_m)}).
\end{aligned} \tag{11}$$



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

Figure 3: Preprocessing face images. (a) Detected face. (b) Cropped face (inner face). (c) Grayscale face. (d) Histogram-equalized face.

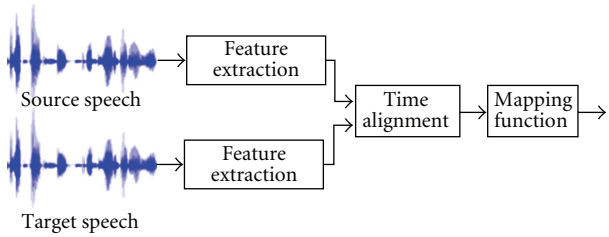


Figure 4: Training.

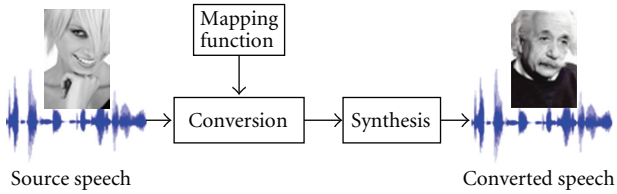


Figure 5: Conversion.

Finding $\{b_k\}$ such that (11) is maximized is found through the use of the EM algorithm. In the expectation "E" step, the probability α_{mt} of component m is calculated. Then, at the maximization "M" step, the log-likelihood is optimized dimension by dimension for a GMM with a diagonal covariance matrix:

$$ll = \sum_{t=1}^{T}\sum_{m=1}^{M}\alpha_{mt}\left[\log\frac{1}{\sigma_m\sqrt{2\pi}} - \frac{1}{2}\frac{\left(\mathbf{X}_t + \sum_{k=1}^{K}\prod_{kt}\mathbf{b}_k - \mu_m\right)^2}{\sigma_m^2}\right]. \tag{12}$$

Maximizing

$$\frac{\partial ll}{\partial b_l} = 0 \Longrightarrow -\sum_{t=1}^{T}\sum_{m=1}^{M}\alpha_{mt}\frac{\left(\mathbf{X}_t + \sum_{k=1}^{K}\prod_{kt}\mathbf{b}_k - \mu_m\right)\prod_{lt}}{\sigma_m^2} = 0, \quad \text{for } l = 1\cdots K, \tag{13}$$

then,

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\frac{\alpha_{mt}P_{lt}}{\sigma_m^2}\left(\mathbf{X}_t - \mu_m\right) = -\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{k=1}^{K}\frac{\alpha_{mt}\prod_{kt}\prod_{lt}\mathbf{b}_k}{\sigma_m^2}, \quad \text{for } l = 1\cdots K,$$

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\frac{\alpha_{mt}\prod_{lt}}{\sigma_m^2}\left(\mathbf{X}_t - \mu_m\right) = -\sum_{k=1}^{K}\mathbf{b}_k\sum_{m=1}^{M}\sum_{t=1}^{T}\frac{\alpha_{mt}\prod_{lt}\prod_{kt}}{\sigma_m^2}, \quad \text{for } l = 1\cdots K, \tag{14}$$

and finally, in matrix notation,

$$-\left(\sum_{m}\sum_{t}\frac{\alpha_{mt}\prod_{lt}\prod_{kt}}{\sigma_m^2}\right)(\mathbf{b}_k) = \left(\sum_{m}\sum_{t}\frac{\alpha_{mt}\prod_{lt}(\mathbf{X}_t - \mu_m)}{\sigma_m^2}\right). \tag{15}$$

This matrix equation is solved at every iteration of the EM algorithm.

4.1.2. Speech Signal Reconstruction. It is known that the cepstral domain is appropriate for classification due to the physical significance of the Euclidean distance in this space [13]. However, the extraction of cepstral coefficients from the temporal signal is a nonlinear process, and the inversion of this process is not uniquely defined. Therefore, a solution has to be found in order to take the advantage of the good characteristics of the cepstral space while applying the transformation in the temporal domain.

Several techniques have been proposed to overcome this problem. In [46], harmonic plus noise analysis has been used for this purpose. Instead of trying to find a transformation allowing the passage from the cepstral domain to the temporal domain, a different strategy is adopted. Suppose that an intermediate space exists where transformation could be directly transposed to the temporal domain. Figure 6 shows the process where the temporal signal goes through a two-step feature extraction process leading to the cepstral coefficients that may be easily transformed into target speaker-like cepstral coefficients by applying the transformation function $\mathcal{T}_\theta(\mathbf{X})$ as discussed previously.

The transformation trained on the cepstral domain cannot be directly projected to the temporal domain since the feature extraction module $(\mathcal{F}_2 \circ \mathcal{F}_1)$ is highly nonlinear.

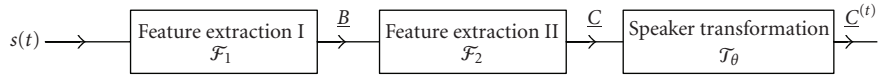


Figure 6: Steps from signal to transformed cepstral coefficients.

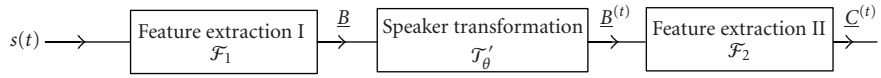


Figure 7: Steps from signal to transformed cepstral coefficients when transformation is applied in a signal-equivalent space.

However, a speaker transformation determined in the $\underline{B}$ space may be directly projected in the signal space, for example, $\underline{B}$ space may be the spectral domain. But, for physical significance it is better to train the transformation in the cepstral domain. Therefore, we suppose that another transformation $\mathcal{T}'_\theta(\mathbf{X})$ exists in the $\underline{B}$ space leading to the same transformation in the cepstral domain satisfying thereby the two objectives: transformation of the signal and distance measurement in the cepstral domain. This is shown in Figure 7.

This being defined, the remaining issue is how to estimate the parameters θ of the transformation $\mathcal{T}'_\theta(\mathbf{X})$ in order to get the same transformation result as in the cepstral domain. This is detailed next.

4.1.3. Estimating Signal Transformation Equivalent to a Calculated Cepstral Transformation. The transformation in the cepstral domain is presumably determined; the idea is to establish a transformation in the $\underline{B}$ space leading to cepstral coefficients similar to the one resulting from the cepstral transformation.

Let $\underline{\widehat{C}}^{(t)}$ represent the cepstral vector obtained after the application of the transformation in the $\underline{B}$ domain, and let $\underline{C}^{(t)}$ represent the cepstral vector obtained when applying the transformation in the cepstral domain. The difference defines an error vector:

$$\underline{e} = \underline{C}^{(t)} - \underline{\widehat{C}}^{(t)}. \tag{16}$$

The quadratic error can be written as

$$E = |e|^2 = \underline{e}^T \underline{e}. \tag{17}$$

Starting from a set of parameters for $\mathcal{T}'_\theta$, the gradient algorithm may be applied in order to minimize the quadratic error E. For every iteration of the algorithm the parameter θ is updated using

$$\theta^{(i+1)} = \theta^{(i)} - \mu \frac{\partial E}{\partial \theta}, \tag{18}$$

where μ is the gradient step.

The gradient of the error with respect to parameter θ is given by

$$\frac{\partial E}{\partial \theta} = -2\underline{e}^T \frac{\partial \underline{\widehat{C}}^{(t)}}{\partial \theta}. \tag{19}$$

Finally, the derivative of the estimated transformed cepstral coefficient with respect to θ can be obtained using a gradient descent

$$\frac{\partial \underline{\widehat{C}}^{(t)}}{\partial \theta} = \frac{\partial \underline{\widehat{C}}^{(t)^T}}{\partial \underline{B}^{(t)}} \frac{\partial \underline{B}^{(t)}}{\partial \theta}. \tag{20}$$

In order to illustrate this principle, let us consider the case of MFCC analysis leading to the cepstral coefficients. In this case, $\mathcal{F}_1$ is just the Fast Fourier Transform (FFT) followed by the power spectral calculation (the phase being kept constant). $\mathcal{F}_2$ is the filterbank integration in the logarithm scale followed by the inverse DCT transform. We can write

$$\begin{aligned} \widehat{C}_l^{(t)} &= \sum_{k=1}^{K} \log\left(\sum_{i=1}^{N} a_i^{(k)} B_i^{(k)}\right) \cos\left(2\pi l \frac{f_k}{F}\right), \\ B_i^{(t)} &= B_i \cdot \theta_i, \end{aligned} \tag{21}$$

where $\{a_i\}$ are the filter-bank coefficients, f_k the central frequencies of the filter-bank, and θ_i is the transfer function at frequency bin i of the transformation $\mathcal{T}'_\theta(\mathbf{X})$.

Using (21), it is straightforward to compute the derivatives in (20):

$$\begin{aligned} \frac{\partial \widehat{C}_i^{(t)}}{\partial B_j^{(t)}} &= \sum_{k=1}^{K} \frac{a_j^{(k)}}{\sum_{i=1}^{N} a_i^{(k) B_i^{(t)}}} \cos\left(2\pi l \frac{f_k}{F}\right), \\ \frac{\partial B_i^{(t)}}{\partial \theta_j} &= B_j \delta_{ij}. \end{aligned} \tag{22}$$

Equations (19), (20), and (22) allow the implementation of this algorithm in the case of MFCC.

Once $\mathcal{T}'_\theta(\mathbf{X})$ completely defined, the transformed signal may be determined by applying an inverse FFT to $B(t)$ and using the original phase to recompose the signal window. In order to consider the overlapping between adjacent windows, the Overlap and Add (OLA) algorithm is used [47].

4.1.4. Initializing the Gradient Algorithm. The previous approach is computationally expensive. Actually, for each signal window, that is, from 10 milliseconds to 16 milliseconds, a gradient algorithm is to be applied. In order to alleviate this high computational algorithm, a solution consists in finding a good initialization of the gradient algorithm. This may be obtained by using an initial value for the transformation $\mathcal{T}'_\theta(\mathbf{X})$, the transformation obtained for the previous signal window.

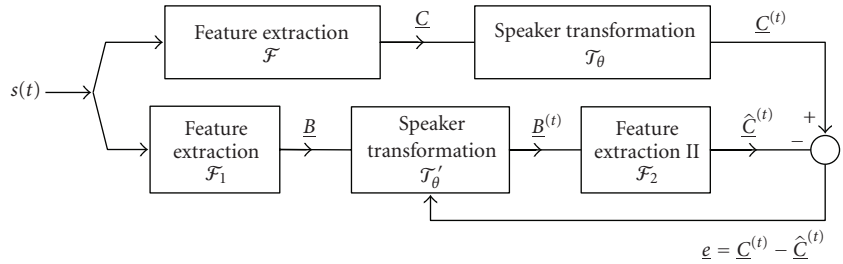


Figure 8: Signal-level transformation parameters tuned with a gradient descent algorithm.

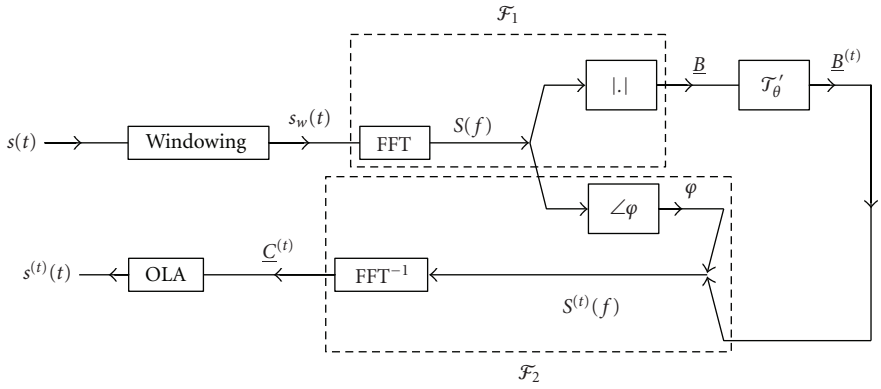


Figure 9: Speech signal feature extraction, transformation, and reconstruction.

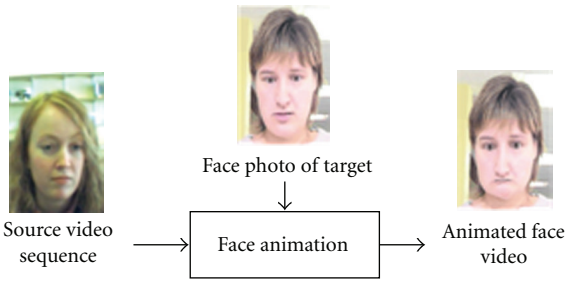


Figure 10: Face animation.

4.2. Face Animation. To complete the scenario of audiovisual imposture, speaker transformation is coupled with face transformation. It is meant to produce synthetically an "animated" face of a target person, given a still photo of his face and some animation parameters defined by a source video sequence. Figure 10 depicts the concept.

The face animation technique used in this paper is MPEG-4 compliant, which uses a very simple thin-plane spline warping function defined by a set of reference points on the target image, driven by a set of corresponding points on the source image face. This technique is described next.

4.2.1. MPEG-4 2D Face Animation. MPEG-4 is an object-based multimedia compression standard, which defines a standard for face animation [48]. It specifies 84 feature points (Figure 11) that are used as references for Facial Animation Parameters (FAPs). 68 FAPs allow the representation of facial expressions and actions such as head motion and mouth and eye movements. Two FAP groups are defined, visemes (FAP group 1) and expressions (FAP group 2). Visemes (FAP1) are visually associated with phonemes of speech; expressions (FAP2) are joy, sadness, anger, fear, disgust, and surprise.

An MPEG-4 compliant system decodes an FAP stream and animates a face model that has all feature points properly determined. In this paper, the animation of the feature points is accomplished using a simple thin-plate spline warping technique and is briefly described next.

4.2.2. Thin-Plate Spline Warping. The thin-plate spline (TPS), initially introduced by Duchon [49], is a geometric mathematical formulation that can be applied to the problem of 2D coordinate transformation. The name *thin-plate spline* indicates a physical analogy to bending a thin sheet of metal in the vertical z direction, thus displacing x and y coordinates on the horizontal plane.

Given a set of data points $\{w_i,\ i = 1, 2, \ldots, K\}$ in a 2D plane—for our case, MPEG-4 facial feature points—a radial basis function is defined as a spatial mapping that maps a location x in space to a new location $f(x) = \sum_{i=1}^{K} c_i \phi(\|x - w_i\|)$, where $\{c_i\}$ is a set of mapping coefficients, and the kernel function $\phi(r) = r^2 \ln r$ is the thin-plate spline [50]. The mapping function $f(x)$ is fit between corresponding sets of points $\{x_i\}$ and $\{y_i\}$ by minimizing the "bending energy" I, defined as the sum of squares of the second derivatives:

$$I[f(x,y)] = \iint_{\mathbb{R}^2} \left[\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right] dx\, dy. \tag{23}$$

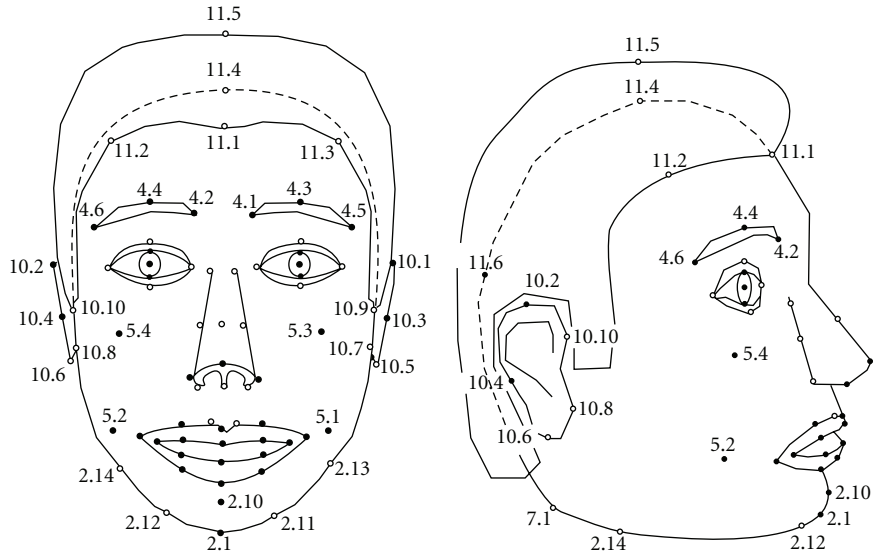


Figure 11: MPEG-4 feature points.

(a) Original sample video frame (BANCA client number 9055)

(b) Annotated face

Figure 12: Feature point annotation on the BANCA database.

5. Effects of Imposture on Verification—Experimental Results

To test the robustness of IV systems, a state-of-the-art baseline audio-visual IV system is built. This system follows the BANCA "P" protocol and is based on a classical GMM approach for both speech and face modalities. It is completely independent from the voice and face transformations described above.

5.1. Verification Experiments

5.1.1. Speaker Verification. For speech, feature extraction and silence detection is first performed, as described in Sections 3.3.1 and 3.3.2. Then GMM speaker classification is performed with 256 Gaussians. The world model of BANCA is adapted using MAP adaptation, and its parameters estimated using the EM algorithm, as discussed in Section 3.3.3 above. The world model is used as a Universal Background Model (UBM) for training to amplify the variability between different speakers. In fact, to improve the performance of the IV system, we use a larger UBM by combining BANCA world model and g2 when training and testing is performed on g1 and vice versa. This is possible because g1 and g2 of BANCA are totally independent. Client models are then adapted from the UBM using speech features from the enrollment set of BANCA. To verify a claimed identity of a test speaker, his/her features are extracted and compared to both the UBM and the GMM of the client. The average log likelihood is calculated, and an acceptance or a rejection decision is taken as described in Section 3.3.3. A total of 234 true client tests and 312 impostor tests (per group) were performed in compliance with BANCA's "P" protocol. Figure 14(a) shows the DET curves for speaker verification on g1 and g2, with an EER of 4.38% and 4.22%, respectively.

(a) Neutral face (b) Joy expression (c) Sad expression (d) Right eye wink (e) Left eye wink

Figure 13: Selected frames from an animated face with various expressions.

5.1.2. Face Verification Experiments. Face verification is based on extracting facial features from a video sequence as described in Section 3.4. First, the face tracking module extracts faces in all frames and retains only 5 of them for training and/or testing. The 5 frames selected are equally distributed across the video sequence so as to have a good sample of faces. These faces are then resized to 48×64, gray-scaled, cropped to 36×40, and then histogram-equalized. Then DCT feature extraction follows. Neighboring blocks of 8×8 with an overlap of 50% is used. M, the number of retained coefficients, is fixed at 15 [29]. In a similar way to speaker verification, GMMs are used to model the distribution of face feature vectors for each person.

For the same BANCA "P" protocol, and a total of 234 true clients and 312 impostor tests (per group per frame—5 frames per video) the DET curves for face verification are shown in Figure 14(b) with an EER of 23.5% and 22.2% for g1 and g2, respectively.

5.1.3. Score Fusion. Figure 14(c) shows an improvement of the verification by score fusion of both modalities, with an EER of 4.22% for g1 and 3.47% for g2. The optimized weights w_s and w_f are integers 8 and 3, respectively, as described in Section 3.5.

5.2. Transformation Experiments. BANCA defines in its protocols imposture attempts during which a speaker proclaims in his/her own voice and face to be someone else. This "zero-effort" imposture is unrealistic, and any text-independent verification system should be able to detect easily the forgery by contrasting the impostor model against the claimed identity model. To make the verification more difficult, transformation of both voice and face is performed.

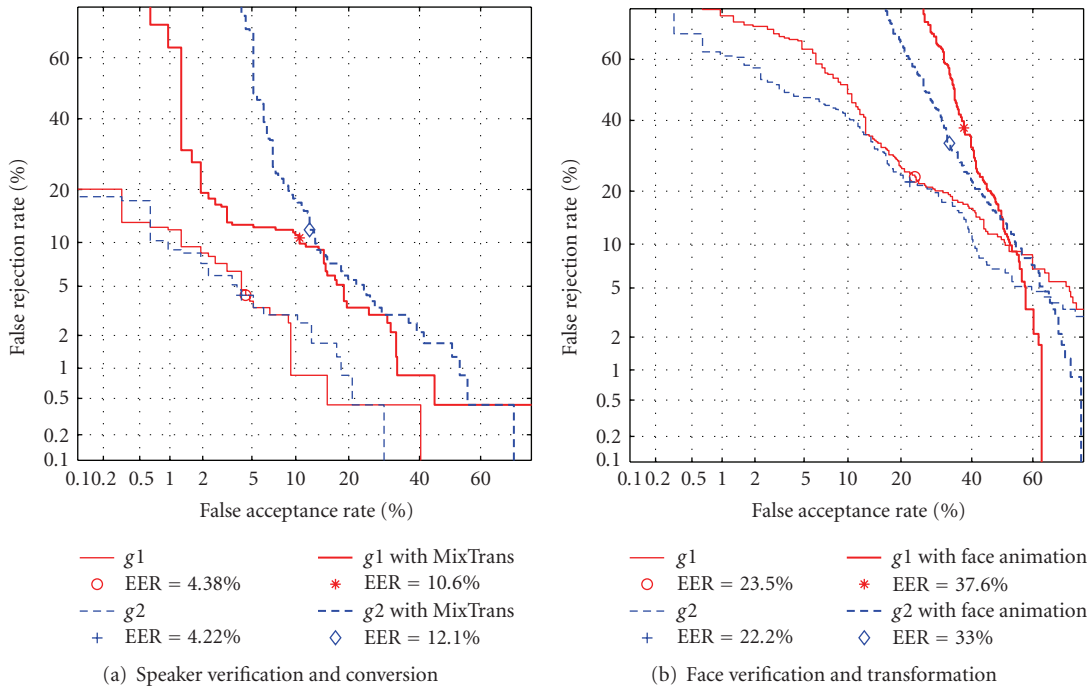
5.2.1. Voice Conversion Experiments. BANCA has total of 312 impostor attacks per group in which the speaker claims in his own words to be someone else. These attempts are replaced by the transformed voices as described in Section 4.1. For each attempt, MFCC analysis is performed, and transformation coefficients are calculated in the cepstral domain using the EM algorithm. Then the signal transformation parameters are estimated using a gradient descent algorithm. The transformed voice signal is then reconstructed with an inverse FFT and OLA as described in Section 4.1.3. The pitch of the transformed voice had to be adjusted to match better the target speaker' pitch. Verification experiments are repeated with the transformed voices. The result is an increase of the EER from 4.38% to 10.6% on g1 and from 4.22% to 12.1% on g2 (Figure 14(a)).

5.2.2. Face Conversion Experiments. Given a still picture of the face of a target person, the MPEG-4 facial feature points are first manually annotated as shown in Figure 12. A total of 61 feature points out of 83 specified by MPEG-4 are annotated, the majority of which belong to the eyes and the mouth regions. Others have less impact on FAPs or do not affect them at all.

The FAPs used in the experiments correspond to a subset of 33 out of the 68 FAPs defined by MPEG-4. Facial actions related to head movement, tongue, nose, ears, and jaws are not used. The FAPs used correspond to mouth, eye, and eyebrow movements, for example, horizontal displacement of right outer lip corner (`stretch_r_cornerlip_o`), vertical displacement of top right eyelid (`close_t_r_eyelid`), and vertical displacement of left outer eyebrow (`raise_l_o_eyebrow`). Figure 13 shows animated frames simulating the noted expressions.

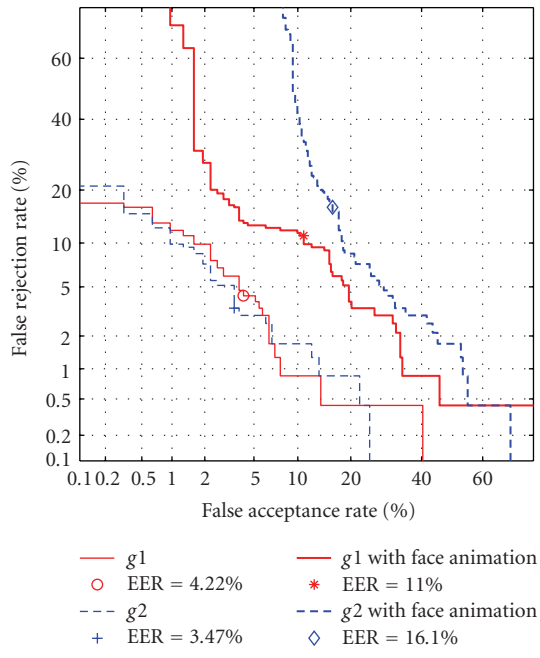
A synthesized video sequence is generated by deforming a face from its neutral state according to determined FAP values at a rate of 25 frames per second. For the experiments presented in this work, these FAPs are selected so as to produce a realistic talking head that is not necessarily synchronized with the associated transformed speech. The only association with speech is the duration of the video sequence, which corresponds to the total time of speech. The detection and the measure of the level of audiovisual speech synchrony is not treated in this work but has been reported in [51–53] to improve the verification performance.

BANCA has total of 312 impostor attacks per group in which the speaker claims in his own words and facial expressions to be someone else. These are replaced by the synthetically animated videos with the transformed speech. The experiments have shown a deterioration of the performance from an EER from (23.5%, 22.2%) on (g1, g2) to (37.6%, 33.0%) (Figure 14(b)) for face, and from (4.22%, 3.47%) to (11.0%, 16.1%) for the audio-visual system (Figure 14(c)).



(a) Speaker verification and conversion

(b) Face verification and transformation



(c) Audiovisual verification and imposture (fused results)

Figure 14: Audiovisual verification and imposture results on BANCA.

6. Conclusion

This paper provides a review of biometric identity verification techniques and describes their evaluation and robustness to imposture. It proposes *MixTrans*, a mixture-structured bias voice transformation technique in the cepstral domain, which allows a transformed audio signal to be estimated and reconstructed in the temporal domain. It also couples the audio conversion with an MPEG-4 compliant face animation system that warps facial feature points using a simple thin-plate spline. The proposed audiovisual forgery is completely independent from the baseline audiovisual IV system and can be used to attack any other audiovisual IV system. The Results drawn from the experiments show that state-of-the-art verification systems are vulnerable to forgery, with an EER average increase from 3.8% to 13.5%. This increase clearly shows that such attacks represent a serious challenge and a security threat to audio-visual IV

systems. The results show that state-of-the-art IV systems are vulnerable to forgery attacks, which indicate more impostor acceptance, and, for the same threshold, more genuine client denial. This should drive more research towards more robust IV systems.

References

[1] E. Bailly-Bailliére, S. Bengio, F. Bimbot, et al., "The BANCA database and evaluation protocol," in *Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '03)*, vol. 2688 of *Lecture Notes in Computer Science*, pp. 625–638, Guildford, UK, June 2003.

[2] K. Messer, J. Matas, J. Kittler, and K. Jonsson, "Xm2vtsdb: the extended m2vts database," in *Proceedings of the 2nd International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA '99)*, pp. 72–77, Washington, DC, USA, March 1999.

[3] J. S. D. Mason, F. Deravi, C. C. Chibelushi, and S. Gandon, "Project: david (digital audio visual integrated database)," Tech. Rep., Department of Electrical and Electronic Engineering, University of Wales Swansea, Swansea, UK, 1996, http://eegalilee.swan.ac.uk.

[4] S. Garcia-Salicetti, C. Beumier, G. Chollet, et al., "Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities," in *Proceedings of the 4th IAPR International Conference on Audio and Video-Based Person Authentication (AVBPA '03)*, vol. 2688 of *Lecture Notes in Computer Science*, pp. 845–853, Guildford, UK, June 2003.

[5] A. C. Morris, J. Koreman, H. Sellahewa, et al., "The securephone pda database, experimental protocol and automatic test procedure for multimodal user authentication," Tech. Rep., The SecurePhone Project, 2006.

[6] J. P. Egan, *Signal Detection Theory and ROC Analysis*, Series in Cognition and Perception, Academic Press, New York, NY, USA, 1975.

[7] T. Fawcett, "Roc graphs: notes and practical considerations for researchers," 2004, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.9777.

[8] S. Bengio, J. Mariethoz, and M. Keller, "The expected performance curve," in *Proceedings of the 2nd International Conference on Machine Learning, ICML, Workshop on ROC Analysis in Machine Learning*, Bonn, Germany, August 2005.

[9] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proceedings of the 5th European Conference on Speech Communication and Technology (EUROSPEECH '97)*, vol. 4, pp. 1895–1898, Rhodes, Greece, September 1997.

[10] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1065–1074, 1999.

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[12] F. Bimbot, J.-F. Bonastre, C. Fredouille, et al., "A tutorial on text-independent speaker verification," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 430–451, 2004.

[13] C. Chouzenoux, *Analyse spectrale à résolution variable. Application au signal de parole*, Ph.D. thesis, Ecole Nationale Superieure des Télécommunications, Paris, France, 1982.

[14] J. Trmal, J. Zelinka, J. Psutka, and L. Müller, "Comparison between GMM and decision graphs based silence/speech detection method," in *Proceedings of the 11th International Conference Speech and Computer (SPECOM '06)*, pp. 376–379, Anatolya, St. Petersburg, Russia, June 2006.

[15] D. R. Paoletti and G. Erten, "Enhanced silence detection in variable rate coding systems using voice extraction," in *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems (MWSCAS '00)*, vol. 2, pp. 592–594, Lansing, Mich, USA, August 2000.

[16] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 291–298, 1994.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[18] R. Blouet, C. Mokbel, H. Mokbel, E. Sánchez Soto, G. Chollet, and H. Greige, "Becars: a free software for speaker verification," in *Proceedings of the Speaker and Language Recognition Workshop (ODYSSEY '04)*, pp. 145–148, Toledo, Spain, May-June 2004.

[19] C. Mokbel, "Online adaptation of HMMs to real-life conditions: a unified framework," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 342–357, 2001.

[20] E. Hjelmås and B. K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[21] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 511–518, Kauai, Hawaii, USA, December 2001.

[23] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '02)*, vol. 1, pp. 900–903, Rochester, NY, USA, September 2002.

[24] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" *International Journal of Computer Vision*, vol. 28, no. 3, pp. 270–277, 1998.

[25] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[26] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[27] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[28] I. Craw and P. Cameron, "Manifold caricatures: on the psychological consistency of computer face recognition," in *Proceedings of the British Machine Vision Conference (BMVC '92)*, pp. 498–507, Springer, Leeds, UK, September 1992.

[29] C. Sanderson and K. K. Paliwal, "Fast feature extraction method for robust face verification," *Electronics Letters*, vol. 38, no. 25, pp. 1648–1650, 2002.

[30] J. Kittler, "Combining classifiers: a theoretical framework," *Pattern Analysis and Applications*, vol. 1, no. 1, pp. 18–27, 1998.

[31] C. Sanderson and K. K. Paliwal, "Identity verification using speech and face information," *Digital Signal Processing*, vol. 14, no. 5, pp. 449–480, 2004.

[32] V. Chatzis, A. G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 29, no. 6, pp. 674–680, 1999.

[33] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez-Rodriguez, "A comparative evaluation of fusion strategies for multimodal biometric verification," in *Proceedings of the 4th IAPR International Conference on Audio and Video-Based Person Authentication (AVBPA '03)*, vol. 2688 of *Lecture Notes in Computer Science*, pp. 830–837, Guildford, UK, June 2003.

[34] A. Kain and M. W. Macon, "Spectral voice conversion for text-to-speech synthesis," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 1, pp. 285–288, Seattle, Wash, USA, May 1998.

[35] B. Yegnanarayana, K. Sharat Reddy, and S. P. Kishore, "Source and system features for speaker recognition using AANN models," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 1, pp. 409–412, Salt Lake City, Utah, USA, May 2001.

[36] A. Kain, *High resolution voice transformation*, Ph.D. thesis, OGI School of Science and Engineering, Oregon Health & Science University, Portland, Ore, USA, 2001.

[37] M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara, "Voice conversion through vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '88)*, vol. 1, pp. 655–658, New York, NY, USA, April 1988.

[38] L. M. Arslan, "Speaker transformation algorithm using segmental codebooks (STASC)," *Speech Communication*, vol. 28, no. 3, pp. 211–226, 1999.

[39] H. Ye and S. Young, "Perceptually weighted linear transformations for voice conversion," in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH '03)*, vol. 4, pp. 2409–2412, Geneva, Switzerland, September 2003.

[40] E. Turajlic, D. Rentzos, S. Vaseghi, and C.-H. Ho, "Evaluation of methods for parameteric formant transformation in voice conversion," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 1, pp. 724–727, Hong Kong, April 2003.

[41] Y. Stylianou, O. Cappé, and E. Moulines, "Continuous probabilistic transform for voice conversion," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 131–142, 1998.

[42] D. Sundermann, H. Ney, and H. Hoge, "Vtln-based cross-language voice conversion," in *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '03)*, pp. 676–681, St. Thomas, Virgin Islands, USA, December 2003.

[43] A. Mouchtaris, J. van der Spiegel, and P. Mueller, "Non-parallel training for voice conversion by maximum likelihood constrained adaptation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 1, pp. 1–4, Montreal, Canada, May 2004.

[44] P. Perrot, G. Aversano, R. Blouet, M. Charbit, and G. Chollet, "Voice forgery using ALISP: indexation in a client memory," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 1, pp. 17–20, Philadelphia, Pa, USA, March 2005.

[45] H. Ye and S. Young, "Voice conversion for unknown speakers," in *Proceedings of the 8th International Conference of Spoken Language Processing (ICSLP '04)*, pp. 1161–1164, Jeju Island, South Korea, October 2004.

[46] Y. Stylianou and O. Cappe, "A system for voice conversion based on probabilistic classification and a harmonic plus noise model," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 1, pp. 281–284, Seattle, Wash, USA, May 1998.

[47] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 1975.

[48] A. M. Tekalp and J. Ostermann, "Face and 2-D mesh animation in MPEG-4," *Signal Processing: Image Communication*, vol. 15, no. 4-5, pp. 387–421, 2000.

[49] J. Duchon, "Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces," *RAIRO: Analyse Numérique*, vol. 10, no. 12, pp. 5–12, 1976.

[50] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[51] H. Bredin and G. Chollet, "Audio-visual speech synchrony measure for talking-face identity verification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07)*, vol. 2, pp. 233–236, Honolulu, Hawaii, USA, April 2007.

[52] H. Bredin and G. Chollet, "Audiovisual speech synchrony measure: application to biometrics," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 70186, 11 pages, 2007.

[53] H. Bredin and G. Chollet, "Making talking-face authentication robust to deliberate imposture," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '08)*, pp. 1693–1696, Las Vegas, Nev, USA, April 2008.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 260148, 16 pages
doi:10.1155/2009/260148

Research Article

Sorted Index Numbers for Privacy Preserving Face Recognition

Yongjin Wang and Dimitrios Hatzinakos

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON, Canada M5S 3G4

Correspondence should be addressed to Yongjin Wang, ywang@comm.utoronto.ca

Received 30 September 2008; Revised 3 April 2009; Accepted 18 August 2009

Recommended by Jonathon Phillips

This paper presents a novel approach for changeable and privacy preserving face recognition. We first introduce a new method of biometric matching using the sorted index numbers (SINs) of feature vectors. Since it is impossible to recover any of the exact values of the original features, the transformation from original features to the SIN vectors is noninvertible. To address the irrevocable nature of biometric signals whilst obtaining stronger privacy protection, a random projection-based method is employed in conjunction with the SIN approach to generate changeable and privacy preserving biometric templates. The effectiveness of the proposed method is demonstrated on a large generic data set, which contains images from several well-known face databases. Extensive experimentation shows that the proposed solution may improve the recognition accuracy.

1. Introduction

Biometric recognition has been an active research area in the past two decades. Biometrics-based recognition systems determine or confirm the identity of an individual based on the physiological and/or behavioral characteristics [1]. A wide variety of biometric modalities have been investigated in the past. Examples of these biometrics include physiological traits such as fingerprint, face, iris, and behavioral characteristics such as gait and keystroke. Each biometric has its strengths and weaknesses. The choice of a biometric is dependent on the properties of the biometric and the requirements of the specific application. Depending on different application context, a biometric system can operate in identification mode or verification mode [1]. Figure 1 depicts the general block diagram of biometric recognition systems.

During enrolment, a feature vector $\mathbf{g}_i$, $i = 1, 2, \ldots, N$, where N is the total number of users, is extracted from the biometric data of each user and stored in the system database as templates. Biometric identification is a one-to-many comparison to find an individual's identity. In identification mode, given an input feature vector $\mathbf{p}$, if the identity of $\mathbf{p}$, $\mathbf{I}_p$, is known to be in the system database, that is, $\mathbf{I}_p \in \{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N\}$, then $\mathbf{I}_p$ can be determined by $\mathbf{I}_p = \mathbf{I}_k = \min_k\{S(\mathbf{p}, \mathbf{g}_k)\}$, $k = 1, 2, \ldots, N$, where S denotes the similarity measure. The performance of a biometric identification system is usually evaluated in terms of correct recognition rate (CRR).

A biometric verification system is a one-to-one match that determines whether the claim of an individual is true. At the verification stage, a feature vector $\mathbf{p}$ is extracted from the biometric signal of the authentication individual $\mathbf{I}_p$, and compared with the stored template $\mathbf{g}_k$ of the claimed identity $\mathbf{I}_k$ through a similarity function S. The evaluation of a verification system can be performed in terms of hypothesis testing [2], $\mathbf{H}_0$: $\mathbf{I}_p = \mathbf{I}_k$, the claimed identity is correct; $\mathbf{H}_1$: $\mathbf{I}_p \neq \mathbf{I}_k$, the claimed identity is not correct. The decision is made based on the system threshold τ, $\mathbf{H}_0$ is decided if $S(\mathbf{p}, \mathbf{g}_k) \leq \tau$, and $\mathbf{H}_1$ is decided if $S(\mathbf{p}, \mathbf{g}_k) > \tau$. A verification system makes two types of errors: false accept (deciding $\mathbf{H}_0$ when $\mathbf{H}_1$ is true), and false reject (deciding $\mathbf{H}_1$ when $\mathbf{H}_0$ is true). The performance of a biometric verification system is usually evaluated in terms of false accept rate (FAR, $P(\mathbf{H}_0 \mid \mathbf{H}_1)$), false reject rate (FRR, $P(\mathbf{H}_1 \mid \mathbf{H}_0)$), and equal error rate (EER, operating point where FAR and FRR are equal). The FAR and FRR are closely related functions of the system decision threshold τ.

While biometric technology provides various advantages, there exist some problems. In the first place, biometric data

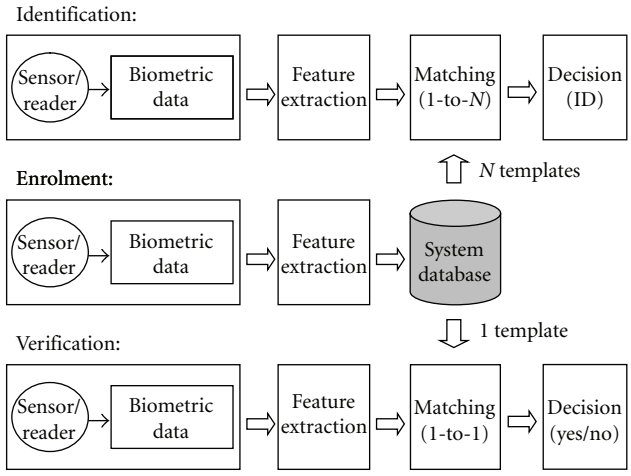


Figure 1: Block diagram of biometric recognition systems.

reflects the user's physiological/behavior characteristics. If the storage device of biometric templates is obtained by an adversary, the user's privacy may be compromised. The biometric templates should be stored in a format such that the user's privacy is preserved even when the storage device is attacked. Secondly, biometrics cannot be easily changed and reissued if compromised due to the limited number of biometric traits that a human has. This is of particular importance in biometric verification scenarios. Ideally, just like password, the biometrics should be changeable. The users may use different biometric representation for different applications. When the biometric template in one application is compromised, the biometric signal itself is not lost forever and a new biometric template can be issued.

A number of research works have been proposed in the recent years to address the changeability and privacy preserving problems of biometric systems. One approach is to combine the biometric technology with cryptographic systems [3]. In a biometric cryptosystem, a randomly generated cryptographic key is bound with the biometric features in a secure way such that both the key and the biometric features cannot be revealed if the stored template is compromised. The cryptographic key can be retrieved if sufficiently similar biometric features are presented. Error correction algorithms are usually employed to tolerate errors. Due to the binary nature of cryptographic keys, such systems usually require discrete representation of biometric data, such as minutia points for fingerprints and iris code. However, the feature vectors of many other biometrics, such as face, are usually represented in continuous domain. Therefore, to apply such a scheme, the continuous features need to be discretized first. It should be noted that such methods produce changeable cryptographic keys, while the biometric data is not changeable. Furthermore, the security level of such methods still needs to be further investigated [4, 5].

An alternative and effective solution is to apply repeatable and noninvertible transformations on the biometric features [2]. With this method, every enrollment (or application) can use a different transform. When a biometric template is compromised, a new one can be generated using a new transform. In mathematical language, the recognition problem can be formulated as follows. Given a biometric feature vector $\mathbf{u}$, the biometric template $\mathbf{g}$ is generated through a generation function $\mathbf{g} = \mathrm{Gen}(\mathbf{u}, \mathbf{k})$. Different templates can be generated by varying the control factor $\mathbf{k}$. During verification, the same transformation is applied to the authentication feature vector, $\mathbf{g}' = \mathrm{Gen}(\mathbf{u}', \mathbf{k})$, and the matching is based on similarity measure in the transformed domain, that is, $\mathrm{S}(\mathbf{g}, \mathbf{g}')$. The major challenge here lies in the difficulty of preserving the similarity measure in the transformed domain, that is, $\mathrm{S}(\mathbf{g}, \mathbf{g}') \approx \mathrm{S}(\mathbf{u}, \mathbf{u}')$. Further, to ensure the property of privacy protection, the generation function $\mathrm{Gen}(\mathbf{u}, \mathbf{k})$ should be nonfinvertible such that $\hat{\mathbf{u}} = \mathrm{Rec}(\mathbf{g}, \mathbf{k}) \neq \mathbf{u}$, where $\mathrm{Rec}(\mathbf{g}, \mathbf{k})$ is the reconstruction function when both the template $\mathbf{g}$ and control factor $\mathbf{k}$ are known.

Among various biometrics, face recognition has been one of the most passive, natural, and noninvasive types of biometrics. Such characteristics of face recognition make it a good choice for some surveillance and monitoring applications. It can also be used in supporting video search and indexing, video-conferencing, interactive games, physical access control, computer network login, and ATM. Many face recognition methods have been proposed in the literature, which can be roughly categorized into holistic template matching-based system, geometrical local feature-based system, and hybrid systems [6]. Promising results have also been reported under controlled condition [7]. In general, the selection of a face recognition scheme is dependent on the specific requirements of a given task [6]. Appearance-based approaches (such as principal component analysis (PCA) and linear discriminant analysis (LDA)) that treat the face image as a holistic pattern are among the most successful methods [6, 8]. In this paper, we first introduce a novel method for face recognition based on sorted index numbers (SINs) of appearance-based facial features. Unlike traditional face recognition methods which store either the original image or facial features as templates, the proposed method stores the SIN vectors only. A matching algorithm is introduced to measure the similarity between two SIN vectors. Because it is impossible to recover the exact values of the original features based on the index numbers, the SIN method is noninvertible. To further enhance the security and address the irrevocable problem, intentional random projection (RP) is applied prior to the sorting operation such that the generated biometrics template is both changeable and privacy preserving. Experimental results on a large data set demonstrate the effectiveness of the proposed solution.

The remainder of this paper is organized as follows. Section 2 provides a review of related works. Section 3 introduces the proposed method. Experimental results along with detailed discussion are presented in Section 4. Finally, conclusions are provided in Section 5.

2. Related Works

To address the privacy and irrevocability problem of biometric systems, many tentative solutions have been introduced

in the literature using various biometrics. Among the earliest efforts, Soutar et al. [9] presented a correlation-based method for fingerprint-based biometric verification, and Davida et al. [10] proposed to store a set of user specific error correction parameters as template for an iris-based system. However, both of these two works are lack of practical implementation and cannot provide rigorous security guarantees [3].

Juels and Wattenberg [11] introduced a fuzzy commitment scheme, which generalized and improved Davida's methods. The fuzzy commitment scheme assumes binary representation of biometric features, and error correction algorithms are used to tolerate errors due to the noisy nature of biometric data. Hao et al. [12] presented a similar scheme on an iris-based problem using a two-level error correction mechanism. Later, a polynomial reconstruction-based scheme, fuzzy vault, is proposed by Juels and Sudan [13]. The fuzzy vault scheme assumes the biometric data being represented by discrete features (e.g., minutia points in fingerprints). In this scheme, error tolerance is achieved by using the property of secret sharing, while the security is obtained by hiding genuine points into randomly generated chaff points. A few implementation works of fuzzy vault have been reported in [14, 15] based on fingerprints. Although the paper proves that this scheme is secure in an information-theoretic sense, it is clear that it is vulnerable to attacks via record multiplicity [5]. Further drawbacks of the method include high computational complexity and high error rate [14, 15].

Dodis et al. [16] presented a theoretical work, fuzzy extractor, for generation of cryptographic keys from noisy biometric data using error correction code and hash functions. Their paper also assumes the biometric features in discrete domain. Different constructions for three metric spaces: Hamming distance, set difference, and edit distance are introduced. Yagiz et al. [17] introduced a quantization-based method for mapping of continuous face features to discrete form and utilized a known secure construction for secure key generation. However, Boyen [18] showed that the fuzzy extractor may be not secure for multiple use of the same biometrics data.

Kevenaar et al. [19] proposed a helper data system for generation of renewable and privacy preserving binary template. A set of fiducial points is first identified from six key objects of human face, and Gabor filters are applied to extract features from a small patch centered around every fiducial point. The extracted features are discretized by a thresholding method, and the reliability of each bit is measured based on statistical analysis. The binary template is generated by combining the extracted reliable bit with a randomly generated key through an XOR operation, and BCH code is applied for error correction. The indexes of the selected reliable bit, the mean vector for feature thresholding, the binary template, and the hash of the key are stored for verification. Their experiments demonstrate that the performance of the binary feature vectors is only degraded slightly comparing with the original features. However, the performance of their system depends on accurate localization of key object and fiducial points.

Savvides et al. [20, 21] proposed an approach for cancelable biometric authentication in the encrypted domain. The training face images are convolved with a random kernel first; the transformed images are then used to synthesize a single minimum average correlation energy filter. At the point of verification, query face image is convolved with the same random kernel and then correlates with the stored filer to examine the similarity. If the storage card is ever attacked, a new random kernel may be applied. They show that the performance is not affected by the random kernel. However, it is not clear how the system preserves privacy if the random kernel is known by an adversary. The original biometrics may be retrieved through deconvolution if the random kernel is known.

Boult [22] introduced a method for face-based revocable biometrics based on robust distance measures. In this scheme, the face features are first transformed through scaling and translation, and the resulting values are partitioned into two parts, the integer part and the fractional part. The integer part is encrypted using Public Key (PK) algorithms, and the fractional part is retained for local approximation. A user-specific passcode is included to address the revocation problem. In a subsequent paper [23], a similar scheme is applied on a fingerprint problem, and detailed security analysis is provided. Their methods demonstrate both improvement in accuracy and privacy. However, it is assumed that the private key cannot be obtained by an imposter. In the case of known private key and transform parameters, the biometrics features can be exactly recovered.

Teoh et al. [24] introduced a two-factor scheme, Bio-Hashing method, which produces changeable non-invertible biometric template, and also claimed good performance, near zero EER. In BioHashing, a feature vector $\mathbf{u} \in \mathbb{R}^n$ is first extracted from the user's biometric data. For each user, a user-specific transformation matrix $R \in \mathbb{R}^{n\times m}, m \leq n$, is generated randomly (associated with a key or token), and the Gram-Schmidt orthonormalization method is applied to R, such that all the columns of R are orthonormal. The extracted feature vector $\mathbf{u}$ is then transformed by $\mathbf{x} = R^T\mathbf{u}$, and the resulting vector $\mathbf{x}$ is quantized by $\mathbf{b}_i = 0$, if $\mathbf{x}_i < t$, and $\mathbf{b}_i = 1$, if $\mathbf{x}_i \geq t$, $i = 1, 2, \ldots, m$, where t is a predefined threshold value and usually set to 0. The binary vector $\mathbf{b}$ is stored as the template. The technique has been applied on various biometric traits [25, 26] and demonstrates zero or near zero equal error rate in ideal case; that is, both the biometric data and the key are legitimate. In the stolen key scenario, the BioHashing method usually degrades the verification accuracy. Lumini and Nanni [27] introduce some ideas to improve the performance of BioHashing in case of stolen key by utilizing different threshold values and fuse the scores. However, as shown in [28], as well as the experimental results in this paper, even in the both legitimate scenario, the performance of BioHashing technique is highly dependent on the characteristics and dimensionality of the extracted features.

In summary, existing works either can not provide robust privacy protection, or sacrifice recognition accuracy for privacy preservation. In this paper, we propose a new method for changeable and privacy preserving template generation

using random projection and sorted index numbers. As it will be shown, the proposed method is also capable of improving the recognition accuracy.

3. Methodology

This section presents the proposed method for privacy preserving face recognition. An overview of the sorted index numbers (SINs) method as well as the similarity measure algorithm is first introduced. Next, the analysis of the SIN algorithm is provided in detail. The random projection-based changeable biometrics scheme is then described. Finally, privacy analysis of the proposed method is presented.

3.1. Overview of SIN Method. The proposed method utilizes sorted index numbers instead of the original facial features as templates for recognition. The procedure of creating the proposed SIN feature vector is as follows.

(1) Extract feature vector $\mathbf{w} \in \mathbb{R}^n$ from the input face image $\mathbf{z}$.

(2) Compute $\mathbf{u} = \mathbf{w} - \overline{\mathbf{w}}$, where $\overline{\mathbf{w}}$ is the mean feature vector calculated from the training data.

(3) Sort the feature vector $\mathbf{u}$ in descending order, and store the corresponding index numbers in a new vector $\mathbf{g}$.

(4) The generated vector $\mathbf{g} \in \mathbb{Z}^n$ that contains the sorted index numbers is stored as template for recognition.

For example, given $\mathbf{u} = \{u_1, u_2, u_3, u_4, u_5, u_6\}$, the sorted vector in descending order is $\hat{\mathbf{g}} = \{u_4, u_6, u_2, u_1, u_3, u_5\}$, then the template is $\mathbf{g} = \{4, 6, 2, 1, 3, 5\}$.

The method for computing the similarity between two SIN vectors is as follows.

(1) Given two SIN feature vectors $\mathbf{g} \in \mathbb{Z}^n$ and $\mathbf{p} \in \mathbb{Z}^n$, where $\mathbf{g}$ denotes the template vector, and $\mathbf{p}$ denotes the probe vector. Start from the first element g_1 of $\mathbf{g}$.

(2) Search for the corresponding element in $\mathbf{p}$, that is, $p_j = g_1$. Record $d_1 = j - 1$, where j is the index number in $\mathbf{p}$.

(3) Eliminate the obtained p_j in the previous step from $\mathbf{p}$, and obtain $\mathbf{p}^1 = \{p_1, p_2, \ldots, p_{j-1}, p_{j+1}, \ldots, p_n\}$.

(4) Repeat steps 2 and 3 on the following elements of $\mathbf{g}$ until g_{n-1}. Record $d_2, d_3, \ldots, d_{n-1}$.

(5) The similarity measure of $\mathbf{g}$ and $\mathbf{p}$ is computed as $S(\mathbf{g}, \mathbf{p}) = \sum_{i=1}^{n-1} d_i$.

Illustration Example.

(1) For two SIN feature vectors $\mathbf{g} = \{4, 6, 2, 1, 3, 5\}$ and $\mathbf{p} = \{2, 5, 3, 6, 1, 4\}$, we first search the 1st element $g_1 = 4$, and find that $p_6 = 4$. Therefore $d_1 = 6 - 1 = 5$. Eliminate p_6 from $\mathbf{p}$ and we form a new vector of $\mathbf{p}^1 = \{2, 5, 3, 6, 1\}$.

(2) Search the 2nd element $g_2 = 6$, and find that $p_4^1 = 6$. Therefore $d_2 = 4 - 1 = 3$. Eliminate p_4^1 from $\mathbf{p}^1$ and form a new vector of $\mathbf{p}^2 = \{2, 5, 3, 1\}$.

(3) Search the 3rd element $g_3 = 2$, and find that $p_1^2 = 2$. Therefore $d_3 = 1 - 1 = 0$. Eliminate p_1^2 from $\mathbf{p}^2$ and form a new vector of $\mathbf{p}^3 = \{5, 3, 1\}$.

(4) Search the 4th element $g_4 = 1$, and find that $p_3^3 = 1$. Therefore $d_4 = 3 - 1 = 2$. Eliminate p_3^3 from $\mathbf{p}^3$ and form a new vector of $\mathbf{p}^4 = \{5, 3\}$.

(5) Search the 5th element $g_5 = 3$, and find that $p_2^4 = 1$. Therefore $d_5 = 2 - 1 = 1$.

(6) Compute $S(\mathbf{g}, \mathbf{p}) = \sum_{i=1}^{n-1} d_i = 5 + 3 + 0 + 2 + 1 = 11$.

3.2. Methodology Analysis. To understand the underlying rationale of the proposed algorithm, we first look into an alternative presentation of the method, named Pairwise Relational Discretization (PRD). The relative relation of different bins has been used to represent histogram shape in [29]. Here, the pairwise relative relation of features is used for Euclidean distance approximation. The procedure of producing the PRD feature vector is as follows.

(1) Extract feature vector $\mathbf{w} \in \mathbb{R}^n$ from the input face image $\mathbf{z}$.

(2) Compute $\mathbf{u} = \mathbf{w} - \overline{\mathbf{w}}$, where $\overline{\mathbf{w}}$ is the mean feature vector calculated from the training data.

(3) Compute binary representation of $\mathbf{u}$ by comparing the pairwise relation of all the elements in $\mathbf{u}$ according to

$$b_{ij} = \begin{cases} 1, & u_i \geq u_j, \\ 0, & u_i < u_j. \end{cases} \tag{1}$$

(4) Concatenate all the generated binary bits into one vector $\mathbf{b} = \{b_{12}, \ldots, b_{1n}, b_{23}, \ldots, b_{2n}, b_{34}, \ldots, b_{n-1,n}\}$. Store the binary vector $\mathbf{b}$ as template for recognition.

The similarity measure of the PRD method is based on Hamming distance. Unlike traditional discretization method, which quantizes individual elements based on some predefined quantization levels, the proposed method takes the global characteristics of the feature vectors into consideration. This is interpreted by comparing the pairwise relation of all groups of two elements in the vector. The intuition behind the idea is to consider an n-dimensional space as combinations of 2-dimensional planes. In n-dimensional subspace, when the similarity of two vectors is evaluated by Euclidean distance, each element of the vectors is treated as coordinates in the corresponding basis $\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\}$, and the similarity is based on the spatial closeness. The elements are essentially the projection coefficients of the vector onto each basis (i.e., lines). Here, instead of projecting onto lines, we explore the projection onto 2D planes. Figure 2 offers a diagrammatic illustration of the PRD method. For two points in n-dimensional subspace, if they are spatially close to each other, then in large number of 2D planes, their projection location should be close to each other, that is, small Hamming distance, and vise versa. Therefore, the Euclidean distance between two vectors can

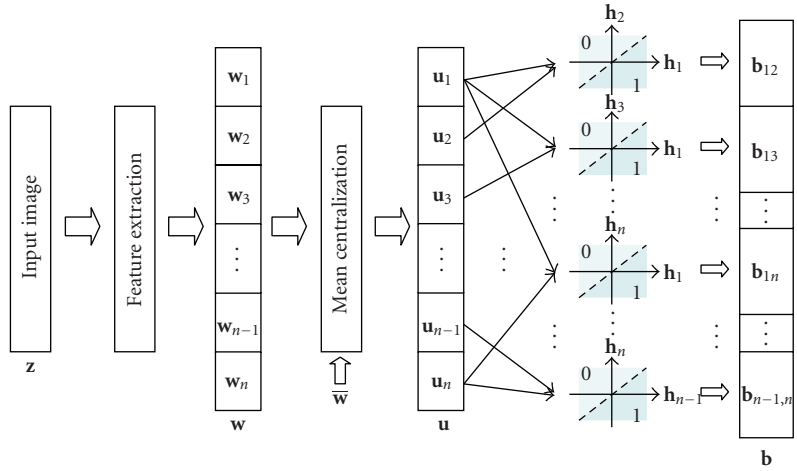


Figure 2: Diagram of Pairwise Relational Discretization (PRD) method.

be approximated by the Hamming distance between the corresponding PRD vectors. The mean centralization step is to leverage the significance of each element such that no single dimension will overwhelm others. The discretization step partitions a plane into two regions by comparing the pairwise relation. It reduces the sensitivity of the variation of individual elements and therefore possibly provides better error tolerance. Figure 3 shows the intra-class and inter-class distribution of 100 PCA coefficients based on 1000 randomly selected images from the experimental data set. The PCA vectors are normalized to unit length, and Euclidean distance and SIN distance are used as dissimilarity measure. Note that the size of the overlapping area of the intra-class and inter-class distributions indicates the recognition error. It can be observed that the SIN method produces smaller error than the original features, therefore will possibly provide better recognition performance.

A major drawback of the PRD method is the high dimensionality of the generated binary PRD vector. For an n-dimensional vector, the generated binary vector $\mathbf{b}$ will have a size of $n(n-1)/2$. For example, for a feature vector with $n = 100$, the PRD vector will have a size of 4950. This problem introduces high storage and computational requirements. This is particularly important for applications with high processing speed demand. To improve this, we note that the PRD method is based on pairwise relation of all the vector elements, and the same information can be exactly preserved from the sorted index numbers of the vector; that is, any single bit in $\mathbf{b}$ can be derived from the SIN vector.

Let $\mathbf{g}$ and $\mathbf{p}$ denote the SIN vector of template and probe images, respectively, $\mathbf{b}_g$ and $\mathbf{b}_p$ represent the corresponding PRD vectors, then we have

$$H\left(\mathbf{b}_g, \mathbf{b}_p\right) = S(\mathbf{g}, \mathbf{p}) = \sum_{i=1}^{n-1} d_i, \tag{2}$$

where $H(\mathbf{b}_g, \mathbf{b}_p)$ and $S(\mathbf{g}, \mathbf{p})$ denote the Hamming distance and SIN distance, respectively, and $d_i, i = 1, \ldots, n$, represents the Hamming distance associated with every single element in $\mathbf{g}$.

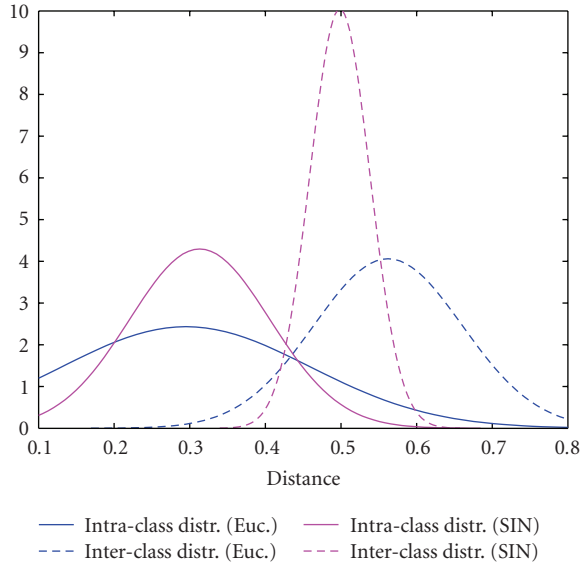


Figure 3: Comparison of intraclass and interclass distribution using Euclidean and SIN distances.

Proof of (2). Since $\mathbf{g}$ and $\mathbf{b}_g$ are derived from the same feature vector, in $\mathbf{b}_g$, there are $n-1$ bits that are associated with the first element of $\mathbf{g}$, g_1. If $p_j = g_1$, where j is the index number of the corresponding element in $\mathbf{p}$, then all the index numbers to the left of p_j will have different bit values in $\mathbf{b}_p$, that is, $d_1 = j - 1$. It should be noted that since the Hamming distance for all the bits associated with $p_j = g_1$ has been computed, the p_j element should be removed for the calculation of next iteration. After the Hamming distances for all the elements in $\mathbf{g}$ and $\mathbf{p}$ are computed, the sum of them will correspond to the Hamming distance of $\mathbf{b}_g$ and $\mathbf{b}_p$, that is, $H(\mathbf{b}_g, \mathbf{b}_p) = S(\mathbf{g}, \mathbf{p}) = \sum_{i=1}^{n-1} d_i$.

Equation (2) shows that the proposed SIN and PRD methods produce exactly the same results. To test the effectiveness of SIN over PRD in computational complexity, we performed experiments on a computer with Intel *Core*2

CPU 2.66 GHz. With an original feature vector of dimensionality 100, the average time for PRD feature extraction and matching is 26.2 milliseconds, while the SIN method only consumes less than 0.9 milliseconds. □

3.3. Changeable Biometrics. To address the changeability problem in biometric verification systems, one solution is to scramble the order of the features before the sorting operation. However, the security of such method is the same as the encryption/decryption key method, where the original SIN vectors will be obtained if the scrambling rule is compromised. In this paper, for the purpose of comparative study, we adopt the random projection-(RP-) based scheme as in [24].

Depending on the requirements of the application, the changeable biometric system can be implemented in two scenarios: user-independent projection and user-dependent projection. In the user-independent scenario, all the users use the same matrix for projection. This matrix can be controlled by the application provider, and therefore the users do not need to carry the matrix (or equivalently a key for matrix generation) for verification. The user-dependent scenario is a two-factor authentication scheme, and requires the presentation of both the biometrics data and projection matrix at the time and point of verification. In both scenarios, the biometric template can be regenerated by changing the projection matrix.

The theory of random projection is first introduced by the Johnson-Lindenstrauss lemma [30].

Lemma 1 (J-L lemma). *For any* $0 < \epsilon < 1$*, and an integer s, let m be a positive integer such that* $m \geq m_0 = O(\epsilon^{-2}\log s)$*. For any set B of s points in* $\Re^n$*, there exists a map f:*$\Re^n \to \Re^m$ *such that for all* $\mathbf{u}, \mathbf{v} \in B$*,*

$$(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \left\|f(\mathbf{u})-f(\mathbf{v})\right\|^2 \leq (1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2. \tag{3}$$

This lemma states that the pairwise distance between any two vectors in the Euclidean space can be preserved up to a factor of ϵ, when projected onto a random m-dimension subspace. Random projection has been used as a dimension reduction tool in face recognition [31], image processing [32], and a privacy preserving tool in data mining [33] and biometrics [24]. The implementation of random projection can be carried out by generating a matrix of size $n \times m$, $m \leq n$, with each entry an independent and identically distributed (i.i.d.) random variable, and applying the Gram-Schmidt method for orthonormalization. Note that when $m = n$, it becomes the random orthonormal transformation (ROT). In user-independent scenario, for two facial feature vectors $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$, since the same ROT matrix $R \in \mathbb{R}^{n \times n}$ is applied, we have the well-known property of ROT:

$$\begin{aligned}\left\|R^T\mathbf{u}-R^T\mathbf{v}\right\|^2 &= \left\|R^T\mathbf{u}\right\|^2 + \left\|R^T\mathbf{v}\right\|^2 - 2\mathbf{u}^T RR^T\mathbf{v} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}^T\mathbf{v} \\ &= \|\mathbf{u}-\mathbf{v}\|^2.\end{aligned} \tag{4}$$

It can be seen that the ROT transform exactly preserves the Euclidean distance of original features. When the projected dimensionality is $m < n$, although exact preservation can not be obtained, the pairwise distance can be approximately preserved. The larger the m, the better the preservation. Since the SIN method also approximates the Euclidean distance, the SIN vectors obtained after RP can also approximately preserve the similarity between two original vectors.

In the user-dependent scenario, different users are associated with distinct projection matrices. The FAR corresponds to the probability of deciding $\mathbf{H}_0$ when $\mathbf{H}_1$ is true, $P(\mathbf{H}_0 \mid \mathbf{H}_1)$, and the FRR corresponds to $P(\mathbf{H}_1 \mid \mathbf{H}_0)$. Note that for the FRR, even in case of a user-dependent scenario, the same orthogonal matrix R is used for the same user, and hence the situation is the same as the user-independent scenario. Therefore we only need to analyze the influence of different projection matrix over the FAR.

Let R_u and R_v represent the RP matrices for feature vectors $\mathbf{u}$ and $\mathbf{v}$, respectively. Let $\mathbf{x} = R_u^T\mathbf{u}$ and $\mathbf{y} = R_v^T\mathbf{v}$, and $\mathbf{g}$ and $\mathbf{p}$ denote the SIN vectors for $\mathbf{x}$ and $\mathbf{y}$, respectively. Due to the randomness of RP, the total number of possible outputs for $\mathbf{g}$ and $\mathbf{p}$ is equal to the number of permutations $m!$. Let γ denote the number of index permutations that have a distance of less than τ to the vector $\mathbf{g}$, then the probability of $\mathbf{p}$ being falsely identified by $\mathbf{g}$ is $P(\mathbf{H}_0 \mid \mathbf{H}_1) = \gamma/m!$. It can be seen that the probability of false accept depends on the characteristics and dimensionality of the features. If the features are well separated, that is, smaller γ value, with relatively higher dimensionality, the false accept rate will be small. The above analysis in user-dependent scenario also applies if the biometrics data is stolen by an adversary, since the $\mathbf{v}$ vector can be exactly the same as $\mathbf{u}$. This also explains the changeability of the method.

Figure 4 shows the distribution of the distance between two feature vectors using user-independent and user-dependent random projections. We randomly selected two PCA features vectors ($n = 100$) of the same subject from the employed data set, performed the same key and different key scenario 2000 times, and plotted the distribution of the Euclidean distance and SIN distance, respectively, at different projection dimensions. The PCA feature vectors are normalized to unit length, and the distances are normalized by dividing the largest value, respectively, 2 for Euclidean distance and $m(m-1)/2$ for SIN. It can be observed that by applying the same key, the mean of the Euclidean distance in the projected domain is centered around the original Euclidean distance, and the variance of the distances decreases as the projected dimensionality increases. This demonstrates better distance preservation at higher projection dimension. When different keys are applied, the mean of the distance distribution shifts to the right, that is larger distance. The clear separation of the distribution indicates the changeability of the proposed method.

3.4. Privacy Analysis. Since the SIN method only stores the index numbers of the sorted feature vector $\mathbf{u}$, the transformation from $\mathbf{u}$ to the corresponding SIN vector $\mathbf{g}$ is non-invertible. There is no effective reconstruction being possible to recover the exact values of $\mathbf{u}$ from $\mathbf{g}$. The most an adversary can do is to estimate the values of $\mathbf{u}$ based on some statistics or his/her own features. By using such method, an

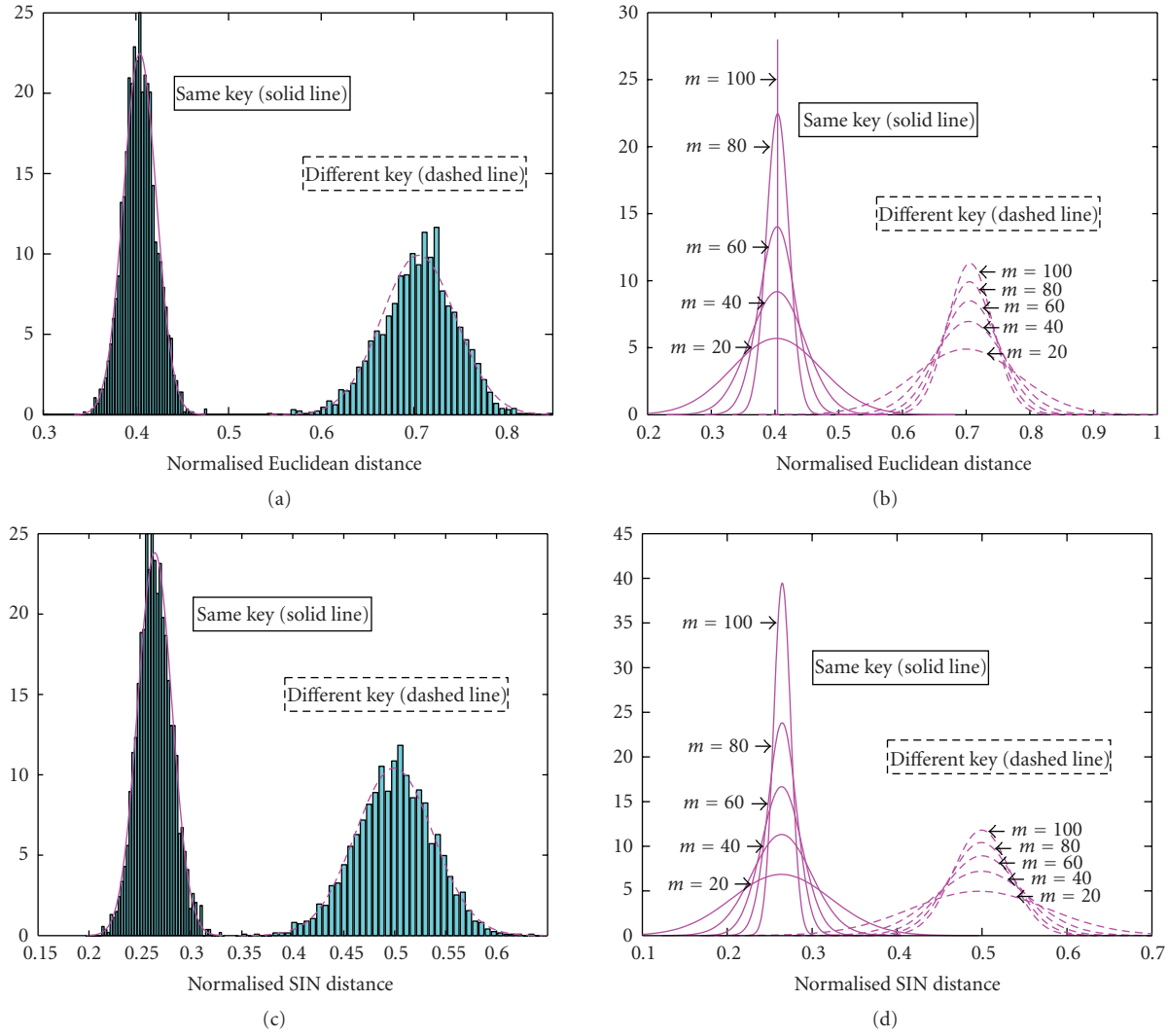


Figure 4: Gaussian approximation of the distribution of (a) normalized Euclidean distance (NED), (c) normalized SIN distance (NSD), with $n = 100$, $m = 80$. Distribution of (b) NED, and (d) NSD, at different projection dimensionality in same key and different key scenarios.

adversary can only produce an approximation of the original features. For RP, when the projected dimensionality m is smaller than the dimensionality n of the original features, even the worst case that the projection matrix is known by an adversary, an estimation will produce an approximation of the original features with variance inverse proportional to m, that is, the smaller the m, the larger the estimation variance [33]. Since both the RP and SIN methods are non-invertible transformations, the combination of these two is expected to produce stronger privacy protection.

To analyze the privacy preserving properties of the proposed method, we introduce the following privacy measures:

Definition 1. A feature vector $\mathbf{u} \in \Re^n$ is called privacy protected at element-wise level α, where α is computed as

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} 1 - [1 - x_i]h(1 - x_i), \quad x_i = \frac{\mathbf{Var}(u_i - \hat{u}_i)}{\mathbf{Var}(u_i)}, \tag{5}$$

where $\mathbf{var}(\cdot)$ denotes variance, $\hat{u}_i$ is the estimated value of element u_i, and $h(x)$ is unit step function, that is, $h(x) = 1$ if $x \geq 0$ and $h(x) = 0$ otherwise. The function $h(x)$ is utilized to regulate the significance of all the elements, such that the variance ratio of any element is maximum 1.

Using the variance of difference between the actual and perturbed values has been widely adopted as a privacy measure for individual attributes in data mining [34]. Similarly, here we take the variance of difference between the original and estimated values as a measure of the privacy protection of individual elements. When the variance ratio of any attribute is greater or equal to 1, that is, $\mathbf{Var}(u_i - \hat{u}_i) \geq \mathbf{Var}(u_i)$, then the estimation of that attribute essentially provides no useful information, and the attribute is strongly protected. The element-wise privacy level α measures the average privacy protection of individual elements. The greater the α value, the better the privacy protection.

Besides measuring the privacy protection of the individual elements, it is also important to measure the global characteristics of the feature vectors such that the estimated vector is not close to the original one up to certain similarity functions. In [35], it is shown that any arbitrary distance functions can be approximately mapped to Euclidean distance domain through certain algorithms. In this paper, we take the squared Euclidean distance between the estimated and original feature vectors as a measure of privacy.

Definition 2. A feature vector $\mathbf{u} \in \Re^n$ is called privacy protected at vector-wise level β, where β is computed as:

$$\beta = \frac{\mathbf{E}\left(\|\hat{\mathbf{u}} - \mathbf{u}\|^2\right)}{\mathbf{E}\left(\|\mathbf{r} - \mathbf{u}\|^2\right)}, \tag{6}$$

where $\mathbf{E}(\cdot)$ denotes expectation, $\|\cdot\|$ denotes the squared Euclidean distance, and $\mathbf{r}$ is any random vector in the estimation feature space. If the average distance between the estimated and original vector is approaching the average distance between any random vector and the original vector, then the estimated vector essentially exhibits randomness, and therefore does not disclose information about $\mathbf{u}$; that is, the larger the β, the better privacy. Without loss of generality, we assume that all the vectors have unit length. Since the vectors are centralized to zero mean, the average distance between any randomly selected vector $\mathbf{r}$ and the original vector $\mathbf{u}$ is

$$\begin{aligned}\mathbf{E}\left(\|\mathbf{r} - \mathbf{u}\|^2\right) &= \mathbf{E}\left(\|\mathbf{r}\|^2 + \|\mathbf{u}\|^2 - 2\mathbf{r}^T\mathbf{u}\right) \\ &= 2 - 2\mathbf{E}\left(\mathbf{r}^T\mathbf{u}\right) = 2,\end{aligned} \tag{7}$$

where we use the fact that $\mathbf{E}(\mathbf{r}^T\mathbf{u}) = \mathbf{E}(\sum_{i=1}^n r_i u_i) = \sum_{i=1}^n \mathbf{E}(r_i u_i) = \sum_{i=1}^n \mathbf{E}(r_i)\mathbf{E}(u_i) = 0$, since r_i is independent of u_i and has zero mean. Therefore, for unit length vectors, (6) can be written as

$$\beta = \frac{\mathbf{E}\left(\|\hat{\mathbf{u}} - \mathbf{u}\|^2\right)}{2}. \tag{8}$$

Figure 5 shows the privacy measures α and β as functions of projected dimension m, with the original dimensionality $n = 100$. Figure 5(a) plots the results generated from 1000 random unit vectors, and Figure 5(b) is obtained from 1000 randomly selected PCA feature vectors in the experimental data set. The random vectors are generated with each element an i.i.d. Gaussian random variable, followed by normalization to unit length. The PCA vectors are normalized to have the same variance and unit length. The estimation $\hat{\mathbf{u}}$ of an original vector $\mathbf{u}$ is performed as follows. For an original vector $\mathbf{u}$ with RP matrix $\mathbf{R}$, we obtain the SIN vector $\mathbf{g}$ by $\mathbf{g} = \text{sort}(\mathbf{R}^T\mathbf{u})$, where sort denotes the operation of getting the sorted index numbers. Given the worst case that an adversary obtains $\mathbf{g}$ and $\mathbf{R}$, he can estimate $\mathbf{u}$ by using a randomly generated unit vector $\mathbf{e}$ according to an i.i.d. Gaussian distribution, mapping to the estimated vector $\hat{\mathbf{e}}$ based on $\mathbf{g}$, then computing $\hat{\mathbf{u}} = \mathbf{R}\mathbf{R}^T\hat{\mathbf{e}}$, and normalizing to unit length. It can be observed from Figure 5 that both the element-wise and vector-wise privacy levels improve as the projected dimension decreases.

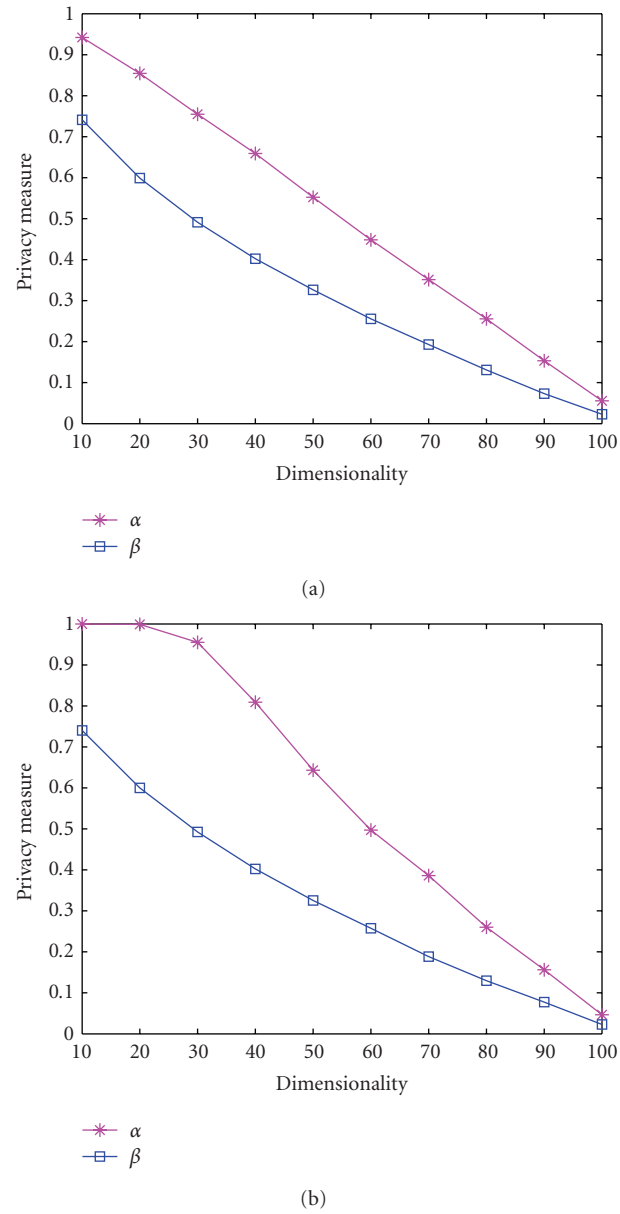


Figure 5: Privacy measure as a function of dimensionality. (a) random vectors, (b) PCA feature vectors.

To provide some insight into the privacy protection property of the proposed method, we compare the reconstructed image with the original image through different methods in Figure 6. The images are randomly selected from the FERET database [36, 37]. A PCA vector $\mathbf{u}$ is first extracted from image $\mathbf{z}$ (Figure 6(a)). A new vector is then generated by $\mathbf{x} = R_u^T\mathbf{u}$, where R_u is a random projection matrix, and the sorted index numbers of $\mathbf{x}$ are stored in a SIN vector $\mathbf{g}$. Here the dimensionality of PCA is selected as $n = 100$, and the projection dimension is $m = 50$. Assuming the worst case that $\mathbf{g}$ and R_u are all compromised, an adversary can only reconstruct the original image based on a vector $\mathbf{v}$,

which is either a PCA feature vector of some other subjects, or a randomly generated vector. The reconstruction can be performed by first sorting and mapping $\mathbf{v}$ to another vector $\tilde{\mathbf{v}}$ based on $\mathbf{g}$, and followed by $\hat{\mathbf{z}} = \Psi(R_u\tilde{\mathbf{v}} + \Psi^T\bar{\mathbf{z}})$. Figure 6(a) shows an original image $\mathbf{z}$ and Figure 6(b) is the reconstructed image from its first 100 PCA coefficients $\mathbf{u}$. The reconstruction is performed by $\hat{\mathbf{z}} = \Psi(\mathbf{u} + \Psi^T\bar{\mathbf{z}})$, where Ψ is the PCA projection matrix, and $\bar{\mathbf{z}}$ is the mean image obtained from the training set. It is obvious that the PCA approach cannot preserve privacy since the original visual information is very well approximated. Figures 6(d) and 6(f) are the reconstructed images from the features of images, Figures 6(c) and 6(d), respectively, while Figure 6(g) and Figure 6(h) are reconstructed from randomly generated vectors, all using the SIN vector $\mathbf{g}$ of image Figure 6(a). All the reconstructed images demonstrate large distortion from the original image. The results in Figure 6 are meant to provide some insight into the privacy preserving property of the proposed method. It can be seen that the original values of the feature vectors can not be recovered, and an estimation can only produce a distorted version of the original image, which has a significant visual difference from the original one. The above analysis, although not exact in the mathematical sense, illustrates that the privacy of the user can be protected by using the proposed method.

4. Experimental Results

To evaluate the performance of the proposed method, we conducted experiments on a generic data set that consists of face images from several well-known databases [38]. In this section, we first give a description of the employed data set. The adopted feature extraction methods are then briefly discussed. Finally, the experimental results along with detailed discussion are presented.

4.1. Generic Data Set. To approach more realistic face recognition applications, this paper tests the effectiveness of the proposed method using a generic data set, in which the intrinsic properties of the human subjects are trained from subjects other than those to be recognized. The generic database was initially organized for the purpose of demonstrating the effectiveness of the generic learning framework [38]. It originally contains 5676 images of 1020 subjects from 5 well-known databases, FERET [36, 37], PIE [39], AR [40], Aging [41], and BioID [42]. All images are aligned and normalized based on the coordinate information of some facial feature points. The details of image selection can be found in [38].

For preprocessing, the color images are first transformed to gray-scale images by taking the luminance component in YC_bC_r color space. All images are preprocessed according to the recommendation of the FERET protocol, which includes: (1) images are rotated and scaled so that the centers of the eyes are placed on specific pixels and the image size is 150×130; (2) a standard mask is applied to remove non-face portions; (3) histogram equalized and image normalized to have zero mean and unit standard deviation. After preprocessing, the face images are converted to an image vector of dimension $J = 17154$. Table 1 illustrates the configuration of the whole data set. Figure 7 shows some example images from the generic data set.

Table 1: Generic data set configuration.

Database	No. of subjects	No. of images per subject	No. of images
FERET	750	≥3	3881
AR	119	4	476
Aging	63	≥3	276
BioID	20	≥6	227
PIE	68	12	816
Total	1020	≥3	5676

4.2. Feature Extraction. To study the effects of different feature extractors on the performance of proposed methods, we compare Principal Component Analysis (PCA) and Kernel Direct Discriminant Analysis (KDDA). PCA is an unsupervised learning technique which provides an optimal, in the least mean square error sense, representation of the input in a lower-dimensional space. In the Eigenfaces method [43], given a training set $\mathcal{Z} = \{\mathcal{Z}_i\}_{i=1}^{C}$, containing C classes with each class $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{C_i}$ consisting of a number of face images $\mathbf{z}_{ij}$, a total of $M = \sum_{i=1}^{C} C_i$ images, the PCA is applied to the training set $\mathcal{Z}$ to find the M eigenvectors of the covariance matrix

$$\mathbf{S}_{\text{cov}} = \frac{1}{M}\sum_{i=1}^{C}\sum_{j=1}^{C_i}\left(\mathbf{z}_{ij} - \bar{\mathbf{z}}\right)\left(\mathbf{z}_{ij} - \bar{\mathbf{z}}\right)^T, \tag{9}$$

where $\bar{\mathbf{z}} = (1/M)\sum_{i=1}^{C}\sum_{j=1}^{C_i}\mathbf{z}_{ij}$ is the average of the ensemble. The Eigenfaces are the first $N(\leq M)$ eigenvectors corresponding to the largest eigenvalues, denoted as Ψ. The original image is transformed to the N-dimension face space by a linear mapping

$$\mathbf{y}_{ij} = \Psi^T\left(\mathbf{z}_{ij} - \bar{\mathbf{z}}\right). \tag{10}$$

PCA produces the most expressive subspace for face representation but is not necessarily the most discriminating one. This is due to the fact that the underlying class structure of the data is not considered in the PCA technique. Linear Discriminant Analysis (LDA) is a supervised learning technique that provides a class specific solution. It produces the optimal feature subspace in such a way that the ratio of between-class scatter and within-class scatter is maximized. Although LDA-based algorithms are superior to PCA-based methods in some cases, it is shown in [44] that PCA outperforms LDA when the training sample size is small and the training images is less representative of the testing subjects. This is confirmed in [38] that PCA performs much better than LDA in a generic learning scenario, where the image samples of the human subjects are not available for training. It was also shown in [38] that KDDA outperforms other techniques in most of the cases. Therefore we also adopt KDDA in this paper.

Figure 6: Comparison of original image with reconstructed images.

Figure 7: Example images for identification (top row) and verification (bottom row).

KDDA was proposed by Lu et al. [45] to address the nonlinearities in complex face patterns. Kernel-based solution find a nonlinear transform from the original image space $\mathcal{R}^J$ to a high-dimensional feature space $\mathcal{F}$ using a nonlinear function $\phi(\cdot)$. In the transformed high-dimensional feature space $\mathcal{F}$, the convexity of the distribution is expected to be retained so that traditional linear methodologies such as PCA and LDA can be applied. The optimal nonlinear discriminant feature representation of $\mathbf{z}$ can be obtained by

$$\mathbf{y} = \Theta \cdot \nu(\phi(\mathbf{z})), \tag{11}$$

where Θ is a matrix representing the found kernel discriminant subspace, and $\nu(\phi(\mathbf{z}))$ is the kernel vector of the input $\mathbf{z}$. The detailed implementation algorithm of KDDA can be found in [45].

4.3. Experimental Results on Face Identification. For face identification, we use all the 5676 images in the data set for experiments. A set of 2836 images from 520 human subjects was randomly selected for training, and the rest of 2840 images from 500 subjects for testing. There is no overlap between the training and testing subjects and images. The test is performed on an exhaustive basis, such that each time, one image is taken from the test set as probe image, while the rest of the images in the test set as gallery images. This is repeated until all the images in the test set were used as the probe once. The classification is based on nearest neighbor.

Table 2 compares the correct recognition rate (CRR) of SIN method with Euclidean and Cosine distance measures at different dimensions. It can be observed that at higher dimensionality, the SIN method may boost the recognition accuracy of PCA significantly, while maintain the good performance of the stronger feature extractor KDDA. The PCA method projects images to directions with highest variance, but not the discriminant ones. This will become more severe in large image variations due to illumination, expression, pose, and aging. When computing the similarity between two PCA vectors, the distance measure is sensitive to the variation of individual element, particularly those directions corresponding to noise. The SIN method, on the other hand, reduces this sensitivity by simply comparing the relative

Table 2: Face identification results (in %).

	PCA			KDDA		
Dim.	Euc.	Cos.	SIN	Euc.	Cos.	SIN
20	56.30	56.31	52.32	40.04	41.09	34.86
40	60.09	61.09	61.94	61.44	65.28	61.94
60	63.52	62.96	66.06	71.73	74.86	74.68
80	64.37	64.44	68.84	81.76	83.27	81.76
100	65.14	65.18	71.27	79.05	80.42	80.07

Table 3: Verification data set configuration.

Database	No. of subjects	No. of images per subject	No. of images
FERET	750	≥2	3029
AR	119	4	476
Aging	63	≥3	276
BioID	20	≥6	227
PIE	68	≥8	658
Total	1020	≥2	4666

Table 4: Obtained EER (in %) for face verification.

	PCA			KDDA		
Dim.	Euc.	Cos.	SIN	Euc.	Cos.	SIN
20	20.05	19.23	13.78	25.22	20.42	20.97
40	19.09	17.81	11.46	21.49	16.22	14.54
60	18.52	17.42	10.28	18.80	13.41	10.97
80	18.50	17.15	9.72	10.96	9.90	7.19
100	18.20	16.94	9.46	10.41	8.84	6.52

relation of the projections, and therefore possibly provides better error tolerance. In the case of strong extractors such as KDDA, the SIN method will approximate the distance between two vectors and hence preserves the recognition accuracy.

4.4. Experimental Results on Face Verification. For face verification, we exclude image samples with large pose variation (>15°) and select 4666 images from 1020 subjects for our experiments. Table 3 illustrates the detailed configuration of the verification data set. In our experiments, we randomly select 2388 images from 520 subjects as the training set, and 2278 images of the rest 500 subjects as the testing set. There is no overlap between the training and the testing subjects and images. The evaluation was also performed on an exhaustive basis, where every single image is used as a template once, and the rest of the images in the test set as the probe images.

Table 4 compares the obtained equal error rate (EER) of SIN with Euclidean and Cosine distance at different dimensions when PCA and KDDA are used as feature extractors. In general, the Cosine metric outperforms the Euclidean distance measure, and the proposed SIN method improves both the verification accuracy of PCA and KDDA at almost all dimensions. This further demonstrates that the sorted index numbers indeed offer better error tolerance and provide more discriminant representation.

4.5. Changeable Face Verification. To enhance the privacy protection level as well as addressing the irrevocable problem of biometric verification systems, this paper adopts the random projection method. For the purpose of comparative study, we compared the performance of the proposed method with that of the BioHashing (BH) technique in this paper. For the BH method, as illustrated in [24], each of the generated BH code should have a probability of 50% to be 1 or 0. To achieve this, we centralize all the feature vectors by subtracting the mean, and then compare with the threshold value $t = 0$.

In the experiments, the same data set as the one for face verification is employed. The images for training and testing are also exactly the same as those for face verification. To minimize the effect of randomness, all the experiments were performed 5 times, and the average of the results is reported. Table 5 gives the obtained EER of BH and SIN methods in both user-independent and user-dependent scenarios at different projected dimension m, with the dimensionality of the original features set to $n = 100$.

In the user-independent scenario, all the users apply the same RP matrix. In the user-dependent scenario, different users have distinct RP matrices. The user-dependent scenario is essentially a two-factor scheme, and it requires correct presentation of both the RP matrix (or a generation key) and biometrics data. The proposed user-dependent scheme assumes that the RP matrix and the biometrics data can not be stolen at the same time. If the RP matrix is stolen, the evaluation can be performed by considering the worst case that the key of all the users is stolen by others. This is equivalent to use the same random projection matrix for all the users. Therefore, the performance of stolen key case will be the same as the user-independent scenario. If only the biometric data is stolen, then the performance will be the same as the both-legitimate case due to the randomness of the transformation, as discussed in Section 3.3.

The experimental results in Table 5 show that the proposed SIN method outperforms the BH method in both user-dependent and user-independent scenarios, at all dimensions, when PCA and KDDA are used as feature extractors. Although the previous works on BH demonstrate near zero EER in both-legitimate cases, the performance of it depends on the characteristics of the data and feature extractors. For an m bit BioHash code $\mathbf{b}$, assume that each bit in $\mathbf{b}$ is independent, let τ be the threshold value in terms of Hamming distance, then the probability of false accept $P(\mathbf{H}_0 \mid \mathbf{H}_1) = \sum_{i=0}^{\tau} \binom{m}{i} / 2^n$. This probability depends on two factors, the system threshold τ and dimension m, which reflect the separability and characteristics of the data and feature extractors. Figure 8 shows the intra-class and inter-class distribution of the generic data set. It can be observed that the SIN method provides better distribution separation than the BH method, in both user-independent and user-dependent scenarios, with both PCA and KDDA feature extractors. This demonstrates that the proposed SIN method

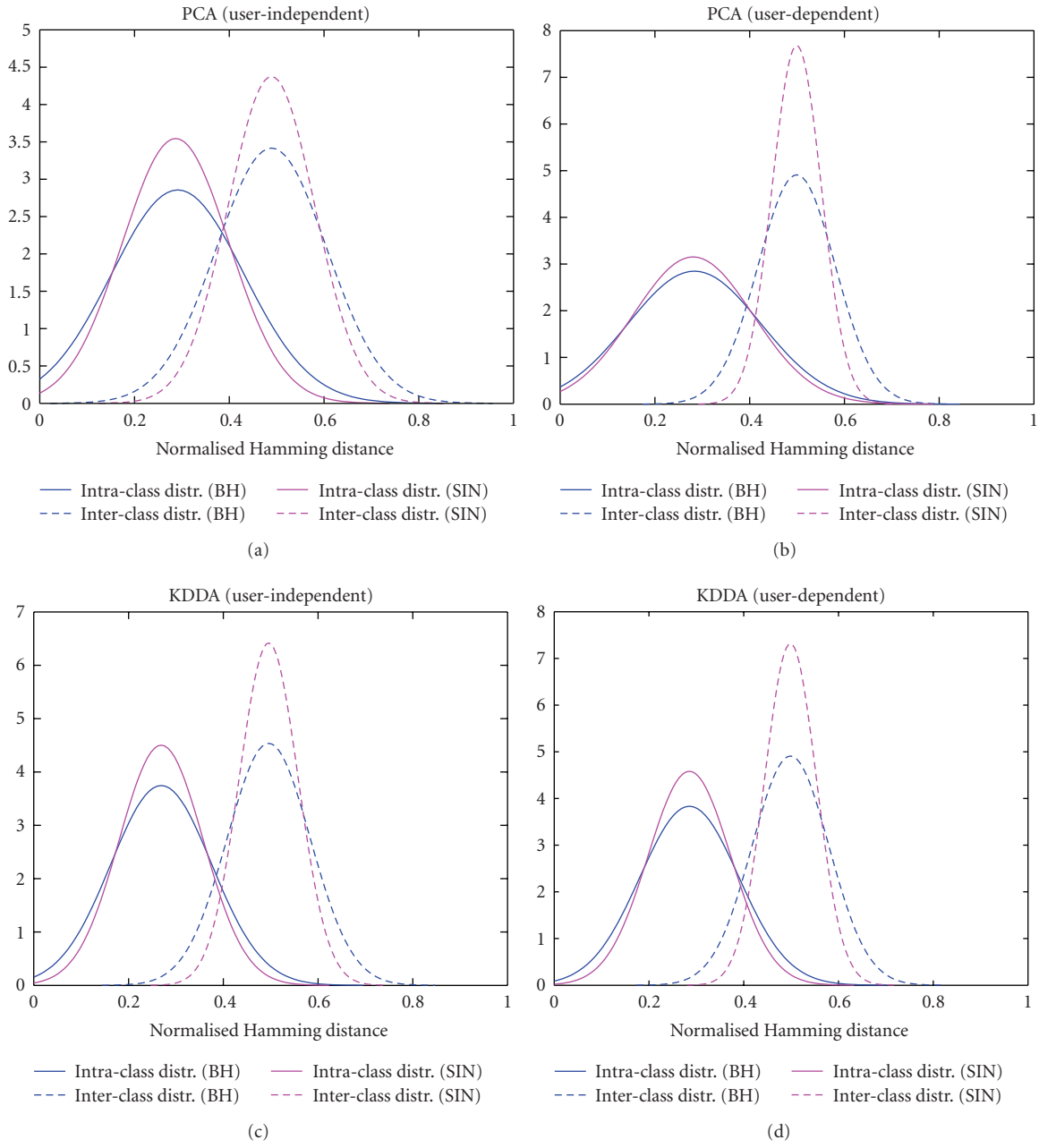


Figure 8: Intra-class and inter-class distributions of SIN and BH, using PCA and KDDA feature extractors, in both user-independent and user-dependent scenarios.

Table 5: Obtained EER (in %) for changeable face verification.

	PCA				KDDA			
	User-dependent		User-independent		User-dependent		User-independent	
Dim.	BH	SIN	BH	SIN	BH	SIN	BH	SIN
20	22.13	16.92	25.25	20.82	18.77	12.96	18.63	13.58
40	17.80	13.44	21.43	18.69	13.03	7.70	13.96	9.23
60	15.54	11.76	19.24	17.63	9.85	5.68	10.92	7.38
80	14.38	10.76	18.34	17.18	7.97	4.54	9.37	6.64
100	12.98	9.89	17.79	16.83	6.84	3.83	8.63	6.05

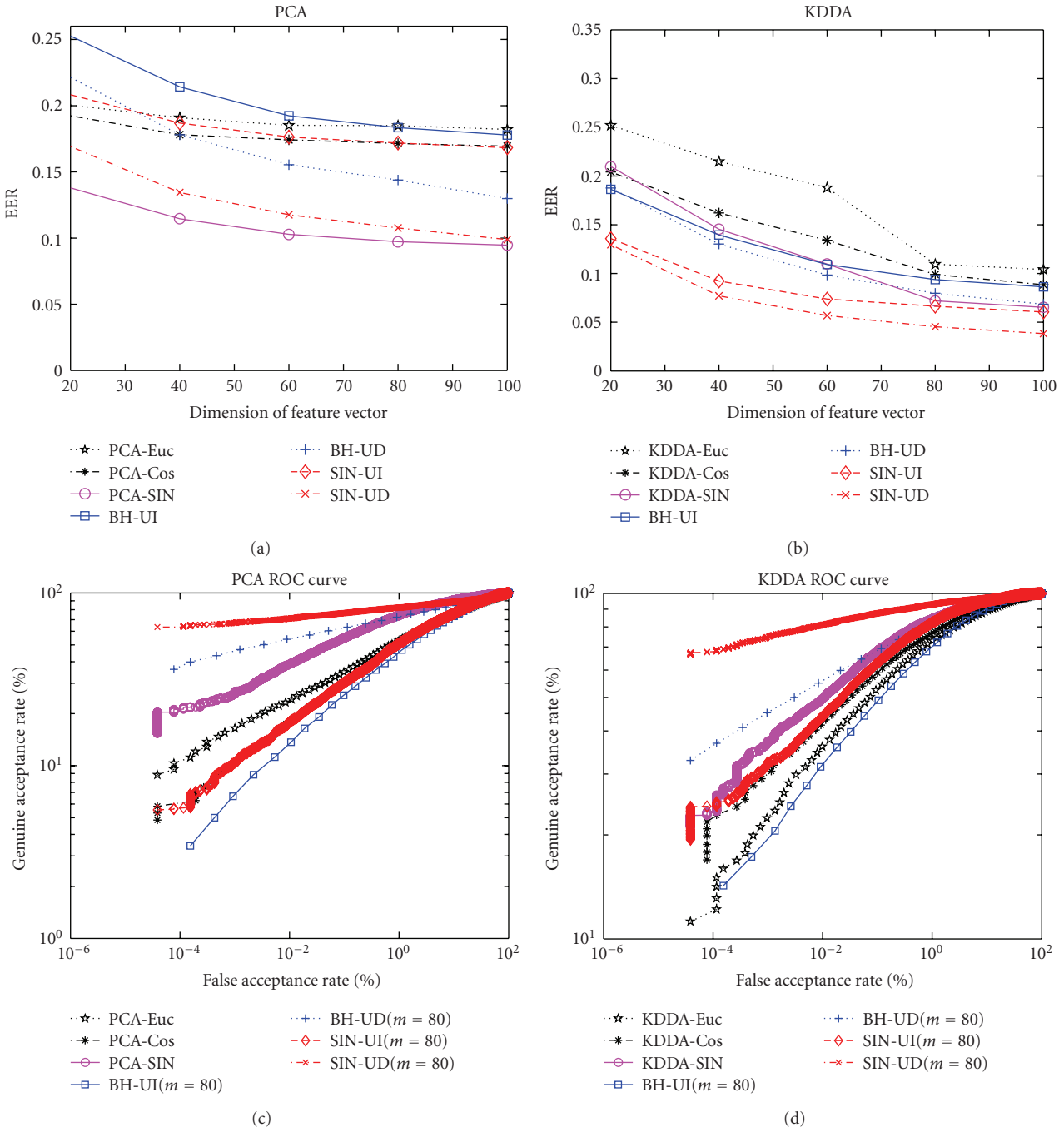


Figure 9: Obtained EER and ROC plots for PCA and KDDA (UI: user-independent, UD: user-dependent).

provides more discriminant representation than the simple thresholding method in BioHashing.

For a complete comparison, Figure 9 plots the EER of all verification scenarios as well as the Receiver Operating Curve (ROC) for both feature extractors at dimensionality of 100. The ROC curve is plotted by Genuine Acceptance Rate (GAR, complement of FRR) against FAR, and the axes are log scaled for better visualization. When the SIN method is applied on facial features directly, it improves the verification accuracy for both feature extractors. In the user-independent scenario of PCA, the verification accuracy is degraded compared to apply SIN directly on PCA features. This is possibly due to the randomness of RP changes the inherent pairwise relations of original PCA features, and therefore the SIN method can not produce more discriminant representation, but approximate the Euclidean distance only. In spite of this, it can be observed that by integrating the RP transform, the proposed SIN method introduces changeability, enhances privacy protection, and achieves better performance than original features, as well as existing work.

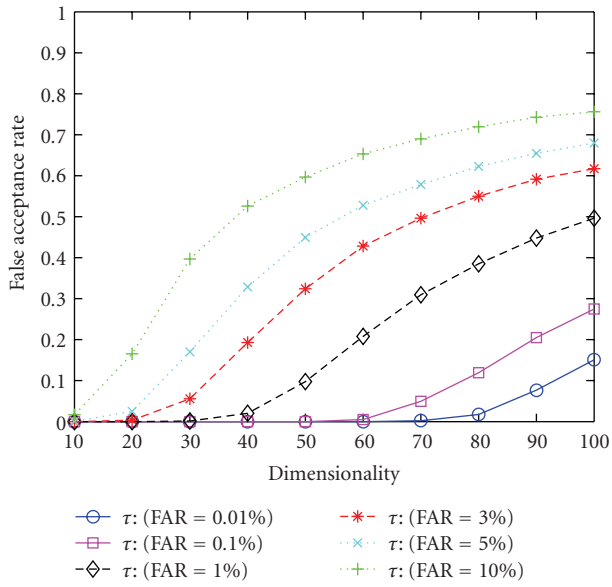


Figure 10: Experimental results based on reconstructed images.

4.6. Experimental Results on Reconstructed Images. To further study the privacy preserving property of the proposed method, we performed experiments on reconstructed images from the estimated PCA coefficients. The original n-dimensional PCA features are projected onto an m-dimensional vector, and the resulting SIN vector is stored as templates. Considering the worst case that the SIN vector, the random projection matrix, the PCA transformation matrix, and the mean image are all obtained, an adversary can reconstruct the original image using the method discussed in Section 3.4. The adversary may then try to compromise the user using the reconstructed image. Figure 10 reports the false acceptance rate obtained when the reconstructed images are utilized to compromise the original PCA-based system. The dimensionality of the PCA vectors is $n = 100$. All the PCA vectors are normalized to unit length, and Euclidean distance is adopted as dissimilarity measure. The system threshold values are selected based on the FAR of the original system. It can be observed that the false acceptance rate decreases as the projection dimension m decreases. This is consistent with our analysis in Section 3.4 that the privacy preserving level increases as m decreases. It can be also seen that the security level is also dependent on the system threshold of the original system, which is closely related to the requirement of the application. In general, applications that require a higher level of security will have a smaller threshold, that is, smaller FAR. In such, the proposed method can provide stronger privacy protection even at a relatively higher projected dimension. On the other hand, when the τ is large, it requires smaller projected dimension m to achieve higher level of security. However, as shown in Figure 9, since the recognition accuracy also degrades as the m getting smaller, the proposed method has a tradeoff between privacy and accuracy. The balancing point of these two is dependent on the requirement of the application.

5. Conclusion

This paper introduced a novel approach for addressing the challenging problem of changeable and privacy preserving face recognition. The proposed method is based on random projection (RP) in conjunction with a sorted index numbers (SINs) approach. A similarity measure is introduced for computing the distance between two SIN vectors. Two different scenarios, namely, user-independent and user-dependent transformation are discussed. In the user-independent scenario, all the users apply the same RP matrix for transformation. Due to the distance preserving property of RP, the similarity of features in the transformed domain can be approximately preserved. The user-dependent scenario is a two-factor authenticator that utilizes user-specific RP matrix for transformation. In both scenarios, the biometrics template can be changed by varying the RP matrix.

Experimental results on a large database demonstrate that the SIN method may improve the recognition accuracy of the original features in both identification and verification scenarios. The combination of RP and SIN method outperforms comparable existing works for all scenarios and feature extractors. In conclusion, the proposed method may improve recognition accuracy, preserve the user's privacy, and generate changeable biometric template. Although we focus on face recognition problem in this paper, the proposed method is general for continuous domain features, and it is expected that such method can also be applied to other biometrics.

Acknowledgment

Y.Wang would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support.

References

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

[2] R. M. Bolle, J. H. Connell, and N. K. Ratha, "Biometric perils and patches," *Pattern Recognition*, vol. 35, no. 12, pp. 2727–2738, 2002.

[3] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: issues and challenges," *Proceedings of the IEEE*, vol. 92, no. 6, pp. 948–959, 2004.

[4] A. Adler, "Vulnerabilities in biometric encryption systems," in *Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '05)*, vol. 3546, pp. 1100–1109, Tarrytown, NY, USA, July 2005.

[5] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proceedings of the Biometrics Symposium (BSYM '07)*, Baltimore, Md, USA, September 2007.

[6] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[7] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and W. Worek, "Preliminary face recognition grand challenge results," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR '06)*, pp. 15–24, Southampton, UK, April 2006.

[8] G. Shakhnarovich and B. Moghaddam, "Face recognition in subspaces," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., Springer, New York, NY, USA, December 2004.

[9] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. K. Vijaya Kumar, "Biometric encryption," in *ICSA Guide to Cryptography*, McGraw-Hill, New York, NY, USA , 1999.

[10] G. I. Davida, Y. Frankel, and B. J. Matt, "On enabling secure applications through off-line biometric identification," in *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 148–157, Oakland, Calif, USA, May 1998.

[11] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proceedings of 6th Conference on Computer and Communication Security (ACM '99)*, pp. 28–36, Singapore, November 1999.

[12] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081–1088, 2006.

[13] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proceedings of the IEEE International Symposium on Information Theory*, p. 408, Lausanne, Switzerland, June 2002.

[14] R. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure smart card based fingerprint authentication," in *Proceedings of the ACM SIGMM Workshop on Biometrics Methods and Applications*, pp. 45–52, Berkley, Calif, USA, November 2003.

[15] U. Uludag, S. Pankanti, and A. Jain, "Fuzzy vault for fingerprints," in *Proceedings of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '05*, vol. 3546, pp. 310–319, Hilton Rye Town, NY, USA, July 2005.

[16] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: how to generate strong keys from biometrics and other noisy data," in *Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT '04)*, pp. 523–540, Interlaken, Switzerland, May 2004.

[17] Y. Sutcu, Q. Li, and N. Memon, "Protecting biometric templates with sketch: theory and practice," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 503–511, 2007.

[18] X. Boyen, "Reusable cryptographic fuzzy extractors," in *Proceedings of the 11th ACM Conference on Computer and Communications Security (CCS '04)*, pp. 82–91, Washington, DC, USA, October 2004.

[19] T. A. M. Kevenaar, G. G. Schrijen, M. van der Veen, A. H. M. Akkermans, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID '05)*, pp. 21–26, Buffalo, NY, USA, October 2005.

[20] M. Savvides, B. V. K. Vijaya Kumar, and P. K. Khosla, "Cancelable biometric filters for face recognition," in *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 922–925, Cambridge, UK, August 2004.

[21] M. Savvides, B. V. K. Vijaya Kumar, and P. K. Khosla, "Authentication-invariant cancellable biometric filters for illumination-tolerant face verification," in *Proceedings of the IEEE International Conference Cancellable Biometric Filters for Face Recognition*, vol. 5404 of *Proceedings of SPIE*, pp. 156–163, Los Alamitos, Calif, USA, 2004.

[22] T. E. Boult, "Robust distance measures for face recognition supporting revocable biometric tokens," in *Proceedings of the 7th IEEE Conference on Face and Gesture*, Southampton, UK, April 2006.

[23] T. E. Boult, W. J. Scheirer, and R. Woodworth, "Revocable fingerprint biotokens: accuracy and security analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, June 2007.

[24] A. B. J. Teoh, D. C. L Ngo, and A. Goh, "BioHashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition*, vol. 37, pp. 2245–2255, 2004.

[25] T. Connie, A. Teoh, M. Goh, and D. Ngo, "PalmHashing: a novel approach for dual-factor authentication," *Pattern Analysis and Applications*, vol. 7, no. 3, pp. 255–268, 2004.

[26] D. C. L. Ngo, A. B. J. Teoh, and A. Goh, "Biometric hash: high-confidence face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 771–775, 2006.

[27] A. Lumini and L. Nanni, "An improved BioHashing for human authentication," *Pattern Recognition*, vol. 40, no. 3, pp. 1057–1065, 2007.

[28] Y. Wang and K. Plataniotis, "Face based biometric authentication with changeable and privacy preserving templates," in *Proceedings of the Biometrics Symposium (BSYM '07)*, Baltimore, Md, USA, September 2007.

[29] S. Xiang, H. Kim, and J. Huang, "Histogram-based image hashing scheme robust against geometric deformations," in *Proceedings of the 9th Workshop on Multimedia and Security*, pp. 121–128, Dallas, Tex, USA, September 2007.

[30] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipshitz mapping into Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.

[31] N. Goel and G. Bebis, "Face recognition experiments with random projection," in *Proceedings of the Defense and Security Symposium (DSS '05)*, vol. 5779 of *Proceedings of SPIE*, pp. 426–437, Orlando, Fla, USA, March 2005.

[32] E. Brigham and H. Maninila, "Random projection in dimensionality reduction: applications to image and text data," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. C245–C250, San Francisco, Calif, USA, August 2001.

[33] K. Liu, H. Kargupta, and J. Ryan, "Random projection-based multiplicative data perturbation for privacy preserving distributed data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 92–106, 2006.

[34] K. Muralidhar, R. Parsa, and R. Sarathy, "A general additive data perturbation method for database security," *Management Science*, vol. 45, no. 10, pp. 1399–1415, 1999.

[35] J. T. Wang, X. Wang, K. I. Lin, D. Shasha, B. A. Shapiro, and K. Zhang, "Evaluating a class of distancemapping algorithms for data mining and clustering," in *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 307–311, San Diego, Calif, USA, August 1999.

[36] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing Journal*, vol. 16, no. 5, pp. 295–306, 1998.

[37] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[38] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognition*, vol. 39, pp. 1746–1762, 2006.

[39] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.

[40] A. M. Martinez and R. Benavente, "The AR face database," CVC Technical report 24, 1998.

[41] Aging Database,http://www.fgnet.rsunit.com/.

[42] BioID Database, http://www.humanscan.de/support/downloads/facedb.php.

[43] M. Turk and A. Pentland, "EigenFaces for recognition," *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71–86, 1991.

[44] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[45] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 980159, 13 pages
doi:10.1155/2009/980159

Research Article

A New User Dependent Iris Recognition System Based on an Area Preserving Pointwise Level Set Segmentation Approach

Nakissa Barzegar and M. Shahram Moin

Multimedia Systems Research Group, Iran Telecom Research Center, IT Faculty, Tehran 14 399 55471, Iran

Correspondence should be addressed to Nakissa Barzegar, barzegar@itrc.ac.ir

Received 30 September 2008; Revised 4 January 2009; Accepted 11 March 2009

Recommended by Kevin Bowyer

This paper presents a new user dependent approach in iris recognition systems. In the proposed method, consistent bits of iris code are calculated, based on the user specifications, using the user's mask. Another contribution of our work is in the iris segmentation phase, where a new pointwise level set approach with area preserving has been used for determining inner and outer iris boundaries, both exclusively performed in one step. Thanks to the special properties of this segmentation technique, there is no constraint about angles of head tilt. Furthermore, we showed that this algorithm is robust in noisy situations and can locate irises which are partly occluded by eyelid and eyelashes. Experimental results, on three renowned iris databases (CASIAIrisV3, Bath, and Ubiris), show that our method outperforms some of the existing methods, both in terms of accuracy and response time.

1. Introduction

The demand for high-confidence authentication of human identity has grown steadily since the beginning of organized society. The identification systems using unique factors of human irises play an important role in this field. In comparison with other biometrics, iris recognition systems have many advantages. Since the degree of freedom of iris textures is extremely high, the probability of finding two identical irises is close to zero; therefore, the iris recognition systems are very reliable and could be used in most secure places [1–3].

A regular iris recognition system consists of different major steps, including image acquisition, iris localization, feature extraction, and matching and classification. In this paper, we have used standard iris datasets; therefore, we have not focused on the image acquisition phase. Other parts of an iris recognition system will be discussed later.

One of the most important steps in iris recognition systems is iris localization, which is related to the detection of the exact location and contour of iris in an image. Obviously, the performance of the identification system is closely related to the precision of the iris localization step [1, 2]. For iris localization, existing methods mainly use circular edge detectors or other standard image processing techniques, to detect the iris location based on derivative operators, which calculate the sum of gray level differences on the vertical arc. It must be mentioned that, since the upper and lower parts of the outer iris boundary are usually obstructed by eyelids, it could be impossible to use a complete circle, instead of two vertical arcs, to represent the iris boundaries. In these methods, the result of localization algorithm depends on the tilt angle of the iris and the quality of the boundaries [1, 2, 4]. For example, if some parts of boundaries are occluded by the eyelid and eyelashes, performance of these algorithms reduces considerably and even in some cases, they fail. Another source of error is the presence of other parts of face in input image.

In [1], Daugman introduces a circular edge detection operator for iris localization, which tries to find a circle in the image with maximum gray level differences with its neighbors. In its method, thanks to a significant contrast between iris and purple regions, the inner boundary is localized. Then, outer boundary is detected using the same operator with different radii and parameters. In order to remove eyelids, Daugman changes the curve of integral to

find an arc which accurately detects iris boundaries. As features, he uses the sign of real and imaginary parts of Gabor Wavelet coefficients of iris image. In matching phase, Hamming distance between binary codes of the query iris and irises in database is calculated. In his recent work [5], Daugman proposed four modifications in his algorithm, including (1) using active contour models (Snake model) for iris localization, (2) handling off-axes gaze samples using Fourier-based methods, (3) using statistical methods for detecting eyelashes, and (4) score normalization in large number databases.

An alternative for iris segmentation and localization has been proposed by Camus and Wildes [3], which is based on edge detection operator, followed by Hough transform. This method has a high computational cost, since it searches among all of the potential candidates. For eyelid detection, Wildes uses some constrains to locate the true edge points.

Snake approach has been used for iris localization in [6]. Using this technique, the boundary of the irises is located without any circularity constraint. In [7], an easy to difficult method has been used for iris localization by, first, determining high-contrast parts of boundary, and then, detecting outer boundary and eyelids. It is obvious that, because of their lower SNR, each step is more challenging than previous ones. For exact inner boundary detection, authors used Harr Wavelet transform followed by modified Hough transform. In the next step, outer boundary is localized with integral differential operators. Since the search space for determining the center and radius of inner boundaries could be limited, the speed of the algorithm is considerably improved. In the last step, for detecting eyelids in the image, a method is utilized based on texture segmentation.

Sun et al. [8] proposed iris localization using texture segmentation. First, they use the information of low frequency of Wavelet transform of iris image for pupil segmentation and also localize the iris with a different integral operator. Then, they detect the upper eyelid next to eyelash segmentation. Finally, the lower eyelid is localized using parabolic curve fitting, based on gray level segmentation.

Huang et al. [9] used a new noise removing approach based on the fusion of edge and region information. The whole procedure includes three steps: rough localization and normalization, edge information extraction based on phase congruency, and the infusion of edge and region information. They proceeded to iris segmentation by simple filtering, edge detection, and Hough transform. This method is specifically proposed for removing eyelash and pupil noises. Boles and Boashah [10] and Lim et al. [11] mainly focused on the iris image representation and feature matching without introducing a new method for segmentation.

Tisse et al. [12] proposed a segmentation method based on integro-differential operators with Hough transform. This approach reduces the computation time and excludes potential centers outside of the eye image. Eyelash and pupil noise have not been considered in this method neither.

Kong and Zhang in [13] presented a method for eyelash detection. Separable and multiple eyelashes are detected using 1D Gabor filters and the variance of intensity, respectively. In this work, specular reflection regions in the eye image are localized using a predetermined threshold value. Thornton et al. [14] used a general probabilistic framework for matching patterns of irises, which improves pattern matching performance, when the iris tissue is subject to in-plane wrapping.

Monro et al. in [15] present a novel iris coding algorithms based on differences of Discrete Cosine Transform (DCT) coefficients of overlapped angular patches with normalized iris image. Iris localization is done using the circularity shape of iris boundaries.

Other methods exist for iris localization, including [12, 16]. However the above mentioned techniques are much more cited in literature. There are also a few papers which survey literature in iris recognition subject; amongst them, Bowyer et al. [2] is one of the best.

We have used active contour based-localization method in [4]. In this paper, we improve our method and test its performance on three famous databases, namely, CASIA-IrisV3 [17], Bath [18], and Proença and Alexandre [19]. The results show the superiority of our proposed method in comparison with other methods, including the method proposed in [6], which is also based on geodesic active contour for iris localization. The details will be discussed in Section 2.

In [19], new approaches for localization have been introduced. In their paper, they use a dataset of irises with heterogeneous characteristics, simulating the dynamics of a noncooperative environment. Their method builds a feature set from pixel position (x, y) and pixel intensity z. They apply a fuzzy clustering algorithm to cluster the pixels. In Section 4 we compare our proposed method to their results.

Considering the above mentioned methods, we can state the following important remarks and drawbacks of existing methods.

(1) Usually, the iris inner and outer boundaries are detected using circle fitting techniques (except the recent works of Daugman [5] and Ross and Shah [6]). This is a source of error, since the iris boundaries are not exactly circles.

(2) In almost all of these methods, inner and outer boundaries, eyelashes, and eyelid are detected in different steps, causing a considerable increase in processing time of the system.

(3) The results of the circle fitting method are sensitive to the image rotation, particularly if the angular rotation of the input image is more than 10 degrees.

(4) In noisy situations, the outer boundary of iris does not have sharp edges.

(5) After detecting iris boundaries, the resulted iris area is mapped into a size independent rectangular shape area.

(6) None of these methods take into account the user specifications.

Considering these remarks, we propose a new user specific iris recognition system with the following contributions.

(i) We use a pointwise area preserving level set approach for iris localization, which guarantees the correct segmentation of iris, even in noisy environment and regardless of the head tilt and occlusion. Although active contours for localization have been also used in [5, 6], our proposed method has many advantages compared to those approaches (we will discuss these advantages in details in Section 2).

(ii) We propose a new user dependent method which improves the system recognition performance.

In [4], we explained how to use pointwise level set with area preserving capability for iris localization purposes. We have also introduced a method for mapping the initial coordinates to polar space based on the estimated location of the center of pupil. In this paper, in order to reduce the complexity of the polar mapping calculations, we propose the improved version of the above mentioned method, which is based on the point trajectory of moving contours. We show the results of the new method on CASIA-IrisV3, Bath, and Ubiris datasets.

The rest of the paper is organized as follows. Section 2 briefly describes the theory of pointwise level set approach with area preserving capability. Section 3 is dedicated to the user dependency in iris recognition systems. Experimental results are presented in Section 4 and Section 5 concludes the paper.

2. Iris Localization with Pointwise Level Set Approach

In this approach, the moving front is defined as a zero level of a higher dimensional potential function [20]. Consequently, the curve corresponding to the zero level set of this potential function is enabled to handle topological changes, such as splitting and merging. Furthermore, it is not necessary to initialize the algorithm very close to the final contours, which is the case of Snakes model. According to the level set model, the initial curve is deformed using the following evolutionary equation:

$$\frac{dC}{dt} = V\vec{N}, \tag{1}$$

where V is any intrinsic quantity and does not depend on parameters, N is the normal vector, and C, as the implicit representation of the curve, is defined as

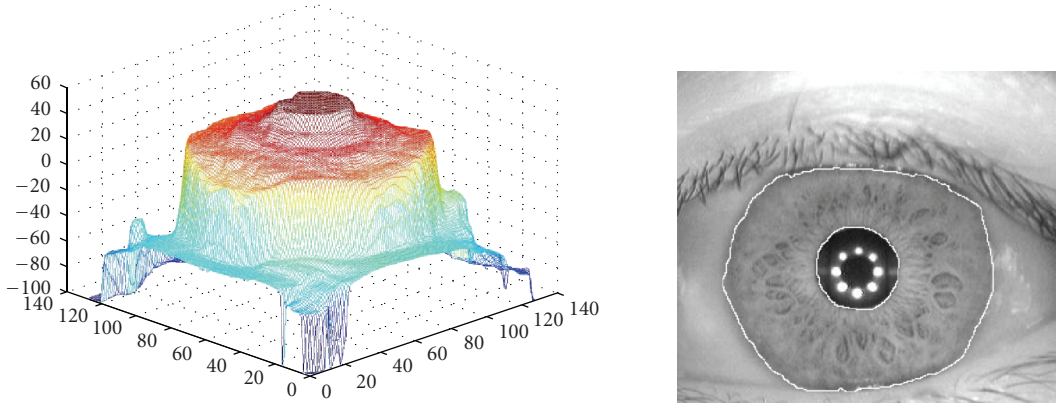
$$C = \{(x, y) : \varphi(x, y) = 0\} : \varphi(x, y) : R^2 \longrightarrow R. \tag{2}$$

A distance measure can be used for initializing the potential function φ. It means that each point of the three-dimensional potential function is initialized with the minimum distance of that point to the contours. More details on this subject are available in [20]. The evolution of φ is such that its zero levels movement corresponds to deformation of the initial curve. This evolution may be described by the following equation:

$$\frac{d\varphi}{dt} = V\,|\nabla\varphi|. \tag{3}$$

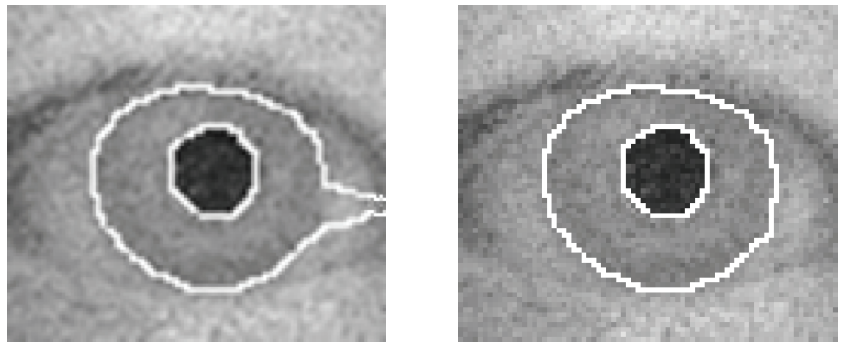
This equation shows that the rate of changes of the potential function φ in time depends on the speed parameter V and the magnitude of the gradient of φ. The speed V has three components: balloon force (which cause all part of contour to move), curvature-based speed, and gradient-based speed [20]. Due to the high performance of active contour-based models for localization purposes, some references in literature are based on these models [4–6]. As we mentioned briefly in Section 1, Daugman, in [5], proposed a method for iris segmentation using Snake model [21]. Despite of the Snakes advantages over the traditional object recognition algorithms, it has some important drawbacks, due to its Lagrangian-based formulas. In Snake model, contour initialization is a crucial point; thus, if the initial contour is far from the target, it may not reach the target. Another important disadvantage of this model is its performance reduction: due to point-based structure of the contour, some unwanted pixels can cause misjudgment of localization results. In order to solve these drawbacks, new models have been introduced based on Euler equations [20]. These models consider moving contours as a level set of a higher dimensional function, which reshape during the different iterations. Very briefly speaking, Euler equations connect the differentiations in time and space together [20]. Because of this capability, if noisy pixels cause some parts of contour to stop, other moving parts prevent the whole contour to stop. Another advantage of this approach is its robustness to contour initialization. Because of the combination of different forces, which cause movement in this approach, almost all kinds of initialization, lead to the same result (Figure 1).

Another related work is Ross and Shah in [6], who use geodesic active contour models for iris segmentation. The structures of geodesic active contour and level set methods are similar; therefore, both can handle noisy situations and initialization problems properly. The major difference between Ross's method and the method proposed in this paper is as follows. Due to the geodesic active contour's structure, it lacks the point correspondence property. Therefore, it is impossible to find the correspondent points in initial and final contours. We used point correspondent level set approach [22], which, in addition to level set's regular abilities, keeps point correspondence during the iterations [4]. This ability enables us to perform both localization and mapping to the dimensionless coordination phases in a single phase, an interesting property which improves the performance of the whole system. Another advantage of our proposed method, in comparison with Ross's work, is that, here, we use an area preserving method [23] for our level set methods, which make our method robust in case of blurred images. If the boundaries of an iris image are blurred, level set method is not able to determine the exact location of blurred parts of the boundaries to stop moving; whilst, in our proposed method, thanks to its area preserving property, even if some parts of boundaries are blurred, the whole contour prevents the unwanted local movement of the contour in blurred image. This property leads us to determine the exact target boundaries (Figure 2). This could be done by defining the application specific normal motion,



(a) (b)

Figure 1: (a) Three-dimensional function of level set approach, (b) Result of application of the zero level set method to an iris image taken from CASIA-IrisV3.



(a) (b)

Figure 2: Iris segmentation with noisy samples (a) without and (b) with area preserving capability.

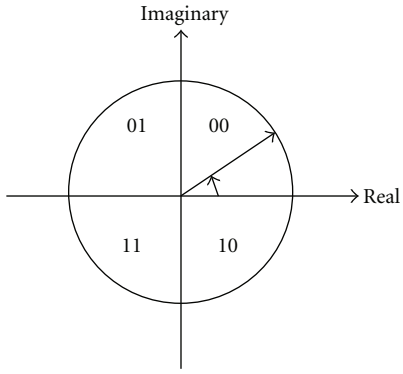


Figure 3: Real and imaginary axes and related binary codes.

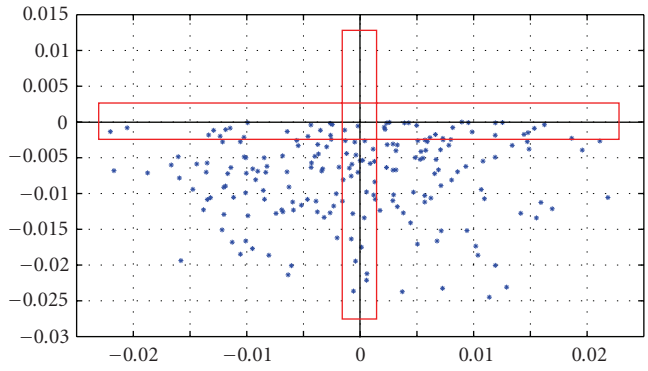


Figure 4: Iris features in the real/imaginary plane. The features near the axes are more inconsistent than others.

combining with adequate tangential speed. More details are available in [23]

3. Template Generating with User Dependency

According to Hallingsworth et al. in [24], it is possible to use weighted iris codes during the Hamming distance estimation phase. This means that different bits in an iris code do not have the same importance. Based on this idea, we propose a new user dependent method for iris recognition. After mapping the segmented area of the iris to the dimensionless polar coordinates, as it has been explained in Section 2, iris texture is transformed into a binary code, using the sign of real and imaginary parts of log Gabor Wavelet

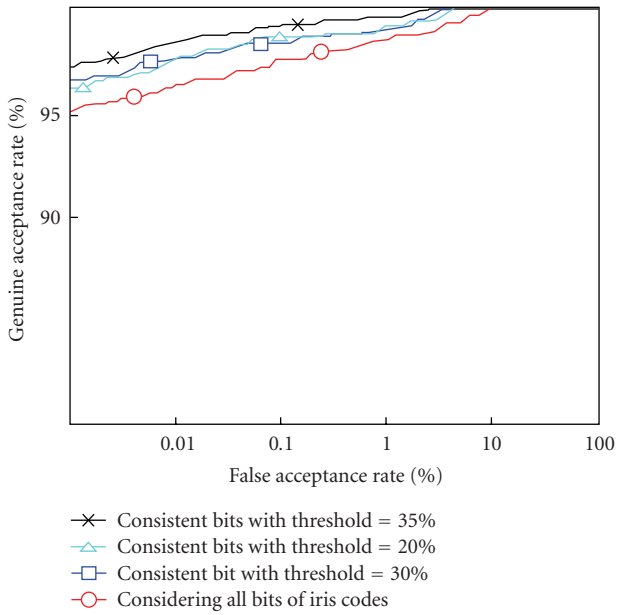


Figure 5: Comparison of ROC curves of our proposed method using all bits of iris code and using only the consistent bits with different thresholds. As it can be seen, the performance of system considering consistent bits with threshold equal to 35% is the best. (Tests using CASIA-IrisV3 dataset).

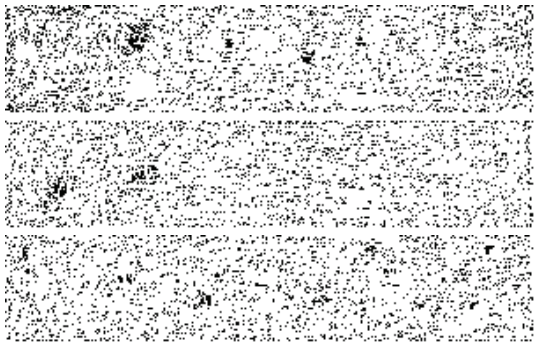


Figure 6: Three samples of masks used for choosing consistent bits in iris codes. Two upper masks are related to two subjects in CASIA-IrisV3, and the last one corresponds to a subject in Bath iris database.

coefficients of the iris image. As it can be seen in Figure 3, considering the quarter of the log Gabor coefficient in the real-imaginary axes, a two-bit binary code can be assigned to each coefficient.

Gabor filters are traditional choices for obtaining localized frequency information, and thanks to their similarity to the human vision system [1], these filters are vastly used in iris feature extraction phase. However, they suffer from two major drawbacks: (1) the maximum bandwidth of a Gabor filter is limited to approximately one octave, and (2) Gabor filters are not optimal, if one is seeking broad spectral information with maximal spatial localization. Considering these points, we used log Gabor filters [25] for feature extraction. Equation (4) shows this filter:

$$G(w) = e^{\left(-\log(w/w_0)^2\right)/\left(2\left(\log(k/w_0)^2\right)\right)}, \quad (4)$$

where w_0 is the filter's center frequency. To obtain constant shape ratio filters, the term k/w_0 must also be held constant for different w_0s.

It must be mentioned that using these filters is not an originality of this work (see [26]). Considering the real and imaginary parts of filters, texture of iris could be mapped to the iris codes, and as mentioned in [24], regarding to the distance of bits from axes, it is possible to choose some probability of bit consistency. For each user, the iris code of different samples is calculated, and by comparing these iris codes, the probability of changing each bit is determined. By choosing a threshold, it could be possible to judge about the consistency of each bit. Details about the consistency of bits in the iris codes can be found in [27].

In [27], existence of fragile bits in iris code has been theoretically proved, and the effect of applying filters, image rotation, and iris alignment has been discussed in details. In our work, we used their idea about the bit consistency in iris code and developed an applied method for iris recognition systems. In Figure 5, the performance of proposed method has been shown with different thresholds for using only the consistent bits in the iris code generation phase. As it can be seen, the best results have been obtained with threshold $T = 35\%$. In addition, the comparison between performances of our system considering all bits of iris code with the same systems considering only consistent bits shows the positive effect of masking fragile bits. For each user the proper rectangular calculated and features inside this rectangular are eliminated from iris code generation process.

For being rotation invariant, in this phase, like Daugman's method [4], the enrolled iris code will be compared with different shifted test iris codes to find the best match.

Figure 6 shows the calculated masks for three persons using samples in CASIA-IrisV3 and Bath iris databases. In this figure, black and white points show consistent and inconsistent bits, respectively.

4. Experimental Results

In our experimentations, we have used all samples of three famous iris databases, that is, CASIA-IrisV3, Bath, and Ubiris. CASIA-IrisV3 includes three subsets which are labeled as CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, and CASIA-IrisV3-Twins. CASIA-IrisV3 contains a total of 22 051 iris images from more than 700 subjects. All iris images are 8-bit gray-level JPEG files, collected under near infrared illumination. Almost all subjects are Chinese except a few ones in CASIA-IrisV3-Interval. Since these three datasets were collected in different times, CASIA-IrisV3-Interval and CASIA-IrisV3-Lamp have a small overlap in subjects. Some samples from this database have been shown in Figure 7(a). Bath iris database includes 20 samples from each eye of 25 subjects. The images are of a very high quality taken with a professional machine vision camera with infrared illumination. Some of these images have been shown in Figure 7(b).

Ubiris iris database version 1 is composed of 1877 images collected from 241 subjects taken in two sessions (Figure 7(c)). Unlike the CASIA-IrisV3 database, it includes

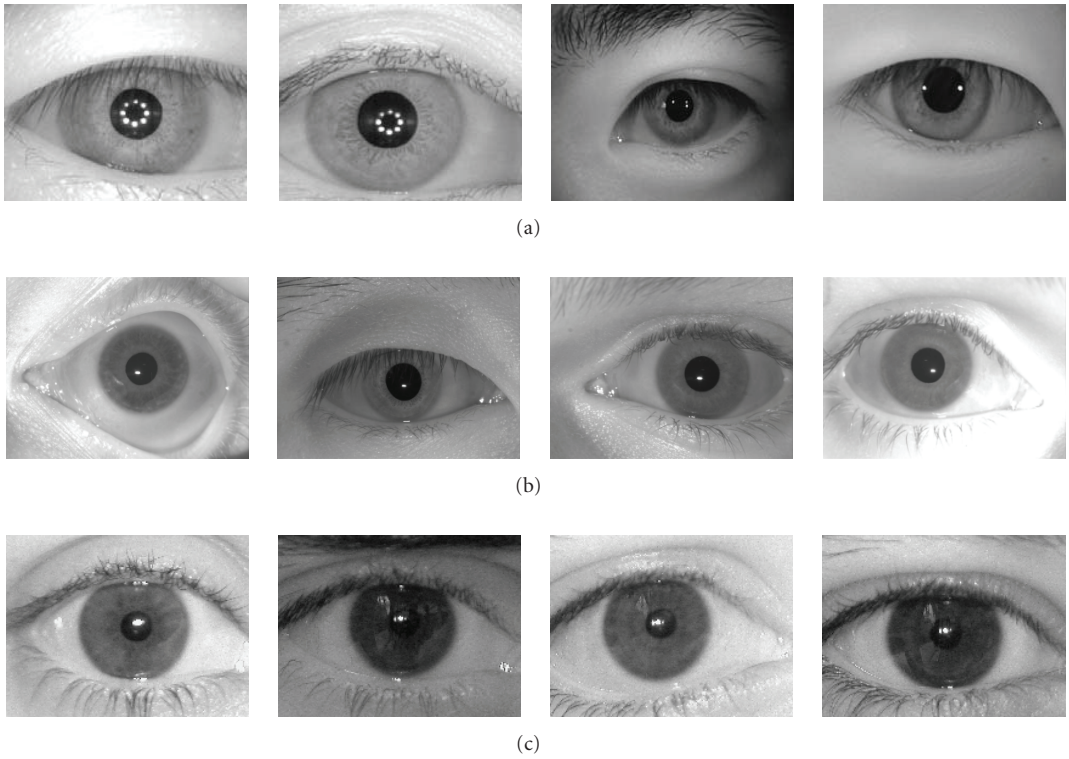


Figure 7: Some samples taken from (a) CASIA-IrisV3 database, (b) Bath database, and (c) Ubiris Version 1 database.

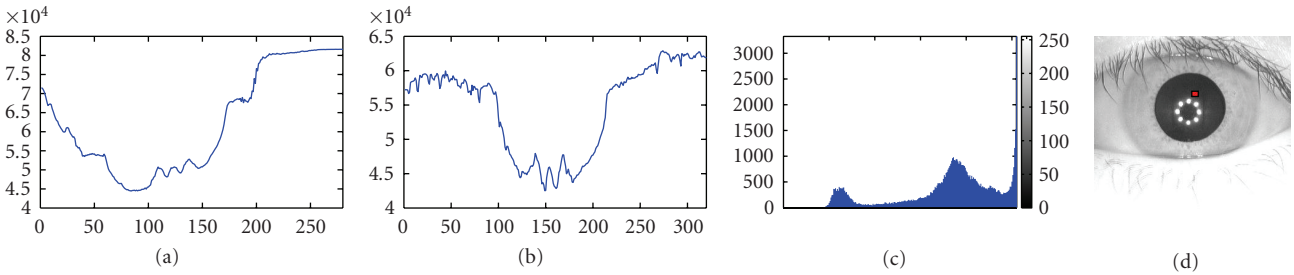


Figure 8: (a) Horizontal histogram, (b) Vertical Histogram, (c) Overall Histogram of the image, and (d) Estimated center.

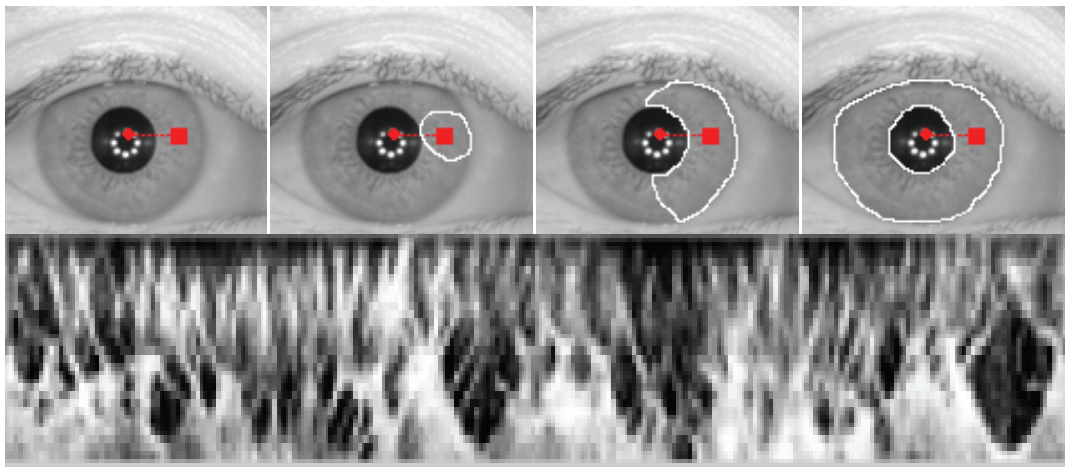
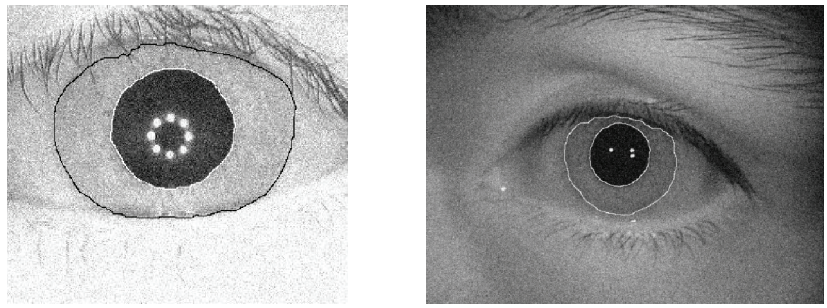
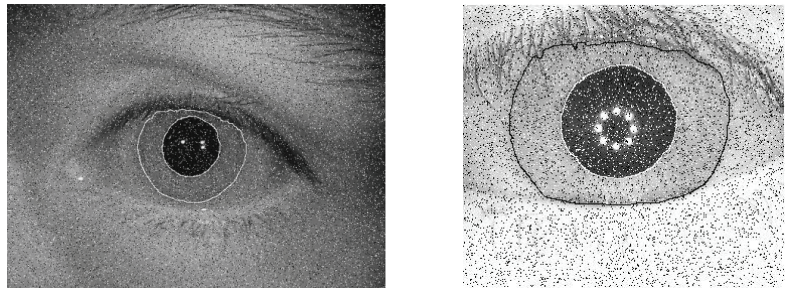


Figure 9: Inner and outer boundaries detection using pointwise level set approach done in one step and related iris codes.



(a) (b)

Figure 10: Performance of proposed algorithm in presence of Gaussian noise. For both images we have added a Gaussian white noise with mean = 0 and variance = 0.007.



(a) (b)

Figure 11: Performance of the proposed algorithm to iris images with (a) 10 percent and (b) 15 percent of salt and pepper noise.

images in different noisy situations, which permits to evaluate the robustness of iris recognition methods in presence of noise [19].

To evaluate the performance of our algorithm, we have used the K-fold cross validation technique. For CASIA-IrisV3 database, for each subject, three-iris samples have been utilized, to extract the user dependent iris code, and the rest of samples to test the algorithm. For Bath database, the number of samples used to extract the code is five. We have repeated this technique in a way that all of the iris images have been used in K-fold cross validation strategy.

In this work, the precise location of an iris is determined using pointwise level set approach with area preserving capability. Generally speaking, active contour models have been used previously in iris recognition systems [6]. Although active contour refers to a family of moving contour methods, in some papers, it corresponds to the Snake techniques [5]. In previous sections, we have described the drawbacks of the Snake model. Geodesic active contours with point correspondence have been used for iris segmentation in [4]. In this paper, we propose a method based on pointwise level set approach with area preserving capability.

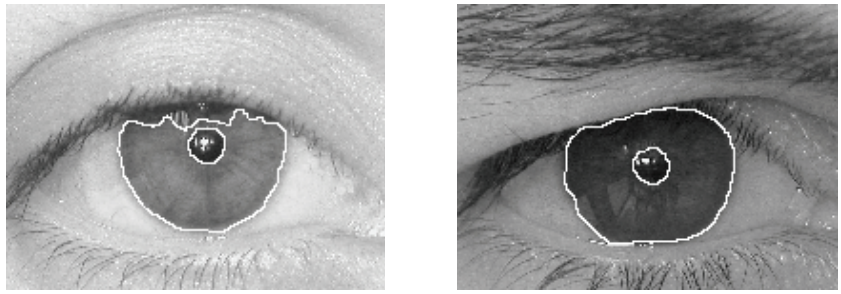
We calculate the approximate center of inner boundary of irises using vertical and horizontal histograms (Figure 8). Using this technique, the initial point of a contour is determined, and the starting point for tracing the contour is selected (for coordinate mapping to dimensionless polar space).

The vertical histogram is calculated as follows: size of the vertical histogram is equal to image's height, and the value of each histogram bin is equal to the sum of gray levels of a row of the image. The minimum of this histogram corresponds approximately to the vertical location of the center of inner boundary circle (almost circle). Indeed, pixels located in the pupil region are always dark; therefore, their values are close to 0. Thus, the minimum of the histogram shows the line that has the lowest number of dark pixels, that is, the diameter of the inner boundary circle. The intersection of this line with the output of the horizontal histogram shows the approximate location of the center point (Figure 8). Our experimental results show that we can locate the center of pupil in a point inside the pupil, even for difficult samples having other dark areas in the eye image. For image samples of datasets used in this paper, all pupils are placed almost in the center of the image.

In order to make the correct contour initialization (X, Y), the estimated center of pupil (x, y) is determined using (5) (Figure 9). In this equation, the contour starts to evolve from this point and is expected to find the whole iris location.

For calculating d from the approximate center, one dimensional derivation in the right horizontal axes has been calculated. d is equal to the length of line between the approximate center and some pixels after the found edge (in our experienced d could be an integer between 10 and 30):

$$X = x + d,$$
$$Y = y. \tag{5}$$



(a) (b)

Figure 12: Localization of two samples from Ubiris database with proposed method.

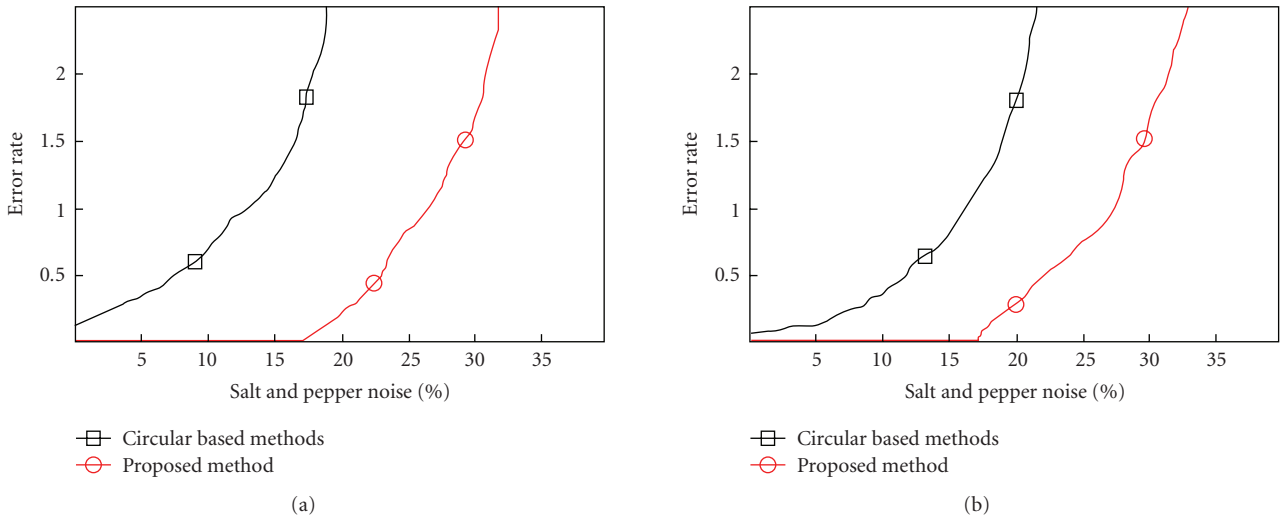


Figure 13: Error comparison between circle-based method and proposed method in noisy situation (salt and pepper noise).

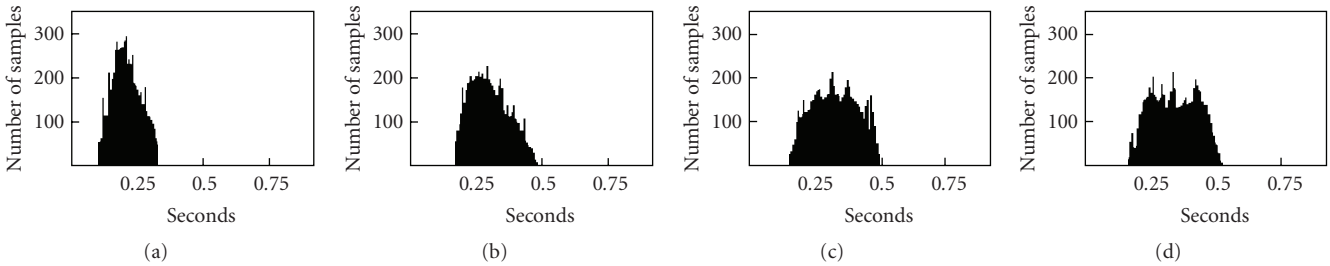


Figure 14: Response times of (a) Proposed, (b) Daugman [5], (c) Monro et al. [15], and (d) Ma et al. [7] methods using CASIA-IrisV3 database.

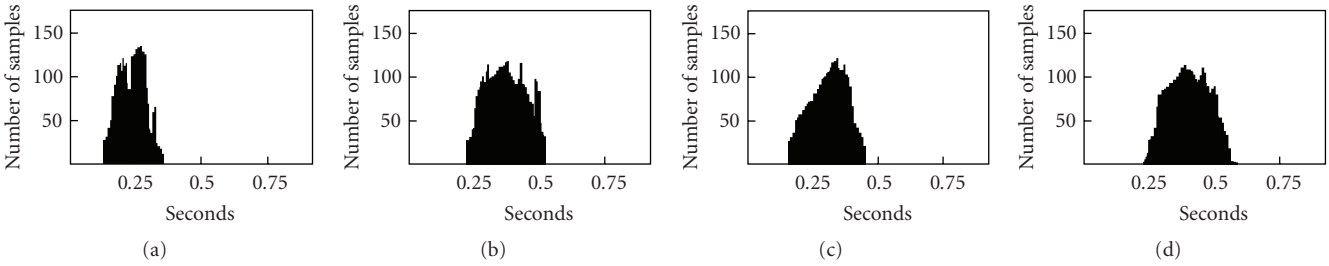


Figure 15: Response times of (a) Proposed, (b) Daugman [5], (c) Monro et al. [15], and (d) Ma et al. [7] methods using Bath iris database.

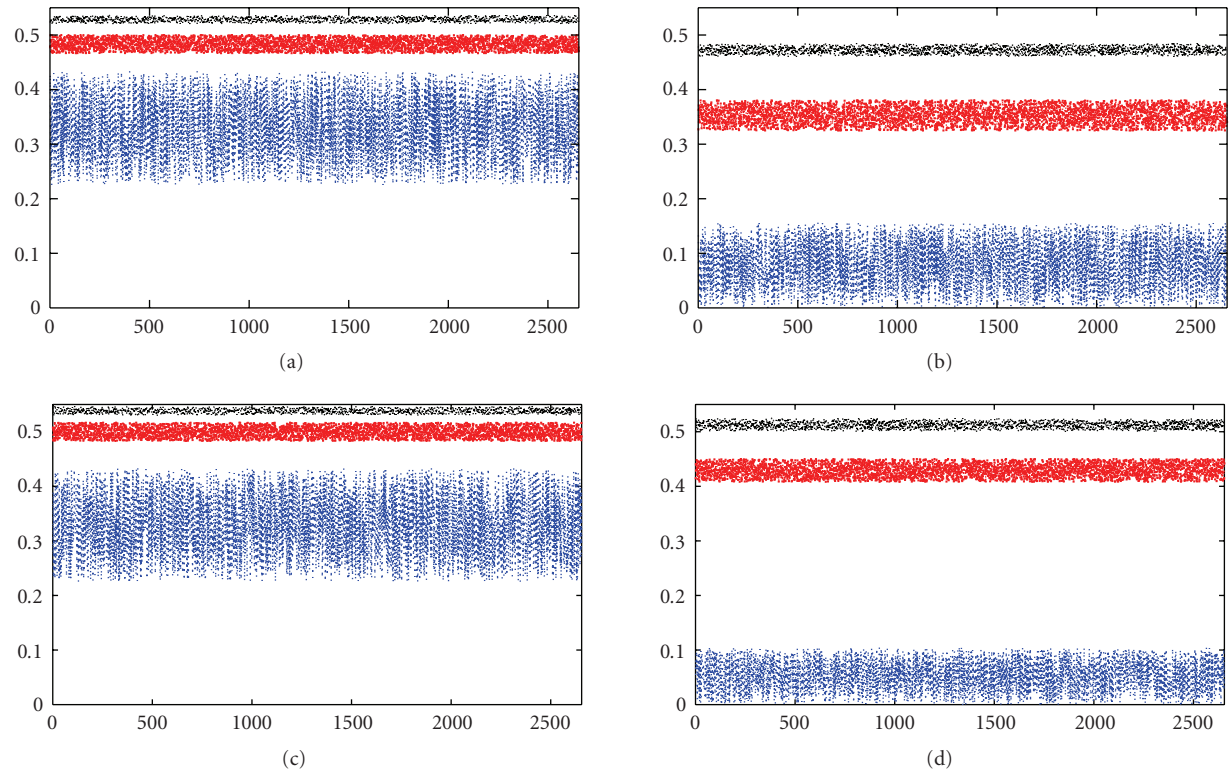


Figure 16: Hamming distance of match (blue,bottom), nearest nonmatch (red, middle), and average nonmatch (black, top) of (a) Daugman [5], (b) Monro et al. [15], (c) Ma et al. [7], and (d) proposed method using CASIA-IrisV3 interval database.

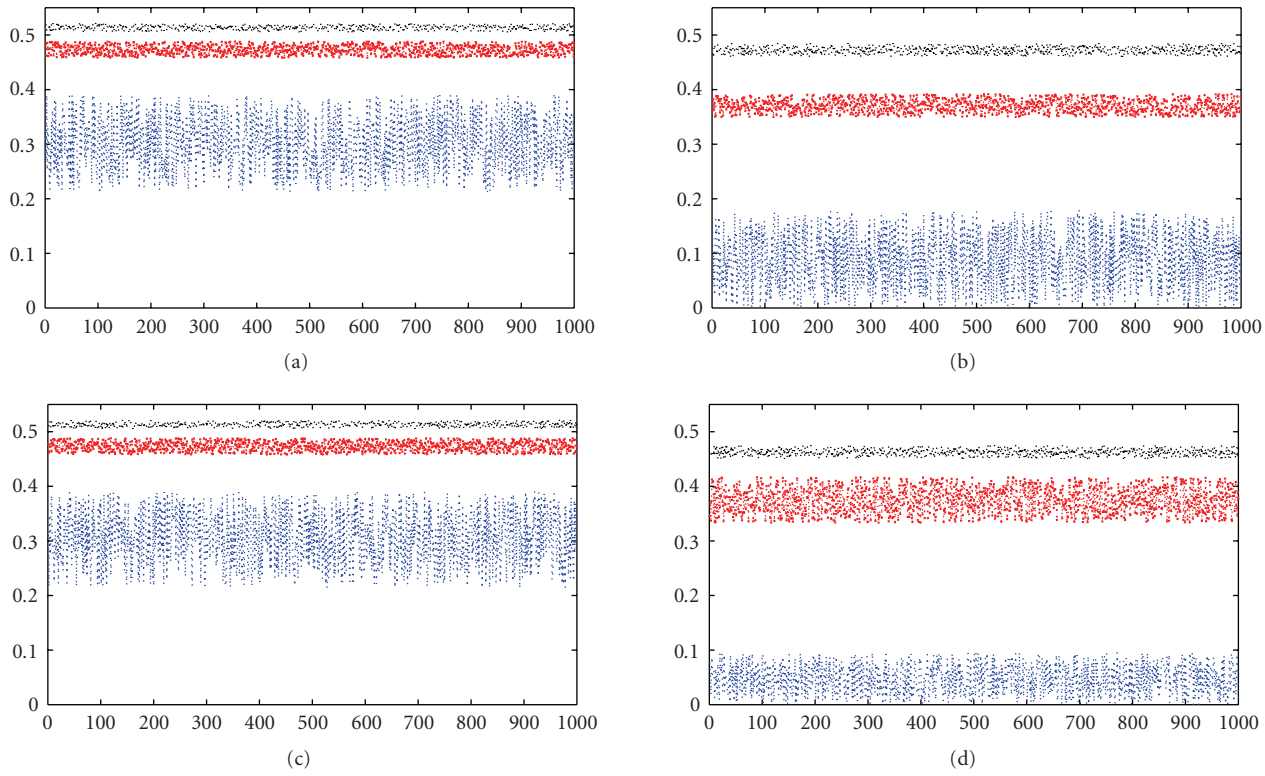


Figure 17: Hamming distance of match (blue, bottom), nearest nonmatch (red, middle), and average nonmatch (black, top) of (a) Daugman [5], (b) Monro et al. [15], (c) Ma et al. [7], and (d) proposed method using Bath iris database.

Table 1: Comparing localization accuracy of different methods using Ubiris database. The whole table entries are taken from reference [19], excluding the last row which contains the results obtained using our approach.

Methodology	Parameters	Session 1, %	Session 2, %	Degradation
Daugman	Original methodology	95.22 ± 0.015	88.23 ± 0.032	6.99
Daugman	Histogram equalization preprocessing	95.79 ± 0.028	91.10 ± 0.028	4.69
Daugman	Threshold preprocessing (128)	96.54 ± 0.013	95.32 ± 0.021	1.22
Wildes	Original methodology	98.68 ± 0.008	96.68 ± 0.017	2.00
Wildes	Shen and Castan edge detector	96.29 ± 0.013	95.47 ± 0.020	0.82
Wileds	Zero crossing edge detector	94.64 ± 0.016	92.76 ± 0.025	1.88
Caumus and Wileds	Original methodology, number of directions = 8	96.78 ± 0.013	89.29 ± 0.030	7.49
Martin-Roche et al.	Original methodology	77.18 ± 0.030	71.19 ± 0.045	5.99
Tuceryan	Total clusters = 5	90.28 ± 0.021	86.72 ± 0.033	3.56
Proenca et al.	Fuzzy K-means + (x, y) = position, z = intensity	98.02 ± 0.010	97.88 ± 0.015	0.14
Our Proposed method	Pointwise level set approach with area preserving capability	99.1 ± 0.01	98.98 ± 0.013	0.1

The proposed one step segmentation approach improves the speed of the whole process in comparison with regular two-step boundary detection methods.

This method is robust in noisy situations. A noisy pixel causes a sudden variation in gray levels and can stop the moving front. However, in this situation, other contour points continue to move and avoid the curve to stop. Figure 10 shows the results of applying our method to an iris image with Gaussian white noise (despite that encoding the iris texture is almost impossible in this image). During the detection process, some parts of the iris boundaries may have low gray level contrast, which may lead the algorithm to inaccurate edge detection results. For solving this problem, we have used a topology preserving algorithm [23], which guarantees the correct iris segmentation. Figure 11 shows the result of applying our algorithm to iris images with 10 and 15 percent salt and pepper noises.

In general, the effect of noncooperative iris images causes serious performance degradation. We used Ubiris iris database version 1 [28] for testing our localization ability dealing with noncooperative iris images. Our experimental results showed that our method is able to handle blurred, occluded images, localizing iris boundaries properly (Figure 12 and Table 1).

We tested our localization algorithm on Ubiris dataset and compared the results with the results published in [19]. The results in [19] were obtained by visual inspection of each segmented image. Although this is not the best for meaningful comparison, we did the same for localization evaluation in our system. Table 1 shows these results that are the proof of performance of our algorithm even for poor quality images. Indeed, in term of the degradation, the lowest accuracy degradation in the presence of noise belongs to our method, depicting low sensitivity of our approach to the image condition.

4.1. Error Definition. In order to measure the error of our method, we compared the points of the detected boundaries with those of the real boundaries. First, the exact boundary contours for inner and outer parts of irises are determined point to point manually. Then, the sum of the distance between the interface points and their nearest point in the correct boundary is calculated. Total error of localization is estimated using

$$E = \frac{\sum_{n=0}^{N} \min\left(\mathrm{dis}(I_n, C)\right)}{N}, \tag{6}$$

where C is the correct boundary, $\mathrm{dis}(I_n, C)$ means the set of distances between nth point of interface and all of the points of correct curve, and N is the total number of interface points. Although a global system performance measure such as ROC curve could be a better measure of performance, by introducing this error measure, we intend to evaluate our segmentation module performance exclusively. Figure 13 shows the localization errors (according to (5)), for proposed method and traditional circular based method, using some samples of CASIA-IrisV3 and Bath iris databases, in noisy situations.

4.2. Response Time. Figures 14 and 15 show the response times of proposed method using CASIA-IrisV3 and Bath iris databases. We implemented Daugman [5], Ma et al. [7], and Monro et al. [15] methods for comparing their results with the results of our proposed method. Our method's average response time in the same situation is less than others. In

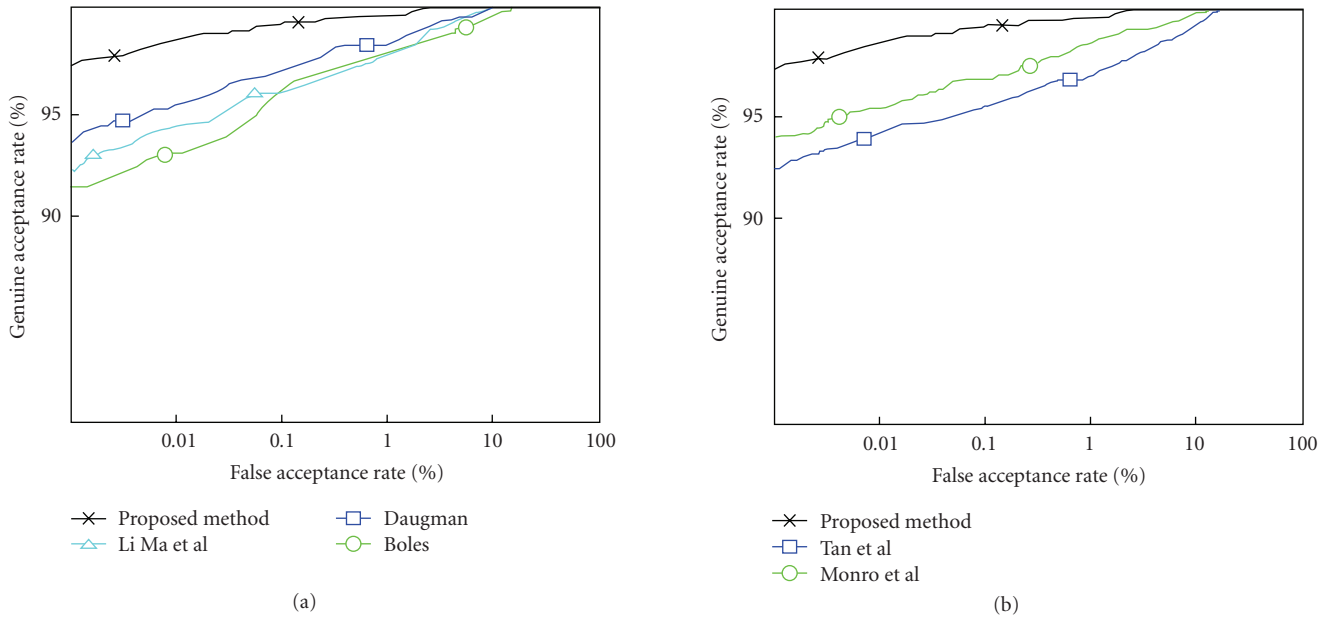


Figure 18: ROC curves of proposed method in comparison with (a) Boles and Boashash [10], Daugman [5], Ma et al. [7], and (b) Monro et al. [15], Halligswroth et al. [27] methods using CASIA-IrisV3 iris database.

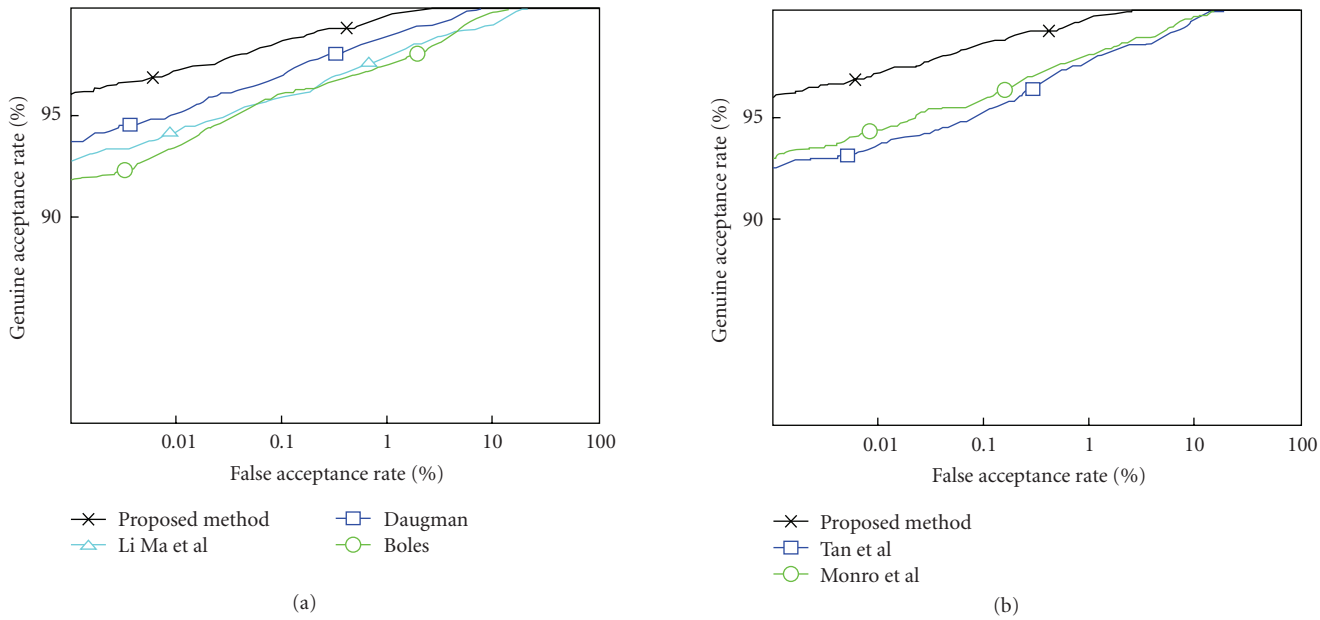


Figure 19: ROC curves of proposed method in comparison with (a) Boles and Boashash [10], Daugman [5], Ma et al. [7], and (b) Monro et al. [15], Halligswroth et al. [27] methods using Bath iris database.

addition, small standard deviation of our method is a proof of its performance for real time applications.

4.3. Hamming Distance. After generating the iris code, the result is compared with iris codes in databases using Hamming distance operators. Depending on the user dependent consistent bits, only the important bits of each iris code are involved in the matching process. Figures 16 and 17 show the calculated Hamming distances for Daugman's [5], Ma et al. [7], Monro et al. [15], and proposed methods, for CASIA-IrisV3 interval and Bath iris datasetss, respectively.

4.4. ROC Curves. ROC curves of proposed method have been compared with those of five different methods, tested on CASIA-IrisV3 and Bath iris databases, respectively, in Figures 18 and 19. The results show the superiority of our method compared to other methods. Figure 20 shows the performance of our method using the iris samples with

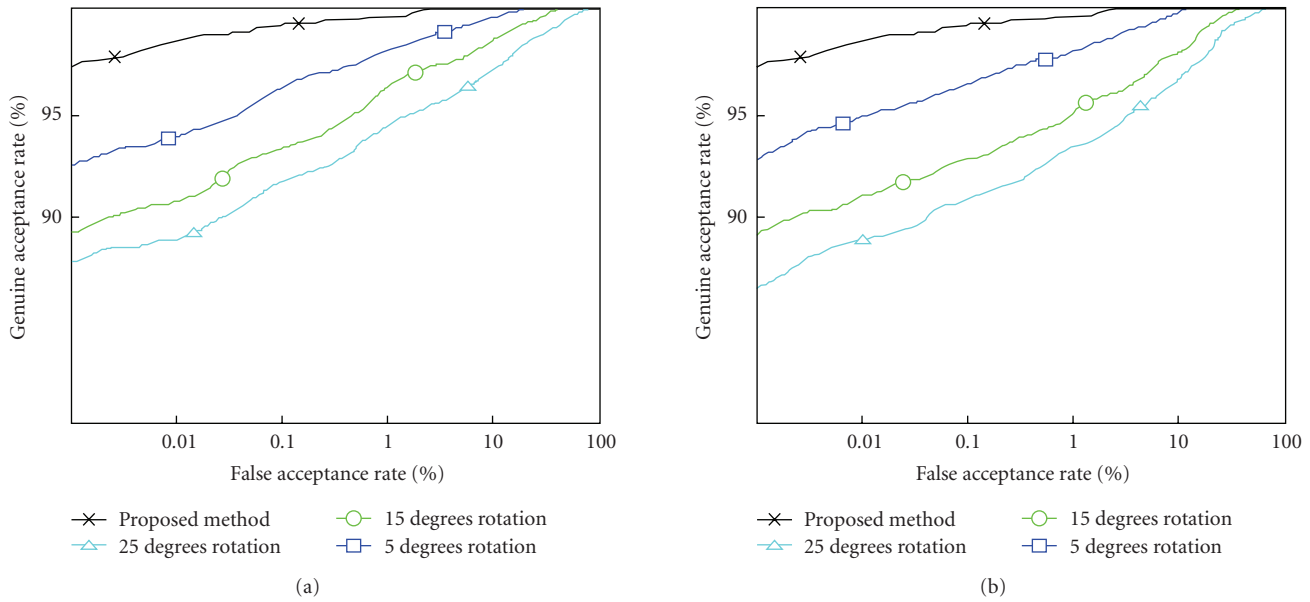


Figure 20: ROC curves of proposed method in comparison with best results of Boles and Boashash [10], Daugman [5], Ma et al. [7], Monro et al. [15] and Sun et al. [29] methods with iris rotations (5, 15, and 25 degrees clockwise) using (a) CASIA-IrisV3 and (b) Bath iris database.

5, 15, and 25 degrees rotation, compared to Boles and Boashash [10], Daugman [5], Ma et al. [7], Monro et al. [15], and Sun et al. [29] methods, tested on CASIA-IrisV3 and Bath iris databases. One of the curves belongs to the proposed method, and in each of the other curves, each point corresponds to the best result obtained from these four methods, for 5, 15, and 25 degrees rotations, respectively. Indeed, we showed only one curve for different rotations applied to our proposed method, which is a proof of robustness of this method against rotation. Concerning the other three curves in Figure 20, as it has been mentioned, each curve is a pointwise combination of the best of the four other methods.

As it can be seen, our method is robust against rotation, while rotation degrades the performance of other methods considerably, due to their circular edge detection nature. In general, circular edge detection process is based on determining the location of the circle with maximum differences of pixel gray levels for two adjacent circular curves. In practice, these differences are calculated using two arches, instead of a whole circle. The performance of the iris localization depends on the location and angle of these arches in relation with the iris axis, and, as a consequence, rotating the image degrades the results of circular edge detection, mainly due to wrong arches used in the process and presence of eyelid and eyelashes. In contrast with these conventional methods, the iris localization in the proposed method is based on geodesic active contour model, which calculates the iris boundaries independently to any geometric shape, including circles and arches; therefore, it is robust to the image rotation problem.

5. Conclusions

We have proposed a new user-dependent iris recognition method. Using a specific mask for each user, inconsistent bits of iris code are omitted during the Hamming distance comparison phase. As the experimental results show, using this approach, the performance of the whole system is improved considerably. Another contribution of this paper is the utilization of pointwise level set approach with area preserving capability for iris segmentation and localization. In this algorithm, the exact location of the iris can be detected using an iterative algorithm based on the active contour model. Comparing our algorithm with other methods, we showed that the new approach is able to solve some of the previous method's drawbacks. For instance, using our method, the iris location can be detected regardless to its angular position and shape, and this is done in only one step. Also, previous methods usually detect iris boundaries using circular edge. One of the disadvantages of this approximation is its sensitivity to the rotation of the iris images. In recent years, active contour model have been used for iris detection purposes. However, our method has some advantages over other methods. Indeed, an area preserving algorithm is used to compensate the problem of incorrect iris boundary detection in presence of noise. Furthermore, even when eyelids occlude some part of iris, our algorithm localizes iris area properly [4]. The experimental results show that our method outperforms the current methods both in terms of accuracy and response time.

References

[1] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.

[2] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: a survey," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 281–307, 2008.

[3] T. A. Camus and R. Wildes, "Reliable and fast eye finding in close-up images," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 1, pp. 389–394, Quebec, Canada, August 2002.

[4] N. Barzegar and M. S. Moin, "A new approach for iris localization in iris recognition systems," in *Proceedings of the 6th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA '08)*, pp. 516–523, Doha, Qatar, March-April 2008.

[5] J. Daugman, "New methods in iris recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 37, no. 5, pp. 1167–1175, 2007.

[6] A. Ross and S. Shah, "Segmenting non-ideal irises using geodesic active contours," in *Proceedings of the Biometric Consortium Conference (BCC '06)*, pp. 1–6, Baltimore, Md, USA, September-August 2006.

[7] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal identification based on iris texture analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1519–1533, 2003.

[8] Z. Sun, Y. Wang, T. Tan, and J. Cui, "Improving iris recognition accuracy via cascaded classifiers," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 35, no. 3, pp. 435–441, 2005.

[9] J. Huang, L. Ma, T. Tan, and Y. Wang, "Learning based enhancement model of iris," in *Proceedings of the 14th British Machine Vision Conference (BMVC '03)*, pp. 153–162, Norwich, UK, September 2003.

[10] W. W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1185–1188, 1998.

[11] S. Lim, K. Lee, O. Byeon, and T. Kim, "Efficient iris recognition through improvement of feature vector and classifier," *ETRI Journal*, vol. 23, no. 2, pp. 61–70, 2001.

[12] C. Tisse, L. Martin, L. Torres, and M. Robert, "Person identification technique using human iris recognition," in *Proceedings of the 15th International Conference on Vision Interface (VI '02)*, pp. 294–299, Calgary, Canada, May 2002.

[13] W.-K. Kong and D. Zhang, "Detecting eyelash and reflection for accurate iris segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 6, pp. 1025–1034, 2003.

[14] J. Thornton, M. Savvides, and V. Kumar, "A Bayesian approach to deformed pattern matching of iris images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 596–606, 2007.

[15] D. M. Monro, S. Rakshit, and D. Zhang, "DCT-based iris recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 586–595, 2007.

[16] Y. Zhu, T. Tan, and Y. Wang, "Biometric personal identification based on iris patterns," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 2, pp. 801–804, Barcelona, Spain, September 2000.

[17] http://www.cbsr.ia.ac.cn/IrisDatabase.htm.

[18] http://www.bath.ac.uk/elec-eng/research/sipg/irisweb.

[19] H. Proença and L. A. Alexandre, "Iris segmentation methodology for non-cooperative recognition," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 153, no. 2, pp. 199–205, 2006.

[20] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, Cambridge, Mass, USA, 2nd edition, 1999.

[21] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," in *Proceedings of the 1st International Conference on Computer Vision (ICCV '87)*, pp. 259–268, London, UK, June 1987.

[22] J.-P. Pons, G. Hermosillo, R. Keriven, and O. Faugeras, "Maintaining the point correspondence in the level set framework," *Journal of Computational Physics*, vol. 220, no. 1, pp. 339–354, 2006.

[23] J.-P. Pons, R. Keriven, and O. Faugeras, "Area preserving cortex unfolding," in *Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, vol. 3216 of *Lecture Notes in Computer Science*, pp. 376–383, Saint-Malo, France, September 2004.

[24] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "All iris code bits are not created equal," in *Proceedings of the 1st IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS '07)*, pp. 1–6, Crystal City, Va, USA, September 2007.

[25] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[26] P. Yao, J. Li, X. Ye, Zh. Zhuang, and B. Li, "Iris recognition algorithm using modified Log-Gabor filters," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 461–464, Hong Kong, August 2006.

[27] K. P. Holligsworth, K. W. Bowyer, and P. J. Flynn, "The Best Bitsin an Iris Code," *IEEE Trensaction on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964–973, 2009.

[28] H. Proença and L. A. Alexandre, "UBIRIS: a noisy iris image database," in *Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP '05)*, vol. 3617 of *Lecture Notes in Computer Science*, pp. 970–977, Cagliari, Italy, September 2005.

[29] Z. Sun, Y. Wang, T. Tan, and J. Cui, "Improving iris recognition accuracy via cascaded classifiers," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 35, no. 3, pp. 435–441, 2005.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 415817, 16 pages
doi:10.1155/2009/415817

Research Article

Gait Recognition Using Wearable Motion Recording Sensors

Davrondzhon Gafurov and Einar Snekkenes

Norwegian Information Security Laboratory, Gjøvik University College, P.O. Box 191, 2802 Gjøvik, Norway

Correspondence should be addressed to Davrondzhon Gafurov, davrondzhon.gafurov@hig.no

Received 1 October 2008; Revised 26 January 2009; Accepted 26 April 2009

Recommended by Natalia A. Schmid

This paper presents an alternative approach, where gait is collected by the sensors attached to the person's body. Such wearable sensors record motion (e.g. acceleration) of the body parts during walking. The recorded motion signals are then investigated for person recognition purposes. We analyzed acceleration signals from the foot, hip, pocket and arm. Applying various methods, the best EER obtained for foot-, pocket-, arm- and hip- based user authentication were 5%, 7%, 10% and 13%, respectively. Furthermore, we present the results of our analysis on security assessment of gait. Studying gait-based user authentication (in case of hip motion) under three attack scenarios, we revealed that a minimal effort mimicking does not help to improve the acceptance chances of impostors. However, impostors who know their closest person in the database or the genders of the users can be a threat to gait-based authentication. We also provide some new insights toward the uniqueness of gait in case of foot motion. In particular, we revealed the following: a sideway motion of the foot provides the most discrimination, compared to an up-down or forward-backward directions; and different segments of the gait cycle provide different level of discrimination.

1. Introduction

Biometric recognition uses humans anatomical and behavioral characteristics. Conventional human characteristics that are used as biometrics include fingerprint, iris, face, voice, and so forth. Recently, new types of human characteristics have been proposed to be used as a biometric modality, such as typing rhythm [1], mouse usage [2], brain activity signal [3], cardiac sounds [4], and gait (walking style) [5]. The main motivation behind new biometrics is that they are better suited in some applications compared to the traditional ones, and/or complement them for improving security and usability. For example, gait biometric can be captured from a distance by a video camera while the other biometrics (e.g., fingerprint or iris) is difficult or impossible to acquire.

Recently, identifying individuals based on their gait became an attractive research topic in biometrics. Besides being captured from a distance, another advantage of gait is to enable an unobtrusive way of data collection, that is, it does not require explicit action/input from the user side. From the way how gait is collected, gait recognition can be categorized into three approaches:

(i) Video Sensor- (VS-) based,

(ii) Floor Sensor- (FS-) based,

(iii) Wearable Sensor- (WS-) based.

In the VS-based approach, gait is captured from a distance using a video-camera and then image/video processing techniques are applied to extract gait features for recognition (see Figure 1). Earlier works on VS-based gait recognition showed promising results, usually analyzing small data-sets [6, 7]. For example, Hayfron-Acquah et al. [7] with the database of 16 gait samples from 4 subjects and 42 gait samples from 6 subjects achieved correct classification rates of 100% and 97%, respectively. However, more recent studies with larger sample sizes confirm that gait has distinctive patterns from which individuals can be recognized [8–10]. For instance, Sarkar et al. [8] with a data-set consisting of 1870 gait sequences from 122 subjects obtained 78% identification rate at rank 1 (experiment B). A significant amount of research in the area of gait recognition is focused on VS-based gait recognition [10]. One reason for much interest in VS-based gait category is availability of large public gait databases, such as that provided by University of South Florida [8], University of Southampton [11] and

Table 1: Summary of some VS-based gait recognitions.

Study	*EER*, %	*#S*
Seely et al. [12]	4.3–9.5	103
Zhao et al. [13]	11.17	—
Hong et al. [14]	9.9–13.6	20
BenAbdelkader et al. [15]	11	17
Wang et al. [16]	3.8–9	124
Wang et al. [17]	8–14	20
Wang et al. [18] (without fusion)	8–10	20
Bazin et al. [19] (without fusion)	7–23	115

Table 2: Summary of several FS-based gait recognitions.

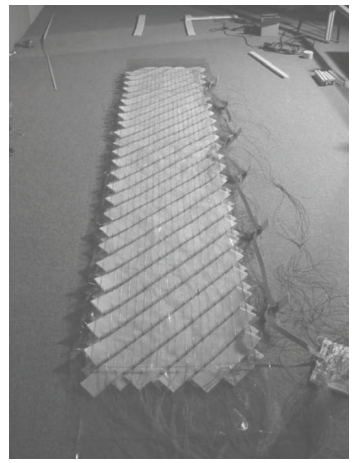
Study	*Recognition rate*, %	*#S*
Nakajima et al. [23]	85	10
Suutala and Röning [24]	65.8–70.2	11
Suutala and Röning [25]	79.2–98.2	11
Suutala and Röning [26]	92	10
Middleton et al. [20]	80	15
Orr and Abowd [27]	93	15
Jenkins and Ellis [28]	39	62

(a) Original image

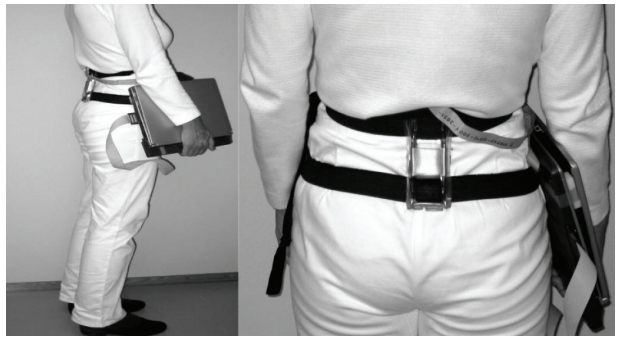
(b) Background

(c) Silhouette

(a) Using video-camera [5]



(b) Using floor sensor [20]



(c) Using wearable sensor on the body [21]

Figure 1: Examples of collecting gait.

Chinese Academy of Sciences [22]. Performance in terms of EER for some VS-based gait recognitions is given in Table 1. In this table (and also in Tables 2 and 3) the column *#S* indicates the number of subjects in the experiment. It is worth noting that the direct comparison of the performances in Table 1 (and also in Tables 2 and 3) may not be adequate mainly due to the differences among the data-sets. The purpose of these tables is to give some impression of the recognition performances.

In the FS-based approach, a set of sensors are installed in the floor (see Figure 1), and gait-related data are measured when people walk on them [20, 24, 27, 28]. The FS-based approach enables capturing gait features that are difficult or impossible to collect in VS-based approach, such as Ground Reaction Force (GRF) [27], heel to toe ratio [20], and so forth. A brief performance overview of several FS-based gait recognition works (in terms of recognition rate) is presented in Table 2.

The WS-based gait recognition is relatively recent compared to the other two mentioned approaches. In this approach, so-called motion recording sensors are worn or attached to various places on the body of the person such as shoe and waist, (see Figure 1). [21, 29–34]. Examples of the recording sensor can be accelerometer, gyro sensors, force sensors, bend sensors, and so on that can measure various characteristics of walking. The movement signal recorded by such sensors is then utilized for person recognition purposes. Previously, the WS-based gait analysis has been used successfully in clinical and medical settings to study and monitor patients with different locomotion disorders [35]. In medical settings, such approach is considered to be cheap and portable, compared to the stationary vision based systems [36]. Despite successful application of WS-based gait analysis in clinical settings, only recently the approach has been applied for person recognition. Consequently, so far not much has been published in the area of person recognition using WS-based gait analysis. A short summary of the current WS-based gait recognition studies is presented in Table 3. In this table, the column "Reg." is the recognition rate.

This paper reports our research in gait recognition using the WS-based approach. The main contributions of the paper are on identifying several body parts whose motion can provide some identity information during gait; and on analyzing uniqueness and security per se (robustness against attacks) of gait biometric. In other words, the three main research questions addressed in this paper are as follows.

(1) What are the performances of recognition methods that are based on the motion of body parts during gait?

(2) How robust is the gait-based user authentication against attacks?

(3) What aspects do influence the uniqueness of human gait?

Table 3: Summary of the current WS-based gait recognitions.

Study	*Sensor(s) location*	*Performance*, %		*#S*
		EER	Reg.	
Morris [29]	shoe	—	97.4	10
Huang et al. [32]	shoe	—	96.93	9
Ailisto et al. [21]	waist	6.4	—	36
Mäntyjärvi et al. [30]	waist	7–19	—	36
Rong et al. [34]	waist	6.7	—	35
Rong et al. [33]	waist	5.6, 21.1	—	21
Vildjiounaite et al. [31] (without fusion)	hand	17.2, 14.3	—	31
Vildjiounaite et al. [31] (without fusion)	hip pocket	14.1, 16.8	—	31
Vildjiounaite et al. [31] (without fusion)	breast pocket	14.8, 13.7	—	31

The rest of the paper is structured as follow. Section 2 presents our approach and results on WS-based gait recognition (research question (1)). Section 3 contains security evaluations of gait biometric (research question (2)). Section 4 provides some uniqueness assessment of gait biometric (research question (3)). Section 5 discusses possible application domains and limitations of the WS-based gait recognition. Section 6 concludes the paper.

2. WS-Based Gait Recognition

2.1. Motion Recording Sensor. For collecting gait, we used so called Motion Recording Sensors (MRSs) as shown in Figure 2. The attachment of the MRS to various places on the body is shown in Figure 3. These sensors were designed and developed at Gjøvik University College. The main component of these sensors was an accelerometer which records acceleration of the motion in three orthogonal directions that is up-down, forward-backward, and sideways. From the output of the MRS, we obtained acceleration in terms of $g(g = 9.8\,\mathrm{m/s^2})$ (see Figure 5). The sampling frequencies of the accelerometers were 16 Hz (first prototype) and 100 Hz. The other main components of the sensors were a memory for storing acceleration data, communication ports for transferring data, and a battery.

2.2. Recognition Method. We applied various methods to analyze the acceleration signals, which were collected using MRS, from several body segments: foot, hip, trousers pocket, and arm (see Figure 3 for sensor placements). A general structure of our gait recognition methods is visualized in Figure 4. The recognition methods essentially consisted of the following steps.

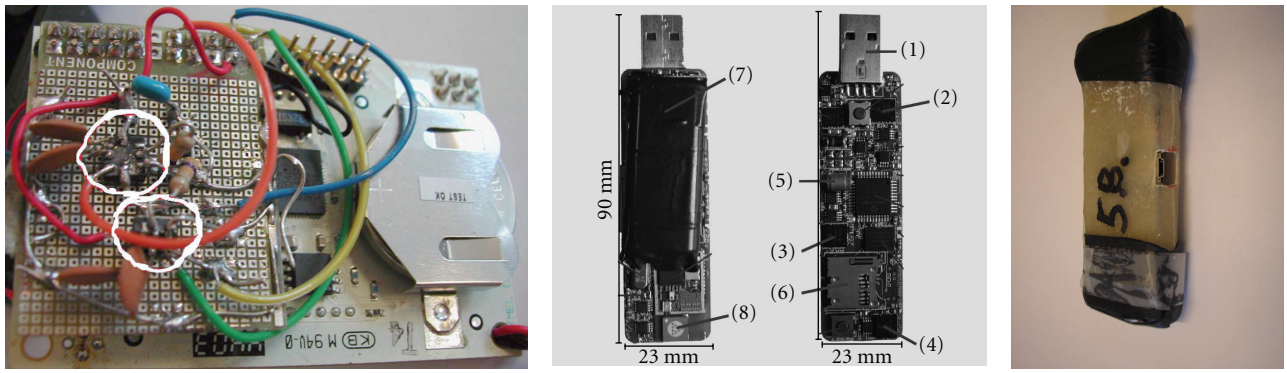
2.2.1. Preprocessing. In this step, we applied moving average filters to reduce the level of noise in the signals. Then, we computed a resultant acceleration, which is combination of acceleration from three directions of the motion. It was computed as follows:

$$R_i = \sqrt{X_i^2 + Y_i^2 + Z_i^2}, \quad i = 1, \ldots, m, \tag{1}$$

where R_i is the resultant acceleration at time i, X_i, Y_i, and Z_i are vertical, forward-backward, and sideway acceleration value at time i, respectively, and m is the number of recorded samples. In most of our analysis, we used resultant acceleration rather than considering 3 signals separately.

2.2.2. Motion Detection. Usually, recorded acceleration signals contained some standing still intervals in the beginning and ending of the signal (Figure 5(a)). Therefore, first we separated the actual walking from the standing still parts. We empirically found that the motion occurs around some specific acceleration value (the value varies for different body locations). We searched for the first such acceleration value and used it as the start of the movement (see Figure 5(a)). A similar procedure could be applied to detect when the motion stops. Thus, the signal between these two points was considered as a walking part and investigated for identity recognition.

2.2.3. Feature Extraction. The feature extraction module analyses motion signals in time or frequency domains. In the time domain, gait cycles (equivalent to two steps) were detected and normalized in time. The normalized cycles were combined to create an average cycle of the person. Then, the averaged cycle was used as a feature vector. Before averaging, some cycles at the beginning and ending of the motion signal were omitted, since the first and last few seconds may not adequately represent the natural gait of the person [35]. An example of selected cycles is given in color in Figure 5(b). In the frequency domain, using Fourier coefficients an amplitude of the acceleration signal is calculated. Then, maximum amplitudes in some frequency ranges are used as a feature vector [37]. We analysed arm signal in frequency domain and the rest of them in time domain.



(a) (b) (c)

Figure 2: Motion recording sensors (MRS).

(a) Ankle (b) Hip (c) Arm

Figure 3: The placement of the MRS on the body.

2.2.4. Similarity Computation. For computing similarity score between the template and test samples we applied a distance metric (e.g., Euclidean distance). Then, a decision (i.e., accept or reject) was based on similarity of samples with respect to the specified threshold.

More detailed descriptions of the applied methods on acceleration signals from different body segments can be found in [37–40].

2.3. Experiments and Results. Unlike VS-based gait biometric, no public data-set on WS-based gait is available (perhaps due to the recency of this approach). Therefore, we have conducted four sets of experiments to verify the feasibility of recognizing individuals based on their foot, hip, pocket, and arm motions. The placements of the MRS in those experiments are shown in Figure 3. In case of the pocket experiment, the MRS was put in the trousers pocket of the subjects. All the experiments (foot, hip, pocket, and arm) were conducted separately in an indoor environment. In the experiments, subjects were asked to walk using their natural gait on a level surface. The metadata of the 4 experiments are shown in Table 4. In this table, the column *Experiment* represents the body segment (sensor location) whose motion was collected. The columns #*S*, *Gender* (*M* + *F*), *Age range*, #*N*, and #*T* indicate the number of subjects in experiment, the number of male and female subjects, the age range of subjects, the number of gait samples (sequences) per subject, and the total number of gait samples, respectively.

For evaluating performance in verification (one-to-one comparison) and identification (one-to-many comparisons) modes we adopted DET and CMC curves [41], respectively. Although we used several methods (features) on acceleration signals, we only report the best performances for each body segment. The performances of the foot-, hip-, pocket- and arm-based identity recognition in verification and identification modes are given in Figures 6(a) and 6(b), respectively. Performances in terms of the EER and identification rates at rank 1 are also presented in Table 5.

3. Security of Gait Biometric

In spite of many works devoted to the gait biometric, gait security per se (i.e., robustness or vulnerability against attacks) has not received much attention. In many previous works, impostor scores for estimating FAR were generated by matching the normal gait samples of the impostors against

Table 4: Summary of experiments.

Experiment	*#S*	*Gender* $(M + F)$	*Age range*	*#N*	*#T*
Ankle	21	12 + 9	20–40	2	42
Hip	100	70 + 30	19–62	4	400
Pocket	50	33 + 17	17–62	4	200
Arm	30	23 + 7	19–47	4	120

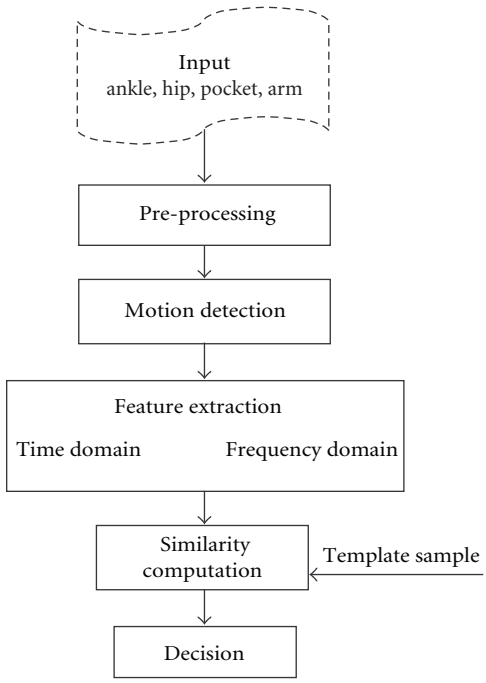


Figure 4: A general structure of recognition methods.

Table 5: Summary of performances of our approaches.

MRS placement	*Performance*, %		*#S*
	EER	P_1 at rank 1	
Ankle	5	85.7	21
Hip	13	73.2	100
Trousers pocket	7.3	86.3	50
Arm	10	71.7	30

the normal gait samples of the genuine users [15, 17–19, 21, 30]. We will refer to such scenario as a "friendly" testing. However, the "friendly" testing is not adequate for expressing the security strength of gait biometric against motivated attackers, who can perform some action (e.g., mimic) or possess some vulnerability knowledge on the authentication technique.

3.1. Attack Scenarios. In order to assess the robustness of gait biometric in case of hip-based authentication, we tested 3 attack scenarios:

(1) minimal-effort mimicking [39],

(2) knowing the closest person in the database [39],

(3) knowing the gender of users in the database [42].

The minimal-effort mimicking refers to the scenario where the attacker tried to walk as someone else by deliberately changing his walking style. The attacker had limited time and number of attempts to mimic (impersonate) the target person's gait. For estimating FAR, the mimicked gait samples of the attacker were matched against the target person's gait. In the second scenario, we assumed that the attackers knew the identity of person in the database who had the most similar gait to the attacker's gait. For estimating FAR, the attacker's gait was matched only to this nearest person's gait. Afterwards, the performances of mimicking and knowing closest person scenarios were compared to the performance of the "friendly" scenario. In the third scenario, it was assumed that attackers knew the genders of the users in the database. Then, we compared performance of two cases, so called same- and different-gender matching. In the first case, attackers' gait was matched to the same gender users and in the second case attackers' gait was matched to the different gender users.It is worth noting that in second and third attack scenarios, attackers were not mimicking (i.e., their natural gait were matched to the natural gait of the victims) but rather possessed some knowledge about genuine users (their gait and gender).

3.2. Experimental Data and Results. We analyzed the aforementioned security scenarios in case of the hip-based authentication where the MRS was attached to the belt of subjects around hip as in Figure 3(b). For investigating the first attack scenario (i.e., minimal-effort mimicking), we conducted an experiment where 90 subjects participated, 62 male and 28 female. Every subject was paired with another one (45 pairs). The paired subjects were friends, classmates or colleagues (i.e., they knew each other). Everyone was told to study his partner's walking style and try to imitate him or her. One subject from the pair acted as an attacker, the other one as a target, and then the roles were exchanged. The genders of the attacker and the target were the same. In addition, the age and physical characteristics (height and weight) of the attacker and target were not significantly different. All attackers were amateurs and did not have a special training for the purpose of the mimicking. They only studied the target person visually, which can also easily be done in a real-life situation as gait cannot be hidden. The only information about the gait authentication they knew was that the acceleration of normal walking was used. Every attacker made 4 mimicking attempts.

As it was mentioned previously in the second and third attack scenarios (i.e., knowing the closest person and gender of users), the impostors were not mimicking. In these

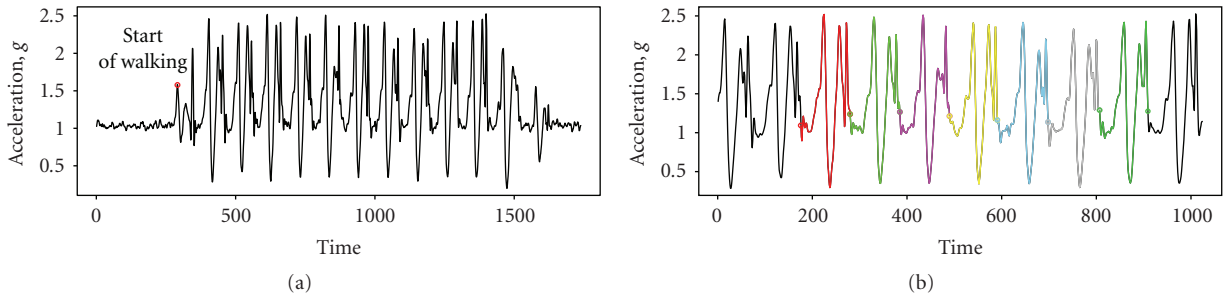
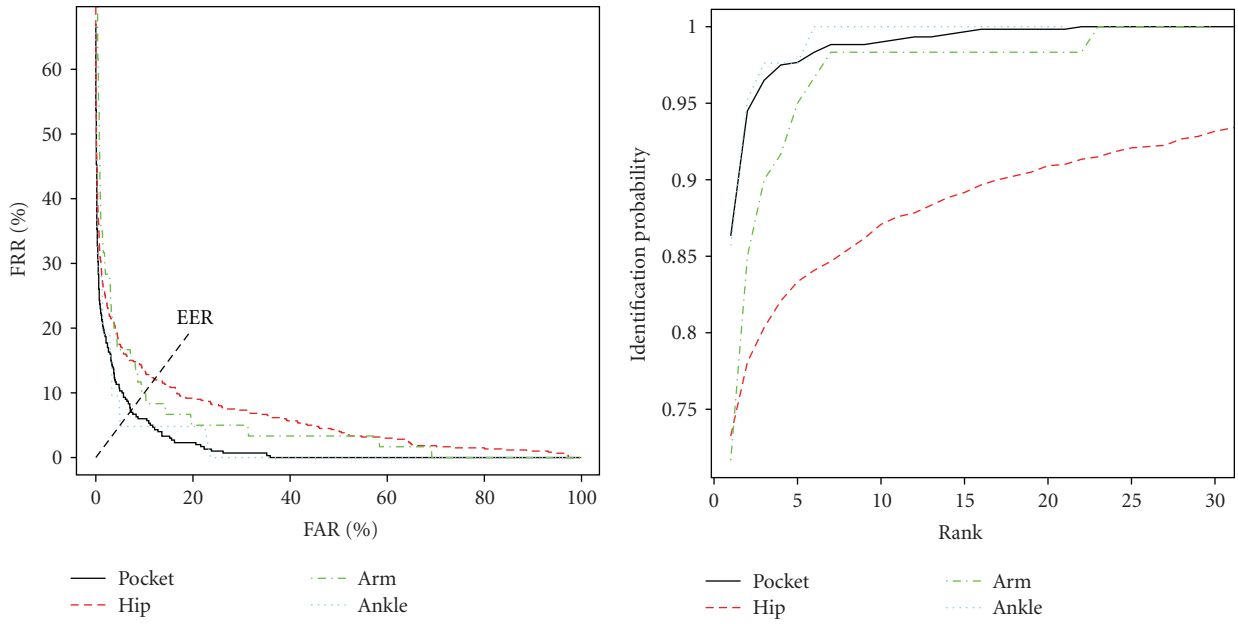


Figure 5: An example of acceleration signal from foot: (a) motion detection and (b) cycle detection.



(a) Decision error trade-off (DET) cuves

(b) Cumulative match characteristics (CMC) curves

Figure 6: Performances in terms of DET and CMC curves.

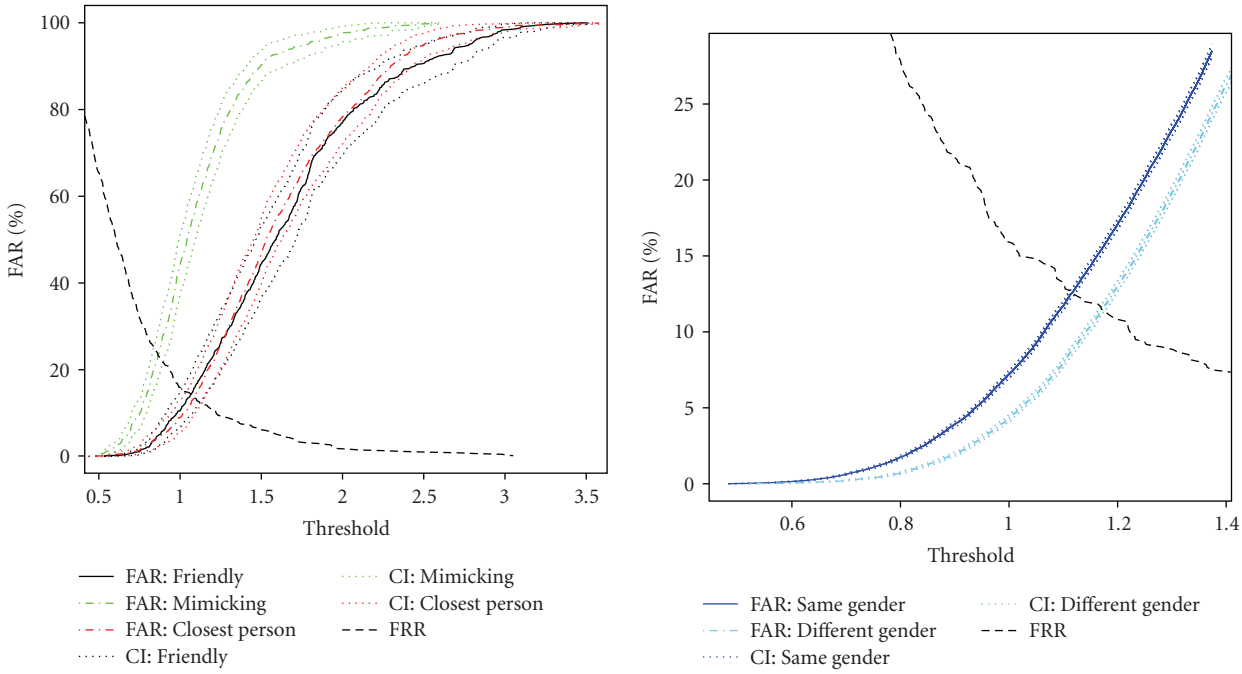
two attack scenarios, the hip data-set from Section 2.3 was used.

In general, the recognition procedure follows the same structure as in Figure 4, and involves preprocessing, motion detection, cycles detection, and computation of the averaged cycle. For calculating a similarity score between two persons' averaged cycle, the Euclidean distance was applied. A more detailed description of the method can be found in [39]. Performance evaluation under attacking scenarios are given in terms of FAR curves (versus threshold) and shown in Figure 7. Figure 7(a) shows the results of the minimal-effort mimicking and knowing the closest person scenarios as well as "friendly" scenario. Figure 7(b) represents the results of security scenario where attackers knew the gender of the victims. In Figures 7(a) and 7(b), the dashed black curve is FRR and its purpose is merely to show the region of EER. In order to get robust picture of comparison, we also computed confidence intervals (CI) for FAR. The CI were computed using nonparametric (subset bootstrap) in Figure 7(a) and parametric in Figure 7(b) techniques as described in [43].

As can been seen from Figure 7(a), the minimal effort mimicking and "friendly testing" FAR are similar (i.e., black and red curves). This indicates that mimicking does not help to improve the acceptance chances of impostors. However, impostors who know their closest person in the database (green FAR curve) can pose a serious threat to the gait-based user authentication. The FAR curves in Figure 7(b) suggest that impostor attempts, which are matched against the same gender have higher chances of being wrongfully accepted by the system compared to the different sex matching.

4. Uniqueness of Gait Biometric

In the third research question, we investigated some aspects relating or influencing the uniqueness of gait biometric in case of ankle/foot motion [44]. The following three



(a) Friendly testing, mimicking and closest person scenarios

(b) Same gender versus different gender

Figure 7: Security assessment in terms of FAR curves.

aspects were studied: footwear characteristics, directions of the motion, and gait cycle parts.

4.1. Experimental Data and Recognition Method. The number of subjects who participated in this experiment was 30. All of them were male, since only men footwears were used. Each subject walked with 4 specific types of footwear, labeled as A, B, C, and D. The photos of these shoe types are given in Figure 8. The footwear types were selected such that people wear them on different occasions. Each subject walked 4 times with every shoe type and the MRS was attached to the ankle as shown in the Figure 3(a). In each of the walking trials, subjects walked using their natural gait for the distance of about 20 m. The number of gait samples per subject was 16 ($= 4 \times 4$) and the total number of walking samples was 480 ($= 4 \times 4 \times 30$).

The gait recognition method applied here follows the architecture depicted in Figure 4. The difference is that in preprocessing stage we did not compute resultant acceleration but rather analyzed the three acceleration signals separately. In the analyses, we used the averaged cycle as a feature vector and applied an ordinary Euclidean distance (except in Section 4.4), see (2), for computing similarity scores

$$s = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}, \quad n = 100. \tag{2}$$

In this formula, a_i and b_i are acceleration values in two averaged gait cycles (i.e., test and template). The s is a similarity score between these two gait cycles.

4.2. Footwear Characteristic. Shoe or footwear is an important factor that affects the gait of the person. Studies show that when the test and template gait samples of the person are collected using different shoe types, performance can significantly decrease [45]. In many previous gait recognition experiments, subjects were walking with their own footwear "random footwear." In such settings, a system authenticates *person plus shoe* rather than the *person* per se. In our experimental setting, all participants walked with the same types of footwear which enables to eliminate the noise introduced by the footwear variability. Furthermore, subjects walked with several types of specified footwear. This allows investigating the relationship of the shoe property (e.g., weight) on recognition performance without the effect of "random footwear."

The resulting DET curves with different shoe types in each directions of the motion are given in Figure 9. The EERs of the curves are depicted in the legend of the figures and also presented in Table 6. In this table, the last two columns, FAR and FRR, indicate the EERs' margin of errors (i.e., 95% confidence intervals) for FAR and FRR, respectively. Confidence intervals were computed using parametric approach as in [43].

Although some previous studies reported performance decrease when the test and template samples of the person's walking were obtained using different shoe types [45], there was no attempt to verify any relationship between the shoe attributes and recognition performance. Several characteristics of the footwear can significantly effect gait of the person. One of such attributes is the weight of the shoe. One of the primary physical differences among shoes was in

A (a) B (b) C (c) D (d)

Figure 8: The footwear types A, B, C, and D.

their weight. The shoe types A/B were lighter and smaller than the shoe types C/D. As can be observed from the curves in Figure 9, in general performance is better with the light shoes (i.e., A and B) compared to the heavy shoes (i.e., C and D) in all directions. This suggests that the distinctiveness of gait (i.e., ankle motion) can diminish when wearing heavy footwear.

4.3. Directions of the Motion. Human motion occurs in 3 dimensions (3D): up-down (X), forward-backwards (Y), and sideway (Z). The MRS enables to measure acceleration in 3D. We analyzed performance of each direction of the motion separately to find out which direction provides the most discrimination.

The resulting DET curves for each direction of the motion for every footwear type are given in Figure 10. The EERs of the curves are depicted in the legend of the figures and also presented in Table 6. From Figure 10 one can observe that performance of the sideway acceleration (blue dashed curve) is the best compared to performances of the up-down (black solid curve) or forward-backward (red dotted curve) for all footwear types.

In addition, we also present performance for each direction of the motion regardless of the shoe type. In this case, we conducted comparisons of gait samples by not taking into account with which shoe type it was collected. For example, gait sample with shoe type A was compared to gait samples with shoe types B, C, and D (in addition to other gait samples with shoe type A). These DET curves are depicted in Figure 11 (EERs are also presented in Table 6, last three rows). This figure also clearly indicates that the discriminative performance of the sideway motion is the best compared to the other two.

Algorithms in VS-based gait recognition usually use frontal images of the person, where only up-down and forward-backward motions are available but not the sideway motion. In addition, in some previous WS-based studies [21, 30, 34], authors were focusing only on two directions of the motion: up-down and forward-backward. This is perhaps due to the fact that their accelerometer sensor was attached to the waist (see Figure 1) and there is less sideways movement of the waist compared to the foot. However, our analysis of ankle/foot motion revealed that the sideway direction of the motion provides more discrimination compared to the other two directions of the motion. Interestingly from biomechanical research, Cavanagh [46] also observed that the runners express their individuality characteristics in medio-lateral (i.e., sideway) shear force.

4.4. Gait Cycle Parts. The natural gait of the person is a periodic process and consists of cycles. Based on the foot motion, a gait cycle can be decomposed into several subevents, such as initial contact, loading response, midstance, initial swing and so on [47]. To investigate how various gait cycle parts contribute to recognition, we introduced a technique for analyzing contribution from each acceleration sample in the gait cycle.

Let the

$$d = \begin{vmatrix} d_{11} & \dots & d_{1n} \\ d_{21} & \dots & d_{2n} \\ \dots & \dots & \dots \\ d_{m1} & \dots & d_{mn} \end{vmatrix}, \qquad \delta = \begin{vmatrix} \delta_{11} & \dots & \delta_{1n} \\ \delta_{21} & \dots & \delta_{2n} \\ \dots & \dots & \dots \\ \delta_{k1} & \dots & \delta_{kn} \end{vmatrix} \tag{3}$$

be genuine and impostor matrices, respectively, ($m < k$, since usually the number of genuine comparisons is less than number of impostor comparisons). Each row in the matrices is a difference vector between two averaged cycles. For instance, assume $R = r_1, \dots, r_n$ and $P = p_1, \dots, p_n$ two feature vectors (i.e., averaged cycles) then values d_{ij} and δ_{ij} in row i in above matrices equal to

(i) $d_{ij} = |r_j - p_j|$, if S and P from the same person (i.e., genuine),

(ii) $\delta_{ij} = |r_j - p_j|$, if S and P from different person (i.e., impostor), where $j = 1, \dots, n$.

Based on matrices 2 and 3, we define weights w_i as follows:

$$w_i = \frac{\text{Mean}(\delta_{(i)})}{\text{Mean}(d_{(i)})}, \tag{4}$$

where $\text{Mean}(\delta_{(i)})$ and $\text{Mean}(d_{(i)})$ are the means of columns i in matrices δ and d, respectively. Then, instead of the ordinary Euclidean distance as in (2), we used a weighted

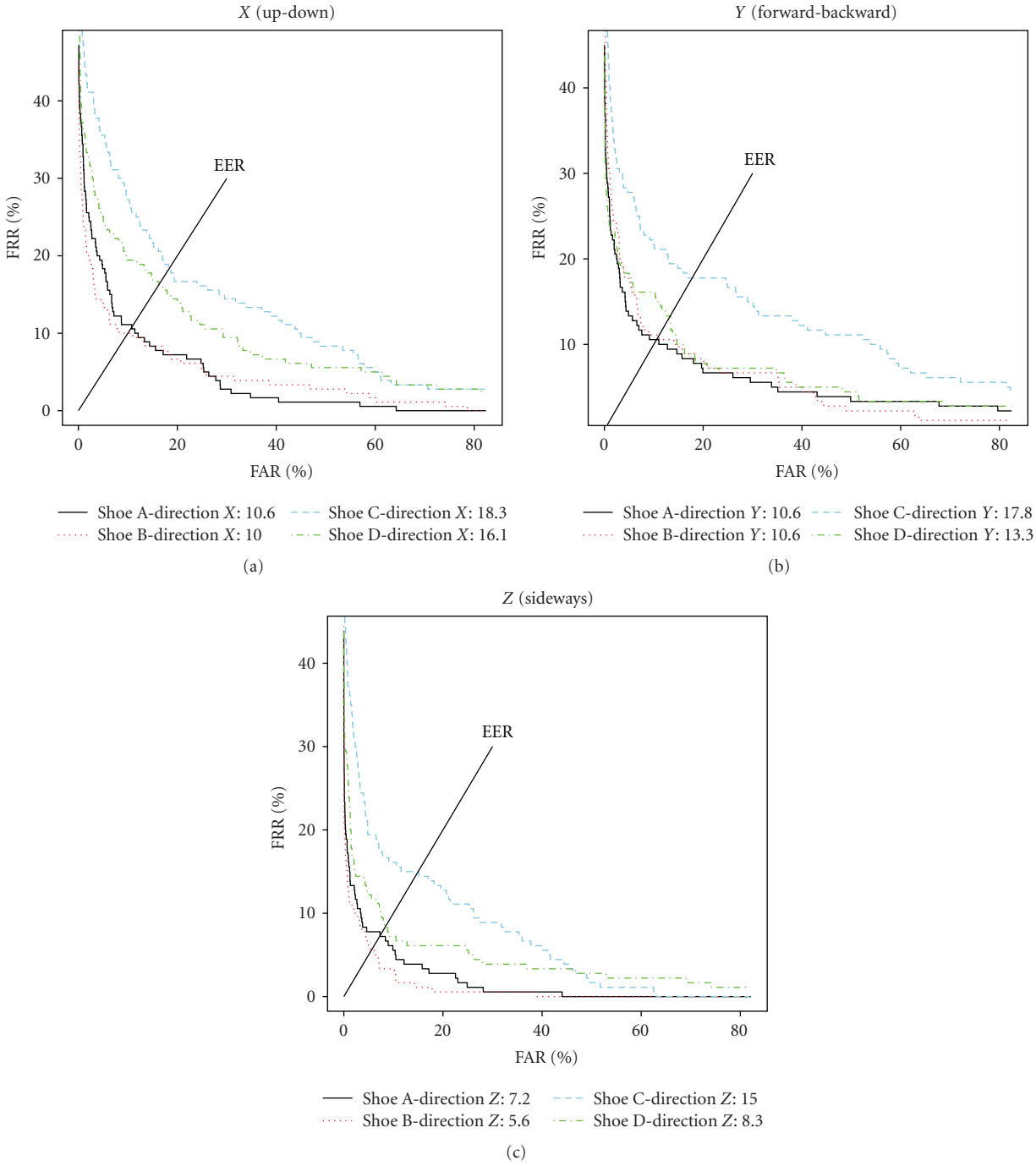


Figure 9: Authentication with respect to footwear types for each direction.

version of it as follows:

$$s = \sqrt{\sum_{i=1}^{n} (w_i - 1) * (a_i - b_i)^2}, \quad n = 100, \tag{5}$$

where w_i are from (4). We subtracted 1 from w_i's because if the Mean($\delta_{(i)}$) and Mean($d_{(i)}$) are equal than one can assume that there is no much discriminative information in that particular point.

We used gait samples from one shoe type (type B) to estimate weights and then tested them on gait samples from the other shoe types (i.e., types A, C, and D). The estimated weights are shown in Figure 12. The resulting DET curves are presented in Figure 13 and their EER are also given in Table 7. The DET curves indicate that performance of the weighted approach (red dotted curve) is better than the unweighted one (black solid curve), at least in terms of EER. This is in its turn may suggest that various gait cycle parts (or gait subevents) contribute differently to the recognition.

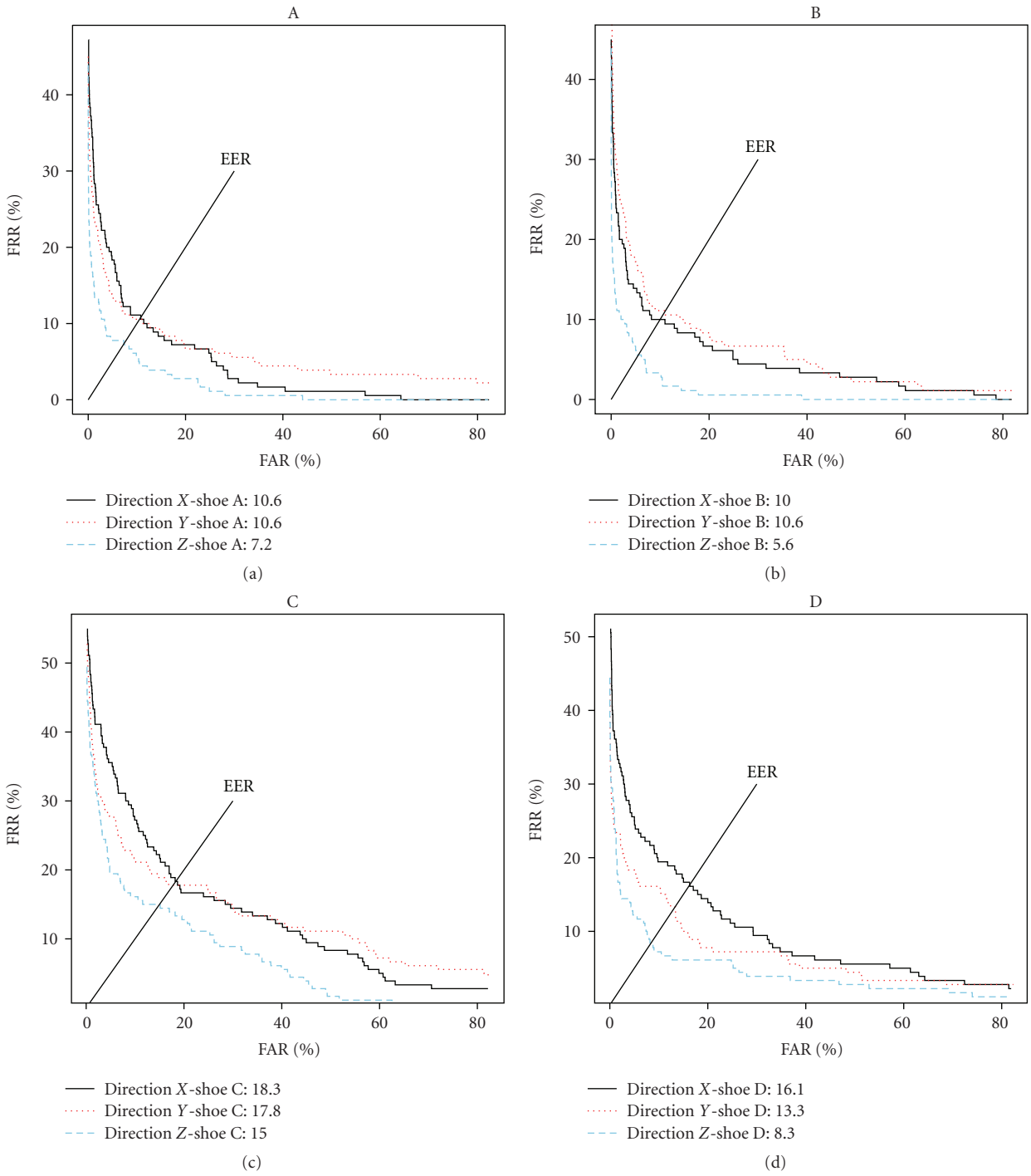


Figure 10: Authentication with respect to directions for shoe types A, B, C, and D.

5. Application and Limitation

5.1. Application. A primary advantage of the WS-based gait recognition is on its application domain. Using small, low-power, and low-cost sensors it can enable a periodic (dynamic) reverification of user identity in personal electronics. Unlike one time (static) authentication, periodic reverification can ensure the correct identity of the user all the time by reassuring the (previously authenticated) identity. An important aspect of periodic identity reverification is unobtrusiveness which means not to be annoying, not to distract user attention, and to be user friendly and convenient in frequent use. Consequently, not all authentication methods are unobtrusive and suitable for periodic reverification.

In our experiments, the main reason for selecting places on the body was driven by application perspectives. For

Table 6: EERs of the methods. Numbers are given in %.

Shoe type	Motion direction	EER	FAR	FRR
Shoe type A	X (up-down)	10.6	± 0.7	± 4.5
Shoe type B	X (up-down)	10	± 0.7	± 4.4
Shoe type C	X (up-down)	18.3	± 0.9	± 5.6
Shoe type D	X (up-down)	16.1	± 0.9	± 5.4
Shoe type A	Y (forw.-backw.)	10.6	± 0.7	± 4.5
Shoe type B	Y (forw.-backw.)	10.6	± 0.7	± 4.5
Shoe type C	Y (forw.-backw.)	17.8	± 0.9	± 5.6
Shoe type D	Y (forw.-backw.)	13.3	± 0.8	± 5
Shoe type A	Z (sideway)	7.2	± 0.6	± 3.8
Shoe type B	Z (sideway)	5.6	± 0.5	± 3.4
Shoe type C	Z (sideway)	15	± 0.8	± 5.2
Shoe type D	Z (sideway)	8.3	± 0.6	± 4
—	X (up-down)	30.5	± 0.3	± 1.5
—	Y (forw.-backw.)	29.9	± 0.3	± 1.5
—	Z (sideway)	23	± 0.2	± 1.4

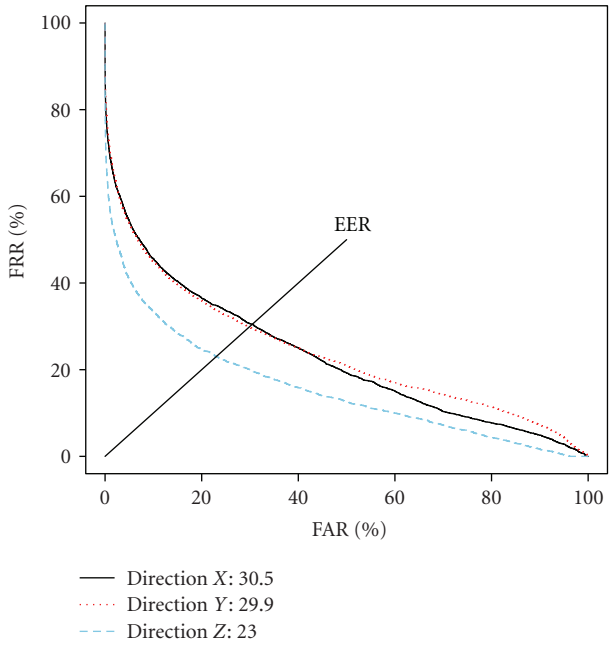


Figure 11: Authentication regardless of the shoe types.

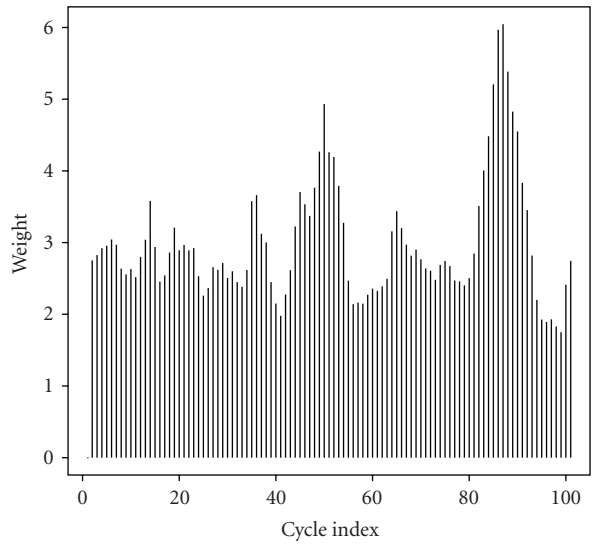


Figure 12: The estimated weights.

Table 7: The unweighted (EER) and weighted distances (EER_w).

Shoe type	Motion direction	EER, %	EER_w, %
Shoe type A	Z (sideway)	7.2	5
Shoe type C	Z (sideway)	15	12.8
Shoe type D	Z (sideway)	8.3	7.8

example, people can carry mobile phone in similar position on the hip or in the pocket. Some models of the mobile phones already equipped with accelerometer sensor, for example, Apple's iPhone [50] has the accelerometer for detecting orientation of the phone. Nowadays the mobile phone services go beyond mere voice communication, for example, users can store their private data (text, images, videos, etc.) and use it in high security applications such as mobile banking or commerce [51, 52]. All of these increase the risk of being the target of an attack not only because of the phone value per se but also because of the stored information and provided services. User authentication in mobile phones is static, that is, users authenticated once and authentication remains all the time (until the phone explicitly is turned off). In addition, surveys indicate high crimes associated with mobile phones [53] and also suggest that users do not follow the relevant security guidelines, for example, use the same code for multiple services [54].

For combating crimes and improving security in mobile phones, a periodic reverification of the authenticated user is highly desirable. The PIN-based authentication of mobile

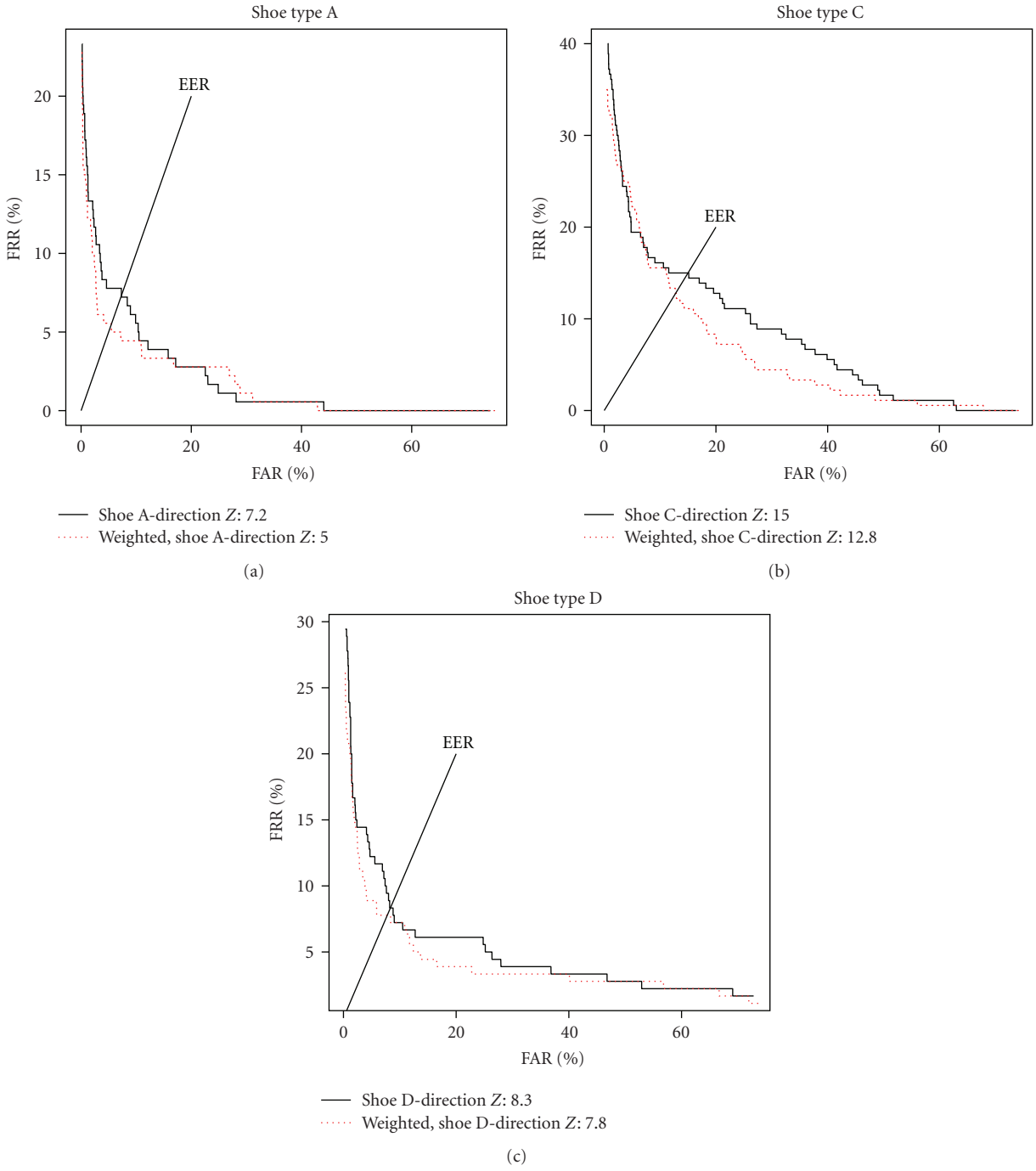


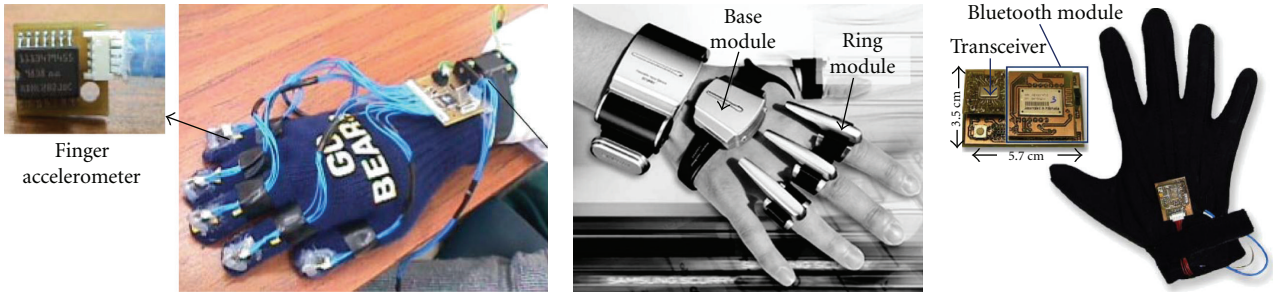
Figure 13: Ordinary (black) vrsus weighted (red) Euclidean distances.

(a) By Chen et al. [48]

(b) By Yamamoto et al. [49]

Figure 14: Examples of smart shoes with integrated accelerometer.

phones is difficult or impossible to adapt for periodic reauthentication because of its obtrusiveness. Indeed, the process of frequently entering the PIN code into a mobile phone is explicit, requires user cooperation, and can be very inconvenient and annoying. Therefore, the WS gait-based analysis can offer better opportunities for periodic identity reverification using MRS embedded in phone hardware or user's clothes (e.g., shoes). Whenever a user makes a few steps his identity is re-verified in a background, without requiring an explicit action or input from the user.



(a) Glove by Sama et al. [55] (b) Glove by Kim et al. [56] (c) Glove by Perng et al. [57]

Figure 15: Examples of the glove like input devices with built-in accelerometer.

Besides the mobile phones and thanks to the rapid miniaturization of electronics, the motion recording/detecting sensors can be found in a wide range of other consumer electronics, gadgets, and clothes. For example,

(i) laptops use accelerometer sensors for drop protection of their hard drive [58];

(ii) various intelligent shoes with integrated sensors are developed (see Figure 14), for example, for detecting abnormal gaits [48], for providing foot motion to the PC as an alternative way of input [49]; Apple and Nike jointly developed a smart shoes that enables the Nike+ footwear to communicate with iPod to provide pedometer functions [59];

(iii) glove like devices with built-in accelerometer (see Figure 15) can detect and translate finger and hand motions as an input to the computer [55–57];

(iv) watches or watch like electronics are equipped with built-in accelerometer sensor [60]. Motion detecting and recording sensors can be built-in even in some exotic applications like tooth brushing [61] or wearable e-textile [62]; and many others.

As the values and services provided by such electronics grow, their risk of being stolen increases as well. Although the motion recording/detecting sensors in the aforementioned products and prototypes are mainly intended for other purposes, it is possible to extend their functionality for periodic re-verification of identity too. Depending on the computing resources, the motion signal can either be analyzed locally (e.g., in case of mobile phones) or remotely in the other surrounding electronics to which data is transferred wirelessly. For instance, a shoe system can transfer the foot motion to the user's computer via wireless network (e.g., Bluetooth).

Furthermore, it is foreseen that such sensors will become a standard feature in many kind of consumer products [63, 64] which implies that WS-based approach will not require an extra hardware. However, it is worth noting that we do not propose the WS-based authentication as a sole or replacement, but rather a complementary one to the traditional authentication techniques (i.e., PIN-code, fingerprint, etc.).

5.2. Limitation. Like the other biometrics, the WS-based gait recognition also possesses its own limitations and challenges. Although the WS-based approach lacks difficulties associated with VS-based approach like noisy background, lighting conditions, and viewing angles, it shares the common factors that influence gait such as walking speed, surface conditions, and foot/leg injuries.

An important challenge related to the WS-based gait recognition includes distinguishing various patterns of walking. Although our methods can differentiate the actual normal walking from the standing still, usually daily activity of an ordinary user involves different types of gait (running, walking fast/slow, walking on stairs up/down, walking with busy hands, etc.). Consequently, advanced techniques are needed for classifying among various complex patterns of daily motion.

The main limitation of the behavioral biometrics including gait is a relatively low performance. Usually, performance of the behavioral biometrics (e.g., voice, handwriting, gait, etc.) is not as accurate as the biometrics like fingerprint or iris. Some ways to improve accuracy can be combining WS-based gait with the other biometrics (e.g., voice [31]), fusing motion from different places (e.g., foot and hip), and/or sensor types (e.g., accelerometer, gyro, etc.). Nevertheless, despite low accuracy of the WS-based gait recognition, it can still be useful as a supplementary method for increasing security by unobtrusive and periodic reverification of the identity. For instance, to reduce inconvenience for a genuine user, one can select a decision threshold where the FRR is low or zero but the FAR is medium to high. In such setting, although the system cannot completely remove impostors of being accepted, it can reduce such risk significantly.

Due to the lack of processing unit in the current prototype of the MRS, our analyses were conducted offline, that is, after walking with MRS, the recorded accelerations were transferred to the computer for processing. However, with computing resources available in some of current electronics we believe it is feasible to analyze motion signals online (i.e., localy) too.

6. Conclusion

In this paper, we presented gait recognition approach which is significantly different from most of current gait biometric research. Our approach was based on analyzing motion signals of the body segments, which were collected by using wearable sensors. Acceleration signals from ankle, hip, trousers pocket, and arm were utilized for person recognition. Analyses of the acceleration signals from these body segments indicated some promising performances. Such gait analysis offers an unobtrusive and periodic (re-)verification of user identity in personal electronics (e.g., mobile phone).

Furthermore, we reported our results on security assessment of gait-based authentication in the case of hip motion. We studied security of the gait-based user authentication under three attack scenarios which were minimal effort-mimicry, knowing the closest person in the database (in terms of gait similarity), and knowing the gender of the user in the database. The findings revealed that the minimal effort mimicking does not help to improve the acceptance chances of impostors. However, impostors who knew their closest person in the database or the gender of the users in the database could pose a threat to the gait-based authentication approach.

In addition, we provided some new insights toward understanding the uniqueness of the gait in case of ankle/foot motion with respect to the shoe attribute, axis of the motion, and gait cycle parts. In particular, our analysis showed that heavy footwear tends to diminish gait's discriminative power and the sideway motion of the foot provides the most discrimination compared to the up-down or forward-backward direction of the motion. Our analysis also revealed that various gait cycle parts (i.e., subevents) contribute differently toward recognition performance.

Acknowledgment

This work is supported by the Research Council of Norway, Grant no. NFR158605/V30(431) "Security of approaches to personnel authentication."

References

[1] N. L. Clarke and S. M. Furnell, "Authenticating mobile phone users using keystroke analysis," *International Journal of Information Security*, vol. 6, no. 1, pp. 1–14, 2007.

[2] A. A. E. Ahmed and I. Traore, "A new biometrie technology based on mouse dynamics," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 3, pp. 165–179, 2007.

[3] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: a machine learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[4] F. Beritelli and S. Serrano, "Biometric identification based on frequency analysis of cardiac sounds," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 596–604, 2007.

[5] Y. Chai, J. Ren, R. Zhao, and J. Jia, "Automatic gait recognition using dynamic variance features," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR '06)*, pp. 475–480, Southampton, UK, April 2006.

[6] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "Eigen-gait: motion-based recognition of people using image self-similarity," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '01)*, Halmstad, Sweden, June 2001.

[7] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '01)*, pp. 272–277, Halmstad, Sweden, June 2001.

[8] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.

[9] T. H. W. Lam and R. S. T. Lee, "A new representation for human gait recognition: Motion Silhouettes Image (MSI)," in *Proceedings of International Conference on Biometrics (ICB '06)*, pp. 612–618, Hong Kong, January 2006.

[10] M. S. Nixon, T. N. Tan, and R. Chellappa, *Human Identification Based on Gait*, Springer, New York, NY, USA, 2006.

[11] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *Proceedings of the 4th International Conference on Recent Advances in Soft Computing*, pp. 66–71, 2002.

[12] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon, "University of southampton multi-biometric tunnel and introducing a novel 3d gait dataset," in *Proceedings of the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008.

[13] G. Zhao, L. Cui, H. Li, and M. Pietikainen, "Gait recognition using fractal scale and wavelet moments," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 453–456, Hong Kong, August 2006.

[14] S. Hong, H. Lee, K. Oh, M. Park, and E. Kim, "Gait recognition using sampled point vectors," in *Proceedings of SICE-ICASE International Joint Conference*, pp. 3937–3940, 2006.

[15] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 357–362, Washington, DC, USA, May 2002.

[16] Y. Wang, S. Yu, Y. Wang, and T. Tan, "Gait recognition based on fusion of multi-view gait sequences," in *Proceedings of the International Conference on Biometrics (ICB '06)*, pp. 605–611, Hong Kong, January 2006.

[17] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1120–1131, 2003.

[18] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 149–158, 2004.

[19] A. I. Bazin, L. Middleton, and M. S. Nixon, "Probabilistic combination of static and dynamic gait features for verification," in *Proceedings of the 8th International Conference of Information Fusion*, 2005.

[20] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon, "A floor sensor system for gait recognition," in *Proceedings of*

the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AUTO ID '05), pp. 171–180, New York, NY, USA, October 2005.

[21] H. J. Ailisto, M. Lindholm, J. Mäntyjärvi, E. Vildjiounaite, and S.-M. Mäkelä, "Identifying people from gait pattern with accelerometers," in *Biometric Technology for Human Identification II*, vol. 5779 of *Proceedings of SPIE*, pp. 7–14, Orlando, Fla, USA, March 2005.

[22] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 441–444, Hong Kong, August 2006.

[23] K. Nakajima, Y. Mizukami, K. Tanaka, and T. Tamura, "Footprint-based personal recognition," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 11, pp. 1534–1537, 2000.

[24] J. Suutala and J. Röning, "Towards the adaptive identification of walkers: automated feature selection of footsteps using distinction sensitive LVQ," in *Proceedings of the International Workshop on Processing Sensory Information for Proactive Systems (PSIPS '04)*, June 2004.

[25] J. Suutala and J. Röning, "Combining classifiers with different footstep feature sets and multiple samples for person identification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 5, pp. 357–360, Philadelphia, Pa, USA, March 2005.

[26] J. Suutala and J. Röning, "Methods for person identification on a pressure-sensitive floor: experiments with multiple classifiers and reject option," *Information Fusion*, vol. 9, no. 1, pp. 21–40, 2008.

[27] R. J. Orr and G. D. Abowd, "The smart floor: a mechanism for natural user identification and tracking," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '00)*, Hague, The Netherlands, April 2000.

[28] J. Jenkins and C. S. Ellis, "Using ground reaction forces from gait analysis: body mass as a weak biometric," in *Proceedings of the International Conference on Pervasive Computing (Pervasive '07)*, 2007.

[29] S. J. Morris, *A shoe-integrated sensor system for wireless gait analysis and real-time therapeutic feedback*, Ph.D. thesis, Division of Health Sciences and Technology, Harvard University-MIT, Cambridge, Mass, USA, 2004.

[30] J. Mäntyjärvi, M. Lindholm, E. Vildjiounaite, S.-M. Mäkelä, and H. J. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 2, pp. 973–976, Philadelphia, Pa, USA, March 2005.

[31] E. Vildjiounaite, S.-M. Mäkelä, M. Lindholm, et al., "Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices," in *Proceedings of the 4th International Conference on Pervasive Computing (Pervasive '06)*, Lecture Notes in Computer Science, pp. 187–201, Dublin, Ireland, May 2006.

[32] B. Huang, M. Chen, P. Huang, and Y. Xu, "Gait modeling for human identification," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '07)*, pp. 4833–4838, Rome, Italy, April 2007.

[33] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, "A wearable acceleration sensor system for gait recognition," in *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications (ICIEA '07)*, Harbin, China, May 2007.

[34] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," in *Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering*, 2007.

[35] M. Sekine, Y. Abe, M. Sekimoto, et al., "Assessment of gait parameter in hemiplegic patients by accelerometry," in *Proceedings of the 22nd Annual International Conference of the IEEE on Engineering in Medicine and Biology Society*, vol. 3, pp. 1879–1882, 2000.

[36] D. Alvarez, R. C. Gonzalez, A. Lopez, and J. C. Alvarez, "Comparison of step length estimators from weareable accelerometer devices," in *Proceedings of the 28th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBS '06)*, pp. 5964–5967, New York, NY, USA, August 2006.

[37] D. Gafurov and E. Snekkenes, "Arm swing as a weak biometric for unobtrusive user authentication," in *Proceedings of IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008.

[38] D. Gafurov, K. Helkala, and T. Sondrol, "Gait recognition using acceleration from MEMS," in *Proceedings of the 1st International Conference on Availability, Reliability and Security (ARES '06)*, pp. 432–437, Vienna, Austria, April 2006.

[39] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, 2007.

[40] D. Gafurov, E. Snekkenes, and P. Bours, "Gait authentication and identification using wearable accelerometer sensor," in *Proceedings of the 5th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID '07)*, pp. 220–225, Alghero, Italy, June 2007.

[41] ISO/IEC IS 19795-1, Information technology, biometric performance testing and reporting—part 1: principles and framework, 2006.

[42] D. Gafurov, "Security analysis of impostor attempts with respect to gender in gait biometrics," in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '07)*, Washington, DC, USA, September 2007.

[43] R. M. Bolle, N. K. Ratha, and S. Pankati, "Error analysis of pattern recognition systems—the subsets bootstrap," *Computer Vision and Image Understanding*, 2004.

[44] D. Gafurov and E. Snekkenes, "Towards understanding the uniqueness of gait biometric," in *Proceedings of the 8th IEEE International Conference Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, September 2008.

[45] S. Enokida, R. Shimomoto, T. Wada, and T. Ejima, "A predictive model for gait recognition," in *Proceedings of the Biometric Consortium Conference (BCC '06)*, Baltimore, Md, USA, September 2006.

[46] Cavanagh, "The shoe-ground interface in running," in *The Foot and Leg in Running Sports*, R. P. Mack, Ed., pp. 30–44, 1982.

[47] C. Vaughan, B. Davis, and J. O'Cononor, *Dynamics of Human Gait*, Kiboho, 1999.

[48] M. Chen, B. Huang, and Y. Xu, "Intelligent shoes for abnormal gait detection," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '08)*, pp. 2019–2024, Pasadena, Calif, USA, May 2008.

[49] T. Yamamoto, M. Tsukamoto, and T. Yoshihisa, "Footstep input method for operating information devices while jogging," in *Proceedings of the International Symposium on*

Applications and the Internet (SAINT '08), pp. 173–176, Turku, Finland, August 2008.

[50] Apple's iphone with integrated accelerometer, April 2008, http://www.apple.com/iphone/features/index.html.

[51] K. Pousttchi and M. Schurig, "Assessment of today's mobile banking applications from the view of customer requirements," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS '04)*, vol. 37, pp. 2875–2884, Big Island, Hawaii, USA, January 2004.

[52] B. Dukić and M. Katić, "m-order- payment model via SMS within the m-banking," in *Proceedings of the 27th International Conference on Information Technology Interfaces (ITI '05)*, pp. 99–104, Cavtat, Croatia, June 2005.

[53] Mobile phone theft, plastic card and identity fraud: findings from the 2005/06 british crime survey, April 2008, http://www.homeoffice.gov.uk/rds/pdfs07/hosb1007.pdf.

[54] N. L. Clarke and S. M. Furnell, "Authentication of users on mobile telephones—a survey of attitudes and practices," *Computers and Security*, vol. 24, no. 7, pp. 519–527, 2005.

[55] M. Sama, V. Pacella, E. Farella, L. Benini, and B. Riccó, "3dID: a low-power, low-cost hand motion capture device," in *Proceedings of Design, Automation and Test in Europe (DATE '06)*, vol. 2, Munich, Germany, March 2006.

[56] Y. S. Kim, B. S. Soh, and S.-G. Lee, "A new wearable input device: SCURRY," *IEEE Transactions on Industrial Electronics*, vol. 52, no. 6, pp. 1490–1499, 2005.

[57] J. K. Perng, B. Fisher, S. Hollar, and K. S. J. Pister, "Acceleration sensing glove (ASG)," in *Proceedings of the 3rd International Symposium on Wearable Computers*, pp. 178–180, San Francisco, Calif, USA, October 1999.

[58] Active protection system, January 2009, http://www.pc.ibm.com/europe/think/en/aps.html?europe&cc=europe.

[59] E. de Lara and K. Farkas, "New products," *IEEE Pervasive Computing*, 2006.

[60] C. Narayanaswami, "Form factors for mobile computing and device symbiosis," in *Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR '05)*, pp. 335–339, Seoul, South Korea, September 2005.

[61] K.-H. Lee, J.-W. Lee, K.-S. Kim, et al., "Tooth brushing pattern classification using three-axis accelerometer and magnetic sensor for smart toothbrush," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '07)*, pp. 4211–4214, Lyon, France, August 2007.

[62] L. Buechley and M. Eisenberg, "The LilyPad Arduino: toward wearable engineering for everyone," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 12–15, 2008.

[63] B. Vigna, "Future of MEMS: an industry point of view," in *Proceedings of the 7th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems (EuroSimE '06)*, Como, Italy, April 2006.

[64] B. Vigna, "More than Moore: micro-machined products enable new applications and open new markets," in *Proceedings of the International Electron Devices Meeting (IEDM '05)*, pp. 1–8, Washington, DC, USA, December 2005.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 896383, 9 pages
doi:10.1155/2009/896383

Research Article

Intersubject Differences in False Nonmatch Rates for a Fingerprint-Based Authentication System

Jeroen Breebaart, Ton Akkermans, and Emile Kelkboom

Philips Research, HTC 34 MS61, 5656 AE Eindhoven, The Netherlands

Correspondence should be addressed to Jeroen Breebaart, jeroen.breebaart@philips.com

Received 4 September 2008; Accepted 7 July 2009

Recommended by Jonathon Phillips

The intersubject dependencies of false nonmatch rates were investigated for a minutiae-based biometric authentication process using single enrollment and verification measurements. A large number of genuine comparison scores were subjected to statistical inference tests that indicated that the number of false nonmatches depends on the subject and finger under test. This result was also observed if subjects associated with failures to enroll were excluded from the test set. The majority of the population (about 90%) showed a false nonmatch rate that was considerably smaller than the average false nonmatch rate of the complete population. The remaining 10% could be characterized as "goats" due to their relatively high probability for a false nonmatch. The image quality reported by the template extraction module only weakly correlated with the genuine comparison scores. When multiple verification attempts were investigated, only a limited benefit was observed for "goats," since the conditional probability for a false nonmatch given earlier nonsuccessful attempts increased with the number of attempts. These observations suggest that (1) there is a need for improved identification of "goats" during enrollment (e.g., using dedicated signal-driven analysis and classification methods and/or the use of multiple enrollment images) and (2) there should be alternative means for identity verification in the biometric system under test in case of two subsequent false nonmatches.

1. Introduction

The use of biometric characteristics for identity verification has been described as security enhancement *on top* of something one has (e.g., a card) and/or something one knows (e.g., a password) in many publications. The main reason for biometrics-enhanced identity management in such two or three-factor authentication approach is to reduce the risk of identity theft by increasing the difficulty of impersonation. In less critical applications, biometrics have also been proposed as *replacement* for passwords. With the ever-increasing number of login codes, passwords, and personal identification numbers (PINs), there is a strong need to reduce the amount of information that individuals have to memorize. Biometrics could provide a convenient solution for this increasing memory burden.

The use of biometrics in forensic context exists for a very long time. Around 1880, Dr. Henry Faulds recognized the importance of fingerprints for identification. In the 1890s, Alphonse Bertillon, a French anthropologist and police desk clerk used multiple body measurements to identify convicted criminals. Later Richard Edward Henry of Scotland Yard started to use fingerprints for the same purpose. These early methods all employed manual measurement and comparison for identification. Only during the last few decades, *automated* biometric identity verification systems have been introduced and have been subject to extensive research. One of the leitmotifs in biometrics research is the verification performance, expressed in average performance characteristics such as false acceptance rates (FARs), false rejection rates (FRR), and equal error rates (EERs). In most applications, the FAR and FRR are subject to a tradeoff; by modifying a comparison threshold value, security (expressed by the FAR) can be enhanced at the expense of a decreased convenience (expressed by the FRR) and vice versa, resulting in a detection error tradeoff (DET) curve. Similarly, performance tests on a subpopulation (excluding failures to acquire or to enroll) are expressed in terms of false match rates (FMRs) and false nonmatch rates (FNMR). In many cases, a biometric verification performance is characterized

by an FRR (or FNMR) at a specific FAR (or FMR) which is typically 0.01 or 0.001. These performance measures heavily depend on the biometric modality, the sensor type, the type of processing, and the corpus that is being used. As examples, various systems for face and iris recognition report an FRR in the range of 0.005–0.05 at an FAR of 0.001 [1–3]; however, significantly worse performance is also reported [4, 5]. For fingerprints, best-in-class solutions typically provide an FRR of approximately 0.04 at an FAR of 0.001 and about 0.03 for an FAR of 0.01 [6, 7].

Although the DET curve provides very meaningful information on the *average* performance of a biometric verification system given a certain population or corpus, it does not describe possible intersubject dependencies in verification performance. Subjects of a biometric system have been categorized into "sheep," "goats," "lambs" and "wolves," depending on their average (intrasubject) genuine and imposter scores [8]. This menagerie has been extended with "worms," "chameleons," "phantoms," and "doves" [9]. There is evidence that a number of these types indeed exist for certain biometric modalities and corpora [8–11] although the presence of "goats" in fingerprint corpora seems subject to debate [12].

The "goats" represent subjects that are difficult to recognize. They account for a disproportionate share of false nonmatches. These false nonmatches may cause these subjects to experience the system as being "inconvenient" possibly resulting in a decreased trust in a certain application which may also have negative consequences for their trust in biometrics in general. As an example, it has been shown that the age band has an effect on the performance of a fingerprint-based biometric system [3, 13] which suggests that elderly people may more often be associated with "goats" than younger people.

Failures to enroll or acquire may also pose challenges on biometric verification systems with regard to convenience. Subjects may have unreliable or absent biometric characteristics or body parts. It has often been assumed that between one and three percent of the general public does not have suitable biometric characteristics (cf. [2, 4]). A further challenge is caused by subjects that *refuse* to enroll. Subjects have in particular circumstances the right to object against the processing of biometric data on compelling legitimate grounds such as privacy concerns [14]. Other concerns may comprise health effects induced by biometric measurements, hygiene issues, the risk of stolen body parts containing a biometric, or negative associations such as fingerprints and crime. Surveys held in the US between 2001 and 2005 indicated that about 6–10% of the Americans found the use of finger and hand scan biometrics for law enforcement and governmental applications not acceptable [15, 16]. In the commercial sector, a similar proportion of the population found it not acceptable to use biometrics for credit card transactions or Automated Teller Machines (ATMs). In Europe, a majority of consumers (92%) now believe that a fingerprint is more secure than a signature, and 84% believes that biometrics are more secure than Chip and PIN [17].

Summarizing, failure of correct authentication in a biometric system may be associated with subjects that do not want to enroll, that cannot enroll, or that experience problems during verification. It is rather obvious that subjects belonging to the first two groups require an alternative means for authentication that is not based on biometrics. For the third group, the "goats," the situation is somewhat more subtle. This type is not easily detectable during enrollment if only a single measurement is available. Depending on the application, the difficulty to detect "goats" in an initial stage may jeopardize the success of a biometric verification system. For example, consider the case of biometrics-enabled ATMs. With billions of ATM transactions per month, a typical false nonmatch rate of 0.01 will result in a tremendous number of complaints, help desk calls, and service costs. Hence an FNMR of 0.01 will most likely not be acceptable for such an application and it will be crucial to understand and to mitigate the risk of false nonmatches. In conventional ATMs based on PIN authentication, a subject has multiple (typically 3) authentication attempts to resolve problems related with erroneously entered PINs. It is of interest to investigate the effect of such multiple verification attempts in a biometric authentication scheme and its influence on the resulting FNMR and FMR.

2. FNMR Analysis

Subject dependencies of FNMR have been found for speaker recognition [8], face recognition [11], and fingerprint recognition [9, 10, 12]. However, it has been argued that "hard-to-match" fingerprints are resulting from properties of a certain (low-quality) measurement, rather than resulting from individual biometric characteristics themselves [12]. Furthermore, although the existence of subject interdependencies has been shown by statistical inference tests, most studies do not provide a clear insight in the distribution of intersubject FNMRs.

2.1. Fingerprint Corpus. An analysis of intersubject FNMRs was carried out based on the Ministerio de Ciencia y Tecnología (MCYT) baseline fingerprint corpus [18]. This database contains 12 images of all 10 fingers from 330 subjects that were located in four different institutions. All combinations of image number, finger, and subject have been measured using two acquisition devices: one optical sensor (UareU from Digital Persona) and one capacitive sensor (model 100SC from Precise Biometrics). Both sensors were operating at a resolution of 500 dpi. All fingerprint capturing was accomplished by the supervision of an operator using three levels of control by the subject that differed in the amount of visual feedback with respect to finger placement provided on a computer screen. In a subjective quality assessment on a subset of the data, 5% of the images was found to be of very bad quality, 20% of low quality, 55% of medium quality, and 20% of high quality (see [18] for details).

The total amount of fingerprint images amounts thus $330 \times 10 \times 12 \times 2 = 79,200$ images. Since 12 measurements are available for each subject, finger, and sensor, the maximum number of unique genuine comparisons per subject, finger, and sensor equals 66 (under the assumption

that for all images a suitable minutiae template could be established). Hence for each finger and sensor, $66 \times 330 = 21\,780$ genuine comparisons can in principle be obtained, resulting in a total number of genuine comparisons per sensor of 217 800, and 435 600 in total.

All 79 200 images were converted to minutiae templates using a state-of-the-art commercially available minutiae-extractor and comparator solution. The minutiae-extractor also provides image quality ratings; the corresponding comparator solution operates symmetrically, that is, a comparison score of A with B is equal to B with A. Some of the images could not be converted to minutiae templates either due to a failure to acquire or a failure to enroll. Since the employed solution does not indicate whether a failure was due to acquisition or enrollment difficulties, we will refer to such failures as failures to enroll in the remainder of this paper. For the capacitive sensor, the system could not enroll one image for one finger of one subject. For four other subjects, none of the images from any finger could be enrolled. Hence out of the 39.600 images, $1 + 4 \times 12 \times 10 = 481$ images resulted in a failure to enroll, corresponding to a failure to enroll rate in terms of the number of images of 0.0121. For the optical sensor, 4 subjects could not be enrolled for one finger and one image. One subject could not enroll one image from two fingers. One subject could not enroll any image from any finger. Hence, in total, $4 + 2 + 120 = 126$ images resulted in a failure to enroll, which corresponds to a rate of 0.0032.

From these data, two databases containing comparison scores were constructed described as follows.

(1) A full database, containing all genuine comparison scores within the same sensor, resulting in 435 600 genuine comparison scores in total (217 800 for each sensor). Comparisons that involved an image that caused a failure to enroll were set to a similarity score of zero to ensure a reject irrespective of the (positive) comparison threshold. The imposter comparisons comprised a subset of 792 000 combinations. Tests on this database describe the FAR-FRR tradeoff (i.e., including the effect of failure to enroll).

(2) A balanced database containing only subjects for which *all* images could be enrolled. Hence the resulting database is fully balanced (i.e., the same number of fingers and measurements per finger for each subject). This process resulted in 214 500 genuine comparisons for the capacitive sensor (325 subjects), and 213 840 genuine comparisons for the optical sensor (324 subjects). The number of imposter comparisons amounted to 384 000 and 381 720, for the capacitive and optical sensors, respectively. Performance tests on this database are more closely related to FMRs and FNMRs, while minimizing the effect of low-quality data that could result in failure to enroll.

2.2. DET Curves. Separate DET curves were constructed for the optical and capacitive sensors from the genuine and imposter comparison scores. The results are visualized in Figure 1 for the full database. The solid line represents the capacitive sensor, the dashed line represents the optical sensor. The EER for the capacitive sensor amounts to 0.024; the EER for the optical sensor amounts to 0.0064. At a FAR of 0.001, the FRR for the capacitive and optical sensors amounts to 0.0295 and 0.0075, respectively. As can be observed, the optical sensor performs significantly better than the capacitive sensor: across the full DET curve, the FRR for the optical sensor is almost 4 times smaller than the FRR of the capacitive sensor for the same FAR. These results confirm earlier statement on quality differences between optical and capacitive sensors [19]. A similar analysis was performed for the balanced database. A comparison between the full and balanced database error rates is provided in Table 1. As can be observed, the EERs and FNMRs for the balanced database are about twice as low as for the full database (FRRs).

Table 1: Error rates for the two sensors and the two databases (the full database providing FRRs and the balanced database providing FNMRs).

Sensor	EER (full)	EER (bal)	FRR@0.001 (full)	FNMR@0.001 (bal)
Capacitive	0.0240	0.0138	0.0295	0.0181
Optical	0.0064	0.0034	0.0075	0.0040

2.3. Statistical Inferences. The existence of "goat-" like behavior is investigated using statistical inference tests. The data is tested to support the null-hypothesis that the genuine comparison scores do not depend on the subject or finger indices. A nonparametric (Kruskal-Wallis) test was employed on the genuine comparison scores from the balanced database. The Kruskal-Wallis test can only be employed to investigate one factor; hence, the test was performed four times to cover all combinations of the two sensors and the two effects under test (subject index and finger index). The results are provided in Table 2. All null hypotheses that the subject or finger index did not have any effect on the comparison scores are rejected based on the observed χ^2 values. Hence, it is concluded that the false nonmatch rates are subject to "goat-" like behavior.

The comparison scores were also subjected to a two-way analysis of variance (including interaction) with the finger index and subject as main effects, and the comparison score as dependent variable. The resulting F values and the corresponding probability of falsely rejecting the null hypothesis, that is, none of the effects or interactions is significant, are provided in the last two columns of Table 2. In line with the results obtained from the Kruskal-Wallis test, both factors and their interaction were found to have a significant effect on the comparison scores. The same analyses were also carried out on the full database which gave the same qualitative result.

2.4. Intersubject Distribution of FNMR. The presence of significant effects of subject and finger index on the comparison scores for both sensors does not provide any insight in the actual distribution of FNMRs across subjects or fingers. To investigate the range of FNMRs between subjects,

Table 2: Results for the Kruskal-Wallis test and analysis of variance (ANOVA) test for the optical and capacitive sensors. The factors that were taken into account were the subject index and the finger index. Tests were performed on the balanced database.

Sensor	Effect	df	χ^2	$p > \chi^2$	F	$p > F$
Capacitive	Subject	324	68276.05	0	510.08	0
Capacitive	Finger	9	5813.31	0	1621.18	0
Capacitive	Interaction	2916	n/a	n/a	44.62	0
Optical	Subject	323	66894.26	0	552.18	0
Optical	Finger	9	17047.04	0	5116.24	0
Optical	Interaction	2907	n/a	n/a	42.15	0

the number of false nonmatches within the set of all 66 genuine comparisons was computed for a threshold value that resulted in a global FMR of 0.001. The threshold was determined separately for each of the two sensors to compensate for performance differences between the sensors and was carried out on the balanced database.

In the following, the number of false nonmatches at a false match rate ϕ_i within a set of N genuine comparisons is given by $x_{i,j,k}[\phi_i]$ for sensor i, subject j, finger k. If one assumes that each of the N genuine comparisons for a given sensor i has a constant probability for a false nonmatch that only depends on the false match rate ϕ_i, the expected number $\mu_i[\phi_i]$ of false nonmatches within a set of $N = 66$ genuine comparisons would be given by

$$\mu_i[\phi_i] = N\psi_i[\phi_i], \tag{1}$$

with $\psi_i[\phi_i]$ the estimate of the probability of a false nonmatch $\psi_i[\phi_i]$ for a false match rate ϕ_i, given by

$$\psi_i[\phi_i] = \frac{\sum_j \sum_k x_{i,j,k}[\phi_i]}{JKN}. \tag{2}$$

In the absence of any intersubject or finger index dependencies, the variable $x_{i,j,k}[\phi_i]$ is then expected to follow a binomial distribution with mean $N\psi_i[\phi_i]$ and variance $N\psi_i[\phi_i](1 - \psi_i[\phi_i])$. This expected distribution is visualized in Figure 2 by the solid lines. Figure 2(a) represents the capacitive sensor; the lower panel represents the optical sensor. In both the upper and lower panels, the horizontal axes indicate the number of nonmatches (in 66 attempts), the vertical axes represent the population proportion. The numbers inbetween the upper and lower panels represent the FNMR corresponding to the number of false nonmatches in 66 attempts. The capacitive sensor (Figure 2(a)) has a maximum at one nonmatch out of 66 which corresponds to the FNMR of 0.0181 at an FMR of 0.001 that was also provided in Table 1. The optical sensor (Fihure 2(b)) has a maximum at zero nonmatches which is caused by the smaller overall FNMR of 0.0040. The far-most right point on the curves represents 9 or more nonmatches out of 66. For the capacitive sensor, the probability of finding 9 or more false nonmatches out of 66 according to the binomial distribution equals $3.01e - 6$; for the optical sensor this value equals to $7.7e - 12$ (not shown in the figure).

The observed FNMRs per subject based on an individual comparison threshold for each sensor to result in an overall FMR of 0.001 are given by the dashed lines in Figure 2. These curves represent the genuine comparisons for all subjects and fingers, that is, different fingers of one subject can be interpreted as *additional* subjects. All number of observations are normalized to sum to +1 to allow direct comparison with the binomial distribution given by the solid line. Interestingly, the curve for the observed number of false nonmatches is quite different from the binomial distributions, for both the capacitive and optical sensors. Two trends can be observed: (1) the number of subjects with zero false nonmatches is larger than expected based on a binomial distribution, and (2) the number of subjects with 9 or more false nonmatches is also significantly larger than expected. The proportion of subjects that obtained 9 or more false nonmatches (which corresponds to an FNMR of 0.136 or more) equals 0.0505 and 0.0145, for the capacitive and optical sensors, respectively. The proportion of subjects with 23 or more nonmatches (an FNMR of 0.33 or larger) amounted 0.0120 and 0.0006, for the capacitive and optical sensors. Hence, the observed frequencies of finding 23 or more nonmatches in a trial of 66 is 3 to 7 orders of magnitude larger than is expected based on a binomial distribution.

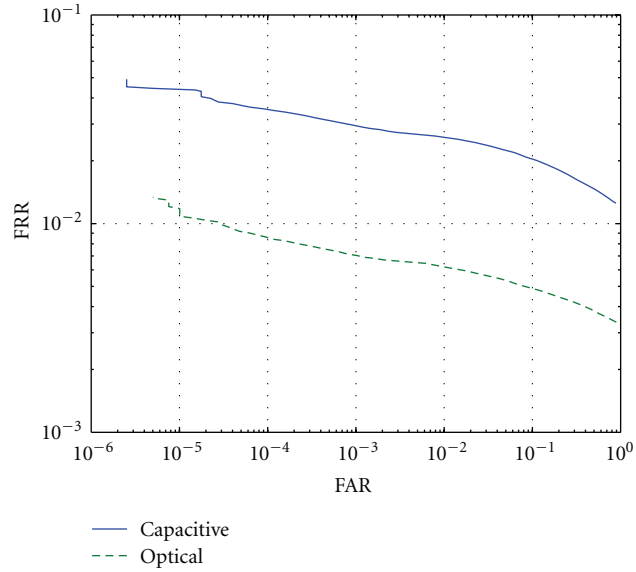


Figure 1: DET curves for the capacitive (solid line) and optical (dashed line) sensors based on the full database (including failures to enroll).

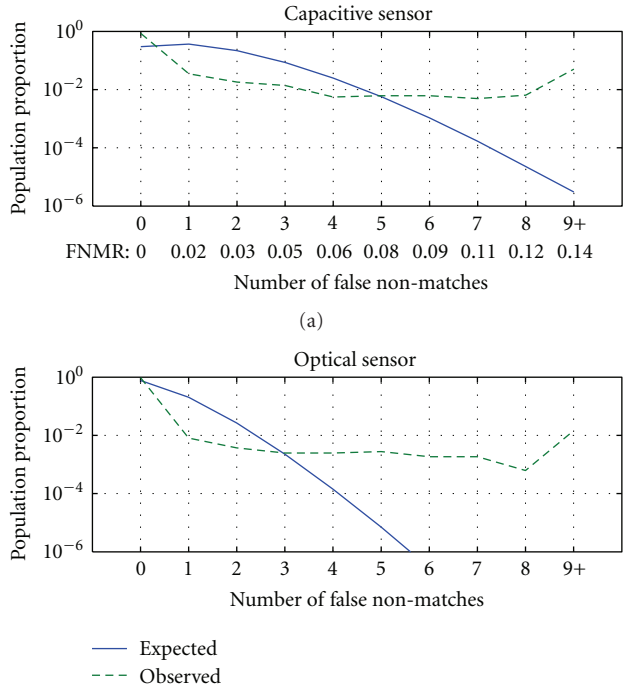


Figure 2: Distribution of the expected (solid lines) and observed (dashed lines) number of false nonmatches across subjects and fingers for the capacitive sensor (a) and the optical sensor (b). The numbers in between panels represent the corresponding false nonmatch rates.

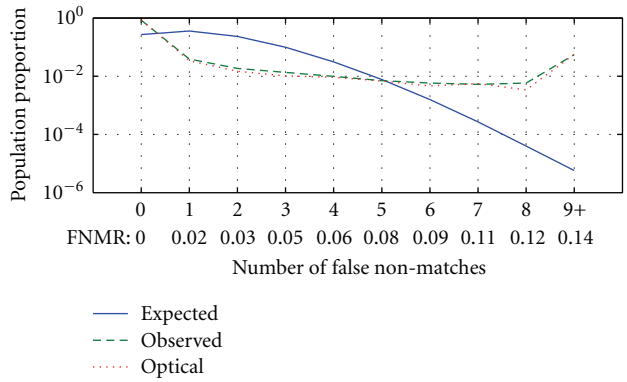


Figure 3: Distribution of expected (solid lines) and observed (dotted and dashed lines) number of false nonmatches across subjects for the capacitive sensor (dashed line) and the optical sensor (dotted line) based on a threshold to result in a mean FNMR of 0.02.

One possible reason for finding a relatively large population of subjects with a high FNMR is that these could be resulting from "weak" fingers that more often causes nonmatches. To investigate the distribution of interclass FNMRs when excluding the effect of different FNMRs per finger, a separate comparison threshold was estimated for each finger index and sensor such that across all subjects, the FNMR was equal to a fixed value of 0.02 when measured for one finger and sensor across all subjects. The distribution of false nonmatches in a set of 66 attempts is shown in Figure 3. The expected values based on the binomial distribution with mean probability of 0.02 are given by the solid line; the observed distributions for the capacitive and optical sensors are shown by the dashed and dotted lines, respectively. Interestingly, using a separate threshold for each sensor and finger to result in the same mean FNMR, the observed distributions of FNMRs across subjects are very similar. Furthermore, there is a significant discrepancy between the expected (binomial) distribution and the observed distribution. More than 5% of the population obtained 9 or more false nonmatches, which is significantly larger than the expected value of $5.8e-6$. Another interesting observation is that for both sensors, about 90% of the subjects has an FNMR which is *smaller* than the population average of 0.02, while only 10% has an FNMR which is (significantly) larger.

2.5. Multiple Verification Attempts. If multiple verification attempts are allowed in a verification system, the expected number of false matches will typically increase if the comparison threshold is kept constant (e.g., assuming that an imposter will use a different finger during each attempt to maximize the false match probability). If the false match probability of the nth trial out of N using sensor i is assumed to be constant across subjects and fingers and given by $\phi_i[n,N]$, the probability that at least one of N attempts will give a false match $\Phi_i[N]$ is given by

$$\Phi_i[N] = 1 - \prod_{n=1}^{N} (1 - \phi_i[n,N]). \tag{3}$$

If one also assumes that the probability $\phi_i[n,N]$ is independent of trial number n and $\phi_i[n,N] \ll 1$, this can be approximated quite accurately by

$$\Phi_i[N] \approx N\phi_i. \tag{4}$$

Said differently, the false match probability increases approximately linearly with the number of attempts if the comparison threshold is kept constant.

The number of false nonmatches will typically *decrease* with the increasing number of attempts. If the false nonmatch probability for attempt n out of N given by $\psi_i[\Phi_i[N],n,N]$, the probability that all N attempts will result in, a false nonmatch is given by

$$\Psi_i[\Phi_i[N],N] = \prod_{n=1}^{N} (\psi_i[\Phi_i[N],n,N]). \tag{5}$$

If one assumes the probabilities $\psi_i[\Phi_i[N],n,N]$ to be independent on trial n, this would result in

$$\Psi_i[\Phi_i[N],N] = \psi_i[\phi_i]^N. \tag{6}$$

Hence, an important consequence of the dependency of both FMR and FNMR on the number of attempts is that the comparison threshold should be dependent on the number of allowed attempts if a fixed FMR is desired.

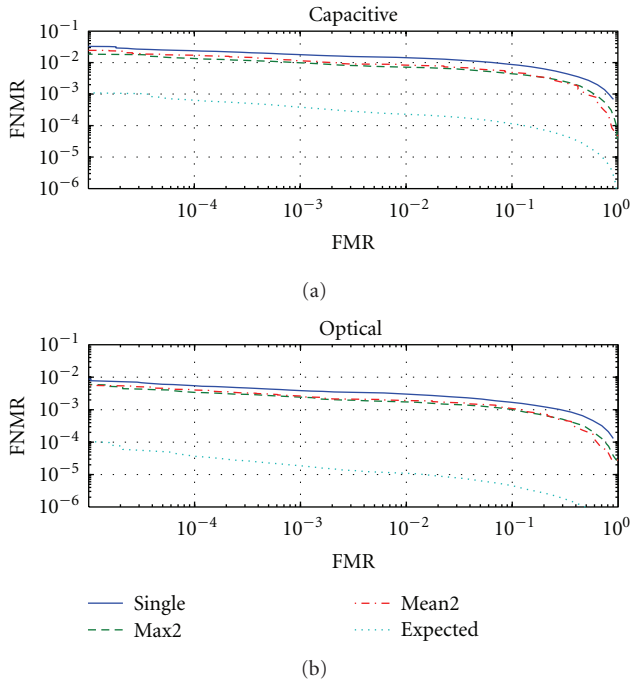


Figure 4: DET curve for the capacitive sensor (a) and the optical sensor (b) based on the balanced database. The solid line represents the FMR/FNMR tradeoff for a single attempt. The dashed lines represent the performance based on the maximum comparison score of 2 attempts (max2); the dash-dotted lines represent the performance for the mean comparison score across 2 attempts (mean2). The dotted curve represents the expected FMR/FNMR tradeoff assuming constant false nonmatch and false match probabilities for each trial.

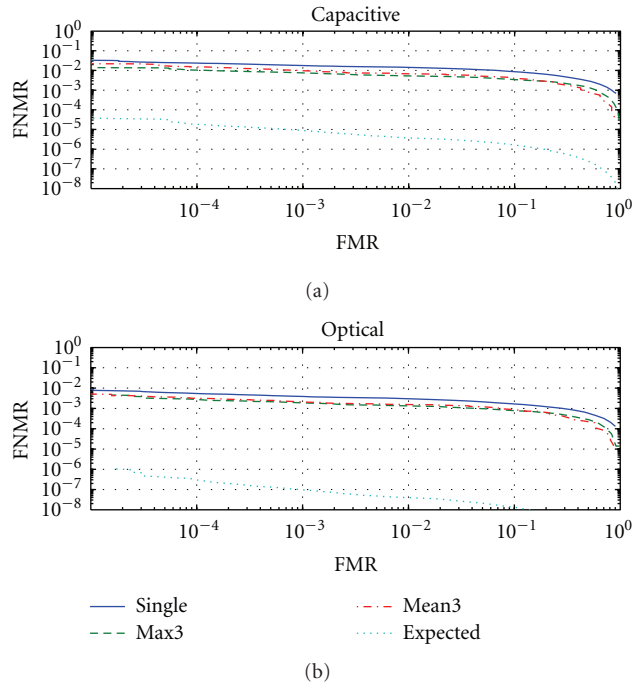


Figure 5: DET curve for the capacitive sensor (a) and the optical sensor (b) based on the balanced database. The solid line represents the FMR/FNMR tradeoff for a single attempt. The dashed lines represent the performance based on the maximum comparison score of 3 attempts (max3); the dash-dotted lines represent the performance for the mean comparison score across 3 attempts (mean3). The dotted curve represents the expected FMR/FNMR tradeoff assuming constant false nonmatch and false match probabilities for each trial.

To investigate the effect of multiple verification attempts, a two and three trial case was simulated by taking the maximum comparison similarity score across two or three genuine comparisons, respectively. The same process was employed for the imposter scores using fingerprints from different subjects across the attempts. The resulting DET curves are visualized in Figures 4 and 5 for the balanced database. Figures 4(a) and 5(a) represent the capacitive sensor, Figures 4(b) and 5(b) represent the optical sensor. The solid, dashed, and dash-dotted lines represent a single trial, the maximum comparison, and the mean comparison scores across multiple attempts, respectively.

As can be observed from Figures 4 and 5, the possibility of multiple verification attempts has a positive influence on the verification performance. For the capacitive sensor, the FNMR of 0.018 at an FMR of 0.001 for a single verification attempt decreases to 0.011 and 0.010 for two attempts according to a "mean" and "max" rule, respectively. For the three-trial case, the respective FNMRs are equal to 0.0096 and 0.0075. The optical sensor shows a similar trend. The FNMRs for a single trial at an FMR of 0.001 correspond to 0.0040. For two attempts, the FNMRs are equal to 0.0026 and 0.0024 (for the "mean" and "max" rules, resp.). For three attempts, these rates are equal to 0.0020 and 0.0018.

For both sensors, the "max" rule provides the lowest FNMR at a given FMR. The ratios of FNMRs at a fixed FMR of 0.001 for two attempts compared to one trial equal to 0.55 and 0.60 (for the capacitive and optical sensors, resp.). For three attempts, these ratios are equal to 0.42 and 0.45, respectively, when compared to the single-attempt case. However, these improvements are significantly smaller than the expected DET curve based on the independence assumption of FNMR and FMR rates for each trial, which is represented by the dotted curves in Figures 4 and 5. This curve was created by transforming the single-attempt curve to a multiple-attempt curve using (6) and (3).

2.6. Discussion. When attempting to enroll the 79,200 images, the failure to enroll rate amounted about 0.012 for the capacitive and 0.003 for the optical sensors, respectively. For the capacitive sensor, the value of 0.012 is quite in line with the assumption that between 1 and 3 % of a population has difficulties or failures to enroll. The value of 0.003 for the optical sensor is relatively low in this respect.

The DET curves based on the full database shown in Figure 1 indicate that the two sensors employed in the test differ considerably in terms of verification performance.

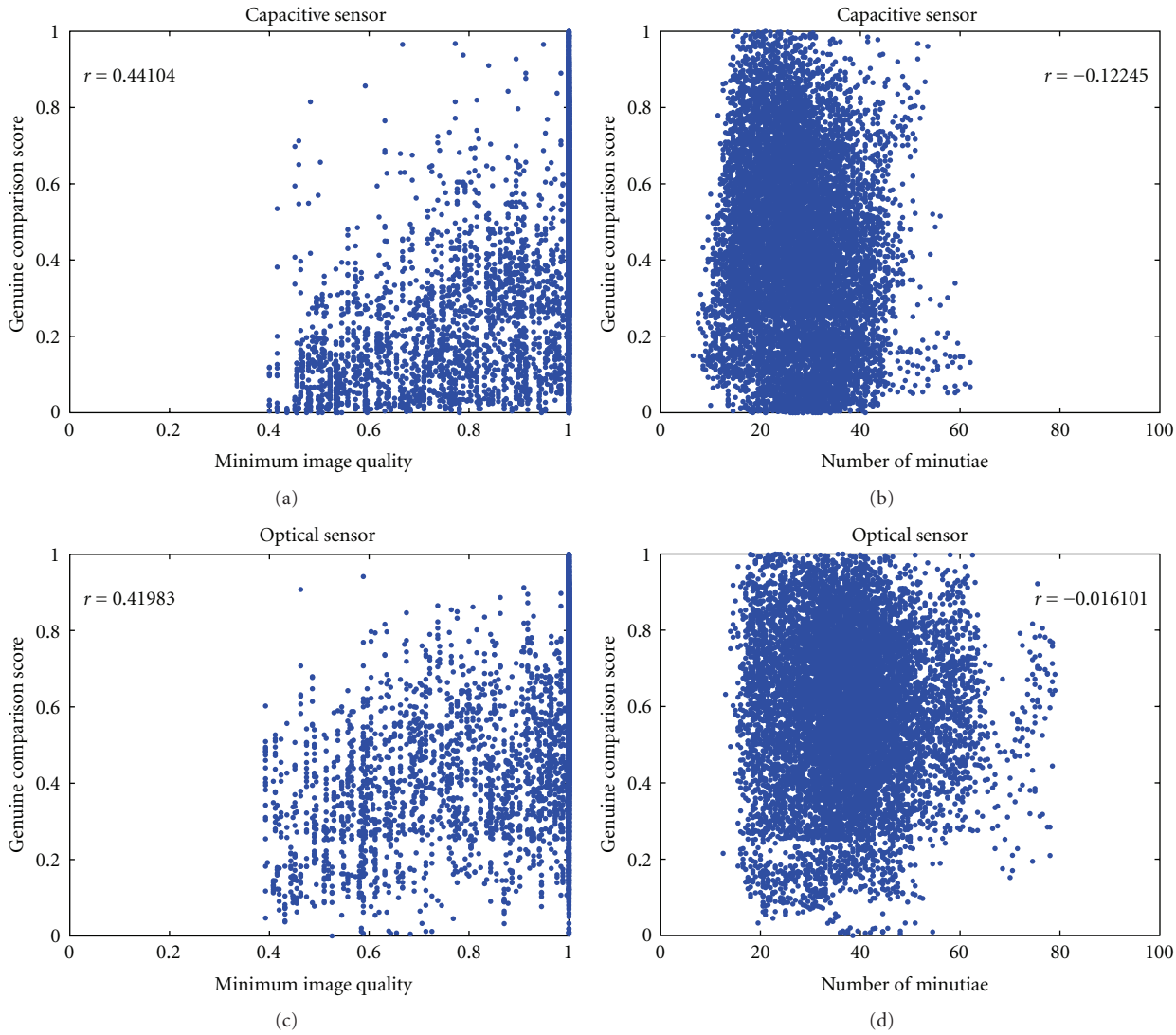


FIGURE 6: Genuine comparison scores as a function of the lowest image quality of the two images under test (a, c) and the number of detected minutiae (b, d). However, (a, b) represent the capacitive sensor; (c, d) represent the optical sensor.

Similar to the ratio of a factor of 4 in terms of failures to enroll, the capacitive sensor has an FRR which is also about 4 times larger than the optical sensor for the same FAR.

When images that caused a failure to enroll are not taken into account in the performance evaluation, the error rates improve by almost a factor of two for both sensors (see Table 1). This indicates that the number of failures to enroll, and the number of false nonmatches is about the same for the current database.

A further analysis on the balanced database revealed statistically significant differences in false nonmatch rates between subjects and fingers. When the thresholds for the capacitive and optical sensors were set to individually achieve an FMR of 0.001 between 1.45% (optical) and 5.05% (capacitive) of the subjects experienced an FNMR of 0.136 or larger. Moreover, when differences between sensors and fingers are accounted for by setting a separate threshold for each finger index and sensor to obtain an average FNMR across the population of 0.02, more than 5% of the population achieved an FNMR of at least 0.136, which is more than 6 times larger than the population mean, and 4 orders of magnitude larger than expected based on subject-independent false nonmatch probabilities. Last but not least, 90% of the population has an FNMR which is smaller than the population average. Said differently, it seems that for this corpus and threshold setting, only 10% of the population is responsible for the majority of the false nonmatches.

In an attempt to explain high false nonmatch rates for certain individuals, the image quality reported by the template extraction algorithm and the number of extracted minutiae were investigated. These experiments were performed on the balanced database. First, for each combination of sensor, subject, and finger, the FNMR (derived from all 66 comparisons) was correlated with the *average* image quality and *average* number of extracted minutiae across all 12 measurements. This correlation thus reflects the

relation between average properties across all observations of a certain subject and finger, and the average FNMR. No significant first-order relations were found. The resulting Pearson correlations between FNMR and image quality, and between FNMR and the number of minutiae were lower than 0.075 for both sensors.

In a second attempt, the individual comparison scores of all genuine template pairs were correlated with the minimum image quality of the two images under test. This test thus aims at discovering a relation between the comparison score and attributes of the individual images. A scatter plot of comparison score versus image quality for the capacitive sensor is shown in Figure 6(a); the scatter plot for the optical sensor is provided in Figure 6(c). Both the comparison scores and image quality data are normalized to an interval between zero and +1. The Pearson correlation coefficients (r) are provided in each panel. As can be observed, there is only a weak correlation between image quality and comparison score ($r = 0.44$ and 0.42, for the capacitive and optical sensors, resp.). Figures 6(b) and 6(d) demonstrate the relation between the number of detected minutiae (as mean value of the two templates under test) and the comparison scores. Given the very low Pearson correlation coefficients ($r = 0.12$ and 0.02), no relation seems to exist between the number of minutiae and genuine comparison score.

When multiple verification attempts are allowed, the number of false nonmatches reduces by a factor of about 1.7 to 1.8 for two attempts and about 2.2 to 2.4 for three attempts (provided that the FMR is kept constant). This increase in performance is roughly in line with results by others (cf. [4]) and is significantly smaller than what would be expected based on independent probabilities for false nonmatches and false matches for each attempt (cf. (3)–(6)). This observation suggests that the false nonmatch probability for a second or third attempt depends on the outcome of the earlier attempts. If we denote the conditional probability for a false nonmatch during the Nth attempt given false nonmatches in all $N-1$ previous attempts by $\psi[\Phi[N], N, N]$, we find the following relation between the overall false nonmatch probability for N and $N-1$ attempts:

$$\Psi[\Phi[N], N] = \psi[\Phi[N], N, N]\Psi[\Phi[N-1], N-1, N-1]. \tag{7}$$

If one assumes that the false match rates $\Phi[N]$ are set to a constant value Φ for every N, this results in

$$\psi[\Phi, N, N] = \frac{\Psi[\Phi, N]}{\Psi[\Phi, N-1, N-1]}. \tag{8}$$

In other words, the conditional probability for a false nonmatch at trial N given false nonmatches during all earlier attempts can be derived from the ratio of the DET curves for N and $N-1$ attempts. For the current database, in which the relative improvement equals to a factor of approximately 1.75, this means that the probability of a false nonmatch during the second trial equals approximately 0.57. Analogously, the conditional probability of a false nonmatch during the third trial, given a false nonmatch during the first and second trial, amounts to approximately 0.75.

It should be noted that these conditional probabilities describe the *average* probability for a second or third false nonmatch (i.e., provided that earlier attempts also resulted in a false nonmatch). This result may erroneously be interpreted as an FNMR that depends on the attempt number for a given subject. Most likely, the FNMR rate for a given subject is more or less constant across attempts. The increase in the conditional probability on a system level is presumably caused by an increase in the probability that the current subject is associated with a high (but constant) FNMR, and hence subsequent attempts will (most likely) also have a high probability of a false nonmatch and hence represents a "goat."

3. Conclusions

The MCYT fingerprint corpus under test, in combination with a state-of-the-art commercially-available fingerprint-matching algorithm, gives rise to subject-dependent false nonmatch rates if single enrollment and verification measurements are used. This result was observed for a capacitive as well as an optical sensor. From the distribution of false nonmatch rates across subjects, it seems that for a threshold setting resulting in an average false nonmatch rate of 0.02, a vast majority of 90% of the population has a probability for a false nonmatch that is smaller than the population average. The average false nonmatch rate seems to be dominated by a small group of subjects that are associated with a disproportionately large number of false nonmatches. When adjusting comparison thresholds as a function of sensor type and finger to result in an average FNMR of 0.02 across the population, at least 5% of all subjects experienced an FNMR of 0.136.

In an attempt to predict which images were associated with high false nonmatch rates, fingerprint image quality, the number of detected minutiae, and the genuine comparison scores were compared. Only a weak correlation (Pearson correlation around 0.4) was observed between image quality and comparison score, and no significant correlation was found between the number of minutiae and comparison score. This indicates that for the system and corpus under test, these measures cannot reliably indicate images associated with high false nonmatch rates.

The consistency in the false nonmatch probability for certain subjects was expressed as conditional false nonmatch rate. It was observed that for the system under test, the conditional probability of a false nonmatch given 2 earlier attempts amounts to approximately 0.75. Hence, for the system and fingerprint database under test, the number of verification attempts is best limited to two, and an alternative biometric modality or authentication method should be provided in case a subject experiences two subsequent false nonmatches.

Acknowledgment

The authors would like to thank the anonymous reviewers and the associate editor for their very helpful comments and suggestions to improve the manuscript.

References

[1] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, et al., "FRVT 2006 and ICE 2006: large-scale results," Tech. Rep. IR 7408, NIST National Institute of Standards and Technology, Gaithersburg, Md, USA, 2007.

[2] B. Toth and T. Mansfield, "Latest biometric test results—performance, quality and interoperability," Tech. Rep., Deloitte, 2006.

[3] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "IBG comparative biometric testing—round6," Tech. Rep., International Biometric Group, Middlesex, UK, 2006.

[4] T. Mansfield, G. Kelly, D. Chandler, and J. Kane, "Biometric product testing—final report," Tech. Rep., Centre for Mathematics and Scientific Computing, National Physics Laboratory, Middlesex, UK, 2001.

[5] P. J. Phillips, P. J. Flynn, T. Scruggs, et al., "Overview of the face recognition grand challenge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 947–954, 2005.

[6] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 3–17, 2006.

[7] Biometric System Laboratory University of Bologna, "FVC2006: the fourth international fingerprint verification competition," 2006, http://bias.csr.unibo.it/fvc2006/default.asp.

[8] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP '98)*, Sydney, Australia, 1998.

[9] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: new additions to the biometrie menagerie," in *Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies (AUTO ID '07)*, pp. 1–6, Alghero, Italy, 2007.

[10] R. M. Bolle, S. Pankanti, and N. K. Ratha, "Evaluation techniques for biometrics-based authentication systems (FRR)," in *Proceedings of the International Conference on Pattern Recognition (ICPR '00)*, pp. 2831–2837, 2000.

[11] M. Wittman, P. Davis, and P. J. Flynn, "Empirical studies of the existence of the biometric menagerie in the FRGC 2.0 color image corpus," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '06)*, p. 33, 2006.

[12] A. Hicklin, C. Watson, and B. Ulery, "The myth of goats: how many people have fingerprints that are hard to match?" Tech. Rep. IR 7271, NIST National Institute of Standards and Technology, Gaithersburg, Md, USA, 2005.

[13] S. K. Modi and S. J. Elliott, "Impact of image quality on performance: comparison of young and elderly fingerprints," in *Proceedings of the 6th International Conference on Recent Advances in Soft Computing (RASC '06)*, K. Sirlantzis, Ed., pp. 449–454, 2006.

[14] European Parliament and European Council, "Directive 1995/46/EC of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," 1995.

[15] ORC, "Public attitudes toward the uses of biometric identification technologies by government and the private sector. Summary of survey findings, opinion research corporation ORC," 2001.

[16] TNS/TRUSTe, "Consumer attitudes about biometrics in ID documents," Tech. Rep., TNS/TRUSTe, August 2005.

[17] Logica CMG, "e-identity—european attitudes towards biometrics," Whitepaper, Logica CMG, 2006.

[18] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, et al., "MCYT baseline corpus: a bimodal biometric database," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[19] F. Alonso-Fernandez, F. Roli, G. L. Marcialis, J. Fierrez, and J. Ortega-Garcia, "Comparison of fingerprint quality measures using an optical and a capacitive sensor," in *Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS '07)*, pp. 1–6, Crystal City, Va, USA, September 2007.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 845893, 12 pages
doi:10.1155/2009/845893

Research Article

Integrating Fingerprint Verification into the Smart Card-Based Healthcare Information System

Daesung Moon,[1] Yongwha Chung,[2] Sung Bum Pan,[3] and Jin-Won Park[4]

[1] *Biometrics Technology Research Team, ETRI, Daejeon 305-700, South Korea*
[2] *Department of Computer and Information Science, Korea University, Jochiwon, Chungnam 339-700, South Korea*
[3] *Department of Information Control and Instrumentation Engineering, Chosun University, Gwangju 501-759, South Korea*
[4] *School of Games, Hongik University, Jochiwon, ChungNam 339-701, South Korea*

Correspondence should be addressed to Yongwha Chung, ychungy@korea.ac.kr

Received 10 October 2008; Revised 13 May 2009; Accepted 14 September 2009

Recommended by Stephanie Schuckers

As VLSI technology has been improved, a smart card employing 32-bit processors has been released, and more personal information such as medical, financial data can be stored in the card. Thus, it becomes important to protect personal information stored in the card. Verification of the card holder's identity using a fingerprint has advantages over the present practices of Personal Identification Numbers (PINs) and passwords. However, the computational workload of fingerprint verification is much heavier than that of the typical PIN-based solution. In this paper, we consider three strategies to implement fingerprint verification in a smart card environment and how to distribute the modules of fingerprint verification between the smart card and the card reader. We first evaluate the number of instructions of each step of a typical fingerprint verification algorithm, and estimate the execution time of several cryptographic algorithms to guarantee the security/privacy of the fingerprint data transmitted in the smart card with the client-server environment. Based on the evaluation results, we analyze each scenario with respect to the security level and the real-time execution requirements in order to implement fingerprint verification in the smart card with the client-server environment.

1. Introduction

A *smart card* is a credit-card-sized plastic card with an embedded chip that can be memory or may include a microprocessor [1]. A microprocessor chip in a smart card can add, delete, or manipulate information in its memory and hence offer complex data security functionalities. These smart cards have been used in many applications such as health cards, e-passport, and e-ID cards for more than 10 years [2–4]. In this paper, we describe an example where a smart card is applied in the healthcare information system. One of the most popular smart card solutions is Gemalto Cryptoflex JavaCard [5], equipped with a 16-bit microcontroller (Infineon SLE66CX322P, compatible with standard SAB 8051 processor) and an additional crypto processor for RSA and DES computations. The card has ROM, EEPROM, and RAM. A Java Applet implements the Chip Authentication Program.

The usual way of obtaining relevant patient data is connecting to the hospital database. However, we may experience the situation where no network connection facility (e.g., ambulance) is available. Also, hospitals cannot open the patient data stored in the hospital database via Internet because of the security/privacy of the patient and/or different network environments. These problems can be solved by using portable storage devices such as smart cards. That is, a doctor may take patient information from the smart card at the time of consultation.

Because of the advantage of the smart card, using the smart card in healthcare becomes popular in such countries as France, Germany, and Taiwan. Also, several solutions are already implemented [5–10]. For example, Taiwan was the first country introducing the nationwide smart card-based healthcare information system in 2003. Over 22 million patient health cards have been issued, as well as over 345 000 health professional cards giving the doctor access to medical

information. In EU, several projects about smart card-based healthcare service have been performed such as Sesam Vitale in France, DENTcard in Germany [7], Transcards [8], NETLINK [9], and NETC@RDS project [10].

Deployed as of 1998, SESAM-Vitale currently links more than 223 000 healthcare professionals with the Health Insurance System, for the benefit of millions of insured persons who have the Vitale card. The NETC@RDS initiative is devoted to establish new improved healthcare administration services for mobile citizens across EU. The actual phase aims at deploying e-health services via the European Health Insurance Card through a wide trans-European network simplifying healthcare access procedures. NETC@RDS successfully tested the electronic version of the European Health Insurance Card (EHIC) during its first and second Market Validation phases in 85 pilot sites of the 10 EU member states. The Initial Deployment phase of the project—launched on June 1, 2007—will deploy operational services in all targeted sites, to total of 305 service sites serving 566 service points across the 15 participating countries.

NETLINK is a project of the fourth framework programme on research and development of European commission HC 4016. Countries involved in the NETLINK consortium (France, Germany, Italy, and the Province of Québec) have set up or are in the process of setting up new nationwide information systems in the healthcare sector based on the use of modern technologies: smart cards (used by health professionals and patients), computers (used by health professionals, hospitals, health Insurance funds), large networks, trusted third parties (for security purpose).

In the security and the privacy terms, several issues should be considered in developing the smart card-based healthcare information system; the ways to certify different devices (card, card reader, terminal), the methods to authenticate users (health professionals and patients), and the amount of information about a patient to be stored in the smart card, for example, patient personal information and emergency contact information. There are many other issues to consider and the details of other issues can be found in [11, 12]. In this paper, we consider the methods to authenticate the cardholder, especially patients in a large-scale hospital database.

In general, the visible information on the health card may contain the cardholder's name, identification number, birth date, photo, and the card serial number. The contents inside the card can be divided into four segments: basic data, health insurance data, medical data, and public health administration data [13, 14]. The *basic data* segment stores the identification information for both the cardholder and the card itself. The contact person information in case of emergency and the cryptographic key for security can also be stored. The *health insurance data* segment is comprised of the cardholder's insurance information and service data for insurance claims. The *medical data* segment contains the information for important physician orders, prescriptions, and drug allergies. With the advance in the smart card technology, more medical information can be stored in the card. Finally, the *public health administration data* segment is used for recording personal data pertaining to public health such as vaccination records and organ donation notes. Additionally, a separate smart card can be used as an identification card for a medical staff.

As more information is stored in the smart card, it becomes important to protect the information it contains. However, the current card holder verification method in the healthcare services is based on the password/PIN such as SESAME-VITALE. Most of people set their passwords/PINs based on words or numbers that they can easily remember. Thus, the passwords are easy to crack by guessing or a simple brute force dictionary attack. Although it is possible and even advisable to keep different passwords for different applications, most of people use the same password across different applications. If a single password is compromised, it may open many doors. "Long and random" passwords are more secure but harder to remember, which prompts some users to write them down in accessible locations. Such passwords also result in more system help desk calls for being forgotten or expired. Cryptographic techniques such as PKI [15, 16] can provide very long passwords that need not to be remembered but are in turn protected by simple passwords, thus defeating their original purpose.

In recent years, there is an increasing trend of using *biometrics*, which refers to the personal biological or behavioral characteristics used for verification or identification [17, 18]. It relies on "something that you are," and can inherently differentiate between a verified person and a fraudulent imposter. The problem of resolving the identity of a person can be categorized into two distinct types, verification and identification. *Verification* matches a person's claimed identity to his/her previously enrolled pattern (i.e., "one-to-one" comparison). However, identification identifies a person from the entire enrolled population by searching a database for match (i.e., "one-to-many" comparison). In this paper, we focus on *fingerprint* because it is relatively mature and its scanning unit is cheaper than other biometrics such as iris [18]. Also, we will focus on verification only because we assume each patient or medical staff holds his/her smart card. (Note that, however, fingerprints can also be applied to healthcare services without the smart card.)

Since fingerprints cannot be lost or forgotten like passwords, fingerprints have the potential to offer higher security and more convenience for user authentication [18]. For example, fingerprints are significantly more difficult to copy, share, and distribute than passwords. That is, the main advantage of a fingerprint recognition solution is the convenience while maintaining sufficiently high security. In general, security with large data is regarded as higher than that with small data although this does depend greatly on the way it is implemented. Fingerprint data size is typically 70 KB for images and 500 B for features, much larger than that of the password with 10B. Furthermore, large fingerprint data need not to be memorized. Especially in healthcare services, fingerprints have additional advantage over the password. The emergency data set such as his/her blood type and contact person information can be accessed from the smart card by using his/her finger in the emergency medical situation when the patient is unconscious.

However, fingerprint-based recognition has some disadvantages as well [19, 20]. A compromised password can be canceled and a new password can be issued as often as desired, whereas people have only 10 fingers. If a fingerprint is compromised repeatedly, it cannot be replaced eventually [18]. Finally, in principle, a fingerprint template stolen from one application may be used in another application. These issues are especially important in pervasive computing where the fingerprint data must be carefully protected because of privacy concerns. Also, we need more computation to authenticate users with the fingerprint data than the password.

In large-scale healthcare services, the computational overhead caused by the deployment of the fingerprint as well as the protection of the fingerprint data should be considered. However, only limited research has been carried out in this direction [21, 22]. In this paper, we consider the possible scenarios to integrate fingerprints into smart cards and evaluate each scenario in terms of security, privacy, and computational overhead (i.e., cost of the smart card). Also, for the cheapest scenario (i.e., the fingerprint data is transmitted to a remote server for verification), the collective performance of fingerprint verification and the authentication protocol on the client-server model is analyzed.

The rest of the paper is structured as follows. Section 2 explains the overview of typical fingerprint verification and three strategies for integrating fingerprints into the smart card. Sections 3 and 4 describe the fingerprint verification scenarios for the fingerprint smart card and the client-server environments, respectively. The results of the performance evaluation are described in Section 5, and the conclusions are given in Section 6.

2. Background

2.1. Biometric-Based User Authentication for Healthcare. The healthcare industry is confronted with solving the legal requirement to protect medical information of patients. Medical information is available in the computer network, thus could be used illegally. Laws call for the protection of patient privacy and also for standardization of medical data. Biometrics are expected to be chosen as possible means for user authentication for healthcare, since biometrics provide a secure method for patient identification and extracting personal data for treatment. In a large hospital, there may occur the following three types of healthcare services that are related to biometric information.

Access to Personal Medical Information. The largest part of biometric usage in healthcare industry may be enhancing individuals' access to their personal information. It seems likely that patients will demand access to their information, while demanding that such information be kept secure. Medical information may be stored on smart cards or on networks; in either case, the biometric is a gateway to personal information.

Emergency Patient Identification. In emergency medical situations, correct and immediate patient authentication is critical. For individuals without identification and unable to communicate, biometric information provides a unique form of authentication. Developing this type of biometric identification system includes enrolling a sufficient number of users to achieve critical mass and the availability of biometric devices in emergency situations and locations. In this paper, we assume each patient holds his/her smart card and the emergency dataset such as his/her blood type and contact person information can be accessed from the smart card by using his/her finger in emergency medical situations.

2.2. Fingerprint Verification. The *fingerprint* is chosen for verification and for identification in this paper. It is more mature in terms of the algorithm availability and feasibility. Fingerprint verification and identification algorithms can be classified into two categories: *image-based* and *minutiae-based* [17, 18].

A minutiae-based fingerprint verification system has two phases: *enrollment* and *verification*. In the off-line enrollment phase, an enrolling fingerprint image for each user is processed, and the features called *minutiae* are extracted and stored in a server. In the online verification phase, the minutiae extracted from an input image is compared to the stored template, and the result of the comparison is returned.

In general, there are six logical modules involved in the fingerprint verification system [18]: *Fingerprint Acquisition module*, *Feature Extraction module*, *Matching module*, *Storage module*, *Decision module*, and *Transmission module*.

The *Fingerprint Acquisition module* contains an input device or a sensor that captures the fingerprint information from the user. It first refines the fingerprint image against the image distortion obtained from the fingerprint sensor. A typical process consists of three stages. The *binary conversion* stage applies a lowpass filter to smooth the high frequency regions of the image and threshold to each subsegment of the image. The *thinning* operation generates a one-pixel-width skeleton image by considering each pixel with its neighbors. In the *positioning* operation, the skeleton obtained is transformed and/or is rotated such that valid minutiae information can be extracted.

The *Feature Extraction module* refers to the extraction of features in the fingerprint image. After this step, some of the minutiae are detected and stored into a pattern file, which includes the position, the orientation, and the type (ridge ending or bifurcation) of the minutiae.

Based on the minutiae, the input fingerprint is compared with the enrolled fingerprint retrieved from the *Storage module*. Actually, the *Matching module* is composed of the *alignment* operation and the *matching* operation. In order to match two fingerprints captured with unknown direction and position, the differences of direction and position between two fingerprints are detected, and the alignment between them needs to be accomplished. Therefore, in this alignment operation, transformations such as translation and rotation between two fingerprints are estimated, and two minutiae are aligned according to the estimated parameters.

If the alignment is performed accurately, the following matching operation can be regarded as a simple point pattern matching. In the matching operation, two minutiae are compared based on their position, orientation, and type. Then, a matching score is computed. The *Decision module* receives the score from the matching module and, using a confidence value based on the security risks and the risk policy, interprets the result of the score, thus reaching a verification decision. The *Transmission module* provides the system with the ability to exchange information between all other modules.

2.3. Integrating Fingerprint into the Smart Card. Fingerprint technologies have been proposed to strengthen the verification mechanisms in general by matching a stored fingerprint template to a live fingerprint features [17, 18]. In the case of verification using a smart card, intuition suggests that the match should be performed by the smart card. However, this is not always possible because of the complexity of the fingerprint information, and because of the limited computational resources available to current smart cards. In general, three strategies of fingerprint verification can be identified as follows [23–26].

Store-on-Card. The fingerprint template is stored on a smart card. It must be retrieved and transmitted to a card reader that matches it to the live template acquired from the user by the fingerprint sensor. Cheap memory cards with no or small operating systems are generally sufficient for this purpose.

Match-on-Card. The fingerprint template is stored on a smart card, which also performs the matching with the live template. Therefore, a microprocessor on the smart card is necessary. The smart card must contain an operating system running a suitable match application. It is not possible to steal information stored in the card since a successful match enables the use of the certificates on the card without the need of stored PINs or passwords. Even in the unlikely event that a card is tampered with; only limited damage is caused since only that specific user's credentials are hacked. An attack on multiple users means that the attacker must get hold of all users' cards. In this strategy, the templates are never exposed to a nontamper proof environment and the user carries his/her own templates.

System-on-Card. This is a combination of the two previous strategies. The fingerprint template is stored on a smart card, which also performs the matching with the live template, and includes the fingerprint sensor to acquire, select, and process the live template. This strategy is the best in terms of the security as everything takes place on the smart card. Embedding a fingerprint acquisition on a smart card provides all the privacy and security solutions but, unfortunately, it is expensive and presents more than one realization problem.

The benefits derived from the Match-on-Card are valuable in themselves: using its own processing capabilities, the smart card decides if the live template matches the stored template closely enough to grant the access to its private data. Nevertheless this scheme presents a danger: we have no certainty that a fingerprint acquisition has been collected through live-scan and there is the risk of an attacker's sniffing the fingerprint data and later using it to unlock the card in a replay attack.

3. Fingerprint Verification Scenarios for the Smart Card-Reader Model

First, simplifying the scenarios between the smart card and the card reader considered, we assume that the symmetric and/or asymmetric keys are distributed to the smart card and the card reader when the system is installed and no further key exchange is required.

We explained three strategies for integrating the fingerprints into the smart card in the previous section. In this section, we consider five different scenarios [20–26, 28, 29]. As shown in Figure 1, SCENARIO 1 and SCENARIO 2 are Store-on-Card strategies because the smart card stores the fingerprint template. Also, as shown in Figure 2, SCENARIO 3 and SCENARIO 4 are Match-on-Card strategies because the matching module takes place on the smart card. Finally, SCENARIO 5 is the System-on-Card strategy (Figure 3). Within the Store-on-Card and the Match-on-Card scenarios, we can differentiate those in which the fingerprint sensor is built into the smart card (SCENARIO 2 and SCENARIO 4) and those where it is in the card reader (SCENARIO 1 and SCENARIO 3).

Store-on-Card: SCENARIO 1 and SCENARIO 2. In SCENARIO 1, the fingerprint sensor is built into the card reader. The user template is transferred from the card to the reader. The reader takes the fingerprint image provided by its built-in fingerprint sensor, performs the feature extraction, and also matches the features to the template provided by the card. The reader then informs the card whether verification has been successful or not.

On the other hand, the fingerprint sensor in SCENARIO 2 is built into the card. The fingerprint image and the user template are transferred from the card to the reader. The reader performs feature extraction and matches the features to the template. The reader then informs the card whether verification has been successful or not.

Match-on-Card: SCENARIO 3 and SCENARIO 4. In SCENARIO 3, the fingerprint sensor is built into the card reader. The reader takes the image provided by the built-in fingerprint sensor and performs feature extraction. The extracted features are sent to the card, which then performs the matching module and reaches the verification decision module.

The fingerprint sensor in SCENARIO 4 is built into the card. The fingerprint image is transferred from the card to the reader. The reader performs the feature extraction module only, and transfers the extracted features back to the card. The card then performs the matching module.

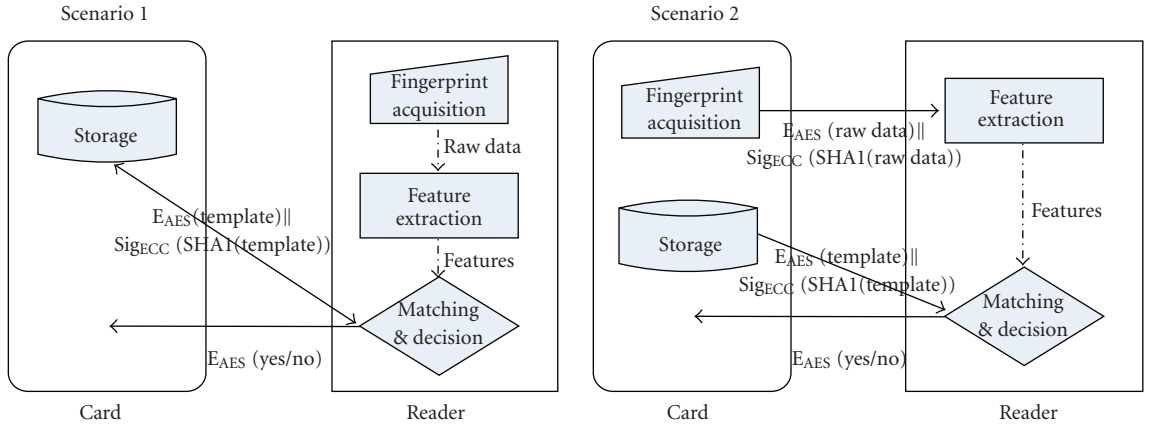


Figure 1: Illustration of the integrating scenarios [23–26] for store-on-card and the corresponding X9.84 [27] implementations.

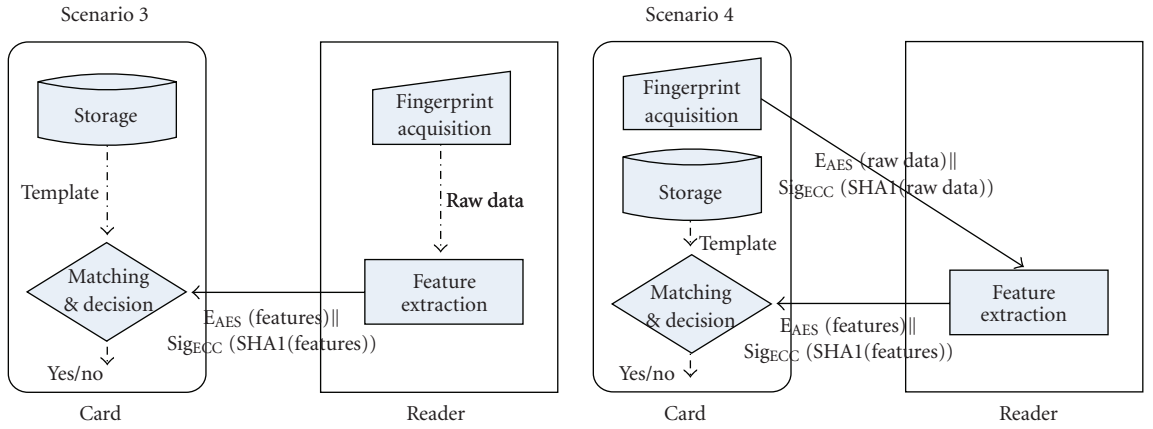


Figure 2: Illustration of the integrating scenarios [23–26] for match-on-card and the corresponding X9.84 [27] implementations.

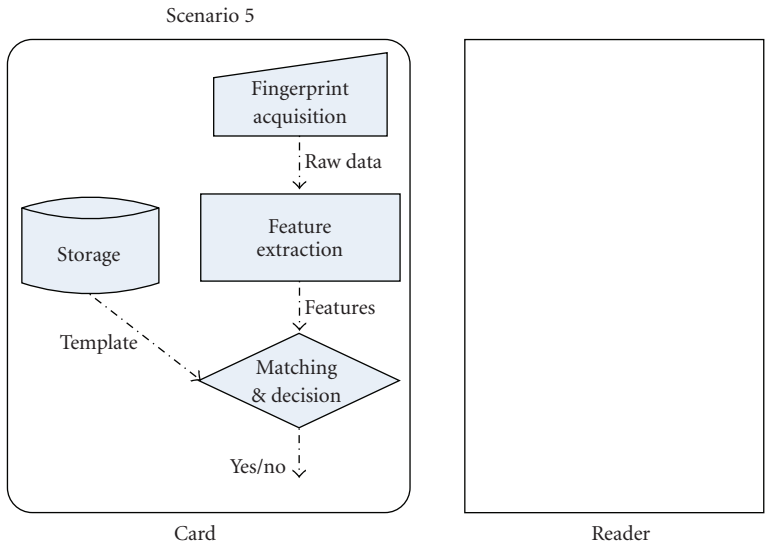


Figure 3: Illustration of the integrating scenarios [23–26] for system-on-card and the corresponding X9.84 [27] implementations.

System-on-Card: SCENARIO 5. SCENARIO 5 is System-on-Card: that is, all fingerprint verification modules take place on the card.

4. Fingerprint Verification Scenarios for the Large-Scale Client-Server Model

As we explained in Section 1, we also consider the client-server model for remote user authentication using fingerprint. Especially, we consider the healthcare information system using the cheapest fingerprint-based smart cards, that is, SCENARIO 1. In spite of guaranteeing higher security level than SCENARIO 1, other scenarios may not be right choices for large-scale applications such as national healthcare services or large hospitals having millions of patients due to the high implementation cost.

The client-server model for remote user authentication using the SCENARIO 1-based fingerprint card must guarantee the security/privacy as well as the real-time execution requirements. To satisfy those requirements, we first consider possible scenarios for remote fingerprint verification in terms of assigning the tasks of the fingerprint verification to each entity (i.e., client and server). Then, we evaluate the performance of each scenario.

Note that, to provide higher security level in the remote healthcare service, we assume the *three-way verification* method among the smart card fingerprint data, the live fingerprint data and the fingerprint data stored in the central DB. Also, we denote the three possible fingerprint verification scenarios for the client-server model as SCENARIO 1_1, SCENARIO 1_2, and SCENARIO 1_3, respectively.

Following assumptions are made to simplify the explanation.

(1) Between the client and the server, the same master key is shared when the system is installed.

(2) Entity authentication is completed using proper methods such as trusted Certificate Authority (CA).

(3) The user authentication service is assumed to be requested by the client at which the board control investigator is working.

(4) The execution time to perform some cryptography mechanisms to protect the fingerprint features stored in the smart card and in the server's database is not considered because this research focuses only on the protection of the fingerprint data transmitted.

In fact, these assumptions are reasonable, since the master key sharing operation (described in Assumption 1) needs to be executed only once, and the time for protecting the stored fingerprint data (described in Assumption 4) is negligible. For the purpose of explanation, we define first the following notations:

N: a nonce generated randomly in the client and used as a "challenge;"

K_m: a master key shared by both the sensor and the client;

$f(N)$: a simple function to generate a "response" for N;

K_s: a shared session key generated for each transmission;

C_Fe: a fingerprint feature stored in the biometric health card;

L_Fe: a fingerprint feature extracted from the live fingerprint image;

S_Fe: a fingerprint feature stored in the DB;

L_Fi: a live fingerprint image;

Mat_CL: a matching result between *C_Fe* and *L_Fe*;

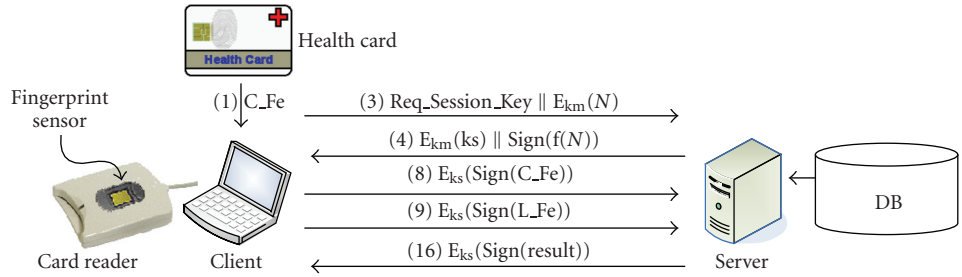
Mat_CS: a matching result between *C_Fe* and *S_Fe.*

SCENARIO 1_1, *Store-on-Card/the Server Does Everything.* In SCENARIO 1_1 as shown in Figure 4, the sensor attached to the client captures a live fingerprint image, and the client extracts some features from the image. Then, the client sends *L_Fe* and *C_Fe* to the server after applying the encryption and digital signature with the same key received from the server.

After verifying the signature for *L_Fe* and *C_Fe* and decrypting these fingerprint features, the server performs two comparisons with *C_Fe* – *S_Fe* and *C_Fe* – *L_Fe*. After checking the two matching results, the server returns a final result to the client. Note that this is a typical scenario of assigning the fingerprint verification tasks to the client-server model and requires five sets of communications for data transmission. This scenario can improve the security level of the fingerprint authentication system because the server can be more secure than the client. A server should be protected by the security experts, while a client maintained by an individual user may be more vulnerable to several attacks such as Trojan Horse [12–14]. On the other hand, the computational workload of the server in this scenario increases as the number of clients increases.

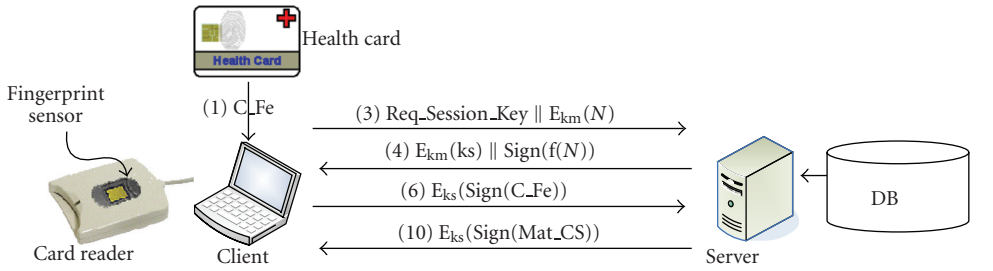
SCENARIO 1_2, *Store-on-Card/Extraction by the Client and Matching by the Server.* Unlike SCENARIO 1_1, in SCENARIO 1_2 as shown in Figure 5, the comparison with *C_Fe* – *L_Fe* and *C_Fe* – *S_Fe* is executed in the client and the server, respectively. In this scenario, the client sends only *C_Fe* to the server and calculates the matching score between *Mat_CS* received from the server and *Mat_CL* resulted in the client.

SCENARIO 1_3, *Store-on-Card/The Client Does Everything.* In SCENARIO 1_3 as shown in Figure 6, all the tasks except fingerprint acquisition are executed in the client. After encrypting the fingerprint features of the requested user stored in the server's database, the server only transmits it to the client. Thus, this scenario can reduce the workload of the server significantly by distributing the fingerprint authentication tasks into the clients. However, the security level of the fingerprint authentication system can be degraded because the client, which is more vulnerable to several attacks than the server, executes most of the tasks and the system depends on the security of keys stored in the client.



(2) Input live fingerprint of patient
(5) Encrypt and generate signature about C_Fe
(6) Feature extraction about L_Fi
(7) Encrypt and generate signature about L_Fe
(10) Decrypt and verify signature about C_Fe
(11) Decrypt and verify signature about L_Fe
(12) Matching between L_Fe and C_Fe
(13) Matching between S_Fe and C_Fe
(14) Fusion between Mat_CS and Mat_CL
(15) Encrypt and generate signature about result

Figure 4: Illustration of SCENARIO 1_1.



(2) Input live fingerprint of patient
(5) Encrypt and generate signature about C_Fe
(11) Feature extraction about L_Fi
(12) Matching between C_Fe and L_Fe
(13) Decrypt and verify signature about Mat_CS
(14) Fusion between Mat_CS and Mat_CL
(7) Decrypt and verify signature about C_Fe
(8) Matching between C_Fe and S_Fe
(9) Encrypt and generate signature about Mat_CS

Figure 5: Illustration of SCENARIO 1_2.

Table 1: Number of instructions required for fingerprint verification [25].

	Number of Instructions
Feature Extraction	451 739 250
Feature Matching	21 164 578

Table 2: Number of instructions and estimated time for fingerprint verification.

Step	Estimated time on ARM7TDMI	Estimated time on 8051
Feature Extraction	7.5 seconds	195 seconds
Feature Matching	0.3 seconds	7.8 seconds

5. Performance Evaluation

5.1. Evaluation of Fingerprint Verification and the Cryptographic Modules. A fingerprint-based smart card system for user verification must guarantee the user's privacy and provide sufficient authentication for access to patient data as well as the real-time execution requirements. To satisfy the requirements, we first evaluate the logical modules involved in the fingerprint verification system and the cryptographic modules for guaranteeing the integrity and confidentiality of the sensitive fingerprint data transmitted between the smart card and the card reader (see Section 3).

For secure transmission of the fingerprint data, we consider ANSI X.9.84, which is the security standard for fingerprint systems. The ANSI X.9.84 Fingerprint Information Management and Security standard [22] covers the requirements for managing and securing biometric data (fingerprint, iris, voiceprint, etc.) for customer identification and employee verification, mainly in the financial industry. In addition, this standard identifies the digital signature and encryption to provide both integrity and privacy of the fingerprint data. Specifically, 128-bit Advanced Encryption Standard (AES) and Elliptic Curve Digital Signature Algorithm (EDCSA) [8, 9] are considered as our symmetric encryption algorithm and digital signature and hash algorithm, respectively (see Figures 1–3). ECDSA is the elliptic curve analogue of the Digital Signature Algorithm (DSA). It is the most widely standardized elliptic curve-based signature scheme, appearing in ANSI X9.62, FIPS 186-2, IEEE 1363-2000, and the ISO/IEC 15946-2 standards as well as several draft standards. Because the most time consuming operations of ECDSA are ECC and the hash operation, we

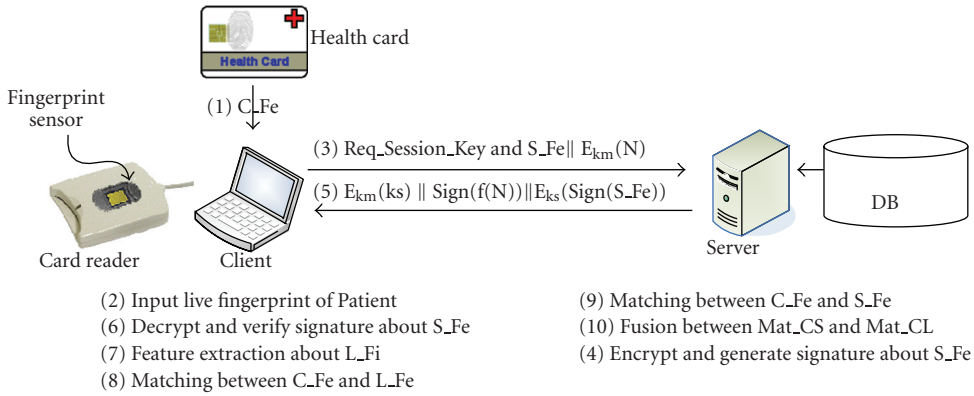


Figure 6: Illustration of SCENARIO 1_3.

Table 3: Number of Instructions and Cycles Required for AES.

		AES (128-bit)			
		Encryption		Decryption	
		No. of instructions	No. of cycles	No. of instructions	No. of cycles
ARM7TDMI	140 KB	292 889 168	485 763 071	406 011 432	690 034 743
	1 KB	2 131 620	3 535 199	2 952 268	5 017 575

confine our evaluation to them. Here, SHA1 is used as the hash algorithm.

Table 1 shows the number of instructions of each task in fingerprint verification measured on an instruction simulator, SimpleScalar [30]. Based on Table 1, we can compute the estimated execution time of each task on each processor. Finally, the time to acquire a fingerprint image through the fingerprint sensor is assumed to be about 1 second. Note that the feature extraction module requires a lot of integer operations for image processing, and the computational workload of this module occupies 96% of the total workload of fingerprint verification.

In order to show the performance requirement of the in-card processor, the number of instructions and the estimated execution time on the 8-bit Intel-8051 and 32-bit ARM7-based smart cards are summarized in Table 2. According to Table 2, it is impossible to assign the feature extraction or the matching step as well as the preprocessing to the 8051 chip. This is because the computation using the fingerprint data requires a large amount of memory and time. Thus, we adopt ARM7 to realize the Match-on-Card [25], which shows an improved result. Actually, the 32-bit smart card is somewhat expensive to be applied for the ordinary system. Nevertheless, it can be a good solution for the system that should guarantee very high level of security such as in E-Health, E-Business, and E-Government.

Because of the limited processing power of the in-card processor, all of the three steps above cannot be assigned to the in-card processor. Instead, we consider assigning only the third step, matching, to the in-card processor. This is because the first two steps involve rigid image processing computation, which is too exhaustive to be executed in the in-card processor. These computation steps can be easily carried out in real-time by a fingerprint capture device or a card reader equipped with at least a 500 MIPS processor. Therefore, all of the computational steps can be performed in real-time, and the smart card can encapsulate the fingerprint data and perform the comparison securely inside the card.

Also, Table 3 and 4 show the number of instructions of the cryptography modules measured on the simulator ARMulator [31]. The cryptographic modules need to guarantee the privacy of the fingerprint data transmitted between the smart card and the card reader. The sizes of the fingerprint image and the features are 140 KB and about 1 KB, respectively. As shown in Table 3, the time to require to encrypt and decrypt the fingerprint image using the AES algorithm are about 9.7 seconds and 13.8 seconds in the ARM7TDMI core, respectively. By comparison, the features require only 0.06 second (encryption) and 0.1 second (decryption). Also, as shown in Table 4, SHA1, and ECC for the digital signature can be executed in real-time for the fingerprint image and the features. If a core of the smart card is the 8051 chip, it is impossible to execute the AES algorithm for the fingerprint image in real-time. Furthermore, the digital signature cannot be executed in real-time because the time to require for the ECC algorithm is about 8 seconds.

Finally, the following configuration is assumed to evaluate the performance of each scenario of the client-server model (see Section 4). First, the client and the server have Pentium 4 (2 GHz) processor and Xeon (3 GHz) processor, respectively. The transmission speed of the Internet is assumed to be 10 Mbps. 128-bit AES, 1024-bit ECC, and SHA1 [15, 16] are used as symmetric encryption algorithm, digital signature algorithm, and hash algorithm, respectively, in order to guarantee both the integrity and the confidentiality of the transmitted information. Also, we examined each scenario on the fingerprint images captured with an

Table 4: Number of instructions and cycles required for the digital signature.

		SHA1		ECC (1024-bit)	
		No. of instruction	No. of cycle	No. of instruction	No. of cycle
ARM7TDMI	140 KB	3 091 169	4 709 469	12 528 848	20 327 978
	1 KB	26 990	42 165	12 528 848	20 327 978

Table 5: Evaluation data of cryptography algorithms (Byte, milisecond).

		Generate signature	Verify signature	Size of signature	AES encrypt	AES decrypt
Pentium 4	1 KB	6.359	33.656	48	3.0	4.0
	140 KB	74.703	98.922	47	234.0	328.0
Xeon	1 KB	3.656	19.797	48	1.0	2.0
	140 KB	42.844	59.046	47	125.0	218.0

optical scanner manufactured by NitGen [32]. The size of the fingerprint image and the fingerprint feature is about 140 KByte and 1 KByte, respectively. Finally, the disk access time of the fingerprint data stored in the server is assumed to be 50 miliseconds.

Table 5 shows the measured execution time of several cryptography algorithms used in our evaluation. These data are measured by arithmetic mean of 1 000 executions on each platform.

5.2. Performance Evaluation Results for the Smart Card-Reader Model. As explained earlier, in the case of the smart card with the 8051 chip, secure transmission of both the fingerprint image and the features for all scenarios cannot be guaranteed because ECDSA, especially the ECC and the SHA1 algorithms, cannot be executed in real time.

When the smart card employs the ARM7TDMI core [29], the performance of each of five scenarios is evaluated as follows.

In SCENARIO 1, the cryptographic processes for guaranteeing the integrity and confidentiality of the sensitive fingerprint data transmitted between the smart card and the card reader can be executed in real-time because only the template stored in the smart card is transferred. It is, however, the Store-on-Card strategy that all the fingerprint verification modules except the storage module are executed in the card reader. Therefore, the fingerprint template stored in the smart card needs to be insecurely released into an external card reader in order to be compared with an input fingerprint.

SCENARIO 2 has the lowest security level because it is also the Store-on-Card strategies that the smart card only stores the fingerprint template. Furthermore, in SCENARIO 2, secure transmission of the fingerprint image captured by the fingerprint sensor within the smart card cannot be guaranteed since the fingerprint sensor is built into the smart card. In this case, ECDSA, especially the ECC and the SHA1 algorithms, cannot be executed in real-time.

On the other hand, SCENARIO 3 is the most proper one to integrate fingerprint verification with the smart card because it guarantees higher security level than with the Store-on-Card. It also executes the cryptographic modules for secure transmission in real time because of transferring only the fingerprint features extracted in the card reader.

SCENARIO 4, like SCENARIO 2 cannot guarantee the security and real-time transmission of the fingerprint image captured by the fingerprint sensor within the smart card.

SCENARIO 5 is the System-on-Card that all the fingerprint verification modules take place on the card. This scenario is the best in terms of security as everything takes place on the smart card. As explained in Section 2.3, it is expensive and presents more than one realization problem. Even if the smart card employs the ARM7TDMI core, SCENARIO 5 cannot be executed in real-time because the feature extraction module of fingerprint verification is time consuming as shown in Table 2.

5.3. Performance Evaluation Results for the Client-Server Model. For each of the fingerprint authentication scenarios described in the previous section, we assume that the response time must be less than 5 seconds for real-time execution. As we expect, the server processes most of the time-consuming tasks in SCENARIO 1_1, whereas the clients have the heavy workload in SCENARIO 1_2. The extreme case is SCENARIO 1_3 where the clients do almost everything. However, the security of SCENARIO 1_3 is weaker.

In this section, we will evaluate the three scenarios in terms of the response time versus workloads imposed on the server. In other words, we will investigate the maximum workload that the server can handle with less than 5 seconds response time, or system time in queueing theory of each scenario. We adopt M/D/1 queueing results assuming that the clients request services to the server in random fashion, but the server processes the jobs in deterministic fashion.

In an M/D/1 system, the response time is given by

$$w = \frac{1}{\mu} + \frac{\lambda}{2\mu(\mu - \lambda)}, \tag{1}$$

where W is the response time (or the system time), μ is the service rate, and λ is the arrival rate. Here, λ is the job request rate to the server by the clients, and λ is assumed to increase as the number of clients increases.

In SCENARIO 1_1, fingerprint acquiring (1 000 milliseconds) is done by the sensor, whereas feature extraction (225 milliseconds), generation of signature for feature (6.359 milliseconds), and encryption (3.0 milliseconds) tasks are done by the client. Additionally, we consider the communication setup time for sending and receiving data from the server as *st* and the data transmission time to be 1.6 milliseconds: that is, 2×0.8 milliseconds or 2 times of 1 KB by 10 Mbps Internet transmission. Thus, the total sums to 1 235.959 + 5 st milliseconds. We take values for st of 1, 5, or 50 millisecond(s) depending on the communication environments. On the server side, decryption for live feature (2.0 milliseconds), decryption for smart card feature (2.0 milliseconds), verifying the signs for live feature and smart card feature (2×19.797 milliseconds), matching twice (2×6.5 milliseconds), and data transmission time of 1.6 milliseconds and the communication setup time (5 st) sum to 58.194 + 5 st, which equal the service time ($1/\mu$). Thus, we may build the response time (W) in the M/D/1 system as

$$w = 1235.959 + 5\,\text{st} + \frac{1}{\mu} + \frac{\lambda}{2\mu(\mu-\lambda)} < 5000. \quad (2)$$

If we take st be 1 millisecond in (2), we have λ being less than 0.01569 in order to meet the total response time being less than 5 seconds. Here, the workload can be interpreted as the number of job requests by the clients to the server in unit time (millisecond).

A similar approach can be applied to SCENARIO 1_2: fingerprint acquiring (1000 milliseconds) is done by the sensor, whereas live feature extraction (225 milliseconds), decryption for smart card feature (4.0 milliseconds), verifying sign for smart card feature (33.656 milliseconds), and matching smart card versus live feature (10 milliseconds) tasks are done by the client. Additionally, we consider the communication setup time for sending and receiving data from the server as st and the data transmission time (0.8 millisecond, 1 time of 1 KB by 10 Mbp*s* Internet transmission). Thus, the total sums to 1 273.456 + 4 st milliseconds. We take values for st of 1, 5, or 50 millisecond(s) depending on the communication environments. On the server side, decryption for smart card feature (2.0 milliseconds), verify the sign for smart card feature (19.797 milliseconds), matching once (6.5 milliseconds) and data transmission time of 0.8 millisecond, and the communication setup time (4 st) sum to 29.097 + 4 st, which plays the service time ($1/\mu$). Thus, we may build the response time (W) in the M/D/1 system as

$$w = 1273.456 + 4\text{st} + \frac{1}{\mu} + \frac{\lambda}{2\mu(\mu-\lambda)} < 5000. \quad (3)$$

If we take st to be 1 millisecond in (3), μ must be less than 0.03008 in order to meet the total response time being less than 5 seconds.

In SCENARIO 1_3, the server is doing practically nothing except encrypting the stored template and sending the encrypted data to the clients. Fingerprint acquiring (1000 milliseconds) is done by the sensor, whereas decryption for DB feature (4.0 milliseconds), verifying signature for DB feature (33.656 milliseconds), decryption for smart card feature (4.0 milliseconds), verifying sign for smart card feature (33.656 milliseconds), live feature extraction (225 milliseconds) and 2 times of matching (2×10 milliseconds) tasks are done by the client. Additionally, we consider the communication setup time for sending and receiving signals from the server as st. Thus, the total sums to 1 320.312 + 2 st milliseconds. We take values for st of 1, 5, or 50 millisecond(s) depending on the communication environments. On the server side, encryption for DB feature (1.0 millisecond), generating signature for DB feature (3.656 millisecond(s)), and the communication setup time (2 st) sum to 4.656 + 2 st, which plays the service time ($1/\mu$). Thus, we may build the response time (W) in the M/D/1 system as

$$w = 1320.312 + 2\,\text{st} + \frac{1}{\mu} + \frac{\lambda}{2\mu(\mu-\lambda)} < 5000. \quad (4)$$

Fixing st be 1 millisecond in (4), we have λ being less than 0.1501 in order to meet the total response time being less than 5 seconds.

Summarizing the results obtained in this section is depicted in Tables 6–8. As we see in Table 6, the server can handle the lowest level of workload with SCENARIO 1_1, whereas the server can handle 9.57 times heavier workload with SCENARIO 1_3. In most reasonable case of SCENARIO 1_2, the server can handle 1.92 times heavier workload compared to the case of SCENARIO 1_1.

Another interesting result comes when we vary the communication setup time from 1 ms to 5 ms and 50 ms, as shown in Tables 7 and 8. As we notice in Tables 7 and 8, the workload that the server can handle within the specified time constraint of 5 seconds changes dramatically as the communication setup time increases. When the communication setup time becomes 5 milliseconds, the server can handle only 5.73 times heavier workload with SCENARIO 1_3 compared to those with SCENARIO 1_1. It becomes worse with the communication setup time being 50 milliseconds, where the server can handle only 3.04

Table 6: Summary of performance evaluation 1 (st = 1 ms).

SCENARIO	1_1	1_2	1_3
Maximum workload	0.01569	0.03008	0.15010
Relative workload	1	1.92	9.57

Table 7: Summary of performance evaluation 2 (st = 5 milliseconds).

SCENARIO	1_1	1_2	1_3
Maximum workload	0.01188	0.02023	0.06810
Relative workload	1	1.70	5.73

Table 8: Summary of performance evaluation 3 (st = 50 milliseconds).

SCENARIO	1_1	1_2	1_3
Maximum workload	0.00310	0.00422	0.00942
Relative workload	1	1.36	3.04

times heavier workload with SCENARIO 1_3 compared to those with SCENARIO 1_1 even though almost everything is performed in the clients with SCENARIO 1_3.

6. Conclusions

We focus on examining which strategies are most appropriate to guarantee the privacy as well as the real time execution requirements for smart card with client-server based fingerprint verification in healthcare information systems. First, five scenarios with three different strategies for integrating fingerprints into a smart card system were examined. Then, three scenarios for implementing of the fingerprint authentication service for large-scale applications were examined with increasing numbers of clients.

Typical task assignment scenarios for fingerprint verification were considered to protect the fingerprint information transmitted and to guarantee both the integrity and the confidentiality of the fingerprint data. The workload characteristics of fingerprint verification and the authentication protocol were obtained by measuring the performance of the primitive operations on an ARM7TDMI-based smart card, a Pentium4-based PC, and a Xeon-based server, respectively. Based on these measured performance, the workload of each scenario of the task assignment was computed and was applied to the smart card reader model. Then, the collective performance of each scenario was analyzed using the workload computed for each scenario.

The analysis results showed that the scenario where the match operation is performed on the smart card with the fingerprint sensor being built into the card reader is the most beneficial in the smart card-card reader model. For large-scale applications, however, this scenario may not be applicable due to the high implementation cost. In the client-server model, the server could handle the lowest level of workload when it does decryption, verification and matching. By comparison, the server can handle 9.57 times heavier workload when the clients do decryption, verification, and matching.

Acknowledgments

This research was supported by The Ministry of Knowledge Economy (MKE), South Korea, under the Home Network Research Center (HNRC) Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA), (NIPA -2009-C1090-0902-0035).

References

[1] H. Dreifus and T. Monk, *Smart Cards*, John Wiley & Sons, New York, NY, USA, 1997.

[2] A. Alkhateeb, T. Takahashi, S. Mandil, and Y. Sekita, "The changing role of health care IC card systems," *Computer Methods & Programs in Biomedicine*, vol. 60, no. 2, pp. 83–92, 1999.

[3] G. Kardas and E. T. Tunali, "Design and implementation of a smart card based healthcare information system," *Computer Methods & Programs in Biomedicine*, vol. 81, no. 1, pp. 66–78, 2006.

[4] M. Marschollek and E. Demirbilek, "Providing longitudinal health care information with the new German Health Card—a pilot system to track patient pathways," *Computer Methods & Programs in Biomedicine*, vol. 81, no. 3, pp. 266–271, 2006.

[5] Gemalto, http://www.gemalto.com/.

[6] Health Smart Card, http://www.healthsmartcard.net/.

[7] C. Pagetti, et al., "A European health card," Final Report, European Parliament, Directorate General for Research, Document for STOA Panel, 2001.

[8] Transcards Project, http://www.sesamvitale.fr/programme/programme_eng.asp.

[9] Netlink Project, http://www.sesam-vitale.fr/netlink/netlk_pres.htm.

[10] Netcards, Trans-European Healthcare Facility Service for Mobile Citizens, http://netcards-project.com/web/frontpage.

[11] G. Kardas and E. T. Tunali, "Design and implementation of a smart card based healthcare information system," *Computer Methods & Programs in Biomedicine*, vol. 81, no. 1, pp. 66–78, 2006.

[12] O. Rienhoff, *Integrated Circuit Health Data Cards (Smart Cards): A Primer for Health Professionals*, PAHO, Washington, DC, USA, 2003.

[13] *Health Insurance Portability and Accountability Act of 1996 (U.S.)*, Public Law no. 104-191, 110 Stat. 1936, HIPAA.

[14] B. Barber, "Patient data and security: an overview," *International Journal of Medical Informatics*, vol. 49, no. 1, pp. 19–30, 1998.

[15] W. Stallings, *Cryptography and Network Security: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, USA, 2003.

[16] D. Hankerson, A. Menezes, and S. Vanstone, *Guide to Elliptic Curve Cryptography*, Springer, New York, NY, USA, 2003.

[17] S. Nanavati, M. Thieme, and R. Nanavati, *Biometrics: Identity Verification in a Networked World*, John Wiley & Sons, New York, NY, USA, 2002.

[18] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, New York, NY, USA, 2003.

[19] R. M. Bolle, J. H. Connell, and N. K. Ratha, "Biometric perils and patches," *Pattern Recognition*, vol. 35, no. 12, pp. 2727–2738, 2002.

[20] B. Schneier, "The uses and abuses of biometrics," *Communications of the ACM*, vol. 42, no. 8, p. 136, 1999.

[21] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: security and privacy concerns," *IEEE Security and Privacy*, vol. 1, no. 2, pp. 33–42, 2003.

[22] S. de Lusignan, T. Chan, A. Theadom, and N. Dhoul, "The roles of policy and professionalism in the protection of processed clinical data: a literature review," *International Journal of Medical Informatics*, vol. 76, no. 4, pp. 261–268, 2007.

[23] Y. S. Moon, H. C. Ho, K. L. Ng, S. F. Wan, and S. T. Wong, "Collaborative fingerprint authentication by smart card and a trusted host," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, vol. 1, pp. 108–112, 2000.

[24] L. Rila and C. Mitchell, "Security analysis of smartcard to card reader communications for biometric cardholder authentication," in *Proceedings of the 5th Smart Card Research and Advanced Application Conference (CARDIS '02)*, pp. 19–28, 2002.

[25] D. Moon, et al., "Performance analysis of the Match-on-Card system for the fingerprint verification," in *Proceedings of the*

International Workshop on Information Security Applications, pp. 449–459, 2001.

[26] Y. Chung, D. Moon, T. Kim, and J.-W. Park, "Workload dispatch planning for real-time fingerprint authentication on a sensor-client-server model," in *Proceedings of the 5th International Conference on Parallel and Distributed Computing: Applications and Technologies (PDCAT '04)*, vol. 3320 of *Lecture Notes in Computer Science*, pp. 833–838, 2004.

[27] X.9.84, http://www.x9.org/home.

[28] A. D. Boyd, C. Hosner, D. A. Hunscher, B. D. Athey, D. J. Clauw, and L. A. Green, "An 'Honest Broker' mechanism to maintain privacy for patient care and academic medical research," *International Journal of Medical Informatics*, vol. 76, no. 5-6, pp. 407–411, 2007.

[29] C. Quantin, O. Cohen, B. Riandey, and F.-A. Allaert, "Unique Patient Concept: a key choice for European epidemiology," *International Journal of Medical Informatics*, vol. 76, no. 5-6, pp. 419–426, 2007.

[30] D. Burger and T. Austin, "The simplescalar tool set, version 2.0," Tech. Rep., University of Wisconsin, Madison, Wis, USA, 1997.

[31] ARM, http://www.arm.com/.

[32] Nitgen, http://www.nitgen.com/.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 967046, 11 pages
doi:10.1155/2009/967046

Research Article

Development of a New Cryptographic Construct Using Palmprint-Based Fuzzy Vault

Amioy Kumar[1] and Ajay Kumar[1,2]

[1] *Biometrics Research Laboratory, Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110 016, India*
[2] *Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

Correspondence should be addressed to Ajay Kumar, ajaykr@ieee.org

Received 7 October 2008; Accepted 16 July 2009

Recommended by Stephanie Schuckers

The combination of cryptology and biometrics has emerged as promising component of information security. Despite the current popularity of palmprint biometric, there has not been any attempt to investigate its usage for the fuzzy vault. This paper therefore investigates the possible usage of palmprint in fuzzy vault to develop a user friendly and reliable crypto system. We suggest the use of both symmetric and asymmetric approach for the encryption. The ciphertext of any document is generated by symmetric cryptosystem; the symmetric key is then encrypted by asymmetric approach. Further, Reed and Solomon codes are used on the generated asymmetric key to provide some error tolerance while decryption. The experimental results from the proposed approach on the palmprint images suggest its possible usage in an automated palmprint-based key generation system.

1. Introduction

Hacking of the information is widely considered as one of the potential attacks on any secure system. Authentication systems should be designed to withstand such attacks when deployed for critical security applications such as e-commerce and accesses to restricted data/buildings. Biometric-based authentication is considered as one of the most secured systems whenever high privacy is demanded. However, such authentication systems itself follow stepwise procedural algorithms, like feature extraction, matching, classification, and so forth, for authentication/verification purposes [1]. As biometric templates are required at each step, it increases the possibilities of intrusion at every step and requires additional security management [2]. For instance, even a most secure authentication system is not reliable if it cannot defy the attacks on the stored database, or if an intruder can intercept the template features generated from the biometric traits. Recent research efforts have developed some promising ideas to resist attacks on biometric authentication system. One of such proposed solutions is to cancel the tainted biometric features and regenerate the new one for authentication purposes (also known as cancelable biometric [3]). BioHashing technique is frequently used to transform (noninvertible) biometric template into some other representations using one-way hash functions. This reissuance of the biometric templates can withstand the attacks on stored templates and widely accepted as a solution to the intrusion in extracted features. The most acknowledgeable work in this area is to provide cryptography-based security at different stages of biometric authentication. Cryptography is one of the most effective ways to enhance the security of the information system via its encryption and decryption modules [4]. Even so, the weakest link of cryptography-based security systems is the associated secret key. While the simple memorized key can be easily intercepted, a long and complex key needs extra storage management like tokens, smart cards, and so forth. Consequently, the smart card-based solutions came in existence. To provide an aid to security, the cryptographic keys are now stored somewhere (e.g., in a computer or on a smart card) and released based on some alternative authentication mechanism. The most popular mechanism used for this purpose is password-based security [5], which

is again a long string and difficult to make secure, as now the whole security depends upon the password given, used for authentication.

As a solution, a secure encryption key can be associated with a biometric signature to ensure *the integrity and confidentiality* of communication in distributed systems. Many of the limitations of the password and PIN-based encryption schemes can be alleviated by using biometric features, which are unique and can be conveniently extracted from every user. The biometric-based encryption requires physical presence of persons to be authenticated and is therefore reliable, convenient, and efficient. The encryption keys can be generated using low-level combination of biometric features and cryptology. Jules and Sudan [6] have proposed the generation of a secure vault using an unordered set, to lock any secret inside and referred it as fuzzy vault. The concept of fuzzy vault has been further explored by Uludag et al. [7], where they used fingerprint templates as an unordered set to create the vault around the secret. They further utilizes error correcting codes, such as Reed and Solomon code to produce some error tolerance in the input biometric templates, while decryption module.

However, the motivation to protect secret key involved in cryptographic modules using biometric based fuzzy vault can have several drawbacks due to different cryptographic approaches. While the symmetric cryptographic approaches suffered authentication problems, asymmetric approaches are computationally intensive (as further discussed in Section 3). We, therefore, proposed the combination of both symmetric and asymmetric cryptographic approaches (which is referred to as double encryption in this paper) into the fuzzy vault to meet high-security standard and utilize the advantages of both approaches in a common domain. In the recent years, biometric features such as face, iris, fingerprint, hand geometry, palmprint, and signature have been suggested for the security in access control. Most of the current research in biometrics has been focused on fingerprint and face. The recent research on face recognition has shown some thorny problems regarding pose, lighting, orientation, and gesture which made it less reliable as compared to other biometrics. Fingerprint identification has successfully implemented and widely accepted in most of the cases for recognition purposes. However, it also has difficulties regarding feature extraction. The fingerprint features are very difficult to extract from the elderly, laborer, and handicapped users. As a result, other biometric characteristics are receiving increasing attention. Moreover, additional biometric features, such as palmprint and hand geometry, can be easily integrated with the existing authentication system to provide enhanced level of confidence in personal authentication. We explored the usage of palmprint biometric to create fuzzy vault. The prior works in this area is summarized in Section 2, while the detail of the earlier cryptographic approaches is presented in Section 3. Double encryption is explored in Section 4. The proposed system is discussed in Section 5. The experimental results from the performed approach are summarized in Section 6. This section also includes a summary of related prior work. Finally, the main conclusions from this paper are summarized in Section 7.

2. Prior Work

The issue of nonrevocable biometric has been investigated by Ratha et al. [3] by introducing the concept of cancelable biometrics. Davida et al. [8] proposed majority decoding and error correcting codes-based technique to generate the cancelable biometric features. The approach is further utilized using optical computation techniques in [9] and using keystroke dynamics in [10]. Sautar et al. [9] were the first to commercialize the concept in to their product *bioscrypt*. They applied Fourier transform and majority coding to reduce the feature variation. A predefined random key is locked by biometric sample using phase angle product, and this product can be further unlocked by other genuine biometrics. The performance analysis is however not reported. Connie et al. [11] used the concept of BioHashing by calculating fisherprojections. However, the results shown by them are based on the assumption that the generated token or keys will be never stolen or shared. This is quite unrealistic and creates doubts about real evaluation. The study of such unrealistic evaluation has been presented by Kong et al. in [12]. One of the innovative works proposed in this area is by [2], where the authors utilized random orientation field into the feature extractor to generate cancelable competitive codes. The authors further considered all the three attacks possible (template reissuance, replay attacks, and database attacks) to provide a complete secure system. To protect the generated cancelable competitive codes (replay attacks) [2], the idea of one-time pad (OTP) ciphers is explored. The OTP [13] is a symmetric cipher (same key is used for both encryption and decryption) generated by applying XOR between the randomly generated key and the plaintext. The decryption can be done using the matched OTP and the key (used for encryption). The advantage with OTP is that each encryption is independent to the next encryption, and random key can be used only once for encryption. Hence, theoretically there is no way to break such encryption just by analyzing a sequence of message. Although OTP encryption has advantages over other encryption algorithms, still it has some open issues like (i) the key involved for decryption should be identical to encryption once and hence required safe communication of key to the decrypting party [13]; (ii) the number of bits in the key is same as in the plaintext which makes the algorithm computationally inefficient for encrypting bulk data; (iii) one of the major requirements of the algorithm is that not part or bit of the key should be ever reused in any other encryption; otherwise it is easy to break it [14]. (Synchronized OTP generator can be employed to counter such problems.) Authors in [15] proposed a new cryptosystem by generating 1024 bits binary string, extracted from the differential operations. The string is then mapped to 128 bits encryption key using a Hash function. The approach is novel and secure in many respects but still has issues to resist against attack on generated encrypting key using Hash function, as raised by Kong et al. in [12].

In most of the works proposed in literature of cancelable biometrics, security of system depends upon the generated unique code from a particular one-way hash functions. Thus the system is secure till the unique code is not compromised

and hence requires extra security management. Juels and Sudan [6] have presented a promising model which was an improvement on the prior study by Juel and Wttenberg in [16]. They have produced a significant improvement by modifying the Scheme of Davida et al. (in using error correcting code size) [8] by introducing Reed and Solomon error correcting coding theory in their fuzzy vault. Their contribution is to hide any secret in fuzzy vault using polynomial construction under unordered set. The secret can be retrieved back by polynomial reconstruction, if certain points of the unordered set can be known at receiving end. The security of the scheme mainly depends upon polynomial construction and reconstruction problem. Uludag et al. [7] have combined the concept of fuzzy vault with biometrics (fingerprint) by using biometric template as an unordered set. Uludag and Jain [17] proposed to use minutiae-based features from the fingerprints for locking and unlocking the vault. However, this approach is limited to its usage due to its inability to eliminate the inherent variability in minutiae feature. Nandakumar et al. [18] have attempted to eliminate such variability using *helper data* and illustrated promising results. Hao et al. [19] use iris biometric for generating cryptographic keys and a combination of Reed and Solomon and Hadamard error correcting theories for error tolerance. Calancy et al. [20] proposed a smart card-based fuzzy vault that employed fingerprints for locking and unlocking. The presumption that acquired fingerprint images are prealigned is not realistic and could be the possible reason for high false rejection rate (30.0%) reported in the paper. Lin and Lai [21] have done remarkable work in order to prevent repudiation but their work still required smart card and password for better implementation and hence reduces its usability. Recently, a modified fuzzy vault scheme is proposed in [19] using asymmetric cryptosystem. Having generated RSA public and private keys, authors have used Reed-Solomon coding to convert the keys in to codes. Further they used two grids, one for codes and the other for biometric features. The elements in the corresponding grids are in same positions. The unlocking of vault only requires the knowledge of the correct positions of the numbers in any of the grids. However, this approach utilizes the asymmetric cryptosystem and has all the problems associated with such systems. Moreover, the database used for the experimental evaluation is too small (9 users) to generate any reliable conclusion on the performance. In summary, a different range of biometrics has been used for fuzzy vaults in literature. However, with few notable exceptions, for example, [15, 19], with small false rejection rates, the average FAR of 15% has been cited.

In contrast to prior work in this area, we proposed [22] fuzzy vault-based security to withstand the attacks on secret key employing palmprint. The secret document/information can be first encrypted using double encryption. The symmetric key approach can be easily employed to encrypt bulk data. The attacks on security of symmetric key (secure communication, authentication, as detailed in Sections 3.1.1 and 3.1.2 in this paper) are reduced by encrypting it again using asymmetric cryptographic approach. Finally, the private key of asymmetric approach (at the end of double encryption) is protected by creating fuzzy vault around it. The approach is to firstly employ double encryption to strengthen the security system and reduce the shortcomings associated with both symmetric and asymmetric cryptographic approaches and finally to utilize the palmprint features to create fuzzy vault around the key at the end of double encryption.

The main contributions of this paper can be summarized as follows. Firstly, this paper investigates a new approach for fuzzy vault using palmprint biometric. Secondly, unlike prior work in literature, this paper proposes a combined cryptosystem which successfully exhibits the advantage of both symmetric and asymmetric cryptography. It may be noted that the asymmetric approach (RSA, named as initials of Ron Rivest, Adi Shamir, and Leonard Adleman) for encryption has been estimated to be very slow as compared to traditional symmetric approach (Data Encryption Standard, abbreviated as DES) [4]. Therefore the proposed approach is to use symmetric cryptography to encrypt the entire document and then we encrypted symmetric key using asymmetric (RSA) approach. The palmprint-based fuzzy vault is then constructed around decryption key. Finally, we investigate the performance of the palmprint-based cryptosystem on a large dataset and achieve promising results.

3. Cryptographic Approaches

The objective of this work is to incorporate both symmetric and asymmetric cryptographic approaches into the fuzzy vault in order to ensure higher security and utilize the advantages of both systems in a common domain. This is referred to as double encryption. The approach is to use symmetric key approach (DES) for encrypting the secret document, and the generated symmetric key is again encrypted by asymmetric approach (RSA). In the next subsections, both symmetric and asymmetric approaches are briefly introduced, and then the proposed approach utilizing the combination of both approaches is discussed.

3.1. The Symmetric Cryptosystem. The symmetric approach is most commonly used cryptosystem, as the system is easy to implement and more importantly it has very fast encryption speed [4]. Symmetric algorithms, such as, DES, Triple DES, and Rijndael [4], provide efficient and powerful cryptographic solutions, especially for encrypting bulk data. Let $X = [x_1, x_2, x_3, \ldots, x_m]$ be the secret message required to be hidden by source A (Lucie). The m letters of message are alphabets. The message is intended to B (Bryan). Lucie generates its symmetric key, say K_{Sim}, and uses this key to lock secret message X:

$$Y = K_{Sim}(X). \tag{1}$$

She then sends the encrypted (locked) message and the respective symmetric key (K_{Sim}) to B (Bryan). Receiver B (Bryan) used the symmetric key to decrypt the message:

$$X = K_{sim}(Y). \tag{2}$$

In the presented work we have used advance encryption standard (AES) as a symmetric cryptosystem, which is advanced version of data encryption standard (DES). The AES is symmetric key-based cryptosystem which is based on the principle of block and substitution cipher. The AES algorithm uses substitution boxes, polynomial matrices, and symmetric key to convert a plain text to cipher text. These are the parameter for AES cryptosystem and required to be generated first before the encryption module [4]. Although symmetric key algorithm is very fast and efficient in bulk data encryption, it can sometimes fail to ensure high-security requirements. There are few shortcomings with the usage of symmetric key cryptography. We now detail the problems associated with symmetric key algorithms.

3.1.1. The Problems behind Authentication. Ensuring the integrity of received data and verifying the identity of the source of that data is of major concern to ensure the security in data communication. A symmetric key can be used to check the identity of the individual, as it requires the presence of symmetric key, but this authentication scheme can have some problems involving trust. The problem is that this scheme cannot discriminate between the two individuals who know the shared key. For example, any person having control on Lucie's private particulars can make any fraud message to her pals by pretending himself as Lucie. This not only allows intruder to do any unauthorized work in place of Lucie but also creates problems for other related persons. This uncertainty with symmetric approaches made them useless whenever high confidentiality required in the communication system. The above discussed issues can lead to the position where there is no stand to deny if the disputes were to arise. The relevant example is of repudiation when Lucie's friend renews the contract signed by Lucie without telling her and repudiates from the fact by claiming that someone else might have stolen the key from Lucie to sign the contract. This concludes the key point that the communication system must present nonrepudiation between communicating parties. The major weakness with symmetric approach is that they sometimes fail to authenticate persons in communication.

3.1.2. The Problems behind Security of Key. The other problem associated with this system is to ensure the security of the involved symmetric key and how to exchange it safely. The security of a signed document depends upon the secret key involved as only secret key can ensure the decryption of this document. Thus for a secure communication system the secret key should be exchanged safely. One of the shortcomings of the cryptographic approaches is that they do not emphasize on key exchange problems. The asymmetric approaches such as RSA, DSA, and ECC are very good substitution of symmetric approach as it eliminates many of its shortcomings. Both of the above discussed problems can be alleviate by using asymmetric approach.

3.2. The Asymmetric Cryptosystem. The conventional symmetric cryptosystem is similar to a lockbox with a combination lock. This combination lock opens and closes with one and the single combination, that is, the key that can be used for both opening and closing the box. However, the asymmetric approach uses a single lock that has two distinct combinations, one for opening and one for losing. This approach allows effective control over who can place or remove the contents in lockbox by assigning one of the combinations as the secret and the other one as public. This added flexibility offers two distinct advantages: confidentiality without prior key exchange and the enforcement of data integrity. Now for this approach, B generates a related pair of keys: a public key K_{pub} and a private key K_{pri}. The K_{pri} is known only to B, whereas K_{pub} is publicly available to everyone and therefore accessible by A also. With the message X and the encryption key K_{pub} as input, A forms the cipher text, denoted as Y, as follows:

$$\begin{aligned} Y &= K_{pub}(X), \\ Y &= [y_1, y_2, y_3, \ldots, y_m]. \end{aligned} \tag{3}$$

The intended receiver in the position to matching is able to invert above using the following transformation:

$$X = K_{pri}(Y). \tag{4}$$

In this work, we have used RSA cryptosystem which is the most commonly used asymmetric approach. A traditional RSA algorithm [23] requires two randomly generated prime numbers [24]. For the security of RSA algorithm, the prime numbers should be bigger (512 bit in our case) and randomly chosen. Any secret encrypted using public key can only be decrypted by using private key and vice versa. The main points involved in encryption and decryption are as follows.

Lucie does the following:

(1) obtains the recipient Bryan's public key,

(2) represents the plaintext message as a positive integer,

(3) computes the ciphertext,

(4) sends the ciphertext to Bryan.

Recipient Bryan does the following:

(1) uses his private key to compute positive integer,

(2) extracts the plaintext from the integer representative.

Using RSA algorithm, asymmetric cryptosystem can be employed to solve a number of problems regarding symmetric cryptographic approach. But as compared to symmetric approach, asymmetric approach also has few drawbacks.

3.2.1. The Problems behind RSA. The private and public key approach of RSA cryptosystem can be substitute of the key exchange problem involved with symmetric approaches, but the major problem regarding this approach is the distribution of public keys. Having signed the secret document with Bryan's secret key, Alice must ensure that the public key available is really Bryan's key but not of intruder Carol. The management and security of private key is also a major concern. The other important problem with asymmetric cryptography is that the processing requires intense use of

the central processing unit as it is computationally intensive and requires a lot of mathematical computations. This may be a real problem when several simultaneous sessions are required. The asymmetric approaches like RSA, DSA, and so forth are generally known to be slower (about 100 times slower) [4] than symmetric approaches like DES, AES, and so forth. As a conclusion one can argue that the symmetric cryptography is highly suitable for encrypting and decrypting the bulk of messages on data lines. However, the associated problem of providing all the recipients with an advanced copy of secret key can be expensive and hazardous. The insecurity associated with the distribution of all the necessary secret keys to all the recipients on a regular basis is very high. In summary, working with RSA cryptosystem can certainly eliminate several drawbacks associated with symmetric approaches. However, this cryptosystem still has some problems regarding complexity of algorithm as it works very slowly (whenever a bulk data encryption is required) due to the fact that it is mathematically intensive and requires extra management for public keys.

4. Double Encryption

One way to alleviate above discussed problems associated with the symmetric and asymmetric cryptographic approaches is to use double encryption. A secret message is encrypted using fast symmetric algorithm; the secret key is then encrypted using asymmetric cryptography; the Ciphertext (encrypted message) and the encrypted keys are finally sent to the recipient. Asymmetric cryptography is slow (computationally intensive), but not too slow to encrypt such a small (as compared to secret message) bits as a symmetric encryption key. Upon receipt, the recipient can easily use his/her private asymmetric key to decrypt the symmetric key. Further that symmetric key can be used to quickly decrypt the message file. This idea not only resolves the problem using both approaches but is also more computationally sound.

4.1. Why Double Encryption? Most of the problems regarding symmetric/asymmetric approaches can be remedied using double encryption. The advantages of the symmetric approaches are utilized to encrypt bulk of the data, while asymmetric approaches are used to provide authentication/verification to secure communication (as discussed in Sections 3.1.1 and 3.1.2 symmetric approaches are sometimes fail in authentication purposes). Using double encryption, a message (may be bulk data) can be encrypted by symmetric key approach, while the key is again encrypted by public/private keys of asymmetric approach. Once the message is encrypted by public key of recipients, it can only be decrypted by its private key. This ensures a safe communication between the source and the verified/authenticated recipient. On the other hand, if the message is encrypted by private key of the recipient, it can only be decrypted by corresponding public key (which is publicly available). This process authenticates the source of encryption and therefore prevents any possible repudiation or denial from the message generator.

4.2. Prior Work in Double Encryption. The concept of double encryption is not new in cryptographic literature [25–27]. However, most of the related work is centered on the implementation of cryptographic encryption and decryption modules [28–30]. Some of these notable efforts can now be outlined. Nishimura et al. [25] in their recent European patent have detailed the concept of encrypting symmetric key with public and private keys of asymmetric approach. Their developed approach ensures that when a doubly-encrypted message is received, it is sent by a particular/authenticated user; also the recipient of this message is a specific/verified user(s). Doh et al. [31] have presented double encryption-based optical security system. They have utilized the facial images by using random-phase patterns in the spatial plane and the Fourier plane and a personal information image consisting of a personal identification number (PIN). With the recognition of PIN, the authentication of the encrypted personal identification card has done by primary classification and recognition of the PIN with the proposed multiplexed MACE phase-encrypted filter. In this technique, the possibility of spoofing is significantly decreased using the double-identification process. Z. Liu and S. Liu [32] proposed Double image encryption based on iterative fractional Fourier transform. They used to encrypt two different images into a single one simultaneously by their amplitudes of fractional Fourier transform with different orders.

In contrast to proposed double encryption schemes, we explored this concept for fuzzy vault. The combination of cryptographic algorithms with biometrics has been presented in several prior publications, for example, [2, 15, 17–19]. Some of these attempts have been focused to hide the secret information in biometric-based fuzzy vault [17, 18] while others used to generate cryptographic keys using biometrics ([15, 19]) to hide the secret information. *Our contribution to literature is that we attempt* to hide secret information using double encryption (via symmetric and asymmetric cryptographic approaches). In order to strengthen the cryptographic approaches, we closed the asymmetric key (at the end of double encryption) by creating palmprint-based fuzzy vault around it. *Our scheme is quite unique in the sense that, it overcomes any dependency on generated secret key (like [11, 33]) in cryptographic approaches and utilized the unique palmprint features to create the fuzzy vault.*

4.3. Motivation to Fuzzy Vault. One of the most important applications of double encryption is that it can overcome many of the problems associated with the symmetric key approach (as the symmetric key is again encrypted by asymmetric approach). In addition, the level of security offered by the resulting asymmetric key, at the end of double encryption, is very high and desired to secure the entire system. *In the cryptographic literature, security of asymmetric key (at the end of double encryption) is generally questioned as the main/key weakness of the double encryption [28].* In the proposed approach, we have utilized the concept of fuzzy vault to overcome this shortcoming of double encryption by locking the private key in the vault. This combination of double encryption with biometrics (fuzzy vault) can

overcome most of the weaknesses regarding symmetric and asymmetric cryptographic approaches.

5. Proposed System

Let X denote the dummy message to be encrypted and let K_{sim} be the symmetric key, used to encrypt the document. In order to encrypt the message X, the symmetric key can be generated using AES algorithm. Let the symmetric key be denoted by K_{sim}. Now for making system more secure and overcome the difficulties of symmetric key approach, (key exchange problem, confidentiality, etc.) the generated symmetric key again is encrypted by asymmetric approach using RSA algorithm. Let the public and private keys associated with the RSA cryptosystem are denoted by K_{pub} and K_{pri}. We will use this generated public key K_{pub} for encryption and the generated private key K_{pri} for decryption. Equation (5) summarize the complete procedure:

$$\begin{aligned} Y &= K_{sim}(X), \\ T &= K_{pub}(Y), \\ Y &= K_{pri}(T), \\ X &= K_{sim}(Y). \end{aligned} \tag{5}$$

Figure 1 illustrates the complete block diagram and includes all the key steps in the double encryption algorithm. For the traditional RSA cryptosystem, the public key has made publicly available while private key has kept private. The cipher text has been generated with the publicly available encryption key while it is decrypted with the private key kept private. The security of the system depends upon the secrecy of private key.

5.1. Palmprint-Based Fuzzy Vault. One of the key objectives of this work is to investigate the usage of palmprint biometric in the development of a cryptographic construct. The palmprint-based cryptosystem can have higher user acceptance and performance. Despite the recent popularity of palmprint-based systems [34–36], there has not been any attempt to investigate its usage for the fuzzy vault. The palmprint literature has cited number of advantages of palmprint biometric: (i) due to large surface area, the region of interest for palmprint is larger as compared to fingerprint and hence more features can be extracted, (ii) the chances of damaged hand are less than damage fingerprint for a person, (iii) even the presence of very less amount of dirt or grease can affect the performance of fingerprint verification, but having little effect in case of palmprint, and importantly (iv) higher user acceptance for palmprint mainly due to the stigma of fingerprints is associated with criminal investigations.

The double encryption method detailed in previous section incorporates both the ideas of symmetric and asymmetric cryptosystem efficiently and minimizes most of the shortcomings associated with both approaches. The other important concern of the system is the management of private key, as at the end of double encryption security of the entire system depends upon the security of private decryption key. The security to private key can be ensured by the use of well-known concept fuzzy vault detailed in [6]. Using the concept of fuzzy vault, our main goal is to hide this decryption key using biometric features to provide some security to the decryption key and make the whole system tailored for its practical usage. The combination of cryptographic keys with biometric offers several advantages including the fact that this removes the extra key management efforts required by the user and ensures that it is nontransferable. This method of protecting the private key not only makes the usage of smart cards redundant but also makes the user self dependent for its key. The difficulties lie in the fact that the cryptographic algorithm expects that the keys should be highly similar for every attempt for successful access, but it is clearly not the case with a typical biometric. The key is to use suitable coding theory scheme which can tolerate errors. We have used Reed and Solomon (RS) coding scheme for providing some error tolerance while decryption. This error tolerance is essentially required to handle inherent variations in palmprint (biometric) features from the same user during decryption. These variations can be attributed to the scale, orientation, and translational variations in the user palmprint due to peg-free imaging. The RS coding scheme has error correcting capacity of $(n - k)/2$, where n is the length of code and k is the length of message, and used to encode decryption key K_{pri}.

We can easily vary (k, n) during the training stage/phase and achieve the best possible combination for minimum false acceptance and rejection rates. The proposed design of palmprint vault is quite similar as for the fingerprint [37]. Let the codes generated by R-S coding theory be of size b. Then we generate a grid of size $b \times 3$ such that ith row of grid contains ith place. The rest two places are filled by random numbers generated during encoding. We designate this grid as grid F. Further, a grid of same size is generated, and the biometric features are placed at the same position as in the case of RS codes. The rest of the two places are filled with numbers such that each row is maintained in the arithmetic progression. Let us designate these numbers as tolerance value. These points are actually the chaff points making the grid fuzzy. We called this grid as grid G. To unlock the vault we only need to know the correct positions of the elements in grid G, which can be achieved by comparing the input palmprint features with all the numbers in the corresponding row. Taking minimum of the distance, we can conveniently locate the positions of actual biometrics from grid F and hence the corresponding positions for the codes in grid G. The idea of generating such random numbers to combine with biometric templates is somewhat similar to as discussed in [2]. However, in contrast to [2], our approach is to add the tolerance value to the feature vectors. Out of the three places on the grid G, only one place is filled by original feature, and the rest two places are filled by original features added with tolerance value. The work presented in [2] has been motivated from the random orientation field, which is inserted into the feature extractor to generate noise-like feature codes. The inverse Reed and Solomon codes are used to decode the codes. One

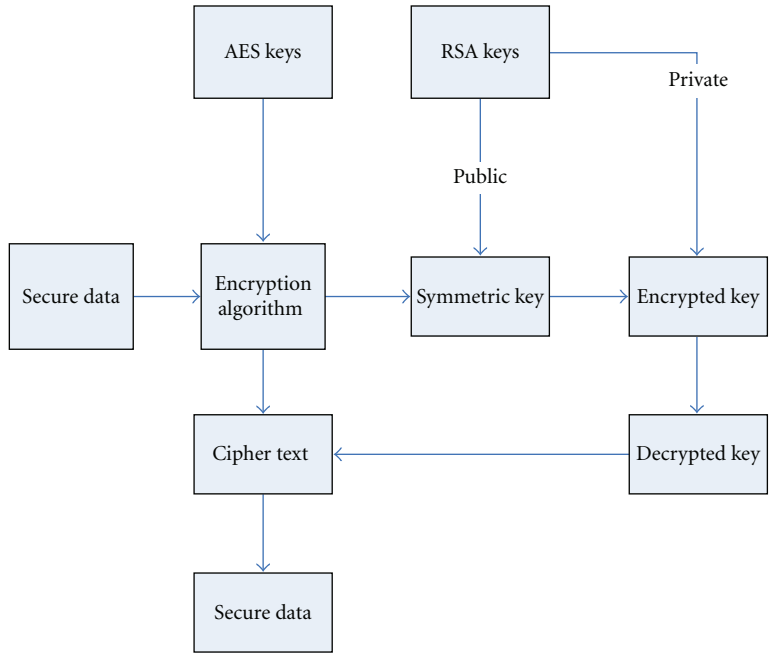


Figure 1: Block diagram for the double encryption.

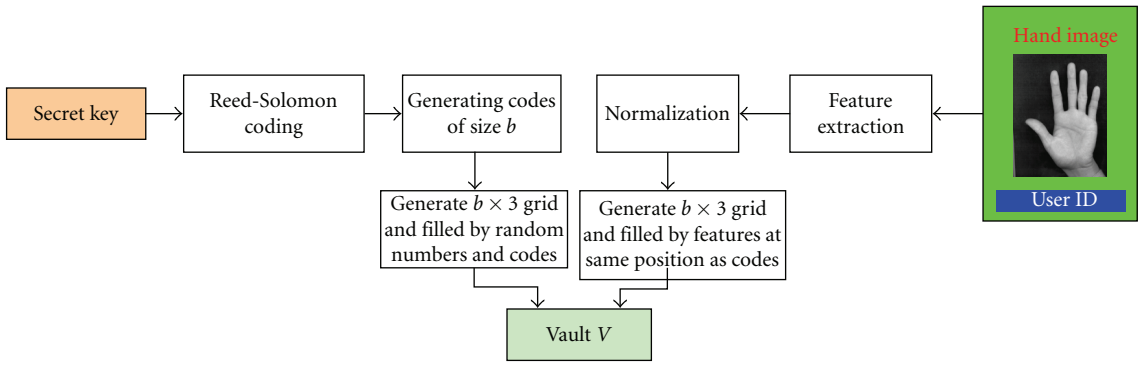


Figure 2: Block diagram for locking of the vault.

can select the suitable values for n and k to control the error occurred due to the variability in palmprint features. *The motivation behind choosing the tolerance for the palmprint features is to make them fuzzy such that an imposter is not able to predict the feature vector just at random*. The block diagram for locking the vault using palmprint features is shown in Figure 2. The corresponding unlocking mechanism is illustrated in Figure 3. Once the procedure for the locking and unlocking of vault is determined, we fix the criteria for the genuine users to successfully open the vault while rejecting the imposter attempts. The vault is said to open successfully, if the codes retrieved from grid F (created by R-S codes) using the query palmprint features will be identically equal to the codes used at the time of locking. The inverse R-S codes can be applied to the retrieved codes to get back the original symmetric decryption key. Finally, this decryption key should successfully decrypt the secret private RSA key.

5.2. Feature Extraction and Normalization. The palmprint features employed in this work were extracted from the palmprint images *acquired from the digital camera* using unconstrained peg-free setup in indoor environment. The extraction of region of interest, that is, palmprint, from the acquired images is similar as detailed in [38]. The Discrete Cosine Transform (DCT) is used for the characterization of unique palmprint texture. The DCT is highly computationally efficient and therefore suitable for any online cryptosystem. (DCT is the basis of JPEG and several other standards (MPEG-1, MPEG-2 for TV/video, and H-263 for video-phones).) As illustrated in Figure 4, each of the 300×300 pixels palmprint image is divided into 24×24 pixels overlapping blocks. The extent of this overlapping has been empirically selected as 6 pixels. Thus we obtain 144 separate blocks from each palmprint image. The DCT coefficients from each of these N square block pixels, that is, $f(x, y)$, are

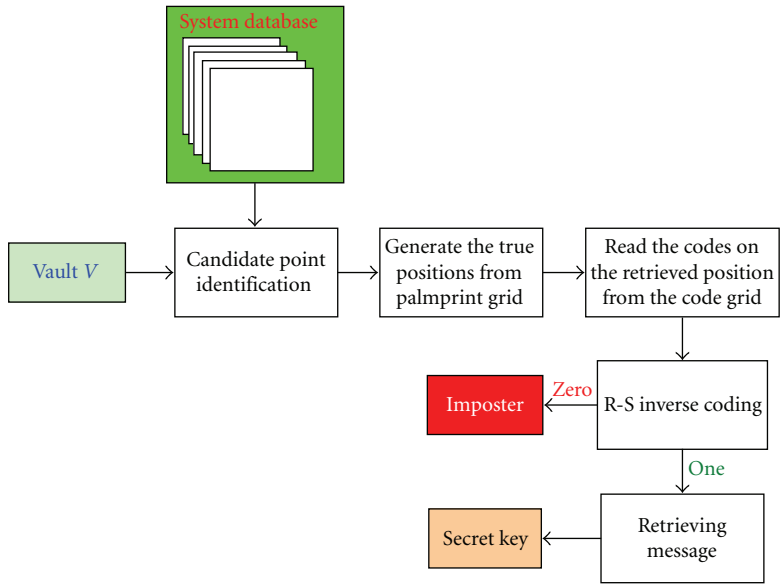


Figure 3: Block diagram for unlocking of the vault.

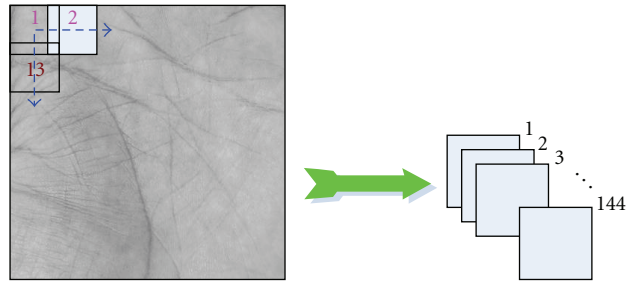


Figure 4: Localization of 144 overlapping palmprint image subblocks for feature extraction.

obtained as follows:

$$C(u,v) = \varepsilon(u)\varepsilon(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{\pi u}{2\cdot N}(2x+1)\right], \times \cos\left[\frac{\pi v}{2\cdot N}(2y+1)\right], \quad \text{where } u,v = 0,1,\ldots,N-1,$$

$$\varepsilon(u) = \varepsilon(v) = \begin{cases} \sqrt{\frac{2}{N}} & \text{for } u \neq 0, \\ \sqrt{\frac{1}{N}} & \text{for } u = 0. \end{cases} \tag{6}$$

The standard deviation of DCT coefficients, obtained from each of the overlapping blocks, is used to characterize the region. Thus we obtain a feature vector of 144 values. High degree of intraclass variability in the palmprint features, mainly due to peg-free imaging, poses serious problems in the unlocking of the constructed vault by the genuine. The variability in feature vectors has been reduced with the help of Z-rule normalization. Corresponding to each feature vector, the training images are normalized, and then their mean and standard deviations are used for feature normalization in the test phase. This normalization reduces the interclass variability of the extracted features and very much helpful in fixing the tolerance for fuzzy vault.

6. Experimental Results

The implementation of the system consists of generation of RSA cryptosystem. A dummy document is then double encrypted using symmetric and asymmetric keys. After double encryption, fuzzy vault is created around the private key by generating grids using R-S codes and palmprint features. The evaluation is based on varying tolerance value over the range, and the corresponding false acceptance rate (FAR) and false rejection rate (FRR) are then computed. The palmprint database consisted of the left-hand images from the 85 users, and two images from each of the users are employed. The first enrolled palmprint image from each of the users was employed to lock the vault. The successful opening with the second enrolled palmprint image of the same user was considered as genuine match while opening with all the other enrolled test images from other enrolled

Table 1: Summary of experimental results.

Key length	EER (%)	Tolerance
306	0.905	1.060
307	0.375	0.995
308	0.752	1.065
309	2.134	1.118

users (i.e., 84 users) was considered as imposter matches. Thus our performance estimation, that is, FAR and FRR, is based on 84×85 imposter and 85 respective genuine attempts. The decisions from the FAR and FRR depend upon choice of tolerance. We performed several experiments to select the best value of this tolerance. Figure 5 illustrates the performance of the proposed palmprint-based vault. Figure 5(a) illustrates the variation of FAR and FRR scores with the tolerance while Figure 5(b) illustrates the receiver operating characteristics (ROC). The RSA cryptosystem used in our program has some variations in key length [39]. The RSA implementation has utilized the string format to generate the RSA keys, and its length varies from 306 to 309 (detailed in Section 6.1) [26]. As cryptographic keys are supposed to be same at each application, authentication rates can vary with each length size of the generated key. Table 1 illustrates the variation in experimental results (equal error rate) with the key length and the corresponding tolerance value.

6.1. Discussion. While the idea of incorporating biometrics within cryptographic constructs has shown promising results than password-based authentication, the system still has open issues. The biometric modalities investigated for the experimental evaluation has been quite limited and most of the prior work is focused on fingerprint. Recently, iris [19], face [33], and signature [40] have also been investigated and yielded promising results. However, summary of prior work presented in Table 2 suggests that much of the work has been simulated on a small dataset, such as [37] has used 9 users, [41] has used 10 users, and [9] has used 20 users, which is quite small to generate a reliable conclusion on performance.

Despite the current popularity of palmprint biometric, there has not been any attempt to investigate its usage for the fuzzy vault. This paper [22] therefore investigated the possible usage of palmprint in fuzzy vault to develop a user friendly and reliable crypto system. The image dataset used for the experiments (85 users) was acquired from unconstrained peg-free setup as such images are more realistic and expected to show large variations.

Our experimental results illustrated the EER up to about 0.3% while achieving the FRR of 0% at 0.35% FAR. However, these results may be less convincing as other approaches [2, 15]; our system is more reliable and robust, as far as attacks on secret key are concerned. The experimental results in BioHashing are dependent upon security of tokenized (pseudo)random number, as reported in [12] and have to put additional efforts to secure these numbers. In contrast, our emphasis is to strengthen the cryptographic approaches for

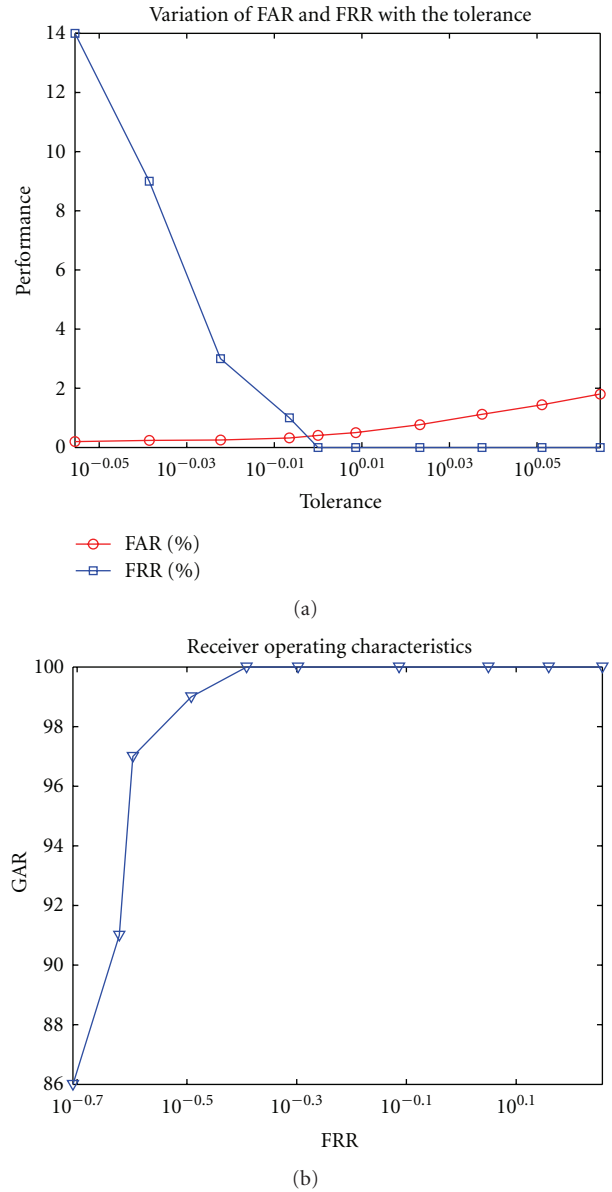


Figure 5: (a) The variations of the FAR and FRR characteristics with the tolerance for the palmprint-based cryptosystem, and (b) corresponding receiver operating characteristics.

encryption (the problems with symmetric and asymmetric approach have been discussed earlier in Sections 3.1.1-3.1.2) and withstand the attacks on secret key. Any secret document/information (of any length) can be encrypted by symmetric cryptographic approach (as symmetric approaches such as, DES, and AES are very efficient for the encryption of bulk data) and the secret symmetric key is again encrypted using asymmetric approach (to overcome dependency on secret symmetric key). Finally, the palmprint-based fuzzy vault is created around the private asymmetric key to prevent unauthorized disclosure of the key. At the decryption end, if the input palmprint template is able to open the vault (using matching criteria), the access to private key is granted. The rest is the conventional cryptographic mechanism as the

Table 2: Summary of related prior work.

Biometric	Feature	Error Correction Code	FRR (%)	FAR (%)	Reference	Database Size
fingerprint	Minutiae Points	RS Code	5	0	[42]	9 Users
Voice	Cepstrum coefficient	Discretization	20	NA*	[41]	10 Users
Signature	Dynamic time wrapping	Feature coding	28	1.2	[40]	25 Users
Iris	Gabor Feature	RS code and Hadamard Codes	0.47	0	[19]	70 Users
Fingerprint	Fourier transform	Majority code	12	35	[33]	20 Users
Fingerprint	Minutiae point	RS code	30	NA	[17]	NA*
Fingerprint	Minutiae points and helper data	RS code	3	0.24	[18]	100 Users (*FVC '02*)
Palmprint	DCT features	RS code	0	0.4	—	85

*NA—Not Available.

private key is used to decrypt symmetric key and finally the secret document. In fact, we propose a mixed cryptosystem which has advantage over both symmetric and asymmetric cryptography. The advantage of the proposed system lies in that it not only attempts to alleviate the shortcomings of symmetric key-based cryptosystem but also solves the problems involved in asymmetric key-based approach. The approach minifies dependency on secret key involved and alternatively investigates a more secure and promising system, as compared to BioHashing-based techniques.

Performance of the proposed system depends upon choice of tolerance chosen for grid of palmprint features. The increase in tolerance could lead to wrong positions in grid, and hence even the genuine user cannot open the vault, which can result in unacceptably high false rejection rate. The low tolerance value could diminish the fuzziness of grid which can cause the imposters to be accepted and hence increase in false acceptance rate. The optimal range for tolerance value is dependent on the range of palmprint features.

The main consideration is on the construction of palmprint-based fuzzy vault around the private key. The private and public keys are generated on publicly available RSA toolbox [26]. The bit length of modulus $m = k * l$, where m, k, and l are prime numbers (Section 3.1.5), is chosen as 1024 bits, and length of the encryption exponent n is 64 bit. The two large primes are chosen to be 512 bits, so that 1024 bit RSA modulus m can be generated. The RSA implementation has utilized the string format to generate the RSA keys and its length varies from 306 to 309 which is equivalent to 1015 to 1024 in binary bits. For the used RSA cryptosystem, the private key sc should be chosen such that it satisfies the following equation:

$$n * sc \equiv 1 \pmod{si}, \quad \text{where } 1 < sc < si. \tag{7}$$

It can be observed from the above equation that more than one value of n can satisfy the congruence, and hence the length of the generated string (key) can vary. The prime numbers are randomly chosen and so are the values of si and n, and therefore the variations in length of keys are not controlled. In our experiments we have observed and accounted for this variation. Our implementation stores the fixed length key and loads it at the time of generating grids to construct the vault. Therefore Table 1 illustrated all the possible variations in key length and the corresponding performance (EER) with the tolerance value. It can be observed from this table that as the key length varies (in the range 306 to 309), the system has different equal error rates at different tolerances. The minimum equal error rate is achieved when the key length is 307.

7. Conclusions

This paper has investigated a new approach to construct the cryptographic vault using palmprint features. In order to combine cryptography with palmprint features we have also incorporated the implementation of double encryption. This can efficiently reduce the possibility of hacking within a cryptosystem. The experimental results presented in Section 6 illustrate that the palmprint-based cryptosystem can operate at low EER (0.375%). The summary of the prior work, presented in Table 2, suggests that the palmprint can be used as a promising biometric in the construction of a cryptosystem. However, the work presented in Table 2 is not directly comparable; our motivation is to mere outline the effectiveness of the proposed work. The cryptosystem investigated in this paper employed localized spectral features from the palmprint. The multiple feature representation, such as detailed in [34], can offer more reliable characterization of features, and therefore cryptosystem based on multiple-palmprint representation can be considered for the extension of this work.

Acknowledgment

This work was partially supported by the research Grant from the Department of Science and Technology, Government of India (Grant no. 100/IFD/1275/2006-2007).

References

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

[2] A. Kong, D. Zhang, and M. Kamel, "Three measures for secure palmprint identification," *Pattern Recognition*, vol. 41, no. 4, pp. 1329–1337, 2008.

[3] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[4] W. Stallings, *Cryptology and Network Security: Principles and Practices*, Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2003.

[5] J. Nam, Y. Lee, S. Kim, and D. Won, "Security weakness in a three-party pairing-based protocol for password authenticated key exchange," *Information Sciences*, vol. 177, no. 6, pp. 1364–1375, 2007.

[6] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proceedings of the IEEE International Symposium on Information Theory*, p. 408, Lausanne, Switzerland, June-July 2002.

[7] U. Uludag, S. Pankanti, and A. K. Jain, "Fuzzy vault for fingerprints," in *Proceedings of the 5th Audio- and Video-Based Biometric Person Authentication (AVBPA '05)*, vol. 3546 of *Lecture Notes in Computer Science*, pp. 310–319, Springer, Hilton Rye Town, NY, USA, July 2005.

[8] G. I. Davida, Y. Frankel, and B. J. Matt, "On enabling secure applications through off-line biometric identification," in *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 148–157, Oakland, Calif, USA, May 1998.

[9] C. Sautar, D. Roberge, A. Stoianov, R. Gilroy, and B. V. K. Vijaya Kumar, "Biometric encryption," *Information Management and Computer Security*, vol. 9, no. 5, pp. 205–212, 2001.

[10] F. Monrose, M. K. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamics," in *Proceedings of the 6th ACM Conference on Computer and Communications Security*, pp. 73–82, 1999.

[11] T. Connie, A. Teoh, M. Goh, and D. Ngo, "PalmHashing: a novel approach for cancelable biometrics," *Information Processing Letters*, vol. 93, no. 1, pp. 1–5, 2005.

[12] A. Kong, K.-H. Cheung, D. Zhang, M. Kamel, and J. You, "An analysis of BioHashing and its variants," *Pattern Recognition*, vol. 39, no. 7, pp. 1359–1368, 2006.

[13] http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci213673,00.html.

[14] http://world.std.com/~franl/crypto/one-time-pad.html.

[15] X. Wu, D. Zhang, and K. Wang, "A palmprint cryptosystem," in *Proceedings of IAPR/IEEE International Conference on Biometrics (ICB '07)*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 1035–1042, August 2007.

[16] A. Juel and M. Wttenberg, "A fuzzy vault commitment scheme," in *Proceedings of the 6th ACM Conference on Computer and Communications Security*, G. Tsudik, Ed., pp. 408–412, 2002.

[17] U. Uludag and A. K. Jain, "Fuzzy fingerprint vault," in *Proceedings of Biometrics: Challenges Arising from Theory and Practice*, pp. 13–16, Cambridge, UK, August 2004.

[18] K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: implementation and performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 744–757, 2007.

[19] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081–1088, 2006.

[20] T. C. Calancy, N. Kiyavash, and D. J. Lin, "Secure smartcard-based fingerprint authentication," in *Proceedings of the ACM SIGMM Multimedia Workshop on Biometrics Methods and Applications*, pp. 45–52, Berkeley, Calif, USA, 2003.

[21] C.-H. Lin and Y.-Y. Lai, "A flexible biometrics remote user authentication scheme," *Computer Standards and Interfaces*, vol. 27, no. 1, pp. 19–23, 2004.

[22] A. Kumar and A. Kumar, "A palmprint-based cryptosystem using double encryption," in *Biometric Technology for Human Identification V*, vol. 6944 of *Proceedings of SPIE*, pp. 1–9, Orlando, Fla, USA, March 2008.

[23] RSA algorithm, http://www.di-mgt.com.au/rsa_alg.html.

[24] http://pajhome.org.uk/crypt/rsa/maths.html.

[25] K. A. Nishimura, S. J. Wenstrand, and G. Panotopoulos, "Biometric identification device," European patent EP1760667, July 2007.

[26] http://www.wipo.int/pctdb/en/wo.jsp?wo=2001092994&IA;=WO2001092994&DISPLAY;=DESC.

[27] P. W. Dent, "Cryptographic method and system for double encryption of messages," US patent no. 6904150, June 2005.

[28] M. J. Fischer, "Cryptography and computer security," Lecture Note-5 CPSC 467a, Department Of Computer Science, Yale University, http://zoo.cs.yale.edu/classes/cs467/2006f/attach/ln05.html.

[29] H. Ng, "Simple pseudorandom number generator with strengthened double encryption (Cilia)," http://eprint.iacr.org/2005/086.pdf.

[30] G. Immega, T. Vlaar, G. Vanderkooy, and K. Tucker, "Method for biometric encryption of email," European patent no. EP1290534, December 2003.

[31] Y.-H. Doh, J.-S. Yoon, K.-H. Choi, and M. S. Alam, "Optical security system for the protection of personal identification information," *Applied Optics*, vol. 44, no. 5, pp. 742–750, 2005.

[32] Z. Liu and S. Liu, "Double image encryption based on iterative fractional Fourier transform," *Optics Communications*, vol. 275, no. 2, pp. 324–329, 2007.

[33] A. Goh and D. C. L. Ngo, "Computation of cryptographic keys from face biometrics," in *Communications and Multimedia Security*, vol. 2828 of *Lecture Notes in Computer Science*, pp. 1–13, Springer, Berlin, Germany, 2003.

[34] A. Kumar and D. Zhang, "Personal authentication using multiple palmprint representation," *Pattern Recognition*, vol. 38, no. 10, pp. 1695–1704, 2005.

[35] A. Kong and D. Zhang, "Compititive coding scheme for palmprint verification," in *Proceedings of the International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 520–523, August 2004.

[36] A. Kumar, "Incorporating cohort information for reliable palmprint authentication," in *Proceedings of the 6th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '08)*, pp. 583–590, Bhubaneswar, India, December 2008.

[37] A. Nagar and S. Chaudhury, "Biometrics based asymmetric cryptosystem design using modified fuzzy vault scheme," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 4, pp. 537–540, Hong Kong, August 2006.

[38] A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain, "Personal authentication using hand images," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1478–1486, 2006.

[39] http://islab.oregonstate.edu/koc/ece575/02Project/Kie+Raj/.

[40] H. Feng and C. C. Wah, "Private key generation from on-line handwritten signatures," *Information Management and Computer Security*, vol. 10, no. 4, pp. 159–164, 2002.

[41] F. Monrose, M. K. Reiter, Q. Li, and S. Wetzel, "Cryptographic key generation from voice," in *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 202–213, May 2001.

[42] F. Monrose, M. K. Reiter, and R. Wetzel, "Password hardening based on keystroke dynamics," *International Journal of Information and Computer Security*, vol. 1, no. 2, pp. 69–83, 1999.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 964746, 12 pages
doi:10.1155/2009/964746

Research Article

A Novel Criterion for Writer Enrolment Based on a Time-Normalized Signature Sample Entropy Measure

Sonia Garcia-Salicetti, Nesma Houmani, and Bernadette Dorizzi

Department of EPH, Institut TELECOM, TELECOM & Management SudParis, 91011 Evry, France

Correspondence should be addressed to Nesma Houmani, nesma.houmani@it-sudparis.eu

Received 15 October 2008; Revised 8 March 2009; Accepted 9 June 2009

Recommended by Natalia A. Schmid

This paper proposes a novel criterion for an improved writer enrolment based on an entropy measure for online genuine signatures. As online signature is a temporal signal, we measure the time-normalized entropy of each genuine signature, namely, its average entropy per second. Entropy is computed locally, on portions of a genuine signature, based on local density estimation by a Client-Hidden Markov Model. The average time-normalized entropy computed on a set of genuine signatures allows then categorizing writers in an unsupervised way, using a K-Means algorithm. Linearly separable and visually coherent classes of writers are obtained on MCYT-100 database and on a subset of BioSecure DS2 containing 104 persons (DS2-104). These categories can be analyzed in terms of variability and complexity measures that we have defined in this work. Moreover, as each category can be associated with a signature prototype inherited from the K-Means procedure, we can generalize the writer categorization process on the large subset DS2-382 from the same DS2 database, containing 382 persons. Performance assessment shows that one category of signatures is significantly more reliable in the recognition phase, and given the fact that our categorization can be used online, we propose a novel criterion for enhanced writer enrolment.

1. Introduction

Handwritten signature is a behavioural biometric modality showing high variability from one instance to another of a same writer. This high variability explains indeed that the best verification approaches, as particularly reflected for Online Signature in the results of the First International Signature Verification Competition SVC2004 [1] and the Signature Evaluation carried out in the framework of BioSecure Multimodal Evaluation Campaign BMEC2007 [2], are those tolerating random local variations of the signature, as elastic matching techniques (Dynamic Time Warping [3–5] or statistical models, as Hidden Markov Models (HMM) [3, 6–13] and Gaussian Mixture Models (GMMs) [14, 15]. Nevertheless, the amount of this variability is writer dependent, in the sense that some writers have a signature by far more variable from one instance to the next than other writers.

An automatic signature verification system involves two steps: the enrolment step and the verification step. In order to provide a given level of security to an individual signer, writer enrolment must guarantee that enrolment signatures are stable and complex enough. Indeed, as studied in [16], when enrolling a writer, his/her signature will be acceptable as a reference signature, or as part of a reference set, for any verification system, only if it is complex enough. In [16], a "difficulty coefficient" estimates the difficulty to reproduce a given signature as a function of the rate of geometric modifications (length and direction of strokes) per unit of time, in other words as a function of complexity of the hand draw. Such study concludes that "problematic" signers in terms of performance of Automatic Verification Systems are those with signatures which have a low "difficulty coefficient" (not complex enough signatures).

On the other hand, when enrolling a writer, his/her signature will be suitable as reference or as part of a reference set for any verification system only if it is not too variable; in [16], enrolment signatures are selected by using a comparison algorithm that computes the spatiotemporal difference between two signatures (elastic matching). By this

way, a "dissimilarity index" is proposed to quantify intraclass variability between different signature samples of a same writer. In [17], a procedure relying on a correlation-based criterion detecting local distortions of the hand draw is proposed to select the reference signatures for a signature verification system. Such correlation criterion measures how much local stability values, computed on different signatures when being matched by elastic matching techniques, are correlated. Finally, the subset of signatures with highest correlation is selected as reference set. Alternatively, in [18], the stability criterion is based on the lowest intraclass Euclidean distance between feature vectors representing globally the candidate reference signatures. Finally, in [19], both complexity and variability criteria were proposed for offline signature verification by a human expert. A human operator labels signatures according to both criteria and their impact on performance is studied. Also in [20], signature analysis by means of fractal geometry led to the emergence of a complexity criterion to categorize writers.

All these works suggest the strong impact of complexity and variability criteria on the classifier performing signature verification. Indeed, stability is required in genuine signatures in order to be able to characterize a given writer, since the less stable a signature is, the more likely it is that a forgery gets dangerously close to genuine signatures in terms of the metric of any classifier. Also, complex enough signatures are required at the enrolment step to generate a certain level of security.

In this work, we propose to exploit for writer enrolment a time-normalized entropy measure that allows quantifying both the stability and the complexity of a writer's genuine signatures. This entropy, measured in bits per second, is computed on portions of the signature, and averaged over such portions. As the entropy of a random variable only depends on its probability density values, a good estimation of this probability density is important [21]. As in online signatures there are local time dependencies in the dynamics of the hand-draw, a local paradigm for density estimation is natural.

In the previous works [22, 23], we proposed to estimate the probability density of a writer's dynamics locally, by a Hidden Markov Model (HMM) trained on a set of ten genuine signatures, to extract a Personal Entropy measure globally from such set. In this work, we follow the same local paradigm, but we compute the time-normalized entropy of a signature sample "Sample Entropy", namely, the average entropy per second of such sample, therefore quantified in bits per second. It is worth noticing that the above mentioned Hidden Markov Model, whose complexity (topology) is related to the length of the genuine signatures of a writer, is only used in our work as a local refined density estimator, devoted to compute the time-normalized entropy of a signature sample, and not as a classifier.

Based on the "Sample Entropy", we then propose to generate for each writer a "Personal Entropy measure" value, by averaging the "Sample Entropy" associated to each of his/her genuine signatures. We show in this work that this measure allows categorizing writers in linearly separable and visually coherent categories, by a K-Means procedure [24]. Moreover, we related this categorisation to variability and complexity measures, this way showing quantitatively the link between our new Personal Entropy measure and some behavioural characteristics of the signature. Our previous performance assessment study [23], carried out only on random forgeries, with different classifiers, showed that system performance changes in function of the different writer categories. In this work, we first extend our performance assessment study to skilled forgeries and confirm this interesting result: there is one category of users, which can be detected by their Personal Entropy, and are "problematic", since their signatures are vulnerable because of their strong variability and low complexity. At the opposite, there is a category of "safer" signatures, highly complex and stable, that can also be detected by their associated writer's Personal Entropy. Our aim in this work is to exploit this entropy measure in order to enhance enrolment in the following ways.

(i) To inform the user of the intrinsic risk related to his/her signature.

(ii) In case of a "problematic" signature, to leave the possibility to the signer of choosing between deciding to pursue enrolment knowing the intrinsic risk of his/her signature, or alternatively to change his/her signature for security purposes.

(iii) To adjust the quality of enrolment data according to the level of security required by the application.

As previously mentioned, writer categories emerge from our entropy measure, by means of a K-means procedure. Given this fact, each writer category is naturally associated to a signature "prototype" or Entropy-Prototype (EP), which corresponds to the mean of the class. We propose in this work to exploit such Entropy-Prototypes to identify beforehand "problematic" signers. We show indeed that after having generated prototypes on a given reduced data set of 104 writers from the complete DS2 database [25], it is possible to generalize the writer categorization process on other writers belonging to the same database. Given the fact that our writer categorization process is totally automatic, independent of any classifier (it only relies on our proposed Personal Entropy measure), and besides can be generalized to new writers acquired in similar conditions (same digitizer, same acquisition protocol, similar sampling frequency, similar resolution), we propose a novel criterion for a better writer enrolment process targeting enhanced signature verification. Indeed, our writer categorization process gives as outputs one Entropy-Prototype per category, which combined to a Nearest Neighbour Rule [24], naturally allows classifying a signature sample during the enrolment step. This classification allows therefore measuring the intrinsic level of security of a user's signature at the enrolment step.

This paper is organized as follows. The next section describes how the "Sample Entropy" measure associated to each genuine signature sample is computed by means of a Writer-HMM, and the resulting "Personal Entropy" value of each writer. Also, we present the automatically generated categories of writers, obtained when performing a K-Means procedure on such "Personal Entropy measure" of each

writer, on a subset of the BioSecure Data Set 2 (DS2-104) and on MCYT-100 database, both captured on a digitizer. In order to give a quantitative interpretation of these categories, we have defined complexity and variability measures, and we have shown the strong relationship between our Personal-Entropy measure and both the complexity and the variability in signatures. Section 3 presents performance assessment across such writer categories, by means of two statistical classifiers of same complexity (number of parameters), namely, a Hidden Markov Model (HMM) and a Gaussian Mixture Model (GMM), on DS2-104 and MCYT-100 databases. Such statistical approaches gave indeed the best signature verification results in the last Signature Evaluation campaign in the framework of BioSecure Multimodal Evaluation Campaign BMEC'2007 [2]. Section 4 describes the generalization of the writer categorization process, relying on Entropy-Prototypes built on a subset of Data Set 2 (DS2-104) and evaluated on the large data set DS2-382 of 382 persons; the resulting global performance on DS2-382 are compared with performance on each category. Finally, the proposed enhanced writer enrolment procedure relying on Personal-Entropy is described in detail.

2. Time-Normalized Sample Entropy and Writer Categories

2.1. Measuring Time-Normalized Sample Entropy with a Hidden Markov Model. We consider in this work a signature as a sequence of two time functions, namely, its raw coordinates (x, y). Indeed, raw coordinates are the only time functions available on all sorts of databases, whether acquired on fixed platforms (as digitizing tablets) or on mobile platforms (as Personal Digital Assistants).

The entropy of a random variable only depends on its probability density values; therefore a good estimation of this probability density must be performed to compute reliably an entropy value. As the online signature is piecewise stationary, it is natural to estimate the probability density locally, namely, on portions of the signature. In this framework, Hidden Markov Models [3] (HMM) appear as a natural tool as they both allow performing a segmentation of the signature into portions and a local estimation of the probability density on each portion.

We thus consider each genuine signature of a given writer as a succession of portions, generated by its segmentation via such writer's Hidden Markov Model (HMM). Therefore, we obtain as many portions in each signature as there are states in the Writer-HMM. Then we consider each point (x, y) in a given portion S_i as the outcome of one random variable Z_i (see the top of Figure 1) that follows a given probability mass function $p_i(z) = \Pr(Z_i = z)$, where z belongs to the Alphabet A of ordered pairs (x, y). Such random variable associated to a given portion of the signature is discrete since its alphabet A has a finite number of values, thus its entropy in bits is defined as

$$H(Z_i) = -\sum_{z \in S_i} p(z) \cdot \log_2(p(z)). \quad (1)$$

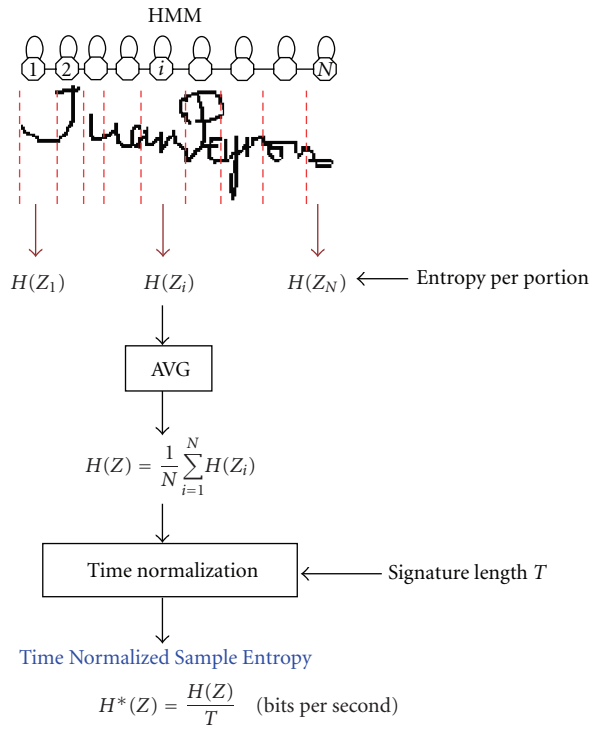


Figure 1: The Time-Normalized Sample Entropy computation.

Nevertheless, the hand-drawing as a time function is a continuous process from which we retrieve a sequence of discrete values via a digitizer. For this reason, although $Z = (x, y)$ is discrete, we take advantage of the continuous emission probability law estimated on each portion by the Writer-HMM. Such density function is modelled as a mixture of Gaussian components.

To compute the Time-Normalized Sample Entropy of a signature sample, we first train the Writer-HMM on 10 genuine signatures of such writer, after computing a personalized number of states, as follows:

$$N = \frac{T_{\text{Total}}}{M * 30}, \quad (2)$$

where T_{Total} is the total number of sampled points available in the genuine signatures, and $M = 4$ is the number of Gaussian components per state. We ensure this way that the number of sample points per state is at least 120, in order to obtain a good estimation of the Gaussian Mixture in each state (four Gaussian components).

Then we exploit the Writer-HMM to generate by the Viterbi algorithm [3] the portions on which the entropy is computed for each genuine signature. On each portion, we consider the probability density estimated by the Writer-HMM to compute locally this entropy. We then average the entropy over all the portions of a signature and normalize the result by the signing time of the signature sample (see Figure 1). This measure is a Time-Normalized Sample Entropy, expressed in bits per second. Our experiments show that in order to get a good estimation of Personal Entropy, it is necessary to have at least 10 signatures of each writer.

Averaging this measure across the 10 genuine signatures on which the local probability densities were estimated by the HMM allows generating a measure of Personal Time-Normalized Entropy, denoted "Personal Entropy" in the following of this paper. Time normalization allows comparing users between them in terms of entropy; indeed, without such time normalization, due to the great difference in length between signatures of different persons, entropy tends to be higher on longer signatures.

2.2. Databases Description. We used three databases in this work: the freely available and the widely used MCYT subset of 100 persons [26], and two subsets from the online signature database acquired in the framework of the BioSecure Network of Excellence [25]: DS2 (for Second Data Set of the whole data collection), acquired on a digitizer. The first subset DS2-104 contains data of 104 persons, and the second subset DS2-382 contains data of 382 persons. The whole BioSecure Signature Subcorpus DS2 [25], acquired on several sites in Europe, is the first online signature multisession database acquired in a digitizer.

DS2 contains data from 667 persons acquired in a PC-based offline supervised scenario and the digitizing tablet WACOM INTUOS 3 A6. The pen tablet resolution is 5080 lines per inch, and the precision is 0.25 mm. The maximum detection height is 13 mm, and the capture area is 270 mm (width) $\times$ 216 mm (height). Signatures are captured on paper using an inking pen. At each sampled point of the signature, the digitizer captures at 100 Hz sampling rate the pen coordinates, pen pressure (1024 pressure levels), and pen inclination angles (azimuth and altitude angles of the pen with respect to the tablet). This database contains two sessions, acquired two weeks apart, each containing 15 genuine signatures. The donor was asked to perform, alternatively, three times five genuine signatures and twice five forgeries. Indeed, for skilled forgeries, at each session, a donor is asked to imitate five times the signature of two other persons after several minutes of practice and with the knowledge of the signature dynamics.

2.3. Writer Categories with Personal Entropy Measure. We performed on the two databases described in Section 2.2 (DS2-104 and MCYT-100), containing around 100 persons, a K-Means procedure [24] on Personal Entropy values for different values of K. We reached a good separation of signatures with $K = 3$ on both databases, as shown in Figure 2 for some signatures in DS2, whose owners authorized their publication.

Figure 3 shows that the obtained three categories are actually linearly separable, as represented by indicative lines reporting the automatic classification results given by the K-Means procedure.

As mentioned before, time normalization allows comparing users between them in terms of entropy since there is a great difference in length between signatures of different persons.

We notice that on the two databases, the first category of signatures, those having the highest Personal Entropy (Figure 2(a)), contains short simply drawn and not legible signatures, often with the shape of a simple flourish. At the opposite, signatures in the third category, those of lowest Personal Entropy (Figure 2(c)), are the longest and their appearance is rather that of handwriting, some being even legible. In between, we notice that signatures with medium Personal Entropy (second category, Figure 2(b)) are longer than those of the first category, often showing the aspect of a complex flourish.

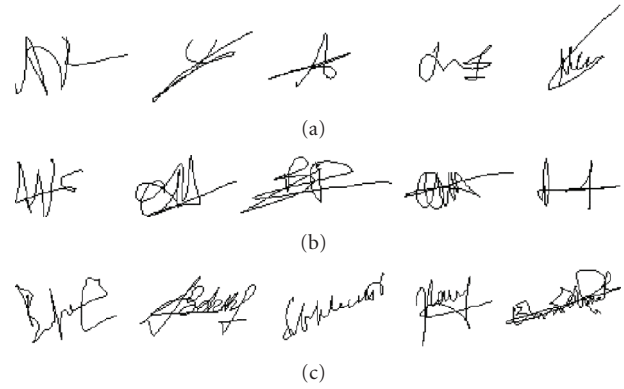


Figure 2: Examples of signatures from DS2-104 of (a) high, (b) medium, and (c) low Personal Entropy (with authorization of the writers).

Categories of signatures seem at this step visually related to complexity and variability criteria. We therefore propose quantitative measures of complexity and variability, with which we will analyze the obtained Entropy-based categories.

2.4. Relation between Our Personal Entropy and Complexity and Variability Measures. In order to measure complexity, we consider a vector of seven components related to the shape of handwriting: numbers of local extrema in both x and y directions, changes of pen direction in both x and y directions, cusps points, crossing points, and "star points" [27]. We consider the Euclidean norm of the vector as the indicator of complexity for each signature. We then average such measure on the 10 genuine signatures in order to generate a complexity measure for a given person.

In order to measure the variability of a client's signature, we use Dynamic Time Warping [3], which relies on a local paradigm to quantify distortions. We compute the distances between all the possible couples of genuine signatures (45 as we consider 10 genuine signatures) and average the obtained distances to get the indicator of signature variability. Four features are extracted locally per point: absolute speed, the angle between the absolute speed vector and the horizontal axis, curvature radius of the signature, and the length to width ratio on a sliding window of size 5.

Figure 4 shows Personal Entropy versus Complexity and Variability indicators, per category on DS2-104 and MCYT-100. We see that signatures of highest Personal Entropy are highly variable and of rather low complexity. At the opposite, signatures of lowest Personal Entropy are by far more complex and more stable (show low variability). We noticed

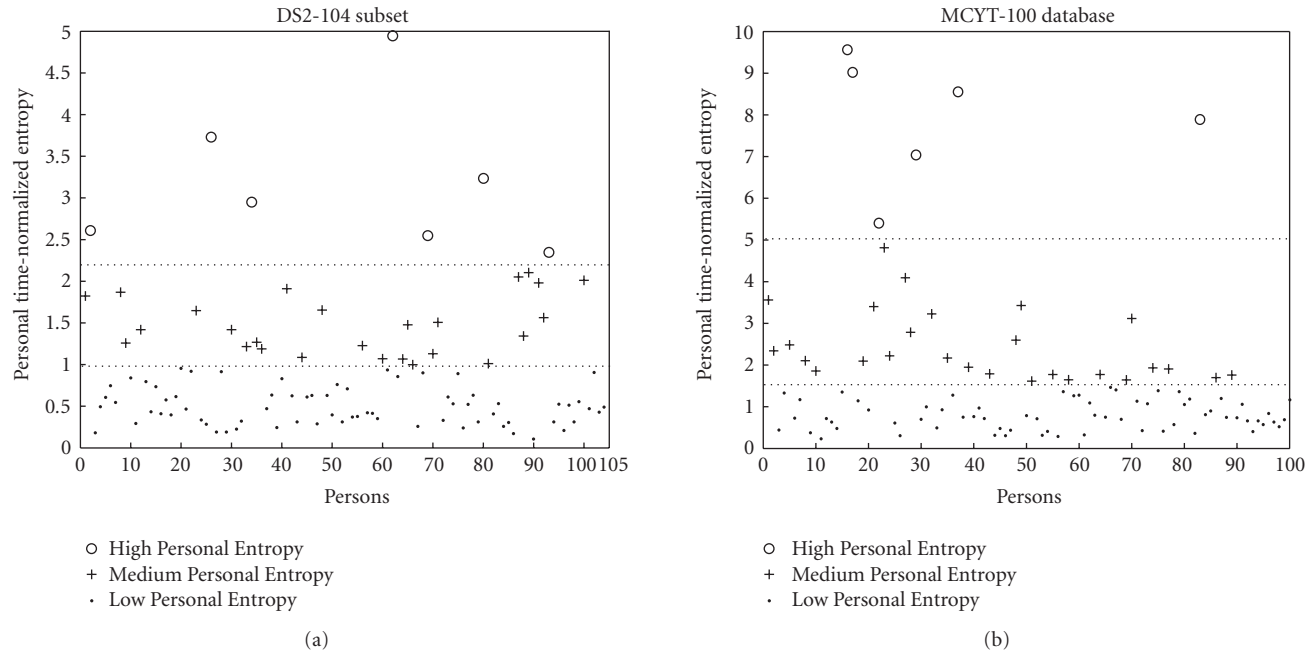


Figure 3: Personal Entropy values on data from DS2-104 (a) and from MCYT-100 (b) across the 3 writer categories. Two indicative lines report the separation between categories obtained by the K-Means procedure.

that this behaviour is verified for all the databases considered in this work. We therefore conclude that our Personal Entropy measure allows quantifying both the complexity and variability of a writer's signatures simultaneously.

3. Verification Performance

In this section, we study the relationship between Personal Entropy-based categories and performance of two different automatic signature verification systems, on two different databases: DS2-104 and MCYT-100.

3.1. Score Computation by the Two Classifiers. Two classifiers are used in this study considering only the raw coordinates description of signatures as input data: a Hidden Markov Model [3] and a Gaussian Mixture Model [14].

For performance assessment, both skilled and random forgeries are considered. Ten random samplings are carried out on genuine and impostor signatures in the following way: each sampling contains five genuine signatures used as the training set for both statistical classifiers. For test purposes, the remaining 25 genuine signatures and 20 skilled forgeries (belonging to two sessions) are used for DS2-104. For MCYT-100, we tested on the remaining 20 genuine signatures and 25 skilled forgeries. Also, 30 impostor signatures randomly sampled in equal number in each Personal Entropy category (10 random forgeries per category) are considered for both databases. The False Acceptance and False Rejection Rates are computed relying on the total number of False Rejections and False Acceptances obtained on the whole ten random samplings.

Concerning the topology of the two statistical models, we used a GMM and a left-to-right HMM of the same complexity in terms of Gaussian components. It is worth noticing that the HMM classifier differs from the HMM used for Personal Entropy computation. Indeed, the former is devoted to classification, while the latter only performs local density estimation. We considered for the HMM classifier a 6 states and 4 Gaussian components per state, as a tradeoff in complexity between the signatures of the two extreme categories. For the GMM, accordingly, we considered 24 Gaussians to model a person's signatures. The dissimilarity matching score for both statistical models is

$$\text{Score} = |\text{LL} - \text{LL}_{BA}|, \tag{3}$$

where LL is the Log-Likelihood of the test signature (normalized by the length of the test signature), and LL_{BA} is the corresponding average Log-Likelihood of the training signatures.

3.2. Performance Assessment on DS2-104 and MCYT-100 with the Two Classifiers. In our experiments, both HMM and GMM classifiers were intentionally not optimized, since our aim is not to improve absolute system performance but to analyze the relative differences in classifiers' performance between writer categories.

We notice on Figures 5 and 6 corresponding to DS2-104, and on Figures 7 and 8 corresponding to MCYT-100, that the results lead to different behaviours in terms of performance according to the category of Personal Entropy that we consider.

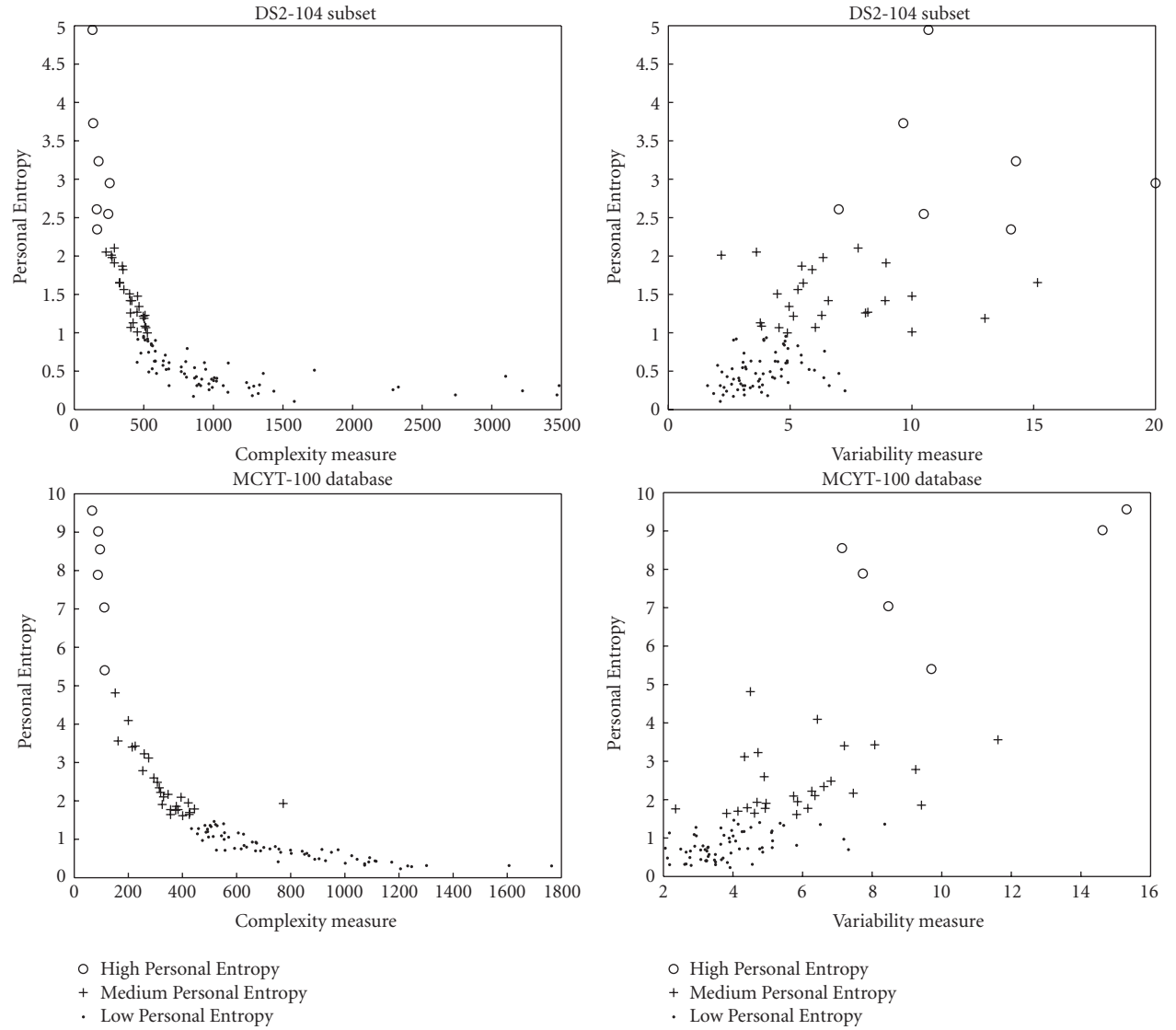


Figure 4: Personal Entropy versus complexity (left) and Personal Entropy versus variability (right) on MCYT-100 and DS2-104 databases, for Personal Entropy-based categories.

There is a significant difference in classifiers' performance between the two extreme categories, for both skilled and random forgeries: GMM and HMM classifiers give the best performance on writers belonging to the category of lowest Personal Entropy, that is, those having the longest most complex and most stable signatures, as those shown in Figure 2(c). At the opposite, HMM and GMM classifiers give the worst performance on writers belonging to the highest Personal Entropy, those having the shortest simplest and most unstable signatures, as those shown in Figure 2(a). We also notice that performance values for the category of writers with medium Personal Entropy are in between those of the two extreme writer categories.

As shown in Tables 1 and 2, for the two classifiers, at the Equal Error Rate functioning point, performance is roughly improved by a factor around 2 for skilled and random forgeries when switching from the highest entropy category to the lowest one, on both DS2-104 and MCYT-100. Confidence Intervals at 95% are given to show the significance of results. At other functioning points, this gap in performance between the two extreme categories is maintained for the two classifiers, as shown in Figures 5, 6, 7, and 8.

For a better insight on the impact of high and medium Personal Entropy categories on system performance, we ordered, in a decreasing way, users from such categories according to their Personal Entropy. Then, we compute when removing the top x% of such users, the relative improvement $\Delta(x)$ of the Equal Error Rate with regard to the average EER on the whole DS2-104 database (denoted by $\overline{\text{EER}}$) defined as follows:

$$\Delta(x) = \frac{\overline{\text{EER}} - \text{EER}(x)}{\overline{\text{EER}}}, \tag{4}$$

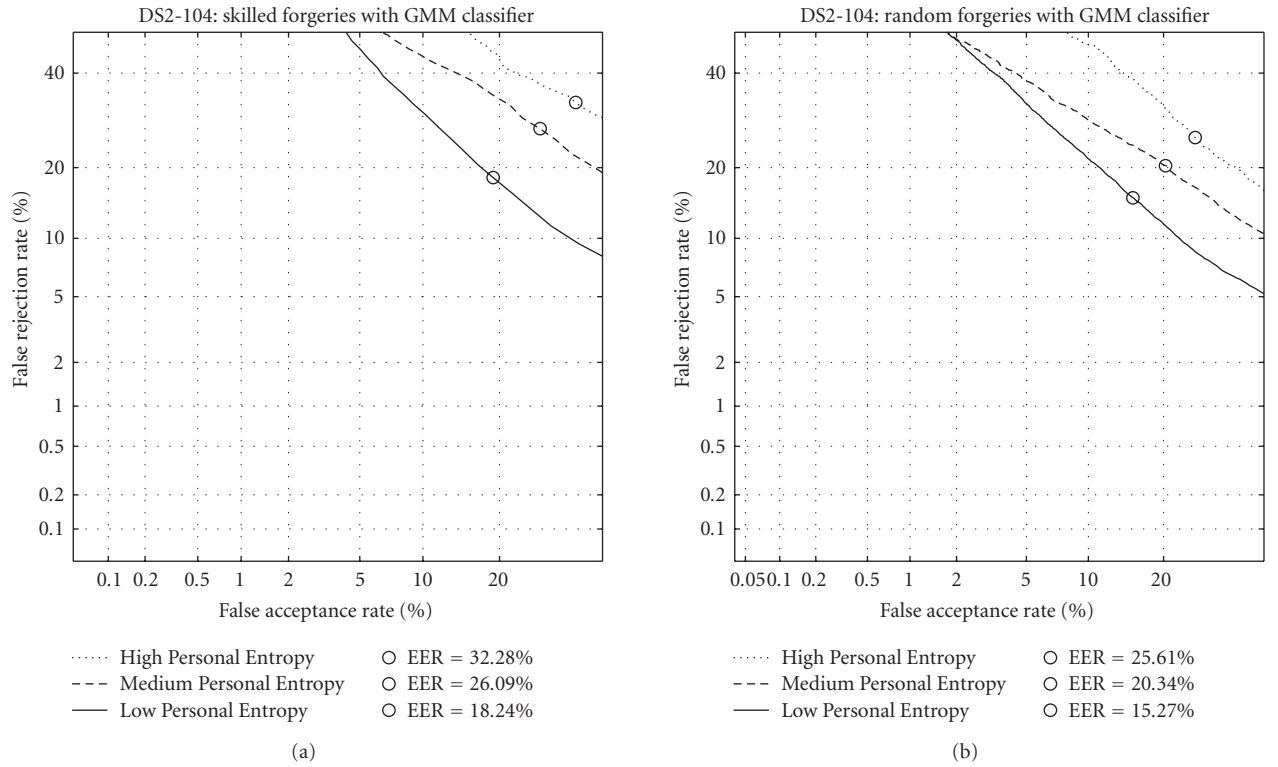


Figure 5: DET-curves considering skilled forgeries (a) and random forgeries (b), on each writer category on DS2-104 subset with the GMM classifier.

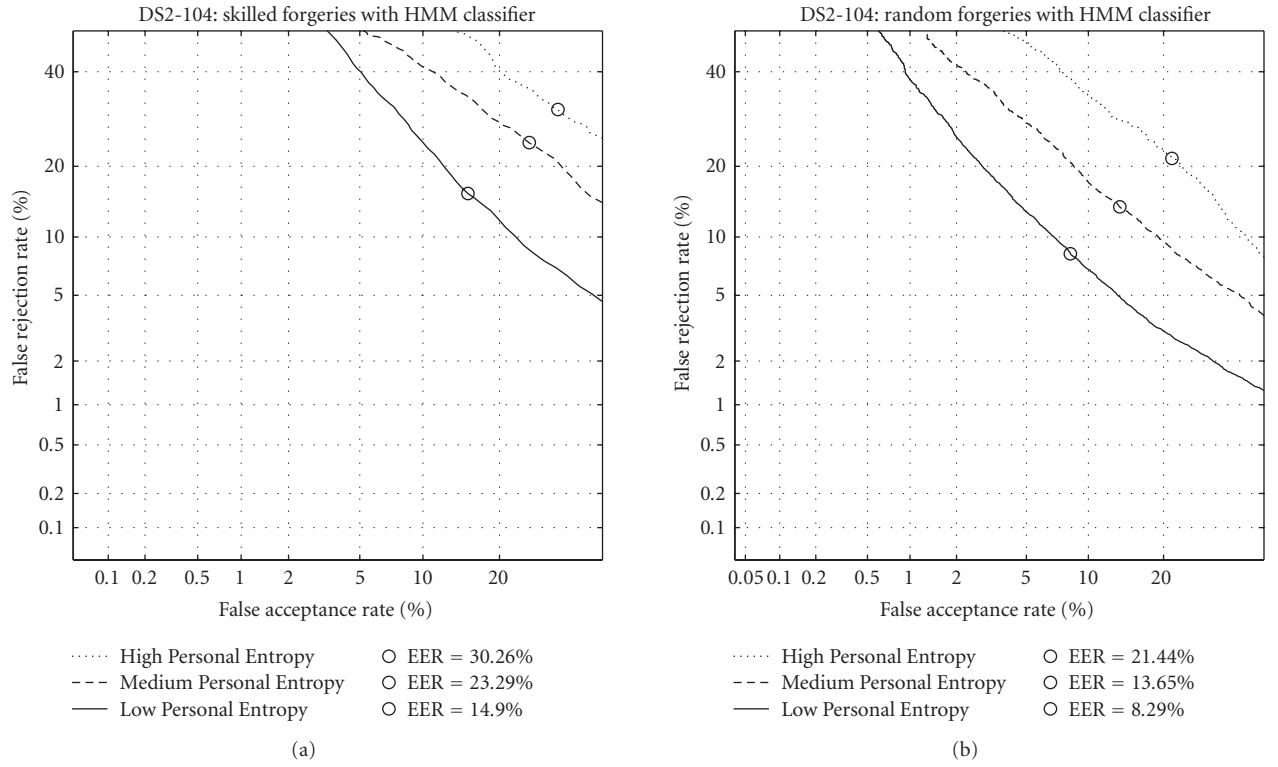


Figure 6: DET-curves considering skilled forgeries (a) and random forgeries (b), on each category on DS2-104 subset with the HMM classifier.

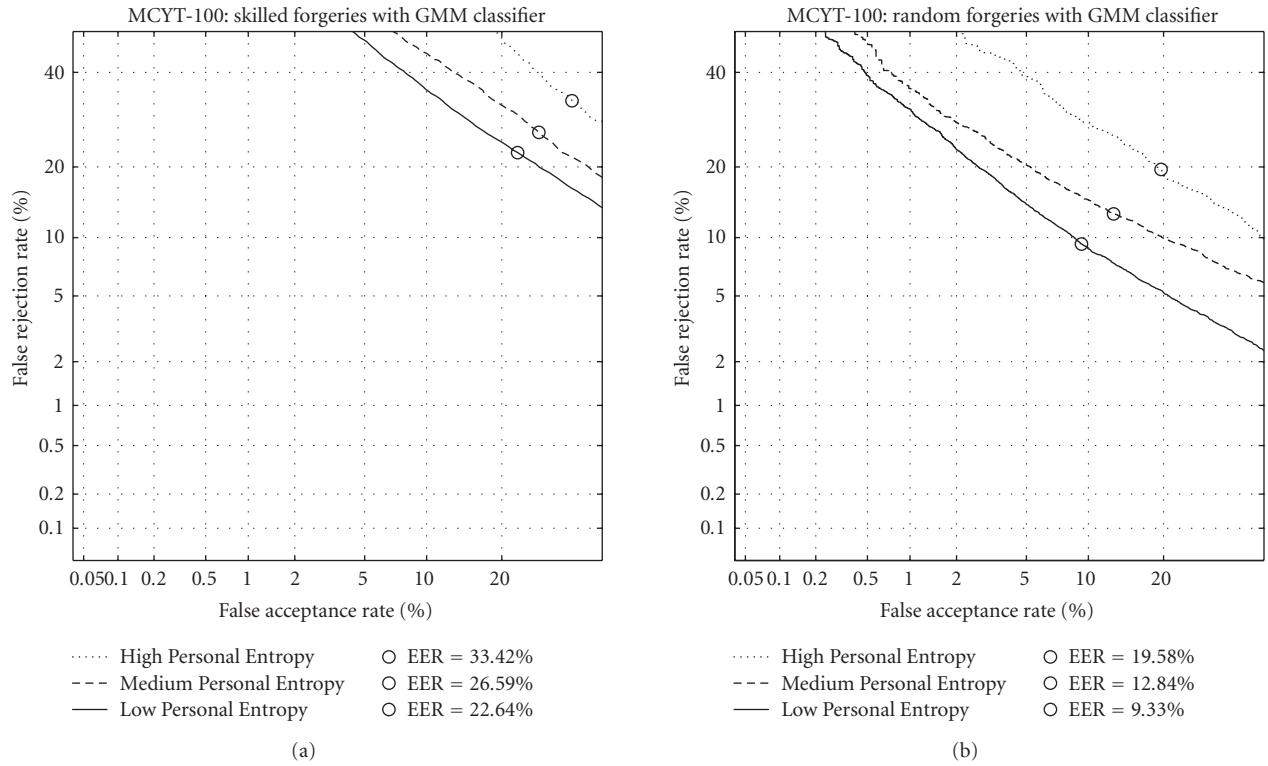


Figure 7: DET-curves considering skilled forgeries (a) and random forgeries (b), on each writer category on MCYT-100 database with the GMM classifier.

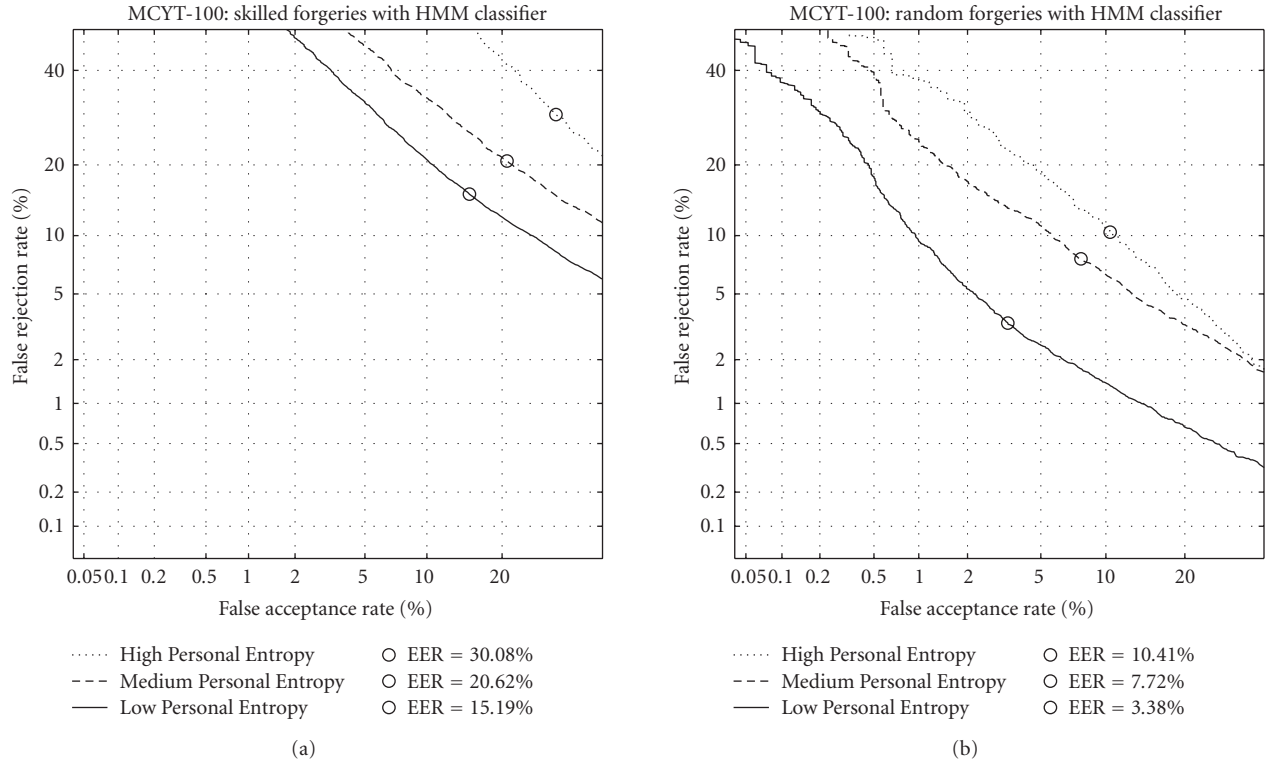


Figure 8: DET-curves considering skilled forgeries (a) and random forgeries (b), on each category on MCYT-100 database with the HMM classifier.

Table 1: Equal Error Rate and Confidence Interval in each writer category on DS2-104 subset, with HMM and GMM classifiers considering skilled and random forgeries.

	DS2-104 subset							
	GMM classifier				HMM classifier			
	Skilled forgeries		Random forgeries		Skilled forgeries		Random forgeries	
	EER (%)	CI (95%)	EER (%)	CI (95%)	EER (%)	CI (95%)	EER (%)	CI (95%)
High entropy	**32.28**	±0.100	**25.61**	±0.057	**30.26**	±0.100	**21.44**	±0.080
Medium entropy	26.09	±0.040	20.34	±0.026	23.29	±0.027	13.65	±0.018
Low entropy	**18.24**	±0.010	**15.27**	±0.007	**14.90**	±0.009	**8.29**	±0.001

Table 2: Equal Error Rate and Confidence Interval in each writer category on MCYT-100 database, with HMM and GMM classifiers considering skilled and random forgeries.

	MCYT-100 database							
	GMM classifier				HMM classifier			
	Skilled forgeries		Random forgeries		Skilled forgeries		Random forgeries	
	EER (%)	CI (95%)	EER (%)	CI (95%)	EER (%)	CI (95%)	EER (%)	CI (95%)
High entropy	**33.42**	±0.170	**19.58**	±0.160	**30.08**	±0.200	**10.76**	±0.120
Medium entropy	26.59	±0.050	12.84	±0.028	20.62	±0.042	7.56	±0.023
Low entropy	**22.64**	±0.018	**9.33**	±0.006	**15.74**	±0.010	**4.13**	±0.003

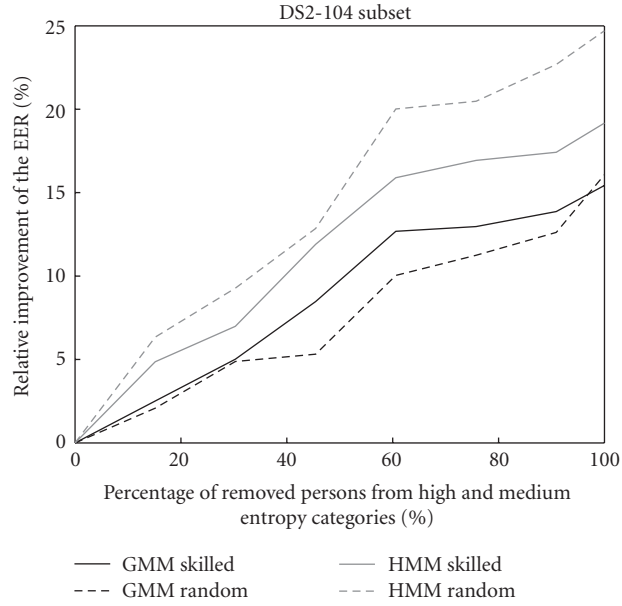


Figure 9: The relative improvement $\Delta(x)$ of the average EER on DS2-104 subset when removing x% of users from high and medium Personal Entropy categories.

where EER(x) represents the average Equal Error Rate on the whole DS2-104 database after removing x% of users from high and medium Personal Entropy categories.

We notice in Figure 9 that for both the GMM and HMM classifiers, and both random and skilled forgeries, when removing progressively an increasing percentage x of users from high and medium Personal Entropy categories (according to their Personal Entropy measure), $\Delta(x)$ increases. When all users from high and medium Personal Entropy categories are removed ($x = 100\%$), this relative improvement $\Delta(x)$ reaches in all cases more than 15%, as reported in detail in Table 3. Moreover, given that the first 21% of users belong to the high Personal Entropy category (7 users), and the remaining 79% belong to the medium Personal Entropy category (26 users), we conclude that the main improvement is obtained when the first 60% of users are removed (that is all users from the high Personal Entropy category and 50% of users from the medium Personal Entropy category).

Table 3: The relative improvement $\Delta(x)$ of the average EER on DS2-104 subset when removing all users from high and medium Personal Entropy categories.

Classifier	Type of forgeries	$\overline{\text{EER}}$	Δ (100%)
GMM	Skilled forgeries	21.57%	15.43%
	Random forgeries	18.19%	16.06%
HMM	Skilled forgeries	18.43%	19.16%
	Random forgeries	10.90%	23.94%

4. Generalizing Writer Categorization

4.1. On Categorizing New Writers Relying on Entropy-Prototypes Obtained Offline. We have this far shown that there is one category of users which are much easier to recognize than others, and much easier to discriminate from skilled and random forgeries, those having a low Personal Entropy value. Alternatively, there is another category of users which are extremely difficult to recognize, those having a high Personal Entropy value.

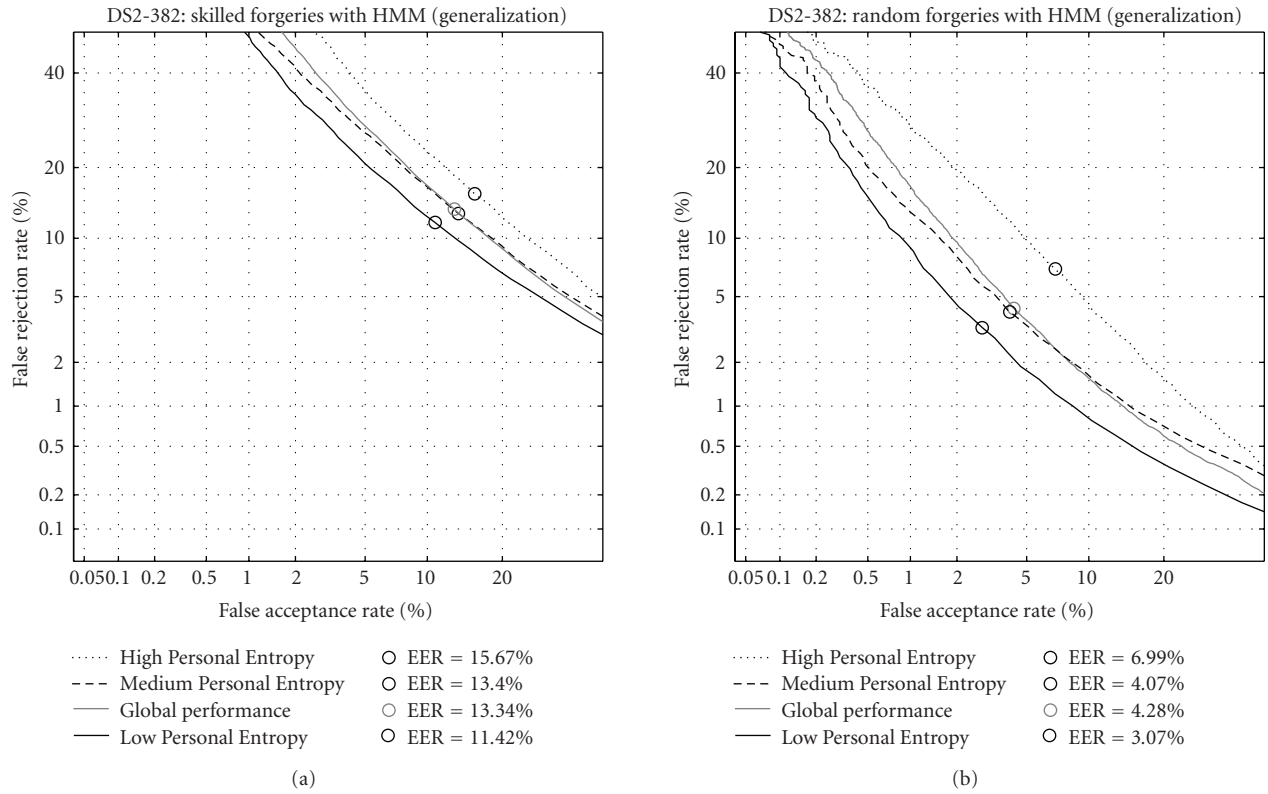


Figure 10: DET-curves considering skilled forgeries (a) and random forgeries (b), on each writer category and globally on DS2-382 database with the HMM classifier, after computing entropy-prototypes on DS2-104.

Table 4: Equal Error Rate and Confidence Interval in each writer category on DS2-382 database, with the HMM classifier considering skilled and random forgeries.

	DS2-382 with HMM classifier			
	Skilled forgeries		Random forgeries	
	EER (%)	CI (95%)	EER (%)	CI (95%)
High entropy	**15.67**	±0.025	**6.99**	±0.015
Medium entropy	13.4	±0.012	4.07	±0.006
Low entropy	**11.42**	±0.014	**3.07**	±0.005
Global performance	13.34	±0.003	4.28	±0.003

Each writer category is naturally associated to an Entropy-Prototype (EP) inherited from the K-Means procedure used to "cluster" writers. Our aim in this section is to study the possibility of categorizing new writers based on previously generated Entropy-Prototypes (EPs), on a data set of limited size. We carry out this study by generating three Entropy-Prototypes on DS2-104, and using such prototypes to categorize writers from another data set: DS2-382.

Indeed, we categorize a writer belonging to such data set as follows:

(1) computing the writer's Personal Entropy with 10 genuine signatures of such writer from DS2-382;

(2) retrieving the three Entropy-Prototypes (one per category) computed offline on DS2-104 database;

(3) associating to such writer from DS2-382 the category of closest Entropy-Prototype by the Nearest Neighbor Rule [24].

In order to study the relevance of the previous protocol, we study performance on the obtained categories after generalization. In order to carry out this study, we only consider in the following an HMM classifier, since the same results are obtained with a GMM classifier.

4.2. Generalization on the Same Database from DS2-104 to DS2-382. Figure 10 and Table 4 show the performance obtained on DS2-382 with an HMM classifier on each of the obtained categories after computing Entropy-Prototypes on DS2-104, with skilled and random forgeries respectively. We also compare results per category to global results on the complete DS2-382 database.

As on DS2-104, on which the Entropy-Prototypes have been computed originally, we notice that also on DS2-382 there is a difference in classifiers' performance between the two extreme categories, for both skilled and random forgeries: the HMM classifier gives the best performance on writers belonging to the category of lowest Personal Entropy. At the opposite, the HMM classifier gives the worst performance on writers belonging to the highest Personal Entropy.

As shown in Table 4, at the Equal Error Rate functioning point, performance is roughly improved by a factor 2 for skilled forgeries and 1.4 for random forgeries when switching from the highest entropy category to the lowest one. We also notice that performance values for the category of writers with medium Personal Entropy are in between those of the two extreme writer categories.

Moreover, the global performance on the whole data set DS2-382 is degraded compared to performance on the category of writers with lowest Personal Entropy.

4.3. Our Proposed Criterion for Writer Enrolment. We have shown that Entropy-Prototypes generated offline on a database of limited size (104 persons) can be used to perform writer categorization on new writers from the same database. We thus propose to exploit such Entropy-Prototypes, which are totally independent of the verification system, to identify beforehand signatures that are not secure in terms of performance. The enrolment procedure that we propose has the following steps.

(1) Ten genuine signatures are requested from the writer to be enrolled.

(2) A Writer-HMM is built for such writer by training the HMM on such ten genuine signatures.

(3) The Personal Entropy of such writer is computed.

(4) The three Entropy-Prototypes computed offline are retrieved.

(5) The category of the closest Entropy-Prototype by the Nearest Neighbor Rule [24] is associated to the writer.

(6) When a writer is classified as belonging to the highest Personal Entropy category, he/she should be informed of the intrinsic risk related to his/her signature. Indeed, this category of writers gives unreliable results relatively to other Personal Entropy categories; we thus propose to the user either to pursue enrolment knowing the intrinsic risk of his/her signature, or alternatively to change his/her signature for security purposes.

(7) When a writer belongs to the category of lowest Personal Entropy, the writer is enrolled.

(8) When a writer belongs to the category of middle Personal Entropy, we recommend to the writer to do a more complex and less variable signature, but still can retain his/her signature.

Based on our experiments, we can assert that the more Personal Entropy lowers, the more reliable is the signature in terms of security. This should be to take into account when using online signature in practical applications.

5. Conclusion

We have proposed a novel criterion for writer enrolment that allows guaranteeing a higher level of security to the individual writer, regardless of the verification system that is used. Such criterion relies on an unsupervised automatic writer categorization process, carried out on a Time-Normalized Personal Entropy measure, quantified in bits per second. We first introduce in this work a "Sample Entropy" measure associated to each enrolment signature sample, computed locally by means of a Writer-HMM trained on ten enrolment signatures. Then we explain how the resulting "Time-Normalized Personal Entropy" value of each writer is retrieved.

We show that a writer can be categorized according to this measure and to Entropy-Prototypes computed offline, into one of three categories of writers. This categorization process is crucial because verification systems' performance is significantly different between the extreme categories of highest and lowest Personal Entropy. Indeed, we show across two data sets that our Personal Entropy measure allows classifying writers automatically into three visually coherent and linearly separable categories, opposing long, complex and stable signatures to short, strongly variable and simple signatures. Moreover, we have quantified the behaviour of the signature in terms of complexity and variability, and we have linked these values to our Personal Entropy measure.

We have shown that Entropy-Prototypes, naturally inherited from the K-Means procedure and performed to generate writer categories, can be generated offline on a data set of limited size (around 100 persons) and be used to perform writer categorization on new writers of another data subset, providing the same acquisition conditions. More generally, a database of roughly 100 persons is enough to generate the categories, then allowing the online categorization of any new user whose signatures are acquired in similar conditions (same digitizer, similar tablet resolution, and same acquisition protocol).

Based on this result, we propose an enrolment writer criterion related to such Entropy-Prototypes, totally independent of the verification system, to identify beforehand signatures which are not secure in terms of performance. Indeed, a Nearest Neighbour Rule on Entropy-Prototypes generated offline, on a database of roughly 100 persons, allows categorizing a writer after requesting from him ten instances of his/her signature. A stable and reliable result emerges of our study: the more Personal Entropy lowers, the more reliable is the signature in terms of security. This statement allows adapting the quality of the enrolment data to the level of security requested by the application.

Acknowledgment

The authors thank Javier Ortega-Garcia and his colleagues for putting at disposal the subset of the first 100 users of MCYT Signature Subcorpus.

References

[1] D.-Y. Yeung, H. Chang, Y. Xiong, et al., "SVC2004: first international signature verification competition," in *Proceedings of International Conference on Biometric Authentication (ICBA '04)*, vol. 3072 of *Lecture Notes in Computer Science*, pp. 16–22, Springer, Hong Kong, July 2004.

[2] http://biometrics.it-sudparis.eu/BMEC2007.

[3] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Signal Processing Series, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[4] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern Recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[5] A. Kholmatov and B. Yanikoglu, "Identity authentication using improved online signature verification method," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2400–2408, 2005.

[6] J. G. A. Dolfing, *Handwriting recognition and verification, a Hidden Markov approach*, Ph.D. thesis, Philips Electronics N. V., 1998.

[7] R. Kashi, J. Hu, W. L. Nelson, and W. Turin, "A Hidden Markov Model approach to online handwritten signature verification," *International Journal on Document Analysis and Recognition*, vol. 1, pp. 102–109, 1998.

[8] G. Rigoll and A. Kosmala, "A systematic comparison of on-line and off-line methods for signature verification with Hidden Markov Models," in *Proceedings of the 14th International Conference on Pattern Recognition (ICPR '98)*, pp. 1755–1757, Brisbane, Autralia, August 1998.

[9] J. Ortega-Garcia, J. Gonzalez-Rodriguez, D. Simon-Zorita, and S. Cruz-Llanas, "From biometrics technology to applications regarding face, voice, signature and fingerprint recognition systems," in *Biometrics Solutions for Authentication in an E-World*, D. Zhang, Ed., pp. 289–337, Kluwer Academic Publishers, Dordrecht, The Netherlands, July 2002.

[10] J. G. A. Dolfing, E. H. L. Aarts, and J. J. G. M. Van Oosterhout, "On-line signature verification with hidden Markov models," in *Proceedings of the International Conference on Pattern Recognition*, pp. 1309–1312, Brisbane Australia, 1998.

[11] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni, "An on-line signature verification system based on fusion of local and global information," in *Proceedings of the 5th IAPR International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '05)*, Lecture Notes in Computer Science, Springer, 2005.

[12] D. Muramatsu and T. Matsumoto, "An HMM on-line signature verifier incorporating signature trajectories," in *Proceeding of the 7th International Conference on Document Analysis and Recognition (ICDAR '03)*, IEEE, Edinburgh, Scotland, August 2003.

[13] B. Ly Van, S. Garcia-Salicetti, and B. Dorizzi, "On using the Viterbi path along with HMM likelihood information for online signature verification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 37, no. 5, pp. 1237–1247, 2007.

[14] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.

[15] J. Richiardi and A. Drygajlo, "Gaussian mixture models for on-line signature verification," in *Proceedings of the ACM SIGMM Workshop on Multimedia Biometrics Methods and Applications (WBMA '03)*, pp. 115–122, Berkley, Calif, USA, November 2003.

[16] J.-J. Brault and R. Plamondon, "How to detect problematic signers for automatic signature verification," in *Proceedings of the International Canadian Conference on Security Technology (ICCST '89)*, pp. 127–132, Zurich, Switzerland, 1989.

[17] V. Di Lecce, G. Di Mauro, A. Guerriero, et al., "Selection of reference signatures for automatic signature verification," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR '99)*, pp. 597–600, Bangalore, India, 1999.

[18] C. Allgrove and M. C. Fairhusrt, "Enrolment model stability in static signature verification," in *Proceedings of International Workshop on Frontiers in Handwriting Recognition (IWFHR '00)*, pp. 565–570, Amsterdam, The Netherlands, 2000.

[19] F. Alonso-Fernandez, M. C. Fairhurst, J. Fierrez, and J. Ortega-Garcia, "Impact of signature legibility and signature type in off-line signature verification," in *Proceedings of the IEEE Biometrics Symposium (BSYM '07)*, Baltimore, Md, USA, September 2007.

[20] V. Boulétreau, *Towards a handwriting classification by fractal methods*, Ph.D. dissertation, Institut National des Sciences Appliquées de Lyon, Lyon, France, 1997.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2006.

[22] S. Garcia-Salicetti, N. Houmani, and B. Dorizzi, "A client-entropy measure for on-line signatures," in *Proceedings of the IEEE Biometrics Symposium (BSYM '08)*, pp. 83–88, Tampa, Fla, USA, September 2008.

[23] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi, "A novel personal entropy measure confronted with online signature verification systems' performance," in *Proceedings of the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '08)*, Washington, DC, USA, September 2008.

[24] R. O. Duda and P. E. Hart, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2000.

[25] http://www.biosecure.info.

[26] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, et al., "MCYT baseline corpus: a bimodal biometric database," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[27] V. S. Chakravarthy and B. Kompella, "The shape of handwritten characters," *Pattern Recognition Letters*, vol. 24, no. 12, pp. 1901–1913, 2003.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 260516, 13 pages
doi:10.1155/2009/260516

Research Article

Online Signature Verification Using Fourier Descriptors

Berrin Yanikoglu[1] and Alisher Kholmatov[2]

[1] *Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Turkey*
[2] *National Research Institute of Electronics and Cryptology (UEKAE), Scientific and Technological Research Council of Turkey (TUBITAK), Gebze, Kocaeli 41470, Turkey*

Correspondence should be addressed to Berrin Yanikoglu, berrin@sabanciuniv.edu

Received 27 October 2008; Revised 25 March 2009; Accepted 25 July 2009

Recommended by Natalia A. Schmid

We present a novel online signature verification system based on the Fast Fourier Transform. The advantage of using the Fourier domain is the ability to compactly represent an online signature using a fixed number of coefficients. The fixed-length representation leads to fast matching algorithms and is essential in certain applications. The challenge on the other hand is to find the right preprocessing steps and matching algorithm for this representation. We report on the effectiveness of the proposed method, along with the effects of individual preprocessing and normalization steps, based on comprehensive tests over two public signature databases. We also propose to use the pen-up duration information in identifying forgeries. The best results obtained on the SUSIG-Visual subcorpus and the MCYT-100 database are 6.2% and 12.1% error rate on skilled forgeries, respectively. The fusion of the proposed system with our state-of-the-art Dynamic Time Warping (DTW) system lowers the error rate of the DTW system by up to about 25%. While the current error rates are higher than state-of-the-art results for these databases, as an approach using global features, the system possesses many advantages. Considering also the suggested improvements, the FFT system shows promise both as a stand-alone system and especially in combination with approaches that are based on local features.

1. Introduction

Signature verification is the task of authenticating a person based on his/her signature. Online (dynamic) signatures are signed on pressure sensitive tablets that capture dynamic properties of a signature in addition to its shape, while offline (static) signatures consist of only the shape information. Dynamic features, such as the coordinates and the pen pressure at each point along the signature's trajectory, make online signatures more unique and more difficult to forge compared to offline signatures.

In online signature verification systems, like in any other biometric verification system, users are first enrolled to the system by providing reference samples. Later, when a user presents a signature claiming to be a particular individual, the query signature is compared with the reference signatures of the claimed individual. If the dissimilarity is above a certain fixed threshold, the user is rejected.

As a behavioral biometric, online signatures typically show more intrapersonal variations compared to physical biometrics (e.g., iris, fingerprint). Furthermore, forging a signature may be relatively easy if the signature is simple and its timing can be guessed from its static image (e.g., short signature showing a strictly left to right progression). Despite these shortcomings, signature is a well-accepted biometric and has potential niche applications such as identity verification during credit card purchases. Also, forging the shape and timing at the same time proves to be difficult in reality, as evidenced by the success of automatic verification algorithms [1].

In this work, we present an online signature verification system based on the spectral analysis of the signature using the Fast Fourier Transform (FFT). The advantage of using the Fourier domain is the ability to compactly represent an online signature using a fixed number of coefficients, which leads to fast matching algorithms. More importantly, the fixed-length is better suited or even necessary in certain applications related to information theory and biometric cryptosystems. For instance, the template protection scheme

by Tuyls et al. [2] requires a fixed-length feature representation of the biometric signal. Similarly, an earlier version of the proposed system was used for assessing the individuality of online signatures, where the fixed-length representation was important for simplifying the analysis [3]. Approaches using global and local features are called feature-based and function-based in literature, respectively. In this work, we also refer to them shortly as global and local approaches.

The challenge of using the Fourier domain representation, on the other hand, is to find the right preprocessing steps and matching algorithm for this representation. We report on the effectiveness of the proposed method, along with the effects of individual preprocessing and normalization steps, on the overall system performance, based on comprehensive tests over two public signature databases. While the current error rates are higher than state-of-the-art results for the used databases, this is to be expected since approaches based on global features of the signature normally underperform those using local information. On the other hand, in addition to the aforementioned advantages, global approaches are good complements to local approaches such as Dynamic Time Warping (DTW) or Hidden Markov Models (HMMs). In fact, we show that the fusion of the proposed system improves the performance of our DTW system by up to about 25%. With regard to the preprocessing, we show that the proposed incorporation of the pen-up durations significantly improves verification performance, while subsampling which is commonly used to obtain equal-length signatures, has the opposite effect. Finally, we discuss potential improvements and conclude that the proposed system has potential both as a stand-alone system and especially in combination with approaches that are based on local features.

This paper is organized as follows. Section 2 describes the previous work in the general area of online signature verification problem, along with some specific work that are more closely related to ours. Section 3 describes the proposed method, including preprocessing, feature extraction, and matching steps. Sections 4 and 5 present and discuss the experimental results using the SUSIG and MCYT databases. Finally Section 6 mentions future work to improve the current performance.

2. Previous Work

Signature verification systems differ both in their feature selection and in their decision methodologies. In fact, more than 70 different feature types have been used for signature verification [4–7]. These features can be classified in two types: global and local. Global features are those related to the signature as a whole, including the signature bounding box dimensions, average signing speed, and signing duration. Fourier Descriptors studied in this work are also examples of global features. Genuine signatures of a person often differ in length due to the natural variations in signing speed. The advantage of global features is that there are a fixed number of measurements (features) per signature, regardless of the signature length; this makes the comparison of two signatures a relatively straightforward task. The fixed-length representation is also better suited or even necessary in certain applications. Xu et al. use the Fourier transform to obtain a fixed-length representation of fingerprint minutiae [8]. Similarly, Yi et al. use the phase information of the Gabor filter to align online signatures and use the temporal shift and the shape dissimilarity measures to represent online signatures using a fixed-length feature vector [9].

In contrast to global features, local features are measured or extracted at each point along the trajectory of the signature and thus vary in number even among genuine signatures. Examples of local features include position, speed, curvature, and pressure at each point on the signature trajectory. In [5, 10], some of these features are compared in order to find the more robust ones for signature verification purposes. When local features are used, one needs to use methods which are suitable to compare feature vectors of different lengths: for instance, the Dynamic Time Warping algorithm [4, 5, 11–13] or Hidden Markov Models [14–19]. These methods are more complicated compared to the relatively simple metrics used with global features but they are generally more successful as well. Methods using global and local features are called feature-based and function-based approaches in literature [7]. Comprehensive surveys of the research on signature verification, including a recent one, can be found in [20–22].

The performance of biometric verification systems are evaluated in terms of false reject rate (FRR) of genuine samples, false accept rate (FAR) of impostors, and equal error rate (EER), where the two types of errors are equal. Due to the differences in databases and forgery qualities, comparing reported performance results is difficult. The First International Signature Verification Competition (SVC2004), organized in 2004, provided a common test set and tested more than 15 online signature verification systems from industry and academia. The results of this competition indicate state-of-the-art results of 2.6% equal error rate in skilled forgery detection and 1.85% equal error rate in random forgery detection tasks, using only position sequence (x, y) of a signature [1]. Our DTW-based system using only positional information, later described in [13], was declared as the winning system (Team 6) for its performance in the skilled forgery tests. We will refer to this system as our DTW system from now on.

Many different features and matching algorithms have been used to compare two signatures but the use of the Fourier Transform has not been widely explored [23–25]. In the work by Lam et al. [23], the signature is first resampled to a fixed-length vector of 1024 complex numbers consisting of the x- and y-coordinates of the points on the signature trajectory. This complex signal then undergoes various preprocessing steps, some of which are suggested by Sato and Kogure [24], including normalization for duration, drift, rotation, and translation, prior to the application of the Fast Fourier Transform (FFT). Feature extraction involves calculating the Fourier Descriptors of the normalized signature and selecting the 15 Fourier Descriptors with the highest magnitudes, normalized by sample variances. Discriminant analysis is then used with the real and imaginary parts of

the 15 selected harmonics, to find the most useful features and their weights. The proposed system was tested using a very small signature dataset (8 genuine signatures of the same user and 152 forgeries provided by 19 forgers), achieving a 0% FRR and 2.5% FAR. In a similar work, Quan et al. [25] use windowed FFT to avoid the discontinuities in the signal, also using discriminant analysis to pick the important FFT coefficients. The authors show that windowing improves performance, resulting in an EER of 7% EER on the MCYT-100 database, using 15 reference signatures.

Similar to the Fourier transform, the Discrete Wavelet Transform (DWT) is recently used for online signature verification by Nanni and Lumini [26]. The results of this system on the MCYT-100 database are 11.5% equal error rate on skilled forgeries, when using only the coordinate information (x- and y-coordinates as a function of time) of the signature. The DWT is also used by Nakanishi et al. [27], with about 4% EER on a small private database.

Recent research on signature verification has concentrated on the fusion of multiple experts [7, 26, 28]. These systems typically combine new methods with proven ones such as DTW and HMMs (e.g., [13, 19] which received the first and second place in the SVC2004 competition). Fusion systems have some of the best results obtained for their respective databases; this is not very surprising because online signature is a complex signal of several dimensions and one method may concentrate on one aspect of the signal (e.g., shape), while another method may focus on another (e.g., timing).

In this paper, we present a novel online signature verification system based on the Fast Fourier Transform. Our work differs from previous work using Fourier analysis [23–25] in preprocessing and normalization steps as well as the matching algorithm. Furthermore, the results of the proposed algorithm and the individual preprocessing steps are comprehensively tested on two large, public databases. The results show the potential of the proposed system and also highlight the importance of the timing information for online signatures, in contrast to previous work where the timing information was discarded to a large extent [23–25].

3. Proposed Method

3.1. Input Signal. An online signature S, collected using a pressure-sensitive tablet, can be represented as a time sequence:

$$S(n) = \begin{bmatrix} x(n) & y(n) & p(n) & t(n) \end{bmatrix}^{\mathrm{T}} \quad (1)$$

for $n = 1, 2, \ldots, N$, where N is the number of points sampled along the signature's trajectory; $x(n)$ and $y(n)$ denote the coordinates of the points on the signature trajectory, while $p(n)$ and $t(n)$ indicate the pen pressure and timestamp, at sample point n. A pressure-sensitive tablet typically samples 100 points in a second (100 HZ) and captures samples only during the interaction of the pen tip with the tablet. Depending on the tablet capabilities, pen azimuth ($az(n)$) and pen altitude ($al(n)$), indicating the angle of the pen with respect to the writing surface, can also be collected. Other features such as local velocity and acceleration may be calculated using the above features, as done by many signature verification systems [5–7, 12, 14].

The positional information consisting of $x(n)$ and $y(n)$ is important because it describes the shape of the signature and it is common to all tablets. The pressure information, on the other hand, had not seem very useful in some previous studies [10, 13, 26], while others have found it useful [29]. In particular, our DTW system [13] using just the positional information achieved the lowest error rates in the skilled forgery tasks of SVC2004, including the task where pressure, azimuth, and altitude were available to participating systems [1]. On the other hand, Muramatsu and Matsumoto [29] tested the discriminative power of the component signals of an online signature both alone and in groups and achieved 10.4% EER when they included the pressure and azimuth information, compared to 12.7% without them, using the SVC2004 database. In the current work, we have also observed that the pressure, azimuth, and altitude information improves the performance, although not drastically. In addition, we propose to use the timestamp information to identify and use the pen-up periods in identifying forgeries.

In the remainder of the paper, we use the sequence index n as if it refers to time (see Section 3.2.1) and describe the methodology concentrating on the positional information, denoted as $s(t)$, while the other input components are used as available.

3.2. Preprocessing. Preprocessing of online signatures is commonly done to remove variations that are thought to be irrelevant to the verification performance. Resampling, size, and rotation normalization are among the common preprocessing steps. While useful in object recognition, our previous research [13] had suggested that preprocessing may decrease biometric authentication performance by removing individual characteristics of the user. Therefore, we keep the amount of preprocessing done to a minimum, preserving as much of the discriminatory biometric information as possible.

In the previous work on online signature verification using FFT [23–25], the signature undergoes various preprocessing steps, consisting of spike and minor element removal to remove noise and extraneous segments; adding ligatures to connect consecutive strokes to reduce discontinuities that would affect FFT results; equi-time subsampling to obtain a fixed-length signature; drift removal; and rotation, translation, and scale normalization. In [23], the effects of drift removal and ligature processing are analyzed and authors report that drift removal significantly improves verification performance, while ligature processing only brings a marginal improvement. They guess that ligature processing that is done to reduce discontinuities is not very helpful because the high-frequency components affected by the discontinuities are discarded in the matching process.

We tested the individual effects of the preprocessing steps found to be important in [23], using two large databases. The results described in Section 4.4 show that subsampling which

is commonly done to normalize the length of a signature significantly reduces verification performance by removing most of the timing information. This was also confirmed in our previous research. On the other hand, mean and drift removal are found to be useful, while scale removal is not needed since our features (Fourier Descriptors) are normalized to be invariant to translation, rotation, and scale changes.

In addition to the steps described above, we propose to use the timestamp information to identify and use the pen-up periods in identifying forgeries. The next sections describe the preprocessing steps used in this work.

3.2.1. Pen-up Durations. Pen-up periods indicate the times when the pen is not in contact with the tablet. These periods may be detected using discontinuities between the timestamps of consecutive points ($t(n)$ and $t(n+1)$) and actual pen-up durations can be calculated using the sampling rate of the tablet and the difference between timestamps.

Forgery signatures often have longer pauses between strokes, compared to genuine signatures, which may help in identifying forgeries. Thus, while the pen-up durations can be useful for verification, such as in detecting a forger's hesitation or recomposition, it is often discarded, keeping just the order of the sampled points. In fact, the timing information is discarded to a large extent by many systems that use resampling to obtain a fixed-length signature, including the previous work using FFT [23–25]. Note that resampling results in keeping only the relative order of the points on the trajectory, while other timing information is discarded.

We propose to fill the pen-up durations with imaginary points, which has a twofold benefit: (i) it incorporates pen-up durations directly into the signature trajectory; (ii) it reduces trajectory discontinuities, which enhances the FFT analysis. For example, if there is a 50 ms wait between two consecutive points of the trajectory using a 100 Hz tablet (corresponding to 10 ms between consecutive samples), we add 4 imaginary points. Imaginary points can be generated through (a) interpolation between the last and first points of the two strokes corresponding to the pen-up event or (b) as if the pen was actually left on the tablet after the stroke prior to the pen-up event. In order for the pen-up events not to dominate the signal, we place imaginary points sparingly (every 30 ms for the 100 Hz tablet). Both methods of adding imaginary points improve the system performance, though the more sophisticated method of interpolation obtains better results, as expected.

Note that after this process, the timestamp information ($t(n)$) itself is basically redundant and discarded. We use the sequence index n and time t interchangeably in the rest of the paper.

3.2.2. Drift and Mean Removal. In signatures that go from left to right, $x(t)$ has a significant drift as time increases and the same can be said for signatures being signed top to bottom and $y(t)$. Drift removal step aims to remove the baseline drift component of a signal, so as to keep only the important information in the signal. We use a linear regression using least squares fit to estimate the drift. Given a discrete time signal y of length n, the drift removed version y' can be computed as

$$y' = y - \beta \times (t - \bar{t}), \tag{2}$$

where

$$\beta = \frac{\Sigma yt - n\bar{y}\bar{t}}{\Sigma t^2 - n\bar{t}^2}. \tag{3}$$

Mean removal on the other hand is simply achieved by subtracting the mean of the signal from itself: $y' = y - \bar{y}$.

3.3. Feature Extraction. We use the Fourier Transform to analyze the spectral content of an online signature. The details of the Fourier transform are out of the scope of this paper but can be found in many references (e.g., [30]). Below we give the basic idea and necessary definitions.

3.3.1. Fourier Transform. Any periodic function can be expressed as a series of sinusoids of varying amplitudes, called the *Fourier Series*. If the signal is periodic with fundamental frequency ω, the frequencies of the sinusoids that compose the signal are integer multiples of ω and are called the *harmonics*. The Fourier Transform is used to find the amplitude of each of the harmonic component, which is called the *frequency spectrum* of the signal. It thus converts a signal from the time domain into the frequency domain.

The Discrete Fourier Transform discrete time signal $f(t)$ is defined as follows:

$$C_k = \frac{1}{N}\sum_{t=0}^{N-1} f(t)e^{-i2\pi kt/N} \quad k = 0, 1, \ldots, N-1, \tag{4}$$

where $f(t)$ is the input signal; N is the number of points in the signature; k indicates the frequency of the particular harmonic; $e^{ix} = \cos(x) + i\sin(x)$.

The amplitude of the kth harmonic found by the Fourier transform is referred to as the kth *Fourier Coefficient*. Given a complex Fourier coefficient $C_k = a_k + ib_k$, the magnitude and phase corresponding to the kth harmonic are given by $|C_k| = \sqrt{a_k^2 + b_k^2}$ and $\tan^{-1}(b_k/a_k)$, respectively.

The Fourier coefficients are normalized to obtain the *Fourier Descriptors* which are the features used in this study, as described in Section 3.3.3.

The Inverse Fourier Transform is similarly defined as

$$f(t) = \sum_{k=0}^{N-1} C_k e^{i2\pi kt/N} \quad t = 0, 1, \ldots, N-1. \tag{5}$$

The Fourier transform has many uses in signal processing. For instance, reconstructing a time signal using the inverse Fourier transform by discarding the high-frequency components of a signal can be done for noise removal.

3.3.2. Input Signal Components. An online signature consisting of x- and y-coordinates can be represented as a complex

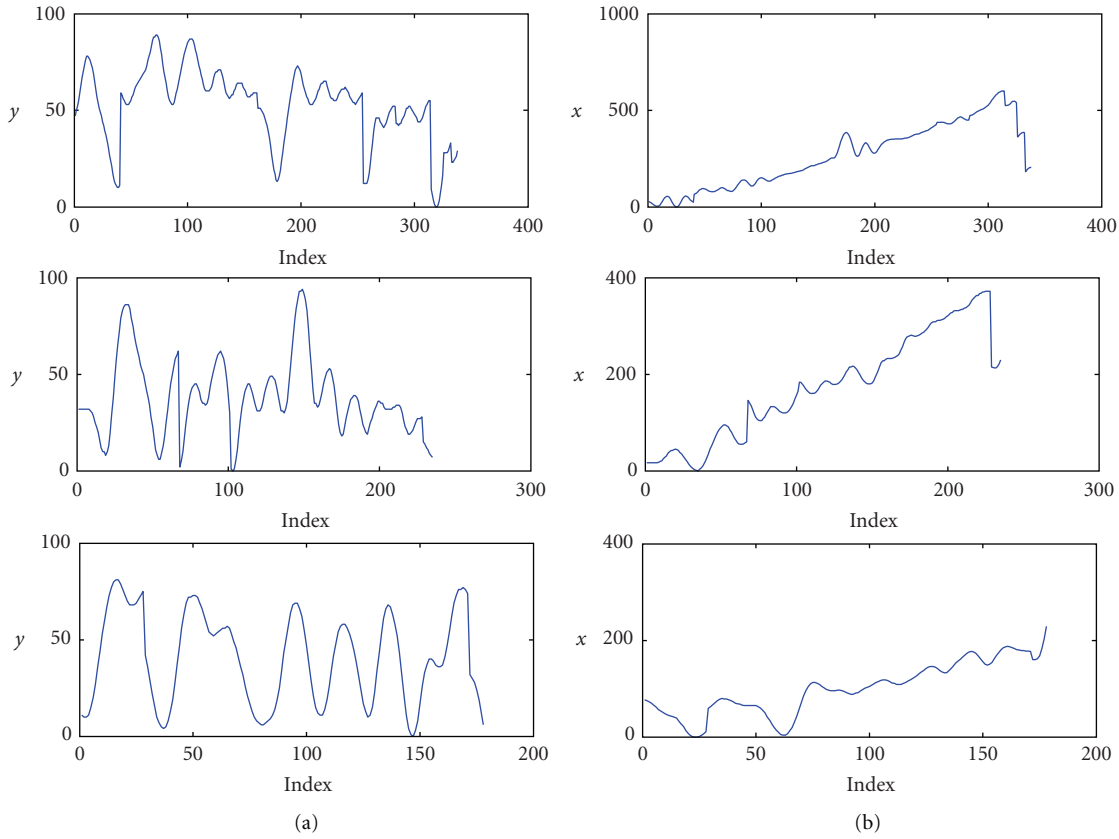


Figure 1: The y-coordinate (a) and x-coordinate (b) profiles belonging to genuine signatures of 3 different subjects from the SUSIG database.

signal $s(t) = x(t) + iy(t)$ where $x(t)$ and $y(t)$ are the x- and y-coordinates of the sampled points. The Fourier transform of the signature trajectory can then be directly computed using the complex signal $s(t)$ as the input, as described in (4).

In signatures which are signed from left to right or right to left, $x(t)$ is a monotonic function for the most part and carries little information, as shown in Figure 1. Based on this observation, we first evaluated the discriminative power of $y(t)$ alone, discarding $x(t)$ for simplicity. Later, we also did the reverse and used only $x(t)$ for completeness. Similarly, we assessed the contribution of other input signal components to the verification performance, by concatenating features extracted from individual component signals (e.g., $x(t)$, $y(t)$, $p(t)$), to obtain the final feature vector. We denote these feature vectors by indicating the individual source signals used in feature extraction: for instance, $x \mid y \mid p$ denotes a feature vector obtained from the x-, y-coordinates and pressure component, respectively. The input signal $f(t)$ in (4) can be any one of these signals ($s(t)$, $y(t)$, $x(t)$, $p(t)$, etc.).

3.3.3. Fourier Descriptors. The extracted Fourier coefficients are normalized to obtain the Fourier Descriptors, using normalization steps similar to the ones used in 2D shape recognition. In particular, the Fourier coefficients obtained by applying the Fourier Transform to the object contour $(x(t), y(t))$ can be normalized to achieve invariance against translation, rotation, and scaling of the original shape [30]. Specifically, translation of a shape corresponds to adding a constant term to each point of the original shape and affects (only) the first Fourier coefficient. By discarding C_0, defined in (4), one obtains translation invariance in the remaining coefficients. Rotation of a shape results in a phase change in each of the Fourier coefficients; rotation invariance is automatically obtained when one uses only the magnitude information of the Fourier Transform. Alternatively, each coefficient can be normalized such that the phase of one of the coefficients (e.g., C_1) is zero; this is equivalent to assuming a canonical rotation that gives a zero phase to C_1. Finally, scaling of a shape corresponds to multiplying all coordinate values of the shape by a constant factor and results in each of the Fourier coefficients being multiplied by the same factor. Therefore, scale normalization is achieved by dividing each coefficient by the magnitude of one of the components, typically $|C_1|$.

An online signature must show adequate match to the reference signatures of the claimed identity in *both* shape and dynamic properties, in order to be accepted. As with the above normalization steps, it is easy to see that by discarding C_0 and using the magnitudes of the remaining coefficients as features, we obtain invariance to translation (position of the signature on the tablet) and rotation (orientation relative to the tablet). Scale invariance is more complicated, due to the additional dimension of time. If a signature is only

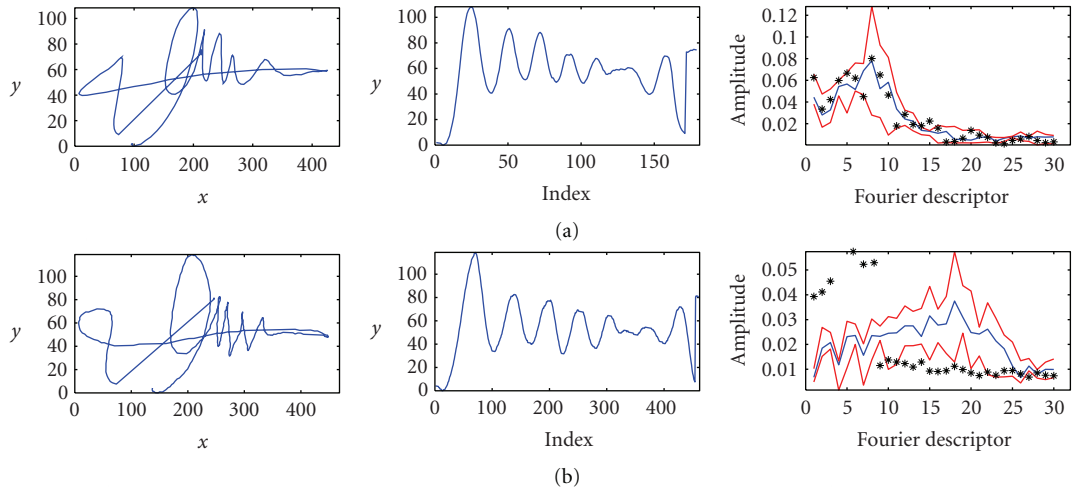


Figure 2: A verification case is shown for illustration, using only the y-profile. From left to right: (a) Genuine signature, its y-profile and its Fourier Descriptors. (b) Forgery signature, its y-profile and its Fourier Descriptors. The Fourier Descriptors of genuine and forgery signatures (shown as dots) are overlaid on top of the envelope showing the min and max values of the reference signatures' descriptors, while the line in the middle denotes the mean reference feature.

scaled in space, while keeping the signing duration the same, dividing each coefficient's magnitude by $|C_1|$ achieves scale normalization. However for the more general case involving both scale and time variations, we have found that a more robust approach is to divide each coefficient by the total magnitude of the Fourier spectrum:

$$m = \sum_{k=0}^{N-1} |C_k| = \sum_{k=0}^{N-1} \sqrt{C_k * C_k^*}, \tag{6}$$

where N is the length of the signature; $|C_k|$ is the magnitude of the complex coefficient C_k; C_k^* is the complex conjugate of C_k.

The total energy of the Fourier spectrum is also commonly used for normalization of the Fourier coefficients:

$$e = \sum_{k=0}^{N-1} |C_k|^2. \tag{7}$$

In our experiments, we have found that the normalization by the total amplitude has outperformed normalization done either by dividing each component by $|C_1|$ or by the total energy of the Fourier Transform (about 3% and 1% percent points less error, resp.).

Using (7), our final features or the Fourier Descriptors F_k are thus obtained as

$$F_k = \frac{|C_k|}{m} \quad k = 1, \ldots, \frac{N}{2}. \tag{8}$$

Notice here that k goes from 1 to $N/2$ since we discard half of the coefficients due to the symmetry of the Fourier transform spectrum.

3.3.4. Zero-Padding. Due to the natural variation in the signing process, genuine signatures of the same user almost never have equal lengths. The length variation results in Fourier domain representation with varying number of components, hence feature vectors of varying lengths. While one can cut out the high-frequency components, leaving only the first k Fourier coefficients, when the signatures are of different lengths, these components do not correspond to the same frequencies.

In order to obtain an equal number of Fourier Descriptors which correspond to the same frequencies, we pad each signature to be compared (reference set + query) with zeros, to match the length of the longest signature in the set, prior to the application of the Fourier Transform. This process is called *zero-padding* and does not affect the amplitudes of the Fourier coefficients but changes the frequency resolution.

3.3.5. Smoothing. We smooth the computed Fourier descriptors F_k by averaging two consecutive descriptors, to account for the normal timing variations between genuine signatures that would result in energy seeping into the neighboring harmonics. The smoothing is found to have a significant effect (roughly 2% point) in overall system performance in both tested databases.

Sample signatures and their forgeries, along with the resultant Fourier descriptors, are shown in Figure 2, using only the y-dimension for simplicity. The figure shows the envelope of the reference set descriptors to indicate the difference between query and reference signature descriptors, while in matching we only use the distance to the mean. The difference in the Fourier descriptors of the reference signatures for the genuine and forgery queries is due to zero-padding used in this example. As explained before, zero-padding does not change the frequency content of a signal but increases the frequency resolution (here note that the forgery signature that is used in determining the padding amount is much longer than the references).

Table 1: The summarizing characteristics of the public databases used in this study. In both of them, the genuine signatures are collected in multiple sessions and there are 5 reference signatures per user.

Dataset	Subjects	Genuine	Skilled forgeries	Input
SUSIG-Visual	100	2000	1000	x, y, p, timestamp
MCYT-100	100	2500	2500	x, y, p, az, al

Table 2: Equal error rates obtained using different components of the input signal. The timestamp is discarded after incorporating the pen-up durations into the trajectory, for the SUSIG database.

Dataset	$x+iy$	y	x	$x \mid y$	$x \mid y \mid p$	$x \mid y \mid p \mid az$	$x \mid y \mid p \mid az \mid al$
SUSIG-Visual	8.37%	9.90%	8.42%	**6.20**%	—	—	—
MCYT-100	17.62%	17.38%	17.42%	14.53%	12.99%	12.61%	**12.11**%

3.4. Matching. When a query signature is input to the system along with a claimed ID, the dissimilarity of its Fourier Descriptors from those of the reference signatures of the claimed person is calculated. Then, this distance is normalized using the reference set statistics of the user, and the query signature is accepted as genuine if this normalized distance is not too large. These steps are explained in detail in the following subsections.

3.4.1. Distance Between Query and Reference Set. During enrollment to the system, the user supplies a number of reference signatures that are used in accepting or rejecting a query signature. To find the dissimilarity between a query signature q and the reference set R_i of the claimed user i, we compute the Euclidian distance between the query features F_q obtained from q and the vector $\overline{F}_{R_i}$ which is the mean of the feature vectors of the reference signatures in R_i:

$$d(q, R_i) = \left\| F_q - \overline{F}_{R_i} \right\|. \tag{9}$$

We have also evaluated different matching algorithms, such as the number of matching Fourier Descriptors between the compared signatures but the presented matching algorithm gave the best results. Ideally, one can apply machine learning algorithms to find the most important descriptors or to decide whether the query is genuine or forgery given the Fourier descriptors of the query and reference set.

3.4.2. User-Dependent Distance Normalization. In order to decide whether the query is genuine or forgery, the distance computed in (9) should be normalized, in order to take into account the variability within the user's signatures. We use a normalization factor computed only from the reference signatures of the user. The normalization factor D_i which is separately calculated for each user i, is the average dissimilarity of a reference signature r to the rest of the reference signatures:

$$D_i = \text{mean}_{r \in R_i} d(r, R_i/r), \tag{10}$$

where R_i/r indicate the set R_i without the element r. The normalization factor D_i is calculated by putting a reference signature aside as query and calculating its dissimilarity d to the *remaining* reference signatures (R_i/r). The resulting normalized distance $d(x, R_i)/D_i$ is compared to a fixed, user-*independent* threshold.

We have previously found that this normalization is quite robust in the absence of training data [13]. Results of similar methods of normalization using slightly different statistics of the reference signatures are shown in Table 5. More conventional normalization techniques using client and impostor score distributions can be used when training data is available [31] and are expected to perform better.

3.4.3. Removing Outliers. Often, there are some important differences (in timing or shape) among the reference signatures of a user. In this work, we experimented with the removal of outliers from the reference set. While the template selection is a research area by itself, we found that eliminating up to one of the outlier from the reference set in a conservative fashion brings some improvement. For this, we sort the reference set distances of a user, as calculated using (9), and discard the last one (the one with the highest distance to the remaining references) if there is a big difference between the last two.

4. Experimental Results

4.1. Databases. The system performance is evaluated using the base protocols of the SUSIG [32] and MCYT [33] databases. The SUSIG database is a new, public database consisting of real-life signatures of the subjects and including "highly skilled" forgeries that were signed by the authors attempting to break the system. It consists of two parts: the Visual subcorpus obtained using a tablet with a built-in LCD display providing visual feedback and the Blind Subcorpus collected using a tablet without visual feedback. The Visual subcorpus used in this study contains a total of 2000 genuine signatures and 1000 skilled (half are highly skilled) forgeries collected in two sessions from 100 people. The data in SUSIG consists of x, y, and *timestamp*, collected at 100 Hz.

The MCYT database is a 330-people database of which a 100-user subcorpus is made public and is widely used for evaluation purposes. The database contains 25 genuine signatures and 25 skilled forgeries signed by 5 different

forgers, for each user. The data in MCYT database consists of consists of x, y, pressure, azimuth, and altitude, collected at 100 Hz. Table 1 summarizes these datasets, while the details can be found in their respective references.

4.2. Results of the Proposed System. We evaluated the usefulness of various preprocessing steps and the different components of the input signal, on the overall verification performance. The results obtained using the best set of preprocessing steps, while varying the input signal, are summarized in Table 2. As can be seen in this table, using only the coordinate information of the signature, we obtained minimum equal error rates of 6.20% and 14.53% for SUSIG and MCYT databases, respectively. These results are obtained using the concatenation of the Fourier descriptors obtained from $y(t)$ and $x(t)$. The pressure, azimuth, and altitude information available in the MCYT-100 database further reduced the EER to 12.11% EER. In addition to the EER results, the DET curves showing how FAR and FRR values change according to changing acceptance thresholds are given for the databases used in the evaluation, in Figure 3.

These results are obtained using 30 normalized Fourier descriptors per signal component (i.e., 30 for $y(t)$, 60 for $y(t) \mid x(t)$, etc.) and the preprocessing steps described in Section 4.4. However, very similar results were obtained with 20 and 25 descriptors. As described in Section 4.4, up to one reference signature was removed from the reference set, if deemed as an outlier. Timestamp information was not available for the MCYT database, and subsequently the pen-up durations were not used for this database.

Considering the effects of the different input signal components, we see that each information source brings the error rate down, from 14.53% using $x \mid y$ to 12.11% using $x \mid y \mid p \mid az \mid al$, for the MCYT database. Notice that the diminishing improvement is not necessarily and indication of the value of an input signal by itself. As for the positional information, we observe that the signature encoded as a complex signal (i.e., $s(t) = x(t) + iy(t)$) which was used in [23] gave significantly worse results compared to the concatenation of the features obtained from the x- and y-components separately (i.e., $x \mid y$). Another interesting observation is that our initial assumption about the x-component being mostly useless was not reflected in the results. While the x-component indeed contains little information in signatures signed strictly from left to right, the results show that it contains enough discriminative information to separate genuine and forgery signatures to a large extent, for the particular databases used.

In order to see the variation of the overall performance with respect to different sets of reference signatures, we ran 25 tests using the proposed method with different sets of 5 reference signatures, on the MCYT database. The mean EER for these tests was 10.89%, while standard deviation was 0.59. In fact, the worst performance was with the original set of references (genuine signatures [0–4]). The better performance with other reference sets can be explained by the fact that reference signatures collected over a wider time span better represent the time variation in the data.

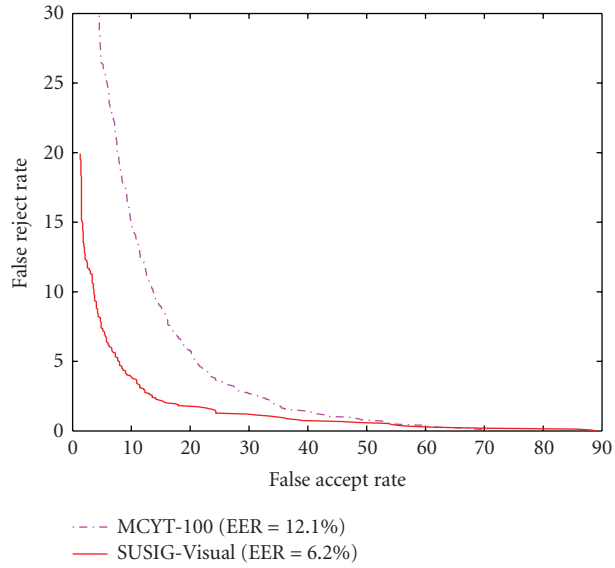


Figure 3: DET curves show how FAR (x-axis) and FRR (y-axis) values change according to changing acceptance threshold, for the tested databases.

The proposed FFT system is very fast: it can process 4500 queries in the MCYT-100 database in 69 seconds of CPU time.

4.3. Effects of Preprocessing Steps. The best results reported in Table 2 were obtained using few preprocessing steps, namely, pen-up duration encoding and drift and mean removal. Some of the other preprocessing steps used in previous work based on FFT [23, 25] were just not useful due to our normalized features (e.g., rotation and scale normalization), while resampling worsened results by removing discriminative information (30.02% versus 6.20% EER for the SUSIG database and 17.82% versus 12.11% EER for the MCYT database). On the other hand, removal of the drift (especially significant in the x-component) was found to improve performance in both our work and in previous work [23], by a few percent points. The effects of drift and mean removal are most apparent when they are used together. Note that mean removal is normally not necessary, since translation invariance is provided when the first Fourier coefficient is discarded; however mean removal affects the outcome due to zero padding.

The proposed incorporation of the pen-up duration is also found to help increase performance (9.09% EER versus 6.20% EER for the SUSIG database).

4.4. Effects of Distance Normalization. Normalization of the query distance, prior to using a *fixed* threshold across all users, has been found to make a significant difference on verification performance, as shown in Table 4. Here, AvgN refers to dividing the distance between the query and the mean descriptor vector by the average distance of the reference signatures. This average is obtained by using a leave-1-out method whereby one of the reference signature

Table 3: Effects of various preprocessing steps on the best configuration. The bold face shows the results of the proposed system, while the last column shows the results if resampling was added to the proposed preprocessing steps (drift and mean removal and pen-up duration incorporation when available).

Dataset	Feature	Raw	Drift	Mean	Drift + Mean	Proposed = Drift + Mean + PenUp	Proposed if resampled
SUSIG-Visual	$y \mid x$	8.18%	7.34%	11.52%	9.09%	**6.20**%	30.02%
MCYT-100	$y \mid x \mid p \mid az \mid al$	20.31%	20.38%	13.51%	**12.11**%	—	17.82%

Table 4: Different methods for user-dependent distance normalization using *only* the reference data.

Dataset	Feature	AvgN	MinN	MaxN	None
MCYT-100	$y \mid x \mid p \mid az \mid al$	**12.11**%	13.2%	14.3%	21.5%
SUSIG-Visual	$y \mid x$	**6.20**%	8.1%	5.8%	14.1%

is treated as query, while the others are used as reference, as described in Section 3.4.2. Similarly, MinN and MaxN refer to dividing the distance between the query and the mean descriptor vector by the minimum and maximum of the reference signature distances (again using the leave-one-out method), respectively. All three of these normalization methods are better than not doing any normalization at all.

Notice that while AvgN gives the best results for the MCYT-100 dataset, MaxN has given the best results for the SUSIG database. This difference highlights an important aspect of the current work, which is the fact that the exact same system is used in testing both databases, without any adjustment. In all of the presented results, we use the AvgN normalization method.

4.5. Results of the Fusion with the DTW System. It has been shown in the last couple of years that the combination of several experts improves verification performance in biometrics [7, 28, 34, 35]. Some of the results, especially as related to the work described here, are summarized in Section 4.6.

In order to show that the proposed FFT system may complement an approach based on local features, we combined the FFT system with a slightly modified implementation of the DTW system described in [13]. The distribution of the DTW and FFT scores in Figure 4 shows that the two systems' scores show a loose correlation, which is an important factor in classifier combination systems. The combination is done using the sum rule, after normalizing the scores of the two systems. The score normalization factor is selected separately for each database, so as to equalize the mean scores of the two systems, as computed over the *reference* signatures in that database. A better selection of the normalization factor can be made when training data is available. Note that using the sum rule with score normalization is equivalent to separating the genuine and forgery classes using a straight line with a fixed slope, where the y-intercept is adjusted to find the equal error rate.

The results given in Table 5 show that the FFT system improves the performance of the DTW system significantly, by 8% or 26% depending on the database. Furthermore, the

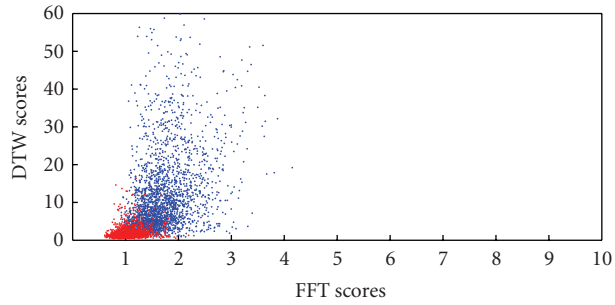


Figure 4: The distribution of the DTW and FFT scores for the MCYT-100 database.

improvement brings the EER rates to state-of-the-art levels given in Table 6 for both databases (3.03% for SUSIG and 7.22% for MCYT-100).

The proposed FFT system is very fast: it can process 4500 queries in the MCYT-100 database in 69 seconds of CPU time. In comparison, the DTW system takes 36 800 seconds for the same task, which corresponds to a factor of more than 500. Theoretically, the time complexity of the DTW system is $O(N \times M)$, where N and M are the lengths of the two signatures being compared, while that of the FFT is $O(N \log N)$ for a signature of length N. Hence, even though using the FFT system in addition to the DTW system results in negligeable time overhead, Figure 4 shows that the systems can also be called in a serial fashion to eliminate the more obvious forgeries using the FFT system and calling the DTW system only for the less certain cases. Using this test with a threshold of 4, the same reported results were obtained while gaining around 10% speed overall.

The DTW approach is probably the most commonly used technique in online signature verification, while quite successful overall and in particular in aligning two signatures, the basic DTW approach has some shortcomings, such as assigning low distance scores to short dissimilar signatures. One such example is shown in Figure 5, along with all of the genuine signatures of the claimed user. As an approach using global features, the FFT-based system is expected to be useful in eliminating some of these errors, when used in fusion with DTW or other local approaches.

4.6. Comparison with Previous Work. Results of previous work tested on the MCYT database are given in Table 6 for comparison. Since SUSIG is a new database, we concentrated on previous work reporting results on the MCYT database. Even with this database, comparing different results is

Table 5: Results of the fusion of the FFT system with our Dynamic Time Warping system.

Dataset	$y \mid x$	$y \mid x \mid p \mid az \mid al$	DTW	DTW + $y \mid x$	$DTW + y \mid x \mid p \mid az \mid al$	Improvement
SUSIG-Visual	6.20%	—	**3.30**%	**3.03**%	—	8%
MCYT-100	14.53%	12.11%	**9.81**%	7.8%	**7.22**%	26%

Table 6: State-of-the-art results on the MCYT database using a priori normalization techniques. Unless otherwise indicated, all dimensions of the input signal are used.

Reference	Dataset	Method	Features	Performance
Garcia-Salicetti et al. [35]	MCYT-280	HMM [18]		5.73%
		HMM [31]		8.39%
		String Matching [36]		15.89%
		Fusion of [18, 31]		3.40%
Faundez-Zanuy [28]	MCYT-280	VQ		11.8%
		DTW		8.9%
		VQ-DTW		5.4% (DCF)
Vivaracho-Pascual et al. [37]	MCYT-280	Length normaliz./p-norm		6.8% (DCF)
Nanni and Lumini [34]	MCYT-100*	SVM	100 global features	17.0%
		SVM-DTW [13]		7.6%
Nanni and Lumini [26]	MCYT-100	Wavelet-DCT	x, y	11.4%
		Wavelet-DCT	x, y, az	9.8%
		Wavelet-DCT fused w/DTW, HMM, GM		5.2%
Quan et al. [25]	MCYT-100*	STFT		7%
This work	MCYT-100	Proposed FFT	x, y	12.11%
		DTW [13]		9.81%
		FFT-DTW		7.22%

difficult due to varying experimental setups. In particular, we have (i) the subset of the MCYT database used: MCYT-280 is the test subset of the full database of 330 people where a 50-people portion is used for training, while MCYT-100 is the publicly available part consisting of 100 people and no allocated training subset; (ii) the number of reference signatures used (most systems use the first 5 genuine signatures as suggested, while others use more, as necessitated by their verification algorithm); (iii) number of available component signals used, such as coordinate sequence, pressure, and azimuth (not counting derived features); and (iv) whether *a priori* or *a posteriori* normalization is used for score normalization, as defined in [31].

In general, the higher the number of references, the better one would expect the results to be, due to having more information about the genuine signatures of a user. Similarly, higher number of signal components normally give better results. Finally, score normalization affects the performance significantly, since the *a posteriori* normalization results are intended to give the best possible results, if all genuine and/or forger statistics in the database were known ahead of time. For this comparison, we tried to included recent results on the MCYT database, using 5 reference signatures as suggested and *a priori* score normalization methods, to the best of our knowledge.

Given the various factors affecting performance and the difficulty in assessing the exact experimental setups of others' work, an exact comparison of different systems is not very easy. Nonetheless, we give the following as indicative results. The best results obtained with the MCYT-100 database is reported by Nanni and Lumini, with 5.2% EER using 3 measured signals (x, y, azimuth) using four experts including Wavelet, DTW and HMM approaches [26]. In that work, the Wavelet based system itself achieves 9.8% EER. The other system developed by the same authors which uses Support Vector Machines (SVMs) with 100 global features obtains 17.0% on the MCYT-100 database (using a 20-people subset for training), while the combination of SVM and DTW (based on our DTW system used in the fusion part of this work [13]) achieves 7.6% [34]. Quan et al. report 7% EER of using windowed FFT on the MCYT-100 but using 15 genuine signatures as reference (instead of 5 which is the suggested number).

On the MCYT-280 database, Garcia-Salicetti et al. evaluates 3 individual systems in a study of complementarity; the individual systems' performance are given as 5.73%, 8.39%, and 15.89%, while the best fusion system obtains 3.40% EER on skilled forgeries [35]. Faundez-Zanuy reports 11.8% and 5.4% using Vector Quantization (VQ) and VQ combined with DTW respectively [28]. However, instead of EER, they

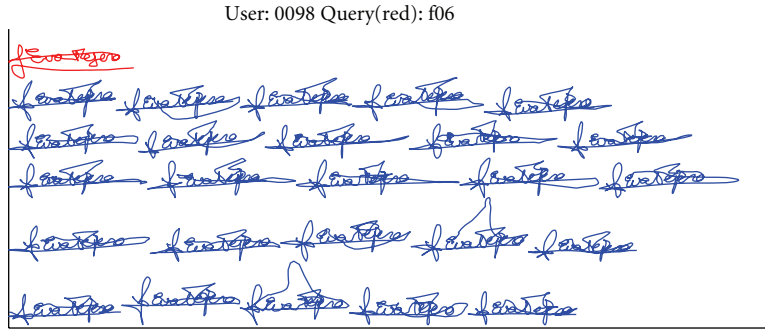


Figure 5: A forgery signature (shown on top) that was misclassified using the DTW system while it was correctly classified using the combined system.

report the main results using the Detection Cost Function (DCF) with 5 genuine and 25 forgery signatures per person. Similarly, Vivaracho-Pascual et al. report a DCF of 6.8%, using the same experimental setup.

The most apparent factor in these results is the effect of classifier combination. Classifier combination or fusion systems are found to be useful in many pattern recognition problems, so the improvement of the results is not surprising and is parallelled in our current results as well. The other important factor affecting performance is the dimensionality of the input signal. In some databases, x- and y-coordinates are the only available dimensions, while pressure, azimuth, and altitude are also available in others. Increasing the number of dimensions generally increases the verification performance, as more relevant information is available to the classifier. One interesting note is that the DTW appears as a component in each of the listed fusion systems.

The performance of the proposed FFT system is lower than the state-of-the-art fusion systems, while it seems to be in par with single engine systems on the same database (12.11% versus 9.8% [26], 17.0% [34], and 9.81% with our DTW approach, on the MCYT database). Approaches using global features typically underperform compared to those using local features. On the other hand, global approaches are necessary in certain applications. Furthermore, due to their speed and complementarity, they are expected to be useful in fusion systems to increase the performance and/or the speed.

We also have to underline the fact that when reporting results on a database, researchers typically report the results of the optimal set of features and algorithm steps, which introduces bias to the results. In fact, often a particular step of an algorithm improves the results on one database, while degrading it on another (e.g., different distance normalization methods gave the best results in SUSIG and MCYT databases, as shown in Table 5). Therefore, the fact that our results are obtained by testing the same exact system on two different databases with different characteristics (e.g., signature types, sensors, measured signals, forgery skills) is important.

As for comparison with previous work using FFT, the system developed by Lam et al. [23] is reported to have 2.5% error rate, however the dataset in their work is very small (8 genuine signatures of the same user and 152 forgeries provided by 19 forgers) and old, making a direct comparison impossible. Similarly, while the improvement of using windowed FFT, suggested by Quan et al. [25] is reasonable, their results are not readily comparable to ours: they report an EER of 7% on the MCYT-100, using 15 genuine signatures as reference instead of 5, presumably necessitated by their use of the Mahalanobis distance. As mentioned before, increased number of reference signatures are expected to increase performance and the resulting test set in their case is significantly different than ours. Furthermore, we have also shown that resampling step used in both of these works significantly degrades verification performance for the proposed method by removing some of the timing information which is useful in discriminating forgery and genuine signatures.

5. Future Work

In the current system, we use only the magnitude of the Fourier coefficients, discarding the phase information for simplicity, while phase information is actually a fundamental part of the signal. We expect that the use of the phase information can improve the system performance. Similarly, other extracted features, such as local velocity, can easily be used and would be expected to improve the system performance based on others' work [35].

Another improvement may be the use of windowed or Short Term Fourier Transform. The STFT aims to give more information about the timing as well as the frequency component of the signal, by breaking the input signal into a number of small segments by a windowing signal prior to the application of the Fourier transform. The size of the window used for this operation is an issue in general but for online signature verification, separate strokes or high curvature points can be used for this purpose.

An analysis of the errors shows that large portion of the errors is due to simple signatures, composed of simple or easily reproducible trajectories. While not much may be done to reduce errors on these types of signatures, one could at least envision a system alerting users when they use simple signatures at enrollment time.

6. Summary and Discussions

We presented a novel approach for online signature verification using global features consisting of Fourier Descriptors that provide a compact and fixed-length representation of an online signature. Our approach is significantly different in preprocessing, feature extraction, normalization and matching steps, compared to previous online signature verification systems that are based on FFT. These steps are carefully designed to retain the full discriminatory information available in the signature; in particular the incorporation of the timestamp information for representing pen-up durations is novel and had significant effects on performance.

The proposed system is extensively tested using two large public databases, both in terms of overall performance and the effects of individual preprocessing steps. The results are inferior to the best results obtained by fusion systems but the system shows potential as a stand-alone system to be used wherever fixed-length representation is needed, and in complementing an approach based on local features. The latter is supported experimentally by the fact the combination of the proposed FFT system improved the results of our state-of-the-art DTW system, resulting in EER of 3.03% for the SUSIG database and 7.22% for the MCYT-100 database. Furthermore, given the previously mentioned factors affecting performance and the difficulty in assessing the exact experimental setups of others' work, an exact comparison of EER results is not always meaningful. This is especially true since the proposed system is tested with *exactly* the same parameters on two different databases with different characteristics (e.g., signature types, sensors, measured signals, forgery skills).

As for overall speed, the proposed system is very fast, about 500 times faster than a dynamic programming approach on the same database. The speed is thus one of the advantages of the proposed system and is especially important in fusion systems and identification problems as well as quickly testing new algorithms or preprocessing steps.

The main aspects of the developed FFT system can thus be summarized as follows:

(i) it is very fast in training, feature extraction, and matching (about 2-3 orders of magnitude faster than the DTW system);

(ii) it uses a fixed-length feature vector comprised of global features of the signature, which is required in certain applications;

(iii) its performance is lower than state-of-the-art results obtained by fusion systems; however its advantages and potential improvements make it a useful alternative in online signature verification, especially in complementing more complex but slower methods based on local features, such as the DTW or HMM approaches.

Given its merits as a global approach and the suggested improvements, we believe that the proposed FFT-based system has potential as a stand-alone system but especially in complementing an approach based on local features. Furthermore, we would expect to have a lower EER by adding more features that are found useful in other studies, such as local velocity or acceleration; this would be done by simply concatenating the new features to the ones used in this work.

Acknowledgments

The authors would like to thank Professor Anil Jain for hosting B. Yankoglu during her sabbatical, Dr. Özgür Gürbüz for valuable help with the Fourier transform, and J. Fierrez-Aguilar and J. Ortega-Garcia for sharing the MCYT-100 database. This work was partially supported by TÜBİTAK (The Scientific and Technical Research Council of Turkey), under project no. 105E165.

References

[1] D. Yeung, H. Chang, Y. Xiong, et al., "SVC2004: first intional signature verification competition," in *Proceedings of the 1st International Conference on Biometric Authentication (ICBA '04)*, pp. 16–22, 2004.

[2] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijen, A. M. Bazen, and R. N. J. Veldhuis, "Practical biometric authentication with template protection," in *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA '05)*, vol. 3546 of *Lecture Notes in Computer Science*, pp. 436–446, 2005.

[3] A. Kholmatov and B. Yanikoglu, "An individuality model for online signatures," in *Defense and Security: Biometric Technology For Human Identification V*, Proceedings of SPIE, Orlando, Fla, USA, March 2008.

[4] T. Ohishi, Y. Komiya, and T. Matsumoto, "On-line signature verification using pen-position, pen-pressure and pen-inclination trajectories," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 4, pp. 45–47, 2000.

[5] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern Recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[6] C. Vielhauer, R. Steinmetz, and A. Mayerhofer, "Biometric hash based on statistical features of online signatures," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 1, p. 10123, 2002.

[7] J. Fieriez-Aguilar, L. Nanni, J. Lopez-Peñalba, J. Ortega-Garcia, and D. Maltoni, "An on-line signature verification system based on fusion of local and global information," in *Proceedings of the 5th International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA '05)*, pp. 523–532, 2005.

[8] H. Xu, R. N. J. Veldhuis, T. A. M. Kevenaar, A. H. M. Akkermans, and A. M. Bazen, "Spectral minutiae: a fixed-length representation of a minutiae set," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '08)*, pp. 1–6, 2008.

[9] J. Yi, C. Lee, and J. Kim, "Online signature verification using temporal shift estimated by the phase of gabor filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 776–783, 2005.

[10] H. Lei and V. Govindaraju, "A comparative study on the consistency of features in on-line signature verification," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2483–2489, 2005.

[11] M. Parizeau and R. Plamondon, "Comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 710–717, 1990.

[12] R. Martens and L. Claesen, "Dynamic programming optimisation for on-line signature verification," in *Proceedings of the 4th International Conference on Document Analysis and Recognition (ICDAR '97)*, vol. 2, pp. 653–656, Ulm, Germany, 1997.

[13] A. Kholmatov and B. Yanikoglu, "Identity authentication using improved online signature verification method," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2400–2408, 2005.

[14] J. J. van Oosterhout, H. Dolfing, and E. Aarts, "On-line signature verification with hidden markov models," in *Proceedings of the 14th International Conference on Pattern Recognition (ICPR '98)*, vol. 2, p. 1309, 1998.

[15] R. Kashi, J. Hu, W. L. Nelson, and W. Turin, "A hidden markov model approach to online handwritten signature verification," *International Journal on Document Analysis and Recognition*, vol. 1, pp. 102–109, 1998.

[16] G. Rigoll and A. Kosmala, "A systematic comparison of on-line and off-line methods for signature verification with hidden markov models," in *Proceedings of the 14th International Conference on Pattern Recognition (ICPR '98)*, pp. 1755–1757, 1998.

[17] D. Muramatsu and T. Matsumoto, "An hmm on-line signature verifier incorporating signature trajectories," in *Proceedings of the 7th International Conference on Document Analysis and Recognition (ICDAR '03)*, 2003.

[18] B. Ly Van, S. Garcia-Salicetti, and B. Dorizzi, "On using the viterbi path along with hmm likelihood information for on-line signature verification," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 37, no. 5, pp. 1237–1247, 2007.

[19] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, "HMM-based on-line signature verification: feature extraction and signature modeling," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2325–2334, 2007.

[20] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification-the state of the art," *Pattern Recognition*, vol. 22, no. 2, pp. 107–131, 1989.

[21] F. Leclerc and R. Plamondon, "Automatic signature verification: the state of the art," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 8, no. 3, pp. 643–660, 1994.

[22] D. Impedovo and G. Pirlo, "Automatic signature verification: the state of the art," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 38, no. 5, pp. 609–635, 2008.

[23] C. F. Lam, D. Kamins, and K. Zimmermann, "Signature recognition through spectral analysis," *Pattern Recognition*, vol. 22, no. 1, pp. 39–44, 1989.

[24] Y. Sato and K. Kogure, "Online signature verification based on shape, motion and writing pressure," in *Proceedings of the International Conference on Pattern Recognition (ICPR '82)*, pp. 823–826, 1982.

[25] Z.-H. Quan, D.-S. Huang, X.-L. Xia, M. R. Lyu, and T.-M. Lok, "Spectrum analysis based on windows with variable widths for online signature verification," in *Proceedings of the International Conference on Pattern Recognition (ICPR '06)*, vol. 2, pp. 1122–1125, 2006.

[26] L. Nanni and A. Lumini, "A novel local on-line signature verification system," *Pattern Recognition Letters*, vol. 29, no. 5, pp. 559–568, 2008.

[27] I. Nakanishi, N. Nishiguchi, Y. Itoh, and Y. Fukui, "Multi-matcher on-line signature verification system in dwt domain," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 1, pp. 178–185, 2006.

[28] M. Faundez-Zanuy, "On-line signature recognition based on VQ-DTW," *Pattern Recognition*, vol. 40, no. 3, pp. 981–992, 2007.

[29] D. Muramatsu and T. Matsumoto, "Effectiveness of pen pressure, azimuth, and altitude features for online signature verification," in *Proceedings of the International Conference on Biometrics*, pp. 503–512, 2007.

[30] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 1992.

[31] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target dependent score normalization techniques and their application to signature verification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 35, no. 3, pp. 418–425, 2005.

[32] A. Kholmatov and B. Yanikoglu, "SUSIG: an on-line signature database, associated protocols and benchmark results," *Pattern Analysis and Applications*, pp. 1–10, 2008.

[33] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, et al., "MCYT baseline corpus: a bimodal biometric database," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[34] L. Nanni and A. Lumini, "Advanced methods for two-class problem formulation for on-line signature verification," *Neurocomputing*, vol. 69, no. 7–9, pp. 854–857, 2006.

[35] S. Garcia-Salicetti, J. Fierrez-Aguilar, F. Alonso-Fernandez, et al., "Biosecure reference systems for on-line signature verification: a study of complementarity," *Annals of Telecommunications*, vol. 62, no. 1-2, pp. 36–61, 2007.

[36] S. Schimke, C. Vielhauer, and J. Dittmann, "Using adapted levenshtein distance for on-line signature authentication," in *Proceedings of the International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 931–934, 2004.

[37] C. Vivaracho-Pascual, M. Faundez-Zanuy, and J. M. Pascual, "An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances," *Pattern Recognition*, vol. 42, no. 1, pp. 183–193, 2009.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 235746, 13 pages
doi:10.1155/2009/235746

Research Article

Retinal Verification Using a Feature Points-Based Biometric Pattern

M. Ortega,[1] M. G. Penedo,[1] J. Rouco,[1] N. Barreira,[1] and M. J. Carreira[2]

[1] *VARPA Group, Faculty of Informatics, Department of Computer Science, University of Coruña, 15071 A Coruña, Spain*
[2] *Department of Electronics and Computer Science, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain*

Correspondence should be addressed to M. Ortega, mortega@udc.es

Received 14 October 2008; Accepted 12 February 2009

Recommended by Natalia A. Schmid

Biometrics refer to identity verification of individuals based on some physiologic or behavioural characteristics. The typical authentication process of a person consists in extracting a biometric pattern of him/her and matching it with the stored pattern for the authorised user obtaining a similarity value between patterns. In this work an efficient method for persons authentication is showed. The biometric pattern of the system is a set of feature points representing landmarks in the retinal vessel tree. The pattern extraction and matching is described. Also, a deep analysis of similarity metrics performance is presented for the biometric system. A database with samples of retina images from users on different moments of time is used, thus simulating a hard and real environment of verification. Even in this scenario, the system allows to establish a wide confidence band for the metric threshold where no errors are obtained for training and test sets.

1. Introduction

Reliable authentication of persons is a growing demanding service in many fields, not only in police or military environments but also in civilian applications, such as access control or financial transactions. Traditional authentication systems are based on knowledge (a password, a pin) or possession (a card, a key). But these systems are not reliable enough for many environments, due to their common inability to differentiate between a true-authorised user and a user who fraudulently acquired the privilege of the authorised user. A solution to these problems has been found in the biometric-based authentication technologies. A biometric system is a pattern recognition system that establishes the authenticity of a specific physiological or behavioural characteristic. Authentication is usually used in the form of verification (checking the validity of a claimed identity) or identification (determination of an identity from a database of known people, this is, determining who a person is without knowledge of his/her name).

Many authentication technologies can be found in the literature, some of them already implemented in commercial authentication packages [1–3]. Other methods are the fingerprint authentication [4, 5] (perhaps the oldest of all the biometric techniques), hand geometry [6], face [7, 8], or speech recognition [9]. Nowadays, the most of the efforts in authentication systems tend to develop more secure environments, where it is harder, or ideally impossible, to create a copy of the properties used by the system to discriminate between authorised and unauthorised individuals. [10–12].

This paper proposes a biometric system for authentication that uses the retina blood vessel pattern. This is a unique pattern in each individual and it is almost impossible to forge that pattern in a false individual. Of course, the pattern does not change through the individual's life, unless a serious pathology appears in the eye. Most common diseases like diabetes do not change the pattern in a way that its topology is affected. Some lesions (points or small regions) can appear but they are easily avoided in the vessels extraction method that will be discussed later. Thus, retinal vessel tree pattern has been proved a valid biometric trait for personal authentication as it is unique, time invariant and very hard to forge, as showed by Mariño et al. [13, 14], who introduced a novel authentication system based on this trait. In that work, the whole arterial-venous tree structure was used as

the feature pattern for individuals. The results showed a high confidence band in the authentication process but the database included only 6 individuals with 2 images for each of them. One of the weak points of the proposed system was the necessity of storing and handling a whole image as the biometric pattern. This greatly facilitates the storing of the pattern in databases and even in different devices with memory restrictions like cards or mobile devices. In [15] a pattern is defined using the optic disc as reference structure and using multi scale analysis to compute a feature vector around it. Good results were obtained using an artificial scenario created by randomly rotating one image per user for different users. The dataset size is 60 images, rotated 5 times each. The performance of the system is about a 99% accuracy. However, the experimental results do not offer error measures in a real-case scenario where different images from the same individual are compared.

Based on the idea of fingerprint minutiae [4, 16], a robust pattern was first introduced in [17] where a set of landmarks (bifurcations and crossovers of retinal vessel tree) were extracted and used as feature points. In this scenario, the pattern matching problem is reduced to a point pattern matching problem and the similarity metric has to be defined in terms of matched points. A common problem in previous approaches is that the optic disc is used as a reference structure in the image. The detection of the optic disc is a complex problem and in some individuals with eye diseases this cannot be achieved correctly. In this work, the use of reference structures is avoided to allow the system to cope with a wider range of images and users.

The paper is organised as follows: in Section 2 a description of the authentication system is presented, specially the feature points extraction and the matching stages. Section 3 deals with the analysis of some similarity metrics. Section 4 shows the effectiveness results obtained by the previously described metrics running a test images set. Finally, Section 5 provides some discussion and conclusions.

2. Authentication System Process

In this work, the retinal vessel pattern for every person is ultimately defined by a set of landmarks, or feature points, in the vessel tree. For the system to perform properly, a good representation of the retinal vessel tree is needed. The extraction of the retinal vessel tree is explained in Section 2.1. Next, the biometric pattern for an individual is obtained via the feature points extracted from the vessel tree (Section 2.2). The last stage in the authentication process is the matching between the reference stored pattern for an individual and the pattern from the acquired image (Section 2.3).

2.1. Retinal Vessel Tree Extraction. Following the idea that vessels can be thought of as creases (ridges or valleys) when images are seen as landscapes (Figure 1), curvature level curves are employed to calculate the creases (crest and valley lines).

Among the many definitions of a crease, the one based on level set extrinsic curvature or LSEC, (1), has useful invariance properties. Given a function $L : \mathbb{R}^d \to \mathbb{R}$, the level set for a constant l consists of the set of points $\{\mathbf{x} \mid L(\mathbf{x}) = l\}$. For $2D$ images, L can be considered as a topographic relief or landscape and the level sets as its level curves. Negative minima of the level curve curvature κ, level by level, form valley curves, and positive maxima form ridge curves:

$$\kappa = (2L_xL_yL_{xy} - L_y^2L_{xx} - L_x^2L_{yy})(L_x^2 + L_y^2)^{-3/2}. \quad (1)$$

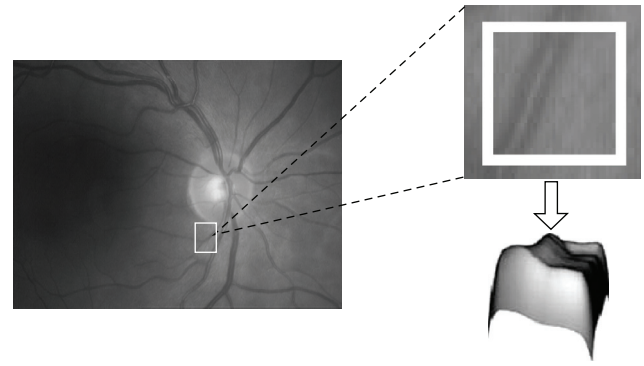


Figure 1: Representation of a region in the image as a landscape. Left side shows the retinal image with the region of interest marked with a white rectangle. In the right side, the zoomed image over the region of interest and the same region represented as a landscape, showing the creaseness feature.

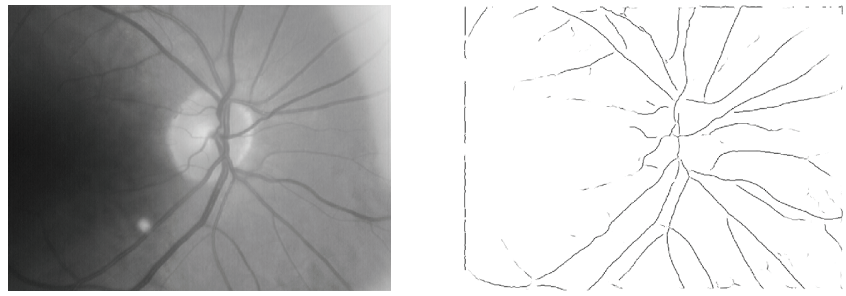
However, the usual discretization of LSEC is ill-defined in a number of cases, giving rise to unexpected discontinuities at the centre of elongated objects. Due to this, the *MLSEC-ST* operator, defined in [18, 19] for 3D landmark extraction of CT and MRI volumes, is used. This alternative definition is based on the divergence of the normalised vector field $\overline{\mathbf{w}}$:

$$\kappa = -\mathrm{div}(\overline{\mathbf{w}}). \quad (2)$$

Although (1) and (2) are equivalent in the continuous domain, in the discrete domain, when the derivatives are approximated by finite-centred differences of the Gaussian-smoothed image, (2) provides much better results. The creaseness measure κ is improved by prefiltering the image gradient vector field using a Gaussian function.

Figure 2 shows the result of the creases extraction algorithm for an input digital retinal image. Once the creases image is calculated, the retinal vessel tree is extracted and can be used as a valid biometric pattern. However, using the whole creases image as biometric pattern has a major problem in the codification and storage of the pattern as we need to store and handle the whole image. To solve this, similarly to the fingerprint minutiae, a set of landmarks is extracted as the biometric pattern in the creases image. These landmarks are representative enough for each individual while consisting of a very reduced set of structures in the retinal tree. In the next subsection, the extraction process of this pattern is described.

2.2. Feature Points Extraction. The goal in this stage is to obtain a robust and consistent biometric pattern easy to



(a) (b)

Figure 2: Example of digital retinal images showing the vessel tree. (a) Input retinal image. (b) Creases image from the input representing the main vessels in the retina.

code and store. To perform this task, a set of landmarks are extracted. The most prominent landmarks in retinal vessel tree are crossovers (between two different vessels) and bifurcation points (one vessel coming out of another one) and they will be used in this work as the set of feature points constituting the biometric pattern for characterising individuals. Thus, the biometric pattern can be stored as a set of feature points.

The creases image will be used to extract the landmarks, as it is a good representation of the vessels in the retinal tree as explained earlier. The landmarks of interest are points where two different vessels are connected. Therefore, it is necessary to study the existing relationships between vessels in the image. The first step is to track and label the vessels to be able to establish those relationships between them.

In Figure 3, it can be observed that creases images show discontinuities in the crossovers and bifurcations points. This occurs because of the two different vessels (valleys or ridges) coming together into a region where the crease direction cannot be set. Moreover, due to some illumination or intensity loss issues, creases images can also show some discontinuities along a vessel (Figure 3). This issue require a process of joining segments to build the whole vessels prior to the bifurcation/crossover analysis.

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

(1) labelling of the vessels segments,

(2) establishing the joint or union relationships between vessels,

(3) establishing crossover and bifurcation relationships between vessels,

(4) filtering of the crossovers and bifurcations.

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

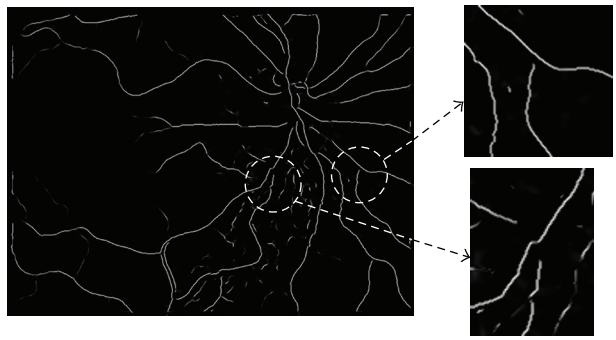


Figure 3: Example of discontinuities in the creases of the retinal vessels. Discontinuities in bifurcations and crossovers are due to two creases with different directions joining in the same region. Also, some other discontinuities along a vessel can happen due to illumination and contrast variations in the image.

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

Once the segment tracking process has started, in every step a neighbour of the last pixel flagged as segment is chosen to be the next. This choice is made using the following criterion: the best neighbour is the one with most nonflagged yet neighbours belonging to the segment. This heuristic contains the idea of keeping the 1pixel width segment to track along the middle of the crease (where pixels have more segment pixels neighbours), keeping also

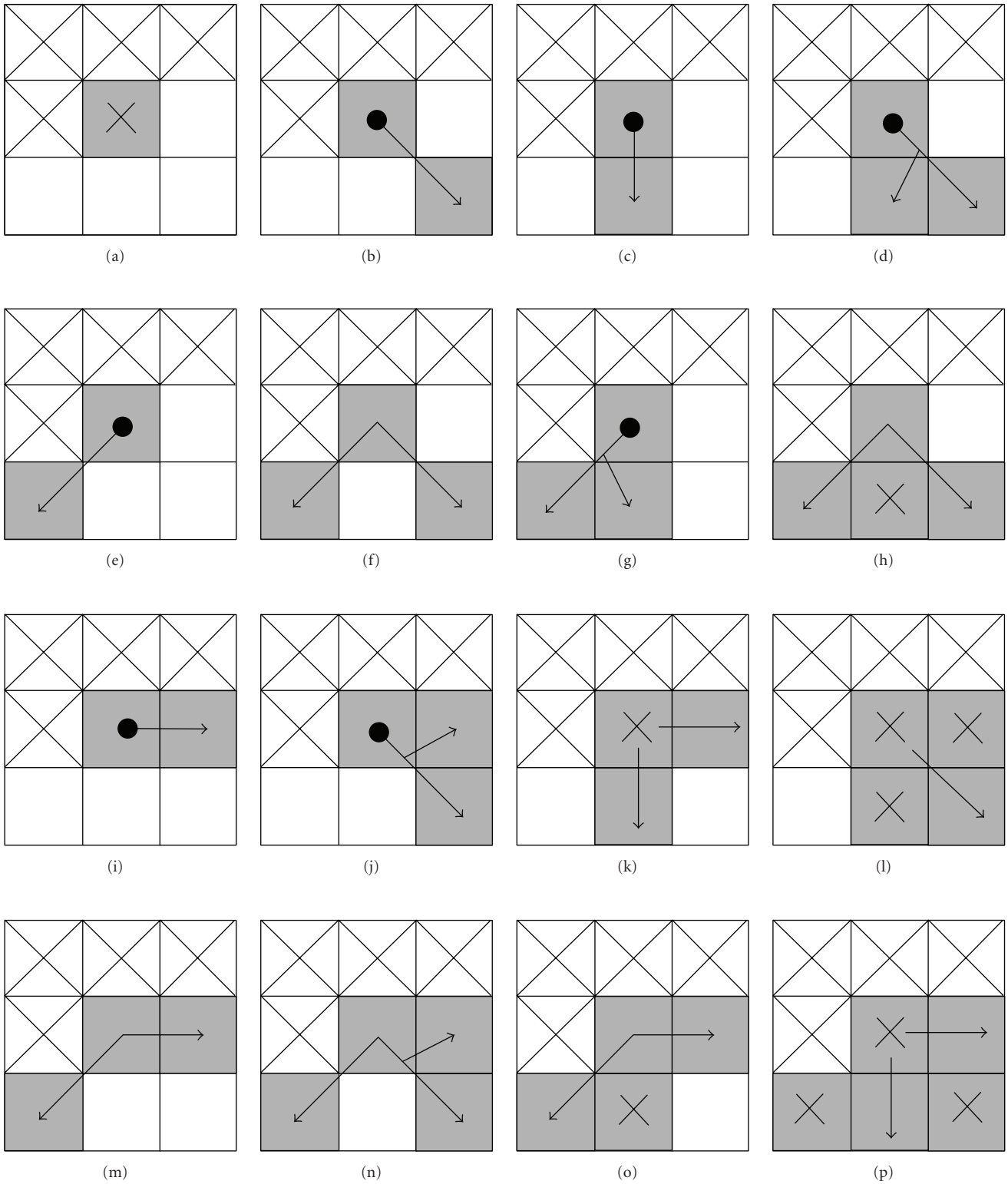


Figure 4: Initial tracking process for a segment depending on the neighbours pixels surrounding the first pixel found for the new segment in a 8-neighbourhood. As there are 4 neighbours not tracked yet (the bottom row and the one to the right), there are a total of 16 possible configurations. Gray squares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

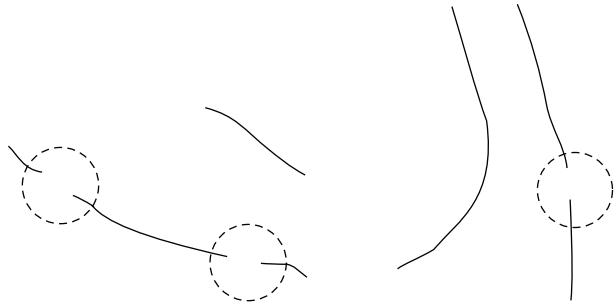


Figure 5: Examples of union relationships. Some of the vessels present discontinuities leading to different segments. These discontinuities are detected in the union relationships detection process.

the original orientations in every step. When the whole image tracking process finishes, every segment is a 1pixel-width line with its endpoints defined. The endpoints are very useful to establish relationships between segments as those relationships can always be detected in the surroundings of a segment endpoint. This avoids the analysis of every pixel belonging to a vessel, considerably reducing the complexity of the algorithm and, therefore, the running time. Finally, to avoid some spurious segments or noise to appear, small segments are removed using a length threshold.

2.2.2. Union Relationships. As stated before, unions detection is needed to build the vessels out of their segments. Aside the segments from the creases image, no additional information is required and therefore is the first kind of relationship to be detected in the image. An union or joint between two segments exists when one of the segments is the continuation of the other in the same retinal vessel. Figure 5 shows some examples of union relationships between segments.

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

An efficient approach to connect the segments is using a straight line between both endpoints. In Figure 6(a), a graphical description of the detection process for an union is showed. The smoothness measurement is obtained from the angles between the straight line and the segment direction. The segment direction is calculated by the endpoint direction. The maximum smoothness occurs when both angles are π rad., that is, both segments are parallel and belong to the straight line connecting it. The smoothness decreases as both angles decrease. A criterion to accept the candidate relationship must be established. A minimum angle $\theta_{\min}$ is set as the threshold for both angles. This way, the criterion to accept an union relationship is defined as

$$\text{Union}(r,s) = (\alpha > \theta_{\min}) \wedge (\beta > \theta_{\min}), \tag{3}$$

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

2.2.3. Bifurcation/Crossover Relationships. Bifurcations and crossovers are the feature interest points in this work for characterising individuals by a biometric pattern. A crossover is an intersection between two segments. A bifurcation is a point in a segment where another one starts from. While unions allow to build the vessels, bifurcations allow to build the vessel tree by establishing relationships between them. Using both types the retinal vessel tree can be reconstructed by joining all segments. An example of this is shown in Figure 6(b).

A crossover can be seen in the segments image, as two bifurcations between a segment and two others related by an union. Therefore, finding bifurcation and crossover relationships between segments can be reduced to find only bifurcations. Crossovers can then be detected analysing close bifurcations.

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

For every endpoint in the image, the process is as follows (Figure 6(c)):

(1) compute the endpoint direction,

(2) extend the segment in that direction a fixed length $l_{\max}$,

(3) analyse the points in and nearby the prolongation segment to find candidate segments,

(4) if a point of a different segment is found, compute the angle (α) associated to that bifurcation, defined by the direction of this point and the extreme direction from step 1.

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

Figure 7 shows one example of results after this stage. Feature points are marked. Also, spurious detected points are identified in the image. These spurious points may occur for different reasons such as wrongly detected segments. In the image test set used (over 100 images) the approximate mean number of feature points detected per image was 28. The mean of spurious points corresponded to 5 points per image. To improve the performance of the matching process is convenient to eliminate as spurious points as possible. Thus, the last stage in the biometric pattern extraction process will be the filtering of spurious points in order to obtain an accurate biometric pattern for an individual.

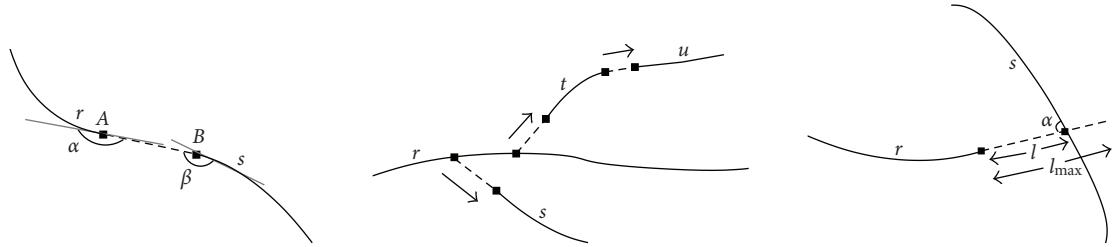


Figure 6: (a) Union of creases segments r and s. The angles between the new segment $\overline{AB}$ and the creases segments r (α) and s (β) are near π rad. so they are above the required threshold ($(3/4)\pi$) and the union is finally accepted. (b) Retinal Vessel Tree reconstruction by unions (t, u) and bifurcations (r, s) and (r, t). (c) Bifurcation between segment r and s. The endpoint of r is prolonged a maximum distance $l_{\max}$ and eventually a point of segment s is found.

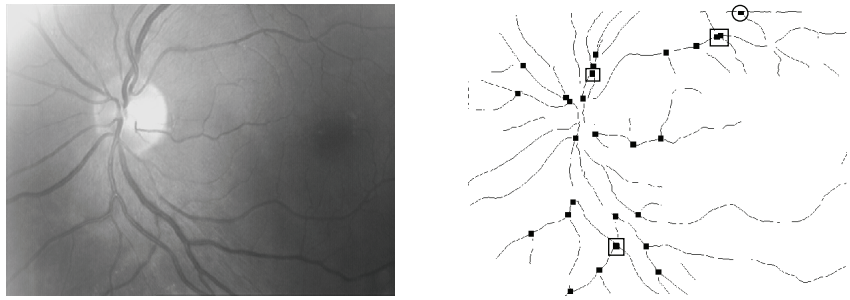


Figure 7: Example of feature points extracted from original image after the bifurcation/crossover stage. (a) Original Image. (b) Feature points marked over the segment image. Spurious points are signalled. Circles surrounding spurious points due to false segments extracted from the image borders and squares surrounding pairs of points corresponding to the same crossover (detected as two bifurcations).

2.2.4. Filtering of Feature Points. As showed in Figure 7(b), the highest feature point detected comes from a bifurcation involving an spurious segment. This segment appears in the creases extraction stage as this algorithm can make some false creases to appear in the image borders.

To avoid these situations, feature points very close to image borders are removed as the vast majority of them correspond to bifurcations involving false segments. A minimum distance to the border threshold of approximately 3% of the width/height of the image is enough to avoid these false features.

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

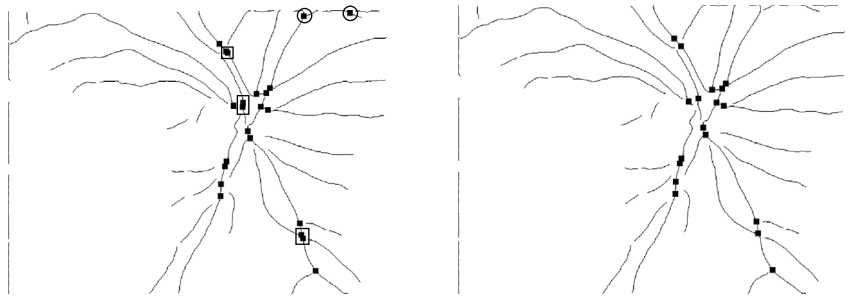
Finally, as crossover points are detected as two bifurcation points, Figure 7(b), these are merged into an unique feature point.

Figure 8 shows an example of the filtering process result, that is, the biometric pattern obtained from an individual. In resume, the average of 5 spurious points per image was reduced to 2 per image after the filtering process. These points are derived from bad extracted regions in the creases stage. The removal of non spurious points with this technique is almost null (around 0.2 points per image in the average).

2.3. Biometric Pattern Matching. In the matching stage, the stored reference pattern, ν, for the claimed identity is compared to the pattern extracted, ν', during the previous stage. Due to the eye movement during the image acquisition stage, it is necessary to align β with α in order to be matched [20–22]. This fact is illustrated in Figure 9 where two images from the same individual, Figures 9(a) and 9(c), and the obtained results in each case, Figures 9(b) and 9(d), are showed.

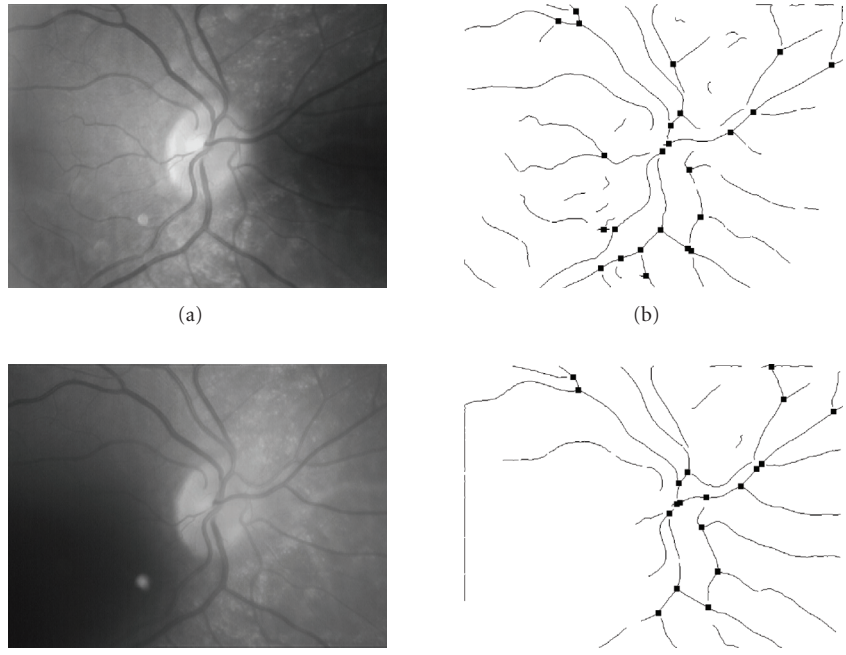
Depending on several factors, such as the eye location in the objective, patterns may suffer some deformations. A reliable and efficient model is necessary to deal with these deformations allowing to transform the candidate pattern in order to get a pattern similar to the reference one. The movement of the eye in the image acquisition process basically consists in translation in both axis, rotation and sometimes a very small change in scale. It is also important to note that both patterns ν and ν' could have a different number of points as seen in Figure 9 where, from the same individual, two patterns are extracted with 24 and 19 points. This is due to the different conditions of illumination and orientation in the image acquisition stage.

The transformation considered in this work is the similarity transformation (ST), which is a special case of the global affine transformation (GAT). ST can model translation, rotation and isotropic scaling using 4 parameters



(a) (b)

Figure 8: Example of the result after the feature points filtering. (a) Image containing feature points before filtering. (b) Image containing feature points after filtering. Spurious points from image borders and duplicate crossover points have been eliminated.



(a) (b)

(c) (d)

Figure 9: Examples of feature points obtained from images of the same individual acquired in different times. (a) (c) Original images. (b) Feature points image from (a). A total of 24 points are obtained. (d) Feature points image from (c). A total of 19 points are obtained.

[23]. The ST works fine with this kind of images as the rotation angle is moderate. It has also been observed that the scaling, due to eye proximity to the camera, is nearly constant for all the images. Also, the rotations are very slight as the eye orientation when facing the camera is very similar. Under these circumstances, the ST model appears to be very suitable.

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

To obtain the four parameters of a concrete ST, two pairs of feature points between the reference and candidate patterns are considered. If M is the total number of feature points in the reference pattern and N the total number of points in the candidate one, the size of the set T of possible transformations is computed using (4):

$$T = \frac{(M^2 - M)(N^2 - N)}{2}, \tag{4}$$

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

Since T represents a high number of transformations, some restrictions must be applied in order to reduce it. As

the scale factor between patterns is always very small in this acquisition process, a constraint can be set to the pairs of points to be associated. In this scenario, the distance between both points in each pattern has to be very similar. As it cannot be assumed that it will be the same, two thresholds are defined, $S_{\min}$ and $S_{\max}$, to bound the scale factor. This way, elements from T are removed where the scale factor is greater or lower than the respective thresholds $S_{\min}$ and $S_{\max}$. However, (5) formalises this restriction:

$$S_{\min} < \frac{\text{distance}(p,q)}{\text{distance}(p',q')} < S_{\max}, \tag{5}$$

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

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

$$\text{SIM}(A,B) = 1 - \frac{\text{distance}(A,B)}{D_{\max}}, \tag{6}$$

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

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

$$P(A_i,B_j) = \frac{\text{SIM}(A_i,B_j)^2}{\left(\sum_{i'=1}^{M}\text{SIM}(A_{i'},B_j)+\sum_{j'=1}^{N}\text{SIM}(A_i,B_{j'})-\text{SIM}(A_i,B_j)\right)}. \tag{7}$$

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

The final set of matched points between patterns is C. Using this information, a similarity metric must be established to obtain a final criterion of comparison between patterns. Performance of several metrics using matched points information is analysed in Section 3.

3. Similarity Metrics Analysis

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

For the metric analysis, a set of 150 images (100 images, 2 images per individual, and 50 different images more) from VARIA database [24] were used. The rest of the images will be used for testing in Section 4. The images from the database have been acquired with a TopCon nonmydriatic camera NW-100 model and are optic disc centred with a resolution of 768×584. There are 60 individuals with two or more images acquired in a time span of 6 years. These images have a high variability in contrast and illumination allowing the system to be tested in quite hard conditions. In order to build the training set of matchings, all images are matched versus all the images (a total of 150×150 matchings) for each metric. The matchings are classified into attacks or clients accesses depending if the images belong to the same individual or not. Distributions of similarity values for both classes are compared in order to analyse the classification capabilities of the metrics.

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

This overlapping is caused by the variability of the patterns size in the training set because of the different illumination and contrast conditions in the acquisition stage. Figure 10(b) shows the histogram for the biometric pattern size, that is, the number of feature points detected. A high variability can be observed, as some patterns have more than twice the number of feature points of other patterns. As a result of this, some patterns have a small size, capping the possible number of matched points (Figure 11). Also, using the matched points information alone lacks a well bounded and normalised metric space.

To combine information of patterns size and normalise the metric, a function f will be used. Normalised metrics are very common as they make easier to compare class separability or establishing valid thresholds [25]. The similarity measure (S) between two patterns will be defined by

$$S = \frac{C}{f(M,N)}, \tag{8}$$

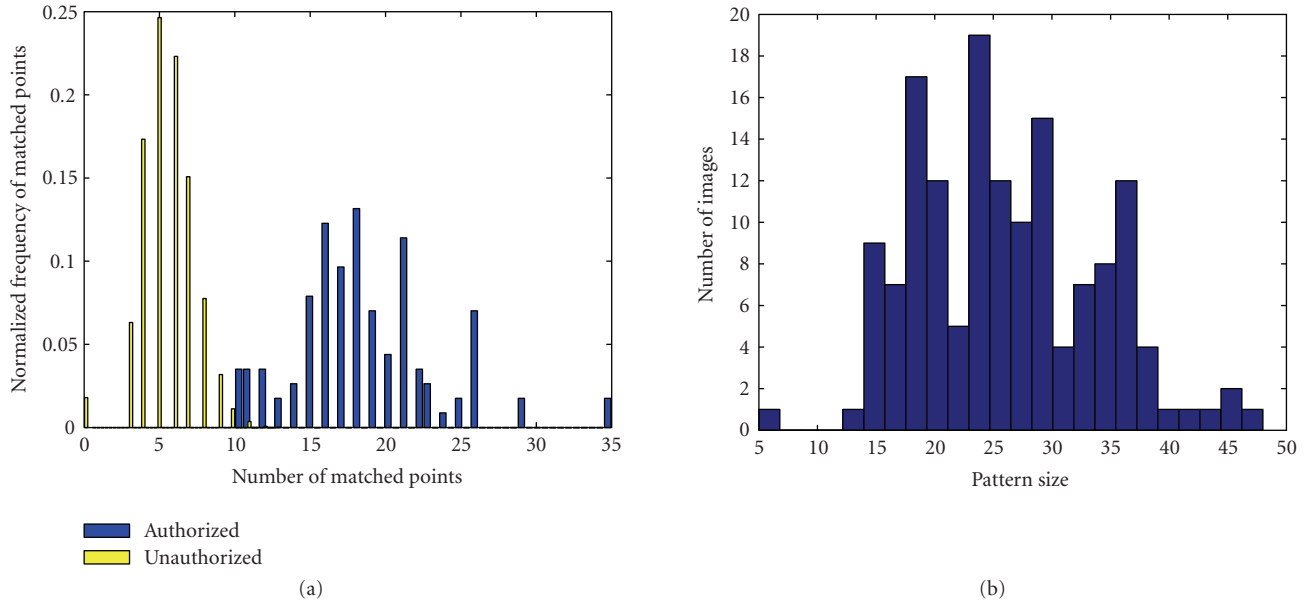


Figure 10: (a) Matched points histogram in the attacks (unauthorised) and clients (authorised) authentications cases. In the interval [10, 13] both distributions overlap. (b) histogram of detected points for the patterns extracted from the training set.

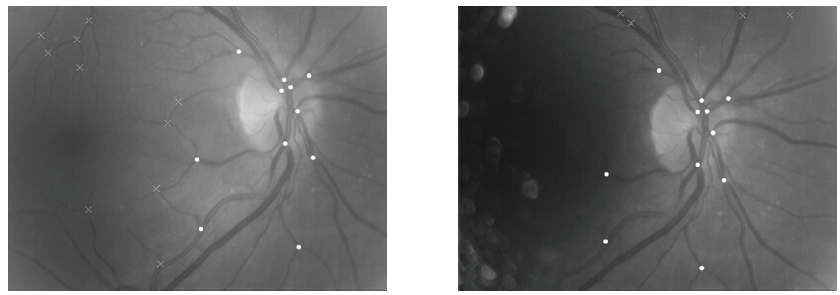


Figure 11: Example of matching between two samples from the same individual in VARIA database. White circles mark the matched points between both images while crosses mark the unmatched points. In (b) the illumination conditions of the image lead to miss some features from left region of the image. Therefore, a small amount of detected feature points is obtained capping the total amount of matched points.

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

$$f(M,N) = \min(M,N). \tag{9}$$

The min function is the less conservative one as it allows to obtain a maximum similarity even in cases of different sized patterns. Figure 12(a) shows the distributions of similarity scores for clients and attacks classes in the training set using the normalisation function defined in (9), and Figure 12(b) shows the FAR and FRR curves versus the decision threshold.

Although the results are good when using the normalisation function defined in (9), a few cases of attacks show high similarity values, overlapping with the clients class. This is caused by matchings involving patterns with a low number of feature points as $\min(M,N)$ will be very small, needing only a few points to match in order to get a high similarity value. This suggests, as it will be reviewed in Section 4, that some minimum quality constraint in terms of detected points would improve performance for this metric.

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

$$f(M,N) = \sqrt{MN}. \tag{10}$$

Figure 13(a) shows the distributions of similarity scores for clients and attacks classes in the training set using the normalisation function defined in (10) and Figure 13(b) shows the FAR and FRR curves versus the decision threshold.

Function defined in (10) combines both pattern sizes in a more conservative way, preventing the system to obtain a high similarity value if one pattern in the matching process contains a low number of points. This allows to reduce the attacks class variability and, moreover, to separate its values away from the clients class as this class remains in a similar values range. As a result of the new attacks class boundaries,

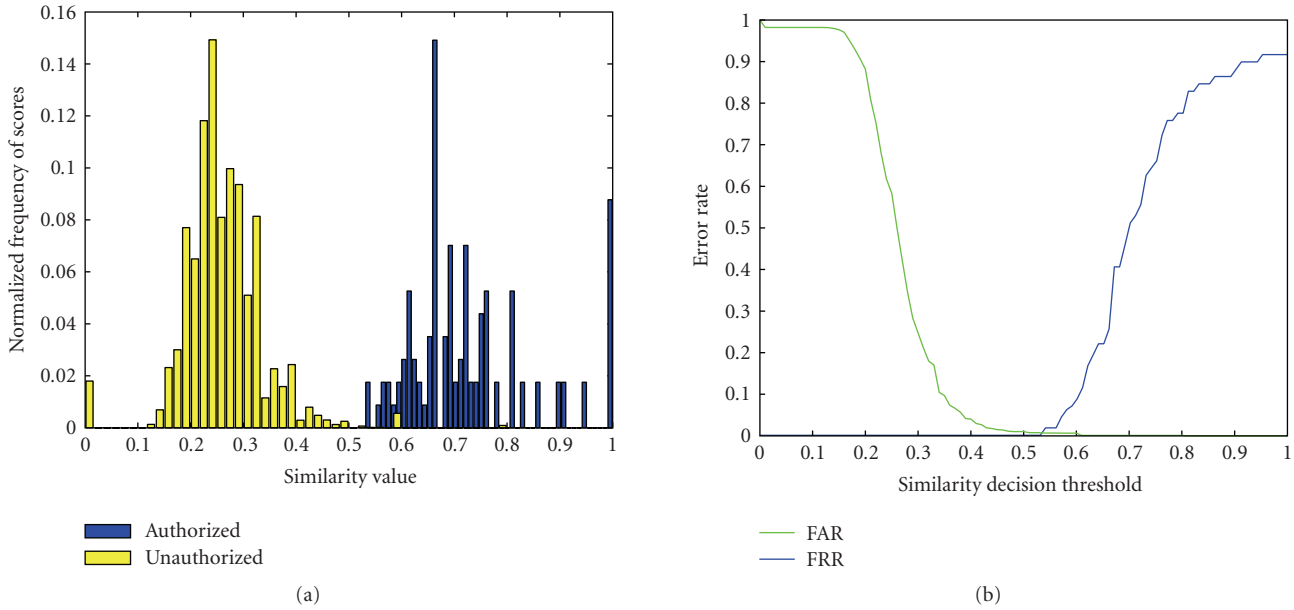


Figure 12: (a) Similarity values distribution for authorised and unauthorised accesses using $f = \min(M, N)$ as normalisation function for the metric. (b) False accept rate (FAR) and false rejection rate (FRR) for the same metric.

a decision threshold can be safely established where FAR = FRR = 0 in the interval $[0.38, 0.5]$ as Figure 13(b) clearly exposes. Although this metric shows good results, it also has some issues due to the normalisation process which can be corrected to improve the results as showed in next subsection.

3.1. Confidence Band Improvement. Normalising the metric has the side effect of reducing the similarity between patterns of the same individual where one of them had a much greater number of points than the other, even in cases with a high number of matched points. This means that some cases easily distinguishable based on the number of matched points are now near the confidence band borders. To take a closer look at this region surrounding the confidence band, the cases of unauthorised accesses with the highest similarity values (S) and authorised accesses with the lowest ones are evaluated. Figure 14 shows the histogram of matched points for cases in the marked region of Figure 13(b). It can be observed that there is an overlapping but both histograms are highly distinguishable.

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

$$S_\gamma = S \cdot C^{\gamma-1} = \frac{C^\gamma}{\sqrt{MN}} \tag{11}$$

with S, C, M, and N the same parameters from (10). The γ correction parameter allows to improve the similarity values when a high number of matched points is obtained, specially in cases of patterns with a high number of points.

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

$$T(x) = \frac{1}{1 + e^{s\cdot(x-0.5)}}, \tag{12}$$

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

$$S'_\gamma = T(S_\gamma). \tag{13}$$

Finally, to choose a good γ parameter, the confidence band improvement has been evaluated for different values of γ (Figure 15(a)). The maximum improvement is achieved at $\gamma = 1.12$ with a confidence band of 0.3288, much higher than the original from previous section. The distribution of the whole training set (using $\gamma = 1.12$) is showed in Figure 15(b) where the wide separation between classes can be observed.

4. Results

A set of 90 images, 83 different from the training set, and 7 from the previous set with the highest number of points, has been built in order to test the metrics performance once their parameters have been fixed with the training set. To test the metrics performance, the false acceptance rate and false rejection rate were calculated for each of them (the metrics normalised by (9), (10) and the gamma-corrected normalised metric defined in (13).

A usual error measure is the equal error rate (EER) that indicates the error rate where FAR curve and FRR curve

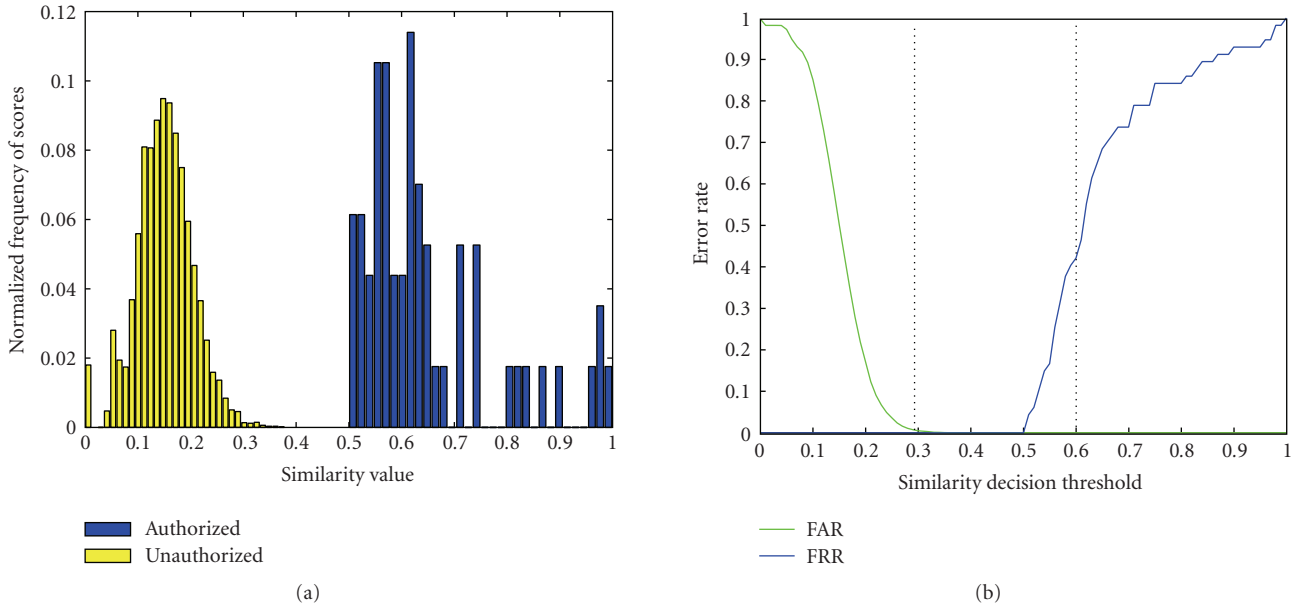


Figure 13: (a) Similarity values distribution for authorised and unauthorised accesses using $f = \sqrt{MN}$ as normalisation function for the metric. (b) False accept rate (FAR) and false rejection rate (FRR) for the same metric. Dotted lines delimit the interest zone surrounding the confidence band which will be used for further analysis.

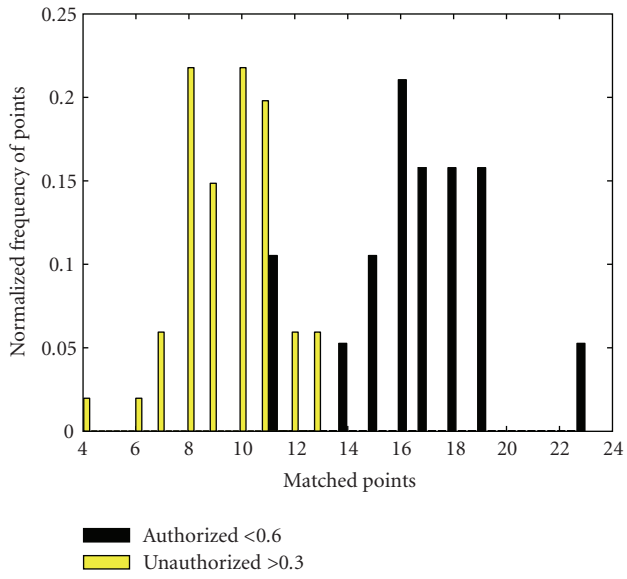


Figure 14: Histogram of matched points in the populations of attacks whose similarity is higher than 0.3 and clients accesses whose similarity is lower than 0.6.

intersect. Figure 16(a) shows the FAR and FRR curves for the three previously specified metrics. The EER is 0 for the normalised by geometrical mean (mean) and gamma corrected (gamma) metrics as it was the same case in the training set, and, again, the gamma corrected metric shows the highest confidence band in the test set 0.2337.

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

Finally, to evaluate the influence of the image quality, in terms of feature points detected per image, a test is run where images with a biometric pattern size below a threshold are removed for the set and the confidence band obtained with the rest of the images is evaluated. Figure 16(b) shows the evolution of the confidence band versus the minimum detected points constraint. The confidence band does not grow significatively until a fairly high threshold is set. Taking as threshold the mean value of detected points for all the test set, 25.2, the confidence band grows from 0.2337 to 0.3317. So removing half of the images, the band is increased only by 0.098 suggesting that the gamma-corrected metric is very robust to low quality images.

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

5. Conclusions and Future Work

In this work, a complete identity verification method has been introduced. Following the same idea as the fingerprint minutiae-based methods, a set of feature points is extracted from digital retinal images. This unique pattern will allow for the reliable authentication of authorised users. To get the set of feature points, a creases-based extraction algorithm is used. After that, a recursive algorithm gets the point features by tracking the creases from the localised optic disc. Finally, a registration process is necessary in order

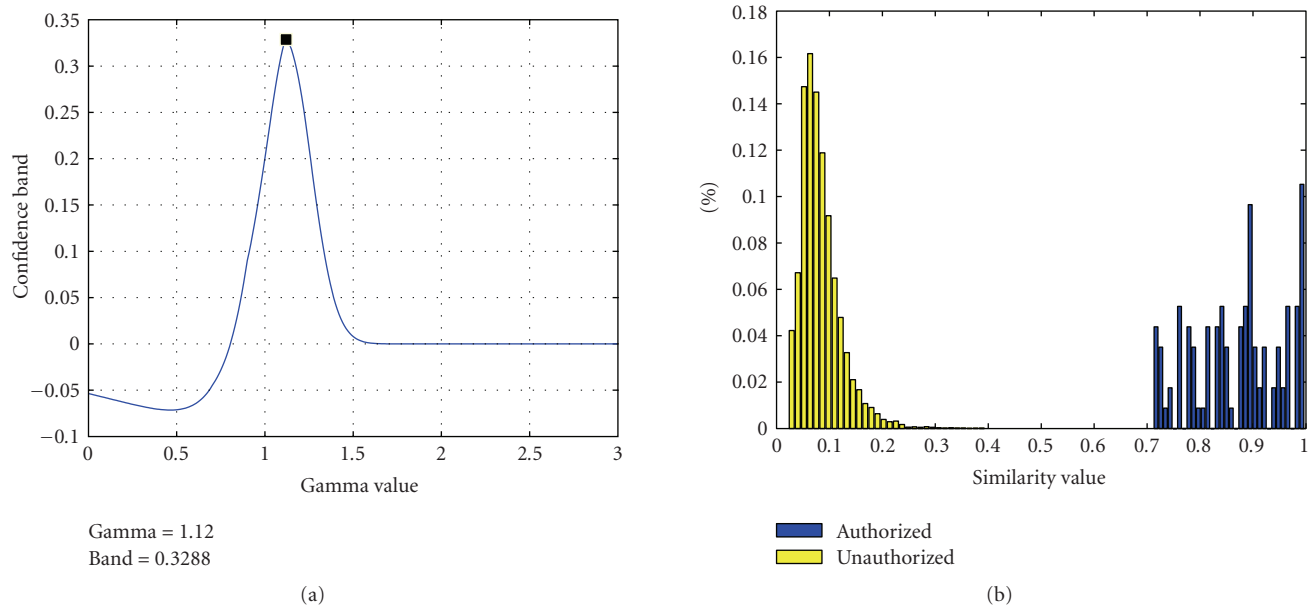


Figure 15: (a) Confidence band size versus gamma (γ) parameter value. Maximum band is obtained at $\gamma = 1.12$. (b) Similarity values distributions using the normalised metric with $\gamma = 1.12$.

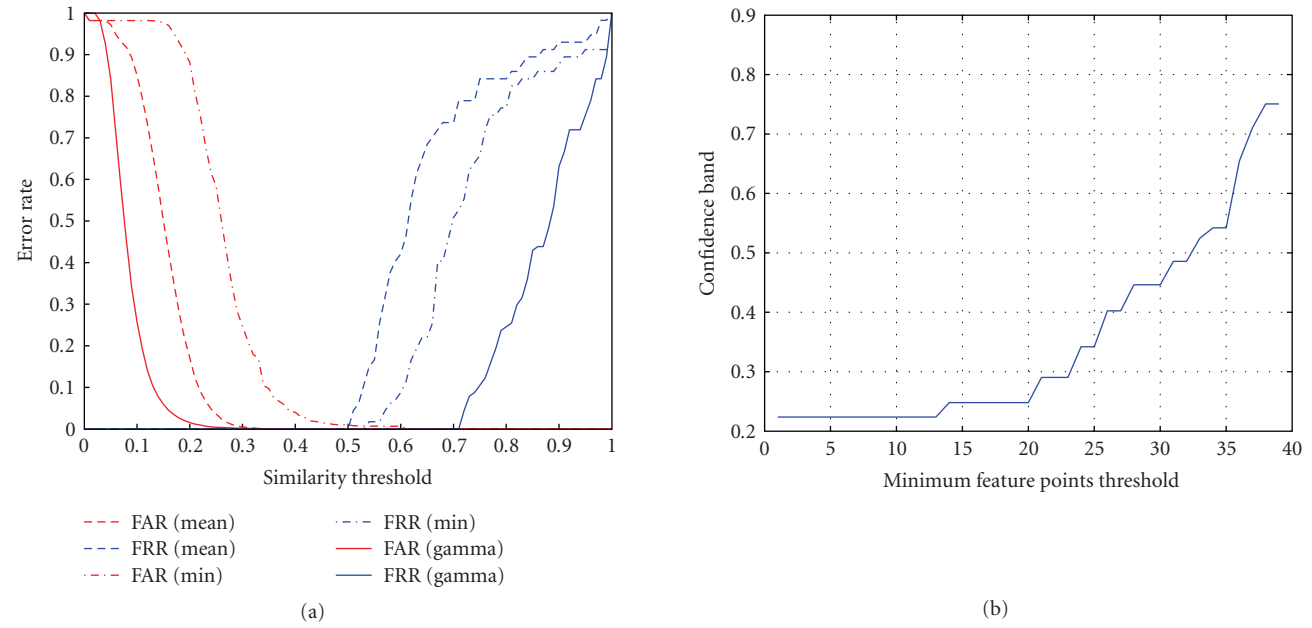


Figure 16: (a) FAR and FRR curves for the normalised similarity metrics (min: normalised by minimum points, mean: normalised by geometrical mean, and gamma: gamma corrected metric). The best confidence band is the one belonging to the gamma corrected metric corresponding to 0.2337.(b) Evolution of the confidence band using a threshold of minimum detected points per pattern.

to match the reference pattern from the database and the acquired one. With the patterns aligned, it is possible to measure the degree of similarity by means of a similarity metric. Normalised metrics have been defined and analysed in order to test the classification capabilities of the system. The results are very good and prove that the defined authentication process is suitable and reliable for the task. The use of feature points to characterise individuals is a robust biometric pattern allowing to define metrics that offer a good confidence band even in unconstrained environments when the image quality variance can be very high in terms of distortion, illumination, or definition. This is also possible as this methodology does not rely on the localisation or segmentation of some reference structures, as it might be the optic disc. Thus, if the the user suffers some structure-distorting pathology and this structure cannot be detected,

the system works the same with the only problem being a possible loss of feature points constrained to that region.

Future work includes the use of some high-level information of points to complement metrics performance and new ways of codification of the biometric pattern allowing to perform faster matches.

Acknowledgment

This paper has been partly funded by the Xunta de Galicia through the grant contracts PGIDIT06TIC10502PR.

References

[1] J. G. Daugman, "Biometric personal identification system based on iris analysis," US patent no. 5291560, 1994.

[2] Retica Systems, "Iris-Retinal multimodal identification," http://www.retica.com/site/technology/irisretina.html.

[3] Digital Persona, "Fingerprint solutions," http://www.digitalpersona.com/index.php.

[4] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity-authentication system using fingerprints," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1365–1388, 1997.

[5] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 3–18, 2006.

[6] R. Zunkel, "Hand geometry based verification," in *BIOMETRICS: Personal Identification in Networked Society*, pp. 87–101, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[7] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: a literature survey," Tech. Rep., National Institute of Standards and Technology, Gaithersburg, Md, USA, 2000.

[8] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: a survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[9] J. Bigun, C. Chollet, and G. Borgefors, Eds., *Proceedings of the 1st International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 1997.

[10] L. Ballard, D. Lopresti, and F. Monrose, "Forgery quality and its implications for behavioral biometric security," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 37, no. 5, pp. 1107–1118, 2007.

[11] R. Cappelli, A. Lumini, D. Maio, and D. Maltoni, "Fingerprint image reconstruction from standard templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1489–1503, 2007.

[12] S. C. Dass, Y. Zhu, and A. K. Jain, "Validating a biometric authentication system: sample size requirements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1902–1913, 2006.

[13] C. Mariño, M. G. Penedo, M. Penas, M. J. Carreira, and F. González, "Personal authentication using digital retinal images," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 21–33, 2006.

[14] C. Mariño, M. G. Penedo, M. J. Carreira, and F. González, "Retinal angiography based authentication," in *Proceedings of the 8th Iberoamerican Congress on Pattern Recognition (CIARP '03)*, vol. 2905 of *Lecture Notes in Computer Science*, pp. 306–313, Havana, Cuba, November 2003.

[15] H. Farzin, H. Abrishami-Moghaddam, and M.-S. Moin, "A novel retinal identification system," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 280635, 10 pages, 2008.

[16] X. Tan and B. Bhanu, "A robust two step approach for fingerprint identification," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2127–2134, 2003.

[17] M. Ortega, C. Mariño, M. G. Penedo, M. Blanco, and F. González, "Personal authentication based on feature extraction and optic nerve location in digital retinal images," *WSEAS Transactions on Computers*, vol. 5, no. 6, pp. 1169–1176, 2006.

[18] A. M. López, D. Lloret, J. Serrat, and J. J. Villanueva, "Multilocal creaseness based on the level-set extrinsic curvature," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 111–144, 2000.

[19] A. M. Löpez, F. Lumbreras, J. Serrât, and J. J. Villanueva, "Evaluation of methods for ridge and valley detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 327–335, 1999.

[20] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[21] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[22] M. S. Markov, H. G. Rylander III, and A. J. Welch, "Real-time algorithm for retinal tracking," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 12, pp. 1269–1281, 1993.

[23] N. Ryan, C. Heneghan, and P. de Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization," *Image and Vision Computing*, vol. 22, no. 11, pp. 883–898, 2004.

[24] VARIA, "VARPA Retinal images for authentication," http://www.varpa.es/varia.html.

[25] M. Tico and P. Kuosmanen, "Fingerprint matching using an orientation-based minutia descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1009–1014, 2003.

Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 243215, 13 pages
doi:10.1155/2009/243215

Research Article

A Sequential Procedure for Individual Identity Verification Using ECG

John M. Irvine[1] and Steven A. Israel[2]

[1] *Advanced Signal Processing and Image Exploitation Group, Draper Laboratory, 555 Technology Square, MS 15, Cambridge, MA 02139, USA*
[2] *Systems and Technology Division, SAIC, 4001 Fairfax Drive, Suite 450, Arlington, VA 22203, USA*

Correspondence should be addressed to Steven A. Israel, steven.a.israel@saic.com

Received 20 October 2008; Revised 14 January 2009; Accepted 24 March 2009

Recommended by Kevin Bowyer

The electrocardiogram (ECG) is an emerging novel biometric for human identification. One challenge for the practical use of ECG as a biometric is minimizing the time needed to acquire user data. We present a methodology for identity *verification* that quantifies the minimum number of heartbeats required to authenticate an enrolled individual. The approach rests on the statistical theory of sequential procedures. The procedure extracts fiducial features from each heartbeat to compute the test statistics. Sampling of heartbeats continues until a decision is reached—either verifying that the acquired ECG matches the stored credentials of the individual or that the ECG clearly does not match the stored credentials for the declared identity. We present the mathematical formulation of the sequential procedure and illustrate the performance with measured data. The initial test was performed on a limited population, twenty-nine individuals. The sequential procedure arrives at the correct decision in fifteen heartbeats or fewer in all but one instance and in most cases the decision is reached with half as many heartbeats. Analysis of an additional 75 subjects measured under different conditions indicates similar performance. Issues of generalizing beyond the laboratory setting are discussed and several avenues for future investigation are identified.

1. Introduction

The biometric verification process can be broken into five major functional blocks: data collection, signal processing, feature extraction, comparison (database lookup), and returning a decision (Figure 1). Verification systems have two competing requirements: (1) quickly processing samples and returning a decision to minimize the user time, and (2) operate at very high probability of detections (Pds) with low false alarm rates (FARs). With the advances in computing, the longest duration function in Figure 1's critical path is the data collection. This paper presents a method for quantifying the minimum number of heartbeats required for verifying the identity of an individual from the electrocardiogram (ECG) signal. The minimum number of heartbeats required provides a user-centric measure of performance for an identity verification system. The outcome of our research forms the basis for selecting elements of an operational ECG verification system.

Since 2001, researchers have identified unique characteristics of the ECG trace for biometric verification, particularly with respect to access control [1–14]. To illustrate, consider the heartbeats from several different individuals (Figure 2). Although each heartbeat follows the same general pattern, differences in the detailed shape of the heartbeat are evident. We exploit these shape differences across individuals to perform identity verification. The last 30 years have witnessed substantial research into the collection and processing of digital ECG signals [15–17]. In addition, a special issue of this journal was devoted to "Advances in electrocardiogram signal processing and analysis" in 2007. We build on this wealth of information and apply it to the development of an ECG verification system.

Hawkins [18] revealed that the traditional biometrics of face, fingerprints, and iris can be forged. The traditional biometrics cited above contain no inherent measure of liveness. The ECG, however, is inherently an indication of liveness and, consequently, is difficult to falsify. Israel et al.

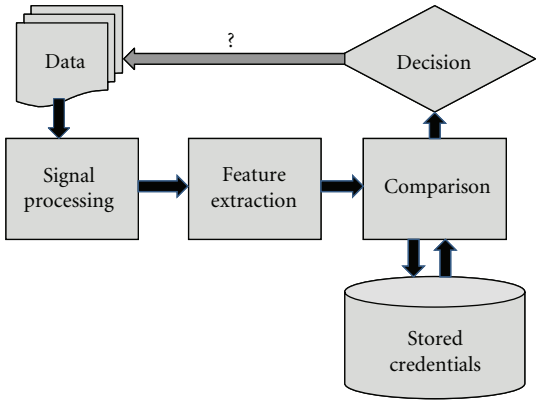


Figure 1: Simplified architecture for an authentication system.

[6] analyzed other cardiovascular modalities and found ECG data most discriminating for human identification.

This paper illustrates a methodology and minimum heartbeat performance metric using data and processing from our previously published research [1–7]. This work extends previous results in two ways. First, it focuses on the identity verification problem, such as would be appropriate for portal access. Second, the method developed here quantifies the minimum number of heartbeats needed for identity verification, thereby fixing the time needed to collect user data. The next section summarizes the utility of applying ECG information as a biometric. The following two sections present the actual methodology, first discussing the processing of the ECG signal and then deriving the actual test statistic used for identity verification. We present results from two data sets to illustrate performance. The final section discusses a number of practical issues related to ECG as a biometric and suggests avenues for further investigation.

2. Background

This paper presents a new approach for processing the ECG for identity verification based on sequential procedures. A major challenge for developing biometric systems based on circulatory function is the dynamic nature of the raw data. Heartrate varies with the subject's physical, mental, and emotional state, yet a robust biometric must be invariant across time and state of anxiety. The heartbeat maintains its structure with changes in heartrate (Figure 2). Irvine et al. [1–3, 5], Israel et al. [4, 6], and Biel et al. [8] identified individuals based upon features extracted from individual heartbeats. Wang et al. [9] followed a similar approach using fiducial features, but then extended the analysis based on a discrete cosine transform (DCT) of the autocorrelation function. Shen et al. [10] and Wubbeler et al. [14] employed a template matching approaches. Additional nonfiducial techniques have exploited principal components analysis (PCA) [19–27] in the same manner as [28] applied to face. Recently, a number of researchers have explored improvements to representations of the ECG signal for human identification [5, 9, 29]. In each case, the extracted ECG attributes performed well for identifying individuals.

Early studies of ECG feature extraction used spectral features to characterize the heartbeat [17]. Later, Biel et al. [8] performed ECG feature extraction by estimating location and magnitude information. Irvine et al. [2] showed that the relative electrode position caused changes in the magnitude of the ECG traces and used only temporal features. To these ends, Israel et al. [4] identified additional fiducial positions to characterize the relative intervals of the heartbeat and performed quantitative feature extraction using radius of curvature features.

Initial experiments for human identification from ECG identified some important challenges to overcome. First, approaches that rely on fiducial attributes, that is, features obtained by identifying specific landmarks from the processed signal have difficulty handling nonstandard heartbeats and high noise floors. Agrafioti and Hatzinakos [30] applied signal processing methods to address common cardiac irregularities. A second challenge is to insure that the identification procedure is robust to changes in the heartrate arising from varying mental and emotional states. Irvine et al. [1–3] and Israel et al. [4, 6] addressed this issue through an experimental protocol that varied the tasks performed by the subjects during data collection. Third, PCA type algorithms must sample a sufficiently wide population to ensure the best generalization of their eigen features.

The ECG measures the electrical potential at the surface of the body as it relates to the activation of the heart. Many excellent references describe the functioning of the heart and the factors affecting the ECG signal [15, 31, 32]. Because the ECG consists of repeated heartbeats, the natural period of the signal is amenable to a wealth of techniques for statistical modeling. We exploit this periodic structure, treating the heartbeat as the basic sampling unit for constructing the sequential method.

3. Signal Processing

We segmented the data into two nonoverlapping, block segmented by time, groups. Group 1 is the training data, where labeled heartbeats are used to generate statistics about each enrolled individual. Group 2 is the test data, which contain heartbeats from the sensor and have known *a posteriori* labels. The computational decision from the system is either a confirmation that the individual is who they say they are; or a rejection that the individual is not who they say they are.

Processing of the ECG signal includes noise reduction, segmentation of the heartbeats, and extraction of the features from each heartbeat (Figure 3). Because the objective is to minimize the data acquisition time for identity verification, the enrollment time was not constrained. Two minutes of data were used for enrollment and to train the verification functions for each individual. Two additional minutes of test data were available to quantify the required number of heartbeats. For our concept of operations, however, the individuals seeking authentication would only need to present the minimum number of heartbeats, which is expected to be on the order of second(s).

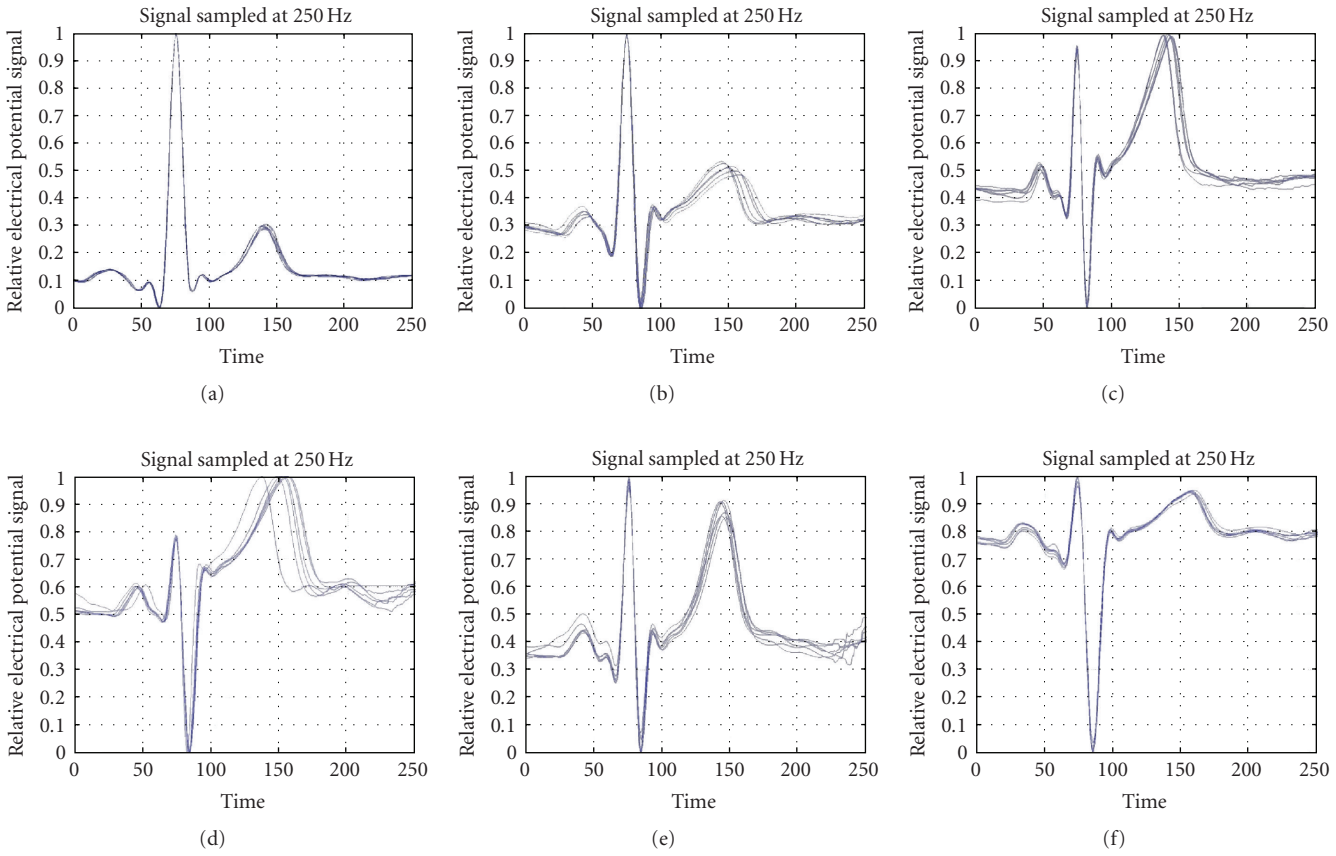


Figure 2: Segmented heartbeats from six individuals.

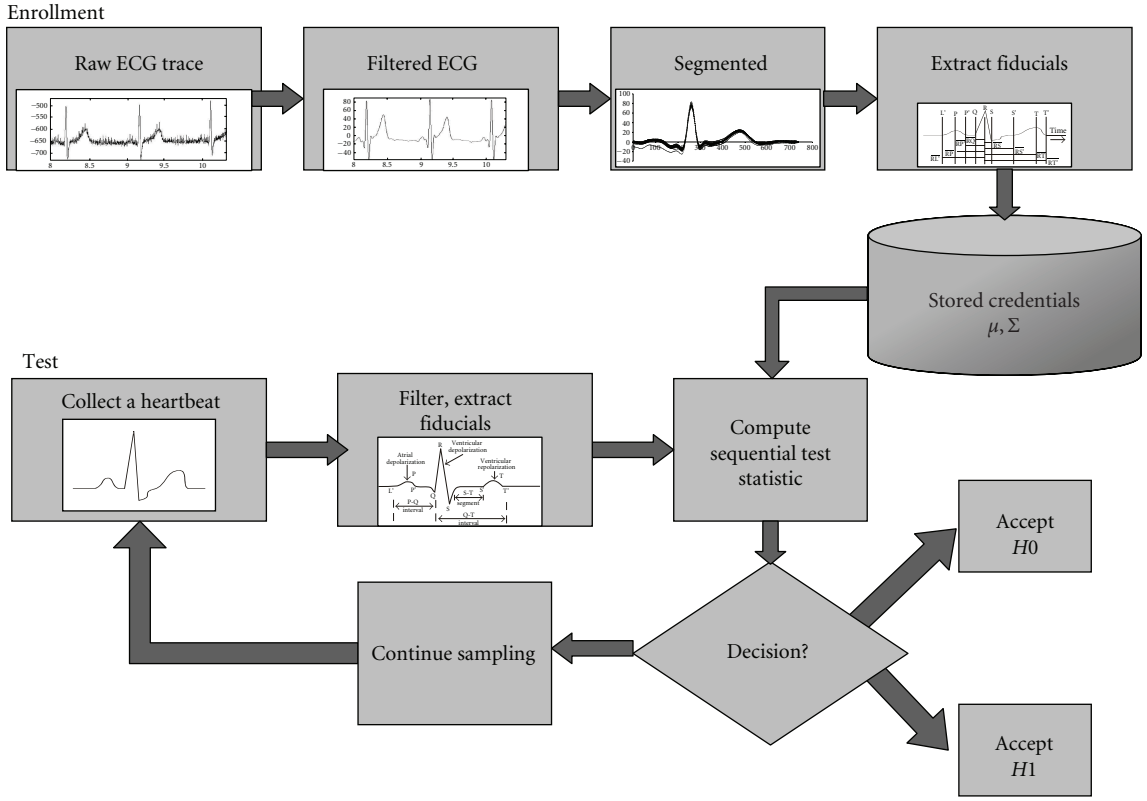


Figure 3: Signal processing for the sequential procedure.

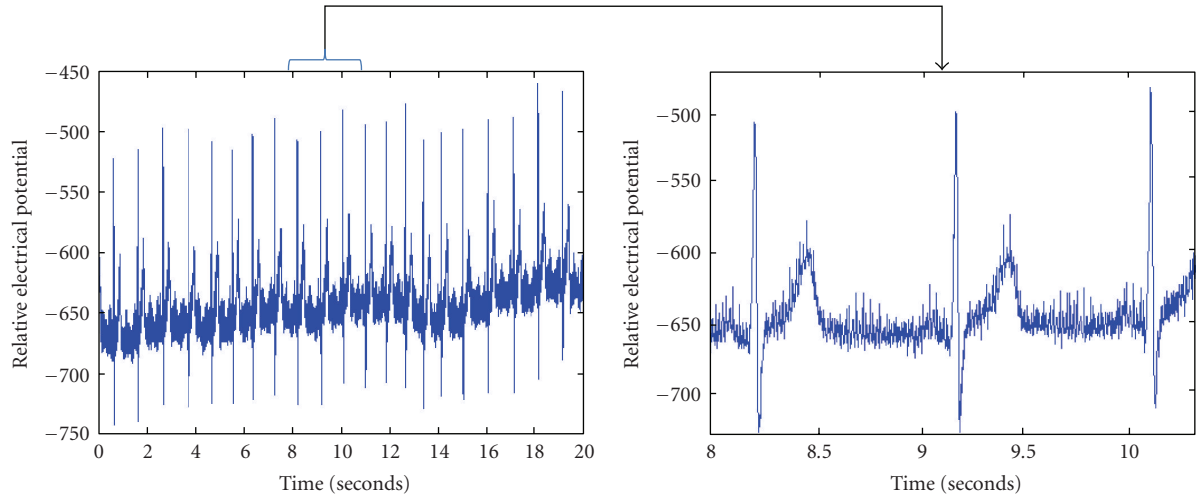


Figure 4: Raw ECG data 1000 Hz (a) 20 seconds (b) 2 seconds.

Figures 4(a) and 4(b) show a sample of high resolution ECG data. The raw data contain both high and low frequency noise components. These noise components alter the expression of the ECG trace from its ideal structure. The low frequency noise is expressed as the slope of the overall signal across multiple heartbeat traces in Figure 4(a). The low frequency noise is generally associated with changes in baseline electrical potential of the device and is slowly varying. Over this 20-second segment, the ECG can exhibit a slowly varying cyclical pattern, associated with respiration, that is known as *sinus arrhythmia* [15]. The high frequency noise is expressed as the intrabeat noise shown in Figure 4(b). The high frequency noise is associated with electric/magnetic field of the building power (electrical noise) and the digitization of the analog potential signal (A/D noise). Additionally, evidence of subject motion and muscle flexure must be removed from the raw traces [33].

Multiple filtering techniques have been applied to the raw ECG traces: heartbeat averaging [34, 35], wavelet [36, 37], least squares modeling [38, 39] and Fourier bandpass filtering [40–42]. For any filtering technique, the design constraints are to maintain as much of the subject-dependent information (signal) as possible and design a stable filter across all subjects.

As previously reported [4], the raw ECG traces were bandpass filtered between 0.2 and 40 Hz. The filter was written with a lower order polynomial to reduce edge effects. Figure 5(a) illustrates the power spectra from a typical 1000 Hz ECG trace. The noise sources were identified, and our notional bandpass filter overlays the power spectrum. Figure 5(b) shows the power spectrum after the bandpass filtering. Figure 6 contains the processed data for heartbeat segmentation and feature extraction.

Commonly, heartbeat segmentation is performed by first locating the R complex. Next, the R position is estimated for the following heartbeats [43, 44]. Our R peak locator used a simple technique of looking at the maximum variance over a 0.2 second interval. The 0.2-seconds represent ventricular depolarization. The metric was computed in overlapping windows to insure that the true R peaks were recovered [4]. The remainder of the heartbeat was realized by locating the P and the T peaks relative to the R position. For the enrollment data, we used autocorrelation techniques to develop an initial estimate of the R-R interval. In the autocorrelation function, the lag for the maximum peak generally corresponds to the mean length of the heartbeat, giving an initial value to guide the heartbeat segmentation.

ECG data are commonly collected by contact sensors at multiple positions around the heart. The change in ECG electrode position provides different information because of the relative position to the heart's plane of zero potential. For nearly all individuals and all electrode locations, the ECG trace of a heartbeat produces three complexes (wave forms). The medical community has defined the complexes by their peaks: P, R, and T (Figure 7). The R-R interval, the time between two successive R peaks, indicates the duration of a heartbeat. Two other fiducials, Q and S, are also identified at the base of the R complex. Israel et al. [4] identified four additional fiducials at the base of the P and T complexes. These are noted with a prime ($'$) symbol (Figure 7). We employ the single channel feature extraction method developed by Israel et al. [4]. The nine features derived from the fiducials are the feature vector used to illustrate the sequential procedure and the minimum number of heartbeats metric.

4. The Sequential Procedure

Abraham Wald developed the sequential procedure for formal statistical testing of hypotheses in situations where data can be collected incrementally [45, 46]. In many instances, the sequential method arrives at a decision based on relatively

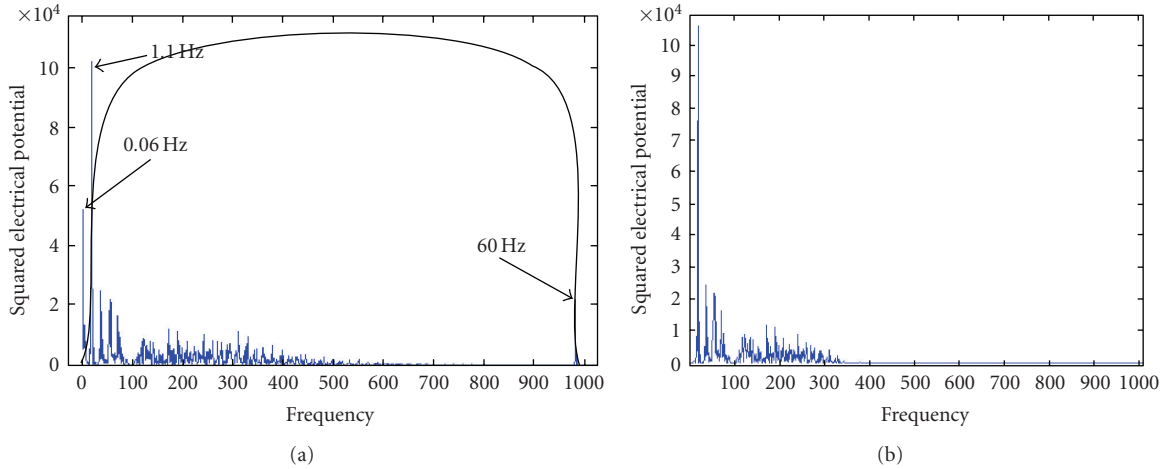


FIGURE 5: Power spectra of frequency filtering: (a) bandpass filter of raw data (b) frequency response of filtered data. (a) shows the noise source spikes at 0.06 and 60 Hz and the information spikes between 1.10 and 35 Hz. (b) shows the filtered data with the noise spikes removed and the subject specific information sources retained. The X-axis is frequency in Hz, and the Y-axis is squared electrical potential.

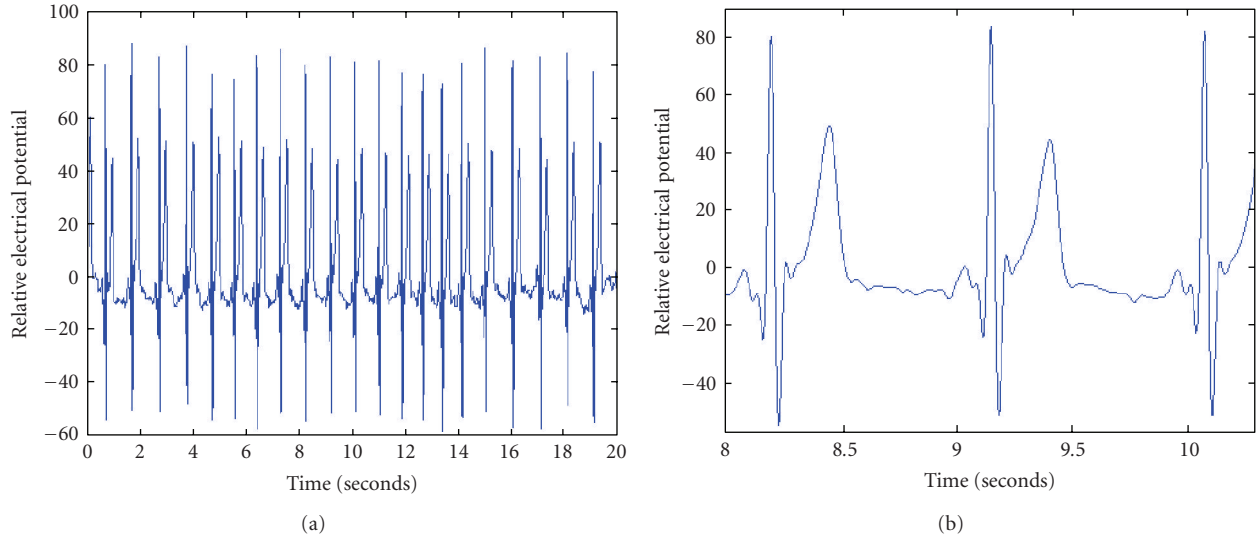


FIGURE 6: Bandpass filtered ECG trace (a) entire range of data (b) segment of data. The results of applying the filter (Figure 5) to the raw (Figure 4) data are shown.

few observations. Consider a sequence of independent and identically distributed random variables $\{X_1, X_2, \ldots\}$ and suppose we wish to test the hypothesis $H0 : X_i \sim f(X, \theta_0)$ against the alternative $H1 : X_i \sim f(X, \theta_1)$. The general approach is to construct the sequential probability ratio statistic for the first T observations:

$$S(T) = \frac{P[X_1, \ldots, X_T \mid H_1]}{P[X_1, \ldots, X_T \mid H_0]} = \frac{\prod_{t=1}^{T} f(X_t, \theta_1)}{\prod_{t=1}^{T} f(X_t, \theta_0)}. \quad (1)$$

At each step in the sequential procedure, that is, for each value of $T = 1, 2, \ldots$, the computed value of $S(T)$ is compared to the decision thresholds A and B, where $0 < A < 1 < B < \infty$. The values of A and B depend on the acceptable level of error in the test of hypothesis. The decision procedure is

$$\begin{aligned} &\text{If } S(T) < A, \text{accept } H0, \\ &\text{If } S(T) > B, \text{accept } H1, \\ &\text{If } A < S(T) < B, \text{continue sampling,} \end{aligned} \quad (2)$$

$S(T)$ is known as the sequential probability ratio statistic. It is often convenient to formulate the procedure in terms of the log of the test statistic:

$$\begin{aligned} S^*(T) &= \log [S(T)] \\ &= \sum_{t=1}^{T} \log [f(X_t, \theta_1)] - \sum_{t=1}^{T} \log [f(X_t, \theta_0)]. \end{aligned} \quad (3)$$

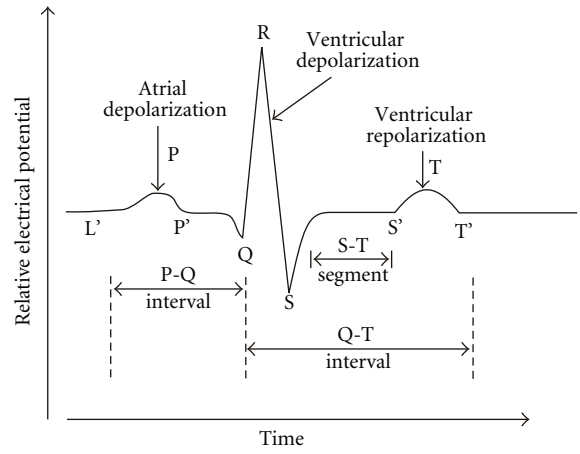


Figure 7: Fiducial features in the heartbeat.

To develop the sequential procedure for our application, we treat identity verification as a test of hypotheses. The two hypotheses are

$$\begin{aligned} H0: &\text{ The subject is who (s)he says} \\ H1: &\text{ The subject is not who (s)he says.} \end{aligned} \tag{4}$$

The data for testing the hypotheses is the series of observed heartbeats presented in the test data. From each test heartbeat the fiducial features are extracted, forming a feature vector. Denote these feature vectors from each heartbeat by $\{H(1), H(2), \ldots\}$. If the person says (s)he is subject i, then $\{H(t) : t = 1, \ldots, T\}$ are drawn from the statistical distribution corresponding to subject i. If (s)he is not who (s)he claims to be, then $\{H(t) : t = 1, \ldots, T\}$ are drawn from a population with a statistical distribution corresponding to subject j, where $i \neq j$. To simplify the procedure, we assume that the feature vectors $\{H(t)\}$ are independent with a K-variate Gaussian distribution, where K is the number of features extracted from each heartbeat. The mean vectors and covariance matrices are estimated from the enrollment data. Using this model for the test data, the hypotheses are restated in statistical terms:

$$\begin{aligned} &H0: \ H(t) \sim \mathcal{N}(Y_i, \Sigma) \quad \text{for } \{H(t) : t = 1, \ldots, T\}, \\ &H1: H(t) \sim \mathcal{N}\left(\mathrm{Y_j}, \Sigma\right) \\ &\qquad\qquad \text{where } i \neq j \text{ for } \{H(t) : t = 1, \ldots, T\}, \end{aligned} \tag{5}$$

where Y_i is the mean feature vector for subject i, and Y_j is the mean feature vector for subject j. The covariance matrix Σ is assumed to be the same across subjects. Implicit in this formulation is the assumption that the $Y_i \neq Y_j$, whenever $i \neq j$, which is a necessary condition for ECG to provide a unique biometric signature. The distance between Y_i and Y_j sets the trade space for selecting ECG attributes and verification algorithms, as it affects the required number of heartbeats needed for making a decision whether the individual is an authentic user or an intruder.

To test the hypotheses $H0$ and $H1$, we calculate the log of the likelihood ratio statistic for whether the first T heartbeats for subject i come from the jth subject. In the classical Neyman-Pearson formulation of hypothesis testing, T would be fixed [47]. In the sequential procedure, we calculate the test statistic for values of T until a decision is reached. Note that the verification methods depend on the Mahalanobis distance, and Y is composed the 9-attribute feature vector. The test statistic as a function of T, the number of heartbeats is:

$$S^*(T) = \sum_{t=1}^{T} \log\left(f_1(H(t))\right) - \log\left(f_0(H(t))\right), \tag{6}$$

where

$$\begin{aligned} f_0(H(t)) &= [2\pi]^{-K/2} |\Sigma|^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2}(H(t) - Y_i)^T \Sigma^{-1} (H(t) - Y_i)\right], \\ f_1(H(t)) &= [2\pi]^{-K/2} |\Sigma|^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2}\left(H(t) - Y_j\right)^T \Sigma^{-1} \left(H(t) - Y_j\right)\right], \end{aligned} \tag{7}$$

where $K = 9$ is the dimensionality of the vectors, and Y_j is the mean for the alternative hypothesis. In principle, we would calculate the statistic $S^*(T)$ for each value of T, starting at $T = 1$. For ECG analysis, at least two heartbeats are required. The features are the distances between fiducial points, normalized by the length of the heartbeat. This normalization insures that the verification procedure is tolerant to changes in overall heartrate attributable to varying physical, mental, or emotional state.

Computing $S^*(T)$ requires calculating (7) for each heartbeat, multiplying, and taking logs to compute the value defined in (6). Computationally, this can be simplified. The term $[2\pi]^{-k/2} |\Sigma|^{-1/2}$ is a constant that gets added and subtracted, so it can be ignored. The test procedure simplifies to calculate the quadratic forms (8) and (9):

$$\begin{aligned} f_0(H(t)) &\propto -\frac{1}{2}(H(t) - Y_i)^T \Sigma^{-1} (H(t) - Y_i), \\ f_1(H(t)) &\propto -\frac{1}{2}\left(H(t) - Y_j\right)^T \Sigma^{-1} \left(H(t) - Y_j\right). \end{aligned} \tag{8}$$

Sum up values to compute $S^*(T)$ for each value of T, that is,

$$S(T) = \sum_{t=1}^{T} \log[f_1(H(t))] - \log[f_0(H(t))]. \tag{9}$$

The result of all this is a series of values for $S^*(T)$ for $T = 1, 2, 3, \ldots$. Because the feature vector $H(t)$ depends on the estimated R-R interval, a minimum of two heartbeats is needed. Thus, in practice, the "0th" heartbeat must be acquired, and $S^*(1)$ is computed from the 0th and 1st heartbeats and $S^*(T)$ for $T \geq 2$ are computed sequentially as each heartbeat is added to the sample.

Comparing $S^*(T)$ to the critical values determines which hypothesis to accept. We define the errors α and β as follows:

$$\begin{aligned}\alpha &= \Pr\{\text{Rejecting } H0 \mid H0 \text{ is true}\} \\ &= \Pr\{S^*(T) > \log(B)\}, \\ \beta &= \Pr\{\text{Rejecting } H1 \mid H1 \text{ is true}\} \\ &= \Pr\{S^*(T) < \log(A)\}.\end{aligned} \tag{10}$$

For a test of simple hypothesis, it has been shown [46] that (11)

$$\begin{aligned}\log(A) &= \log\left[\frac{\beta}{1-\alpha}\right], \\ \log(B) &= \log\left[\frac{1-\beta}{\alpha}\right].\end{aligned} \tag{11}$$

To illustrate the application of the sequential procedure to the ECG signal, consider the example shown in Figure 8. Suppose the person presenting his/her credentials claims to be person i. Then the enrollment data for the ith subject gives the estimated mean Y_i under $H0$. If the true identity is j, where $j \neq i$, then one could use Y_j for the mean value under $H1$. We consider 5 cases in which $H0$ is false and the data come from five different individuals, labeled 1–5 in Figure 8(a). In all cases, the test statistic quickly exceeds the decision threshold $\log(B)$ for $\alpha = \beta = 0.01$. Comparing the behavior of the test statistics (Figure 8(a)) to the distance between the mean vector for the true identity and the mean vector for the declared identity (Figure 8(b)) reveals a direct correspondence. Note that these distances are computed from the training/enrollment data, while the test statistic depends on the enrolled means and the actual heartbeats observed in the test data. As one might expect, a large difference between the enrolled means for the true and declared identities corresponds to a large value of $S^*(T)$ and a rapid acceptance of $H1$. When the true mean is close to the mean of the declared identity, $S^*(T)$ increases more slowly.

This leads to the final step in the formulation of the sequential procedure, namely, the selection of i and j for constructing the test statistic. The choice of i is clear—it always corresponds to the declared identity of the individual presenting the credentials. To select j, we use the "closest imposter," that is, the enrolled individual with credentials closest to the declared individual. In other words, we select j such that as

$$\left\|Y_i - Y_j\right\| = \min\{k \ni k \neq i : \|Y_i - Y_k\|\}, \tag{12}$$

where $\|Y_i - Y_j\|$ is distance defined by

$$\left\|Y_i - Y_j\right\| = \left(Y_i - Y_j\right)^T \Sigma^{-1} \left(Y_i - Y_j\right), \tag{13}$$

and Σ is the pooled covariance matrix. When $H0$ is true, we use the nearest imposter to calculate the test statistic shown in Figure 8. The procedure determines that the $S^*(T)$ falls below the decision boundary, and $H0$ is accepted.

5. Results

We present performance results for two data sets. The first data set, consisting of 29 subjects, was acquired under a strict protocol documented previously [1–4]. The second data set merges recordings from two data acquisitions discussed by Israel et al. [6]. Both datasets are single channel collections. Together, these data sets suggest the performance that can be expected for a moderate size population. In practice, however, a range of issues require further investigation: the effects of varying mental and emotional states on the ECG signal, the sensor placement and efficient data acquisition, generalization to larger populations, and the long-term stability of the ECG credentials. These issues are explored in the next section.

5.1. First Data Set. The ECG data analyzed in the work of Israel et al. [4] and Irvine et al. [5] provides a target performance for the sequential procedure. For this experiment, the single channel ECG data were collected at the base of the neck at a sampling rate of 1000 Hz with an 11-bit dynamic range. The population consisted of 29 males and females between the ages of 18 and 48, with no known cardiac anomalies. During each session, the subject's ECG was recorded while performing seven 2-minute tasks. The tasks were designed to elicit varying stress levels and to understand stress/recovery cycles. The results shown here used data from the subject's low stress tasks. The next section presents results for one of the high-stress tasks.

Setting the decision threshold based on $\alpha = \beta = 0.01$, all 29 subjects were analyzed using the sequential procedure. When $H0$ is true, that is, the test data comes from the subject who is declared to be subject i, the results show that $H0$ is accepted in all cases (Figure 9). We stopped processing at 15 heartbeats. In all cases, the decision was reached within that time span, and usually much sooner.

Similarly, when $H1$ is true, the correct decision is generally reached in fewer than 15 heartbeats (Figure 10). In this set of results, the true identity for the test data is, in fact, the closest imposter. In only one case did the test procedure fail to reject an imposter within 15 heartbeats. In addition, we have computed the sequential tests when data for other subjects are used for the test set and the correct decision is always made in fewer heartbeats. Essentially, Figure 10 represents a worst case in which the subject trying to pose as someone else has a heartbeat that is fairly similar to the declared identity.

The sequential procedure performs well for the test data. An important practical issue is the number of heartbeats required to reach a decision. Figure 11 depicts the number of heartbeats required for a decision when $H0$ is true (Figure 11, left side) and when $H1$ is true (Figure 11, right side). In both cases, most of the individuals were identified using only 2 or 3 heartbeats. In cases where there is some ambiguity, however, additional heartbeats are needed to resolve the differences.

The number of heartbeats needed to reach a decision depends on the level of acceptable error. The results presented in Figures 9, 10 and 11 assume $\alpha = \beta = 0.01$.

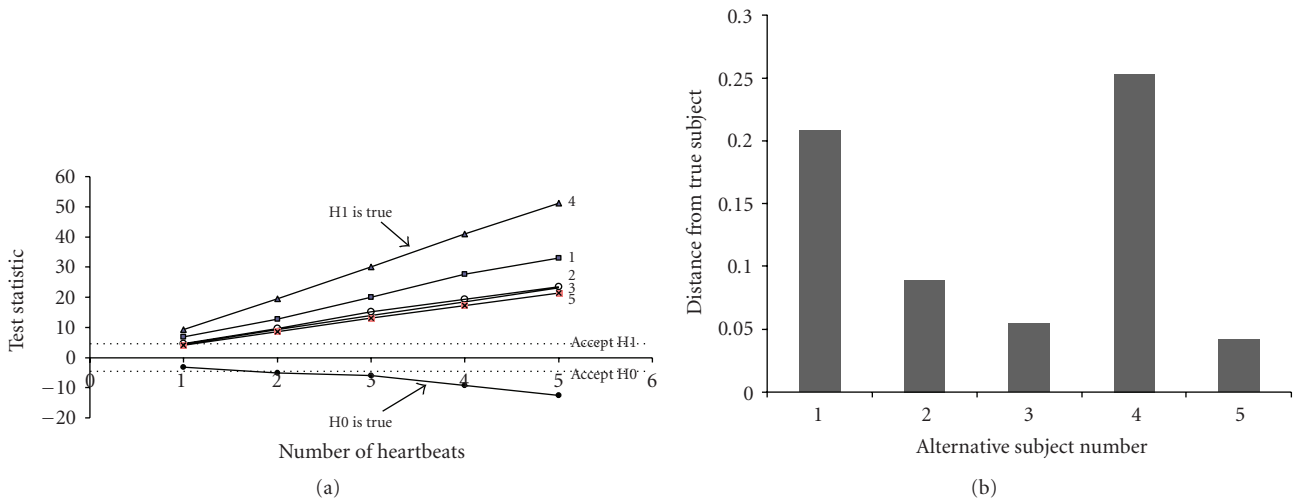


Figure 8: Example of a sequential procedure. (a) Sequential test statistic for a single declared identity when $H0$ is true and for five imposters. (b) The distance of the declared identity to the five imposters.

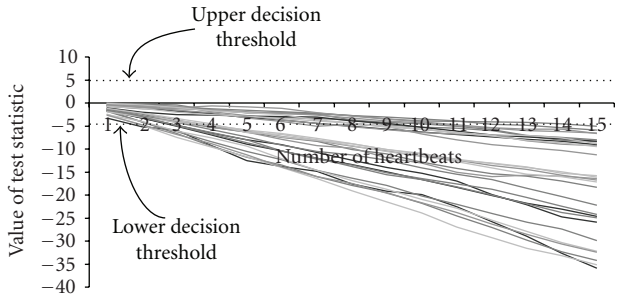


Figure 9: Sequential test statistics for all subjects when $H0$ is true. The test data are from the declared individual.

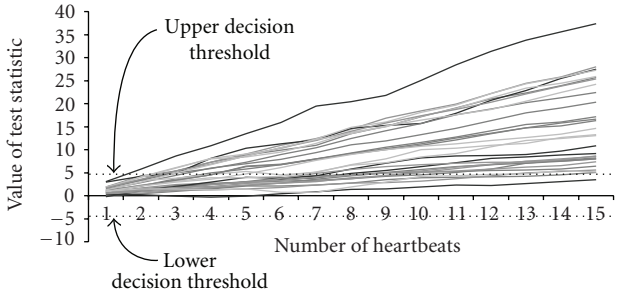


Figure 10: Sequential test statistics for all subjects when $H1$ is true. The test data are from the subject closest to the declared individual, that is, the nearest imposter.

An inverse relationship exists between acceptable error rate and required number of heartbeats. Smaller levels of acceptable error will drive the decision process to require more data. Table 1 summarizes the performance for $\alpha = \beta$ ranging from 0.1 to 0.0001. More stringent constraints on α and β, for example, $\alpha = \beta = 0.001$or $\alpha = \beta = 0.0001$, generally require more heartbeats. As the acceptable error reduces, a decision is not always realized within 15 heartbeats. For the case of $\alpha = \beta = 0.0001$, the procedure was run until a decision was reached for all subjects. When $H0$ is true, the maximum number of heartbeats needed was 33. When $H1$ was true, the maximum was 37 heartbeats. In all cases, the correct decision was reached.

5.2. Second Data Set. Two additional ECG data collection campaigns used a simplified protocol and a standard, FDA approved ECG device. The clinical instrument recorded the ECG data at 256 Hz and quantized it to 7 bits. These data were acquired from two studies: one which collected single channel data from 28 subjects with the sensor placement at the wrist and one which collected single lead data from 47 subjects using a wearable sensor. The result is an additional 75 subjects.

The analysis followed the same procedure as with the first data set. Application of the sequential procedure for all 75 subjects was performed under both $H0$ and $H1$. Table 2 summarizes the results for the two cases $\alpha = \beta = 0.05$ and $\alpha = \beta = 0.01$, where the procedure ran for a maximum of 24 heartbeats. The results show that in a few instances a decision is not reached within the 24 heartbeats. For $\alpha = \beta = 0.05$, when $H0$ is true the procedure fails to decide for 2 subjects and 2 additional subjects are classified incorrectly. When $H1$ is true, the procedure failed to decide for 1 subject and decided incorrectly for 1 subject.

A comparison of the results from the two data sets shows good consistency. A statistical comparison reveals no significant difference. Consider, for example, performance when $\alpha = \beta = 0.05$. Under $H0$, a statistical comparison of the correct acceptance rates yields a t-statistic of 1.39. The corresponding t-statistic under $H1$ is 0.58. In short, performance for the two experiments is statistically indistinguishable.

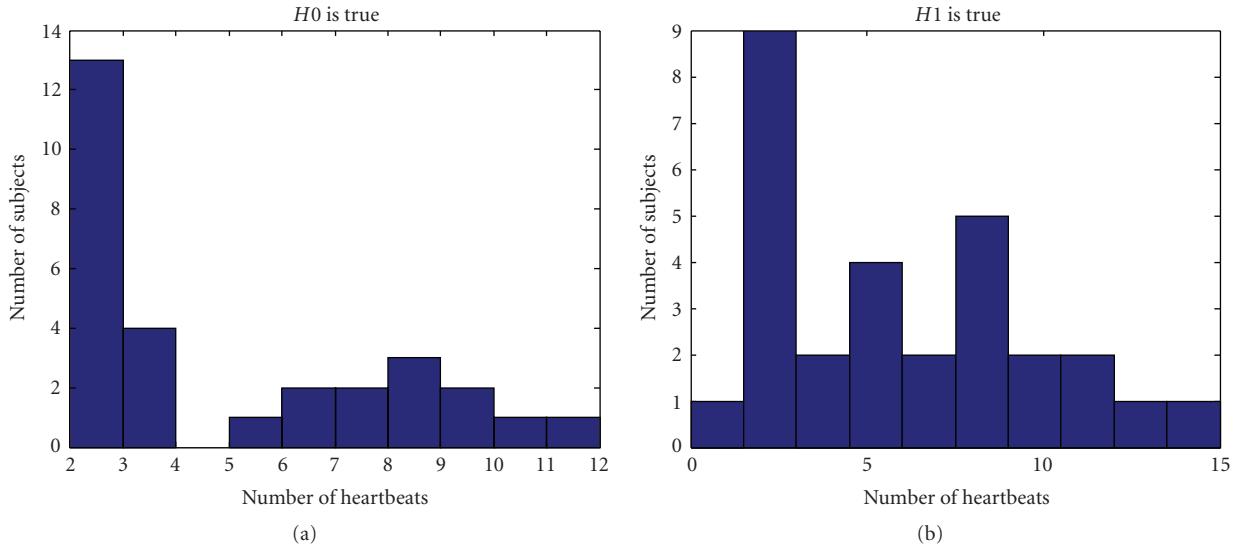


Figure 11: Histograms showing the number of heartbeats needed to reach a decision where the acceptable level of error is $\alpha = \beta = 0.01$.

Table 1: Summary statistics for the number of heartbeats needed to reach a decision for varying levels of the acceptable error.

	$H0$ is true					$H1$ is true			
Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent resulting in decision	Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent resulting in decision
0.1	3.38	2	8	100	0.1	3.655	2	11	100
0.05	4.24	2	9	100	0.05	4.621	2	14	100
0.01	6.07	2	15	100	0.01	6.500	2	15	96.6
0.005	6.68	3	14	96.6	0.005	7.000	2	14	93.1
0.001	7.28	3	13	86.2	0.001	7.792	3	15	82.8
0.0005	7.96	4	15	86.2	0.0005	8.174	3	15	79.3
0.0001	7.55	4	15	69.0	0.0001	7.647	4	14	58.6

6. Issues and Concerns

The results presented in the previous section, while promising, were obtained from modest data sets collected under controlled conditions. To be operationally viable, a system must address performance across a range of conditions. Key issues to consider are

(i) heartrate variability, including changes in mental and emotional states,

(ii) sensor placement and data collection,

(iii) scalability to larger populations,

(iv) long-term viability of the ECG credentials.

Heartrate Variability. Heartrate, of course, varies with a person's mental or emotional state. Excitement or arousal from any number of stimuli can elevate the heartrate. Under the experimental protocol employed to collect the first data set, subjects performed a series of tasks designed to elicit varying mental and emotional states [1–4]. The subjects exhibited changes in heartrate associated with these tasks. The fiducial features, however, show relatively small differences due to the variation in heartrate. To illustrate, consider Figure 12. For a single subject, Figure 12 presents 6 heartbeats from the baseline task in which the subject is

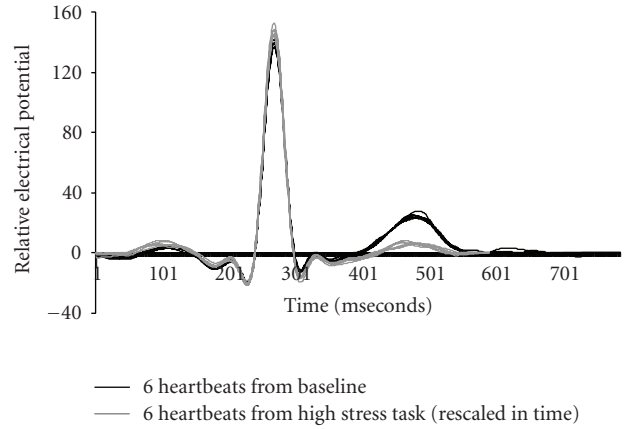


Figure 12: Aligned heartbeats from high stress and low stress tasks.

Table 2: Analysis of second data set.

(a) Heartbeats required to reach a decision

$H0$ is true					$H1$ is true				
Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent resulting in decision	Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent resulting in decision
0.05	3.04	2	22	97.3	0.05	3.10	2	22	98.7
0.01	4.93	2	24	92.0	0.01	4.99	2	24	92.0

(b) Correct decision rates

$H0$ is true		$H1$ is true	
Allowable error (α, β)	Percent resulting in correct decision	Allowable error (α, β)	Percent resulting in correct decision
0.05	94.7	0.05	97.3
0.01	89.3	0.01	90.7

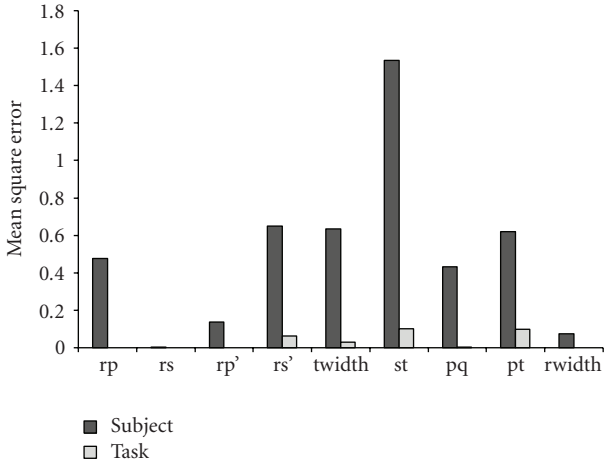


Figure 13: Comparison of variance attributable to subject and task.

seated at rest. In addition, 6 heartbeats from a high stress task (a virtual reality driving simulation) were temporally rescaled and overlaid on the same graph. For this particular subject, the mean R-R interval for the baseline task was 0.715 seconds and for the high stress task it was 0.580 seconds. However, by a linear rescaling, the high-stress heartbeats align well with the baseline heartbeats. A difference in the height of the T wave is evident but the fiducial features depend on the relative positions of the peaks, not the heights.

Delving deeper than the visual evidence for a single subject, we conducted a systematic analysis of the sources of variance in the fiducial features using a multivariate analysis of variance (MANOVA). The 29 subjects performed all seven tasks in the experimental protocol eliciting a range of stimulation. The MANOVA shows that there are small, but statistically significant, differences in the fiducials across the various tasks, indicating that there are subtle differences in the ECG signal that are more complex than a linear rescaling. This source of variance, however, is typically one or two orders of magnitude smaller than the variance across subjects. Figure 13 shows the relationships between the two mean square errors for each fiducial, and the variation across subjects is far more pronounced than the variation due to task. This relationship is why the fiducial-based features are likely to provide good information about a subject's identity across a range of conditions.

To verify this hypothesis, we explored the effect of varying the level of arousal of the subject. The protocol used for collecting Dataset 1 included a set of tasks designed to elicit varying levels of stimulation or arousal [1–4]. Using the baseline, low stress task for training, we processed data from one of the high-stress tasks for testing. Specifically, the subjects performed an arithmetic task designed to affect both stress and cognitive loads. The effectiveness of the task is evident in that the mean R-R interval decreased from a baseline of 0.83 to 0.76 for this task. Nevertheless, the sequential procedure yielded good performance on these data (Table 3).

If alternative attributes are evaluated in the trade space, such as wavelets [35] or Legendre coefficients [48], then their sensitivity must also be evaluated in the same manner as above. Likewise, incorporating other verification algorithms such as PCA [5, 49] or Gaussian modeling [50] will require substituting their characteristics into the sequential process. Regardless, the minimum number of heartbeats is appropriate for comparing systems.

Sensor Placement. Dataset 1 collected ECG traces from the base of the neck. Dataset 2 collected ECG traces on the forearms. Both collections used medical quality single use electrodes. However, any operational system must design a more robust collection method. This method must have reusable electrodes, a concept of employment for locating electrodes on normally exposed skin, and other human factors. These issues are outside the scope of this paper. However, the concept of employment does raise significant concerns about the noise floor for an operational system. As the noise floor increases the separability between the subject and the nearest imposter reduces.

Table 3: Effects of varying levels of stimulation.

	$H0$ is true					$H1$ is true			
Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent correct decision	Allowable error (α, β)	Mean no. of heartbeats	Minimum no. of heartbeats	Maximum no. of heartbeats	Percent correct decision
0.01	5.57	2	17	96.6	0.01	4.41	2	10	93.1

Scalability. Depending on the application, an ECG-based identity verification system may need to store credentials for hundreds or thousands of individuals. The recent experiments lack the sample sizes needed to determine large-scale performance, and the next step is to assess performance over much larger data sets. Because our approach compares the credentials for the declared subject to the nearest imposter, the separability among members of the training set is critical. By always choosing j to be the closest imposter, we guard against accepting a person's credentials too readily. Fortunately, determining the closest imposter is performed using training data, offline, which greatly improves the processing efficiency and system usability. It does, however, raise a concern about extending these methods to applications involving large enrolled populations. An alternative approach is to select j based on the features extracted from the first heartbeat. One could select j to be the member of the enrollment set closest to the first heartbeat from the test data, where $j \neq i$. In terms of the statistical formulation, $H1$ is no longer a simple hypothesis, since j is chosen to minimize a criterion over the full enrollment set. A simple experiment on a subset of the data revealed mean decision times of approximately 8 heartbeats for $\alpha = \beta = 0.01$, compared to 6.5 using the nearest imposter. Further investigation of this issue is still needed.

Long-Term Viability. Characteristics of an individual's ECG can change for a variety of medical reasons, including cardiovascular disease and changes in medication. Research has examined these issues from a clinical perspective, but further investigation is needed to understand how these factors affect ECG as a biometric for identification. For the data analyzed in this paper, the time difference between the training and test sets ranged from minutes to months, but no truly long-term differences have been studied. Such a study needs to be conducted, and existing clinical measurements are likely to be the most readily available source of data. Depending on the concept of employment, however, periodic re-enrollment may be one strategy for addressing long-term changes in an individual's ECG signal.

7. Discussion

This research builds on previous investigations into the viability of ECG as a biometric for human identification. We focus specifically on a procedure for exploiting the ECG signal for identity verification, with the optimization metric being the number of heartbeats needed for the system to make a decision. By using a method based on a sequential procedure for statistical hypothesis testing, data acquisition time is minimized. For the two data sets analyzed here, the approach generally yields the correct decision given enough heartbeats.

For modest levels of acceptable risk ($\alpha = \beta = 0.01$ or 0.05), the decision is often made after only 3 or 4 heartbeats and is almost always made within 15 heartbeats. In practice, this implies a data acquisition time of approximately 5 to 15 seconds. Lower risk tolerance (e.g., $\alpha = \beta = 0.0001$) could require 30 seconds or more to reach a decision for some individuals.

Whether the data acquisition time is acceptable in practice will depend, of course, on the specific application. One attractive approach is to use ECG in conjunction with other biometrics, such as fingerprint and hand geometry. This multimodality approach could support less demanding performance limits for the ECG (e.g., $\alpha = \beta = 0.01$), while providing a high level of security that will be very difficult to forge or circumvent.

Further investigations are still needed to refine and validate the methods presented here. Specific avenues for future research include the following.

(i) Assessment of performance over a much larger population of test subjects: larger data sets, including data collected at greater time intervals, are necessary to characterize the behavior of these methods.

(ii) Investigation of robustness to physical, mental, and emotional states and longer baselines between visits: heartrate will vary with a variety of stimuli. Irvine et al. [3] and Israel et al. [4] demonstrated that with proper normalization, the fiducial-based features are robust to mental and emotional states. Further validation that similar results hold for the sequential procedures would be useful.

(iii) Exploration of alternative feature extraction methods and verification algorithms: researchers have proposed a variety of alternative feature extraction methods, including variations on the fiducial features, principal component analysis, template matching, and frequency-domain approaches. These methods can be integrated into the sequential procedure framework, and a comparison of different approaches could prove enlightening.

Acknowledgments

This research was sponsored by the Defense Advanced Research Projects Agency (DARPA) under Contract no. DABT63-00-C-1039, with additional support from Charles Stark Draper Laboratory and SAIC. The data were collected

at the Virtual Reality Medical Center under the supervision of Drs. Mark and Brenda Wiederhold. Additional assistance was provided by Dr. Rodney Meyer, Dr. Lauren Gavshon, Ms. Shannon McGee, and Ms. Elizabeth Rosenfeld. The authors also wish to thank Dr. P. Jonathon Phillips, formerly of DARPA, for valuable comments concerning the development of this work. Finally, the authors wish to thank the anonymous reviewers for valuable insights and suggestions. The views expressed here are those of the authors and do not necessarily represent the positions of DARPA, SAIC, Draper Laboratory, or VRMC.

References

[1] J. M. Irvine, B. K. Wiederhold, L. W. Gavshon, et al., "Heart rate variability: a new biometric for human identification," in *Proceedings of the International Conference on Artificial Intelligence (IC-AI '01)*, pp. 1106–1111, Las Vegas, Nev, USA, June 2001.

[2] J. M. Irvine, S. A. Israel, A. Cheng, M. D. Wiederhold, B. K. Wiederhold, and S. McGehee, "Validation of new biometrics for human identification," in *Proceedings of the Joint Statistical Meetings (JSM '02)*, New York, NY, USA, August 2002.

[3] J. M. Irvine, S. A. Israel, M. D. Wiederhold, and B. K. Wiederhold, "A new biometric: human identification from circulatory function," in *Proceedings of the Joint Statistical Meetings of the American Statistical Association*, pp. 1–7, San Francisco, Calif, USA, August 2003.

[4] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition*, vol. 38, no. 1, pp. 133–142, 2005.

[5] J. M. Irvine, S. A. Israel, W. Todd Scruggs, and W. J. Worek, "eigenPulse: robust human identification from cardiovascular function," *Pattern Recognition*, vol. 41, no. 11, pp. 3427–3435, 2008.

[6] S. A. Israel, J. M. Irvine, B. K. Wiederhold, and M. D. Wiederhold, "The heartbeat: the living biometric," in *Biometrics: Theory, Methods, and Applications*, N. V. Boulgouris, E. Micheli-Tzanakou, and K. N. Plataniotis, Eds., Wiley-IEEE Press, New York, NY, USA, 2009.

[7] M. D. Wiederhold, S. A. Israel, R. P. Meyer, and J. M. Irvine, *Human Identification by Analysis of Physiometric Variation*, SAIC, San Diego, Calif, USA, 2006.

[8] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[9] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 148658, 11 pages, 2008.

[10] T. W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for identity verification," in *Proceedings of the 2nd Joint Conference of the IEEE Engineering in Medicine and Biology Society and the 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society (EMBS/BMES '02)*, vol. 1, pp. 62–63, Houston, Tex, USA, October 2002.

[11] F. R. Yu, H. Tang, V. C. M. Leung, J. Liu, and C.-H. Lung, "Biometric-based user authentication in mobile *ad hoc* networks," *Security and Communication Networks*, vol. 1, no. 1, pp. 5–16, 2008.

[12] R. Palaniappan and S. M. Krishnan, "Identifying individuals using ECG beats," in *Proceedings of the International Conference on Signal Processing and Communications (SPCOM '04)*, pp. 569–572, Bangalore, India, December 2004.

[13] M. Kyoso and A. Uchiyama, "Development of an ECG identification system," in *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '01)*, vol. 4, pp. 3721–3723, Istanbul, Turkey, October 2001.

[14] G. Wubbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, and C. Elster, "Verification of humans using the electrocardiogram," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1172–1175, 2007.

[15] G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Artech House, Norwood, Mass, USA, 2006.

[16] H. C. Bazett, "An analysis of time relations of electrocardiograms," *Heart*, vol. 7, pp. 353–370, 1920.

[17] D. P. Golden Jr., R. A. Wolthuis, and G. W. Hoffler, "A spectral analysis of the normal resting electrocardiogram," *IEEE Transactions on Biomedical Engineering*, vol. 20, no. 5, pp. 366–372, 1973.

[18] D. Hawkins, "Body of evidence," *U.S. News & World Report*, vol. 132, no. 5, pp. 60–62, 2002.

[19] M. P. S. Chawla, H. K. Verma, and V. Kumar, "A new statistical PCA-ICA algorithm for location of R-peaks in ECG," *International Journal of Cardiology*, vol. 129, no. 1, pp. 146–148, 2008.

[20] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 686–696, 2003.

[21] M. Kotas, "Robust projective filtering of time-warped ECG beats," *Computer Methods and Programs in Biomedicine*, vol. 92, no. 2, pp. 161–172, 2008.

[22] K. Noponen, J. Kortelainen, and T. Seppänen, "Invariant trajectory classification of dynamical systems with a case study on ECG," *Pattern Recognition*, vol. 42, no. 9, pp. 1832–1844, 2009.

[23] R. Nygaard, G. Melnikov, and A. K. Katsaggelos, "A rate distortion optimal ECG coding algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 28–40, 2001.

[24] N. C. Oza and K. Tumer, "Classifier ensembles: select real-world applications," *Information Fusion*, vol. 9, no. 1, pp. 4–20, 2008.

[25] R. Palaniappan and K. V. R. Ravi, "Improving visual evoked potential feature classification for person recognition using PCA and normalization," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 726–733, 2006.

[26] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee, "ECG biometric recognition without fiducial detection," in *Proceedings of the Biometric Consortium Conference (BCC '06)*, pp. 1–6, Baltimore, Md, USA, September 2006.

[27] A. van Oosterom, R. Hoekema, and G. J. H. Uijen, "Geometrical factors affecting the interindividual variability of the ECG and the VCG," *Journal of Electrocardiology*, vol. 33, no. 3, pp. 219–227, 2000.

[28] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[29] F. M. Bui, F. Agrafioti, and D. Hatzinakos, "Electrocardiogram biometric," in *Biometrics: Theory, Methods, and Applications*, N. V. Boulgouris, E. Micheli-Tzanakou, and K. N. Plataniotis, Eds., Wiley-IEEE Press, New York, NY, USA, 2009.

[30] F. Agrafioti and D. Hatzinakos, "ECG biometric analysis in cardiac irregularity conditions," *Signal, Image and Video Processing*, pp. 1683–1703, 2008.

[31] D. Dubin, *Rapid Interpretation of ECGs*, Cover, Tampa, Fla, USA, 6th edition, 2000.

[32] E. N. Marieb, *Essential of Human Anatomy and Physiology*, Benjamin Cummings, San Francisco, Calif, USA, 7th edition, 2003.

[33] T. Pawar, N. S. Anantakrishnan, S. Chaudhuri, and S. P. Duttagupta, "Transition detection in body movement activities for wearable ECG," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 6, pp. 1149–1152, 2007.

[34] R. Jane, H. Rix, P. Caminal, and P. Laguna, "Alignment methods for averaging of high-resolution cardiac signals: a comparative study of performance," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 6, pp. 571–579, 1991.

[35] P. Laguna, R. Jane, O. Meste, et al., "Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: comparison with signal averaging techniques," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 10, pp. 1032–1044, 1992.

[36] M. Alfaouri and K. Daqrouq, "ECG signal denoising by wavelet transform thresholding," *American Journal of Applied Sciences*, vol. 5, no. 3, pp. 276–281, 2008.

[37] W. Zhang, X. Wang, L. Ge, and Z. Zhang, "Noise reduction in ECG signal based on adaptive wavelet transform," in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '05)*, pp. 2699–2702, Shanghai, China, January 2005.

[38] V. Almenar and A. Albiol, "A new adaptive scheme for ECG enhancement," *Signal Processing*, vol. 75, no. 3, pp. 253–263, 1999.

[39] M. Elena, J. M. Quero, and I. Borrego, "An optimal technique for ECG noise reduction in real time applications," in *Proceedings of the Annual International Conference on Computers in Cardiology (CIC '06)*, vol. 33, pp. 225–228, Valencia, Spain, September 2006.

[40] A. K. Barros, M. Yoshizawa, and Y. Yasuda, "Filtering noncorrelated noise in impedance cardiography," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 3, pp. 324–327, 1995.

[41] A. K. Barros and N. Ohnishi, "Heart instantaneous frequency (HIF): an alternative approach to extract heart rate variability," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 8, pp. 850–855, 2001.

[42] P. S. Hamilton, "A comparison of adaptive and nonadaptive filters for reduction of power line interference in the ECG," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 1, pp. 105–109, 1996.

[43] A. H. Al-Khalidi, M. E. Lewis, J. N. Townened, R. S. Bonser, and J. H. Coote, "A novel and simple technique to allow detection of the position of the R-waves from intraventricular pressure waveforms: application to the conductance catheter method," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 606–610, 2001.

[44] B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42–57, 2002.

[45] M. Ghosh, *Handbook of Sequential Analysis*, CRC Press, Boca Raton, Fla, USA, 1st edition, 1991.

[46] A. Wald, *Sequential Analysis*, Dover, New York, NY, USA, 1994.

[47] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, NY, USA, 3rd edition, 2008.

[48] J. Faganeli and F. Jager, "Automatic distinguishing between ischemic and heart-rate related transient ST segment episodes in ambulatory ECG records," in *Proceedings of the 35th Annual International Conference on Computers in Cardiology (CIC '08)*, vol. 35, pp. 381–384, Bologna, Italy, September 2008.

[49] G. B. Moody and R. G. Mark, "QRS morphology representation and noise estimation using the Karhunen-Loeve transform," in *Proceedings of the Annual International Conference on Computers in Cardiology (CIC '89)*, vol. 16, pp. 269–272, Jerusalem, Israel, September 1989.

[50] D. Clifford, A. Shoeb, P. E. McSharry, and B. A. Janz, "Model-based filtering, compression and classification of the ECG," *International Journal of Bioelectromagnetism*, vol. 7, no. 1, pp. 158–161, 2005.