An Automated Multimodal Face Recognition System

Based on Fusion of Face and Ear

Lorenzo Luciano

A Thesis
in
The Department
of
Computer Science and Software Engineering

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

February, 2009

Your file Votre référence
ISBN: 978-0-494-63339-7
Our file Notre référence
ISBN: 978-0-494-63339-7

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Lorenzo Luciano**

Entitled: **An Automated Multimodal Face Recognition System Based on Fusion of Face and Ear**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Hovhannes Harutyunyan	Chair
Dr. Peter Grogono	Examiner
Dr. Thomas Fevens	Examiner
Dr. Adam Krzyżak	Supervisor

Approved By ______________________
Chair of Department or Graduate Program Director

____
Dr. Robin A.L. Drew, Dean
Engineering and Computer Science

Date ______________________

ABSTRACT

An Automated Multimodal Face Recognition System Based on Fusion of Face and Ear

Lorenzo Luciano

This thesis presents an automated system for the detection and recognition of humans using a multimodal approach. Face recognition is a biometric method which has in recent years become more relevant and needed. With heavy research, it is achieving respectable recognition rates and is becoming more mature as a technology. It is even being deployed in certain situations such as with passports and credit cards.

Our multimodal biometric system uses both a person's face and ear to improve the recognition rate of individuals. By combining these two biometric systems we are able to achieve significantly improved recognition rates, as compared to using a unimodal biometric system.

The system is totally automated, with a trained detection system for face and one for ear. We look at recognition rates for both face and ear, and then at combined recognition rates, and see that we have significant performance gains from the multimodal approach. We also discuss many existing methods of combining biometric input and the recognition rates that each achieves.

Experimental results indicate that a multimodal biometric system has higher recognition rates than unimodal systems. This type of automated biometric recognition system can easily be used in installations requiring person identification such as person recognition in mugshots. It can also be used by security agencies and intelligence agencies requiring robust person identification systems.

ACKNOWLEDGEMENTS

I would like to thank all people who have helped and inspired me during my studies.

I want to thank my supervisor, Dr. Adam Krzyżak from the Department of Computer Science and Software Engineering at Concordia University, for his support throughout this thesis. His guidance, help, stimulating suggestions and encouragement helped me complete the research for, and writing of this thesis.

I would like to thank my family, on whose constant encouragement and love I have relied on throughout my studies.

I cannot end without a special thank you to my wife, Shelley, for always believing in me and encouraging me every step of the way.

To my wife Shelley, for her constant love, support, and encourgement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this master's thesis, we present a multimodal framework for the automatic detection and recognition of individuals using face and ear as input. The main motivation behind our research is towards the design of a fast, automatic detection and recognition system of humans, especially applicable to security services such as intelligence agencies, investigative services, commercial protection, terrorist tracking, human tracking, etc.

Given the prevalent criminal and terrorist activity in our society it would be beneficial to have a mechanism of human detection which is fast and reliable. We decided to use face and combine it with the ear because they are both usually exposed and do not require user participation.

1.1 Motivation and Objectives

The research and implementation lead us towards an automated system for the detection and recognition of humans using a multimodal approach. Our multimodal biometric system uses both a person's face and ear to improve the recognition rate of individuals. By combining these two biometric systems we are able to achieve significantly improved recognition rates, as compared to using a unimodal biometric system. The system is totally automated, with a trained detection system for face and one for ear.

We look at recognition rates for both face and ear, and then at combined recognition rates, and see that we have significant performance gains from the multimodal approach. We also discuss many existing methods of combining biometric input and the recognition rates that each achieves.

Experimental results indicate that a multimodal biometric system has higher recognition rates than unimodal systems. This type of automated biometric recognition system can easily be used in installations requiring person identification such as person recognition in mugshots. It can also be used by security agencies and intelligence agencies requiring robust person identification systems.

1.2 Overview of Biometrics

Biometrics is a science dedicated to the identification and classification of individuals, based on some physical or behavioural attributes of that individual. Recognition is a task easily accomplished by humans, but is an enormously complex task for the computer.

Falsification of individuals is a problem in today's digital world, passwords and identity cards can only go so far in protecting us and assuring us of the protection of our identity[36]. Imagine a system where falsification was impossible or nearly impossible, that is what we can someday hope biometrics will achieve. The necessity and requirement of a reliable identification system has put demand on biometric researchers to develop something useful and practical, in recent years.

There are many different human traits that are being used in biometric systems, any human trait that can be used to identify and recognize an individual with a certain degree of accuracy is being researched and used in an attempt to accurately recognize individuals, see Table 1.1 for some of these biometric traits.

One of the oldest biometric techniques is fingerprint recognition, still heavily used

Biometric Trait
Character
Ear
Face
Fingerprint
Gait
Hand Geometry
Iris
Signature
Vein Pattern
Voice

Table 1.1: Examples of some biometric traits being used in biometric systems.

today by almost all law enforcement agencies, although with newer modern techniques. Iris has also proved to be a very reliable biometric, although like fingerprint requires the full cooperation of the user. Iris technology is now being used at many airports world wide for frequent flyers instead of passport identification. Some of the airports using this technology are in Canada, the United Kingdom and the United States[14]. See Figure 1.1 for some biometric systems that are being developed using various biometric traits and deployed in many different areas.

Unlike these techniques, face recognition does not require full participation of the user, although it has not reached the recognition levels that fingerprint and iris have achieved thus far. Face recognition is still a very difficult problem, due to variabilitys in the human face, plus other variabilitys such as illumination, ageing, eyeglasses, camera location and distance etc.[1]. The fact that it is a non-intrusive biometric method however, makes it very interesting and practical as a biometric.

Ear as a biometric is not as well known or recognized as many others are such as face, iris and fingerprint. Ears have however, been used as a forensic science[19] for a long time. The advantage of ear over other biometrics is that they have a rich and stable structure which does not change with time and age. It also does not change with a person's facial expressions and can be captured at a distance unlike fingerprint and iris, which was very important in our research.

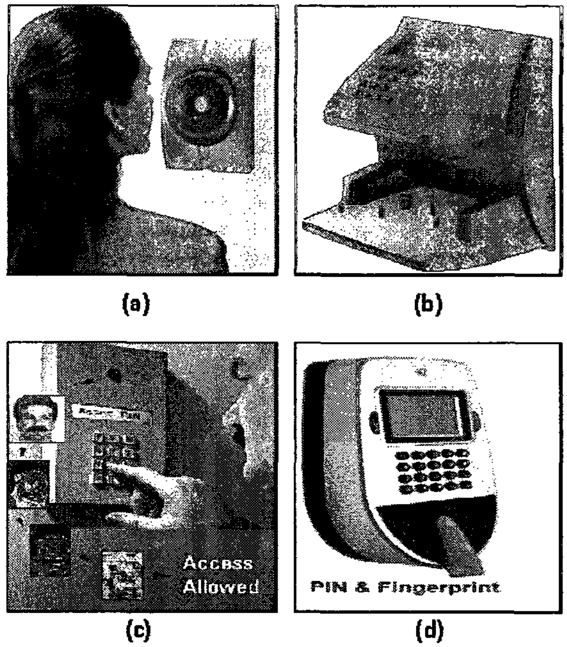


Figure 1.1: Biometric systems are being developed and deployed in many different areas, here are a few examples of some biometric devices using various biometric inputs. (a) Iris (b) Hand (c) Face (d) Fingerprint and Pin

Non-intrusive biometric methods have gained much interest in recent years due to the increased concern for public security. Since there is a tremendous increase in demand for useful and practical biometric systems, such a biometric system would be of high practical use and demand in both public and private sectors.

1.3 Approaches and Contribution

Human recognition using biometric means is a highly researched area, the need for such a system is incredibly high and the demand for one just as high. We have seen many advances in biometric uses such as fingerprint, iris, face, voice, gait, hand, ear etc.[1].

What makes face recognition an interesting biometric is that it is intuitive and does not require user participation, in other words it can be implemented without the user being aware that there is detection and recognition happening. Due to these factors, face recognition becomes a very interesting biometric for many applications in real world situations, such as intelligence agencies, investigative organizations, surveillance, tracking and apprehension of criminals, against terrorism etc.

There has been a tremendous amount of progress and research in recent years in the area of face recognition, especially in controlled environments. However, face recognition remains a very difficult and challenging problem for researchers. This is mainly due to the variation in the human face under various conditions such as illumination, expressions, view etc. These various conditions make it difficult to accurately match faces in real situations.

The human ear as a biometric, compared to the face has not received as much attention, although in recent years interest has increased due to its enormous potential. Ears have a tremendous potential as a biometric because of a number of factors. It is relatively stable over a persons lifetime, it does not change with a person's expression, it is a small and very rich area and is usually exposed[5, 40]. Along with the last issue stated, the ear is large enough to be captured at a distance, therefore does not require user participation.

These factors led us to our research in combining face and ear biometric to possibly increase the recognition rate that they could achieve individually.

1.4 Research Contribution

The detection, recognition, and authentication of individuals without their full co-operation would be a valuable tool for security and intelligence agencies requiring a robust person identification system. Biometric systems that do not have full person cooperation are still not as robust as other systems requiring complete person cooperation such as fingerprint technology and iris technology, now widely used at airports and other installations.

However, a robust system not requiring full person cooperation such as face recognition using video or images would facilitate the identification of individuals and would allow person identification from reasonable distances without the subjects knowledge. Such an unobtrusive, robust biometric system would have a great demand and implication for law enforcement agencies and other commercial installations requiring person identification.

We believe a system such as this would also have to be fully automated with the detection and recognition done without manual intervention, to be of the greatest possible value and use. Towards such a system, we have decided to combine face recognition with ear recognition in a multibiometric system to perhaps achieve a more robust recognition rate.

There are many methods for face recognition as this is a heavily researched area. Some of these more popular methods are Eigenface [43], Gabor features [38], Fisherface [4] and Local Feature Analysis [33]. Eigenfaces are a fast technique, but do have some issues with lighting and scale.

Due to the fact that we wanted a fast automated system, we used eigenfaces with eigenears to see if we could improve the recognition rate by using multibiometric, yet still maintain a fast, uncooperative fully automated system. Many other biometrics modalities such as iris, hand, gait, voice, fingerprint are given in Handbook of Biometrics [1].

We decided to use ear in combination with face as it still enables us to have an uncooperative biometric system. In most instances, the face and ear are both exposed and we are able to detect and use these in a person recognition and identification system.

The ear as a biometric seems less intuitive than the face, it is not what humans normally use to identify each other. Ear does however, make for a very interesting biometric and was first used as a biometric by Iannarelli[21]. He attempted to use ear as a biometric in a manual system by identifying important points on the ear and then using measurements to see if it could be used to identify individuals uniquely. In the end he came up with twelve features and could uniquely identify 10,000 individuals.

Ear has some very interesting properties lending itself to be a very good biometric candidate[37]. Given these properties, it seems like a very good complementary biometric to use in combination with the face.

Many of the methods we mentioned for face recognition can also be used for ear recognition. In addition, there are some other methods which were developed specifically for ear recognition. There are also geometric approaches to ear recognition such as those described by Coras[11], who in his studies uses geometric properties such as width and length of the ear to create a feature vector.

Another method to ear recognition is the Gaussian approach described by Hurley et al.[20]. In this approach the ear is modeled using a Gaussian force field. The pixels of the ear image create a magnetic like force field as they exert forces against one another. These force field lines which are created by the magnetic like force on the pixels generate channels which are then used for identification.

There are researchers also studying the validity of using 3D ear shape for recognition[46, 24] and also in multibiometrics[47]

The method we used for ear recognition is eigenears, which like eigenfaces, uses the method of PCA for comparing ears for recognition.

The difficulty in such as system is occlusions, which could be a limitation for both face and ear, also lighting, has shown to cause difficulties. Combining both biometrics, we hope to overcome some of these difficulties and allow for a system with greater robustness.

Some of the research and experimental results in the area of face and ear multibiometrics have shown to be very promising.

Chang et al.[23] used PCA on face and ear using multi-instance biometric, using a manual land marking method. With the largest dataset of 111 subjects, they were able to achieve a combined recognition rate of 90%. Unimodal recognition rates for face and ear were 70.5% and 71.6%, respectively, for one experiment and 64.9% and 68.5%, respectively for another.

Rahman and Ishikawa[39] also used PCA for combining face and ear, moreover they used profile images and manually extracted features. On a dataset of 18 subjects of profile face and ear, the recognition rate was 94.44%. In these experiments, the authors used the profile image to capture face(profile) and ear.

Middendorff and Bowyer[29] used PCA/ICP for face/ear, manually annotating feature landmarks. On a 411 subject dataset they were able to achieve a best fusion rate of 97.8%. Face had a recognition rate of 88.1% and ear had a recognition rate of 62.2%

It must be mentioned that none of these systems is an automated system, requiring manually intervention in all cases to either manually landmark or annotate features.

The research contribution of this thesis is the development of a multibiometric system using face and ear as biometrics, which is fully automated [26]. It requires no manual intervention at any point and was able to achieve a higher recognition rate(with a best rate of 100.0%) [27] on two separate experiments using two different databases(100 individual subset of FERET[35] and 114 person CVL database[32]), which is higher than previously mentioned research papers[23, 39, 29].

The automation includes a trained face and ear detector, extraction, cropping, and pre-processing. An automated system such as this would find immediate applications in many areas where identification and authentication are crucial.

We will demonstrate and describe how fusion of face and ear using an optimized weighted scheme can significantly improve recognition levels.

Our proposal, presents a multimodal framework for the detection and recognition of humans. The presented multimodal approach is able to achieve significantly higher recognition rates as we will demonstrate in this thesis.

1.5 Organization of Thesis

The remainder of this thesis is organized as follows; In Chapter 2, we look at automatic object detection, namely for face and ear. In Chapter 3, we illustrate face and ear biometrics as unimodal biometrics. In Chapter 4, we explore face and ear recognition as a multimodal biometric, combining face and ear for recognition. In Chapter 5, we review the experiments performed and the data achieved, measures, databases and tests used along with results. In Chapter 6, review and some concluding remarks.

Chapter 2

Object Detection

2.1 Introduction

The first step in our automated multimodal biometric system was the detection of the regions of interest. We are interested in extracting the objects which will form the basis of the recognition system; in this case the regions of interest are the objects, face and ear. The regions of interest are extracted using a Haar like features based object detector provided by the open source project OpenCV library[31]. This form of detection system is based on the detection of features that display information about a certain object class to be detected.

Haar like features encode the oriented regions in images whenever they are found, they are calculated similarly to the coefficients in Haar wavelet transformations. These Haar like features can be used to detect objects in images, in this case the human face and the human ear. This Haar like object detector was originally proposed by Viola and Jones[44] and later extended by Lienhart and Maydt[25].

A cascade of boosted classifiers using Haar like features is trained using positive (containing object to detect) images and negative (arbitrary, not containing object to detect) images. Cascade implies that the resultant classifier consists of many simple stages applied subsequently. It allows regions of non-interest or background regions to be discarded at every stage, so that computing time is not wasted in these areas of non-interest[44]. Boosted implies the classifiers at every stage are built using boosting

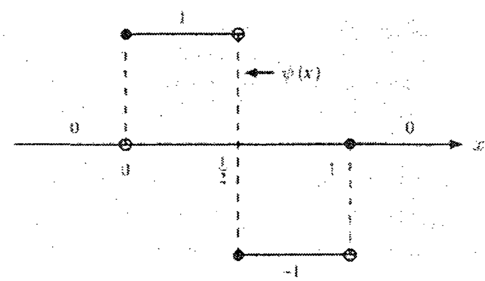


Figure 2.1: Haar Wavelet

techniques such as AdaBoost.

2.2 Haar Wavelets

To understand our detection system it is crucial to have an understanding of Haar wavelets, which form the basis of the object detector. The Haar wavelet was developed by Alfred Haar in 1909[16], it is the first wavelet ever discovered. Haar first used it to describe an orthonormal system, however the study of wavelets and the term wavelets only came to be much later.

The Haar wavelet is a simple wavelet which is not continuous and not differentiable. You can see a visual representation of a Haar wavelet in Figure 2.1.

The Haar wavelet function $\psi(t)$ can be described as,

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & otherwise \end{cases}$$

Its scaling function $\phi(t)$ can be described as,

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & otherwise \end{cases}$$

2.2.1 Haar Wavelet Properties

The Haar wavelet has a number of properties which we will describe below[12].

1) All functions can be approximated linearly by,

$$\phi(t), \phi(2t), \phi(4t), ...\phi(2^k t).$$

their shifted functions.

And also by,

$$\psi(t), \psi(2t), \psi(4t), ...\psi(2^k t).$$

their shifted functions.

2) In orthogonal representation,

$$\int_{-\infty}^{\infty} 2^m \psi(2^m t - n)\psi(2^{m_1} t - n_1)dt = \delta_{m,m_1}, S_{n,n_1}.$$

where $\delta_{i,j}$, represents the Kroneker delta.

3) Functions which represent scaling and have different scale m have a functional relationship as described,

$$\phi(t) = \phi(2t) + \phi(2t - 1).$$

$$\psi(t) = \phi(2t) - \phi(2t - 1).$$

4) We can calculate the coefficients of scale m, by using the coefficients of scale $m + 1$, if

$$X_w(n, m) = 2^{\frac{m}{2}} \int_{-\infty}^{\infty} x(t)\phi(2^m t - n)dt.$$

and,

$$X_w(n, m) = 2^{\frac{m}{2}} \int_{-\infty}^{\infty} x(t)\psi(2^m t - n)dt.$$

then,

$$X_w(n,m) = \sqrt{1/2}(X_w(2n, m+1) + X_w(2n+1, m+1)).$$

and,

$$X_w(n,m) = \sqrt{1/2}(X_w(2n, m+1) - X_w(2n+1, m+1)).$$

2.2.2 Haar Matrix

We can describe the Haar matrix associated with the Haar wavelet as follows,

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

using a discrete wavelet transformation, we can transform a sequence,

$$(a_0, a_1, ..., a_{2n}, a_{2n+1}).$$

into a sequence of two vectors,

$$((a_0, a_1), ..., (a_{2n}, a_{2n+1})).$$

If we then multiply each vector by the matrix H_2 we get,

$$((s_0, d_0), ..., (s_n, d_n)).$$

which is the first stage of the last Haar wavelet transform.

With a sequence of length four, combining two stages the fast Haar wavelet transform we would get,

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

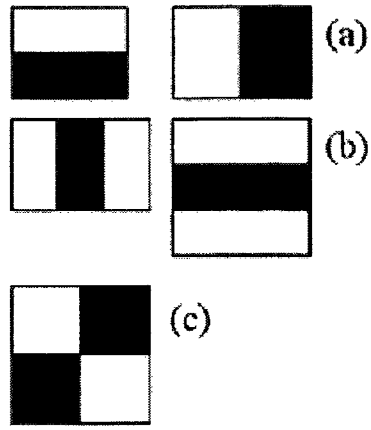


Figure 2.2: In this figure we see examples of Haar like rectangular features, a) two rectangular features, b) three rectangular features, c) four rectangular features

2.3 Haar like Features

The object detector uses features as opposed to pixels, features make it easier to encode ad-hoc domain knowledge which would be difficult to learn using a large quantity of training data[44]. The second reason for using features is the feature based system is a lot faster than a pixel based one.

The system uses three types of features, see Figure 2.2 for a visual representation. A two rectangle feature, see Figure 2.2(a), has two rectangle regions. The two regions are the same size and shape. There is also a three rectangle feature, see Figure 2.2(b), and a four rectangle feature, see Figure 2.2(c). Given a detector with a width and height of 24x24, the set of features would be very large with over 180,000 [44].

2.3.1 An Integral Image

With the large set of features and heavy computation involved, Viola and Jones[44] propose an intermediate representation for an image called an integral image.

The integral image $ii(x, y)$ is given by,

$$\sum_{x' \leq x, y' \leq x} i(x', y')$$

where, the location of the integral image is (x, y), and the original image is $i(x, y)$. Using the following pair of recurrences,

$$s(x, y) = s(x, y-1) + i(x, y)$$

$$ii(x, y) = ii(x-1, y) + s(x, y)$$

where, $s(x, y)$ is the row sum $s(x, -1) = 0$ and $ii(-1, y) = 0$.

Rectangular features provide an image representation which is rich and easily trainable. Together with the integral image concept, the rectangular features become very efficient.

2.4 AdaBoost Learning

AdaBoost learning is a method to boost the performance of a weak learning algorithm. In their research, Freund and Schapire[15] proved that AdaBoost is capable of increasing the performance of a simple classifier. They showed that the training error rate of a strong classifier is capable of reaching zero exponentially with the number of rounds used, in later experiments the results proved this result[42]. AdaBoost is capable of achieving greater performance because it produces large margins rapidly.

To compute every feature for every sub-window is very computationally expensive, therefore we need a method to make this more efficient. Without using every feature possible in a sub-window we are still capable of producing an effective classifier[44]. This question is which features do we use?

To accomplish this the simple learning classifier is designed to select the rectangular feature which best separates positive and negative examples. Using this premise, the simple classifier determines the threshold function where the least amount of examples are misclassified.

Therefore, a weak classifier $h_j(x)$, with feature f_j, threshold θ_j and a parity p_j will equal;

$$\begin{cases} 1 & if p_j f_j(x) < p_j \theta_j \\ 0 & otherwise \end{cases}$$

where x is a 24×24 pixel sub-window of an image.

In practice, it has been shown that a single feature cannot perform the classification with a low error rate.

In early stages of the boosting process the error rate achieved are between 0.2 and 0.3[44], later stages the rates achieved were between 0.4 and 0.5.

2.4.1 AdaBoost Algorithm

In this section we describe the AdaBoost algorithm for classifier learning. This algorithm is capable of selecting one feature in each boosting stage from the potential 180,000 features[44].

With a set of images $(x_1, y_1), ..., (x_n, y_n)$ where $y_i = 0$ for negative examples and $y_i = 1$ for positive examples.

If we initialize weights $w_{1,i}$ as,

$$i = \frac{1}{2m}, \frac{1}{2l}$$

for y_i, where m is the number of negatives and l is the number of positives.

Now, for $t = 1, ..., T$ we,

1)

Normalize the weights $w_{t,i}$ as,

$$\frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$

giving, w_t as a probability distribution.

2)

For every feature j, we train a classifier h_j, which can only use one feature, with the error rate ϵ_j with respect to w_t given as,

$$\sum_i w_i |j_j(x_i) - y_i|$$

we choose the h_t with the lowest ϵ_t

3)

next, we update the weights for the next round,

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where, $e_i = 0$ if x_i is correct, otherwise $e_i = 1$, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

In the end, we have a strong classifier given as,

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$.

2.5 Cascade Classifiers

A cascade of classifiers is capable of achieving increased performance and decreased computational time. It is capable of achieving this by rejecting most of the negative sub-windows while correctly detecting the positive sub-windows, therefore making it much smaller and more efficient.

The name cascade comes from the fact that the process is sort of like a decision tree, where at each level of the process we make decisions depending on whether we have a negative or positive result. If we have a positive result, we proceed to a second classifier, if we get a positive result from the second classifier we proceed to a third classifier, we continue with the process in this manner. A negative result will cause

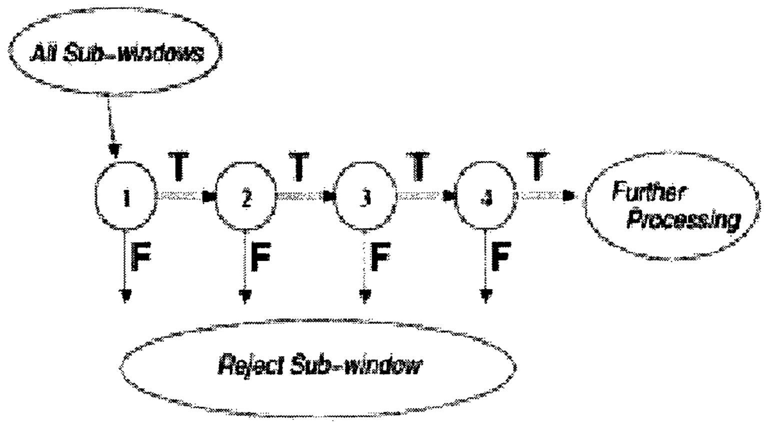


Figure 2.3: Visual depiction of the Detection Cascade.

the sub-window to be rejected, see figure 2.3 for a visual of the cascade process. The stages in the cascade use AdaBoost to construct the training classifiers.

The theory behind the cascade is that many of the sub-windows which are generated from an image are negative. Therefore, by rejecting these negative sub-windows early on, we can improve performance significantly. While at the same time, if we have a positive result, we evaluate every classifier in that cascade.

In the following stages, the classifiers uses for its training only the examples which remain. Therefore, the task of the classifier gets harder as we go on into the stages, as compared to the first stages. It is very much like a decision tree as we narrow in on the solution. In Figure 2.4 you can see a complete visual representation of the AdaBoost Learning algorithm.

2.6 Face Detection

In this section, we will describe and look at how we went about developing and training our face detector. The process was tedious and long, due to the amount of images that were required to get a decent detector and the computing power also required for the training of the detector throughout all of its stages.

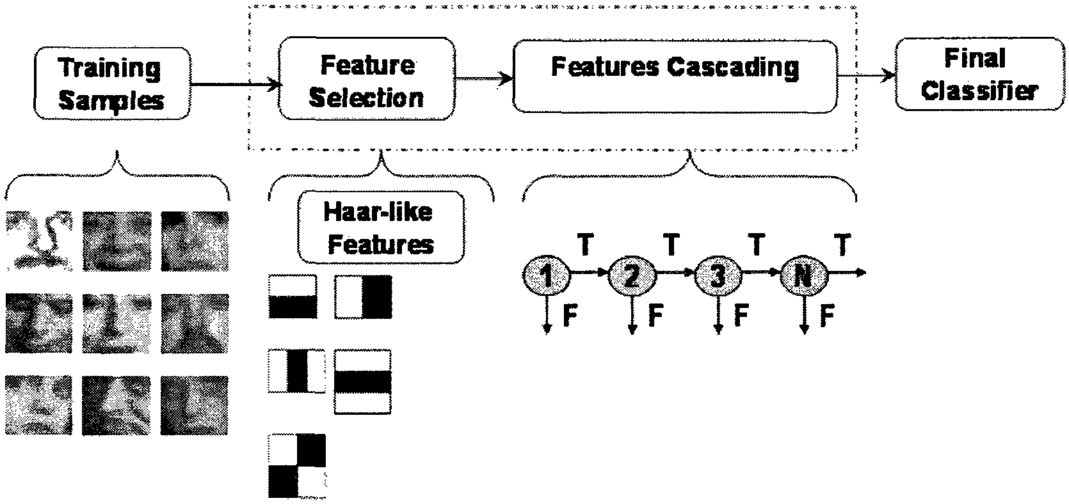


Figure 2.4: Visual depiction of the AdaBoost Learning Algorithm.

2.6.1 Data Set

For our face detection, the positive data set was built with 2000 face images. These face images were made using an image clipping tool which allowed us to select an area from within the original image. We used the FERET database for the images as they provided many profile images from which we could use to crop the ear. Many of the images in the positive data set were also created with a tool which allowed us to use a face image and create many more positive sets from it, by for example rotating it a little. You can see some examples of the face images from Figure 2.5.

The negative data set was made from about 5000 images acquired from various sources found on the internet. The only requirement was that they did not contain any faces. Some of the images included were of animals, landscapes, people(where the face was not visible) trains, automobiles etc. You can see some examples of these negative images in Figure 2.6.

Figure 2.5: Face Samples. Some samples of face images clipped from an image of a person.

2.6.2 Preprocessing

The positive face images we eventually gathered after clipping them from the original image, came in many different sizes and intensities. These positive samples were scaled to the same size of 24x24; this seems to have yielded the best and fastest results. Once we had these positive input images scaled and normalized to a size of 24x24, we proceeded with the training. Due to the heavy processing involved at each stage of the training process, it took about a week to run the complete training of the detector on a Duo CPU 2.66GHz PC with 2 GB of RAM.

2.6.3 Performance

The performance of the AdaBoost Haar face classifier was tested against a set of 200 frontal images. It detected all faces, but we did have six falsely detected faces. To remedy this, we always chose the largest detected face in the image, with the belief that the largest detected face would actually be the face since the image was of a person, from about shoulder height and above, see Figure 2.7.

This eliminated the problem of false detections and all faces were correctly detected. At this point, our detector worked well, so we were able to proceed to the next level of our multimodal research.

2.7 Conclusions

The face detector developed was fast and accurate for the purpose intended. The face detector was automatic, fast and worked great for our non-intrusive multi modal biometric system. It was capable of detecting all faces, with the falsely detected faces being eliminated by selecting the largest detected face in an image.

Figure 2.6: Negative Samples. Some samples of images used in the negative data set, the images contain no faces or ears.

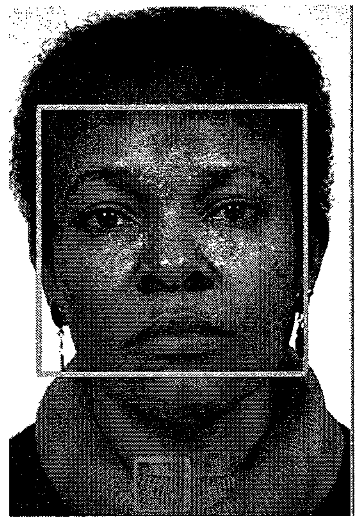


Figure 2.7: This shows an image where there was a falsely detected face.

2.8 Ear Detection

In this section, we look at and discuss the process we undertook to create our ear detector, In the end, we did have an ear detector which worked well. The process, like for the face detector, was long and tedious due to the amount of images required and the computing power it took to run through all of its stages.

2.8.1 Data Set

For our ear detector, the positive data set was built with 2000 ear images. These ear images were made using an image clipping tool which allowed us to select an area from within the original image. We used the feret database for the positive images as they provided many profile images from which we could use to crop the ear. Many of the images in the positive data set were also created with a tool which allowed us to use an ear image and create many more positive sets from it, by for example rotating it a little. You can see some examples of the ear images from Figure 2.8.

The negative data set was made from about 5000 images acquired from various sources found on the internet. The only requirement was that they did not contain any ears. Some of the images included were of animals, landscapes, people (where

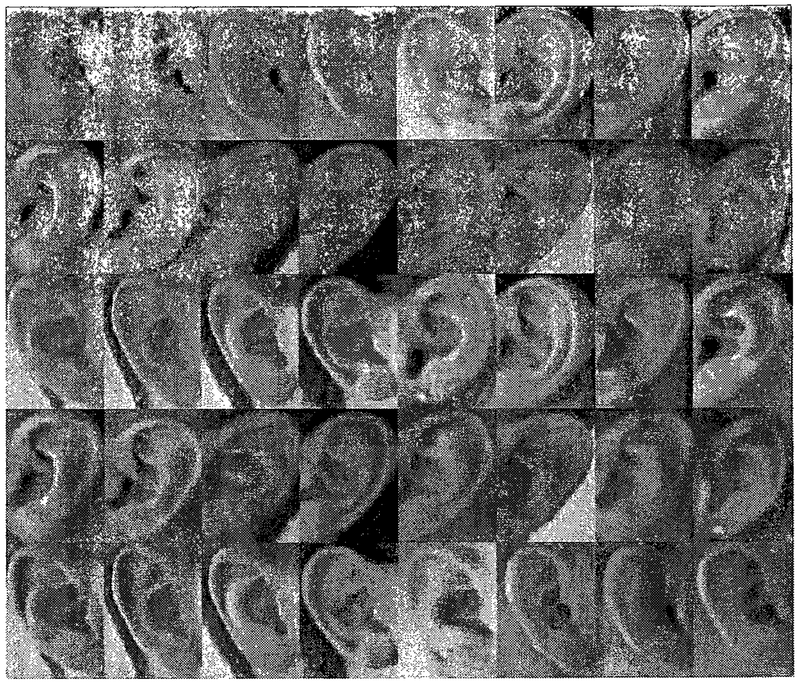


Figure 2.8: Ear Samples. Some samples of ear images clipped from an image of a person.

the ear was not visible) trains, automobiles etc. You can see some examples of these negative images in Figure 2.6.

2.8.2 Preprocessing

The positive images we eventually came up with after clipping them from the original image, came in many different sizes and intensities. Therefore, these input images were scaled to a size of 16x24, this was to reflect the rectangular dimensions of the ear. Due to the heavy processing involved in each stage of the training process, it took about a week to run on a Duo CPU 2.66GHz PC with 2 GB of RAM.

2.8.3 Performance

The performance of the AdaBoost Haar ear classifier was tested against a set of 200 profile images. The ear detector worked well with a few falsely detected ears, again this problem was overcome by selecting the largest detected object, see Figure 2.9.

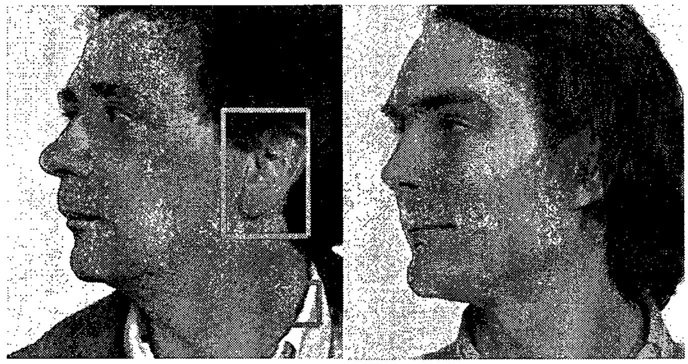


Figure 2.9: This shows an image where on the left we have a falsely detected ear.

We did get images where the ear was not detected, this was due to too many occlusions around the ear, see Figure 2.9 for an example. In these cases, the image was simply not included.

It failed to detect 5 of the images and this is due to occlusions around the ear, we could not improve on these results. For the purpose of this thesis, which was to research the capabilities of improving on recognition rates of persons using a multi-modal approach with face and ear, the ear detector suited our purpose. There are better ear detectors which have been discussed in detail by many researchers[22], but our detector worked well enough for us to proceed with our multimodal research.

2.9 Conclusions

The ear detector developed, was fast and accurate for the purpose intended. Our goal was to come up with an ear detector which was automatic and fast for our non-intrusive multi modal biometric system. It was capable of detecting all ears which were not occluded and to detect all but a few of the ears which were occluded, these being almost entirely occluded.

Chapter 3

Unimodal Biometrics

3.1 Introduction

A biometric system which uses face recognition is in great demand for the fight against crime and terrorism and for other various applications requiring a non-intrusive biometric recognition system. Such a system is however still a very challenging problem for researchers. This is mainly due to the variability in the human face under various conditions such as expressions, ageing, glasses etc. and external conditions such as lighting, background, angle of camera etc. These varying conditions are a challenge in trying to develop a robust face recognition system.

In analyzing these systems we must look at false acceptance rates (FAR), which is the probability that the biometric system incorrectly accepts someone, when it is in fact false. We also have to look at the false rejection rate (FRR), which is the probability that a biometric system incorrectly rejects a person when in fact it is true.

With the intention of developing a robust face recognition system, there have been many methods or approaches to face recognition which have been proposed, researched and tested [34]. We will look at and analyze some of these methods, with the intention of seeing what are the benefits and disadvantages of each of the different methods.

Some of the algorithms which have been proposed through the years include Principal Component Analysis/Eigenface(PCA)[43][3], Linear Discriminant Analysis(LDA)[13][7], Independent Component Analysis(ICA)[30], Local Feature Analysis(LFA)[33], and Support Vector Machines[17].

Recent developments in face recognition have also explored the area of 3D face recognition[6][10][8], in which the 3D surface area of the face is used for recognition.

In the following sections we will look at some of these face recognition methods and then discuss in depth the method we will be using in our multimodal biometric system.

3.1.1 Principle Component Analysis/Eigenface(PCA)

Principal Component Analysis(PCA) or Eigenface is a recognition algorithm which attempts to find the least mean squared error linear subspace which is projected from the original N dimensional data space to an M dimensional feature space[43].

Doing this, the eigenface is capable of achieving a reduction in dimensionality by using the M eigenvectors of the covariance matrix which corresponds to the largest eigenvalues. We then attempt to find the vectors which best fit the data, in affect the vectors which maximize the total variance of the projected data.

3.1.2 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis(LDA) attempts to find the best projection vectors that will maximize the separation between classes in the projected space[13]. it accomplishes this by finding the best projection vectors that will maximize the ratio between the different class data and the same class data.

3.1.3 Independant Component Analysis(ICA)

In PCA, we attempt to find an orthogonal projection for the face images, so that we achieve an uncorrelated transformation of the features. Independent Component

Analysis(ICA) on the other hand, attempts to find a non-orthogonal projection in an attempt to get transformation features which are statistically independent.

In PCA, the basis images depend on second order statistics, whereas in ICA, the concept is generalized to a model of higher order statistical relationships[41].

3.1.4 Local Feature Analysis(LFA)

For its method of recognition, Local Feature Analysis(LFA) uses eigen-subspace decomposition to construct a family of locally correlated feature detectors[33]. In the selection phase, LDA produces a minimally correlated subset of features which are topographically indexed that will define the subspace we are interested in.

In this system robustness against variability will come from the local representation of the subspace. In this respect, the features used in LDA are less sensitive to variability such as illumination. This method of face recognition is also used in a commercial system[34].

3.1.5 Elastic Bunch Graph Matching(EBGM)

In Elastic Bunch Graph Matching(EBGM) recognition is accomplished by constructing a dynamic link architecture using image graphs as a representation of face images[45]. In effect, this is a geometrical representation of a face image.

The image graph represents the face by using nodes and edges. The nodes in the image graph are used to represent facial landmarks such as nose, mouth and pupils. For the training images, a set of image graphs is used.

At each node, which represent local features, a set of Gabor wavelet coefficients are used. In the Gabor, their contains information about the orientations and frequencies at every node. To perform recognition, the facial image is matched against every graph in the training set and the closest match is chosen as the identity of the person.

3.1.6 Support Vector Machines(SVM)

Support Vector Machines(SVM) have been successful in face recognition systems[18]. This system works by mapping the data onto higher dimensional feature spaces, this is sometimes called the kernel trick[17].

The SVM then attempts to find the hyper plane which maximizes the margin of separation. This is in attempt to minimize the risk of misclassifying an image. In this manner, the SVM is not only able to classify training samples but also develop a better generalization of data it has not seen yet.

3.2 Face Recognition using Principle Component Analysis

PCA is one of the most successful methods used in face recognition and is also widely used for other image recognition applications such as the ear. PCA is a classical statistical technique working in the linear domain, making it suitable for many areas of research such as image processing, signal processing and control theory. PCA uses a statistical approach to recognition, it attempts to reduce the larger dimensionality of the data space to a much smaller dimensionality of feature space, which is how PCA is capable of describing the data economically.

For our research, once we have our objects detected, the next step is the preprocessing of the images. First, we extract only the portion of the image which was detected. For the face, the detected portion was further cropped in width to remove some of the unwanted and unneeded areas not making up the face. The ear was also extracted and further cropped to get a more accurate ear representation. This was all done automatically with experiments to determine the best cropping techniques and settings. For both face and ear we used Principle component Analysis (PCA) for recognition purposes.

PCA is a successful method for recognition in images and is largely a statistical method. PCA reduces the image space to a feature space; the feature space is then used for recognition. PCA translates the pixels of an image into principal components, which is called an eigenspace projection. Eigenspace is determined by the eigenvectors of the covariance matrix derived from the images.

Let a face/ear image be represented by $N \times N$ matrix

$$I(x, y).$$

Let the training database be represented by images

$$I_1, I_2, ..., I_M.$$

The average face Υ is

$$\Upsilon = \frac{1}{M}\sum_{n=1}^{M} I_n.$$

Each face differs from the average face Υ by vector

$$\phi_i = I_i - \Upsilon.$$

Set of vectors is subject to PCA seeking a set of M orthonormal vectors μ_n and eigenvalues λ_k. Let C be a covariance matrix

$$\begin{aligned} C &= \frac{1}{M}\sum_{n=1}^{M} \phi_n \phi_n^T \\ &= AA^T \end{aligned}$$

where μ_k are its eigenvectors and λ_k are its eigenvalues and

$$A = [\phi_1, \phi_2, ..., \phi_M].$$

The eigenproblem with $N^2 \times N^2$ matrix C is computationally intensive, so instead we can determine M eigenvectors μ'_k and M eigenvalues λ'_k by solving a smaller $M \times M$ matrix A^TA. Observe that $A\mu'_k$ are eigenvectors of $C = AA^T$. We then use linear combination of M training faces to form eigenfaces u_l

$$u_l = \sum_{n=1}^{M} \mu'_{l,n}\phi_n.$$

We usually use only a subset of M' eigenfaces corresponding to the largest eigenvalues.

For classification, an unknown face image I is resolved into weight components by the transformation

$$\omega_k = u_k^T(I - \Upsilon), \quad k = 1, ..., M'$$

and we form a new weight vector Ω_{new}^T

$$\Omega_{new}^T = [\omega_1, ..., \omega_M'].$$

Let Ω_k be a vector describing k-th face class. Then we compute the Euclidean distance

$$\epsilon_k = ||\Omega - \Omega_k||$$

and we classify face I to class k, where ϵ_k is minimum.

3.3 Cropping Techniques

We noticed during our experiments with the unimodal face and ear biometrics that cropping an image before recognition would drastically change the recognition rates. We experimented with techniques of shortening the width and length at different intervals to see how this would affect the results.

For face, bringing in the width and length so that most of the surface except for the part containing features such as eyebrows, eyes, nose, and lips was left, worked best. With the ear removing any excess of the image that was not correlated with the eye gave best results.

3.4 Face Recognition Implementation

Intuitively, the face seems like the most reasonable biometric, it is what humans use to recognize other humans, and they do it very well[49]. Computers on the other hand, struggle to achieve the accuracies that humans are able to achieve and at the speeds which humans are capable of recognizing other people.

One popular method which we use here is the technique of eigenfaces [43], this method uses PCA for comparing faces for recognition. Once we have detected, cropped and preprocessed our images we extract the portion which will be used for recognition. At this point, PCA is performed, it computes 'face space' represented by vectors, these eigenvectors which are computed by PCA, contain variance information. These eigenvectors, called eigenfaces can be thought of as features. A face is then projected onto the 'face space' to determine eigenface coefficients.

First, for each face in the training set, the face is transformed into 'face space' and the data for each face is stored into a 'face space'. For each test face, each image is projected onto this 'face space' and the system computes the distance from each face in the 'face space'. The smallest distance is assumed to be the match to the training set.

3.4.1 Face Recognition Remarks

We implemented our face recognition system using PCA, because it is a proven recognition method which is fast and was well suited for our automated multimodal system. We wanted something which was fast and proven and we felt we could improve upon it with our face detector and our researched and tested cropping methods.

Once the face was detected and extracted we went about cropping the image before we fed it to our PCA recognition algorithm, which we believed would increase substantially the recognition rates.

We experimented with many different cropping styles, in the end we choose a method which gave us the best recognition results for face. There is however much more research to be done in this area, as we feel cropping is crucial to achieving decent recognition rates.

Once the face was detected it was extracted from the original image. Then, this extracted image of the face was cropped and then fed into our PCA recognition system. We will discuss results for these recognition levels and others in the next section.

3.5 Ear Recognition Implementation

The ear as a biometric seems less intuitive than the face, it is not what humans normally use to identify each other. However, it does have some very interesting properties lending itself to be a very good biometric. A few of these being that it is a small area, it is normally visible, and does not change with a persons expression or mood, and stays relatively stable throughout a persons life. Given these properties, it seems like a good biometric to use in combination with the face.

Ear makes for a very interesting biometric and was first used as a biometric by Iannarelli[21]. He attempted to use ear as a biometric in a manual system by identifying important points on the ear and then using measurements to see if it could be used to identify individuals uniquely. In the end he came up with twelve features and could uniquely identify 10,000 individuals.

Due to its small size and many distinguishing features ear is an interesting candidate for a biometric system. It also exhibits a uniform colour, which is a desirable trait for a biometric candidate[37]. What attracted us to the ear as a possible candidate for our multimodal system with face, was that it is a non-intrusive biometric, so that both of our inputs into our automatic multimodal system of face and ear were non-intrusive, therefore our system could be used in a non-intrusive application.

Many of the methods we mentioned for face recognition can also be used for ear recognition. In addition, there are some other methods which were developed specifically for ear recognition. There are also geometric approaches to ear recognition such as those described by Coras[11], who in his studies uses geometric properties such as width and length of the ear to create a feature vector. For comparison and recognition he then uses the measurements of these features against other features vectors to determine to best candidate.

Another method to ear recognition is the Gaussian approach described by Hurley et al.[20]. In this approach the ear is modeled using a Gaussian force field. The pixels of the ear image create a magnetic like force field as they exert forces against one another. These force field lines which are created by the magnetic like force on the pixels generate channels which are then used for identification.

There are researchers also studying the validity of using 3D ear shape for recognition[46, 24] and also in multibiometrics[47]

The method used for ear recognition is eigenears, which like eigenfaces, uses the method of PCA for comparing ears for recognition.

We chose to use PCA or eigenears in conjunction with our PCA method for face, because it allowed us to maintain a fast and proven recognition system with the hopes that recognition rates would increase once we applied a multimodal technique to the recognition system.

3.5.1 Ear Recognition Remarks

Like for face, for ear we implemented our ear recognition system using PCA because mostly due to the fact that it is a proven recognition method which is fast and was well suited for our automated multimodal system. We wanted something which was fast and proven and we felt we could improve upon it with our face detector and researched and tested cropping methods.

The ear was first detected using our ear detector, we then went about cropping to ear to get the best recognition results. Once the ear was detected, extracted, and cropped we had one additional step to perform and that was to reflect the ear horizontally if it was a left ear, that way all ears had the same direction, or in other words the curvature for all ears was located on the same side. Without this step we would be unable to properly recognize the ears as the ears would be pointed differently and we would get incorrect values.

When all of the preprocessing was accomplished, we then used this final preprocessed ear image for our recognition process. We thus, fed it to our PCA recognition algorithm, which we believed would produce substantially better recognition rates with the preprocessing done.

We experimented with many different cropping styles, in the end we choose a method which gave us the best recognition results for ear. There is however more research to be done in this area as we feel cropping is crucial to achieving decent recognition rates.

Therefore, once the ear was detected it was extracted from the original image. Then, this extracted image of the face was cropped and fed into our PCA recognition system. We will discuss results for these recognition levels and others in the next section.

3.6 Conclusions

In the end, we decided to use PCA for both face and ear, so we created eigenface and eigenear recognition algorithms for our system. We already mentioned many of the reasons for this, but basically because PCA is a proven method, which, with the properly applied preprocessing is capable of good recognition rate and that the algorithm fits quite well into our automated multimodal system. Another reason behind our decision was that our research was centered on improvement in recognition rates using a multimodal approach and less on trying to get the best recognition rates using a unimodal approach.

By combining face and ear as biometrics we hoped to increase recognition rates as opposed to the unimodal recognition rates. Both face and ear being non-intrusive biometrics our system was capable of maintaining the non-intrusive requirement we had for our system.

The face and ear was also beneficial as combined modes because of the fact that, the face being larger, but which changes with expressions and could be occluded with certain items such as eye glasses, sun glasses, hats, moustaches, beards, scarves, etc. and which could change over time. The ear is of a small area, does not change much overtime, does not change with expressions, on the other hand is more easily occluded with some of the items mentioned for face, but also by hair, hats, ear muffs, ear phones, etc. By combining both these biometrics we hoped to increase the coverage we achieved, as an example if one is occluded we can still do recognition on the other. However, the system does become more complicated as we move to a more complex multimodal system.

Let us now move on to our mulitmodal research and see what we were able to achieve, and review the advantages and disadvantages of such a system.

Chapter 4

Multimodal Biometrics

4.1 Introduction

The basic idea behind multibiometrics is to combine two or more biometrics with the hope of achieving better recognition results. The method we use for our multibiometric system is to use a multimodal approach (face and ear) with a single algorithm (PCA).

The goal of a multibiometric system is to increase or improve the recognition rates achieved over unimodal biometric system, this is achieved using many methods. A multibiometric system normally overcomes many of the factors that plague a unimodal biometric system such as noise, variability and error rates [9]. Apart from the benefit of a higher recognition rate, a multimodal biometric system can also help in lowering false rejection error rates.

4.2 Multimodal Biometric Methods

In a multimodal biometric system the face and ear can be combined to create a more robust biometric system. The ear can be considered as another face component and therefore is a natural component in a multimodal system with face, giving us a more robust system with respect to occlusions and recognition rates.

There are many methods in which two or more modes can be combined in a multimodal system. I will discuss two of these multimodal methods.

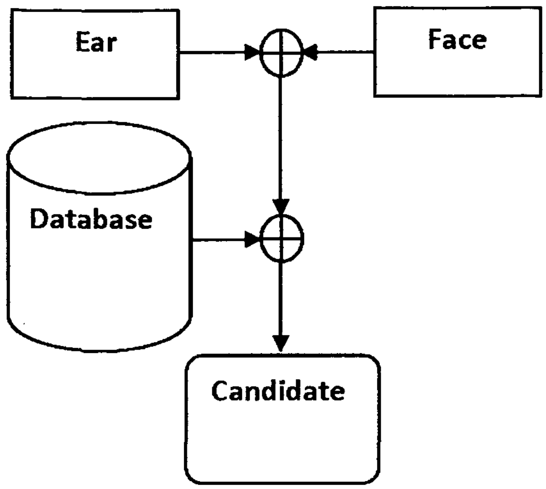


Figure 4.1: Parallel Multimodal System.

The first method we will discuss is the parallel method in which both modes are run simultaneously and the results of these unimodal methods are then merged to get a unique measurement. As a simple example, using the PCA method, we can take the two distances from both face and ear and add them to create a new distance measurement for the combined modal system. There are of course, much more elaborate methods than this, which we will describe later in our experiments section.

This kind of multimodal method is advantageous for compensating modes which are occluded or where some other information is lost in one of the modes. The other mode in these instances can compensate for the lost information of the mode. In our case, the ear is capable of filling in for face when it is occluded or in aiding recognition rate when they are both visible. You can see a visual diagram of this type of multimodal system in Figure 4.1.

The second method is the cascade method combining more than one mode for a biometric system[23]. In this kind of a multimodal system, one of the modes is used to remove from the dataset non-candidates according to that mode. Then, the other mode will use only the rest of the dataset to determine which candidate best fits the criteria. You can of course start with either one of the modes and end with the other.

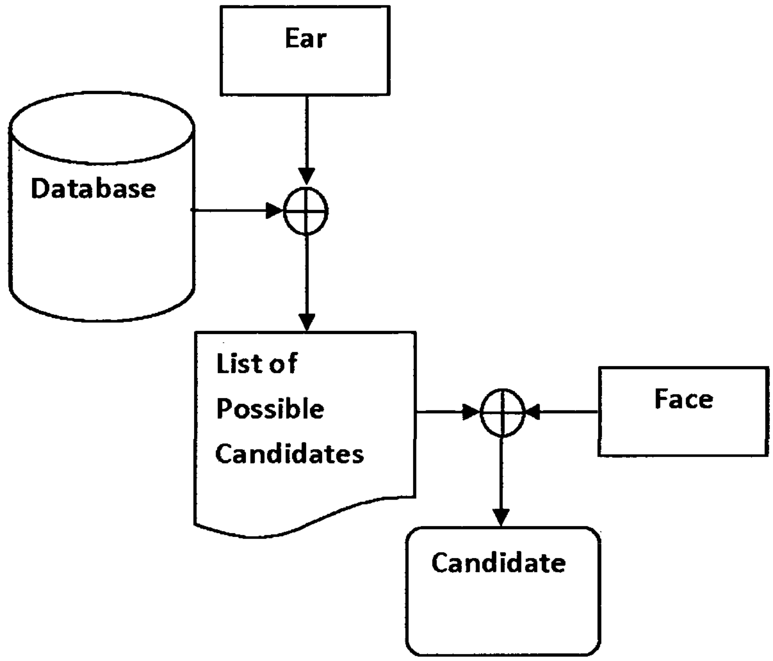


Figure 4.2: Cascade Multimodal System

As an example in our case we can prune the dataset with face and then determine the best fit candidate with ear, or vice versa of course.

Research will determine which mode is best as a pruner and which is best as a determiner of the best candidate[2]. You can see a visual diagram of this type of multimodal system in Figure 4.2.

We decided to do our research of multimodal biometrics using a parallel system. Using this method we were capable of achieving some decent recognition improvements over unimodal methods for face and ear.

4.3 Individual Recognition and Evaluation

We performed Principal Components Analysis individually for both face and ear. For our dataset we used a 100 subject gallery where both frontal and profile images were available. We also needed to have two of both frontal images and profile images if the subject was to be included in the dataset. We gathered our dataset from the FERET database where both frontal and profile are available.

For each of the features, the gallery of gathered subjects is used to generate the PCA space, in other words we a had a PCA 'face space' and a PCA 'ear space'. Each run generated a table of distance measurements from which the least distance is seen as the best candidate match for the subject. The distance measurements for each candidate are the distances from the candidate to the subject, therefore the least distance is seen as the best match.

Each mode was first run separately, specifically we ran face through to the recognition phase and then did the same for ear. As the distance measurements we use the Euclidean distance and the Mahalanobis distance. In the preceding sections we will briefly describe these two distance measurements and then look at our individual recognition results for both face and ear.

We achieved better results for face than for ear, the best result for face was 95.2% while the best result for ear was 75.8%, see Table 4.1 and Table 4.3 for results.

4.3.1 Euclidean Distance

The Euclidean distance as mathematically defined as the distance between two points, sometimes also referred to as the 'ordinary' distance or the distance between two points we can measure with a ruler. The Euclidean distance can be proven using the Pythagorean theorem.

More formally, the Euclidean distance between points $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ in Euclidean n-space would be defined as follows;

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2}$$

Or, as;

$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

4.3.2 Mahalanobis Distance

The Mahalanobis distance measurement was first introduced by P.C. Mahalanobis in 1936[28]. It is used in applications trying to determine similarity between an unknown set and a known one. The Mahalanobis distance is often used in applications such as cluster analysis and in other applications requiring classification.

To use the Mahalanobis distance for classification, for example to see if a test point belongs to any one of a group of N classes. First, we need to take the covariance matrix of each class. Then with a test sample, we compute the Mahalanobis distance to each class, the class with the minimal distance is the class which the test point belongs to.

More formally, given a group of values where the mean $\mu = (\mu_1, \mu_2, ...\mu_\rho)^T$, and with a covariance matrix Σ for a multivariate vector $x = (x_1, x_2, ...x_\rho)^T$, the Mahalanobis distance is defined as,

$$D_M(x) = \sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

4.3.3 Face Recognition and Evaluation

For face recognition, after the image was preprocessed we ran it through our PCA algorithm. As mentioned before, our dataset contained 100 subjects. The best recognition rate we achieved was 95.2% using a Mahalanobis distance measurement, see Table 4.1 for results. You can see that the Euclidean distance performed slightly lower than the Mahalanobis distance.

Distance	Face Recognition Rate
Euclidean	93.6%
Mahalanobis	95.2%

Table 4.1: Unimodal face recognition rates using different distance models.

Once we had our projected face we then calculated the distance from this face to every face in our training dataset. The image in our training dataset which had the lowest distance is the one which was classified as being the best candidate for our subject. You can see a sample of these calculated distance measurements using both Euclidean and Mahalanobis distances for a few of the candidates in our dataset in Table 4.2. The least distance measurement being the best candidate, in this instance the best candidate would be 00721 for both distance measurements.

Candidate	Euclidean Distance	Mahalanobis Distance
00706	2.09334E+07	1.75387E+09
00707	2.07378E+07	1.53092E+09
00708	2.06042E+07	1.73069E+09
00711	5.02336E+07	1.76720E+09
00712	1.80437E+07	1.55883E+09
00713	2.13500E+07	1.44996E+09
00715	7.12372E+07	1.59484E+09
00718	1.73229E+07	1.70514E+09
00721	**1.06490E+06**	**2.13263E+08**
00722	3.61059E+07	1.72997E+09

Table 4.2: Sample Face Distances.

The distance measurement gives us a possible candidate for our recognition application, in Figure 4.3 we see images of possible candidates for a subject with their distance measurements. We are only displaying 5 of the possible candidates, but we ran it against the complete dataset. In this sample we can determine how a subject is matched to possible candidates using distance measurements. In this instance both distance measurements correctly choose a best match candidate, as both have the lowest distance.

However, that is not always the case (otherwise our jobs as researchers in face recognition would be accomplished), and we do have cases where subjects are incorrectly matched to candidates. That is the premise of our research in multimodal biometrics, to see if in cases like these, the other mode can compensate and actually

Subject: 708

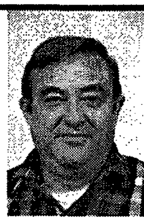
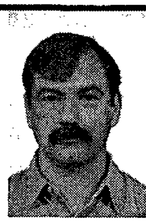
	Candidate	Euclidean Distance	Mahalanobis Distance
	706	1.43216e+007	1.61465e+009
	707	1.46384e+007	1.48886e+009
	**708**	**1.47136e+006**	**2.33209e+008**
	711	3.51303e+007	1.58248e+009
	712	1.14182e+007	1.44013e+009

Figure 4.3: Face Image Distance Measurements with Correct Match.

correctly match a candidate to a subject.

That said, our face recognition algorithm performed quite well with peak recognition rates of 95.2%, there are however incorrect matches. In Figure 4.4, we see a sample of a subject incorrectly matched to a candidate. In this instance, both distance measurements have incorrectly matched the subject to candidates. It will be interesting to see if these incorrectly matched subjects can improve with a multimodal biometric system.

4.3.4 Ear Recognition and Evaluation

For ear recognition, we first preprocessed the image and then ran it through out PCA algorithm. As mentioned before, our dataset contained 100 subjects. The best recognition rate we achieved using ear as an input was 75.8% using a both the Euclidean and Mahalanobis distance measurements, see Table 4.3 for results. You can see from the table the results did not change between distance measurements.

Distance	**Ear Recognition Rate**
Euclidean	75.8%
Mahalanobis	75.8%

Table 4.3: Unimodal ear recognition rates using different distance models.

The next step in our ear recognition process was to take the projected ear and then calculate the distance from this ear to every ear in our training dataset. The image in our training dataset which had the lowest distance from the projected ear is the one which was classified as being the best candidate for our subject. You can see a sample of these calculated distance measurements using both Euclidean and Mahalanobis distances for a few of the candidates in our dataset in Table 4.4. The least distance measurement being the best candidate, in this instance the best candidate would be 00713 for both distance measurements.

Subject: 812

	Candidate	Euclidean Distance	Mahalanobis Distance
	785	**4.58518E+06**	1.00979E+09
	798	9.07750E+06	**7.37066E+08**
	799	3.77199E+07	1.35259E+09
	***812***	5.19798E+06	8.84530E+08
	816	2.64810E+07	1.37970E+09

Figure 4.4: Face Image Distance Measurements with Incorrect Match.

Candidate	Euclidean Distance	Mahalanobis Distance
00706	1.48560E+07	1.11942E+09
00707	2.17146E+07	1.19981E+09
00708	1.52913E+07	1.25766E+09
00711	3.00794E+07	1.25473E+09
00712	3.53754E+07	1.21968E+09
00713	**1.93287E+06**	**3.75250E+08**
00715	5.32190E+07	1.24166E+09
00718	1.90368E+07	1.14052E+09
00721	2.28253E+07	1.23895E+09
00722	4.73500E+07	1.30703E+09

Table 4.4: Sample Ear Distances.

The distance measurement gives us a possible candidate for our recognition application, in Figure 4.5 we see images of possible candidates for a subject with their distance measurements. We are only displaying 5 of the possible candidates, but we ran it against the complete dataset. In this sample we can determine how a subject is matched to possible candidates using distance measurements. In this instance both distance measurements correctly choose a best match candidate, as both have the lowest distance.

As for face, we do not always get correct matches for subjects for ear too, and do have cases where subjects are incorrectly matched to candidates.

Our ear recognition algorithm did not perform as well as the face recognition algorithm, the best we are able to achieve for ear recognition was 75.8% see Table 4.3 for results, whereas we were able to achieve peak recognition rates of 95.2% for face.

To illustrate some of these mismatches you can look at Figure 4.6. In this table, we see a sample of a subject incorrectly matched to a candidate.

In this instance, both distance measurements have incorrectly matched the subject to candidates. Again, it will be interesting to see if these incorrectly matched subjects can improve with a multimodal biometric system.

Subject: 00722

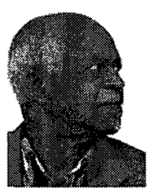
	Candidate	Euclidean Distance	Mahalanobis Distance
	00813	1.43216e+007	8.17085e+008
	00814	3.98075e+007	1.14601e+009
	00815	2.80185e+007	1.07254e+009
	00816	3.59356e+007	9.96727e+008
	**00722**	**3.90871e+006**	**4.99585e+008**

Figure 4.5: Ear Image Distance Measurements with Correct Match.

Subject: 775

	Candidate	Euclidean Distance	Mahalanobis Distance
	706	**6.83835E+06**	9.53180E+08
	713	9.14893E+06	**7.21720E+08**
	**775**	7.68272E+06	8.99598E+08
	796	1.62874E+07	1.05904E+09
	801	2.87700E+07	1.11909E+09

Figure 4.6: Ear Image Distance Measurements with Incorrect Match.

4.4 Fusion Recognition and Evaluation

The fusion of a multibiometric system is where the results for each individual biometric system is used to determine the classification. There are many methods of achieving this, one simple method is the sum of both biometric results to determine the best classification. We will experiment with many methods of fusion which will be discussed in detail in the experimental section.

In our multimodal biometric system we will experiment with many methods of fusion to see what results we can achieve. We will experiment with from as simple methods as summing the two results and choosing the candidate with the least distance to more elaborate methods of normalizing the distances and then applying some kind of fusion method. We will also look at some weighted sums and see if this will lead us to some superior results.

The key factor to any multibiometric system is to use a fusion technique that increases recognition rates, which method to use will depend on the algorithms and inputs used. We want to optimally combine the results from both biometrics to increase the levels of correct recognition.

The best result in our multibiometric system was achieved using a Mahalanobis distance for each individual biometric and then a combined normalized sum using a weight of 0.7 for face and a weight of 0.3 for ear, with this we achieved a 100% recognition rate[27].

These results that were achieved are definitely very exciting and interesting to note at how combining two modes can drastically improve results, and confirms in this case the benefits of using a multimodal recognitions system as opposed to using any one of the modes used in a unimodal system. In later sections, we will describe these experiments in detail and look at how they were achieved.

The main concern or issue that we will try to understand, explain and analyze in detail, is how this multimodal system was able to improve the recognition rates as compared to using the same modes in a unimodal fashion.

4.5 Contributions

In researching the development of an automated multimodal biometric system we experimented with many techniques in an attempt to achieve significantly higher recognition rates. The face and ear are automatically detected, extracted, cropped and preprocessed in preparation for our recognition algorithm. In [26], we describe fully the methods and experiments we used in researching and developing our automated multimodal biometric system.

We also experimented with many fusion techniques, in an attempt to discover which methods yielded the best results. We will not only present the experimental data, but we will also try to describe and analyze the data so that a better understanding of multimodal biometrics, more specifically face and ear multimodal biometrics can be gleaned from research we did in this area of multimodal biometrics[27].

Experimental results will show that significant improvement in recognition rates can be achieved using various fusion techniques in multimodal biometrics and that results will vary greatly depending on fusion technique implemented.

4.6 Conclusions

In this section, we describe and look at the results of our multimodal biometric system. We went went into detail in describing the recognition results we achieved for face and ear individually. Looking at both the successes and shortcomings for each mode and their respective results.

We also described our multimodal technique and some of the fusion methods we applied to get better results. As mentioned, we were able to achieve 100% multimodal recognition rate for one of the fusion techniques. These results definitely confirm previously stated literature that a multimodal biometric system using face and ear can improve results over using either mode on its on.

The results will show that a multimodal biometric system is capable of achieving superior recognition rates over any of the modes employed on its own. Face and ear seem to be very complimentary modes, while the face is larger it does change easily with expressions. The ear is much smaller than the face but it maintains is structure with expressions and is less likely to change with time. Together these two modes seem to offer a very good biometric as the results of our experiments show.

In the next section, we will describe our experiments and all results in detail. We will look into all the fusion techniques we used in our implementation and all other techniques we might have employed in coming up with our results.

Chapter 5

Experiments

5.1 Introduction

In this section, we will describe and review our experiments and also all results from these experiments. The main purpose of this research and these experiments was to ascertain whether combining the two recognition modes of face and ear would lead to a better recognition rate as opposed to running the two respective modes individually.

In an attempt for completeness, we experimented with many techniques even though we suspected the outcome of some of these experiments, and were found to be sometimes surprising. Many experiments were done with a multiple of fusion techniques, in an attempt to discover which methods yielded the best results. The experiments yielded many interesting results, some expected and some surprising. We will attempt to describe all in full detail so that they can be understood.

In the preceding sections, we will first try to describe the system setup for the experiments. We will look at all aspects of our system setup. A look at the database will follow and then the experiments. We will not only present the experimental data, but we will also try to describe and analyze the data so that a better understanding of multimodal biometrics, more specifically face and ear multimodal biometrics can be understood from the research we did in this area of multimodal biometrics.

Experimental results will show that significant improvement in recognition rates can be achieved using a multimodal biometric approach to recognition.

5.2 Database

To demonstrate the concepts in this thesis, we ran many experiments using many different methods in an attempt to get the best and fullest understanding possible of face and ear multimodal biometrics. We will attempt to explain fully the experiments we went through in applying these concepts.

For our dataset we used a subset of the FERET [35] see Figure 5.1 for samples of images. We used a set of 100 persons, for each person there must have been at least two frontal images and two profile images. This was so that we had enough of both frontal and profile images for each person, so that a training dataset and a testing dataset could be created.

The training dataset consisted of one frontal image and one profile image for each of the 100 persons. The frontal images were used for the face biometrics and the profile images were used for the ear biometrics. This training dataset, which contained a frontal and profile image of each person, allowed us to run our detection, preprocessing and PCA algorithm on both face and ear for each person in the training dataset. At this point, we would have a training data file for both face and ear, containing face and ear information for each person.

The testing dataset included at least one frontal image and one profile image for each of the 100 persons, in this respect we could get, for each person a face biometric and an ear biometric, so that we could test the multibiometric system. The test dataset is used to take a subject in the test dataset and then see what type of recognition rates we can get against the training dataset.

5.3 System Setup

Our automated multimodal biometric system using face and ear as input modes was built and run on a PC with a Intel dual core CPU @2.66GHz and two GB of RAM. The operating system of the computer was Windows XP with service pack 2. The

Figure 5.1: Samples of images used, 1st row train face, 2nd row test face, 3rd row train ear, 4th row test ear.

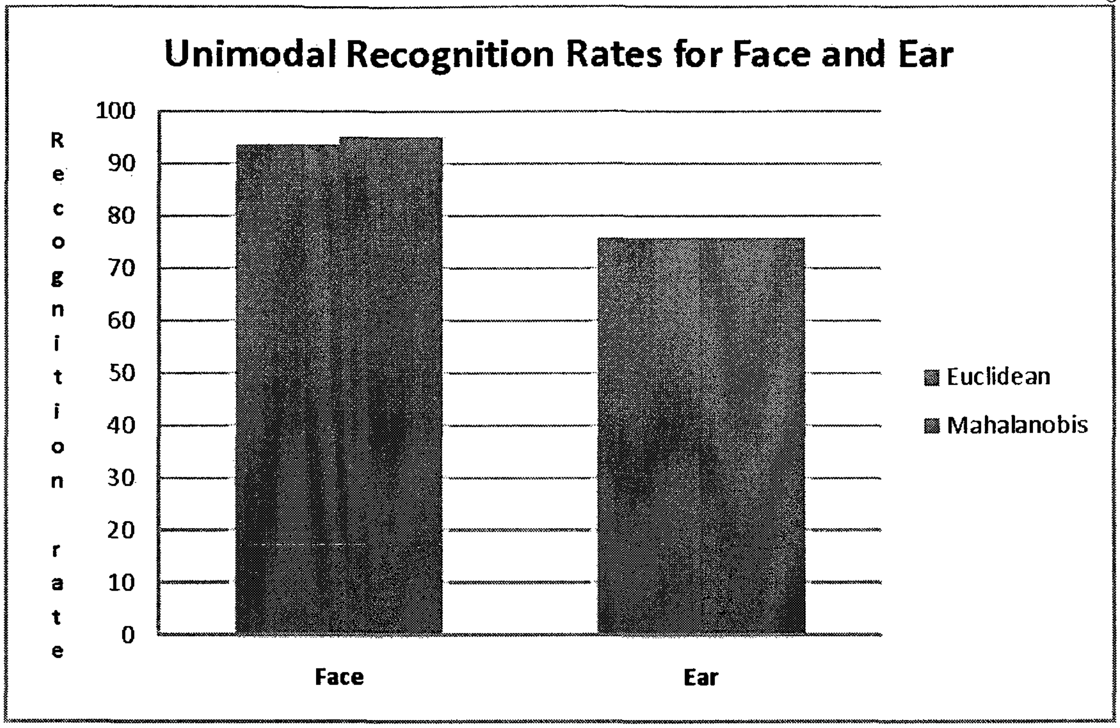


Figure 5.2: Unimodal Recognition Rates for Face and Ear.

application was written in C++ using Microsoft Visual Studio 2008.

5.4 Unimodal Recognition

Face and ear recognition as unimodal biometrics have already been discussed, but it is worth mentioning again before we proceed to the multimodal section for future comparisons in later sections. Keeping these rates in mind, we will see later whether they have any affect on the rates for our automated multimodal biometric system using face and ear as input modes.

In Figure 5.4, you can get a visual depiction of the recognition rates for face and ear as unimodal biometrics. This graph presents the unimodal recognition rates for face and ear in an easy to read Bar Graph fashion for both Euclidean and Mahalanobis distances.

5.4.1 Unimodal Face Recognition

As mentioned and explained in a previous section, for unimodal face recognition our results were pretty good and in line with many face recognition algorithms. Using the Euclidean distance measurement the recognition rate is 93.6% and using a Mahalanobis distance measurement the recognition rate achieved is 95.2%, a slight improvement, see Table 4.1 for results.

5.4.2 Unimodal Ear Recognition

For unimodal ear recognition our experimental testing did not achieve the same level of rates achieved using face. The recognition rate for both the Euclidean and Mahalanobis distances was 75.8%, see Table 4.3 for results.

5.5 Multimodal Recognition

At this point, once we had the results from the unimodal biometrics for face and ear, it was time to fuse the two to see if we could achieve better results.

In order to properly compare and fuse distances from different modes, there is a need for an accurate normalization technique we can apply on the distances. To normalized the distances in our experiments calling for normalization we used the min-max normalization[29]. As an exercise, we complete a few experiments and show the results where normalization is not done.

More formally, to normalize the distance x in the dataset; we get the normalized value $x_i^{'}$ by;

$$x_i^{'} = (x_i - min_i)/(max_i - min_i)$$

where, min and max are the minimum and maximum values for each dataset.

Using this normalization we will get values in the range of $[0, 1]$ for each distance. This will allow us to fuse face and ear values with more accurate comparisons.

Candidate	**Distance**	
	Mahalanobis	**Normalized Mahalanobis**
00706	1.36678e+008	0
00707	1.25624e+009	0.918055
00752	1.20509e+009	0.876106
00768	1.27156e+009	0.930616
00774	1.33894e+009	0.985867
00775	1.14325e+009	0.825395
00776	1.35618e+009	1
00814	1.22822e+009	0.895073
00815	1.23459e+009	0.9003
00816	1.14159e+009	0.824037

Table 5.1: Sample Normalized Mahalanobis Face Distances.

As an example, if you look at the data in Table 5.1, you will see how the min-max normalization transforms the distances into values between 0 and 1. The value of 0 will always go to the distance with the lowest value, and 1 will always go to the distance with the greatest value. Everything else will fit somewhere in between.

As you can see, we are now able to compare apples with apples and oranges with oranges, or in our case face and ear.

Candidate	**Face**		**Ear**	
	Distance	**Normalized**	**Distance**	**Normalized**
00713	1.50731e+009	0.795371	1.07424e+009	0.791158
00718	1.50145e+009	0.791631	1.06808e+009	0.782685
00721	1.66943e+009	0.898813	9.62873e+008	0.637837
00722	2.60779e+008	0	4.99585e+008	0
00732	1.54012e+009	0.816305	9.97497e+008	0.685506
00736	1.82802e+009	1	1.05802e+009	0.768834
00739	1.43991e+009	0.752363	9.96477e+008	0.684102
00792	1.39405e+009	0.723105	8.38748e+008	0.466947
00793	1.55724e+009	0.82723	1.22593e+009	1
00816	1.44869e+009	0.757967	9.96727e+008	0.684446

Table 5.2: Sample Face and Ear Normalized Distances.

As another example, take a look at Table 5.2 which shows us the face and ear distances before and after normalization. We can easily see that with the normalized values we get better comparisons between face and ear. The normalized values for both face and ear are all within the [0,1] range, with the best match for both having a normalized distance of 0.

Without normalization it is hard to compare and hard to fuse distances coming from different modes, because they may have completely different min's and max's. Therefore, making it hard to compare the two with these values, it is easier to fuse the two together using some fusion methods as we will look at in later sections.

In our experiments, we will included some experiments with distances that are not normalized for comparison purposes only.

Now that we have a good understanding of the min-max normalization technique employed in our experiments, we can start looking at some experimental data.

5.5.1 Non-Normalized Sum

As mentioned previously, we will do experiments using non-normalized distances as an exercise.

We first did a sum using non-normalized distances; the first was a simple sum, where the distance measurements for face and ear were summed and the lowest summed distance was considered a best match. This was done for both distance measurements, see Table 5.3.

Distance Type	Recognition Rate
Euclidean	95.2%
Mahalanobis	98.4%

Table 5.3: Combined Non-Normalized Recognition Rates.

As can be seen from the results, the recognition rates achieved were higher than in the unimodal systems, with Mahalanobis distance achieving greater success rates than the Euclidean distance.

We were able to achieve greater recognition rates for both distances as compared to the same distances for unimodal recognition.

If we look at the data for some of the candidates who were recognized incorrectly in the unimodal biometric, but correctly in the multimodal biometric, you will see how this is possible.

In certain cases one mode can compensate for the other even if the distances are not normalized. If you look at the data in Table 5.4 you can see how this may be possible. In this case, the face recognition incorrectly identifies the subject as candidate 00708, with its second best candidate actually being the subject. Ear recognition however, correctly matches the subject to candidate 00707.

Candidate	**Distance**		
	Face	**Ear**	**Sum**
00706	1.25376E+09	1.05692E+09	2.31068E+09
00707	8.93037E+08	**4.70358E+08**	**1.36340E+09**
00708	**7.05484E+08**	1.16739E+09	1.87287E+09
00711	1.15751E+09	1.20618E+09	2.36369E+09
00712	1.30226E+09	8.62152E+08	2.16441E+09
00713	1.13408E+09	9.56125E+08	2.09021E+09
00715	1.33495E+09	1.10457E+09	2.43952E+09
00718	9.81345E+08	1.07699E+09	2.05834E+09
00721	1.10393E+09	9.69211E+08	2.07314E+09
00722	1.13453E+09	9.96581E+08	2.13111E+09

Table 5.4: Sample Face and Ear Non-Normalized Sum.

When the distances are summed we get the correct match, partly due to the fact that faces second lowest distance was the correct match and also because the ear distance for the correct match compensates for the incorrect match for face.

Weight(Face/Ear)	Distance Type	
	Euclidean	Mahalanobis
(0.9/0.1)	95.2%	98.4%
(0.8/0.2)	95.2%	98.4%
(0.7/0.3)	95.2%	98.4%
(0.6/0.4)	95.2%	98.4%
(0.5/0.5)	95.2%	98.4%
(0.4/0.6)	95.2%	98.4%
(0.3/0.7)	95.2%	98.4%
(0.2/0.8)	95.2%	98.4%
(0.1/0.9)	95.2%	98.4%

Table 5.5: Non-Normalized Face/Ear Weighted Sum Recognition Rates.

5.5.2 Non-Normalized Weighted Sum

As a second exercise on non-normalized distances, we tried to see if results would improve using weighted sum. In our experiments, the weights on non-normalized distances did not seem to have any influence on the recognition rate.

Table 5.5 presents recognition rates for different weights between face and ear for non-normalized distances. As can be seen from the data the weights in our case did not have any influence on the recognition rates we got. We experimented with distances from 0.9 face and 0.1 ear to 0.1 face and 0.9 ear, and did not receive any change in the recognition rates.

The process of how we calculated the non-normalized weighted sum from the face and ear distances can be seen in Table 5.6. This table presents a few samples of non-normalized weighted sum for a combined face/ear. We first take the distances and multiply them by the weights, in this case 0.7 for face and 0.3 for ear. Then we add the results of these two operations and get the non-normalized weighted sum. As mentioned before, in our case the weights did not seem to have any influence on recognition rates.

Candidate	Distance				
	Non-Weighted		Weighted(Face 0.7/Ear 0.3)		
	Face	Ear	Face	Ear	Sum
00706	1.254E+09	1.057E+09	8.776E+08	3.171E+08	1.195E+09
00707	8.930E+08	**4.704E+08**	6.251E+08	1.411E+08	**7.662E+08**
00708	**7.055E+08**	1.167E+09	4.938E+08	3.502E+08	8.441E+08
00711	1.158E+09	1.206E+09	8.103E+08	3.619E+08	1.172E+09
00712	1.302E+09	8.622E+08	9.116E+08	2.586E+08	1.170E+09
00713	1.134E+09	9.561E+08	7.939E+08	2.868E+08	1.081E+09
00715	1.335E+09	1.105E+09	9.345E+08	3.314E+08	1.266E+09
00718	9.813E+08	1.077E+09	6.869E+08	3.231E+08	1.010E+09
00721	1.104E+09	9.692E+08	7.728E+08	2.908E+08	1.064E+09
00722	1.135E+09	9.966E+08	7.942E+08	2.990E+08	1.093E+09

Table 5.6: Sample Face/Ear Non-Normalized Weighted Sum.

5.5.3 Normalized Sum

In this section we will explore normalized distances and see how normalization effects the recognition rates of our multimodal biometric system using face and ear as modes. Using this method the distances for face and ear are first normalized using the min-max normalization technique, then we sum the two normalized distances to get a normalized combined sum. The candidate with the least distance, is considered to be the best candidate.

Distance Type	Recognition Rate
Euclidean	95.2%
Mahalanobis	98.4%

Table 5.7: Combined Face/Ear Normalized Recognition Rates.

It is worth first looking at the recognition rates we achieved using normalization for the total dataset, these can be seen in Table 5.7. This table presents the recognition rates achieved using both the Euclidean and Mahalanobis distances. With the Euclidean distance the recognition rate achieved was 95.2% and with the Mahalanobis distance a recognition rate of 98.4% was achieved.

For a better understanding and clarification, we will look at a couple of examples of how normalization affects the selection of a best fit candidate for a subject. We will look at a case were the normalized sum correctly selects the best candidate even where one of the modes had a different candidate match, and we will look at a case where the normalized sum candidate is completely different from the face or ear unimodal selection and is incorrect.

Candidate	**Normalized Distance**		
	Face	**Ear**	**Sum**
00718	0.677116	0.447693	1.124809
00721	0.470315	0.276721	0.747036
00722	0.703521	1	1.703521
00750	0.728494	0.350975	1.079469
00786	1	0.523457	1.523457
00793	0.652536	0.549456	1.201992
00794	0.860332	0	0.860332
00796	0.565252	0.757541	1.322793
00803	0	0.173009	**0.173009**
00816	0.489258	0.296058	0.785316

Table 5.8: Sample Face/Ear Sum of Normalized Distances with Correct Match.

In the first example, from Table 5.8 we can see the data from some of the candidates for subject 00803. I have included candidates who had the least and most distances for both face and ear, and some other random candidates. From the data, we can see that face correctly matches the subject to the candidate, while we see that for ear, there is an incorrect match to candidate 00794, which has the least distance.

Combined however the lowest normalized distance is candidate 00803 which is the correct match. If you look at the data, you will see that for face the second best match had a lot higher distance at 0.470315. For ear, the second best match had a normalized distance of 0.173009 and was actually the correct match and had a very low distance. These two factors lead to a correct match when the two are combined with a distance of 0.173009, where the second best distance had a much higher rate of 0.747036.

In our next example we will look at a situation where one of the modes in unimodal mode correctly identifies the correct candidate and the other doesn't, and combined the normalized sum selects an incorrect candidate.

We will look at the data to try to ascertain why this situation occurs. The data can be seen in Table 5.9, from the table we can see how face correctly select the candidate for subject 00800, it is important to note however the low distance of the second best candidate for face at a rate of 0.16967.

Ear incorrectly select candidate 00792, again however it is important to note the second best candidate which happens to be the same candidate as face's second best candidate with a very low distance of 0.0632297.

You can probably guess now, how when these distances are combined the second best candidates for both face and ear will lead to the lowest combined distance and incorrectly selecting candidate 00751 as the best match for subject 00800.

In the next section on weighted normalized sum, we will re-visit this example and see if a weighted normalized sum can remedy this situation by correctly selecting the candidate for subject 00800, even though the situation we just describes exists.

Candidate	**Normalized Distance**		
	Face	**Ear**	**Sum**
00715	0.170573	0.56654	0.737113
00722	0.83388	1	1.83388
00751	0.16967	0.0632297	**0.2328997**
00792	0.528418	0	0.528418
00793	0.398919	0.794843	1.193762
00794	1	0.295133	1.295133
00797	0.288109	0.689195	0.977304
00800	0	0.327908	0.327908
00803	0.520331	0.600329	1.12066
00813	0.559398	0.420601	0.979999

Table 5.9: Sample Face/Ear Sum of Normalized Distances with Incorrect Match.

5.5.4 Weighted Normalized Sum

In this section, we will look at weighted normalized sum and see if weighted values for face and ear can improve the recognition rate of our algorithm. Using weighted values, the best recognition rate was achieved using a normalized Mahalanobis distance with a weight of 0.7 for face and 0.3 for ear, see Table 5.10 for all results. From this table, we can see the affects of different weight values for face and ear.

These recognition results presented in Table 5.10 improve the results obtained in [29].

Weight(Face/Ear)	**Distance Type**	
	Euclidean	**Mahalanobis**
(0.9/0.1)	98.4%	95.2%
(0.8/0.2)	98.4%	96.8%
(0.7/0.3)	96.8%	100%
(0.6/0.4)	96.8%	98.4%
(0.5/0.5)	95.2%	98.4%
(0.4/0.6)	91.9%	93.6%
(0.3/0.7)	91.9%	91.9%
(0.2/0.8)	90.3%	85.5%
(0.1/0.9)	85.5%	79%

Table 5.10: Combined Face/Ear Normalized Weighted Sum Recognition Rates.

In Figure 5.3, we see a graph of the different recognition rates achieved using normalized sum of face/ear with many different face/ear weights. The data is presented for both the Euclidean and Mahalanobis distance. From the graph we can see how for the Mahalanobis distance the line peaks at face/ear weights of 0.7/0.3 respectively, and then declines after that. From the graph, we can see that all points on the line above the point marked by the 0.5/0.5 point show an improvement in the recognition rate.

Figure 5.4 presents the data in Table 5.10 in a bar graph format. This graph clearly indicates the levels achieved by the Euclidean and Mahalanobis distances for

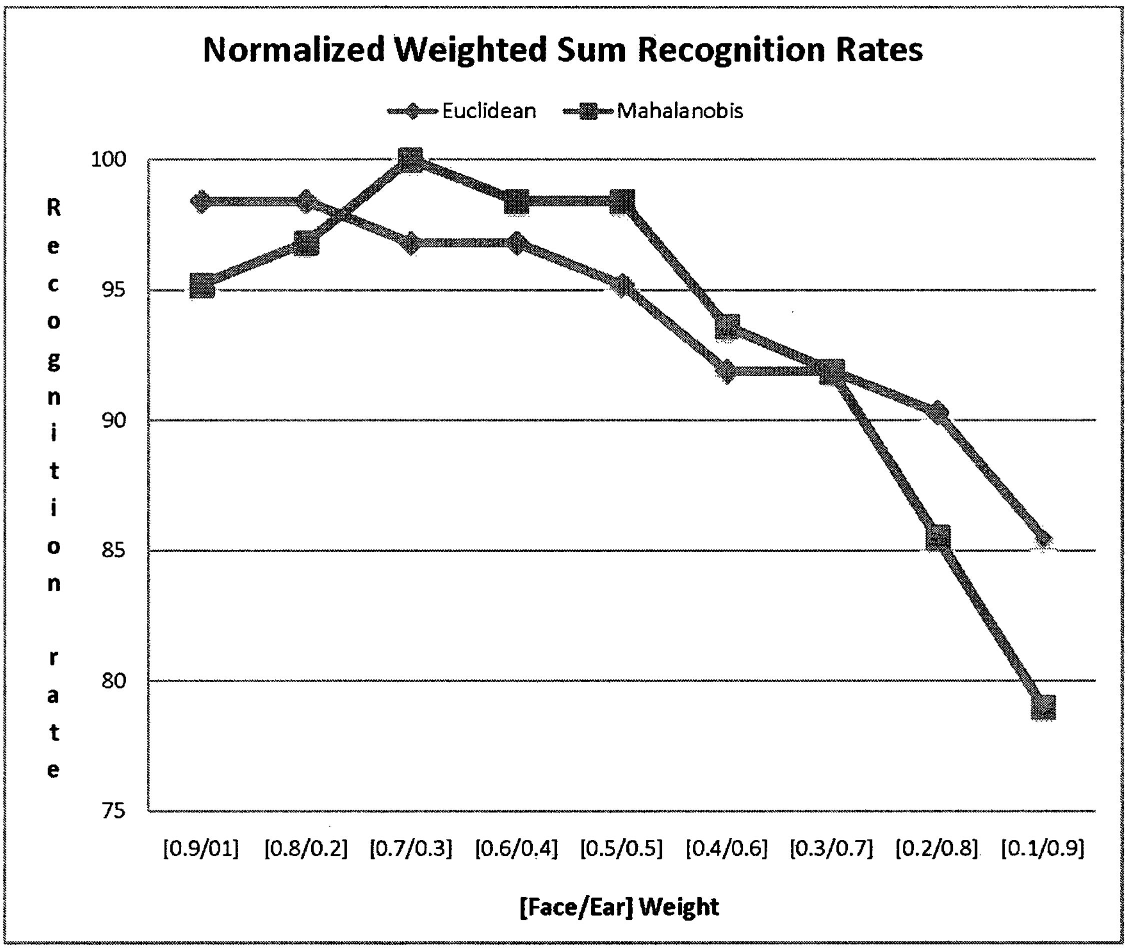


Figure 5.3: Graph of Recognition Rates for Different Face/Ear Weights Using Sum.

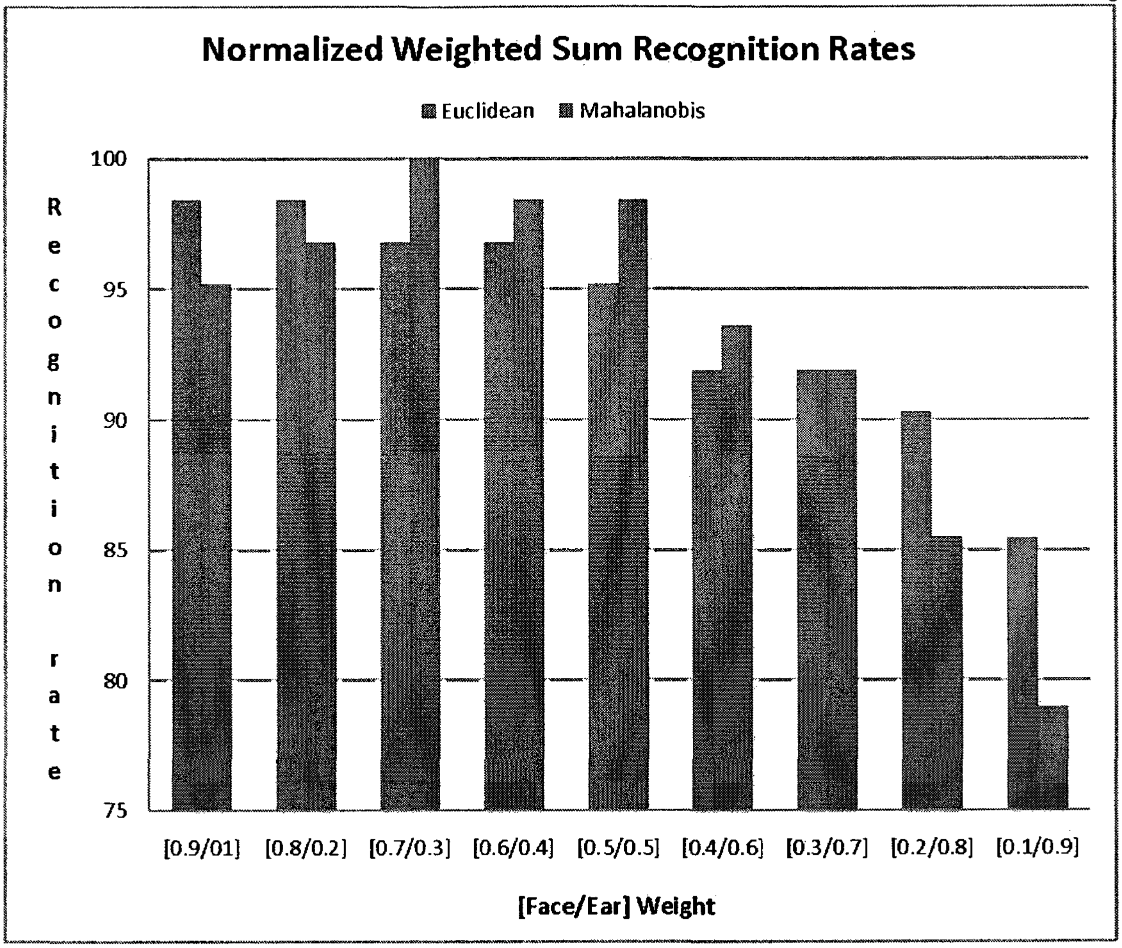


Figure 5.4: Graph of Recognition Rates for Different Face/Ear Weights Using Sum.

various face/ear weights in comparable fashion.

We will now re-visit the example of subject 00800 we first introduced in the previous section. Recall how the sum of the normalized distances incorrectly selected a candidate even though face had the correct candidate selected.

In Table 5.11 is presented the data for a weighted normalized Mahalanobis distance with weights of 0.7 for face and 0.3 for ear for subject 00800. You can see from the data that in unimodal mode, face correctly selects candidate 00800, but ear incorrectly selects candidate 00792.

With a weighted sum however, the algorithm is capable of selecting the correct candidate. The weighted sum for the lowest distance is 0.09837 which is for the candidate 00800. The second lowest distance is the candidate that previously gave us trouble 00751, its weighted sum is 0.13774, still however not close to the distance for the best candidate.

Candidate	Normalized Mahalanobis Distance				
	Non-Weighted		Weighted(Face 0.7/Ear 0.3)		
	Face	Ear	Face	Ear	Sum
00715	0.17057	0.56654	0.11940	0.16996	0.28936
00722	0.83388	1	0.58372	0.30000	0.88372
00751	0.16967	0.06323	0.11877	0.01897	0.13774
00792	0.52842	0	0.36989	0	0.36989
00793	0.39892	0.79484	0.27924	0.23845	0.51770
00794	1	0.29513	0.70000	0.08854	0.78854
00797	0.28811	0.68920	0.20168	0.20676	0.40843
00800	0	0.32791	0	0.09837	**0.09837**
00803	0.52033	0.60033	0.36423	0.18010	0.54433
00813	0.55940	0.42060	0.39158	0.12618	0.51776

Table 5.11: Sample Face/Ear Weighted Sum of Normalized Distances. Bold values indicate best candidate from respective algorithm.

In this section we presented the fusion technique which gave us the best recognition rates for our automated multimodal biometric system. We demonstrated many techniques such as using different distance measurements, normalization and weights to achieve better and better results. Also, explaining the details behind these techniques using many examples.

In the next section we will briefly look at an interval technique for fusion, and see what it can accomplish in multimodal biometrics.

5.5.5 Interval

In our experiments we also attempted to use a distance measurement between the first and second best match assuming that a greater distance between the first and second would indicate a greater reliability, we called this the Interval-Euclidean and the Interval-Mahalanobis distances, see Table 5.12.

Distance Type	Recognition Rate
Euclidean	95.26%
Mahalanobis	96.8%

Table 5.12: Combined Face/Ear Normalized Interval Recognition Rates.

The thinking behind this is that if there is a greater distance between the first and second best matches then it is an indication that the selection of the first is a surer thing or more reliable selection. As opposed to the first and second distances being very close, where this might indicate the selection is not so sure and was a close call.

The results using this technique with both a Euclidean and Mahalanobis distance can be seen in Table 5.12. We can see that the results are much in line with the recognition rates obtained using a normalized distance, with a slight decrease in recognition for the Mahalanobis distance.

It will be interesting to see if these results can hold up using a weighted technique, which we will explore in the next section.

5.5.6 Weighted Interval

In this section we will experiment with using weights on the interval recognition algorithm. We basically ran experiments with the same weights we did for the normalized weighted sum. The results can be seen in Table 5.13. As can be seen from the resultant recognition rates there is very little if any improvement in using weights on an interval fusion technique.

In Figure 5.5, we see a graph of the different recognition rates achieved using weights on an interval based fusion system. From this visual representation you can see that there really is no improvement before or beyond the 0.5/0.5 mark, indicating that weights have no beneficial affect on recognition rates in an interval fusion system

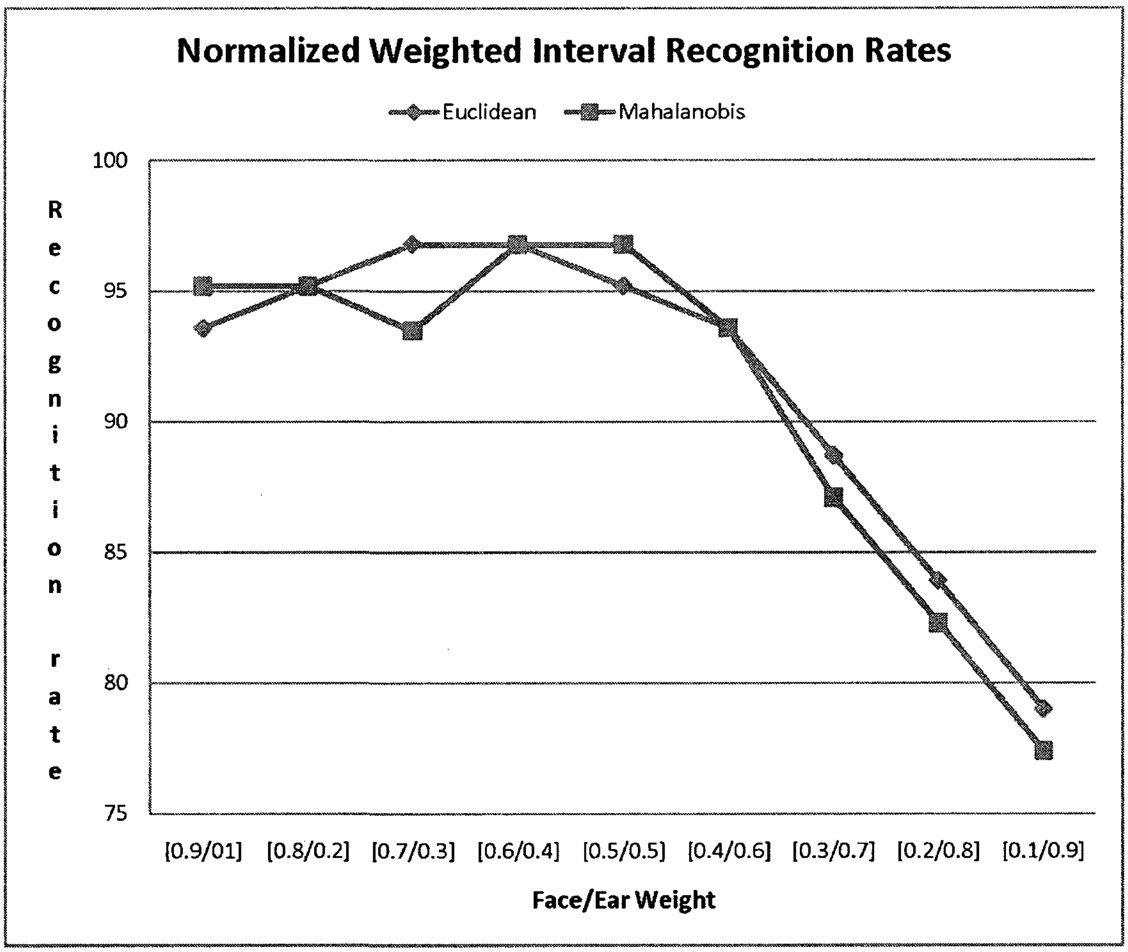


Figure 5.5: Graph of Recognition Rates for Different Face/Ear Weights Using Interval.

Weight(Face/Ear)	Distance Type	
	Euclidean	Mahalanobis
(0.9/0.1)	93.6%	95.2%
(0.8/0.2)	95.2%	95.2%
(0.7/0.3)	96.8%	93.5%
(0.6/0.4)	96.8%	96.8%
(0.5/0.5)	95.2%	96.8%
(0.4/0.6)	93.6%	93.6%
(0.3/0.7)	88.7%	87.1%
(0.2/0.8)	83.9%	82.3%
(0.1/0.9)	79.0%	77.4%

Table 5.13: Combined Face/Ear Normalized Weighted Interval Recognition Rates.

for combining face and ear in a multibiometrics system.

Figure 5.6 presents the data in Table 5.5 in a bar graph format. This graph clearly indicates the levels achieved by the Euclidean and Mahalanobis distances for various face/ear weights in comparable fashion.

5.5.7 Experiments with CVL Database

We decided to test our system on another face database in later experiments using the CVL Face database[32], which consists of a dataset of 114 people with 7 images per person at various angles. Using this database, we were able to achieve very similar results to our original experiments using a subset of FERET[35].

Using an Euclidean distance, our best recognition rate was 99.2% with a normalized weighted sum for face/ear of 0.8/0.2 respectively[26].

Figure 5.14 presents results for a combined face/ear normalized weighted sum recognition rates for the CVL database.

With the CVL Database, and using a Mahalanobis distance we achieved a best result of 100% with a normalized weighted sum of (0.8 to 0.7)/(0.2 to 0.3) for face/ear, respectively[27], which matched our experiments using FERET.

Figure 5.15 presents results for a combined face/ear normalized weighted interval recognition rates for the CVL database.

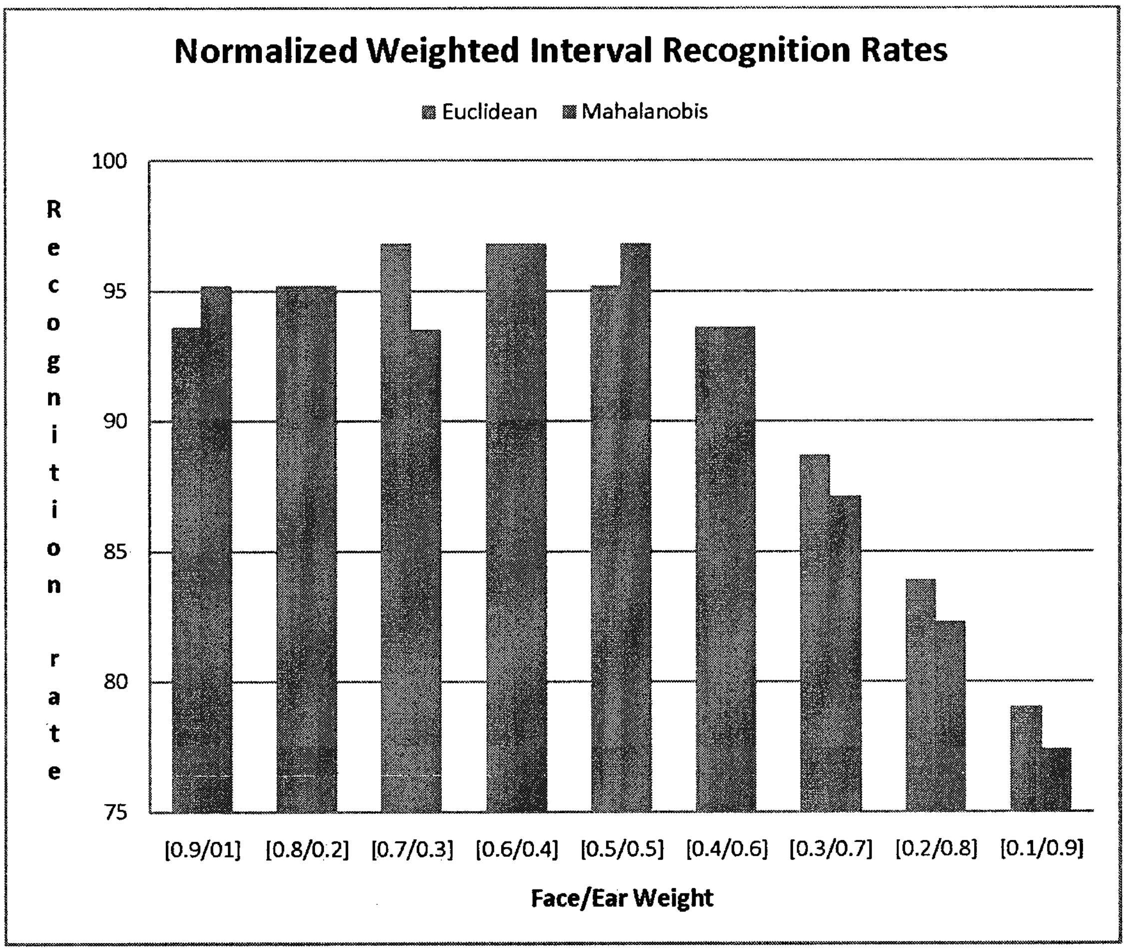


Figure 5.6: Graph of Recognition Rates for Different Face/Ear Weights Using Interval.

Weight(Face/Ear)	Distance Type	
	Euclidean	Mahalanobis
(0.9/0.1)	98.9%	97.1%
(0.8/0.2)	99.2%	100.0%
(0.7/0.3)	97.6%	100.0%
(0.6/0.4)	97.1%	99.6%
(0.5/0.5)	96.1%	99.6%
(0.4/0.6)	93.8%	95.2%
(0.3/0.7)	92.2%	93.9%
(0.2/0.8)	91.3%	87.4%
(0.1/0.9)	87.1%	81.2%

Table 5.14: Combined face/ear normalized weighted sum recognition rates for CVL database.

Weight(Face/Ear)	Distance Type	
	Euclidean	Mahalanobis
(0.9/0.1)	94.8%	96.7%
(0.8/0.2)	96.3%	96.9%
(0.7/0.3)	97.5%	94.3%
(0.6/0.4)	97.2%	97.2%
(0.5/0.5)	95.6%	97.2%
(0.4/0.6)	94.2%	95.2%
(0.3/0.7)	91.1%	90.8%
(0.2/0.8)	85.5%	84.4%
(0.1/0.9)	80.6%	79.2%

Table 5.15: Combined face/ear normalized weighted interval recognition rates for CVL database.

5.6 Conclusions

In this section, we presented a lot of data on the experiments we performed for the research leading to this thesis. We used many different techniques in an attempt to fully explore and contribute to multimodal research.

From the experimental data it is clear that a multimodal system is capable of augmenting the results achieved with a unimodal recognition using the same modes. The results from the experimental data clearly indicate that combining ear and face biometrics improves the recognition rate over using one of the modes on its own.

Both face and ear present advantages and disadvantages that combining them in a multimodal biometric system seems to take advantage of both their strong points. Another advantage of the face and ear combination is that both can be used as non-invasive biometrics.

We can see why interest in researching recognition techniques for person recognition is increasing using face and ear. With face and ear as multibiometrics we also get increased coverage.

We tried to present full details on all our experiments and on how they were conducted. Also presented, were many examples on detailed calculation and how these calculations were made. These details illustrate some of the obstacles that must be overcome in order to achieve even better recognition rates in a biometric system.

The results from the experiments presented here indicate that the best recognition rates are achieved using a normalized weighted sum. We do suspect however that the optimum weights will change depending on the rates achieved with those modes in unimodal mode, and on what modes are used as part of the multimodal biometric system.

For our experiments however, a normalized weighted sum of 0.7 for face and 0.3 for ear achieved the best results.

Chapter 6

Conclusions

In this thesis, we presented a framework for the automatic detection and recognition of individuals using face and ear as input modes. We were motivated to undertake this research due to the great demand for such a system by security services, agencies, investigative services etc.

Towards this goal, we developed an automatic face detector and an automatic ear detector to use in our system. Our detectors worked very well and we overcame certain obstacles by using some innovative techniques as explained in the thesis.

Our system uses a multimodal approach to improve the recognition rates, and used face and ear as input modes. By combining these two biometrics we were able to significantly improve the recognition rates as was shown in the experimental section, over the same metrics used in an individual unimodal basis.

One novel achievement of our system, is that the multimodal biometric system system presented in this thesis is totally automated, requiring no manual intervention. This type of automated biometric recognition system can be easily used in installations requiring person identification.

The research presented in this thesis, displays the possibilities for a system of detection and recognition of individuals without their full cooperation. These type of biometric systems are still not as robust as systems such as fingerprint and iris, and much work is still to be done to create such a system which is as robust as the ones mentioned.

However, such a robust system not requiring person cooperation would be invaluable to applicable areas already mentioned.

6.1 Summary

Interest in multimodal biometrics is clearly increasing amongst researchers, face and ear seems to be a natural fusion between two biometrics. Combining face and ear biometrics can improve the recognition rate of humans compared to unimodal face and ear biometrics as can be seen from the experiments in this thesis and in research carried out by other researchers[48, 23, 2].

Face is more intuitive but difficult because it changes with expression and also overtime. To compensate for this ear is smaller than face and seems to retain its properties through different expressions and overtime. Another advantage of face and ear multimodal system is that they can compensate for one another, such as if one is occluded(e.g. hair, sunglasses, hat etc.) you can still achieve a biometric by using the other.

In this thesis, we looked at many fusion methods and the different results achieved, we also compared fusion methods amongst themselves. Among all of the different fusion methods experimented a normalized Mahalanobis sum weighted (0.7/0.3) distance achieved the best result at 100%.

In Figure 6.1, we see a overall graph of the recognition rates for different face/ear weights using sum and interval. The line graph clearly displays the rates achieved using both Euclidean and Mahalanobis distances, you can see from this line graph the achievements of each fusion type.

It is worth noting that this system is totally automated, no manual intervention was done including at the preprocessing stage. For biometric systems to be useful this is of crucial importance. There is still much research to be done in this area, such

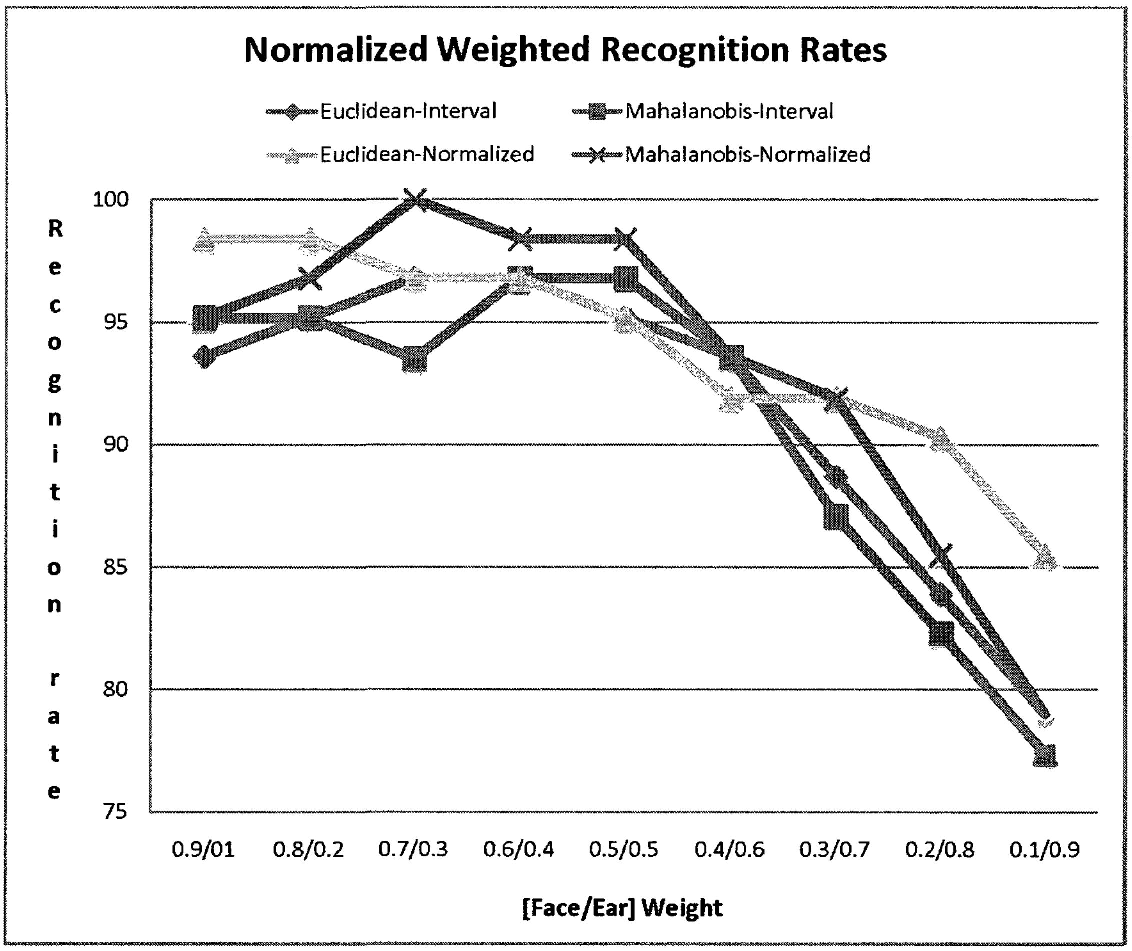


Figure 6.1: Graph of Recognition Rates for Different Face/Ear Weights Using Sum and Interval.

as more effective preprocessing, so that these results can be achieved over a wider spectrum.

In our research, we focused on person recognition using a multimodal biometric system where little or no person cooperation will be required. The system is fast and automated, and as experiments show achieves high recognition rates. There is however much research still to be done in testing other types of inputs, preprocessing and classification methods, so as to achieve these results in all environments.

6.2 Future Work

Our work and research was motivated by the need for a person identification system which is non-invasive and accurate. We also tried to present and research this need with the idea that the system would be totally automated. While we achieved some excellent results from our research and our experiments, from these also came new ideas and areas which could be explored in some future research and work.

Pre-processing is crucial for any recognition algorithm and system. We noticed dramatic changes in the recognition rate using different methods of pre-processing. Specifically in cropping an image before it is run through a recognition system, we feel that there is still much work to be done in this area. It would be interesting to explore new techniques of cropping that would lead to the optimal recognition rates.

In our case, we used face and ear as biometrics, but perhaps there are other metrics that can be combined that will lead to a more robust system. It would be interesting to see what kind of results could be achieved in a system such as this with other metrics.

Also perhaps, the inclusion of a third metric could lead to very good recognition rates with a higher level of robustness. These are all areas in which future work is possible and very exciting.

In our research, we were able to prove that a multimodal biometric system is capable of better recognition rate than those same modes run as unimodal recognition. There is much research to be done to improve these rates under all conditions such as different illumination, gestures, ageing, occlusion, etc. How can we get a constant superior recognition rates.

6.3 Final Thoughts

The research carried out and described in this thesis was very interesting and exciting. We were able to develop an automated multibiometric system which achieved excellent recognition rates. The fusion of face and ear in s multimodal biometric system seems natural, and the experimental data in our research proved that this is possible and that it leads to superior recognition rates as opposed to using face and ear as unimodal biometrics.

There is much research to be done in the field of mulitbiometrics and we plan to continue to carry out research in this exciting and interesting field.

Bibliography

[1] P. Flynn A. Jain and A.A. Ross. *Handbook of Biometrics.* Springer, 2008.

[2] Andrea F. Abate, Michele Nappi, and Daniel Riccio. Face and ear: A bimodal identification system. *Image Analysis and Recognition*, 4142:297–304, 2006.

[3] F. R. Bach, M. I. Jordan, J. Marques, J. Min, and W. Worek. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

[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*, pages 711–720, 1997.

[5] Alphonse Bertillon. *La Photographie Judiciaire, avec un appendice sur la classification et l'identification anthropomtriques.* Gauthier-Villars et fils, 1890.

[6] Paul J. Besl and Ramesh C. Jain. Invariant surface characteristics for 3d object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 33:30–80, 1986.

[7] R. Beveridge, K. She, B. Draper, and G. Givens. Evaluation of face recognition algorithm (release version 5.0). *In URL: www.cs.colostate.edu/evalfacerec/index.html.*

[8] Kevin W. Bowyer, Kyong I. Chang, and Patrick J. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101:1–15, 2006.

[9] U.M. Bubeck and D Sanchez. Biometric authentication: Technology and evaluation. *Technical report, San Diego State University, USA*, 2003.

[10] K. Chang, K. Bowyer, and P. Flynn. Face recognition using 2d and 3d facial data. *In Workshop on Multimodal User Authentication*, pages 25–32, 2003.

[11] Michal Choras. Ear biometrics based on geometrical feature extraction. *ELCVIA. Electronic letters on computer vision and image analysis*, 05(3):84–95, 2005.

[12] C. K. Chui. *An Introduction to Wavelets.* Elsevier Publishers, 1992.

[13] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition).* Wiley-Interscience, November 2000.

[14] J. G. Daugman F. Hao and P. Zielinski. A fast search algorithm for a large fuzzy database. *IEEE Transactions on Information Forensics And Security*, 3(2):203–212, 2008.

[15] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Proceedings of the Second European Conference on Computational Learning Theory*, 904:23–37, 1995.

[16] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.

[17] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: Global versus component-based approach. *Proceedings of the IEEE International Conference on Computer Vision*, 2:688–694, 2001.

[18] J. Heo, M. Savvides, and B. V. K. Vijaya Kumar. Large scale face recognition with distance measure in support vector machine spaces. *Proceedings Of The International Conference On Enterprise Information Systems*, 2006.

[19] Arbab-Zavar B. Hurley, D. J. and M. S. Nixon. The ear as a biometric. *European Signal Processing Conference, EUSIPCO*, pages 25–29, 2007.

[20] Davide Hurley, Mark S. Nixon, and John Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98:491–512, 2005.

[21] Alfred V. Iannarelli. *Ear Identification.* Paramont Pub, 1989.

[22] S.M.S. Islam, M. Bennamoun, and R. Davies. Fast and fully automatic ear detection using cascaded adaboost. *Applications of Computer Vision, WACV*, pages 1–6, 2008.

[23] Chang K., Bowyer K. W., and Victor B. Sarkar S. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1160–1165, 2003.

[24] S. Lao, Y. Sumi, M. Kawade, and F. Tomita. 3d template matching for pose invariant face recognition using 3dfacial model built with isoluminance line based stereo vision. *Proceedings of the 15th International Conference on Pattern Recognition, ICPR 2000*, 2:911–916, 2000.

[25] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. *Proceedings of International Conference on Image Processing*, pages 900–903, 2002.

[26] L. Luciano and A. Krzyżak. Automated multimodal biometrics using face and ear. *The International Conference on Image Analysis and Recognition*, 2009, submitted.

[27] L. Luciano and A. Krzyżak. Multibiometrics system based on fusion of face and ear. submitted to. *International Conference on Image Analysis and Processing*, 2009, submitted.

[28] P.C. Mahalanobis. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55, 1936.

[29] C. Middendorf and K. W. Bowyer. Multibiometrics using face and ear. *Handbook of Biometrics*, pages 315–341, 2008.

[30] M.S.Bartlett, J.R. Movellan, and T.J Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[31] Opencv. http://sourceforge.net/projects/opencvlibrary/.

[32] P. Peer. CVL Face Database. http://www.lrv.fri.uni-lj.si/facedb.html.

[33] P. S. Penev and J. J. Atick. Local feature analysis: a general statistical theory for object representation. *Network: Comput. Neural Syst*, pages 477–500, 1996.

[34] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. *Evaluation report, National Institute of Standards and Technology*, 2003.

[35] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

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

[37] K. H. Pun and Y. S. Moon. Recent advances in ear biometrics. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 04:164–169, 2004.

[38] J. Qin and Z. S. He. A svm face recogntion method based on gabor-featured key points. *Proc. Fourth Int. Conf. Machine Learning and Cybernetics*, pages 5144–5149, 2005.

[39] M. M. Rahman and S. Ishikawa. Proposing a passive biometric system for robotic vision. *Proc. 10th International Symp. on Artificial Life and Robotics*, 2005.

[40] G.N. Rutty, A. Abbas, and D. Crossling. Could earprint identification be computerised? an illustrated proof of concept papercould earprint identification be computerised? *International Journal of Legal Medicine*, 119(6):335–343, 2005.

[41] Marios Savvides, Jingu Heo, and Sung Won Park. Face recognition. *Handbook of Biometrics*, pages 43–70, 2008.

[42] Robert E. Schapire, Yoav Freund, Peter Bartlett, and Wee S. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5):1651–1686, 1998.

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

[44] P. Viola and M. Jones. Rapid object detection using boosted cascade of simple features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 511–518, 2001.

[45] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[46] P. Yan and K. W. Bowyer. Icp-based approaches for 3d ear recognition. volume 5779, pages 282–291, 2005.

[47] P. Yan and K. W. Bowyer. Biometric recognition using three-dimensional ear shape. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(8):1–12, 2007.

[48] Li Yuan, Zhi-Chun Mu, and Xiao-Na Xu. Multimodal recognition based on face and ear. *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition*, 2(4):1203–1207, 2007.

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