Multimodal Biometric System Based on Face and Hand Images Taken by a Cell Phone

Joanna Rokita

A Thesis
in
The Department
of
Computer Science and Software Engineering

Presented in Partial Fullfilment of the Requirements
for the Degree of Master of Computer Science at
Concordia University
Montreal, Quebec, Canada

March 2008

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-40952-7
Our file Notre référence
ISBN: 978-0-494-40952-7

ABSTRACT

Multimodal Biometric System Based on Face and Hand Images Taken by a Cell Phone

Joanna Rokita

One of the methods to improve the recognition rate of humans is multimodal biometrics, which is based on more than one physiological or behavioral characteristics to identify an individual. Multimodal biometrics improves not only the performance, but also nonuniversality and spoofing that are commonly encountered in unibiometric systems. In this thesis, we built a multibiometric system that works on face and hand images taken by a camera built into a cell phone. The multimodal fusion is done at the feature extraction level. The nine facial models are built according to the number of features / points extracted from the face. Active shape models method is applied in order to find the concatenated string of facial points in the eyes, nose, and mouth areas. The face feature vector is constructed by applying Gabor filter to the image and extracting the key points found by an active shape model. The hand feature vector contains nine geometric measurements, including heights and widths of four fingers, and the width of the palm. Support vector machine is used as a classifier for a multimodal approach. One SVM machine is built for each person in the database to distinguish that person from the others. The database contains 113 individuals. As the experiments show, the best accuracy of up to 99.82% has been achieved for the multibiometric model combining 8 eye points, 4 nose points, and 9 hand features.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisors Dr. Adam Krzyżak and Dr. Ching Y. Suen for providing me with the opportunity to work in the exciting and challenging areas of pattern recognition and biometrics. Without their broad knowledge in the field, I would not be able achieve my goals.

I would like to thank also all members of Centre for Pattern Recognition and Machine Intelligence (CENPARMI). Special thanks to Marleah Blom for professional editorial help, and Nicola Nobile for technical and hardware support.

I would like to thank my parents who have been the pillars of strength in all my endeavors. I am always deeply indebted to them for all that they have given me.

Finally, I thank my loving husband Artur for being with me all the way.

To My Parents

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

New technologies are designed to make our life easier and safer; however, they are often not secure enough. One of the most common identification methods is a Personal Identification Number (PIN), which can be easily eavesdropped by other people or forgotten by the user. That is why researchers started to work on biometric methods in order to identify people. Biometrics offers a natural and reliable solution to certain aspects of identity management by utilizing fully automated or semi-automated schemes to recognize individuals based on their inherent physical and/or behavioral characteristics [34]. A biometric system is a pattern recognition system that requires biometric data from an individual, such as fingerprint, iris, face, hand, voice, etc. Identification systems which are based on human characteristics such as fingerprint, hand geometry, face, iris, palm print, etc., have many advantages over the traditional authentication techniques based on what you know or what you possess [34, 6].

A wide variety of systems require person recognition to confirm or determine the identity of an individual requesting their services. The purpose of such systems is to ensure that only legitimate users can access the system. One such system that needs security access is a cellular phone. Nowadays, the new technology of cell phones allows us to connect to the internet, do banking transactions, maintain address books, send personal information, etc. Our research is focused on the security of login to the cell phones based on face and hand images. Since most cell phones are equipped with a camera, we built a user verification system based on face, hand, and multimodal characteristics.

1.2 Our Approach

The first step of our research was to build the database of face and hand images taken by the cell phone camera. From 113 individuals, we collected 20 pictures per person, 16 of them are face images and the remaining 4 are hand images. We used a Sony-Ericsson K750i cell phone with 2.0 mega pixel camera.

The system described in the thesis consists of three modules:

1. Face authentication system

2. Hand authentication system

3. Multimodal biometrics

Face authentication is based on facial points extracted from the face, such as points which describe eyes, nose and mouth. We find these points semiautomatically using active shape models. To be able to construct the feature vector, we convolve these face models with Gabor filters. We analyze what combinations of facial models yield the best accuracy against how many impostors are involved in the testing phase. To be able to compare our method, we run the same experiments using a well-known method called eigenfaces, which is based on principal component analysis.

Hand authentication utilizes hand geometry features, like width and height of fingers and width of the palm. These points are found by active shape models.

At the last level of experiments, we combined the face and hand authentication methods together. Hand features are concatenated with the face Gabor feature vectors. The experiments were run the same way as for the face authentication system, meaning according to the number of impostors involved in the tests and combinations of facial points extracted.

In this research we use a support vector machine (SVM) to verify the identity of a cell phone user. We use radial basis function (RBF) kernel for face and hand authentication systems as well as for a combination of the two. During training we use labeled facial images, hand geometry or the combination of both, to design a classifier which is capable of separating a genuine user from an intruder.

1.3 Our Goal

Our goal is to build a biometric system that works on face and hand images taken by a camera already built into a cell phone. Images taken by a cell phone camera are usually of poor quality. They contain the noise resulting from the movement of the camera or a person because the cell phone camera does not have an image stabilizer.

Also, the images are taken in different environments, outdoors, in the office with the daylight, fluorescent or incandescent light, in the dark environment using the flash light. We show that multimodal biometric system obtains the highest accuracy compared to separate face and hand authentication methods. As the experiments show, a high accuracy of up to 99.82% has been achieved for multibiometrics on our database. The system is simple and fast and can be incorporated into cell phones in future. We also show that when we apply only face authentication, the results are reasonable. The approach that is presented in this thesis is an improvement of the preliminary authentication systems introduced in [59, 58].

1.4 Organization of the Thesis

Chapter 2 presents an overview of the field of biometrics and multibiometrics. It describes the architecture, as well as biometric applications. Chapter 3 then discusses the face and hand authentication methods developed in this thesis. Also, the multimodal biometrics system is presented. Experimental results and analysis are then presented in Chapter 4. A cell phone face and hand database is illustrated in Chapter 4. Conclusions and directions of possible future work are discussed in Chapter 5.

Chapter 2

Biometrics

2.1 Introduction

The term biometric comes from the Greek words bios (life) and metrikos (measure). It is well-known that humans use inputs such as face, voice or gait to recognize each other. Recognition of people based on their characteristics is important in many emerging technologies. These days, biometrics is used in a wide variety of applications that require the verification schemes to confirm the identity of an individual.

In this chapter, we present an overview of biometric methods.

2.2 Biometrics: Overview

Biometrics is a constantly evolving field that is becoming more widespread in the industry. The term biometrics is described as an automatic personal recognition system based on physiological or behavioral characteristics [32, 54, 6]. Biometrics use biological properties of a human, such as fingerprints, iris, voice recognition, face recognition, and hand geometry to identify individuals. Figure 2.1 shows examples

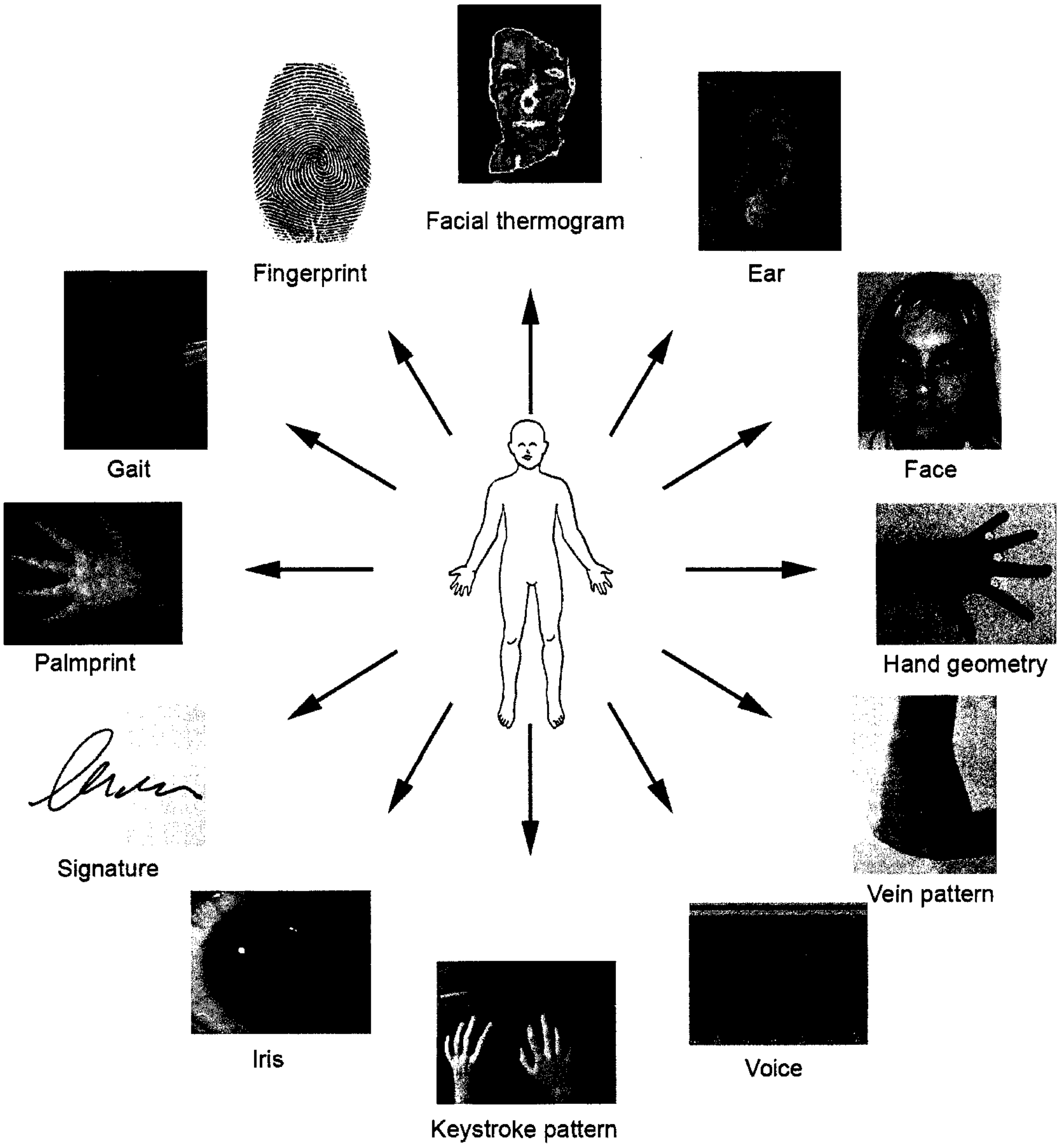


Figure 2.1: Biological properties used for biometrics.

of the biological properties used for biometrics.

Biometrics is no longer confined to criminal law enforcement. In addition, more businesses use biometrics to regulate access to buildings and information. Governments are contemplating the inclusion of biometric identifiers in passports, driver's licenses, and possibly a future national ID card. Also, digital video surveillance is spreading in private and public places.

A biometric system is basically a pattern recognition system that recognizes a person based on specific physiological or behavioral features. A biometric system usually runs in two modes:

1. Verification mode, where the system validates a person's identity by comparing the captured biometric characteristic with the individual's biometric template, which is stored in the system database. Identity verification is typically used for positive recognition, where the aim is to prevent different people from using the same identity [54, 68].

2. Identification mode, the system recognizes an individual by searching the entire template database for a match. The system conducts a one-to-many comparison to establish an individual's identity. Identification is a critical component of negative recognition applications, in which the system establishes whether the person is who he/she claims to be [54, 68]. The purpose of negative recognition is to prevent a single person from using multiple identities.

However, each biometric system, regardless if it operates on verification or identification mode, contains the following parts:

- Biometric readers or sensors
- Feature extractors
- Feature matchers for comparing two sets of biometric features

Physiological biometrics are based on measurements and data derived from direct measurement of a part of the human body. Here we list some examples of those biometrics:

- **DNA** - deoxyribonucleic acid: It is the one-dimensional ultimate unique code for one individual, except for the identical twins. DNA is currently used mostly in the forensic applications for person recognition.
- **Ear recognition**: It is based on matching the distance of salient points on the pinna from a landmark location on the ear. The evidence from two studies [31, 32] supports the hypothesis that the ear contains unique physiological features, since in both studies all examined ears were found to be unique though identical twins were found to have similar, but not identical, ear structures especially in the Concha and lobe areas. Having shown uniqueness, it remains to ascertain if the ear provides biometrics which are comparable over time.
- **Facial, hand, and hand vein infrared thermogram**: The pattern of heat radiated by the human body is a characteristic of an individual and can be captured by an infrared camera in an unobtrusive way much like a regular (visible spectrum) photograph [56].

- **Fingerprint**: A fingerprint is the pattern of ridges and valleys on the surface of a fingertip used for recognition. Fingerprints are unique for each person. In addition, fingerprints of identical twins are different and so are the prints of each finger of the same person [48].

- **Palmprint**: The palms of human hands contain patterns of ridges and valleys much like the fingerprints. The area of the palm is much larger than the area of a finger and as a result, palmprints are expected to be even more distinctive than the fingerprints. Some of palmprint techniques distinguish between identical twins [40].

- **Retina recognition**: The retinal vasculature is rich in structure and is supposed to be a characteristic of each individual and each eye. It is claimed to be the most secure biometric since it is not easy to change or replicate the retinal vasculature. Retina is unique for each individual, even for identical twins [32].

- **Hand and finger geometry**: Hand geometry recognition systems are based on a number of measurements taken from the human hand, including its shape, size of palm, and lengths and widths of the fingers. The geometry of a hand and fingers it is not very distinctive, and cannot be used for systems requiring identification of an individual from a large population [32].

- **Iris recognition**: Iris recognition systems scan the surface of an iris in order to compare patterns. Each iris is distinctive and, like fingerprints and retinas, even the irises of identical twins are different [21].

- **Face recognition**: Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [37, 65, 4, 53, 69, 63].

Since face recognition is a non-intrusive method, and facial images are probably the most common biometric characteristic used by humans to make personal recognition, we decided to explore this biometric method in our research. Face recognition is presented in detail in chapter 3.

Behavioral characteristics are based on an action taken by a person. On the other hand, behavioral biometrics are based on measurements and data derived from an action and indirectly measure characteristics of the human body. The following are examples of biometric techniques based on behavioral characteristics:

- **Gait**: Gait is the way one walks and is a complex spatio-temporal biometric. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications [32].

- **Signature recognition**: The way a person signs his/her name is known to be a characteristic of that individual. Signatures change over a period of time and are influenced by physical and emotional conditions of the signatories [50].

- **Voice recognition**: Voice recognition systems use the characteristics of the voice in order to recognize a person. The behavioral part of the speech of a person changes over time due to age, medical conditions (such as common cold), emotional state, etc; therefore, voice is not very distinctive and may not be appropriate for large-scale identification [10].

- **Keystroke**: It is hypothesized that each person types on a keyboard in a characteristic way. It is not unique to each individual but it offers sufficient discriminatory information to permit identity verification [49].

There are seven factors defined by Jain, Bolle, and Pankanti [32] that determine the suitability of a physical or a behavioral trait to be used in a biometric application.

1. **Universality**: each person accessing the application should posses the trait.

2. **Uniqueness**: the given trait should be sufficiently different across individuals comprising the population.

3. **Permanence**: the characteristic should be sufficiently invariant with respect to the matching criterion over a period of time.

4. **Collectability**: the characteristic should be measured quantitatively.

5. **Performance**: the recognition accuracy and the resources required to achieve that accuracy should meet the constraints imposed by the application.

6. **Acceptability**: individuals in the target population that will utilize the application should be willing to present their biometric trait to the system.

7. **Circumvention**: this reflects how easily the system can be fooled using fraudulent methods.

Table 2.1 presents a brief comparison of the physiological and behavioral biometric techniques based on these seven factors described above.

Biometrics	**Universality**	**Uniqueness**	**Permanence**	**Collectability**	**Performance**	**Acceptability**	**Circumvention**
DNA	High	High	High	Low	High	Low	Low
Ear	Medium	Medium	High	Medium	Medium	High	Medium
Thermogram	High	High	Low	High	Medium	High	Low
Fingerprint	Medium	High	High	Medium	High	Medium	Medium
Palmprint	Medium	High	High	Medium	High	Medium	Medium
Retina	High	High	Medium	Low	High	Low	Low
Hand geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	High	Medium	High	Low	Low
Face	High	Low	Medium	High	Low	High	High
Gait	Medium	Low	Low	High	Low	High	Medium
Signature	Low	Low	Low	High	Low	High	High
Voice	Medium	Low	Low	Medium	Low	High	High
Keystroke	Low	Low	Low	Medium	Low	Medium	Medium

Table 2.1: Comparison of biometric technologies [34].

2.2.1 Biometric applications

Several biometric characteristics are in use in various applications. These applications of biometrics can be divided into three main groups [54, 34]:

1. Commercial applications, such as computer network login, electronic data security, ecommerce, Internet access, ATM, credit card, physical access control, cellular phone.

2. Government applications, such as national ID card, correctional facility, drivers license, social security, welfare-disbursement, border control, passport control, etc.

3. Forensic applications, such as corpse identification, criminal investigation, missing children, etc.

Traditionally, commercial applications have used knowledge-based systems, such as PIN and passwords; however, these methods are not secure enough because passwords or PINs are easy to crack and easy to forget. In addition, the password can be shared by the user with his/her colleagues and then there is no way for the system to know who the actual user is. Government applications have used token-based systems, such as ID cards and badges. These systems have their problems as well. Firstly, the token or ID card can be stolen, shared, duplicated or lost; whereas, biometrics cannot be stolen, lost or forgotten. Only forensic applications have relied on human experts to match biometric features. These days, biometric methods are also used more often for security purposes. The first airport that applied biometrics for passengers verification was the Schipol airport in Amsterdam (Netherlands). This airport is equipped with iris scans that validates passengers (Privium) [55]. The border passage identifies a passenger using iris recognition. It is safe and considerably faster compared with manual passport control. Such a system also exists in Canada (CANPASS) [11] at the airports in the following cities: Edmonton, Winnipeg, Calgary, Halifax, Ottawa, Montreal, Toronto and Vancouver.

2.2.2 Limitation of unimodal biometric systems

In real world applications there are several problems with unimodal biometric systems which operate on a single biometric modality. The limitations of unimodal biometric systems are as follows [34]:

- Noise in sensed data: Noise can be present in the acquired biometric data mainly due to defective or improperly maintained sensors. For example, accumulation of dirt or the residual remains on a fingerprint sensor can result in a noisy fingerprint image. Failure to focus the camera appropriately can lead to blurring in face and iris images.

- Intra-class variations: Biometric data acquired from an individual during an authentication session may be different from the data that was used to generate the template during enrollment. The variations may be due to improper interaction of the user with the sensor (e.g., changes due to rotation, translation and applied pressure when the user places his finger on a fingerprint sensor, changes in pose and expression when the user stands in front of a camera, etc.), use of different sensors during enrollment and verication, changes in the ambient environmental conditions (e.g., illumination changes in a face recognition system) and inherent changes in the biometric trait (e.g., appearance of wrinkles due to aging or presence of facial hair in face images, presence of scars in a fingerprint, etc.).

- Distinctiveness: While a biometric trait is expected to vary significantly across individuals, there may be large inter-class similarities in the feature sets used to represent these traits. This limitation restricts the discrimination power provided by the biometric trait. Inter-user similarity refers to the overlap of the biometric samples from two different individuals in the feature space.

- Non-universality: While every user is expected to possess the biometric trait being acquired, in reality it is possible that some users do not possess that particular biometric characteristic. The National Institute of Standards and Technology (NIST) has reported that it is not possible to obtain a good quality fingerprint from approximately two percent of the population (people with hand-related disabilities, manual workers with many cuts and bruises on their fingertips, and people with very oily or dry fingers) [51]. Hence, such people cannot be enrolled in a fingerprint verification system. Similarly, persons having long eye-lashes and those suffering from eye abnormalities or diseases like glaucoma, cataract, aniridia, and nystagmus cannot provide good quality iris images for automatic recognition [3].

- Spoof attacks: An individual may attempt to fake the biometric trait. It is easy for behavioral characteristics, such as when signature and voice are used as an identifier.

Some of the limitations imposed by unimodal biometric systems can be overcome by using multiple biometric modalities [5, 39, 7]. Multibiometric systems are described in the next section.

2.3 Multibiometrics

Biometric systems can also be designed to recognize a person based on information received from multiple biometric sources. Such systems, known as multibiometric systems, can be expected to be more accurate due to the presence of multiple pieces of evidence [29, 43].

Multimodal biometric systems address the problem of non-universality, since multiple traits ensure sufficient population coverage. In addition, multimodal biometric systems provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a legitimate user.

The multibiometric systems are divided into five groups. The grouping depends on the various sources of information. Figure 2.2 shows multibiometric sources, which are divided into following groups:

1. **Multi-sensor**, in this system a single biometric trait is imaged using multiple sensors in order to extract information from registered images. For instance, the face images of an individual obtained using a thermal infrared camera and a visible light camera [14].

2. **Multi-modal**, these systems combine the evidence presented by different body traits for establishing identity. The cost of these systems is high since new sensors must be added [52, 13, 28, 58].

3. **Multi-instance**, these systems use multiple instances of the same body trait. For example, the left and right index fingers. These systems are cost-effective, because they require neither new sensors nor new algorithms for feature extraction [34].

4. **Multi-algorithm**, in this system the same biometric data is processed using multiple algorithms. This system does not require the use of new sensors and therefore is cost-effective. For example, a texture-based algorithm and minutiae-based algorithm can operate on the same fingerprint image [34].

5. **Multi-sample**, in these systems a single sensor is used to obtain multiple samples of the same biometric trait. For example, face pictures, frontal profile, left and right profiles.

Multibiometric system architectures are categorized into three system architectures according to the strategies used for information fusion [29, 60]:

- Fusion at the feature extraction level
- Fusion at the matching score level
- Fusion at the decision level

It means that the system is classified depending on how early in the authentication process the information from the different sensors is combined. Fusion at the feature extraction level stands for immediate data integration at the beginning of the processing chain. The information extracted from the different sensors is encoded into a joint feature vector, which is then compared to an enrollment template (which itself is a joint feature vector stored in a database) and is assigned a matching score as in a single biometric system (see Figure 2.3).

In the fusion at the matching scores, feature vectors are created independently for each sensor and then are compared to the enrollment templates, which are stored separately for each biometric trait. Based on the proximity of feature vector and template, each subsystem now computes its own matching score. These individual scores are finally combined into a total score, which is handed over to the decision module. The whole process is shown in Figure 2.4. Examples of feature level fusion

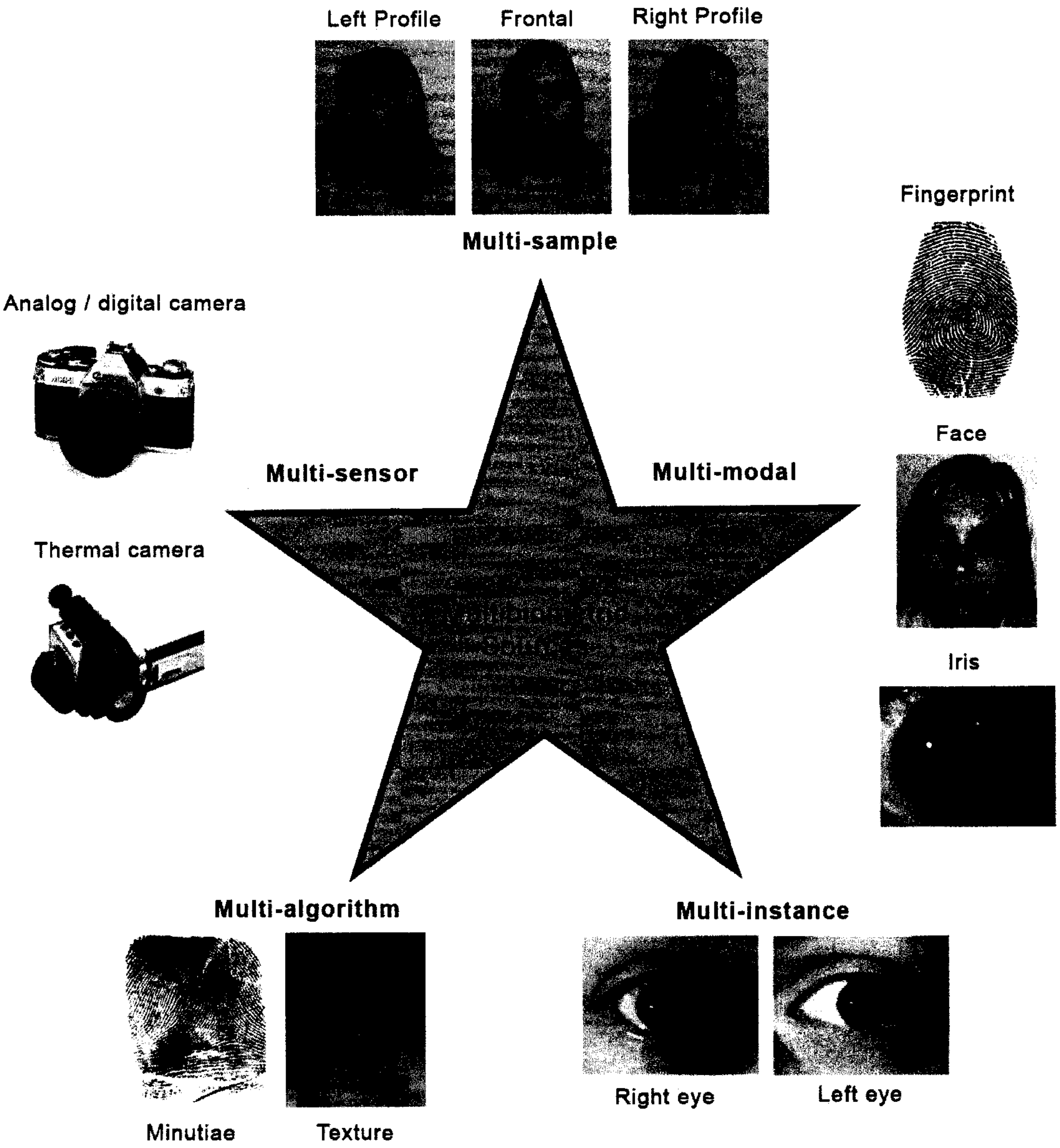


Figure 2.2: Multibiometric sources.

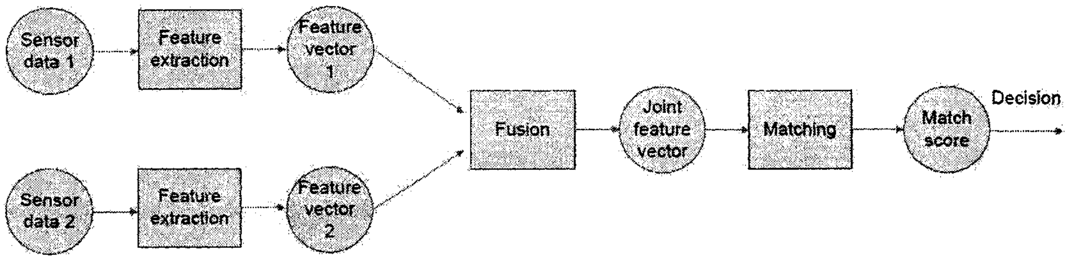


Figure 2.3: Block diagram of the fusion at the feature extraction level.

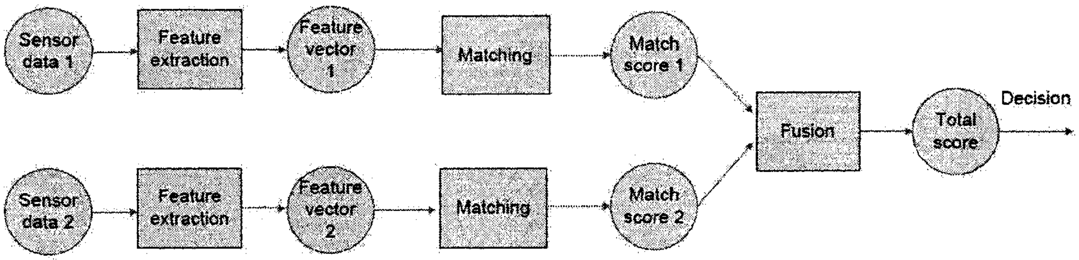


Figure 2.4: Block diagram of the fusion at the matching score level.

schemes proposed in the literature can be found in [16] (voice and lip shape), [64] (face and iris), [41] (hand geometry and palmprint), [59, 58] (face and hand geometry).

In the decision level fusion architecture, a separate authentication decision is made for each biometric trait. These decisions are then combined into a final vote, as shown in Figure 2.5. Methods proposed in the literature for decision level fusion include AND

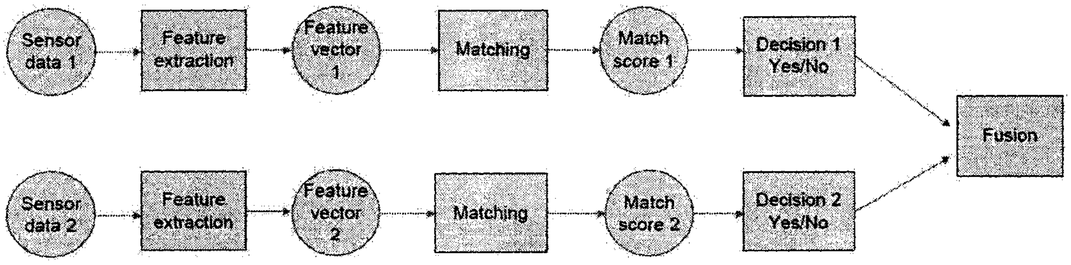


Figure 2.5: Block diagram of the fusion at the decision level.

and OR rules [17], majority voting [45], weighted majority voting [42], Bayesian decision fusion [70], the Dempster-Shafer theory of evidence [70] and behavior knowledge space [30].

It is generally believed that a combination scheme applied as early as possible in the recognition system is more effective. For example, an integration at the feature level typically results in a better improvement than at the matching score level. This is because the feature representation conveys the richest information compared to the matching score of a matcher, while the abstract labels contain the least amount of information about the decision being made. However, it is more difficult to perform a combination at the feature level because the relationship between the feature spaces of different biometric systems may not be known and the feature representations may not be compatible.

Multibiometric systems also have a few disadvantages when compared to unibiometric systems. They are more expensive and require more resources for computation and storage than unibiometric systems. Multibiometric systems generally require additional time for user enrollment, causing some inconvenience to the user; however multibiometric systems achieves better accuracies.

Chapter 3

Face and Hand Authentication

3.1 Introduction

This chapter describes the approach that we implemented for this research, including face authentication, hand authentication, and multimodal biometrics. Section 3.2 deals with recent face recognition methods and briefly presents our approach for face recognition. The next section presents a well-known method for face recognition, principal component analysis (Eigenfaces). The detection of facial points is described in section 3.2.2. We concatenate facial picture with Gabor filters, as shown in section 3.2.3. A Support vector machine (SVM), explained in section 3.5, is applied to verify the identity of the user based on Gabor features for the face and geometry features for the hand. One SVM machine is built for each person in the database to distinguish that person from others.

3.2 Face Recognition

Face recognition involves computer recognition of personal identity based on geometrical or statistical features derived from face images [71, 8, 23]. Humans can easily detect and identify faces in a scene; however, to build an automated system that recognizes faces is still a challenging problem. The problem becomes even more difficult when one considers the large variations in lighting conditions, poses, facial expressions, aging, glasses or cosmetics [68].

A face recognition system contains four subsystems that carry out face detection, face alignment, feature extraction and feature matching. Figure 3.1 shows the overall face recognition system.

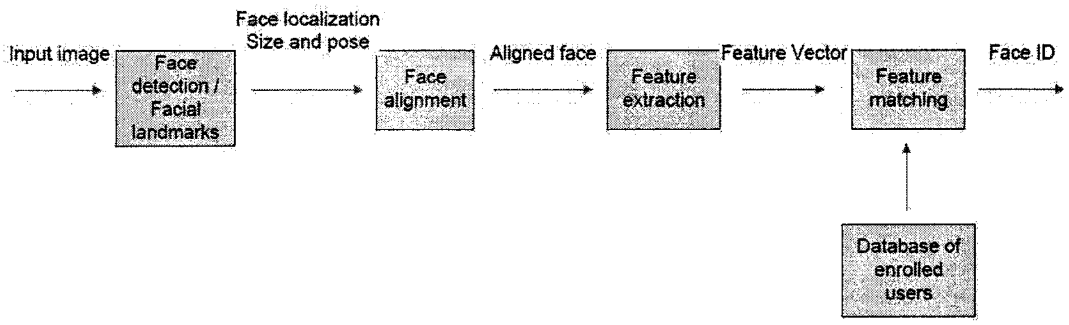


Figure 3.1: Block diagram of the overall face recognition system.

Face detection and coarse facial landmark detection are involved in the first step. Secondly, face alignment is achieved. Also at this stage the facial components such as eyes, nose, and mouth are located. Then, the system extracts features that are suitable for the face verification system. Finally, the system performs matching. It matches the feature vector of the input face against those face vectors enrolled in the database.

Face recognition is one of the primary biometric technologies. It has several advantages over the biometrics methods: it is natural, nonintrusive, and easy to use. Among the six biometric indicators considered in [27], facial features scored the highest compatibility, shown in Figure 3.2, in a Machine Readable Travel Documents (MRTD) system based on a number of evaluation factors [27], such as enrollment, renewal, machine requirements, and public perception.

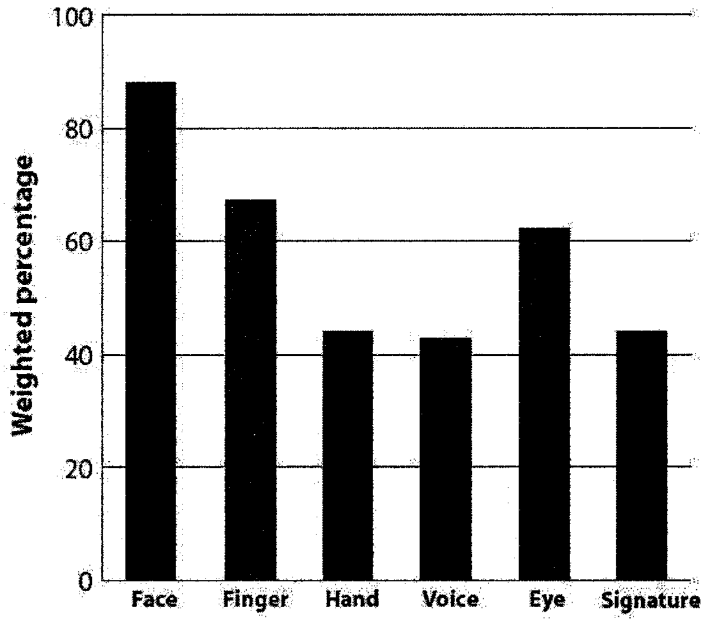


Figure 3.2: Comparison of various biometric features based on MRTD compatibility.

Face recognition scenarios can be classified into three types:

- face verification (or authentication),

- face identification (or recognition),
- watch list recognition

Face verification is a one-to-one match that compares a query face image against a template face image whose identity is being claimed; whereas, face recognition is a one-to-many matching process that compares a query face image against all the template images in a face database to determine the identity of the query face. The watch list method is where a query face image is matched to a list of suspects (one-to-few matches).

The first face recognition system was introduced by Kanade [37] in 1973. Since then, face recognition and authentication have become a very active area of research. Image based face recognition techniques can be divided into three groups: appearance-based (holistic), feature-based (analytic) and hybrid methods. The appearance-based methods use the whole face region as a raw input to a recognition system. Eigenface approach [65] and Fisherface approach [4, 46] are the best known appearance based methods. Feature based approach [37, 8, 61] used in the early days, is based on facial feature analysis such as eyes, nose and mouth. These features are first detected, then the properties and relations, such as areas, distances and angles between the features are fed into a structural classifier. Hybrid methods use both the local features and whole face region to recognize a face. Local Feature Analysis (LFA) [53], Elastic Bunch Graph Matching (EBGM) [69], and methods involving Gabor features [63, 57] are examples of holistic methods. The last method (Gabor features) has especially become very popular since it exhibits desirable characteristics of spatial locality and orientation selectivity [35]. Among classification techniques one of the most popular is Support Vector Machine (SVM) [20, 1]. SVM has been used for face recognition [25, 26] and face verification [36].

Our approach belongs to a class of holistic methods. Figure 3.3 represents the flowchart of our face identification system. It finds facial feature points that describe eyes, nose and mouth and concatenates those points with Gabor filters. In order to find those facial points, we use active shape models described in section 3.2.2. We use the support vector machine as a classifier, see section 3.5. We compare our system with the PCA approach, described in section 3.2.1.

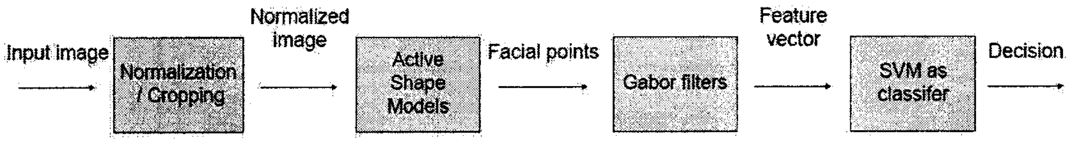


Figure 3.3: Flowchart of the face identification system, implemented in this thesis.

3.2.1 Principal Component Analysis

The PCA for face analysis and representation was first used in [38]. The first application of PCA to face recognition was described in [65]. Because the basis vectors constructed by PCA have the same dimension as the input face images, they were named "eigenfaces". Figure 3.4 shows the average eigenface for our database.

Principal component analysis (PCA) is a dimensionality reduction technique. The main idea behind the PCA is that each original image of the training set can be transformed into corresponding eigenfaces. Each eigenface represents only certain features

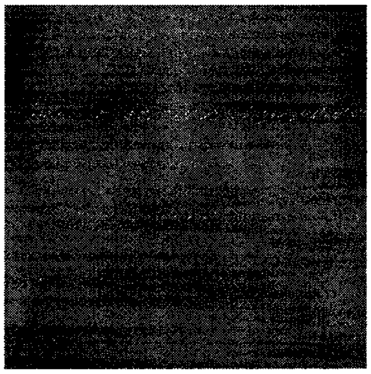


Figure 3.4: The average face from our database.

of the face, which may or may not be present in the original image. If the feature is present in the original image to a higher degree, the share of the corresponding eigenface in the sum of the eigenfaces should be greater. If, on the contrary, the particular feature is not (or almost not) present in the original image, then the corresponding eigenface should contribute a smaller (or not at all) part to the sum of eigenfaces.

Each digital image can be represented as an array of pixel values. In the m-by-n image size, the pixel array can be represented as a point in a mn-dimensional image space by simply writing its pixel values in a fixed position. For example, an 8x8 image may be unwrapped and treated as a vector of length 64. The image is said to live in m-dimensional space, where m is the number of pixels (and the length of the vector). This vector representation of the image is considered to be the original space of the image. A face image of size NxN pixels can be viewed as a vector of dimension N^2 as shown in Figure 3.5. However, such a high-dimensional representation is too detailed. Since facial features are similar in overall configuration across individuals, it is possible to describe faces quite compactly using PCA. The PCA approach to face recognition includes following steps [65]. First, the original images of the training

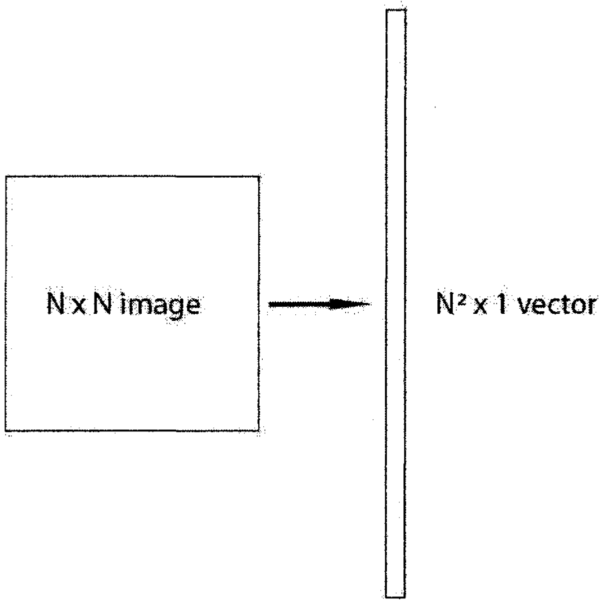


Figure 3.5: Mapping NxN image into a vector of dimension N^2.

set are transformed into eigenfaces. Then, the weights are calculated for each image of the training set. Upon observing an unknown image, the weights are calculated for this particular image. In the last step the difference between the weights in the training set and the weights for the unknown image are compared and the closest difference based on the threshold θ is considered as the recognized image.

Suppose the training set of face images $\Gamma_1, \Gamma_2, \Gamma_3, \ldots, \Gamma_M$ are an N^2x1 vectors. The average face of the set of M images is defined by:

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{3.1}$$

where M is the number of face images in the training set. Then each face Γ_i differs from the average face Ψ by Φ_i

$$\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Psi} \tag{3.2}$$

In the next step, the covariance matrix C is calculated according to:

$$C = \frac{1}{M}\sum_{n=1}^{M} \mathbf{\Phi}_n \mathbf{\Phi}_n^T = AA^T \tag{3.3}$$

where $A = [\mathbf{\Phi}_1 \mathbf{\Phi}_2 \ldots \mathbf{\Phi}_M]$ is MxN^2 matrix. When the covariance matrix is calculated, then the eigenvectors (eigenfaces) $\mathbf{u_i}$ and the corresponding eigenvalues λ_i of the covariance matrix have to be calculated. From M eigenvectors (eigenfaces) $\mathbf{u_i}$, only M' should be chosen, which have the highest eigenvalues.

The covariance matrix C is N^2xN^2, and determining the N^2 eigenvectors and eigenvalues is not very efficient; therefore, a simplified calculation has to be adopted. If the number of data points in the image space is less than the dimension of the space ($M < N^2$) there will be only M rather than N^2 eigenvectors. To find eigenfaces, first the eigenvectors of M by M matrix A^TA have to be calculated. Consider eigenvectors $\mathbf{v_i}$ of A^TA such that

$$A^TA\mathbf{v_i} = \mu_i \mathbf{v_i} \tag{3.4}$$

Multiplying both sides by A, we get:

$$AA^TA\mathbf{v_i} = \mu_i A\mathbf{v_i} \tag{3.5}$$

We observe that $A\mathbf{v_i}$ are eigenvectors of $C = AA^T$. We can construct the M by M matrix $L = A^TA$, and find M eigenvalues that correspond to the M largest eigenvalues of AA^T. Then we have to compute eigenvectors $\mathbf{u_i} = A\mathbf{v_i}$of AA^T.

According to this approach, the calculations are reduced from the order of the number of pixels in the image (N^2) to the order of the number of images in the training set (M). Usually, we will use only a subset of M eigenfaces, the K eigenvectors corresponding to the K largest eigenvalues.

In order to represent the face onto this basis, each face (minus the mean) Φ_i in the training set can be represented as a linear combination of the best K eigenvectors.

$$\Phi_i - mean = \sum_{j=1}^{K} \omega_j u_j, \omega_j = u_j^T \Phi_i \tag{3.6}$$

where u_j is an eigenface. Each normalized training face Φ_i is represented in this basis by a vector:

$$\Omega_i^T = [\omega_1^i \omega_2^i \ldots \omega_K^i], i = 1, 2, \ldots, M \tag{3.7}$$

The process of classification of a new (unknown) face Γ to one of the classes, known faces, proceeds in two steps. A new face Γ is transformed into its eigenface components, projected into "face space", by a simple operation:

$$\omega_k = u_k^T(\Gamma - \Psi), k = 1, 2, \ldots, K \tag{3.8}$$

This describes a set of point-by-point image multiplications and summations.

The weights for a vector $\Omega^T = [\omega_1 \omega_2 \ldots \omega_K]$ that describes the contribution of each eigenface in representing the input face image, treating the eigenfaces as a basis set of face images. The vector is used to find which of a number of pre-defined face classes, if any, best describes the face. The best method to check which face class provides the best description of a new face image is to find the face class k that minimizes the Euclidean distance

$$\epsilon_k = ||\Omega - \Omega_k|| \tag{3.9}$$

where Ω_k is a vector describing the kth face class. A face is classified to belong to class k when the minimum distance ϵ_k is below some chosen threshold θ; otherwise the face is unknown. If the Euclidean distance exceeds on average some threshold value θ, we can assume that a new image is not a face at all.

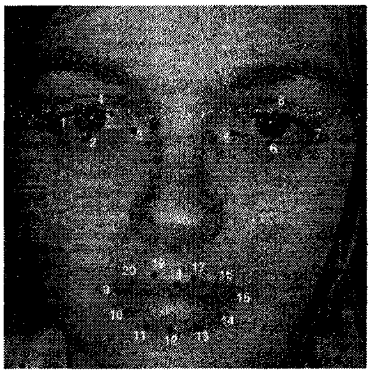


Figure 3.6: A face image with labeled points.

3.2.2 Active Shape Models

Active shape models [19, 18] are statistical models of the shapes of objects which iteratively deform to fit into an example of the object in a new image. The shape is constrained by a point distribution model. We use an active shape model to find facial feature points. We require a set of training images to be able to build a statistical model. Each picture of a face for a training set must be annotated with a set of points defining the key facial features. Each shape in the training set is represented by a set of n labeled landmark points, which must be consistent from one shape to the next, see Figure 3.6. We place our points around main facial features such as: eyes, nose and mouth. These points are used to define the correspondences across the training set and represent the shape of the face in the image.

For instance, on a face example, the first point may correspond to the right corner of the right eye. Given a set of such labeled training examples, we align them into a common coordinate frame. This is accomplished by translating, rotating and scaling each training shape so as to minimize the sum of squared distances to the mean of the set of points. When we use n points to describe the face shape, the shape can be represented as the 2n element vector $\mathbf{x}$ where

$$\mathbf{x} = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T \tag{3.10}$$

To be able to perform statistical analysis on these vectors, we have to represent the shapes in the same coordinate frame. In our project, we use Procrustes approach [24]. To align single shape x to the mean $\bar{\mathbf{x}}$, choose the parameters t which minimize $| S_t(x) - \bar{\mathbf{x}} |^2$, where $S_t(x)$ specifies a transformation defined by parameters in vector $\mathbf{t}$.

To be able to do the alignment of a set of shapes, we use an iterative approach, defined as follows:

1. Translate each shape so its center of gravity is at the origin.
2. Choose one example as an initial estimate of the mean shape and scale so $| \bar{\mathbf{x}} |= 1$.
3. Record the first estimate as $\bar{\mathbf{x}}_0$ to define the default reference frame.
4. Align all the shapes with the current estimate of the mean shape.
5. Reestimate the mean from aligned shapes.
6. Apply constraints on the current estimate of the mean by aligning it with $\bar{\mathbf{x}}_0$ and scaling so $| \bar{\mathbf{x}} |= 1$.
7. If not converged, return to step 4.

On convergence, all the samples are aligned in a common coordinate frame and can be analyzed for shape change.

Suppose now we have s sets of n points $\mathbf{x}_i$ in d dimensions that are aligned into a common coordinate frame. These vectors form a distribution in nd dimensional space.

We can simplify the problem by reducing the dimensionality of the data from nd using principal component analysis (PCA). The data form a cloud of points in the nd-D space. PCA is used to pick out the main axes of the cloud, and model only the first few, which account for most of the variation. It is done by:

1. Compute the mean of the data

$$\bar{\mathbf{x}} = \frac{1}{s}\sum_{i=1}^{s}\mathbf{x}_i \tag{3.11}$$

2. Compute the covariance of the data

$$\mathbf{S} = \frac{1}{s-1}\sum_{i=1}^{s}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{3.12}$$

3. Compute the eigenvectors ϕ_i and corresponding eigenvalues λ_i of $\mathbf{S}$

The shape model is then given by:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P_s b_s} \tag{3.13}$$

where $\mathbf{P_s} = (\phi_1|\phi_2|\ldots|\phi_t)$ contains the t eigenvectors corresponding to the largest eigenvalues, and $\mathbf{b_s}$ is a t dimensional vector given by:

$$\mathbf{b_s} = \mathbf{P}_s^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{3.14}$$

$\mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}$ because the eigenvectors are orthogonal.

By choosing a set of shape parameters $\mathbf{b_s}$ for the model we define the shape of the object in an object-oriented coordinate frame. We can create an instance $\mathbf{X}$ of the model in the image frame by defining the position, orientation, and scale parameters $\mathbf{t}$.

The iterative approach to improving the fit of the instance $\mathbf{X} = S_t(\mathbf{x})$ to an image is done as follows:

1. Examine a region of the image around each point X_i to find the next nearby match X_i' point,
2. Update the parameters $(\mathbf{t}, \mathbf{b_s})$ to best fit the new points X',
3. Repeat until convergence.

3.2.3 Gabor filters

Since Gabor features are robust against local distortions caused by expression, pose and illumination, Gabor filter has been successfully applied to face recognition and verification.

Gabor wavelet is a complex planar wave restricted by a two-dimensional Gaussian envelope. We use 2D Gabor wavelet, which is usually applied for image representation. Gabor filters are also used for image analysis and in computer vision. Gabor filters have a high computational cost because use of Gabor filters requires convolution of the image with Gabor filters in different scales and orientations.

2D Gabor functions were proposed by Daugman [22] to model the spatial summation properties of the receptive fields of simple cells in the visual cortex. A Gabor filter is constructed by modulating a sine / cosine wave with a Gaussian. A 2D Gabor filter over an image domain (x, y) is defined as follows:

$$G(x,y) = e^{-[\frac{(x-x_0)^2}{\alpha^2}+\frac{(y-y_0)^2}{\beta^2}]}e^{-2\pi i[u_0(x-x_0)+\nu_0(y-y_0)]} \tag{3.15}$$

where (x_0, y_0) specify position in the image, (α, β) specify filter's effective width and lenght, and (u_0, ν_0) specify the filter's modulation wave vector, which can be interpreted in polar coordinates as spatial frequency $\omega_0 = \sqrt{u_0^2 + \nu_0^2}$ and orientation $\theta_0 = \arctan(\frac{\nu_0}{u_0})$.

Usually, Gabor wavelets are used with the following parameters:

$$\omega_0 \in \{0, \ldots, 4\} \tag{3.16}$$

$$\theta_0 \in \{0, \ldots, 7\} \tag{3.17}$$

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor wavelets. The convolution of an image I and a Gabor filter $G(x, y)$ is defined as follows:

$$G_{\omega_0,\theta_0}(x,y) = I(x,y) * G(x,y) \tag{3.18}$$

where $z = (x, y)$.

In our project, we first convolved each image with Gabor filters; however, adopting the whole Gabor image feature representation, the computation time for classification was very high. Therefore, to avoid that, we extracted only the pixels that represent facial points from the Gabor magnitude part that represent dominant features of the face, such points describe eyes, mouth, and nose. The size of the feature vector after convolution with Gabor filters depends on how many points are extracted from the face. When the image is convolved with Gabor wavelets with 5 frequencies and 8 orientations, the size of feature vector for 1 pixel is 40. Table 3.1 represents the length of the feature vector according to the number of points extracted from the face.

Feature Points	**Size of Feature Vector**
8 points for eyes	320
16 points for eyes	640
8 points for eyes, 4 points for nose	480
16 points for eyes, 4 points for nose	800
12 points for mouth, 8 points for eyes	800
12 points for mouth, 16 points for eyes	1120
12 points for mouth, 4 points for nose	640
12 points for mouth, 4 points for nose, 8 points for eyes	960
12 points for mouth, 4 points for nose, 16 points for eyes	1280

Table 3.1: The size of face feature vector according to the number of points extracted from face.

When a set of 40 complex Gabor wavelets are used, local features are represented by the set of convolution point z, which contains important information at different orientations and scales. Therefore, we obtain two Gabor components for every image pixel: a real part and an imaginary part. In the same way we get two kinds of features: Gabor magnitude features and Gabor phase features. In our experiments we use only the magnitude part. In our experiments we use Gabor wavelets with 5 frequencies and 8 orientations, see Figures (3.7) - (3.8) for the whole set of 40 Gabor wavelets. The Gabor wavelets with those parameters exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity.

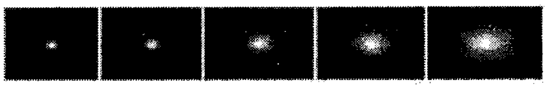


Figure 3.7: The magnitude at five scales.

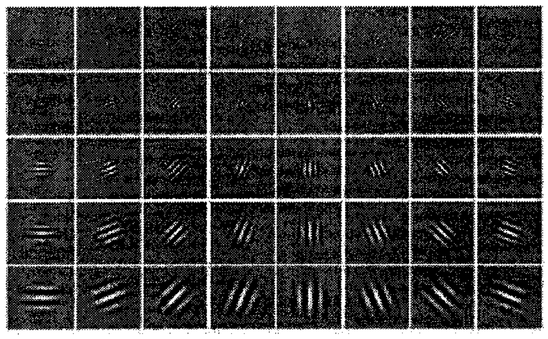
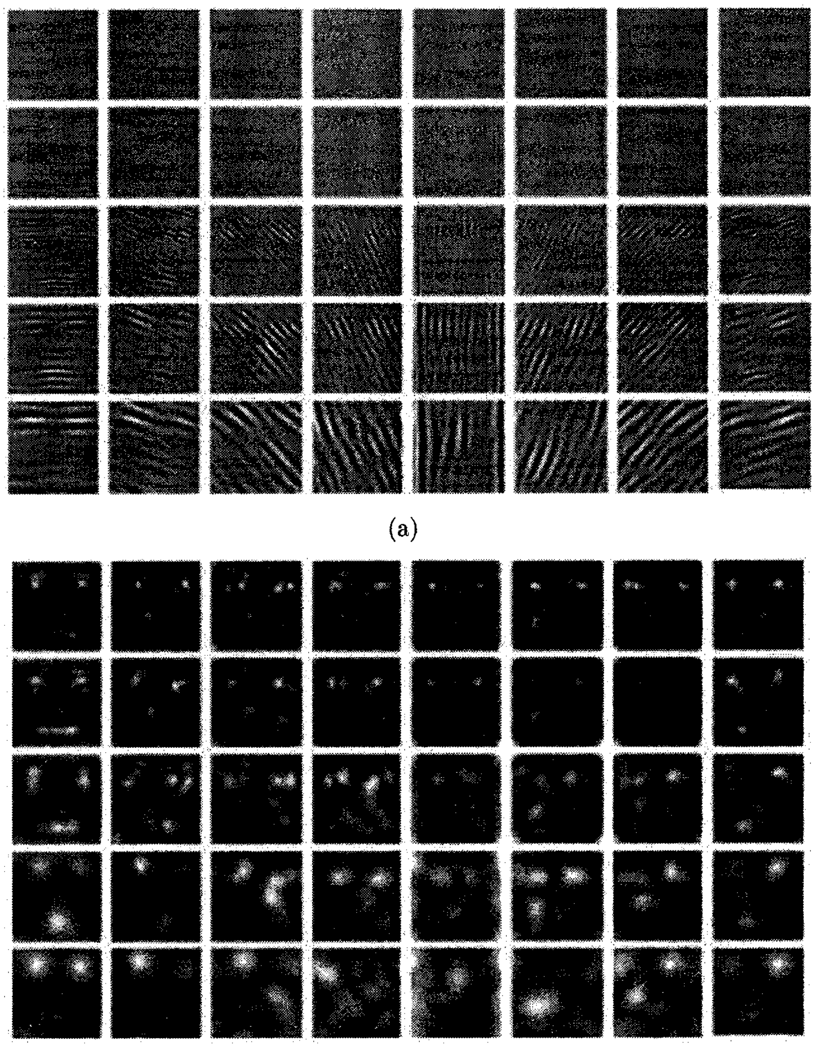


Figure 3.8: The real parts at five scales and eight orientations.

Figure 3.9 shows the Gabor wavelet representation (the real part and the magnitude) of a sample image. These representation results display scale, locality, and orientation properties corresponding to those displayed by the Gabor wavelets in Figures 3.7 - 3.8.

3.3 Hand Recognition

Hand geometry is an example of biometric based on physiological characteristics. Hand geometry refers to the geometric structure of human hand. Hand geometry



(a)

(b)

Figure 3.9: Gabor wavelet representation (the real part and the magnitude) of a sample image. (a) The real part of the representation and (b) the magnitude of the representation.

recognition systems are quite good for environments where medium security is required, and where a medium cost for equipment, and relative low computational costs are needed [47]. Hand recognition systems are not new; some initial work in this domain can be traced since the 1970s along with the existence of commercial systems in the market [2]. Most of the hand-based biometric schemes in the literature fall into the broad category of geometric features of the hand such as length and width of fingers, width of palm, aspect ratio of the palm or fingers, thickness of the palm, etc. [33]. For instance [62] selects 25 features, such as finger widths at different latitudes, finger and palm heights, finger deviations and the angles of the interfinger valleys with the horizontal, and model them with Gaussian mixtures; whereas, [33] obtained 16 features, which include length and width of the fingers, aspect ratio of the palm to fingers, and thickness of the hand. Also [9] used 30 geometric hand features including the length and height of fingers and palm.

In our approach, a hand image is captured by the same cell phone camera that was used to take a face image. The cell phone is fixed above the platform on which the hand is placed, as shown on Figure 4.1. The platform and location of the cell phone are designed to put the hand at a fixed location. In our experiments, only left hand is taken into consideration. We asked people to remove rings and other jewelery from their fingers.

To detect points in a hand image, we use the same algorithm (active shape models) as for face points, described in section 3.2.2. We built the system that detects nine points from the hand image. As shown in Figure 3.10, four of those points are at the end of four fingers (excluding thumb), and the next five are located at the base of each finger. These points represent vertices that we use in measuring the hand. Also in our approach, we use the basic measurements for a hand. Figure 3.11 shows nine geometric features that are extracted from the hand image. They are:

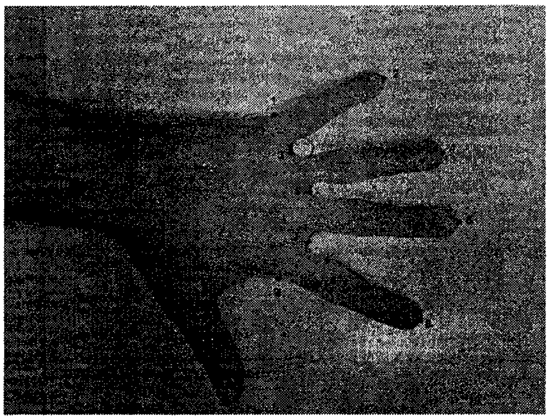


Figure 3.10: Point distribution model for hand.

1. The lengths of four fingers: index finger, middle finger, ring finger, and small finger.
2. The widths of four fingers: index finger, middle finger, ring finger, and small finger.
3. The width of the palm.

The width of each of the four fingers is extracted by calculating the distance between the two base points. In order to calculate the lengths, we have to find the middle point of the two base points. When this point is found, the length of each finger is calculated by finding the distance between the end point and the middle point. We repeat this procedure for four fingers.

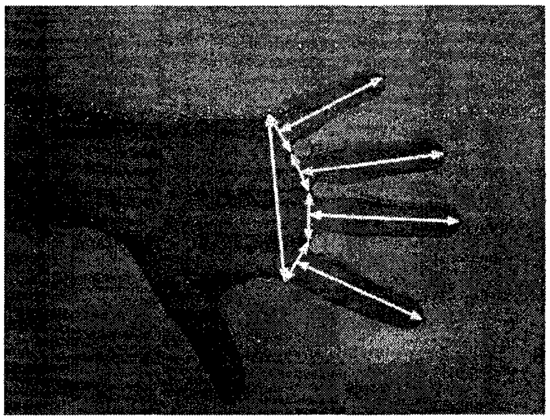


Figure 3.11: Nine hand geometry features.

3.4 Multimodal Biometric System

To make the system more secure, researchers often combine different biometric methods together. The drawback of multiple biometrics is that the user has to use multiple sensors, eg. face and voice [52] (the face has to be captured by a camera, and voice has to be recorded), or face and iris [13] (to take those pictures, two different cameras have to be used), or face and fingerprint [28] (where, the face is captured by a camera, and a fingerprint by a special scanner). In our approach we use the same sensor, a cell phone camera, to take pictures of both the user's face and hands.

The face feature vectors and hand geometry are combined at the feature extraction level. The combined feature vector is built by concatenating the face features with hand features. Figure 3.12 represents the block diagram of the fusion for face and hand feature vectors.

The face feature vectors contain the imaginary part of the Gabor filters convolved with the points extracted from the face, ie. eyes, nose, mouth. The hand feature vector includes the basic geometry feature for the hand, described in section 3.3. The

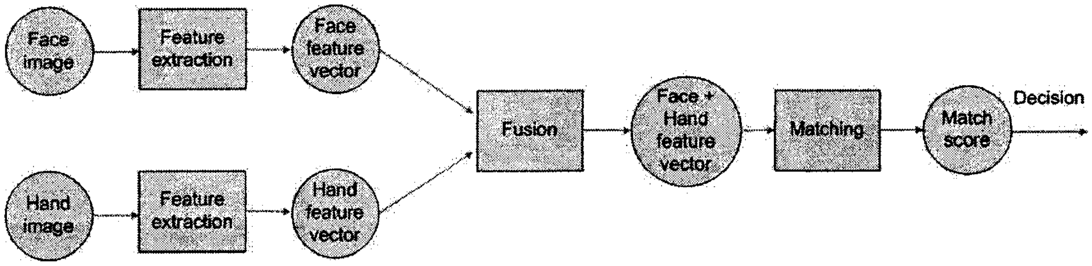


Figure 3.12: The block diagram of the fusion for face and hand.

same classifier, support vector machines, is used for multimodal biometric system.

To the best of our knowledge, the multibiometrics based on face and hand geometry does not exist. We decided to combine these two characteristics, because only one sensor is needed to capture both face and hand images. The experiments in section 4.5 shows that the multimodal system improved the accuracy rates as compared to face verification and hand verification separately.

3.5 Support Vector Machines

Support vector machines (SVM) have been accepted as a powerful tool for developing pattern classification and function approximation systems [1]. The support vector classifiers devise a computationally efficient way of learning good separating hyperplane in a high dimensional feature space [67]. In this research, we apply SVM to verify the identity of a cell phone user. During training we use labeled facial images to design a classifier which is capable of separating a genuine user from an intruder. For a two-class problem, the support vector machine is trained so that direct decision function maximizes the generalization ability [66, 44, 20]. In SVM, the data that satisfy the equalities are called support vectors. In Figure 3.13 the data corresponding to the filled circles and the filled rectangle are support vectors. SV are selected from the training data so that the distance between two hyperplanes represented by lines passing through the support vectors is at maximum. For linearly separable cases, the main decision hyperplane, which is the midway line between two hyperplanes on Figure 3.13 should minimize classification error [15].

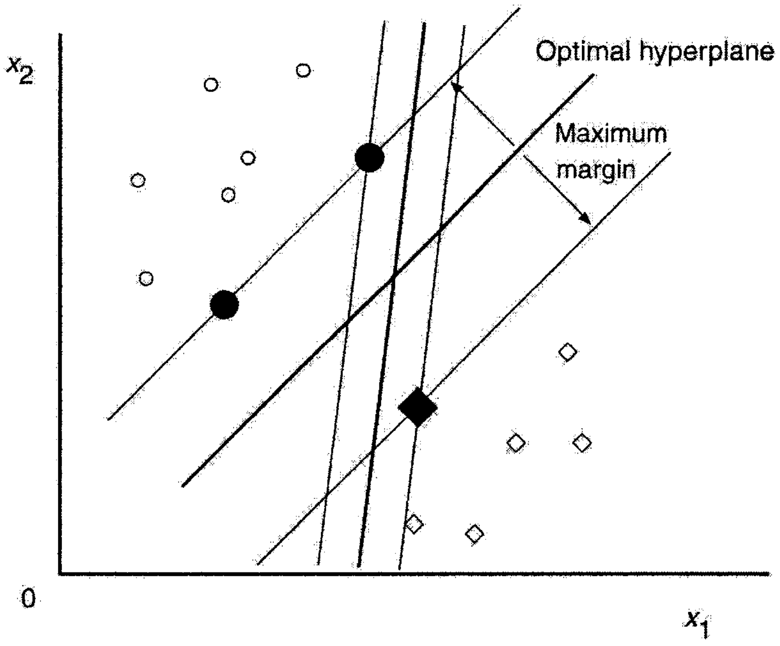


Figure 3.13: Optimal separating hyperplane in a two-dimensional space.

Suppose we are given a training set of input/output pairs $(x_i, y_i), x_i = 1, 2, \ldots, M$ where $x_i \in R^n$ and $y \in \{-1, 1\}^l$ are the labels. The decision function in figure (3.13) is determined as:

$$D(x) = w^T x + b \tag{3.19}$$

where w is an m-dimensional vector, b is a bias term. For $i = 1, \ldots, M$

$$w^T x_i + b \begin{cases} \geq 1 & for \quad y_i = 1 \\ \leq -1 & for \quad y_i = -1 \end{cases} \tag{3.20}$$

The hyperplane:

$$D(x) = w^T x + b = c \; for \; -1 < c < 1 \tag{3.21}$$

forms a separating hyperplane that separates $x_i (i = 1, \ldots, M)$. When $c = 0$, the separating hyperplane is in the middle of the two hyperplanes with $c = 1$ and $c = -1$. The distance between the separating hyperplane and training data nearest to the hyperplane is called the margin. The region $\{x | -1 \leq D(x) \leq 1\}$ is the generalization region for the decision function. The optimal separating hyperplane can be obtained by minimizing

$$Q(w) = \frac{1}{2} ||w||^2 \tag{3.22}$$

with respect to w and b subject to the constraints by combining (3.20):

$$y_i (w^T x_i + b) \geq 1 \; for \; i = 1, \ldots, M \tag{3.23}$$

In the training phase, the goal is to find support vectors that maximize the margin of separation, or to minimize $||w||^2$. It can be solved by:

$$Q(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{M} \alpha_i \{ y_i (w^T x_i + b) - 1 \} \tag{3.24}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ and α_i are the nonnegative Lagrange multipliers. The optimal solution of the equation (3.24) is obtained when the (3.24) is simultaneously minimized with respect to w and b and maximized with respect to α_i.

$$\frac{\partial Q(w, b, \alpha)}{\partial w} = 0 \quad and \tag{3.25}$$

$$\frac{\partial Q(w,b,\alpha)}{\partial b} = 0 \tag{3.26}$$

According to Karush-Kuhn-Tucker (KKT) theorem, the following conditions must be satisfied at all of the saddle points [1]:

$$\alpha_i\{y_i(w^T x_i + b) - 1\} = 0 \quad for \quad i = 1, \ldots, M \tag{3.27}$$

$$\alpha_i \geq 0 \quad for \quad i = 1, \ldots, M \tag{3.28}$$

Thus, only those data points for which $y_i(w^T x_i + b) - 1 = 0$ can be support vectors because only these points have a nonzero value of multipliers α_i.

From (3.27), $\alpha_i = 0$, or $\alpha_i \neq 0$ and $y_i(w^T x_i + b) = 1$ must be satisfied. The training data x_i with $\alpha \neq 0$ are called support vectors and these data are considered to be most important for the classification problem. Using (3.24), we reduce (3.25) and (3.26), respectively to:

$$w = \sum_{i=1}^{M} \alpha_i y_i x_i \quad and \tag{3.29}$$

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \tag{3.30}$$

By substituting (3.29) and (3.30) into (3.24), we obtain the following dual problem. Maximize

$$Q(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{3.31}$$

with respect to α_i subject to the constraints:

$$\sum_{i=1}^{M} y_i \alpha_i = 0 \quad and \tag{3.32}$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, M \tag{3.33}$$

The obtained support vector machine is called the hard-margin support vector machine because:

$$\frac{1}{2} \sum_{i,j=1} M \alpha_i \alpha_j y_i y_j x_i^T x_j = \frac{1}{2} \left(\sum_{i=1}^{M} \alpha_i y_i x_i \right)^T \left(\sum_{i=1}^{M} \alpha_i y_i x_i \right) \geq 0 \tag{3.34}$$

maximizing (3.32) under the constraints (3.33) is a concave quadratic programming.If a solution exists then the global optimal solution $\alpha_i (i = 1, \ldots, M)$ exists.

From equation (3.29) the decision function is given by:

$$D(x) = \sum_{i \in S} \alpha_i y_i x_i^T x + b \tag{3.35}$$

where S is the set of support vectors indices, and from (3.27), we obtain threshold b:

$$b = y_i - w^T x_i \tag{3.36}$$

In hard-margin support vector machines, we assume that the training data is linearly separable; however, that does not always happen. For the data that is not linearly separable, the soft margin support vector machines are introduced. In order to solve the problem of inseparability, the nonnegative slack variables $\xi_i \geq 0$ are introduced. Figure 3.14 shows the inseparable case in a two-dimensional space. To add slack variables to equation (3.23), we obtain:

$$y_i(w^T x_i + b) \geq 1 - \xi \quad for \quad i = 1, \ldots, M \tag{3.37}$$

In this case, we have to minimize the following:

$$Q(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{M} \xi_i \tag{3.38}$$

subject to the constraints:

$$y_i(w^T x_i + b) \geq 1 - \xi \quad for \quad i = 1, \dots, M \tag{3.39}$$

where $\xi = (\xi_1, \dots, \xi_M)^T$ and C is the margin parameter that determines the trade-off between the maximization of the margin and minimization of the classification error. Now, the new cost function including nonnegative Lagrange multipliers α_i and β_i becomes:

$$Q(w, b, \xi, \alpha, \beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}\xi_i - \sum_{i=1}^{M}\alpha_i(y_i(w^T x_i + b) - 1 + \xi_i) - \sum_{i=1}^{M}\beta_i\xi_i \tag{3.40}$$

where $\alpha = (\alpha_1, \dots, \alpha_M)^T$ and $\beta = (\beta_1, \dots, \beta_M)^T$.

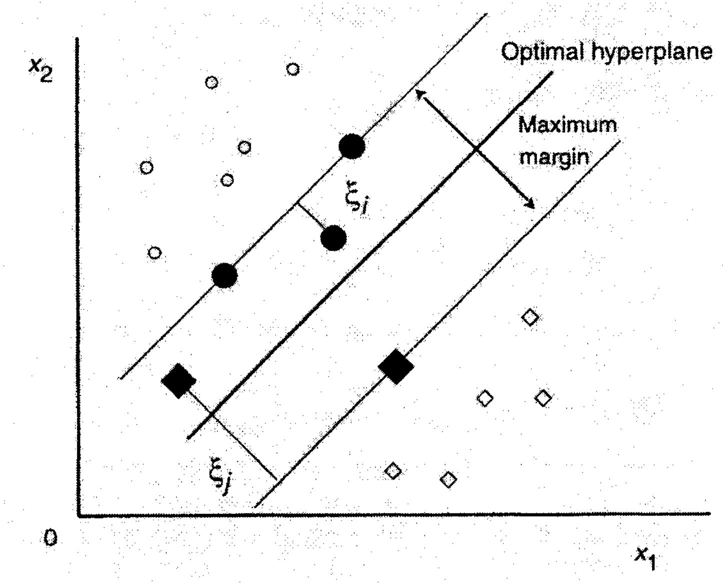


Figure 3.14: Inseparable case in a two-dimensional space.

For the optimal solution, the KKT conditions must be satisfied:

$$\frac{\partial Q(w, b, \xi, \alpha, \beta)}{\partial w} = 0 \quad and \tag{3.41}$$

$$\frac{\partial Q(w,b,\xi,\alpha,\beta)}{\partial b} = 0 \quad and \tag{3.42}$$

$$\frac{\partial Q(w,b,\xi,\alpha,\beta)}{\partial \xi} = 0 \tag{3.43}$$

$$\alpha_i(y_i(w^T x_i + b) - 1 + \xi) = 0 \quad for \quad i = 1,\dots,M \tag{3.44}$$

$$\beta_i \xi_i = 0 \quad for \quad i = 1,\dots,M \tag{3.45}$$

$$\alpha_i \geq 0, \quad \beta_i \geq 0, \quad \xi_i \geq 0 \quad for \quad i = 1,\dots,M \tag{3.46}$$

Using (3.40), we reduce (3.41)- (3.43), respectively:

$$w = \sum_{i=1}^{M} \alpha_i y_i x_i \tag{3.47}$$

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \tag{3.48}$$

$$\alpha_i + \beta_i = C \quad for \quad i = 1,\dots,M \tag{3.49}$$

Thus, substituting (3.47)- (3.49) into (3.40), the following dual problem is obtained:

$$\max_{\alpha} \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\alpha_i \alpha_j y_i y_j x_i^T x_j \tag{3.50}$$

subject to $0 \leq \alpha_i \leq C$, $\sum_{i=1}^{M} y_i \alpha_i = 0$ for $i = 1,\dots,M$.

When the classification problem is not linearly separable in the input space, we need to use a kernel function $K(x_,x_j)$ to map the data x from the input space to the new higher dimensional feature space, where the mapped points become linearly separable. In this thesis, we used radial basis function kernel (RBF), which is defined as follows:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma \geq 0 \tag{3.51}$$

In our experiments, we used RBF kernel for soft margin support vector machine for face and hand verification, and for a combination of the two. We use the LIBSVM software [12], which is an integrated software for support vector classification.

Chapter 4

Experiments

4.1 Introduction

This chapter describes experiments that were performed for this thesis. This chapter starts describing the process of building the database described in section 4.2, followed by experiments and analysis for the face authentication method. Section 4.3 shows an analysis of experiments based on different facial models and a well-known method called eigenfaces. Also in this section, the comparison between all face authentication methods is presented. Section 4.4 represents experiments that were run for the hand authentication system. The last section discusses multimodal biometric system based on face and hand cell phone images.

4.2 Face and Hand Database

An important part of the research was gathering the database by using a cell phone digital camera. The cell phone that was used to take pictures was a Sony Ericsson K750i. The telephone is able to take pictures up to 1632x1224 pixels; however the resolution that we used for the pictures was 640x480. The camera is equipped with zoom 4x. White balance in the camera was set up in automatic mode, and depending on the lighting conditions, the images were taken with or without flash. The images for one person were taken during one session. First we collected 16 face images and then 4 hand images for each person.

The database contains images from 113 individuals, men and women. The individuals in the database differ by race: Caucasian, Asian, and African. Their ages also differ. Most of the people are between 18 - 60 years old. The oldest person is 75; however, the average age is between 30 - 40 years old. Some of the men have a beard or mustache.

For face pictures, we asked people to take pictures of themselves, more precisely of their faces, holding the camera in front of them. The camera was held in front of the face and zoom was set at 1.5 times. To take the picture, a subject (person) would use a small mirror built into the camera next to the lens.

Hand pictures were captured by the same cell phone camera that was used to take the face images. To be able to capture hand images at a fixed position and distance, the cell phone was located above the platform on which the hand is placed, as shown on Figure 4.1. The platform and location of the cell phone were designed to put the hand at a fixed location. We asked people to remove rings and other jewelery from their fingers. The zoom was also set at 1.5 times, and the flash light was on.

Pictures were taken under different lighting conditions. As you can see on Figure 4.2, the first 16 images were taken in a dark environment with the flash light which has 6000K. The last 16 pictures were taken under the tungsten light which has

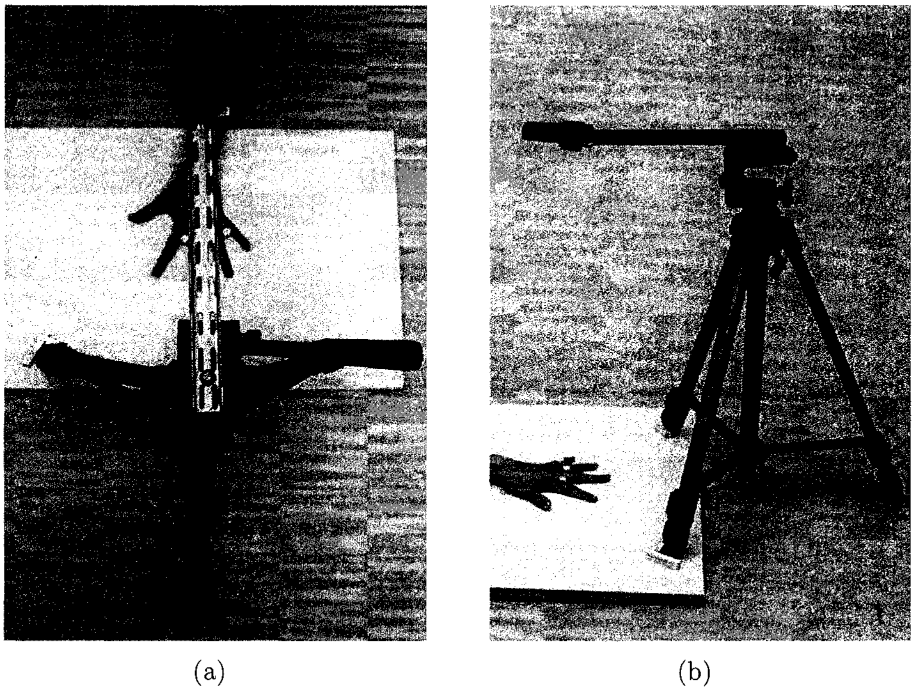


Figure 4.1: The platform that was used to keep the cell phone in a fixed distance to be able to capture the hand images. (a) top view, (b) side view

3200K. In comparison, the day light has 5000K. Appendix A contains sample face and hand images for each individual involved in our experiments.

In some of the images, the subject moved while taking the pictures which resulted in blur and lack of sharpness. The camera used had auto focus but it is not as good as the auto focus in digital cameras.

4.3 Face Authentication

In order to apply active shape models to face images, we have to preprocess the face images. First, the color image is converted to grey scale. Then, each image is manually cropped to 350x350 pixels in order to eliminate the background, and hair. By cropping the image, we extracted only part of the head that contains the eyes,

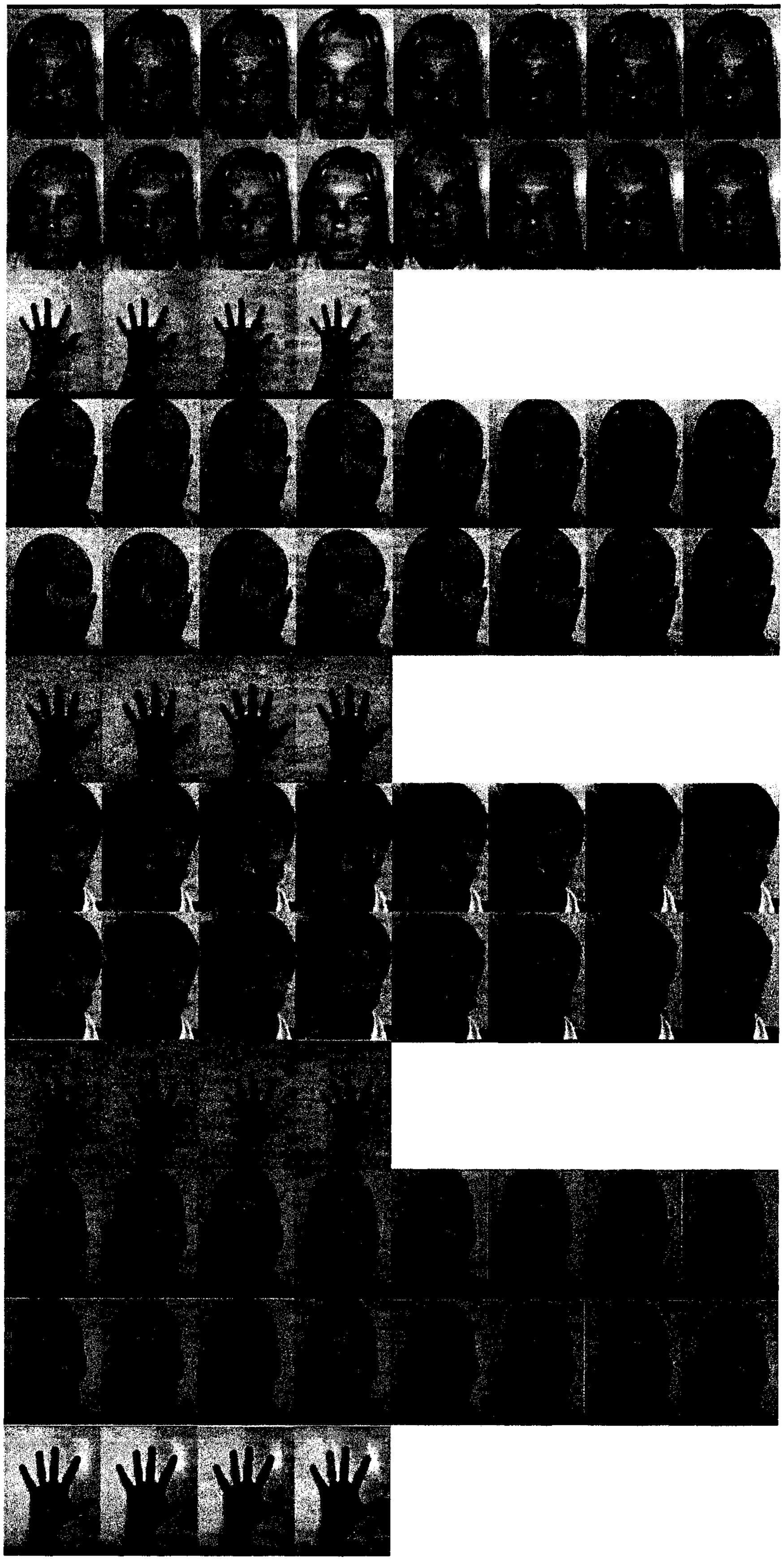
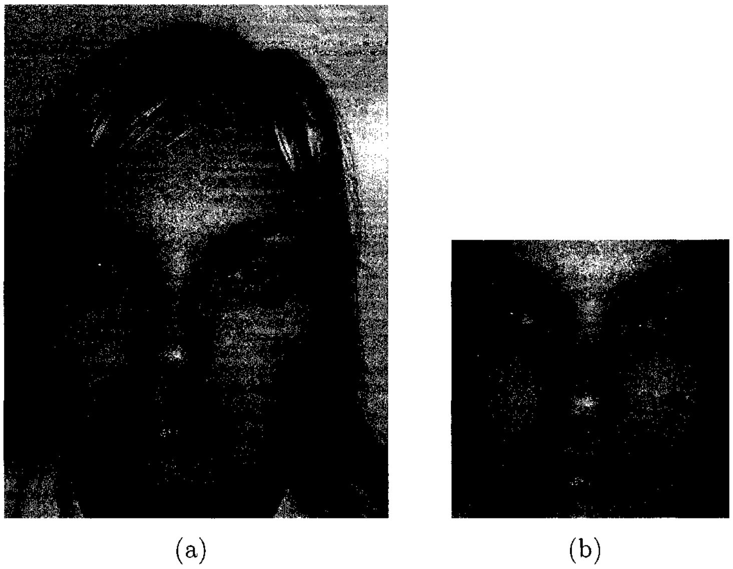


Figure 4.2: Sample face and hand pictures from our database.



(a) (b)

Figure 4.3: (a) The original image of a face. (b) The cropped image of the same face picture.

nose and mouth, see Figure 4.3. We use active shape models from section 3.2.2 on this preprocessed image to detect the facial points. Then the image is convolved with Gabor filters described in section 3.2.3 in order to obtain face feature vectors.

We built nine different face models to detect various facial features. To build these models, we manually labeled 4 out of 16 images for each person. Figure 4.4 shows the point distribution model for 9 different face models. They include:

- 4 points for each eye,
- 8 points for each eye,
- 4 points for each eye plus 4 points for a nose,
- 8 points for each eye plus 4 points for a nose,
- 12 points for a mouth plus 4 points for each eye,

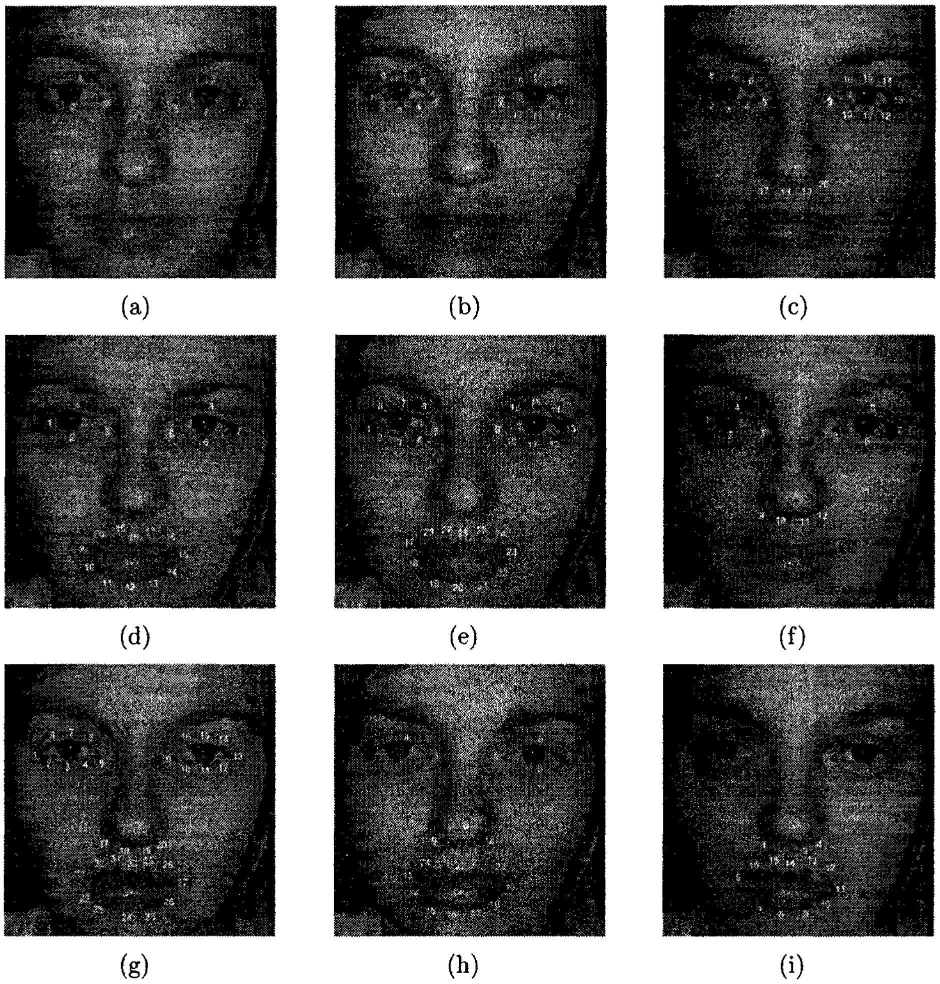
- 12 points for a mouth plus 8 points for each eye,
- 12 points for a mouth plus 4 points for a nose,
- 12 points for a mouth plus 4 points for a nose plus 4 points for each eye,
- 12 points for a mouth plus 4 points for a nose plus 8 points for each eye.

We convolve these face models with Gabor filters, as discussed in section 3.2.3, and according to the face model, we obtain following feature vectors shown in Table 4.1:

Feature Points	**Size of Feature Vector**
8 points for eyes	320
16 points for eyes	640
8 points for eyes, 4 points for nose	480
16 points for eyes, 4 points for nose	800
12 points for mouth, 8 points for eyes	800
12 points for mouth, 16 points for eyes	1120
12 points for mouth, 4 points for nose	640
12 points for mouth, 4 points for nose, 8 points for eyes	960
12 points for mouth, 4 points for nose, 16 points for eyes	1280

Table 4.1: The size of face feature vectors according to the number of points extracted from face.

In our experiments, we used RBF kernel for face and hand verification, and for a combination of the two. We use the LIBSVM software [12], which is an integrated software for support vector classification. The SVM is built for each person. 50% of the face pictures are used for training and the remaining 50% are used for testing. Training and testing are performed according to face models; therefore, we obtained nine different classifications. In addition to those classifications, we also tested one positive person against 5, 10, 15, 20, 25 and 30 impostors. Figure 4.5 shows the



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 4.4: Nine different face point distribution models: (a) 8 eye points, (b) 16 eye points, (c) 16 eye points, 4 nose points (d) 8 eye points, 12 mouth points, (e) 16 eye points, 12 mouth points, (f) 8 eye points, 4 nose points, (g) 16 eye points, 4 nose points, 12 mouth points, (h) 8 eye points, 4 nose points, 12 mouth points, (i) 12 mouth points, 4 nose points.

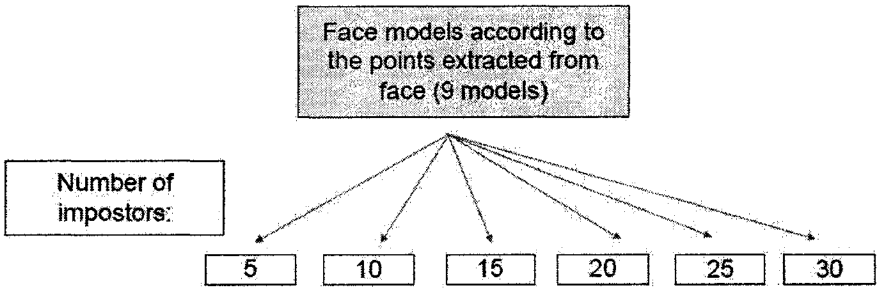

Face models according to the points extracted from face (9 models)
Number of impostors:
5
10
15
20
25
30

Figure 4.5: The diagram of the experiments performed according to the facial point models and the number of impostors.

graphical interpretation of the experiments we ran for this thesis.

The Table 4.2 presents face verification results according to the nine facial point models extracted from the face and number of impostors.

As we can see from Table 4.2, in the facial model, which contain 8 eye points, the best average result, 99.43%, is obtained when one person is tested against 10 impostors. The minimum recognition rate achieved is 90.56%. It is obtained when the tests are run against 30 impostors. On the other hand, in every group the highest accuracy is 100%. We notice that the average accuracy has its peak when the system is tested against 10 negative individuals, and then the average accuracy decreases obtaining 99.19%, 99.07%, 98.82% and 97.58% for 15, 20, 25, 30 impostors respectively.

The next model from the Table 4.2 is based on 16 eye points - 8 points from each eye. Here, also the maximum accuracy that is obtained is 100% for all groups. The minimum accuracy is achieved when the tests are performed against 5 impostors, and it is 95.83%. We observe that the average accuracy is increasing, and obtains its maximum value 99.45%, when the tests are run against 15 negative individuals. After it reaches the maximum value, it then starts decreasing up to 98.51% for 30

Feature Points			Accuracy	Number of impostors					
Eyes	Nose	Mouth	%	5	10	15	20	25	30
8	-	-	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	96.09	98.11	96.85	90.56
			average	98.95	99.43	99.19	99.07	98.82	97.58
16	-	-	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.43	98.74	98.11	97.48
			average	99.14	99.37	99.45	99.04	98.82	98.51
8	4	-	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.43	98.74	98.11	98.11
			average	99.33	99.54	99.68	99.20	99.01	98.85
16	4	-	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.48	98.11	97.48	96.22
			average	99.24	99.43	99.60	99.08	98.67	98.14
8	-	12	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.48	98.74	98.11	97.48
			average	99.14	99.54	99.72	99.23	99.01	98.51
16	-	12	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	96.09	96.22	97.48	96.22
			average	99.05	99.08	99.29	98.82	98.53	98.38
-	4	12	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.48	98.74	98.74	97.48
			average	99.04	99.21	99.33	99.09	98.97	98.78
8	4	12	maximum	100	100	100	100	100	100
			minimum	95.83	97.72	98.48	98.74	97.48	97.48
			average	99.14	99.60	99.72	99.20	98.98	98.70
16	4	12	maximum	100	100	100	100	100	100
			minimum	93.75	97.72	96.87	98.74	98.11	96.85
			average	98.57	99.08	99.29	99.04	98.66	98.55

Table 4.2: The maximum, minimum, and average accuracy based on 9 facial models extracted from the face and the number of impostors involved in testing.

impostors involved in the tests.

The next facial model, which is used in our experiments, contains 12 facial points. Eight of them represent eyes, and the next four represent the nose. This facial model obtains better average accuracy rates than the models described in previous sections. However, the increasing tendency is the same, and again the maximum average accuracy (99.68%) is achieved when tests are performed against 15 impostors. After reaching the peak, the average accuracy then drops off to 98.85% for 30 negative individuals involved in the test. Again, 100% is gained for all tests that were run for this facial model. The same minimum value, as for the previous model, is reached. It is 95.83%, and again it is achieved for the same number of impostors (5).

Another facial model is based on 16 eye points and 4 nose. The results are slightly poorer, as the ones when 8 eye points and 4 nose points were extracted. We can observe the same curve tendency for this facial model. Again, the average accuracy is growing from 99.24% for 5 impostors in order to reach its peak value at 99.60% for 15 impostors, and then goes down to 98.14% for 30 impostors. Also, the minimum accuracy, 95.83%, is achieved for 5 impostors.

Facial model, which contains 8 eye points, and 12 mouth points obtained the best average result of 99.72% when one person is tested against 15 impostors. The minimum recognition rate achieved is 95.83%. It is obtained when the tests are run against 5 impostors. Also, in every group the highest accuracy is 100%. We notice that the average accuracy has its peak when the system is tested against 15 negative individuals, and then the average accuracy decreases obtaining 99.23%, 99.01%, 98.51% for 20, 25, 30 impostors respectively.

Face verification results based on 16 eye points and 12 mouth points extracted from the face obtained worse average accuracy rates than other models. However, the increasing tendency is the same, and again the maximum average accuracy (99.29%) is achieved when tests are performed against 15 impostors. After reaching the peak, then the average accuracy drops off to 98.38% for 30 negative individuals involved in the test. The same minimum value, as for the previous model, is reached. It is 95.83%, and again it is achieved for the same number of impostors (5). Also, 100% is reached for all tests ran for this facial model.

The next facial model that is presented in the Table 4.2 is based on 12 mouth points and 4 nose points. As we see, the same minimum average of 95.83% is obtained for 5 impostors. The same increasing tendency is observed, in which it reaches its top average accuracy for 15 impostors, and then it decreases. The highest average accuracy achieved for this model is 99.33%; whereas, the smallest is 98.78%, obtained for 30 impostors. Again, as in previous facial models in every group of impostors, the 100% has been accomplished.

For the last two models, we combine three main parts of the face together, such as the eye, the nose and the mouth. The facial model based on 8 eye points, 12 mouth points and 4 nose points proved to be a good model because the highest average accuracy achieved for 15 impostors is 99.72%. The same growing tendency as in previous facial models is observed here. The average accuracy is increasing starting at 99.14% for 5 impostors, 99.60% for 10 impostors, and finally reaches the peak of 99.72% for 15 impostors. Afterwards, the average accuracy decreases to 99.20% for 20 impostors, 98.98% for 25 impostors and finally 98.70% for 30 impostors. The maximum accuracy of 100% is found in every group.

The last model that is based on points extracted from the face is a model that contains 16 eye points, 4 nose points and 12 mouth points. This model obtains poorer average accuracy rates than the model that contains the same parts of the face, but a less number of points extracted from the eyes. The minimum accuracy obtained is 93.75% for 5 impostors; however, in this facial model the 100% is achieved in every group. The highest average accuracy of 99.29% is reached for 15 impostors. After reaching the peak, the average accuracy then drops off to 98.55% for 30 negative individuals involved in the test.

In order to compare results obtained from the method that we developed in this thesis, we used a well-known method called eigenfaces for face verification based on cell phone images. The next table shows results obtained using the PCA analysis (eigenface) method.

Accuracy	**Number of impostors**					
	5	**10**	**15**	**20**	**25**	**30**
maximum	95.5%	95.25%	96.5%	96.48%	96.48%	97.58%
minimum	90.25%	90.25%	90.25%	94.01%	94.01%	94.01%
average	94.5%	93.65%	95.25%	95.66%	96.03%	96.82%

Table 4.3: The maximum, minimum, and average accuracy based on eigenfaces and the number of impostors involved in testing.

As we expected, the results for eigenfaces are much poorer the the ones that used facial points convolved with Gabor filters. Here, the maximum average of 97.58% is achieved for 30. In addition, 100% was never obtained for all tests that we performed. The highest average accuracy of 96.82% was also obtained for 30 impostors. 93.65%

is the minimum average accuracy achieved for 10 impostors. The minimum values reached are low compared to minimum values obtained in our approach, and they are between 90.25% to 94.01%.

Figure 4.6 represents the graphical comparison of the face verification methods - all nine facial models and eigenfaces.

As we observe, the best average results are obtained when 15 impostors were involved in the testing phase. The highest average accuracy of 99.72% is achieved for two facial models. One of these two models is based on 8 eye points and 12 mouth points, and the other one is based on 8 eye points, 4 nose points and 12 mouth points. From our observations, we concluded that the facial model built from the points extracted from the eyes (8 points), the nose (4 points) and the mouth (12 points) is the best model for face verification. The facial model that contains 8 eye points and 4 nose points is also a good model; in addition, this model achieved the highest average accuracy rates for 30 impostors, and its feature vector is much smaller than the one for the facial model that contain eyes, nose and mouth. From the experiments, we noticed that in general, the facial models that contain the combination of 16 eye points performed worse than the ones that contain 8 eye points. It is clearly visible from the Diagram 4.6, that the eigenface approach is not as good as our approach for face verification based on cell phone images. In the eigenface approach, the highest average accuracy is obtained for 30 impostors. The average accuracy decreases to 93.65% when 10 impostors are involved in the testing phase. So, we proved that the method we designed using Gabor filters and points extracted from the face was a good choice.

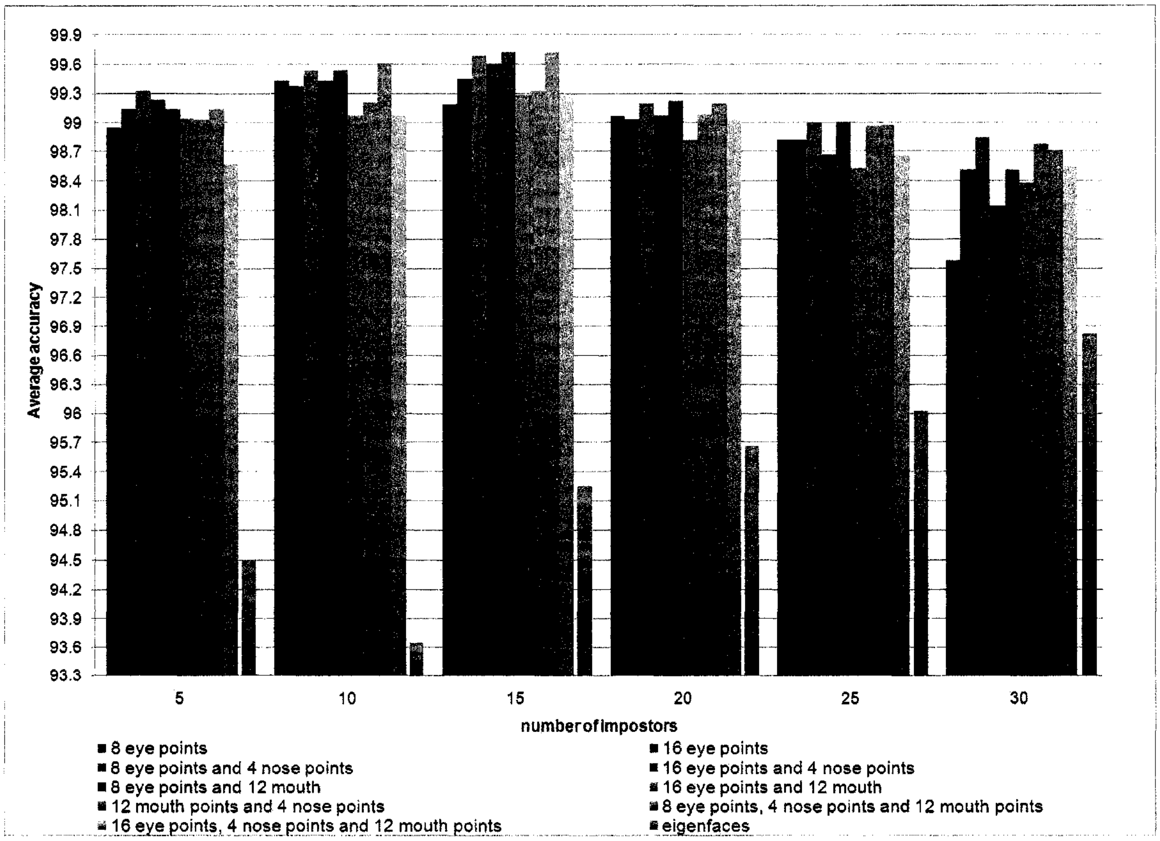


Figure 4.6: The comparison of face authentication methods including 9 different facial models and eigenfaces.

4.4 Hand Authentication

Hand authentication method is another unimodal biometric system used in this thesis. As described in section 3.3, we extracted 9 hand geometry features, including the width of four fingers, the hight of four fingers, and the width of the palm. The SVM is built for each person where 50% of the hand pictures are used for training and the remaining 50% are used for testing. We ran the experiments according to the number of impostors involved in the testing phase. The following table presents results obtained for hand verification system.

Accuracy	**Number of impostors**					
	5	**10**	**15**	**20**	**25**	**30**
maximum	100%	100%	98.76%	98.11%	97.61%	94.26%
minimum	91.91%	89.86%	87.43%	84.43%	81.48%	77.48%
average	96.26%	93.51%	91.67%	89.38%	87.26%	85.13%

Table 4.4: The maximum, minimum, and average accuracy based hand verification and the number of impostors involved in testing.

As we see in Table 4.4, the hand authentication system does not perform as well as face authentication system. The highest average accuracy of 96.26% is achieved when 5 impostors are involved in the testing. Than the average accuracy decreases and reached 85.13% for 30 impostors. Only in two groups the 100% was obtained, when 5 and 10 impostors were involved. When the number of impostors increases, the average decreases. The hand authentication system based on 9 hand geometry features is not a strong biometric method. When more people are involved in the testing phase, the accuracy drops off dramatically.

4.5 Multimodal Face and Hand Authentication

In this thesis, we also combined two biometric methods together, face and hand. The architecture for multibiometrics that is used in this thesis is fusion at the feature level. We concatenate facial features with hand features. Again, we received nine multibiometrics models that are based on nine different facial models plus hand features. The same facial models are used as for the face authentication system. Also, the same experiments used for the face verification system are run for multibiometrics. Table 4.5 presents multibiometric results according to the nine facial point models extracted from the face and hand geometry features, and number of impostors involved in the testing phase.

This section describes multibiometric results based on eight eye points extracted from the face and nine geometry features extracted from the hand. We ran the experiments according to the number of impostors involved in the testing phase. The following table represents the results obtained for this combination of eight facial points and nine hand geometry features.

The multibiometric model, which contains 8 eye points and 9 hand geometry features, proved to be a good combination because the highest average accuracy achieved for 15 impostors is 99.69%. The average accuracy starts increasing at 99.3% for 5 impostors, 99.55% for 10 impostors, and finally reaches its peak of 99.69% for 15 impostors. Afterwards, the average accuracy decreases to 99.54% for 20 impostors, 99.47% for 25 impostors and finally 98.91% for 30 impostors. The maximum accuracy of 100% is found in every group. Figure 4.7 shows the average accuracies for

Feature Points				Accuracy	Number of impostors					
Eyes	Nose	Mouth	Hand	%	5	10	15	20	25	30
8	-	-	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	99.21	98.11	98.11	94.96
				average	99.3	99.55	99.69	99.54	99.47	98.91
16	-	-	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	98.43	98.74	98.11	96.22
				average	99.19	99.43	99.60	99.44	99.37	99.26
8	4	-	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	98.43	98.74	98.74	98.11
				average	99.53	99.69	99.78	99.68	99.58	99.51
16	4	-	9	maximum	100	100	100	100	100	100
				minimum	95.03	97.72	98.48	98.43	98.74	98.11
				average	99.27	99.46	99.64	99.58	99.51	99.44
8	-	12	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	98.48	98.43	98.74	96.22
				average	99.65	99.74	99.82	99.65	99.58	99.54
16	-	12	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	97.65	98.43	97.48	97.48
				average	99.07	99.36	99.60	99.47	99.33	99.12
-	4	12	9	maximum	100	100	100	100	100	100
				minimum	97.91	98.86	98.43	98.43	97.48	97.48
				average	99.26	99.51	99.67	99.38	99.26	99.13
8	4	12	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	98.48	98.43	98.11	97.48
				average	99.42	99.69	99.76	99.65	99.54	99.37
16	4	12	9	maximum	100	100	100	100	100	100
				minimum	95.83	97.72	97.65	97.48	97.48	96.22
				average	98.95	99.36	99.52	99.44	99.23	98.91

Table 4.5: The maximum, minimum, and average accuracy for multimodal bimoetrics based on 9 facial models extracted from the face hand geometry features and the number of impostors involved in testing.

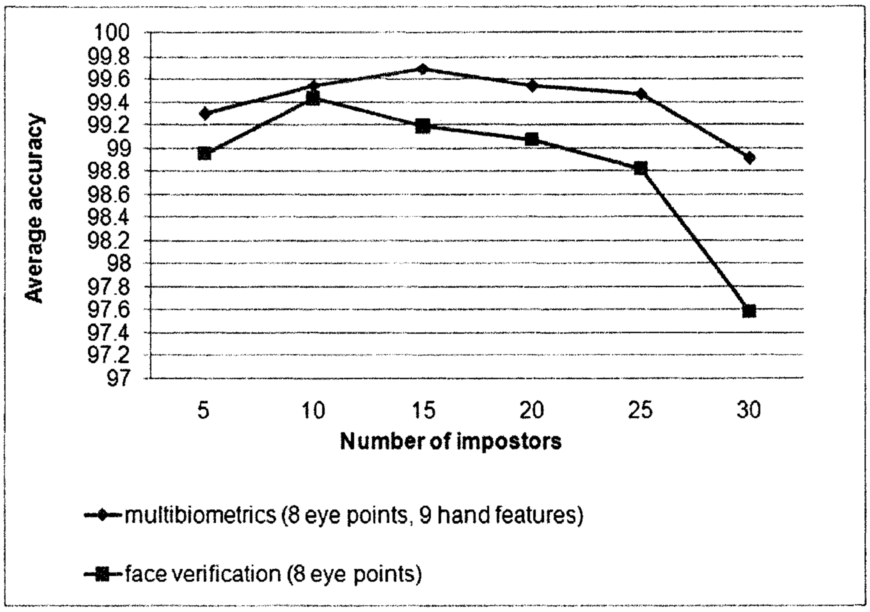


Figure 4.7: The comparison of the face verification method based on 8 eye points and multibiometrics based on 8 eye points and 9 hand features.

face verification based on 8 eye features, and the same facial model combined with hand geometry features. As we can see, this multibiometric method obtains better average results than unimodal biometric system that is based on 8 facial eye points. The average accuracy increased in all categories, but the best increase of 1.33% was obtained for 30 impostors. The smallest increase was achieved in the category where 10 impostors were involved in the testing phase and reached 0.12%.

Next multibiometric model, which is used in our experiments, contains 16 eye points and 9 hand geometry features. The minimum accuracy obtained for multibiometrics is 95.83% for 5 impostors; however, in every group the maximum accuracy of 100% was obtained. The highest average accuracy of 99.60% is reached for 15 impostors. After reaching the peak, the average accuracy then drops off to 98.26% for 30 negative individuals involved in the test. After comparing the multibiometric system based on 16 eye points and 9 hand features with the face verification system based

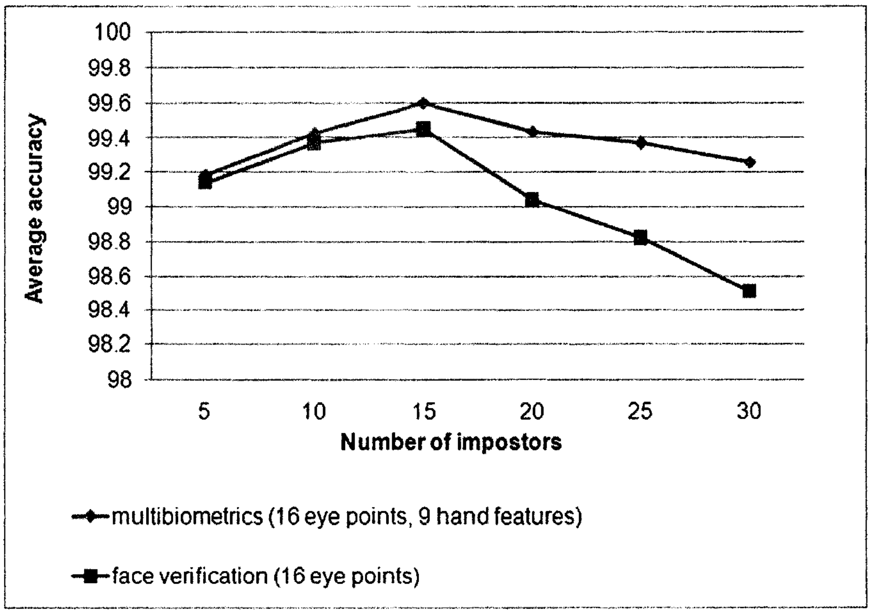


Figure 4.8: The comparison of the face verification method based on 16 eye points and multibiometrics based on 8 eye points and 9 hand features.

only on 16 eye points, we observe that again multibiometric system obtains better average accuracy. The graphical comparison between the multibiometric system and the unimodal facial system based on 16 eye points is presented in Figure 4.8. We observe that there are slight improvements of 0.05%, 0.06% and 0.15% in average accuracy for multibiometrics over face verification based on 16 eye points for 5, 10, and 15 impostors respectively. The average accuracies increase more for 20, 25, and 30 impostors by 0.4%, 0.55%, and 0.75% correspondingly for multibiometrics based on 16 eye points and hand features.

Another multibiometric model, which contains 8 eye points, 4 nose points and 9 hand geometry features, achieved its highest average accuracy of 99.78% for 15 impostors. The same growing tendency, as in previous multibiometrics models, is observed here. The average accuracy starts increasing at 99.53% for 5 impostors, 99.69% for 10 impostors, and finally reaches the peak of 99.78% for 15 impostors.

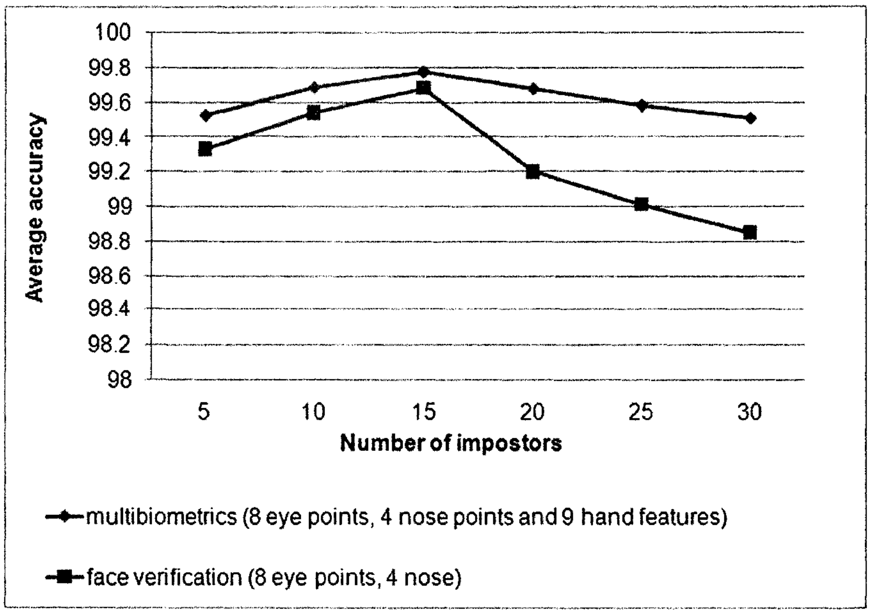


Figure 4.9: The comparison of the face verification method based on 8 eye points and 4 nose points and multibiometrics based on 8 eye points, 4 nose points and 9 hand features.

Afterwards, the average accuracy decreases to 99.68% for 20 impostors, 99.58% for 25 impostors and finally 99.51% for 30 impostors. The maximum accuracy of 100% is found in every group. This multibiometric system obtained better results than the face verification system based on 8 eye points, 4 nose points, as shown on Figure 4.9. The average accuracy increased in all categories, but the best increase of 0.66% was obtained for 30 impostors. The smallest increase was achieved in the category where 15 impostors were involved in the testing phase and reached 0.1%.

The next multibiometric model is based on 16 eye points, 4 nose points and 9 hand geometry features. As we can see, in this category the best average result, 99.64%, is obtained when one person is tested against 15 impostors. The minimum recognition rate achieved is 95.03%, and it is obtained when the tests are run against 5 impostors. The highest accuracy reaches 100% in every group. We notice that the average

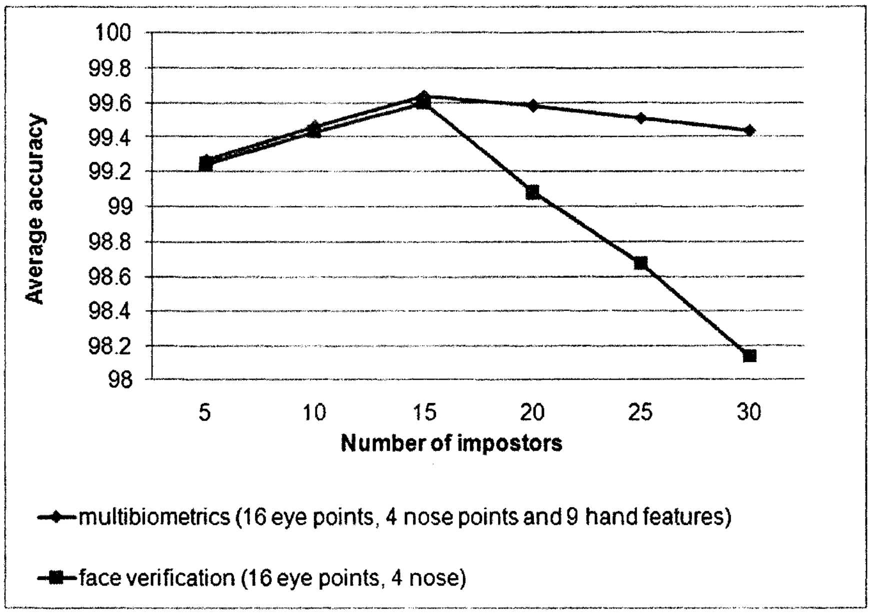


Figure 4.10: The comparison of the face verification method based on 16 eye points and 4 nose points and multibiometrics based on 16 eye points, 4 nose points and 9 hand features.

accuracy has its peak, when the system is tested against 15 negative individuals, and the average accuracy then decreases, obtaining 99.58%, 99.51%, 98.44% for 20, 25, 30 impostors respectively. When we compare the multibiometric system based on 16 eye points, 4 nose points and 9 hand geometry features with the face verification system based only on facial points that include 16 eye points, and 4 nose points, we observe that again the multibiometric system obtains better average accuracy. Figure 4.10 presents the comparison of the multibiometric system and the unimodal facial system based on 16 eye points and 4 nose points. We observe that there are slight improvements of 0.03%, 0.03% and 0.04% in average accuracies for multibiometrics over the face verification based on 16 eye points and 4 nose points for 5, 10, and 15 impostors respectively. The average accuracies increase more for 20, 25, and 30 impostors by 0.5%, 0.84%, and 1.3% respectively for multibiometrics based on 16 eye points, 4 nose

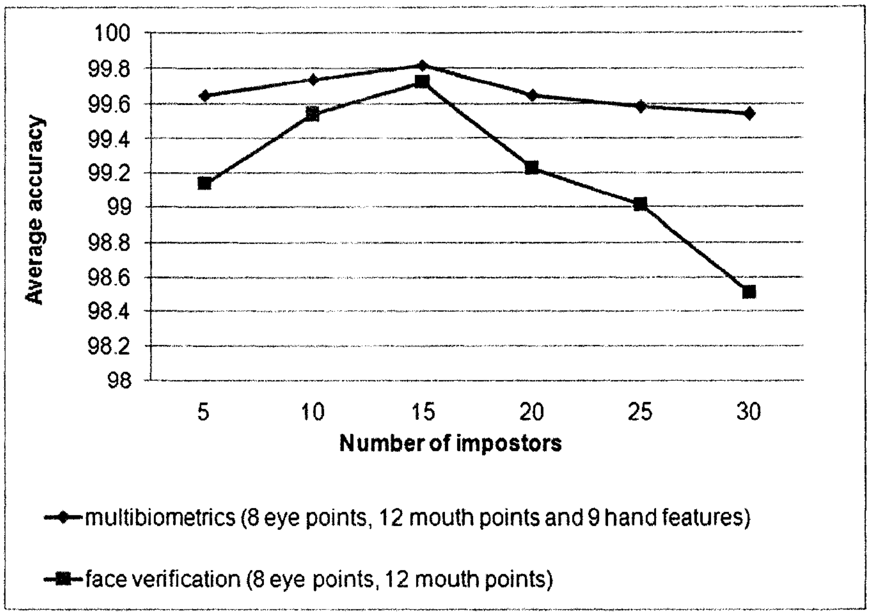


Figure 4.11: The comparison of the face verification method based on 8 eye points and 12 mouth points and multibiometrics based on 8 eye points, 12 mouth points and 9 hand features.

points and hand features.

Multibiometric model, which contains 8 eye points, and 12 mouth points obtained better average accuracy rates than the multibiometric models described previously. The highest average accuracy of 99.82% is reached for 15 impostors. After reaching the peak, the average accuracy then drops off to 99.54% for 30 negative individuals involved in the test. The minimum accuracy obtained is 95.83% for 5 impostors; however, in this facial model 100% is achieved in every group. Figure 4.11 shows the average accuracies for face verification based on 8 eye points and 12 mouth points, and the same facial model combined with hand geometry features. As we can see, this multibiometric method obtains better average results than unimodal biometric system. The average accuracy increased in all categories, but the best increase of 1.03% was obtained for 30 impostors. The increases varies from 0.1% for 15 impostors, 0.2% for 10 impostors, 0.42% for 20 impostors, 0.51% for 5 impostors, 0.57% for 25 impostors, and finally 1.03% for 30 impostors.

Another multibiometric system is based on 16 eye points, 12 mouth points and 9 hand geometry features. As shown in Table 4.5, this model obtains poorer average accuracy rates than the multibiometric model based on 8 eye points, 12 mouth points and 9 hand geometry features. The minimum accuracy obtained is 95.83% for 5 impostors; however, in this facial model the 100% is achieved in every group. The highest average accuracy of 99.60% is reached for 15 impostors. After reaching the peak, the average accuracy then drops off to 99.12% for 30 negative individuals involved in the test. Again, we observe that the multibiometric system obtains better results than the face verification system based on the same facial points as multibiometric system. As shown on Figure 4.12, the highest improvement of 0.8% is received for 25 impostors. The smallest increase in average accuracy of 0.02% is obtained for 5 impostors.

The multibiometric model, which contains 12 mouth points, 4 nose points and 9 hand geometry features, achieved its highest average accuracy of 99.67% for 15 impostors. The same growing tendency, as in previous multibiometrics models, is observed here. The average accuracy starts increasing at 99.26% for 5 impostors, 99.51% for 10 impostors, and finally reaches the peak of 99.67% for 15 impostors. Afterwards, the average accuracy decreases to 99.38% for 20 impostors, 99.26% for 25 impostors and finally 99.13% for 30 impostors. The maximum accuracy of 100% is found in every group.

Figure 4.13 shows the average accuracies for face verification based on 12 mouth

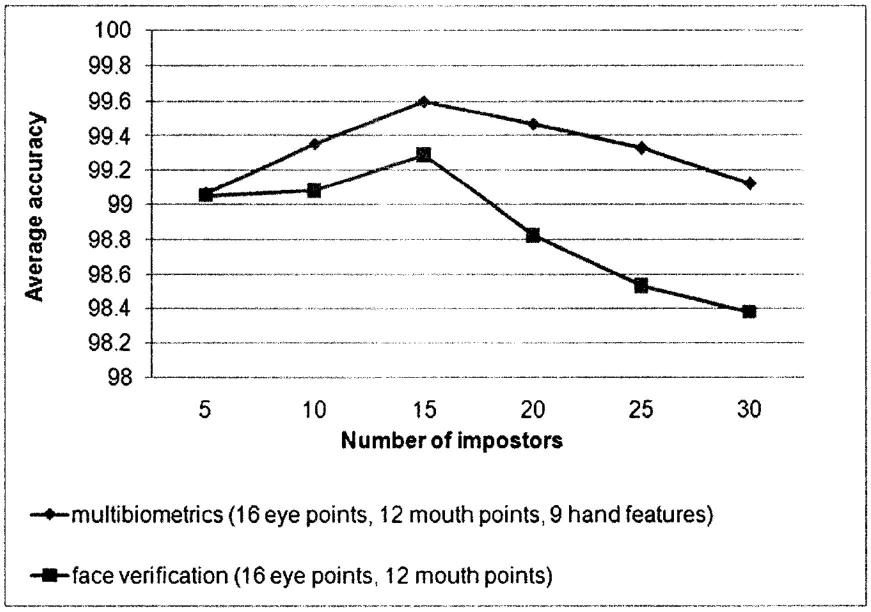


Figure 4.12: The comparison of the face verification method based on 16 eye points and 12 mouth points and multibiometrics based on 16 eye points, 12 mouth points and 9 hand features.

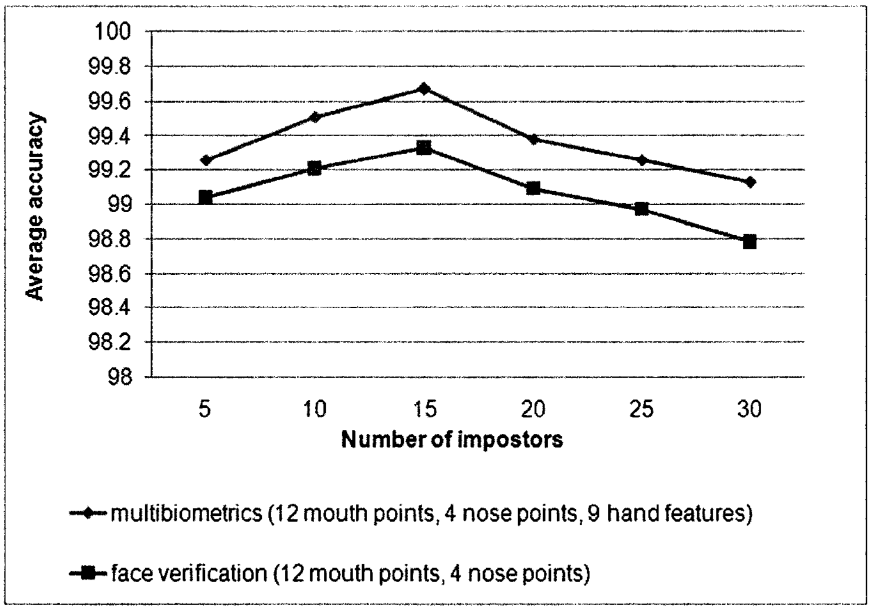


Figure 4.13: The comparison of face verification method based on 12 mouth points and 4 nose points and multibiometrics based on 12 mouth points, 4 nose points and 9 hand features.

points and 4 nose points, and the same facial model combined with hand geometry features. As we can see, this multibiometric method obtains better average results than the unimodal biometric system. The average accuracy is increased in all categories, but the best increase of 0.35% was obtained for 30 impostors. The smallest increase was achieved in the category where 5 impostors were involved in the testing phase and reached 0.22%.

For the last two models, we combine three main parts of the face together, such as the eyes, the nose, and the mouth, with hand geometry features. In this section, the multibiometric model based on 8 eye points, 12 mouth points, 4 nose points and 9 hand features is described. As in previous multibiometric models, the experiments were run against different numbers of impostors involved in the testing phase.

This multibiometric model, which contains 8 eye points, 4 nose points, 12 mouth points and 9 hand features, achieved the highest average accuracy of 99.76% for 15 impostors. The same growing tendency as in previous multibiometric models is observed here. The average accuracy increases starting at 99.42% for 5 impostors, 99.69% for 10 impostors, and finally reaches the peak of 99.76% for 15 impostors. Afterwards, the average accuracy decreases to 99.65% for 20 impostors, 99.54% for 25 impostors and finally 99.37% for 30 impostors. The maximum accuracy of 100% is found in every group. Again, the multibiometric system based on 8 eye points, 4 nose points, 12 mouth points, and 9 hand geometry features compared with the face verification system based only on facial points that include 8 eye points, 4 nose points, and 12 mouth points, we observe that the multibiometric system obtains better average accuracy. Figure 4.14 shows the comparison of the multibiometric

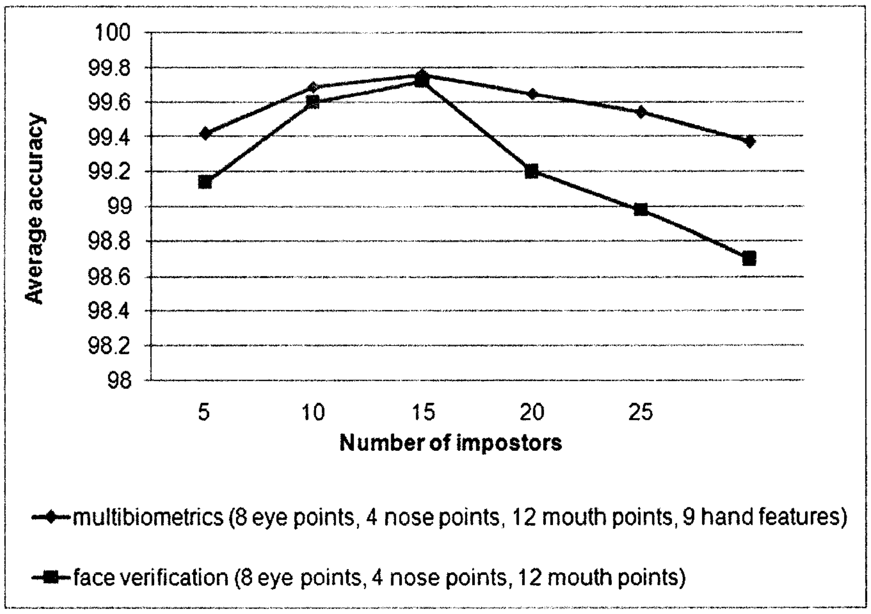


Figure 4.14: The comparison of face verification method based on 8 eye points, 4 nose points and 12 mouth points and multibiometrics based on 8 eye points, 4 nose points, 12 mouth points and 9 hand features.

system and the unimodal facial system based on 16 eye points, 4 nose points, and 12 mouth points. The highest increase in average accuracy of 0.67% is observed when 30 impostors are involved in the testing. The smallest increase is noticed for 15 impostors and it is 0.04%

The last multibiometric model is based on 16 eye points, 4 nose points, 12 mouth points, and 9 hand geometry features. This multibiometric model obtains poorer average accuracy rates than the model that contains the same parts of the face, but less numbers of points extracted from the eyes. The minimum accuracy obtained is 95.83% for 5 impostors; however, in this facial model the 100% is achieved in every group. The highest average accuracy of 99.52% is reached for 15 impostors. The average accuracies vary between 98.91% for 30 impostors to 99.52% for 15 impostors.

When we compare the multibiometric system based on 16 eye points, 4 nose

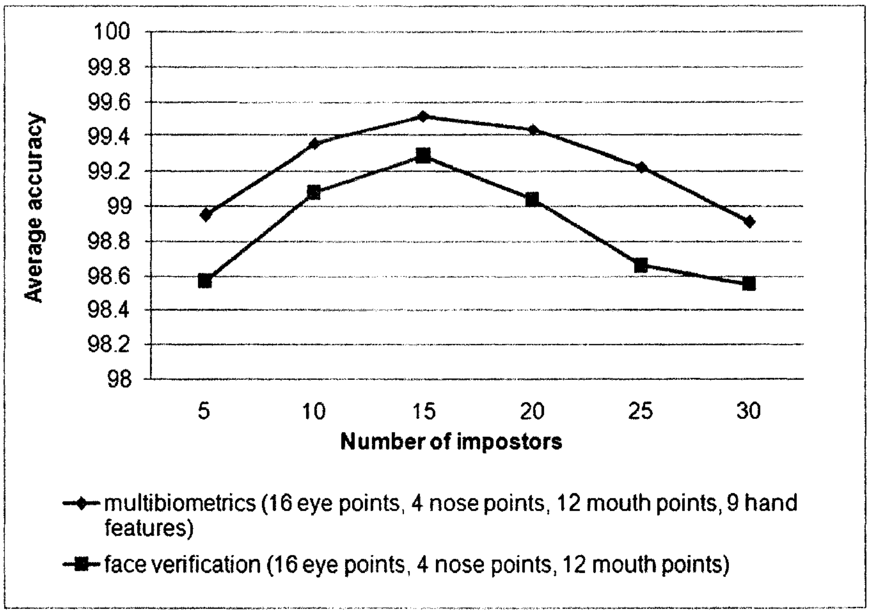


Figure 4.15: The comparison of face verification method based on 16 eye points, 4 nose points and 12 mouth points and multibiometrics based on 16 eye points, 4 nose points, 12 mouth points and 9 hand features.

points, 12 mouth points, and 9 hand geometry features, with the face verification system based only on facial points that include 16 eye points, 4 nose points, and 12 mouth points, we observe that again the multibiometric system obtains better average accuracy. Figure 4.15 shows the comparison of the multibiometric system and the unimodal facial system based on 16 eye points, 4 nose points, and 12 mouth points. The average accuracy increases by 0.38%, 0.28%, 0.23%, 0.4%, 0.57%, and 0.36% for 5, 10, 15, 20, 25, 30 impostors respectively.

The multibiometric system, which combines face features with hand geometry features, obtains better results than the face verification method alone. In some models, the increase of 1.3% for multibiometrics was achieved. The greatest improvements were obtained when more impostors were involved in the testing mode. The combination of face and hand images proved to be a good model, because the best average accuracy of 99.82% was achieved for the model that is based on 8 eye points, 12 mouth points and 9 hand geometry features. Also the system that contains the combination of 8 eye points, 4 nose points and 9 hand features is considered to be the second best, since the average accuracy obtained for this model reached 99.78%.

Figure 4.16 represents the graphical comparison of all 9 multibiometric authentication methods. As we see, all systems perform well. Most of them, except two models, obtain average accuracies more than 99% in all categories. The best average results are obtained when 15 impostors were involved in the testing phase. The highest average accuracy of 99.82% is achieved for the model that contains 8 eye points, 12 mouth points and 9 hand geometry features. From the experiments, we noticed that in general, the multibiometric models that contain the combination of 16 eye points performed worse than the ones that contain 8 eye points. It is clearly visible from Figure 4.16 that the multibiometric system obtained much better results than the unimodal system based on facial features, shown in Figure 4.6.

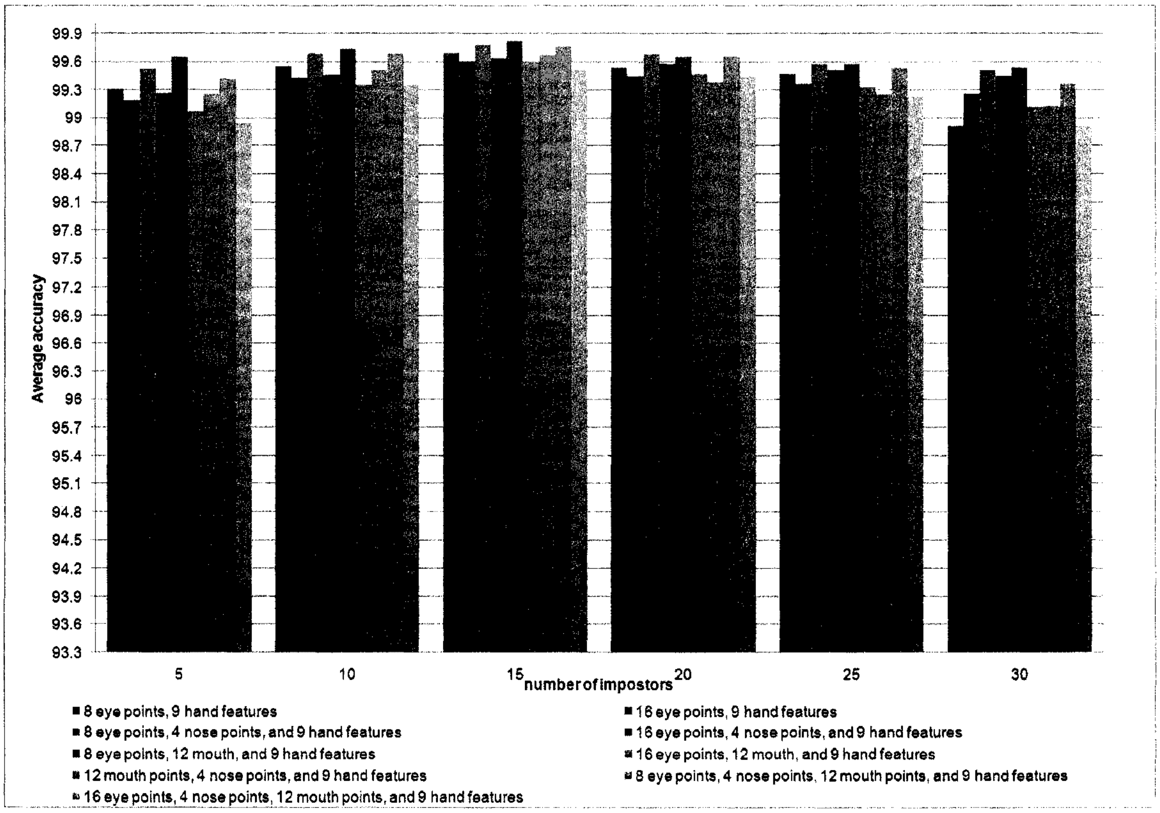


Figure 4.16: The comparison of multimodal biometrics including 9 different facial models and hand features.

Chapter 5

Conclusions

5.1 Summary

Biometrics refers to an automatic recognition of a person based on his/her behavioral and/or physiological characteristics. Many businesses have already applied biometric methods in practice, mostly identification or authentication purposes; however, more of work is left to improve the accuracy rates. Biometrics has been adopted in a variety of large scale identification applications, such as border control, criminal investigations and security.

One of the methods to improve the recognition rate is multimodal biometrics, which is based on more than one physiological or behavioral characteristics to identify an individual. Multimodal biometrics improves not only the performance, but also nonuniversality and spoofing.

As introduced earlier, a face recognition system consists of several components, such as face detection, tracking, alignment, feature extraction and matching. Face authentication is one of the primary biometric technologies, and it became more important in new technologies such as mobile devices, and increased demands on security. Hand recognition mostly based on hand geometry is also well-known and quite good for environments where medium security is required. In order to improve the accuracy, researchers combine different unibiometric methods together. In our approach, we combined at the feature level fusion face and hand authentication systems.

As shown in face authentication sections, the best average accuracy of 99.72% is achieved for two facial models. One model is based on 8 eye points and 12 mouth points, and the other one is based on 8 eye points, 4 nose points and 12 mouth points. The hand authentication method alone showed that it is a weaker verification system than the face verification method. As the number of impostors increases, the average accuracy for hand verification drops off to 85.13%. However, when we combine those two systems together, we obtain better accuracies. The best average accuracy of 99.82% is obtained for a multibiometric model that is based on 8 eye points, 12 mouth points and 9 hand geometry features. The multibiometric models improved accuracies in all groups of impostors. The improvement rates vary from 0.02% - 1.3% according to face verification system.

A user authentication system based on cell phone images is very important not only for security reasons but also convenience. If such a system could be built into the cell phone, people would no longer need to memorize the personal identification numbers. In our research, we focused on person authentication from cell phone images by extracting facial points using point distribution model and active shape models, and geometry hand features obtained from hand images. In order to construct the face feature vector, we convolved nine face models with Gabor filters. We analyzed which combination of facial models obtains the highest accuracy against how many impostors are involved in the testing phase. Hand authentication is focused on hand geometry features, like width and height of fingers and width of the palm. We wanted to build a multibiometric system based on face and hand biometrics. To the best of our knowledge it is the first system involving multibiometrics based on face and hand of cell phone pictures. The system is simple and fast. As the experiments show, the high accuracy of up to 99.82% has been achieved for multibiometrics.

5.2 Future Work

Right now, the system is performing well and achieves high accuracy rates. Development of a fully automatic system and incorporating it into a cell phone is left for future research. Localizing facial feature points and hand critical points are semi-automated since we have to manually label some of the images in order to build a model. Automatic detection of points of interest will allow to incorporate the system into a cell phone. Moreover, this is a challenging problem to design a chip with multibiometric system, which can be built into the phone.

Bibliography

[1] S. Abe. *Support Vector Machines for Pattern Classification.* Springer, 2005.

[2] M. Arif, T. Brouard, and N. Vincent. Personal identification and verification by hand recognition. In *Proceedings of the 2006 IEEE International Conference on Engineering of Intelligent Systems*, 2006.

[3] BBC News. Long Lashes Thwart ID Scan Trial. `http://news.bbc.co.uk/2/hi/uk_news/politics/3693375.stm`, 2004. Available online.

[4] P.N. Belhumer, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[5] E. S. Bigun, J. Bigun, and S. Fisher. Expert conciliation for multimodal person authentication systems using baysian statistics. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication*, pages 291–300, 1997.

[6] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. *Guide to Biometrics.* Springer-Verlag New York, 2004.

[7] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 12(10):955–966, 1995.

[8] R. Brunelli and T. Poggio. Face recognition: Features vs. templates. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 15(10):1042–1053, 1993.

[9] Y. Bulatov, S. Jambawalikar, P. Kumar, and S. Sethia. Hand recognition using geometric classifiers. In *International Conference on Bioinformatics and its Applications*, pages 753–759, 2004.

[10] J.P. Campbell. Speaker recognition: a tutorial. 85(9):1437–1462, 1997.

[11] CANPASS: Streamlines customs clearance for frequent travellers. `http://www.cbsa-asfc.gc.ca/travel/canpass/menu-e.html`, 2007. Available online.

[12] C. C. Chang and C. J. Lin. LIBSVM: a Library for Support Vector Machines. `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 2001. Available online.

[13] C. H. Chen and C. T. Chu. Fusion of face and iris features for multimodal biometrics. In *Advances in Biometrics: International Conference, ICB 2006*, pages 571–580, 2006.

[14] X. Chen, P. J. Flynn, and K. W. Bowyer. Ir and visible light face recognition. *Computer Vision and Image Understanding*, 99(3):332–358, 2005.

[15] V. Cherkassky and F. Mulier. *Learning from Data: Concepts, Theory, and Methods.* John Wiley and Sons, New York, 1998.

[16] C. C. Chibelushi, J. S. D. Mason, and F. Deravi. Feature-level data fusion for bimodal person recognition. In *Proceedings of the Sixth International Conference on Image Processing and Its Applications*, volume 1, pages 399–403, 1997.

[17] Combining Multiple Biometrics. `http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html`, 2000. Available online.

[18] T. F. Cootes and C. J. Taylor. *Statistical Models of Appearance for Computer Vision.* University of Manchester, 2004.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[20] N. Cristianini and J. Shawe Taylor, editors. *An Introduction to Support Vector Machines.* Cambridge University Press, 2000.

[21] J. Daugman. How iris recogntion works? *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[22] J. G. Daugman. Uncertainty relation for resolution in space, spatial - frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7):1160–1169, 1985.

[23] J. G. Daugman. Face and gesture recognition: Overview. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 19(7):675–676, 1997.

[24] C. Goodall. Procrutes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society Series B*, 53(2):285–339, 1991.

[25] G. Guo, S. Z. Li, and C. Kapluk. Face recognition by support vector machines. *Vision and Image Computing*, 19:631–638, 2001.

[26] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: Global versus component-based approach. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, pages 668–694, 2001.

[27] R. Hietmeyer. Biometric identification promises fast and secure processing of airline passengers. *The International Civil Aviation Organization Journal*, 55(9):10–11, 2000.

[28] L. Hong and A. K. Jain. Intergrating faces and fingerprints for personal identification. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 20:1295–1307, 1998.

[29] L. Hong, A. K. Jain, and S. Pankanti. Can multibiometrics improve performance? In *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, pages 59–64, 1999.

[30] Y.S. Huang and C. Y. Suen. Method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):90–94, 1995.

[31] A. Iannarelli. Ear identification. In *Forensic Identification Series.* Paramont Publishing Company, Fremont, California, 1989.

[32] A.K. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identification in a Networked Society.* New York Kluwer Academic Publishers, 2002.

[33] A.K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verfification system. In *2nd IEEE International Conference on Audio- and Video-based Biometric Person Authentication*, pages 166–171, 1999.

[34] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics*, 14(1):4–20, 2004.

[35] J. Jones and L. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, pages 1233–1258, 1987.

[36] K. Jonsson, J. Kittler, Y. Li, and J. Mattas. Support vector machines for face authentication. *Vision and Image Computing*, 20:269–275, 2002.

[37] T. Kanade. *Picture processing by Computer Complex and Recognition of Human Faces.* Ph.D. Thesis, Kyoto University, 1973.

[38] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 12(1):103–108, 1990.

[39] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[40] A. Kong, D. Zhang, and G. Lu. A study of identical twins' palmprints for personal verification. *Pattern Recognition*, 39:2149–2156, 2006.

[41] A. Kumar and D. Zhang. Personal authentication using multiple palmprint representation. *Pattern Recognition*, 38(10):1695–1704, 2005.

[42] L. I. Kuncheva. *Combining Pattern Classifiers - Methods and Algorithms.* Wiley, 2004.

[43] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Is independence good for combining classifiers? In *Proceedings of the International Conference on Pattern Recognition(ICPR)*, volume 2, pages 168–171, 2001.

[44] S. Y. Kung, M. W. Mak, and S. H. Lin. *Biometric Authentication: A Machine Learning Approach.* Prentice Hall, 2005.

[45] L. Lam and C. Y. Suen. Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 27(5):553–568, 1997.

[46] C.J. Liu and H. Wechesler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[47] C. Lopez-Ongil, R. Sanchez-Reillo, J. Liu-Jimenez, F. Casado, L. Snchez, and L. Entrena. FPGA implementation of biometric authentication system based on hand geometry. *Lecture Notes in Computer Science*, pages 43–53, 2004.

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

[49] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proceedings of Fourth ACM Conference on Computer and Communications Security*, pages 48–56, 1997.

[50] V.S. Nalwa. Automatic on-line signature verification. *Proceedings of the IEEE*, 85(2):215–239, 1997.

[51] NIST Report to the United States Congress. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability. `ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf`, 2002. Available online.

[52] C. Park, T. Choi, Y. Kim, S. Kim, J. Namkung, and J. Paik. Multi-modal human verification using face and speech. In *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*, pages 54–60, 2006.

[53] P.S. Penev and J.J. Atick. Local feature analysis: A general statistical theory for object representation. *Network: Comput. Neural Syst*, 7(3):471–500, 1996.

[54] S. Prabhakar, S. Pankanti, and A. K. Jain. Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy Magazine*, 1(2):33–42, 2003.

[55] Privium - fast border passage with iris scan. `http://www.schiphol.nl/privium/privium.jsp`, 2007. Available online.

[56] F.K. Prokoski. Disguise detection and identification using infrared imagery. In *Proceedings of SPIE, Optics, and Images in Law Enforcement II*, pages 27–31, 1982.

[57] J. Qin and Z. S. He. A SVM face recogntion method based on Gabor-featured key points. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, volume 8, pages 5144 – 5149, 2005.

[58] J. Rokita, A. Krzyżak, and C. Y. Suen. Cell phone personal authentication systems using multimodal biometrics. *Proceedings of International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science (to appear)*, 2008.

[59] J. Rokita, A. Krzyżak, and C. Y. Suen. Multimodal biometrics by face and hand images taken by a cell phone camera. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(3), 2008.

[60] A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24:2115–2125, 2003.

[61] A. Samal and P. A. Iyengar. Automatic recognition and analysis of human faces and facial expresions: a survey. *Pattern Recognition*, 25:65–77, 1992.

[62] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos. Biometric identification through hand geometry measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1168–1171, 2000.

[63] L. Shen and L. Bai. A review on gabor wavelets for face recognition. *Pattern Analysis and Application*, 9:273–292, 2006.

[64] B. Son and Y. Lee. Biometric authentication system using reduced joint feature vector of iris and face. In *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 513–522, 2005.

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

[66] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1995.

[67] V. N. Vapnik. *Statistical Learning Theory.* John Wiley and Sons, 1998.

[68] J. Wayman, A. Jain, D. Maltoni, and D. Maio. *Biometric Systems: technology, Design and Performance Evaluation.* Springer-Verlag London, 2005.

[69] L. Wiskott, J.M. Fellous, N. Kruger, and C. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[70] L. Xu, A. Krzyżak, and C. Y. Suen. Methods for combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435, 1992.

[71] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399 – 458, 2003.

Appendix A

Face and Hand Database

The two figures (A.1) - (A.2) present sample images from all individuals presented in our database. A sample of one face and one hand picture per each individual is shown.

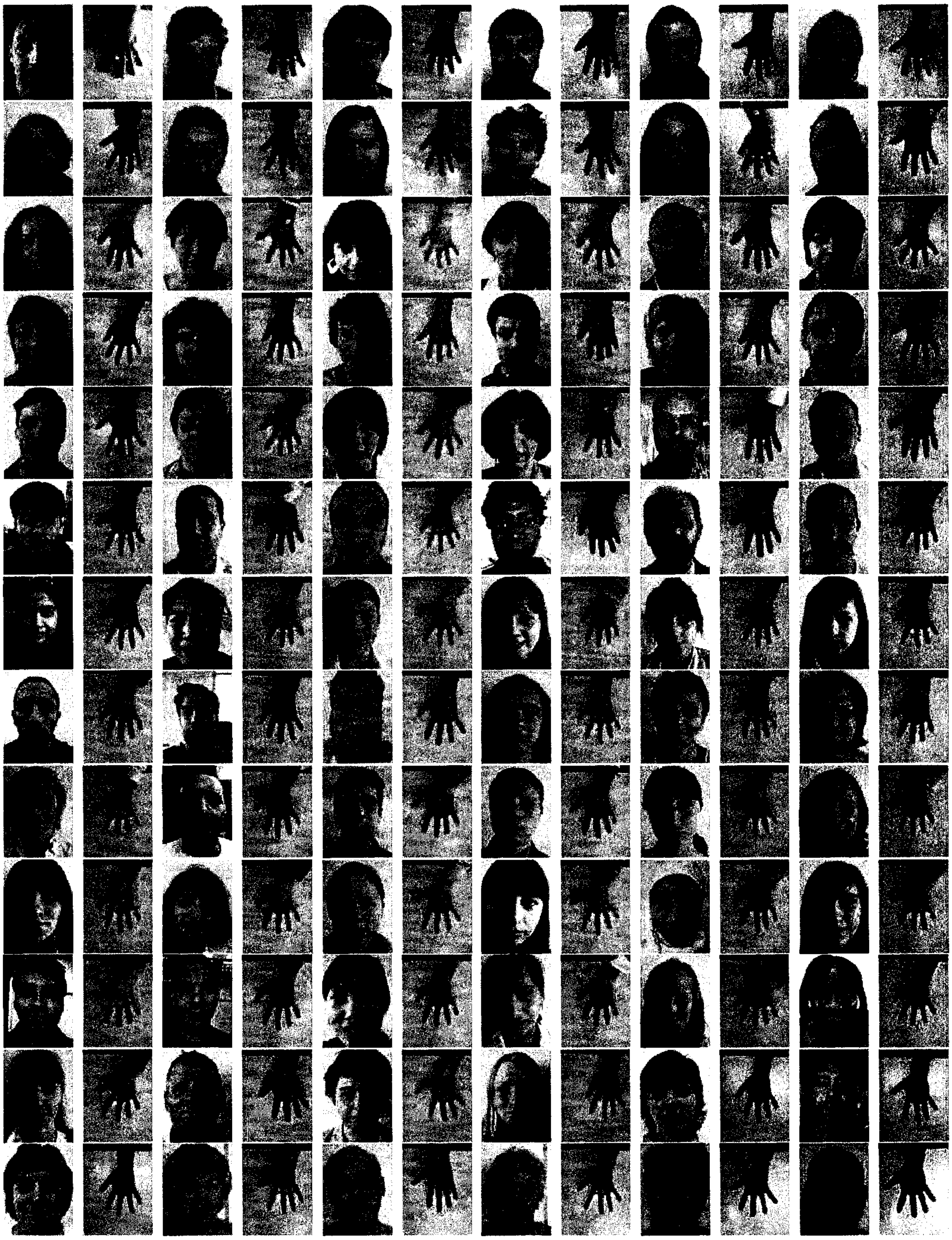


Figure A.1: Sample face and hand pictures from our database for all individuals involved in experiments.

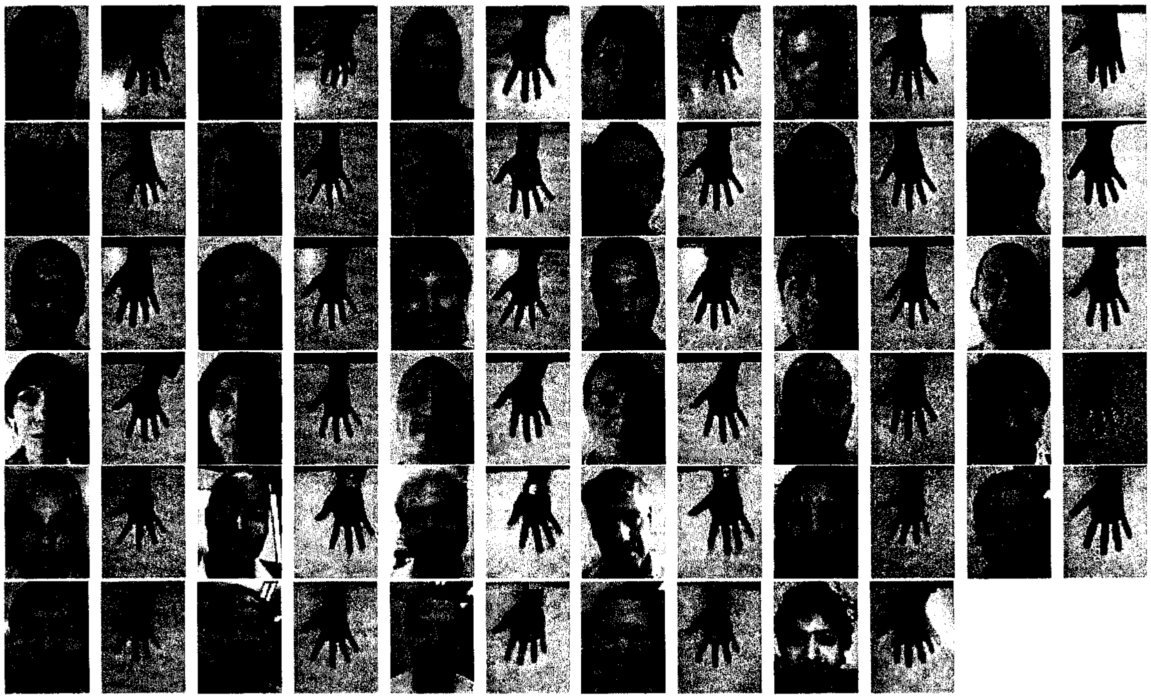


Figure A.2: Continuation of sample face and hand pictures from our database for all individuals involved in experiments.