USING EARS FOR HUMAN IDENTIFICATION

Mohamed Ibrahim Saleh

Thesis submitted to the faculty of
Virginia Polytechnic Institute and State University
In partial fulfillment of the requirements for the degree of

Master of Science
In
Computer Engineering

A. Lynn Abbott, Chairman
Chris L. Wyatt
Patrick Schaumont

May 7, 2007
Blacksburg, Virginia

Keywords: Biometric, Pattern Recognition, Ear, Ear Recognition, Face Recognition, Principal Components Analysis (PCA), Segmentation, Multimodal Biometrics, Eigen-Face, Eigen-Ears.

USING EARS FOR HUMAN IDENTIFICATION

Mohamed Ibrahim Saleh

ABSTRACT

Biometrics includes the study of automatic methods for distinguishing human beings based on physical or behavioral traits. The problem of finding good biometric features and recognition methods has been researched extensively in recent years. Our research considers the use of ears as a biometric for human recognition. Researchers have not considered this biometric as much as others, which include fingerprints, irises, and faces. This thesis presents a novel approach to recognize individuals based on their outer ear images through spatial segmentation. This approach to recognizing is also good for dealing with occlusions. The study will present several feature extraction techniques based on spatial segmentation of the ear image. The study will also present a method for classifier fusion. Principal components analysis (PCA) is used in this research for feature extraction and dimensionality reduction. For classification, nearest neighbor classifiers are used. The research also investigates the use of ear images as a supplement to face images in a multimodal biometric system. Our base eigen-ear experiment results in an 84% rank one recognition rate, and the segmentation method yielded improvements up to 94%. Face recognition by itself, using the same approach, gave a 63% rank one recognition rate, but when complimented with ear images in a multimodal system improved to 94% rank one recognition rate.

Acknowledgment

First and foremost, I would like to thank my advisor, Dr. A. Lynn Abbott, for his guidance, patience, and valuable input throughout my research work. He is a great teacher and greater human being.

I would also like to thank Dr. Atalla Hashad from Arab Academy for Science and Technology for his guidance and valuable discussions.

An acknowledgement must be made to all the people that made it possible for me to pursue a degree in a great university as Virginia Tech through the VT-MENA program, and on top of these people, Dr. Sedki Riad, Dr. Yasser Hanafy, Dr. Ayman Abdel Hamid, and the president of the Arab Academy for Science and Technology and Maritime Transport, Dr. Gamal Mokhtar.

I would like to thank Dr. Patrick J. Flynn and Dr. Kevin Bowyer, from the computer vision research laboratory at the University of Notre Dame for providing their public biometrics database used in this research.

I also would like to thank all my friends in Egypt and Blacksburg for their support and friendship. For my friends in Egypt, thank you very much for all your support and encouragement throughout the period of my study. For my friends in Blacksburg, it was an honor to meet such great people, and I would like to thank you for all your help.

Last but not least, I would like to thank my family for all their support, specially my father who gave me the courage and support to pursue my dreams, and I hope I make him proud. Also thanks to Sherifa and Hadeel for their encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

Chapter 1
Introduction

1.1 Motivation

Biometrics is the study of methods for measuring physical or behavioral traits of an individual that can be used for uniquely recognizing or verifying that individual's identity [3], [13], [27]. Physical biometrics based on fingerprints, eye retinas and irises, facial patterns, and hand shapes are more common since they are easier to measure than behavioral biometrics such as handwriting, gait, and typing patterns.

Biometric methods are divided to physical and behavioral methods, which in turn can be divided into invasive and noninvasive methods. Invasive methods are those that require the cooperation of the individual in order to acquire data needed to compare his biometric features to the ones stored in a database. Noninvasive biometrics does not require the cooperation of the individuals; in fact data capture may be done without their knowledge. Applications for biometrics are most common in security, medical, and robotics areas related to fingerprint, face, iris, and gait. These biometric areas have gained the most attention among the research community.

The potential for using the ear's appearance as a means of personal identification was recognized and advocated as early as 1890 by the French criminologist Alphonse Bertillon [1]. Ear-based recognition is of particular interest because unlike face recognition it is not affected by factors such as mood or health. Significantly, the appearance of the auricle (outer ear) is nearly unaffected by aging [2], making it better suited for long-term identification when compared to other non-invasive techniques such as face recognition. Figure 1.1 shows the anatomy of the external ear, and figure 1.2 shows a sample of the ear images used in our research.

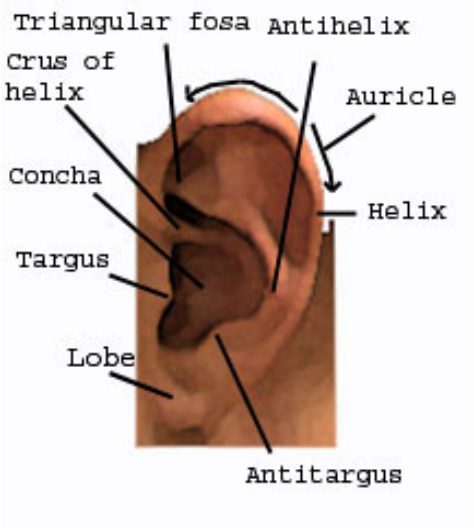

Figure 1.1 Anatomy of the external ear

Figure 1.2 Sample of images used in our research [7]

Fingerprint and iris recognition are considered to be invasive biometrics, which makes them unsuitable for some applications such as surveillance. Furthermore, fingerprint and iris based methods need good quality scanners or cameras so as to produce good results. On the other hand, face and gait recognition methods are noninvasive biometric techniques, and they do not need such high quality cameras as needed for fingerprint or iris biometrics. Ear-based biometrics is similar to face-based methods. Both are noninvasive techniques, and both can utilize a camera that is the same as the one used for face recognition.

Ear recognition has received considerably less attention than many alternative biometrics, including methods based on face, fingerprint, iris, and gait recognition. However, in recent years a few researchers have begun to consider the problem of developing a computational approach for ear recognition. Although the ears are often subjected to occlusions, our research will show that this can be overcome to some extent by the methods presented in this study.

1.2 Overview of Work

Automated human identification systems have become increasingly important in society. Human identification is involved in many applications such as gaining access to secure locations, or for surveillance. Our goal is to advocate the use of ears as a noninvasive biometric technique for human identification.

Our work is based on classifying segments of ear images individually and arriving at a unified decision to correctly identify the individual in question. We present a novel approach of spatial segmentation of 2D ear images, where each segment is classified independently and a final result is reached by a weighted classifier fusion system. We present different segmentation techniques aiming to enhance the recognition rate and finding a good way to deal with occlusions.

A multimodal approach is also investigated where face and ear images are combined to enhance the identification process of the individuals. Another approach is also presented by combining two segmentation methods together, and also combining face images with ear segments.

1.3 Previous Work

Previous research has suggested the use of ears as a biometric for human identification [1]. Researchers have advocated that the shape and appearance of the outer ear for humans is unique, and relatively unchanged throughout the lifetime of an individual. Although no one has proved that each person's ears are unique, studies in [12], [13] gave empirical supporting evidence. The most prominent work was done by Iannarelli [2], who examined over 10,000 ears and found them all to be distinguishable. Innarelli developed an anthropometric method where 12 measurements are used as features to distinguish individuals. Other work was that of Carreira-Perpiñán in 1995 [11], in which artificial neural networks (ANN) with linear nodes were used for feature extraction. Because of the linear nodes, the approach was closely related to singular value decomposition, and a decision rule was proposed based on exceeding a given threshold of reconstruction error.

More recently, Chang et al. [9] compared ear recognition with face recognition using a standard principal components analysis (PCA) technique known as the "eigen-face" approach on face and ear images acquired at the University of South Florida. They reported that recognition performance was not significantly different for the two modalities; with 71.6% and 70.5% accuracies obtained in the baseline experiment for ear and face recognition, respectively. Chang et al. also presented results for a lighting variation experiment where the reported rates were 64.9% for face and 68.5% for ear. In a multimodal experiment for combining ear and face images, they reported recognition results of a 90.9% rank one recognition, which is a significant gain in performance over their baseline experiment.

In another study, Hurley et al. [6] considered a "force field" feature extraction approach that is based on simulated potential energy fields. This force field transformation treats the image as an array of mutually attracting particles that act as the sources of a Gaussian force field. A potential energy field is extracted from the force field transformation which is used to discover potential energy wells. The experiments were performed on a database of ear images selected from the XM2VTS face database [10].

Using 63 individuals with 4 images per individual, they reported improved performance over PCA-based methods with 99% rank one recognition rate using template matching.

Choras [15] presented a study on ear biometrics using geometric feature extraction. The method is divided into image normalization, contour extraction (edge detection), calculation of the centroid, coordinate normalization, and 2 steps of geometrical feature extraction. The geometric features extracted are based on intersection points between circles of different radius with the calculated centroid as their center and the contours extracted from the ear image. The experiment was conducted on very high quality images with ideal conditions for recognition, and without illumination changes. With this ideal environment for recognition, an error free recognition was reported.

Moreno et al. [5] performed two experiments with neural network classifiers, where the features were extracted from ear images by performing edge detection for the first experiment and extracting seven known feature points of the outer ear to form the feature vector. For the second experiment a "morphology vector" is formed from intersection points between h horizontal cuts, v vertical cuts, and $2(h+v)$ diagonal cuts performed over the profile of an $h \times v$ size image. Another approach presented in [5] was recognition by using compression networks which was previously suggested for face recognition by Cottrell [14]. In Moreno's research, six images per individual were used; these were divided into three images for training, one image for validation, and two images for testing. For the classification technique based on feature points a 43% recognition rate was reported, and for the ear morphology approach with a neural network classifier a recognition rate of 83% was reported.

Other research has been done at the University of Notre Dame by Ping Yan et al. [26], where they reported 63.8% rank one recognition rate for an "eigen-ear", PCA-based approach. The same approach was performed on range images achieving 55.3% rank one recognition rate. They also performed Hausdorff matching of edge images achieving 67.5%, and an experiment for edge matching on 3D ear images which achieved 98.7% rank one recognition rate.

In our early investigative research [16] we performed experiments using different feature extraction methods and different classification techniques. The research was performed on a smaller data set [11] that comprised seventeen individuals with six images per individual. In this study, four classifiers utilized feed-forward artificial neural networks (ANN), and three classifiers that were based on a nearest neighbor rule. Image subdivision was first investigated in this study using the discrete cosine transform as a feature extraction technique, and feed-forward artificial neural networks for classification. The seven experiments performed in this study had rank one recognition rates ranging from 76.5% to 94.1%.

Although out of the scope of this research, but it should be mentioned that other research was done on 3D image recognition in [26],[4],[8],[24],[22],[23],[20],[21].

1.4 Contributions of this Research

The main contributions of the research are as follows:

- The thesis presents new insights and experimental results for the use of ears as a noninvasive biometric for human identification.
- The thesis presents a novel approach for spatial segmentation of ear images.
- The approach is well suited for dealing with occlusions.
- The thesis presents experimental results for using ear biometrics in a multimodal approach with face images.
- The thesis presents a novel approach of feature vector concatenation in the PCA space.

1.5 Organization of Thesis

The remainder of this thesis describes a biometric method for ear recognition that is robust and adaptive. In chapter 2 we present a description of the principal components analysis, and how we calculate the transformation matrices that are used to transform the images to the PCA space. A description of how we perform the dimensionality reduction is also presented in this chapter. Chapter 3 explains the extraction and normalization steps performed on ear and face images to prepare them to be introduced to our recognition system. Chapter 4 describes the main segmentation process and the main recognition process performed in the segmentation experiments. An eigen-ears experiment is presented in this chapter, and is followed by experiments presenting different segmentation schemes. Chapter 5 discusses the multimodal approach that we investigated for combining face and ear images, and for combining face images with ear segments. Chapter 6 is a general discussion of the work results, and chapter 7 presents the concluding remarks. Appendix I contains receiver operating curves and a comparison of the area under the curves for all the experiments in our research. Appendix II contains the cumulative match characteristics curves for all the experiments performed.

Chapter 2
Principal Components Analysis

2.1 Principal Component Analysis Calculations

In this section we will explain how the principal components are calculated and used as a transformation matrix to transform the training set images and test set images to the PCA space as a feature extraction technique which follows the seminal work of Turk and Pentland [17].

Each image's pixels are taken row by row from top to bottom and converted to a row vector containing the gray scale or intensity values of that image. These row vectors are then concatenated in a single matrix so that each row in that matrix corresponds to an image. This process is done to training images as well as test images, keeping them in two separate matrices.

The covariance matrix is then calculated for the training images where each row represents an image (observation) and each column represent a pixel position (variable). Covariance is the measure of how much two variables vary together. If two variables tend to vary together (that is, when one of them is above its expected value, then the other variable tends to be above *its* expected value too), then the covariance between the two variables will be positive. On the other hand, if when one of them is above its expected value, the other variable tends to be *below* its expected value, then the covariance between the two variables will be negative. The covariance is calculated using the following formula

$$\text{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \quad \text{For i and j = 1,2,....,} n$$

where E is the mathematical expectation and $\mu_i = Ex_i$, and x is the training images matrix. If the size of x is (m x n), where m is the number of images (rows) and n is the

number of pixels per image (columns), then the resulting covariance matrix will be of size (n x n).

Principal components analysis (PCA) is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. This is done on the covariance matrix (C) by satisfying the relation $Ce_i = \lambda_i e_i$ where e_i and λ_i, for $i = 1,2,...,n$ are the corresponding eigenvectors and eigenvalues respectively. Now we construct a matrix A from the eigenvectors sorted by decreasing eigenvalues.

2.2 Transformation to the PCA space

The resulting matrix A containing the sorted eigenvectors is used as a transformation matrix to transform the images to the PCA space. This is done by substituting in the following formula

$$y = A(p - m_x)^T$$

Where p is a vector representing one image from the training set and m_x is a vector of mean values of each pixel position of all training set images. Now the resulting vector y is the image p transformed to the PCA space. This is called the *Hotelling transform.* And also known as the *principal components transform*. This will be done to all training images transforming them to the PCA space. If an individual has two images in the training set then this means that this individual will be represented by two multidimensional points in the PCA space.

Using this transformation, we take a new test image t and apply the same transformation to it.

$$r = A \times (t - m_x)^T$$

The resulting vector *r* is the mapping of that image to the PCA space.

2.3 Dimensionality reduction

The eigenvectors matrix *A* has a size of (*n* x *n*), which means that we have *n* eigenvectors (where *n* is the number of pixels per image), and by transforming the images into the PCA space, we will have an *n* dimensional space. To reduce the dimensionality of that space we can take the top *k* eigenvectors corresponding to the highest *k* eigenvalues which will produce a transformation matrix A_k that we can use to transform images to the new space.

The value of *k* can be chosen to have an arbitrary number or can be chosen to represent a specific percentage of total variance that the eigenvectors correspond to. In our research we chose to reduce dimensionality approximately by half using the following formula to calculate the size of the transformation matrix

$$k = \lfloor number\ of\ \ pixels/2 \rfloor$$

We also exclude the first eigenvector corresponding to the greatest eigenvalue since it primarily reflects the difference in illumination between images [9], [19].

The size of the images in the database used in our research is 60 x 33, which is 1980 pixels per image. After the dimensionality reduction and transformation to the PCA space with the new reduced transformation matrix, each image will be represented by 989 pixels.

To find out the percentage of total variance that the taken eigenvectors represent we calculate it using the following formula:

$$m_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \times 100$$

The resulting vector *m* contains the percentages of each corresponding eigenvalue sorted by decreasing variances.

To find the percentage represented from participating eigenvectors in A_k we use the following formula:

$$TP = \sum_{j=2}^{k} m_j$$

The resulting value *TP* is the percentage of total variance represented in the transformation matrix A_k. This is the main process that is followed in all the experiments that uses the PCA transformation performed in our research.

Chapter 3
Normalization

3.1 Ear image extraction and normalization

The ear images for all individuals are extracted from left profile images. Ear images have to be extracted, and normalized with respect to scale and rotation. Hence, the normalization process is done on two stages. The first stage is geometric normalization, in which the images are normalized with respect to scale and rotation. This is done by selecting a two point landmark (figure 3.1) in every image and transforming the image with respect to those landmarks to a new location. This transformation will result in getting the landmark points in all images to have the same distance between them and that orientation of the line connecting the two landmark points will be vertical in the image. This transformation will normalize the images with respect to scale and rotation (figure 3.2). The landmark points are selected manually for each image. The landmark points that where selected are the Triangular Fossa and Incisure Intertragica in the ear [26]. These locations in the ear were selected because they prominently visible points in most ears.

The second stage is cropping around the ear to produce a new image that is tightly cropped to minimize the appearance of background information in the image (figure 3.3). All images are cropped to the same size. Conversion to gray scale is a preprocessing process that is done to all images and histogram equalization is performed on each image to compensate for lighting variations.

Figure 3.1 Profile image with Landmark points selected [7]

Figure 3.2 Image after scaling and rotation [7]

Figure 3.3 Ear image cropped [7]

3.2 Face image extraction and normalization

The extraction and normalization of face images is done according to the standard followed in the Eigen-face approach done on the FERET database. The normalization is also done on two stages. The first stage is the geometric normalization where all images are normalized with respect to scale and rotation. A two point landmark is manually selected in each image, and all images are transformed to a new location with respect to those points so that all landmark points will have the same distance between them, and the line connecting those points will have the same orientation. The landmark points selected are the centre of the eyes (figure 3.4). The orientation of the line connecting the two landmark points in this case after transformation will be vertical in the image (figure 3.5).

The second stage is cropping around the head to produce a new image that is tightly cropped to minimize the appearance of background information in the image. All images are cropped to the same size (figure 3.6). Conversion to gray scale is a preprocessing process that is done to all images and histogram equalization is performed on each image to compensate for lighting variations.

Figure 3.4 Face image with Landmark points selected [7]

Figure 3.5 Face image after rotation and scaling [7]

Figure 3.6 Face image cropped [7]

Chapter 4

Recognition Experiments

4.1 The datasets

The work done in this research was done using datasets provided by the University of Notre Dame [7]. The datasets were composed of face images and profile images. Images for individuals that we considered for our research were taken on different sessions, different days, and at different times of day. Figure 4.1 shows an example of images used in our research. The face dataset acquired had images for a lot more individuals than we used in our research but we where restricted to the number of good profile images that we can use. Ear images were extracted from profile images. Some of the images where excluded from the dataset that we used due to poor quality or movement distortions. Figure 4.2 shows some examples of excluded images. The final dataset that was used in this research was composed of 115 individuals with 4 images per individual.

Figure 4.1 Example of images from the database [7] used in our research group F

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

Figure 4.2 Sample for excluded images from the database [7] group F. (a), (b) and (c) are examples of images excluded due to motion blur, and (d) Images excluded for partial occlusion.

4.2 Main segmentation process

In this work, spatial segmentation is dividing the ear image into regions. Each segment is treated as a separate recognition problem. This means that if we decide on segmenting each image into two segments, we will have for each individual two segments representing each image to be recognized separately. To recognize that individual correctly each segment will have to be classified correctly to be belonging to that individual. In the following sections of this chapter we will present different segmentation methods with their corresponding recognition rates. We will also be presenting a method for combining the results of classification of individual segments to come to a unified decision about the classification of the image in question.

4.3 Performance Evaluation Methodology

The performance metrics that we are using and reporting our results upon are the ones used by National Institute of Standards and Technology (NIST) in performing the Face Recognition Vendor Test (FRVT) [19], [25]. We have also followed the definitions stated by the National Science and Technology Council's (NSTC) committee on Technology, Committee on Homeland and National Security, Subcommittee on Biometrics [18].

According to [19] the database of images is divided into subsets of training images, and test images or probes. Training set images are used to train the classifiers used to identify individuals based on certain features chosen for that purpose. Test set images consist of images for individuals for whom there exist images in the training set, and images for individuals for whom there are no images present in the training set. The later are called imposters.

In a verification scenario, an individuals biometrics are compared to his biometrics that are stored in a database and if a match is decided then the verification process is a success otherwise the individual is rejected, this is also known as a closed-set identification. The open-set identification or the watch list identification is a

generalization for both identification and verification. In the watch list problem a probe *p* is compared to a database of images, the probe ranks the database so we state the performance as an identification rate. A significant issue is that the system should not identify individuals that are no in the database (watch list), so we measure a false alarm rate.

To measure the recognition rate or true accept rate of the system the following formula is used

$$P_{RI}(t,r) = \frac{\left|\left\{p_j : rank(p_j) \le r, d_{ij} \le t, id(p_j) = id(g_i)\right\}\right|}{\left|P_g\right|} \quad \forall p_j \in P_g$$

where *p* is a probe image, *r* is a predefined rank, *d* is the similarity distance measure, and *t* is the threshold. The recognition rate is defined as the number of test images for which there are corresponding training images that have distances below the threshold for a predefined rank and are correctly matched with their corresponding individual, divided by the total number of images in the test set images that the system is trained to identify.

To measure false alarm rate for the system we use the following formula

$$P_{FA}(t) = \frac{\left|\left\{p_j : \max_i d_{ij} \le t\right\}\right|}{\left|P_N\right|} \quad \forall p_j \in P_N \ , \ \forall g_i \in G$$

where p is a probe image from the test set images that does not have a corresponding image in the training set, or as called in [19], [25] the set of imposters. The false alarm rate is calculated by comparing the set of imposters to the database and counting the occurrences where a match is found. This is decided if the similarity distance measure was found to be below the threshold for any image in training set. The number of occurrences that this happened is then divided by the total number of imposter images.

Since the watch list problem according to [19] and [25] is defined to be a generalization for the identification and verification problems, we chose to follow that definition in our research. And apply the above performance evaluation methodologies.

4.4 Main recognition process for segments

Our image database is composed of 115 individuals with 4 images per individual. These images are divided into training set images and test set images. Images for 15 individuals were put in the test set to construct the set of imposters, and the images for the other 100 individuals where divided into training and test sets as follows: The first two images per individual will construct the training set, and the last two images per individual will be part of the test set. Any chosen segmentation method will be applied on both the training and test set images. After segmentation, each segment is treated as an image representing the individuals. This means that if we are performing an experiment that involves segmentation of the image into four segments, instead of recognizing the whole image we will be recognizing each segment on its own. Principal components will be calculated for each segment separately and the segments will be transformed to the PCA space using their corresponding transformation matrix as discussed in chapter 2.

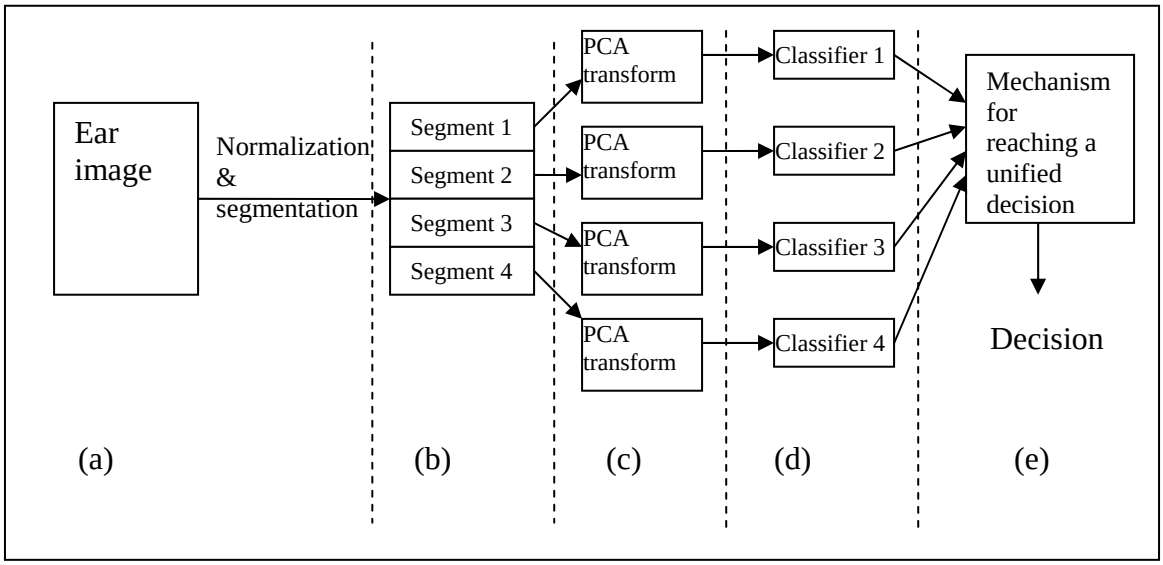

Figure 4.3 A general view of the recognition process for image segments. (a) The image is normalized and preprocessing operations performed. (b) Segmentation is done according to the experiment performed. (c) individual segments are transformed to the PCA space. (d) every segment is recognized with its corresponding classifier. (e) a mechanism is applied for reaching a unified decision.

For all individuals, with images participating in the training set used to calculate the principal components and obtain the transformation matrix, they are also used to train the nearest neighbor classifiers that use the Euclidian distance as a distance measure. The

third image belonging to those individuals that is taken from the test set will be used to test the classifiers and modify the weights for different segment classifiers. The rest of the images in the test set will be used to test the system for which the recognition rates are reported.

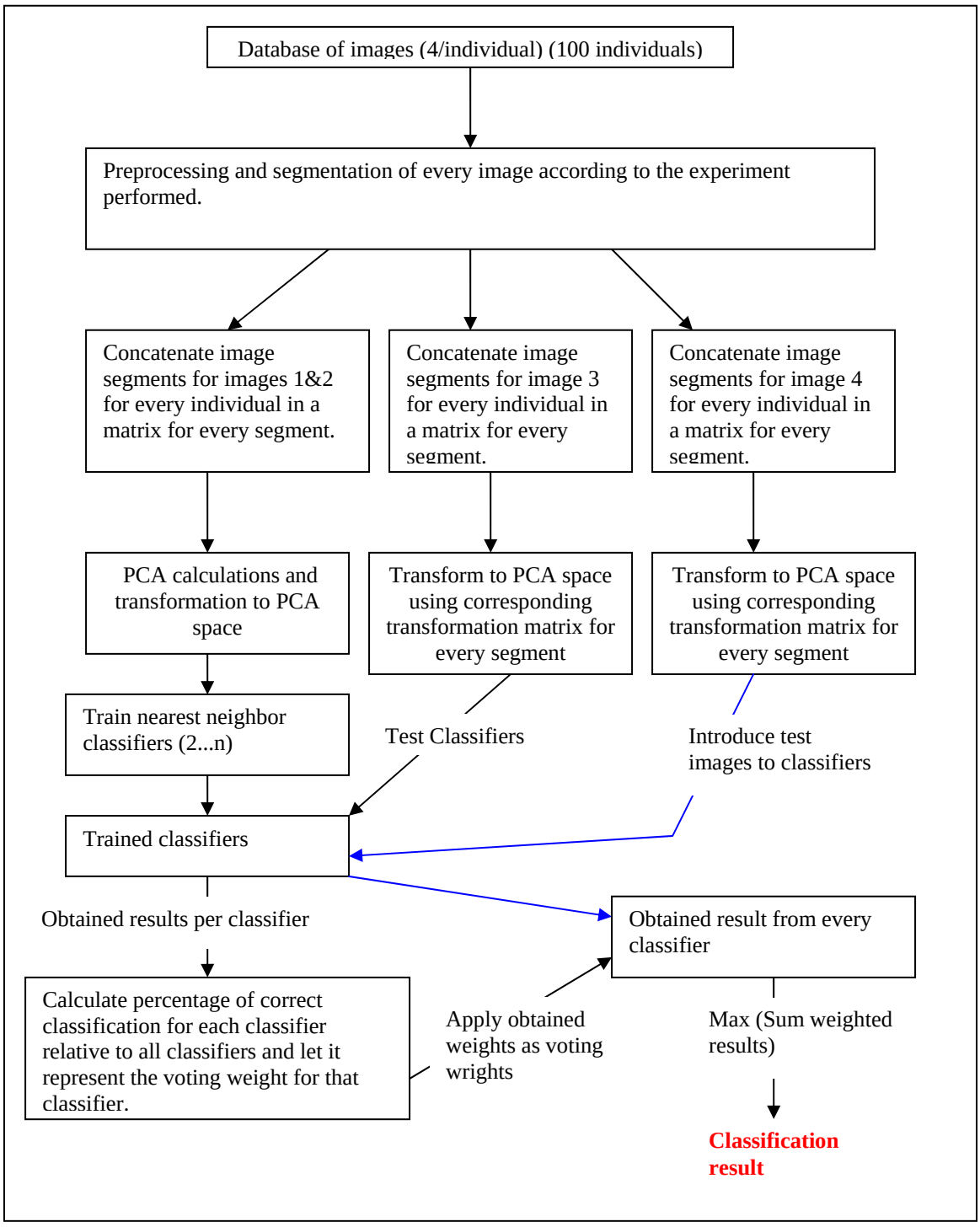

Figure 4.4 Main recognition process for segmentation experiments.

As it can be seen in figure 4.4 each classifier for every segment will be trained using the corresponding training segments of the training images transformed to the PCA space. The segments of the third image for all individuals will be used to test the classifiers. The resulting recognition rate of each classifier is used to set the weight of that classifier by substituting in the following formula

$$W_i = \frac{R_i}{\sum_{j=1}^{n} R_j}$$

where R is the correct recognition rate of a classifier resulting from testing with the third image's segments, and n is the number of classifiers used which is equivalent to the number of segments.

$$D_j = \sqrt{\sum_{i=1}^{m} \left(x_{k,i} - y_{j,i}\right)^2} \qquad (4.1)$$

The Euclidian distance measured from the test image to the rest of training set images is presented as equation 4.1, where D is the Euclidean distance, x is the test image, and y is an image from the training set. We use the subscript i to indicate an element in the feature vector, the subscript j to indicate an image from the training set, and the subscript k to indicate an image from the test set.

$$Decision = \min(D) \qquad (4.2)$$

When the fourth image for an individual is used to test the classifiers, each classifier will label that segment to belong to an individual according equation 4.2. Each classifier then will add its weight to a zero initialized accumulation vector of size m, where m is the number of individuals which is 100 in our case. The index of the maximum number in the accumulation vector is the result of that voting system. Switching roles between the third and fourth images from the test set has a negligible effect on the results, and in 90% of the experiments had no effect.

To further clarify the process, consider performing an experiment in which we are dividing the image to 3 segments. This means that we will have 3 classifiers, one for each segment. After the classifiers are trained using the first 2 images from the training data set, we test the classifiers with the third image for all individuals from the test set. Suppose that the resulting recognition rates for the 3 classifiers were 70%, 80%, and 90% respectively, then the calculated weights will be

$$W_1 = \frac{70}{70+80+60} = 0.3333$$
$$W_2 = \frac{80}{70+80+60} = 0.3810$$
$$W_3 = \frac{60}{70+80+60} = 0.2857$$

The fourth image is then used to test the classifiers. When an image is introduced to the system the classifiers will vote with their weighted result as illustrated in table 4.1

Table 4.1 classifier recognizes segment to belong to 3rd individual

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5
Classifier1	0.3333	0	0	0	0
Classifier2	0	0	0.3810	0	0
Classifier3	0	0.2857	0	0	0
Sum	0.3333	0.2857	0.3810	0	0

In the previous example in table 4.1, each classifier identified its corresponding image segment to belong to a different individual but the maximum from the accumulation vector gives a result that this image belongs to the 3rd class (individual).

In the example illustrated in table 4.2, classifiers 1 and 3 identified their corresponding image segments to belong to the 2nd class but classifier 2 identified its corresponding image segment to belong to the 4th class. Although the second classifier has the highest voting weight but the maximum from the accumulation vector (majority vote) gives a result that this image belongs to the 2nd class (individual).

Table 4.2 classifier recognizes segment to belong to 2nd individual

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5
Classifier1	0	0.3333	0	0	0
Classifier2	0	0	0	0.3810	0
Classifier3	0	0.2857	0	0	0
Sum	0	0.619	0	0.3810	0

4.5 Base experiment (Eigen-ears)

This experiment followed the procedure performed on the FERET [19] database for face recognition. Training set images are used for the PCA calculations and the formation of the transformation matrix which is used to transform the images from the training set and test set to the PCA space as explained in chapter 2. A nearest neighbor classifier is trained using the training set images, and the test set images are then introduced to the classifier for classification. The percentage of the total variance used in the transformation matrix is 79.048%. This experiment gave an 84.5% rank one recognition rate. Figure 4.5 shows the confusion matrix that resulted from this experiment, where the correctly identified images are represented on the diagonal and the off diagonal points represent the wrongly classified images.

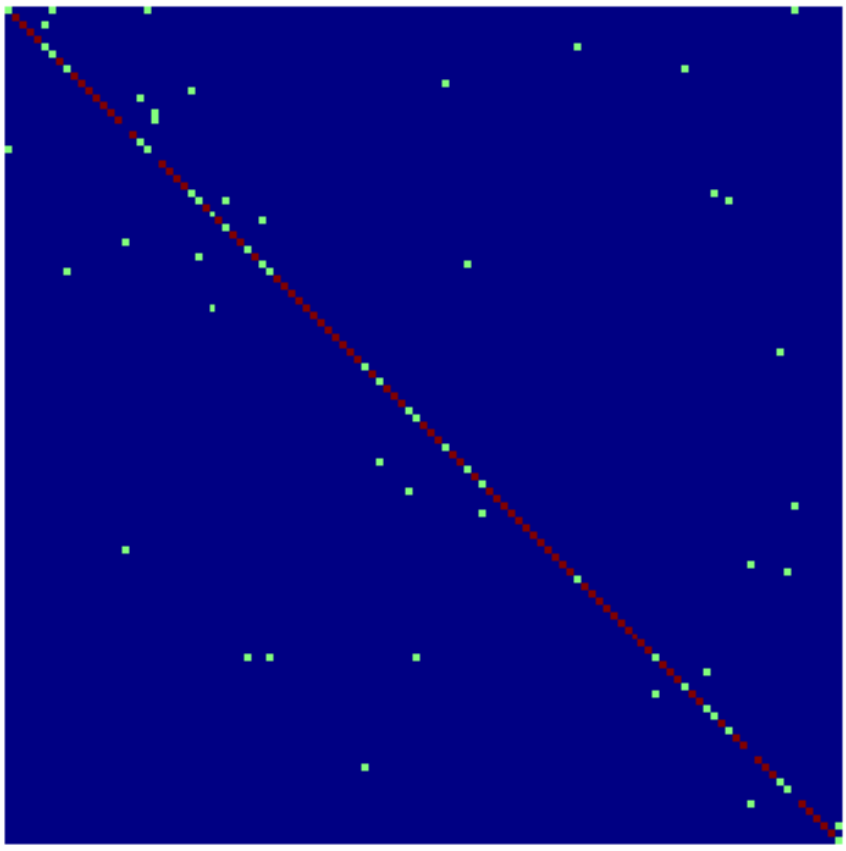

Figure 4.5 The confusion matrix for eigen-ear experiment. The columns represent the individuals' classes and the rows represent the classification of test images. Red pixels symbolize the correct classification of the 2 test images for the corresponding individual, and green pixels on the diagonal represent the correct classification of 1 image for the corresponding individual, while the green pixels off the diagonal represent wrong classification.

4.6 Uniform Horizontal Segmentation

Uniform segmentation means that the segments are approximately equal in size. But a segment might be smaller than other segments due to the division process since the smallest part of on image is the pixel, and we can not divide the pixel so if the division results in fraction it will be rounded. From the title of this section the following experiments will test the feasibility of segmenting the ear image to uniform horizontal segments.

4.6.1 Four Segments

In this experiment each ear image is divided into 4 uniform horizontal segments, as shown in figure 4.6.

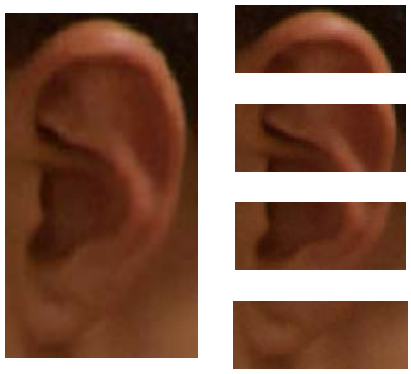

Figure 4.6 Ear image [7] segmented into 4 uniform horizontal segments

The number of rows in each of the top three segments is $n = \lceil R/4 \rceil$ where R is the number of rows of the ear image. For the last segment the number of rows is $n_{Last} = R - 3n$.

The recognition process described in section 4.4 is followed using four classifiers. The percentages of the total variance used in the transformation matrices for segments one to four are 74.2%, 79.5%, 72.1%, and 58.1% respectively. This experiment gave a rank one recognition rate of 86%.

4.6.2 Three Segments

In this experiment each ear image is divided into 3 uniform horizontal segments as shown in figure 4.7.

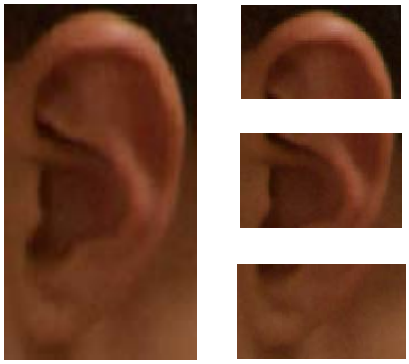

Figure 4.7 Ear image [7] segmented into 3 uniform horizontal segments

The number of rows in each of the top three segments is $n = \lceil R/3 \rceil$ where R is the number of rows of the ear image. For the last segment the number of rows is $n_{Last} = R - 2n$.

The recognition process described in section 4.4 is followed using three classifiers. The percentages of the total variance used in the transformation matrices for segments one to three are 76.1%, 75.7%, and 74.5% respectively. This experiment gave a rank one recognition rate of 88%.

4.6.3 Two Segments

In this experiment each ear image is divided into 2 uniform horizontal segments as shown in figure 4.8.

Figure 4.8 Ear image [7] segmented into 2 uniform horizontal segments

The number of rows for the top segment is $n = \lceil R/2 \rceil$ where R is the number of rows of the ear image. For the bottom segment the number of rows is $n_{Last} = R - n$.

The recognition process described in section 4.3 is followed using two classifiers. The percentages of the total variance for segments one and two are 79.1%, and 69.6% respectively. This experiment gave a rank one recognition rate of 79%.

4.7 Uniform Vertical Segmentation

In this section the following experiments tests the feasibility of segmenting the ear image to uniform vertical segments.

4.7.1 Four Segments

In this experiment each ear image will be divided into 4 uniform vertical segments as shown in figure 4.9.

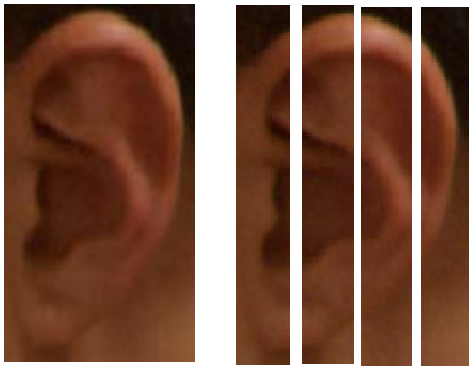

Figure 4.9 Ear Image [7] Segmented into 4 uniform vertical segments

The number of columns in each of the left three segments is $m = \lceil C/4 \rceil$ where C is the number of columns of the ear image. For the right most segment the number of columns is $m_{Last} = C - 3m$.

The recognition process described in section 4.4 is followed using four classifiers. The percentages of the total variance for segments one to four are 80.5%, 78.4%, 75.7%, and 61.1% respectively. This experiment gave a rank one recognition rate of 78%.

4.7.2 Three Segments

In this experiment each ear image is divided into 3 uniform vertical segments as shown in figure 4.10.

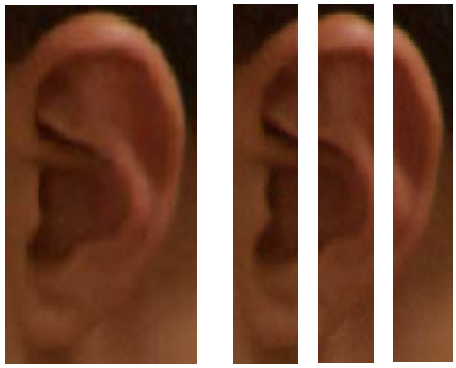

Figure 4.10 Ear Image [7] Segmented into 3 uniform vertical segments

The number of columns of the left two segments is $m = \lceil C/3 \rceil$ where C is the number of columns of the ear image. For the right most segment the number of columns is $m_{Last} = C - 2m$.

The recognition process described in section 4.4 is followed using three classifiers. The percentages of the total variance for segments one to three are 80.5%, 79.4%, and 69.1% respectively. This experiment gave a rank one recognition rate of 83%.

4.7.3 Two Segments

In this experiment each ear image is divided into 2 uniform vertical segments as shown in figure 4.11.

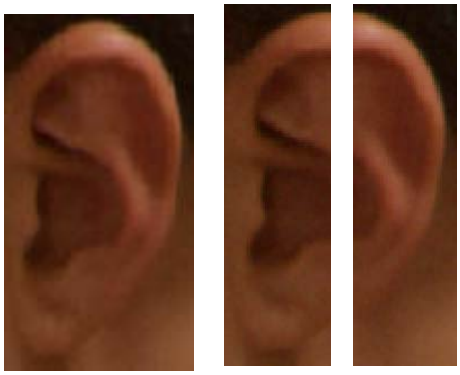

Figure 4.11 Ear Image [7] Segmented into 2 uniform vertical segments

The number of columns of the left segment is $m = \lceil C/2 \rceil$ where C is the number of columns of the ear image. For the right segment the number of columns is $m_{Last} = C - m$.

The recognition process described in section 4.4 is followed using two classifiers. The percentages of the total variance for segments one and two are 81.9%, and 73.1% respectively. This experiment gave a rank one recognition rate of 81%.

4.8 Uniform Grid Segmentation

Combining horizontal segmentation with vertical segmentation results in a grid like form of segmentation. In this section uniform means that the number of horizontal lines and vertical lines that form the grid are approximately equally spaced.

4.8.1 Four segments

In this experiment each ear image is divided into 4 grid segments as shown in figure 4.12. Here we are combining the horizontal and vertical segmentation methods by dividing the image into two horizontal segments and two vertical segments. The result is having 4 segments numbered as shown in figure 4.13.

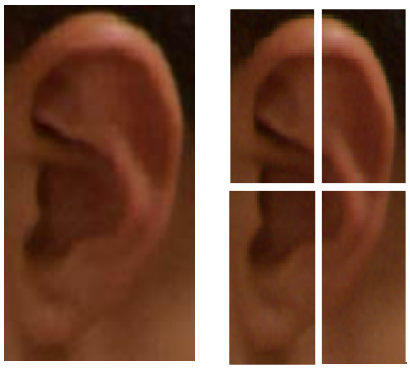

Figure 4.12 Ear Image [7] Segmented into 4 uniform grid segments

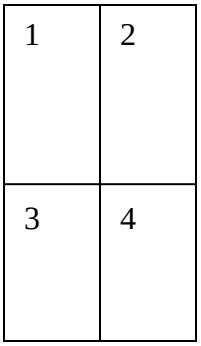

Figure 4.13 numbering of 4 segments

The number of columns and rows of each segment is $m = \lceil C/2 \rceil$, $n = \lceil R/2 \rceil$, where C is the number of columns, and R is the number of rows of the ear image. Segment one's size is (n x m), segment two's size will be (n x C-m), segment three's size will be (R-n x m), segment four's size will be (R-n x C-m).

The recognition process described in section 4.4 is followed using four classifiers. The percentages of the total variance for segments one to four are 77.6%, 71.3%, 77.1%, and 57.5% respectively. This experiment gave a rank one recognition rate of 86%.

4.8.2 Six Segments

In this experiment the each image is divided into 6 grid segments as shown in figure 4.14. Here we are combining the horizontal and vertical segmentation methods by dividing the image into two vertical segments and three horizontal segments. The result is having 6 segments numbered as shown in figure 4.15.

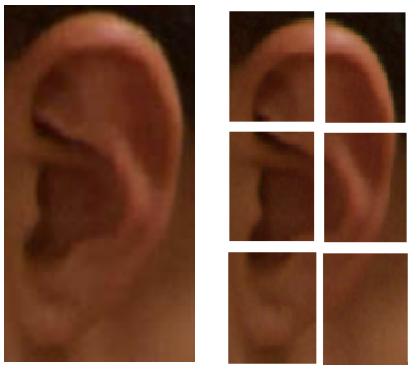

Figure 4.14 Ear Image [7] Segmented into 6 uniform of 6 grid segments

1	2
3	4
5	6

Figure 4.15 Numbering segments

The number of columns and rows of each segment is $m = \lceil C/2 \rceil$, $n = \lceil R/2 \rceil$, where C is the number of columns, and R is the number of rows of the ear image. Segment one and three's size will be (n x m), segment two and four's size will be (n x C-m), segment five's size will be (R-$2n$ x m), and segment six's size will be (R-$2n$ x C-m).

The recognition process described in section 4.4 is followed using six classifiers. The percentages of the total variance for segments one to six are 72.2%, 63.7%, 76.6%, 69.5%, 76.3% and 48.3% respectively. This experiment gave a rank one recognition rate of 87%.

4.9 Non Uniform Segmentation

In this section we present variations on the previous experiments by altering the size of the segments so that some segments are larger in size than the other segments. We also present non uniform topological segmentation by combining vertical and horizontal segmentation at different rates.

4.9.1 Three Horizontal Segments

In this experiment each ear image is divided into 3 non uniform horizontal segments where the second (middle) segment is larger than the other two as shown in figure 4.16.

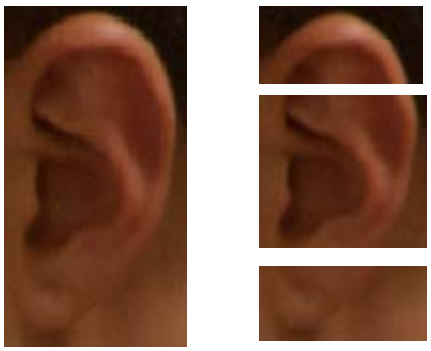

Figure 4.16 Ear Image [7] Segmented into 3 non uniform horizontal segments

The number of rows for the top segments is $n = \lceil R/4 \rceil$ where R is the number of rows of the ear image. For the middle segment the number of rows is $n_{middle} = R - 2n$, which makes the size of the middle segment ($2n$ x C), where C is the number of columns of the ear image. The number of rows for the last segment is $n_{Last} = R - 3n$. This segmentation method makes the first and last segments approximately equal in area while the middle segment is double their area.

The recognition process described in section 4.4 is followed using three classifiers. The percentages of the total variance for segments one to three are 73.7%, 77.6%, and 58.3% respectively. This experiment gave a rank one recognition rate of 91%.

4.9.2 Three Vertical Segments

In this experiment each ear image is divided into 3 non uniform vertical segments where the second (middle) segment is larger than the other two as shown in figure 4.17.

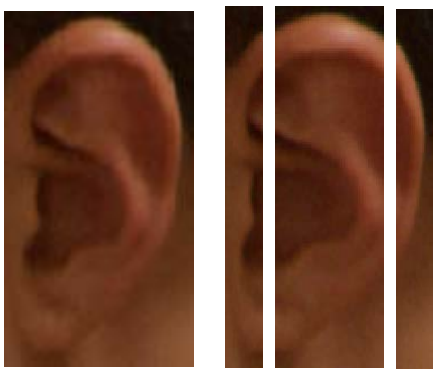

Figure 4.17 Ear Image [7] Segmented into 3 non uniform vertical segments

The number of columns for the left segments is $m = \lceil C/4 \rceil$ where C is the number of columns of the ear image. For the middle segment the number of columns is $m_{middle} = C - 2m$, which makes the size of the middle segment (R x $2m$), where R is the number of rows of the ear image. The number of columns for the last segment is $m_{Last} = C - 3m$. This segmentation method makes the first and last segments approximately equal in area while the middle segment is double their area.

The recognition process described in section 4.4 is followed using three classifiers. The percentages of the total variance for segments one to three are 80.5%, 80.7%, and 61.1% respectively. This experiment gave a rank one recognition rate of 83%.

4.9.3 Three Horizontal Segments and one Vertical

In this experiment each ear image is divided into 4 non uniform segments where we combine horizontal and vertical segmentation to form a new topological segmentation scheme as shown in figure 4.18.

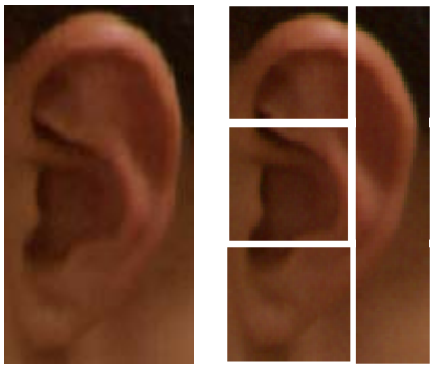

Figure 4.18 Ear Image [7] Segmented into 4 non uniform Segments

The size of the segments is the result of manipulating horizontal and vertical segmentation methods. The following formulas are used to come up with the segmentation scheme, $m = \lceil C/3 \rceil$, $n = \lceil R/3 \rceil$, where C is the number of columns, and R is the number of rows of the ear image. For the top two segments on the left their size will be (n x $2m$), for the bottom left segment, its size will be (R-$2n$ x $2m$). The right segment's size will be (R x C -$2m$).

The recognition process described in section 4.4 is followed using four classifiers. The percentages of the total variance for segments one to four are 77.2%, 77.9%, 76.8%, and 69.1% respectively. This experiment gave a rank one recognition rate of 88%.

4.9.4 Two Horizontal Segments and one Vertical

In this experiment each ear image is divided into 3 non uniform segments where we also combine horizontal and vertical segmentation to form a new topological segmentation scheme as in the previous experiment. This can be shown in figure 4.19.

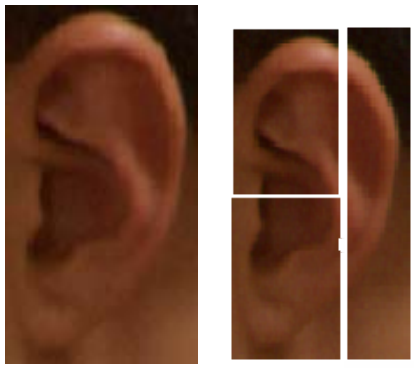

Figure 4.19 Ear Image [7] Segmented into 3 non uniform Segments

The size of the segments is the result of manipulating horizontal and vertical segmentation methods. The following formulas are used to come up with the segmentation scheme $m = \lceil C/3 \rceil$, $n = \lceil R/2 \rceil$, where C is the new number of columns and R is the new number of rows of the ear image. For the top left segment, its size will be (n x $2m$), for the bottom left segment, its size will be (R-n x $2m$). The right segment's size will be (R x R -$2m$).

The recognition process described in section 4.4 is followed using three classifiers. The percentages of the total variance for segments one to three are 81.1%, 77.9%, and 69.1% respectively. This experiment gave a rank one recognition rate of 82%.

4.9.5 One Circular Segment

This experiment is a result of observing the performance of individual segments from all previous experiments. In this experiment we only use one segment that we believe has the most important features to distinguish between individuals. We have chosen an ellipse shape for the segment as shown in figure 4.20.

Figure 4.20 Taking 1 circular segment from ear image [7]

Since the ear images are tightly cropped around the ear from the normalization step discussed in chapter 3, the circular ellipse that represents the segment has a diameter which is equal to the number of columns of the image. The center of the circle is the spatial center of the image which is obtained by $Center = \lceil R/2 \rceil$, where R is the number of rows of the ear image. The feature vector is the result of the PCA transformation of all the pixels inside the boundary of the circular ellipse.

The recognition process used here is the same process used for the base PCA experiment since we only have one segment. The PCA calculations and transformation matrix are done in the same manner as the base PCA. No weights were calculated since there is no classifier fusion done in this experiment. The percentage of the total variance used in the transformation matrix is 78.4%. This experiment gave a rank one recognition rate of 89%.

4.10 Segmentation by Threshold

This experiment is a non spatial segmentation experiment. We try a new method of segmentation in this experiment using the intensity values of pixels. The gray scale value of each pixel in the image varies from 0 to 255, so in our experiment here we segment the image into 4 segments by dividing the gray scale into 4 regions. The first segment will include all pixels that has an intensity value in the range [0…63], the second segment will include all pixels that has an intensity value in the range [64…127], the third segment will include all pixels that has an intensity value in the range [128…191], the last segment will include all pixels that has an intensity value in the range [192…255]. A bitmap of the regions of different segments is drawn on a new image that has the same size as the original image as shown in figure 4.21. Figure 4.22 shows the 4 segments collected in one image and color coded for clarification.

For classification, we use nearest neighbor classifiers that use the Euclidean distance measure. We use 4 classifiers, one for each segment. We do not perform PCA calculations in the experiment, but rather we use the bitmaps directly to train the classifiers. So the process described in section 4.4 is also used here but with a variation that skips the step of PCA calculation and transformation to the PCA space. The bitmap images are also transformed to row vectors by taking each row from the bitmap image from left to right and from top to bottom, concatenating them together for every bitmap to form the row vector. These vectors from the training set images are used to train the classifiers. The third and forth images as described in section 4.4 are used to set the classifier's weights and test the classification system. Since we are using bitmap images with one and zero values for pixels then the Euclidean distance is transformed to the Hamming distance. And the nearest neighbor classifiers now act as if we are performing template matching. This experiment gave a rank one recognition rate of 89%.

Figure 4.21 Original ear image [7] and the resulting 4 bitmap images

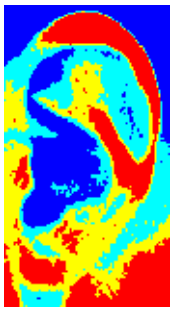

Figure 4.22 Segments collected and color coded

Chapter 5
Multimodal Biometrics

5.1 Face Recognition (Eigen-Face)

In this section we implement the face recognition method known as Eigen-face as performed in the FERET [19] experiment. This is the method that inspired the Eigen-ear experiment described in section 4.5. The image database we acquired from the University of Notre Dame had a lot of face images with more individuals than we have for the ear images, but for comparison reasons we will restrict the number of individuals to the same number of individuals we have for ear images which is 115 individuals and 4 images per individual.

We used two images for training and two for testing for each individual of the first 100 individuals keeping all four images of the last 15 individuals as an imposters set. The normalization explained in section 3.2 was performed on all images from the training set, the test set, and the imposters set. The PCA calculations and transformation described in chapter 2 was followed including the dimensionality reduction as performed in the eigen-ear experiment. A nearest neighbor classifier was trained using the training set images and images from the test set were introduced to the classifier for classification. The percentage of the total variance used in the transformation matrix was 84.1%. This experiment gave a 63.5% rank one recognition rate.

This recognition rate compared to the rank one recognition rate from the eigen-ear experiment which was 84.5% is significantly low. In the coming sections we present experiments to enhance the face recognition rate by combining face features with ear features.

5.2 Ear-Face Combination

In the following subsections we describe the experiments performed to combine face features with ear features using the principal components analysis method.

5.2.1 Combining images pre PCA

In this experiment we combine face images with corresponding ear images from both the face images database and ear images database to form a single image as shown in figure 5.1. This process is done on both the training set images and the test set images maintaining our standard throughout the research by having two images for training and two images for testing for each individual.

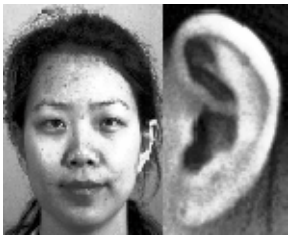

Figure 5.1 Face and ear images [7] combined to form one image

The combining process is done to all training images as well as the test images, without including any of the images used in the training set as a part of the images used in the test set for ear or face images. The resulting combined images will represent the 115 individuals with 4 new combined images per individual. PCA calculations described in chapter 2 were performed on the new images after the combination as performed by K. Chang et al. [3].

A nearest neighbor classifier that uses the Euclidian distance measure was trained with the training set images after transformation to the PCA space. Test set images transformed to the PCA space were then introduced to the classifier for identification. This experiment gave a rank one recognition rate of 91.5%.

This experiment was repeated but this time the face and ear images where concatenated as shown in figure 5.2 where the ear image was appended below the face image. All other steps taken were the same. The reason for doing this was to verify that the location of appended image does not matter especially with the method of transforming the images to vectors. This experiment gave a 91% rank one recognition rate.

Figure 5.2 Second Method for Combining face and ear images [7].

As it can be seen from the recognition rate, the performance was enhanced significantly by combining ear features and face features.

5.2.2 Combining images post PCA

In the previous experiment face and ear images were combined before calculating the PCA and transformation to the PCA space. This means that the transformation matrix used to transform images to the PCA space resulted from performing the principal components analysis on the combined ear-face image.

In this experiment the combination will be done in the PCA space. PCA calculations described in chapter 2 will be performed on face images and ear images

independently. This means that the transformation matrix for the face images will result from performing principal components analysis on the face images only. And the transformation matrix for the ear images will result from performing principal components analysis on the ear images only. This can be thought of as performing the Eigen-face experiment and eigen-ear experiment separately. After transforming the face images and ear images to the PCA space with their corresponding transformation matrices, they are combined. Face vectors from the training set are combined with their corresponding ear vectors from the training set, and face vectors from the test set are combined with their corresponding ear vectors from the test set.

A nearest neighbor classifier that uses the Euclidean distance measure is trained using the newly formed vectors from the training set letting every individual be represented by two multidimensional points in the PCA space. Combined vectors from the test set are then introduced to the classifier for identification. This experiment gave a rank one recognition rate of 92.5%.

This recognition rate is very close to the recognition rate obtained from the previous experiment where we combined the images before performing the principal components analysis. Since the face and ear images' size are the same for images used in both experiments, and we are using the same method for dimensionality reduction in both experiments, the resulting feature vector is also the same size in this experiment and the previous ones.

5.2.3 Combining images post PCA with permutation

From the previous two experiments we conclude that combining the images before performing the principal components analysis or combining them after performing the principal components analysis individually is the same. In this experiment the new combined images are formed after the transformation of the face and ear images to the

PCA space as the previous experiment. The difference in this experiment is that the way the face and ear images are combined together.

Since we have 4 images per individual for face and ear images, and we are using 2 for training and the other 2 for testing, we combined the face and ear images from both training sets by using all possible combinations without repetition. The first face image for an individual was combined with the first ear image (for the same individual) to form a new combined image. Then the first face image was also combined with the second ear image for that individual to form a second combined image. The same was done for the second face image from the training set. The same process was performed on the training set images. This method for combining the face and ear images yielded 4 combined images for training per individual and 4 combined images per individual for testing. We did not use any of the face or ear images that were used to produce the test images to produce any of the training images. Figure 5.3 illustrates the process of combining training images for an individual which yielded four combined images.

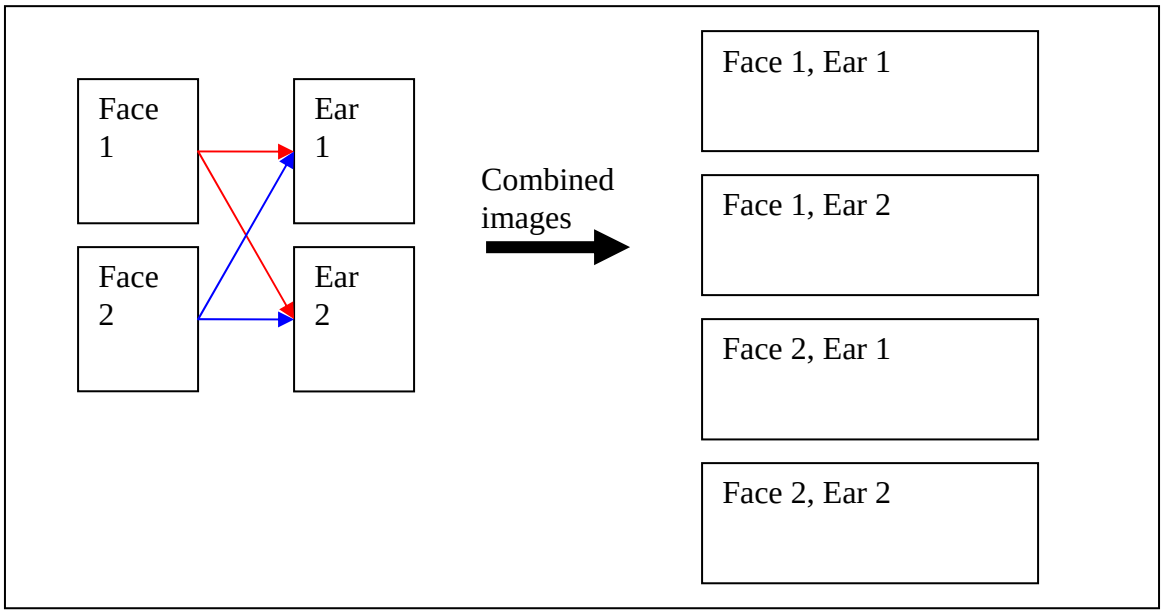

Figure 5.3 Combining face and ear images with permutation

The logic behind combining the images in that way is to cover all matching possibilities. If we consider the experiment of the eigen-faces alone or the eigen-ears

alone, it is possible that for a certain individual's test image to yield a correct classification as a result of being the closest neighbor of the first training image, or as a result of being the closest neighbor of the second training image. This goes for both the face and ear experiment. Now suppose that as in the second multimodal experiment each face image is coupled with its corresponding ear image. That is, the first face image for an individual is combined with the first ear image for that individual and the second face image with the second ear image and so on. Now the first half of the feature vector represents the face image and the second half represents the ear image. Suppose that we introduce a test image to the classifier and the first part of the feature vector is close to the first part of the first point for that individual but the second part is close to the second part of the second point for the same individual. The result is unexpected and can lead to wrong classification. By permuting the images we give the classifier a chance to find the best class for which the introduced image is closest to. This experiment gave a 92.5 % rank one recognition rate.

Another experiment was performed following the same procedure but with the permutation done for training set images only. The test set images were combined on a one to one basis, combining each face image from the test set with its corresponding ear image. This means that each individual was represented by four points in the multidimensional PCA space, and the testing was done by two combined images per individual. This experiment gave a 94% rank one recognition rate.

5.3 Combined Segmentation Methods

In this experiment we will present a combination of two segmentation methods explained previously with a variation. This is considered a new segmentation method but since we are doing the combination pre PCA and post PCA we saw fit to consider it a multimodal experiment. In the following sections we will present the combination of segments from the six segment uniform grid segmentation method and the one circular segment. These two methods were chosen because they are among the best experiments

achieving a high success rate, and the way we are combining them minimizes redundant information.

As it can be seen from the illustration in figure 5.4 the circular segment is combined with the four corner segments of the six segments grid method leaving out the two middle segments. Using the six segment grid method and not the four segment grid method allowed us to minimize redundant information by eliminating the middle part already represented in circular segment. Combining the segmentation methods, results in getting unusual segment shapes as illustrated in figure 5.4.

Since the circular segment's performance was the highest obtained among the segmentation methods by combining it with other segments it will enhance the individual recognition performance of each segment. This results in an overall improvement in the final reported recognition rate.

5.3.1 Combining Pre PCA

In this experiment we combined the segments from the six segment uniform grid segmentation method with the one circular segment method pre PCA calculations.

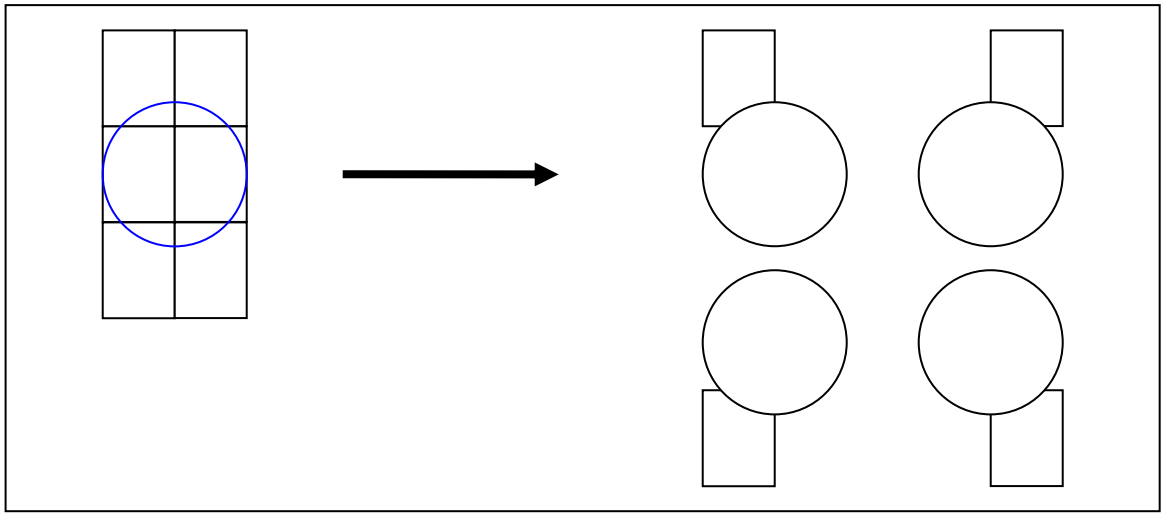

Figure 5.4 Illustration of combined segments from both segmentation methods

The combined segments shown in figure 5.4 are converted into row vectors and the procedure explained in chapter 2 is followed. The procedure described in section 4.2 was followed as in the other segmentation experiments done before. Since the combination resulted in four combined segments, we used four classifiers. As described in section 4.2 the third image is used for weights calculation, and as mentioned before, swapping the role of the third and fourth images does not affect the recognition rate. In this experiment it only affected the recognition rate with a negligible percent. The experiment gave a 93% rank one recognition rate.

5.3.2 Combining Post PCA

In this experiment the PCA calculations explained in chapter 2 are calculated for every segment from the chosen four segments and the circular segment individually. The segments are then concatenated in the PCA space for the training set and the test set images. The procedure in section 4.3 is followed using four classifiers. This experiment gave a 94% rank one recognition rate.

These experiments yielded a great improvement in the recognition rate from the base eigen-ears experiment by giving 94% rank one recognition rate instead of 84% from the eigen-ears experiment.

5.4 Combining Face images with Ear segments

In the following sections we present the results of the multimodal experiments of combining face images with ear segments. The segmentation methods chosen are some of the methods that yielded the best recognition rates.

5.4.1 Combining with one circular segment pre PCA

In this experiment we combined the face images with the one circular segmentation method of ear images. The face image is concatenated with the corresponding circular segment of the ear image to form a single row vector for each image. This is done to all images in the training and test sets. The procedure explained in chapter 2 for doing the PCA and dimensionality reduction is followed and the images are transformed to the PCA space. The recognition process explained in section 4.4 does not apply to this experiment since we only have one segment from the ear image. Instead the procedure followed for the eigen-ears experiment is followed.

This experiment gave an 85.5% rank one recognition rate. This rate is less than the recognition rate for the circular segment by itself but better than the base ear recognition experiment (eigen-ears), but it is a significant improvement over the face recognition experiment (eigen-face).

5.4.2 Combining with one circular segment post PCA

This experiment is the same as the previous experiment but the feature vectors are created in the PCA space. The face images and the circular ear segments are transformed to the PCA space separately as described in chapter 2 including the dimensionality reduction. The feature vectors are formed from the vectors in the PCA space that represent the face images and the circular segments representing the ear images. This is done for the training and test set images.

The recognition process is the same as the previous experiment, where a nearest neighbor classifier is trained using the feature vectors formed from the training set face and ear images and then tested using the feature vectors formed from the test set face and ear images. This experiment gave an 87% rank one recognition rate.

5.4.3 Combining with two segmentation methods

In this experiment we combined the face images with the four corner segments from the six grid segments method and the one circular segment method. We did this experiment as before by combining pre PCA calculation and combining post PCA calculations.

5.4.3.1 Combining with two segmentation methods Pre PCA

In section 5.2 we presented the procedure of combining two segmentation methods. These segmentation methods were the four corner segments from the six grid segments segmentation method and the one circular segment. This combination resulted in four new segments as shown in figure 5.4. After performing this combination the newly formed segments are converted to vectors and concatenated with the corresponding vectors representing the face images.

Every face image is concatenated with the corresponding newly formed four segments. The result is four combined segments representing each image in the database for training and test set images. The procedure performed previously and explained in chapter 2 for PCA calculation and transformation to the PCA space is done on the newly formed vectors.

The recognition process described in section 4.4 is followed using four nearest neighbor classifiers one for each combined face-segment. This experiment gave an 88% rank one recognition rate.

5.4.3.2 Combining with two segmentation methods Post PCA

To do the combination post PCA we will calculate the PCA separately for every segment involved in this experiment. The PCA calculations are done separately for every

segment of the four corner segments of the six segments grid method, the one circular segment, and the face images. These calculations include the dimensionality reduction and the conversion to the PCA space.

The representation of the four corner segments in the PCA space and the representation of the circular segment in the PCA space are concatenated together as performed in section 5.3.2. This process yields newly formed four vectors representing every image from the ear images database. The newly formed segments are combined with the corresponding representation of the face images in the PCA space, by combining the face image with every segment representing the corresponding ear image.

The recognition process described in section 4.4 is followed also using four nearest neighbor classifiers. This experiment gave an 89% rank one recognition rate.

Chapter 6
Discussion

In this chapter we will discuss the pros and cons of using ears as a biometric for identifying humans. We will also explain the logic behind combining the images post PCA calculations, and the weakness of facial recognition and the benefit of using a multimodal system for human identification.

6.1 What's wrong with using face recognition?

Face recognition has been researched a lot in the past years and a lot of algorithms, feature extraction techniques, and classification techniques have been developed for that purpose, but it all comes down to the efficiency of the feature extraction.

Facial features are susceptible to many factors such as mood, health, facial hair, and facial expressions. This is a natural barrier in using face as a reliable means for human identification. The feature extraction technique used will have to deal with the material at hand, so no matter how good the feature extraction process used is, the condition of the face presented will determine the outcome.

Figure 6.1 Facial hair affects feature extraction [7]

Figure 6.1 gives an example of the same person at different stages of hair and facial hair growth. Although us humans most often can still recognize each other under

these conditions but a computerized system will easily be fooled and produce unexpected or false results if subjected to cases like this one.

Figure 6.2 presents another example were facial expressions may affect the feature extraction for an individual. Although some feature extraction techniques may be resilient to a point to facial expression changes, but this is still considered an obstacle in the reliability of any face recognition system.

Figure 6.2 Facial expressions may affect the feature extraction [7]

Figure 6.3 presents profile images of two identical twins. The case of identical twins may very easily fool the best face recognition system, since it fools human beings most often. Although identical twins have similar ears but they are not identical, especially in the lobe, concha, and targo areas of the ear which are mostly located in the middle and lower part of the ear which is the most parts that do not suffer from occlusions as shown in figure 6.4 and figure 6.5.

Figure 6.3 Identical twins have similar ears but not identical [7]

Figure 6.4 Close up of twins ears

Figure 6.5 High pass filter applied to twins ear images to emphasize the difference

The face as imperfect for use as a biometric, there has been some successful systems for face recognition developed in the past few years. The face is also the most visible biometric and suitable as a non invasive biometrics.

6.2 Why use ears?

The use of ears as a biometric for human identification has not been researched as intensively as other biometrics has been researched. Although research in this area is relatively small, the research that has been done showed a lot of promise in using the ear as a biometric for human identification.

The ear much like the face is a visible part of the human body that can be used for a non invasive biometric technique. Humans most likely will have to keep their ears uncovered to be able to hear. The ears unlike the face are unaffected by ageing, in fact the ear undergoes very slight changes from infancy to adulthood, in fact the only change that happens is elongation due to gravity. The ears also do not suffer the change in appearance by hair growth like the face does.

Although these are all pros for using the ears as a biometric, but using the ears for human identification has some disadvantages. These disadvantages are embodied in occlusions. Sources of occlusion may be long hair, earrings and multiple piercings.

6.3 How to overcome the occlusions?

The recognition systems presented in our work give a partial solution to overcome these occlusions. This can be achieved by using the segmentation method and using a separate classifier for each segment. Using separate classifiers give a chance to the segments that are not occluded to give a correct classification and successfully identify the individual in question.

This argument is valid only if individual segments give a high enough recognition rate to be able to correctly identify an individual.

6.4 What is the performance of individual segments?

Table 6.1 presents some of the performance rates achieved for individual segments in different segmentation experiments performed in this research. As it can be seen from these figures the performance of the individual segments is highly acceptable, keeping in mind that they are only segments from the ear achieving these success rates without the assistant of information from other segments and with a nearest neighbor classifier.

Table 6.1 Performance of individual segments

Experiment	Segment number	Success Rate (%)
Segment 4p Horizontal	2	79.1
Segment 3p Unequal Horizontal	2	86.2
Segment 3p Unequal Vertical	2	85.2
Segment 2p Horizontal	2	80
Segment 2p Vertical	1	77.3
Segment 4p Mesh	3	71.3
Segment 4p (3H1V) Mesh	2	80.8
Segment 3p (2H1V) Mesh	2	80.8

It is also noticed from the segment numbers that achieved theses rates that they are the least segments of the ear susceptible to occlusions. This makes this technique very suited for the ear recognition problem. It is also notices that some segments on their own are achieving a higher recognition rate that the base eigen-ears experiment which more evidence of the success of this segmentation method.

6.5 How flexible and robust is the system?

The proposed recognition system in our research is very flexible because of the use of different classifiers for different segments. this gives the system the ability to be

able to get input from what ever segment that is available for identifying a particular individual. This fact makes the recognition system robust since it will not fail because of lack of input information.

The flexibility of the system also comes from the ability to combine any number of segments from any segmentation technique. The combinations are endless and the segments used can be of any size or shape.

6.6 What is the difference between combining images Pre and Post PCA?

Combining the segments (images) pre PCA will result in getting a new image for which we calculate the covariance matrix and the rest of the PCA calculations. The transformation matrix is calculated using the new combined image and the resulting transformation of the images to the PCA is a representation of that image in multidimensional space. So in the case of combining the images pre PCA every image will be represented by a multidimensional point in space.

Combining the segments or images post PCA calculations occurs in the PCA space. This is achieved by transforming the first set of images that we want to combine to the PCA space and then transforming the second set of images that we want to combine to the PCA space. Each set is transformed to the PCA space independent from the other set with its own transformation matrix. This means that now every image from the first set is represented by a multidimensional point in the PCA space and every image from the second set is represented by a multidimensional point in the PCA space.

Combining these two sets in the PCA space can be thought of as representing each combined image by two multidimensional points. This combination makes the system more flexible, and makes it faster to make the calculations and transformation to the PCA space. The performance of combining the images post PCA is not less than the

performance of combining the images pre PCA if not better. Table 6.2 presents some of the rank one recognition rates of the multimodal experiments we did pre and post PCA.

Table 6.2 Rank one recognition rates of multimodal experiments pre and post PCA

Experiment	Pre PCA	Post PCA
Ear and Face combined	91.5%	92.5%
1p Circular seg. combined with 4 seg. From 6 seg.	93%	94%
Face & 1p Circular segment	85.5%	87%
Face & grid 6p & 1p Circular segment	88%	89%

Chapter 7
Conclusion and Future Work

7.1 Conclusion

In our research we performed several experiments for ear recognition and multimodal approaches for human identification. The performed experiments that are presented in this thesis yielded good results that indicate the eligibility of using ears as a reliable biometric for human identification. Just as no one can prove the uniqueness of finger prints we cannot also prove the uniqueness of outer ear shapes for all individuals. This can only be presented as empirical test results like we have done in this thesis.

The results obtained from the multimodal experiments we performed also suggests that the ear is a very suitable biometric as a noninvasive technique to be combined with other noninvasive techniques like face recognition, or gait.

The novel approach presented in our work using spatial segmentation for dividing the recognition problem and recognizing segments of the ear proved to be successful and yielded higher recognition rates than the base eigen-ear experiment. The segmentation technique is also very useful for dealing with occlusions. The segmentation technique also gives flexibility to the system which makes it possible to recognize individuals with what ever segment of the ear that is available. The segmentation of the ear also makes the calculations for the PCA faster since the images are divided to smaller images where the PCA calculations are done independently for every segment.

The approach of combining images post PCA calculations or in the PCA space for the multimodal experiments, so that each combination of images will be represented by two points instead of one point in the PCA space, makes the system even more flexible. The multimodality could be applied when ever the appropriate information from the biometrics involved is available or could be ignored and each system will work on its own without the need to have more than one implementation to accommodate each case.

This approach also does not negatively affect the recognition rate. If the recognition rate does not improve it will not deteriorate.

The rank one recognition rates reported for the ear segmentation methods ranged from 78% to 91%, which got higher by combining two different segmentation methods achieving 94%. These recognition rates are a significant improvement over the base eigen-ears experiment which had an 84.5% rank one recognition rate.

Comparing the ear recognition rates to the face recognition rate which was a rank one recognition of 63% for the eigen-face experiment, we can see that the systems performance using ears is significantly better for the given databases. This implies that the ears are more distinctive for a computer based system to differentiate between individuals.

Multimodal experiments gave success rates that ranged from 85.5% rank one recognition rate for the combination of the face images and the one circular ear segments, to 94% rank one for the face and ear images combination post PCA.

Receiver operating curves for all the performed experiments are presented in appendix I which shows the performance of the experiments performed with respect to the recognition rate versus the false alarm rate. Cumulative match characteristics curves for all the performed experiments are presented in appendix II which measures the performance of the experiments performed with respect to recognition rate at different ranks.

7.2 Future work

Future work may pursue different areas of the research. The areas of research that could be pursued include feature extraction from ear images. We used principal components analysis and transformation to the PCA space also known as the eigen-face approach, as our feature extraction technique. Other feature extraction techniques like edge images, force field extraction, or contour detection could be used to extract features for use as feature vectors for ear recognition.

In the normalization step explained in chapter 3 landmark selection was done manually. Future work may include a method for automatically selecting landmark points for normalization.

The segmentation techniques presented in our work are not the only possible segmentation methods; in fact the possibilities are endless. Different segmentation schemes can be pursued and tested to obtain a more optimum segmentation scheme.

The segmentation technique also involves the fusion of results from individual segment classifiers. A different approach for classifier fusion may be implemented that could give better results. Also we used a nearest neighbor classifier on our research; different types of classifiers can replace the nearest neighbor classifier.

In the multimodal area of our research we presented a multimodal experiment using face and ear images, and also using face and ear segments, and an example of combining two segmentation methods. This area could be pursued, testing more combination of ear segmentation methods or combining face images with more ear segmentation methods.

Bibliography

[1] A. Bertillon, *La Photographie Judiciaire, avec un Appendice sur la Classification et l'Identification Anthropometriques*, Gauthier-Villars, Paris, 1890.

[2] A. Iannarelli, *Ear Identification*, Paramont Publishing Company, 1989.

[3] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, 2002.

[4] B. Bhanu and H. Chen, "Human Ear Recognition in 3D," in *Proc. Workshop Multimodal User Authentication*, 2003, pp. 91-98.

[5] B. Moreno, A. Sanchez, and J. F. Velez, "On the Use of Outer Ear Images for Personal Identification," in *Proc. IEEE International Carnahan Conf. on Security Technology*, 1999, pp. 469-476.

[6] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force Field Feature Extraction for Ear Biometrics," *Computer Vision and Image Understanding*, vol. 98, pp. 491-512, June 2005.

[7] Face and profile image databases F and G, computer vision research Laboratory at the University of Notre Dame.

[8] H. Chen and B. Bhanu, "Contour Matching for 3D Ear Recognition," in *Proc. Seventh IEEE Workshop on Application of Computer Vision*, 2005, pp. 123-128.

[9] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and Combination of Ear and Face Images in Appearance-based Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25 pp. 1160-1165, Sept. 2003.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: the extended M2VTS Database," in *Proc. Audio and Video Based Person Authentication*, Washington DC, 1999.

[11] M. Á. Carreira-Perpiñán, "Compression Neural Networks for Feature Extraction: Application to Human Recognition from Ear Images" (in Spanish), M.S. thesis, Faculty of Informatics, Technical University of Madrid, Spain, 1995.

[12] M. Burge and W. Burger, "Ear biometrics", chapter 13, *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, 2002.

[13] M. Burge and W. Burger. “Ear Biometrics in Computer Vision”. *Proc. 15th International Conf. of Pattern Recognition*, vol. 2, pp. 822–826, 2000.

[14] M. K. Fleming, G. W. Cottrell. “Categorization of Faces Using Unsupervised Feature Extraction”. *Proc. International Joint Conf. On Neural Networks*, vol. 11, pp. 65-70, 1990.

[15] M. Choras, “Ear Biometrics Based on Geometric Feature Extraction”, *Electronic Letters on Computer Vision and Image Analysis 5(3)*, 84-95, 2005.

[16] M. Saleh, S. Fadel, and L. Abbott, “Using Ears as a Biometric for Human Recognition”, *Proc. International Conf. on Computer Theory and Applications*, pp. 311-314, Sept. 2006.

[17] M. Turk, A. Pentland. ”Eigenfaces for Recognition”. *Journal of Cognitive Neuroscience*, Vol 3, No. 1, pp. 71-86, 1991.

[18] National Science and Technology Council (NSTC), Committee on Technology, Committee on Homeland and National Security, Subcommittee on biometrics, “Biometrics Glossary”, September 2006.

[19] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, Oct. 2000, pp.1090-1104.

[20] P. Yan and K.W. Bowyer, “A Fast Algorithm for ICP-Based 3D Shape Biometrics,” *Proc. IEEE Workshop Automatic Identification Advanced Technologies*, pp. 213-218, 2005.

[21] P. Yan and K.W. Bowyer, “An Automatic 3D Ear Recognition System,” *Proc. Third International Symp. 3D Data Processing, Visualization, and Transmission*, pp. 213-218, 2006.

[22] P. Yan and K.W. Bowyer, “Ear Biometrics Using 2D and 3D Images,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Advanced 3D Imaging for Safety and Security*, 2005.

[23] P. Yan and K.W. Bowyer, “ICP-Based Approaches for 3D Ear Recognition,” *Proc. SPIE Conf. Biometric Technology for Human Identification*, pp. 282-291, 2005.

[24] P. Yan and K.W. Bowyer, “Multi-Biometrics 2D and 3D Ear Recognition,” *Proc. Audio and Video-Based Biometric Person Authentication*, pp. 503-512, 2005.

[25] P. J. Grother, R. J. Micheals and P. J. Phillips, “Face Recognition Vendor Test 2002 Performance Metrics”, *Proc. fourth International Conf. on Audio Visual Based Person Authentication*, 2003, pp. 937-945.

[26] P. Yan and K. W. Bowyer, “Empirical Evaluation of Advanced Ear Biometrics”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Empirical Evaluation Methods in Computer Vision*, 2005, pp. 41-48.

[27] Wikipedia the free encyclopedia. http://en.wikipedia.org/wiki/Biometrics

Appendix I

Receiver Operating Curves

This appendix contains the receiver operating curves for all experiments performed in our research. These curves are shown in figures AI.1 to AI.27. The curve is drawn as a parametric relation between the true accept rate (recognition rate) and the false alarm rate at different thresholds.

Table AI.1 and figure AI.28 introduce the area under the curve for all experiments performed to compare the performance of different methods to the base eigen-ear and eigen-face experiments with respect to the false alarm rate.

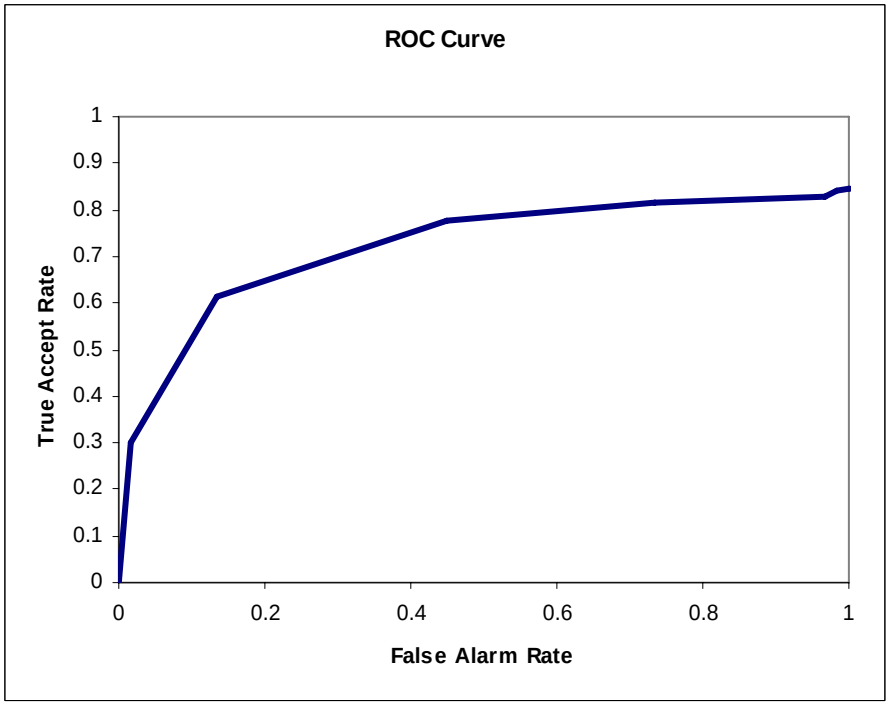

Figure AI.1 ROC curve for base eigen-ear experiment

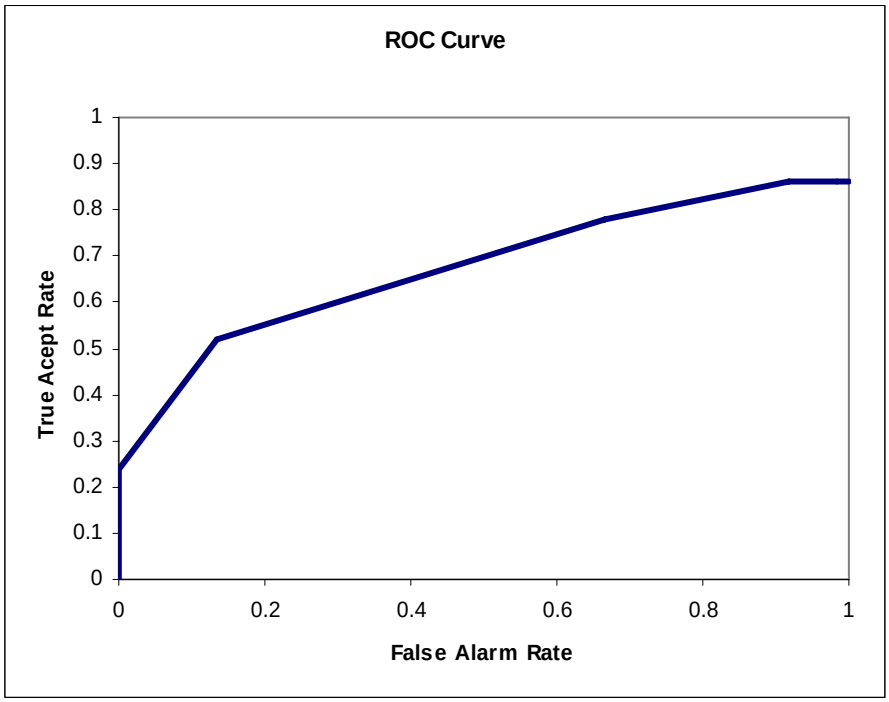

Figure AI.2 ROC curve for 4 uniform horizontal segments experiment

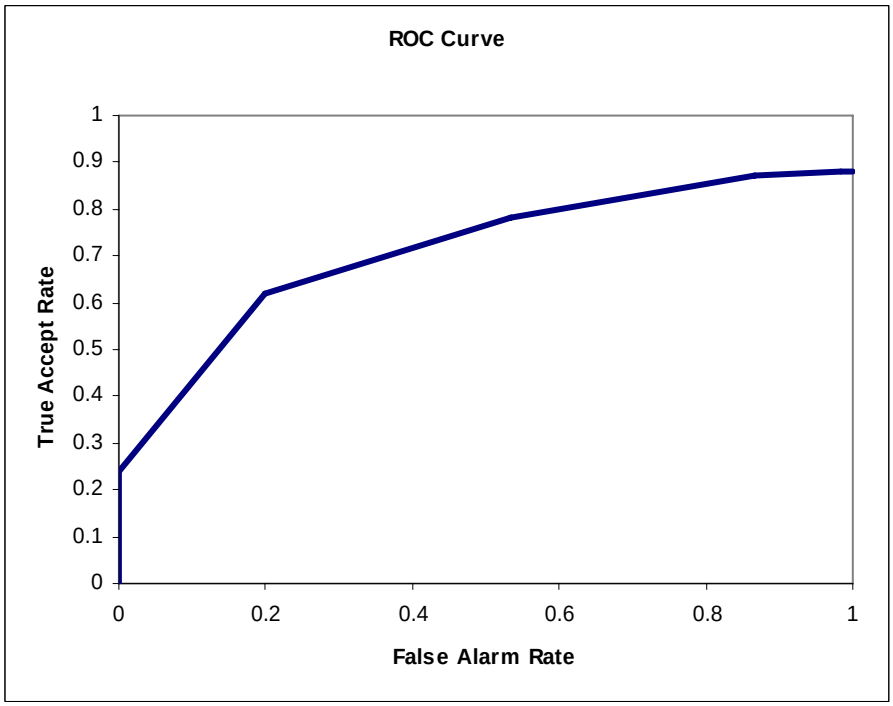

Figure AI.3 ROC curve for 3 uniform horizontal segments experiment

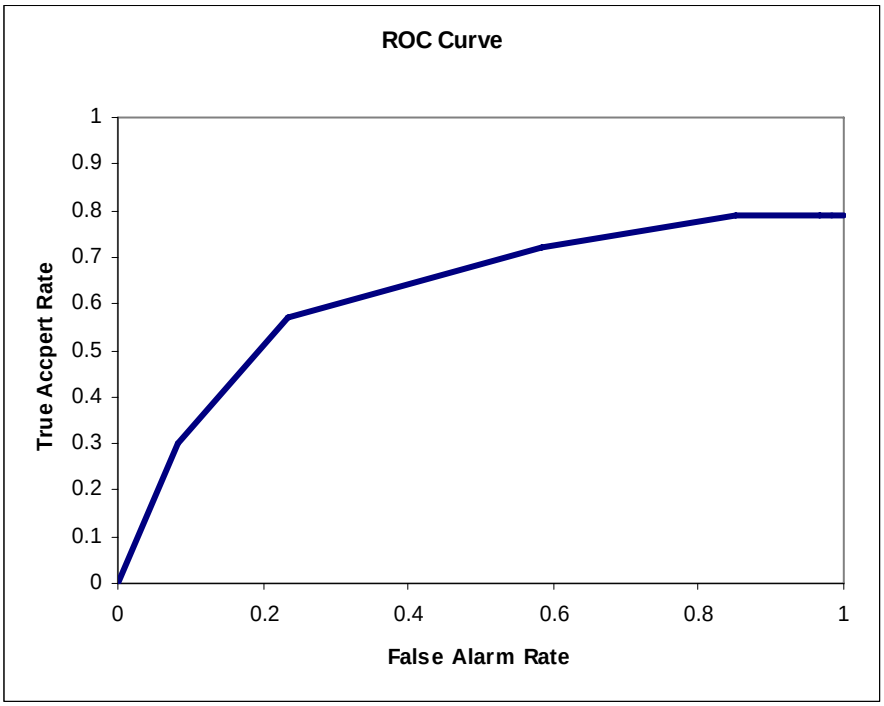

Figure AI.4 ROC curve for 2 uniform horizontal segments experiment

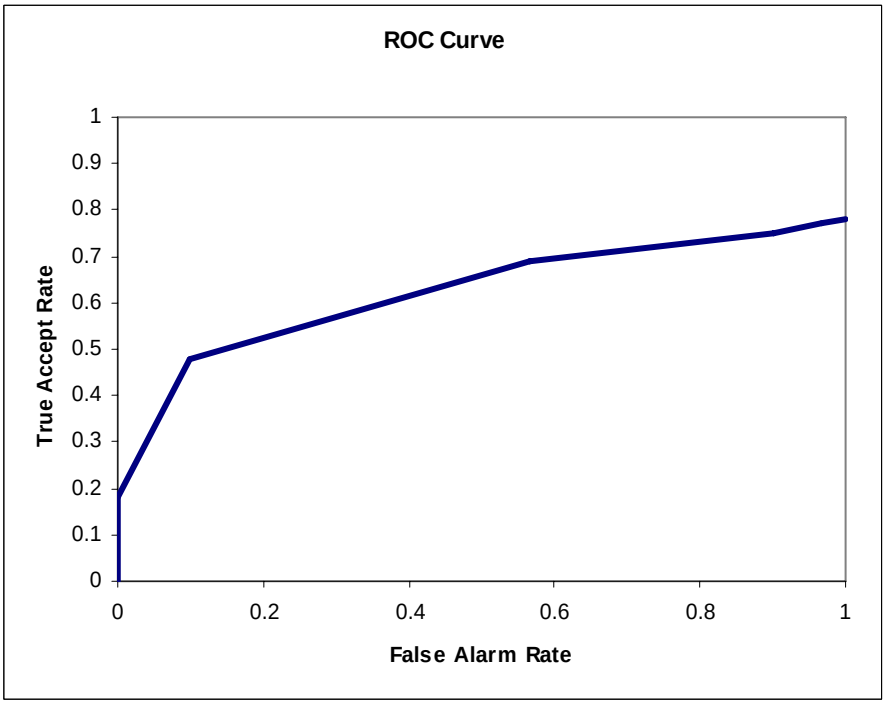

Figure AI.5 ROC curve for 4 uniform vertical segments experiment

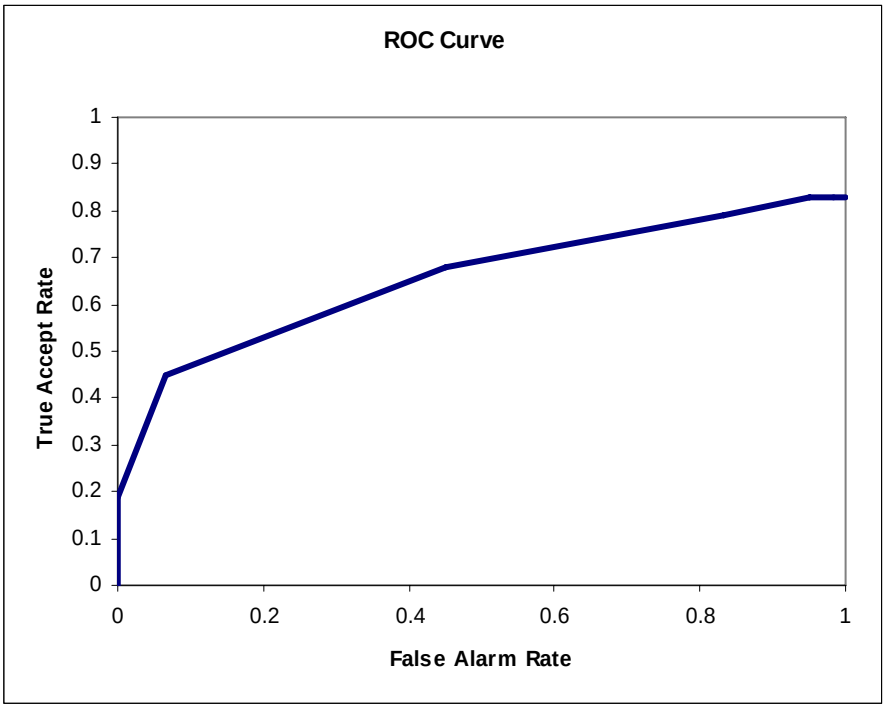

Figure AI.6 ROC curve for 3 uniform vertical segments experiment

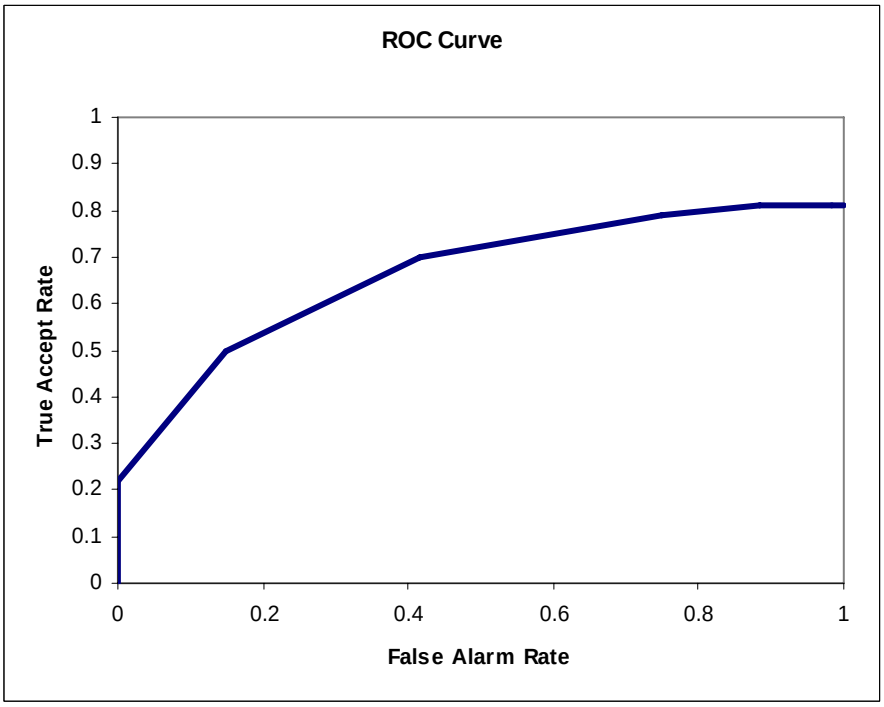

Figure AI.7 ROC curve for 2 uniform vertical segments experiment

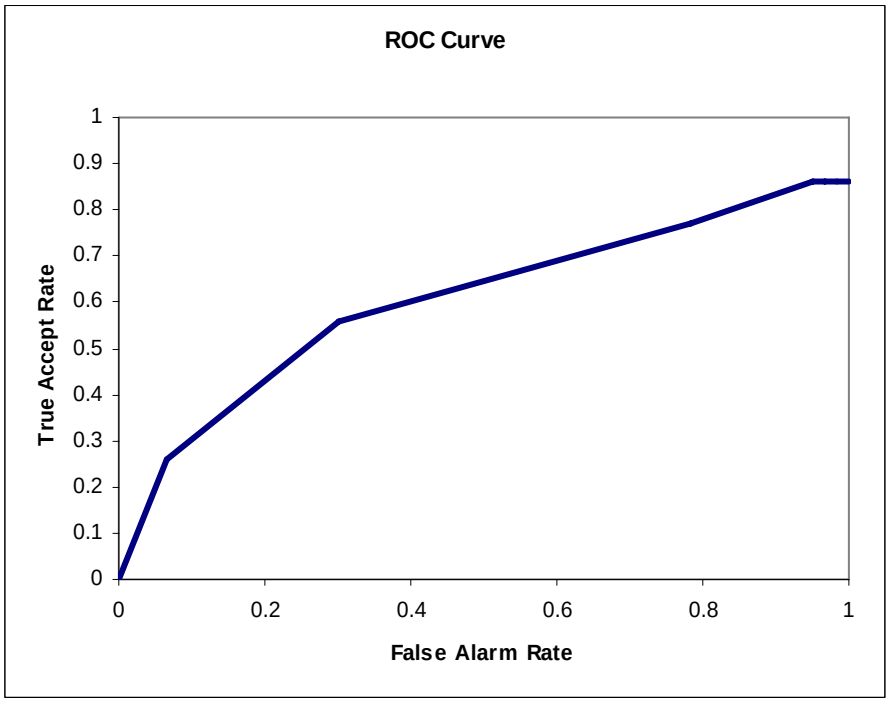

Figure AI.8 ROC curve for 4 uniform grid segments experiment

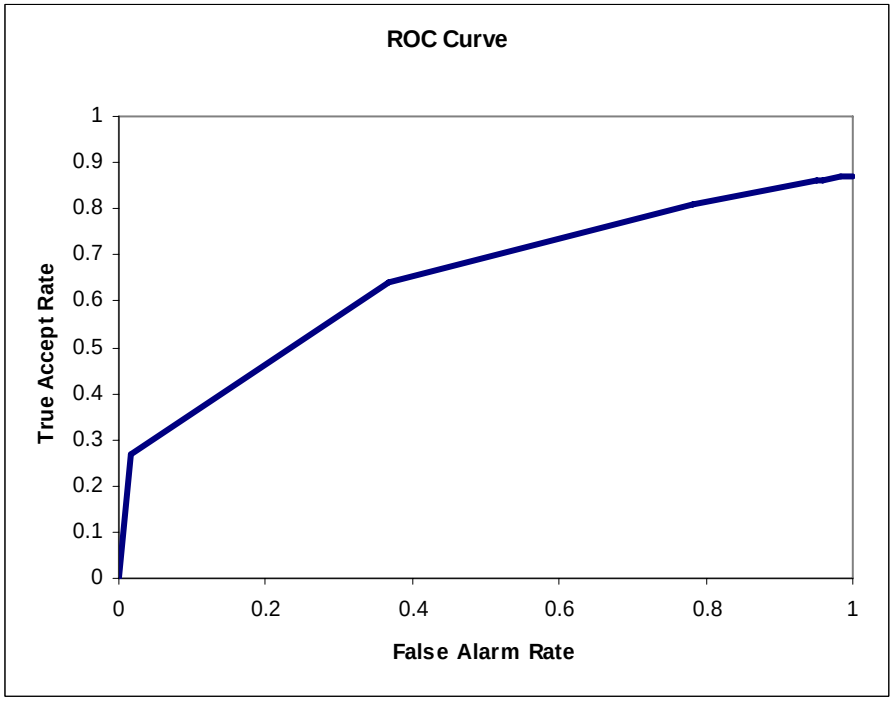

Figure AI.9 ROC curve for 6 uniform grid segments experiment

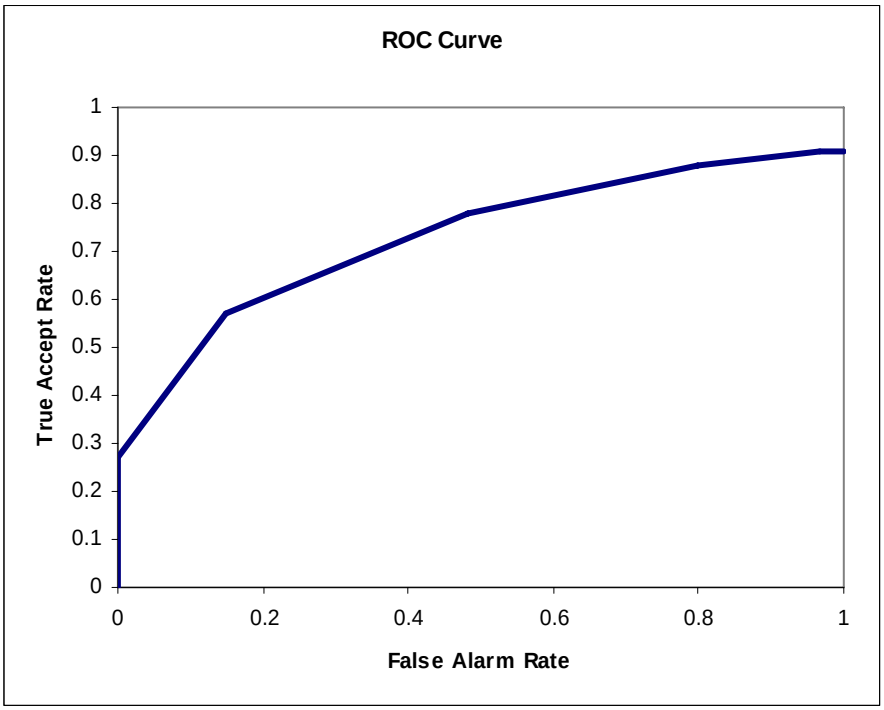

Figure AI.10 ROC curve for 3 non-uniform horizontal segments experiment

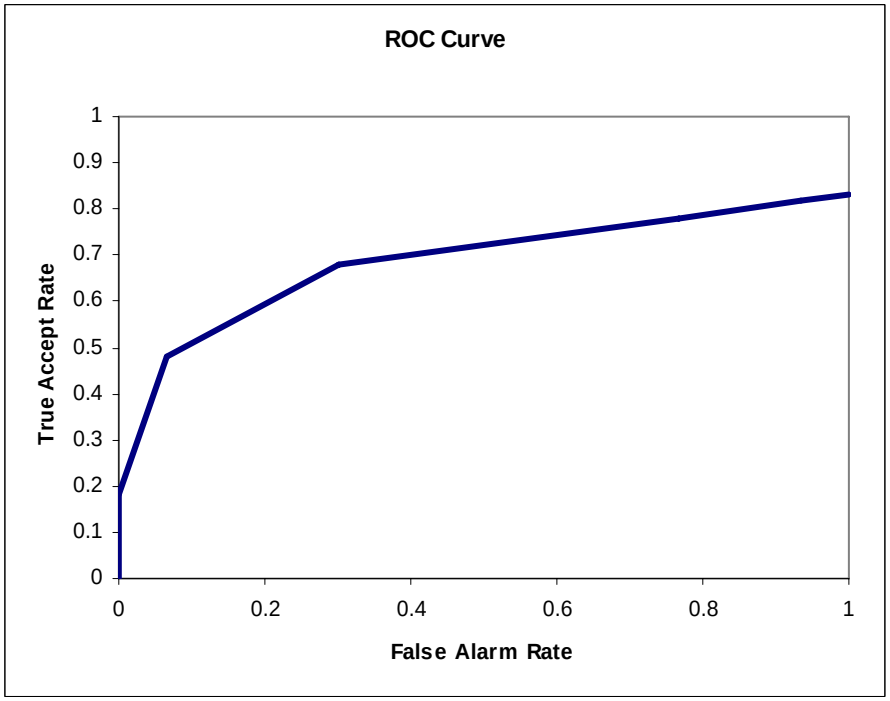

Figure AI.11 ROC curve for 3 non-uniform vertical segments experiment

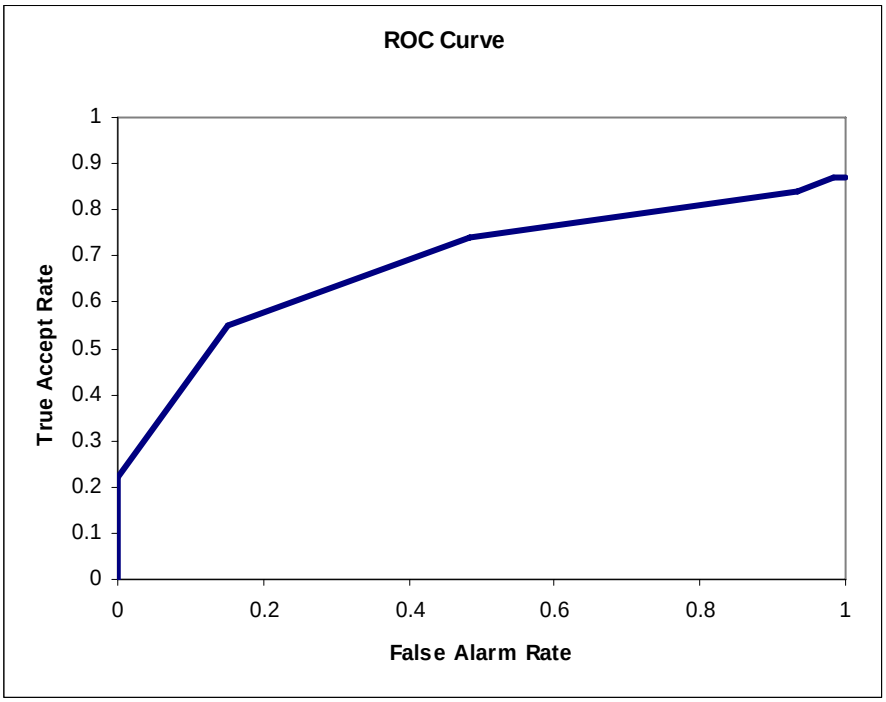

Figure AI.12 ROC curve for 4 (3 horizontal and 1 vertical) non-uniform segments experiment

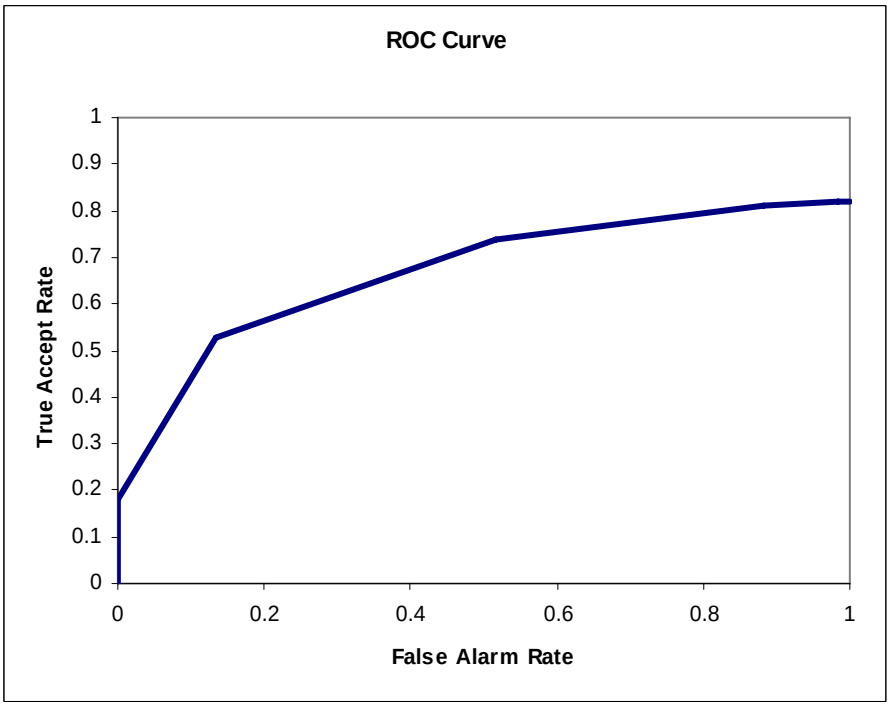

Figure AI.13 ROC curve for 3 (2 horizontal and 1 vertical) non-uniform segments experiment

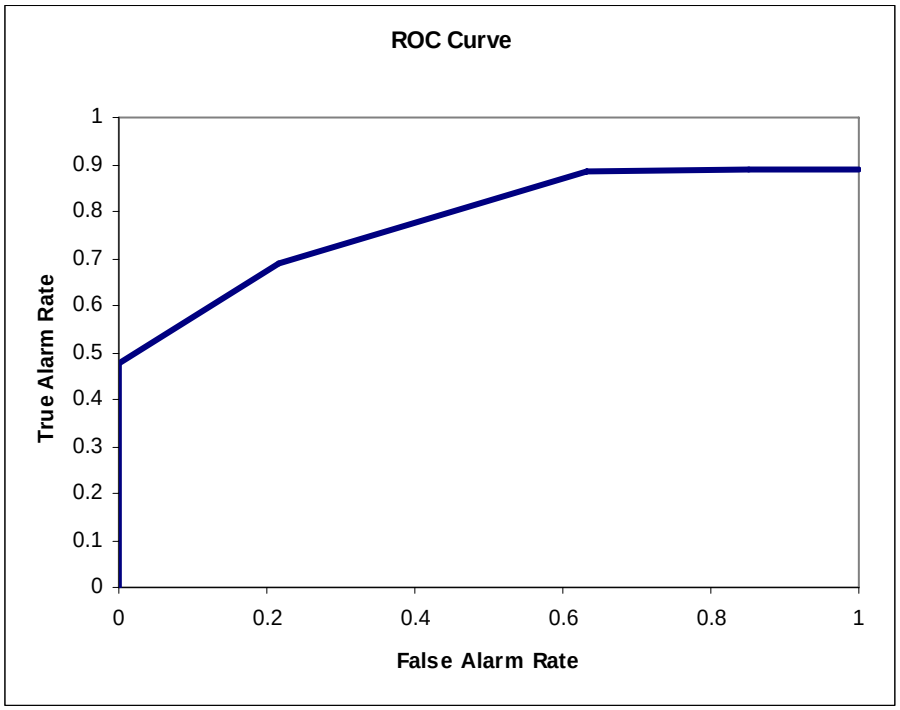

Figure AI.14 ROC curve for 1 circular segment experiment

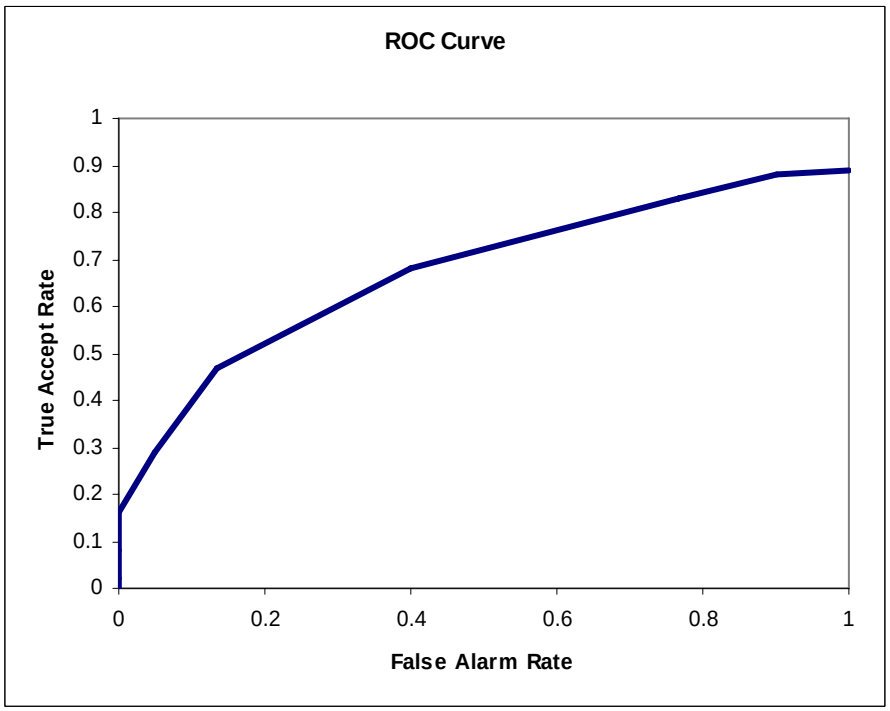

Figure AI.15 ROC curve for threshold segmentation experiment

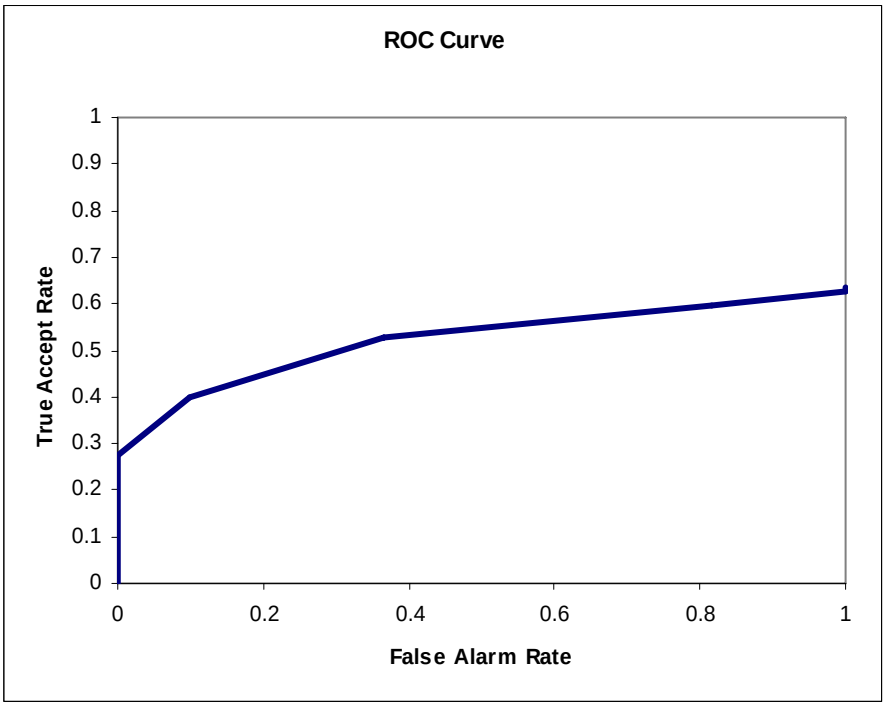

Figure AI.16 ROC curve base eigen-face experiment

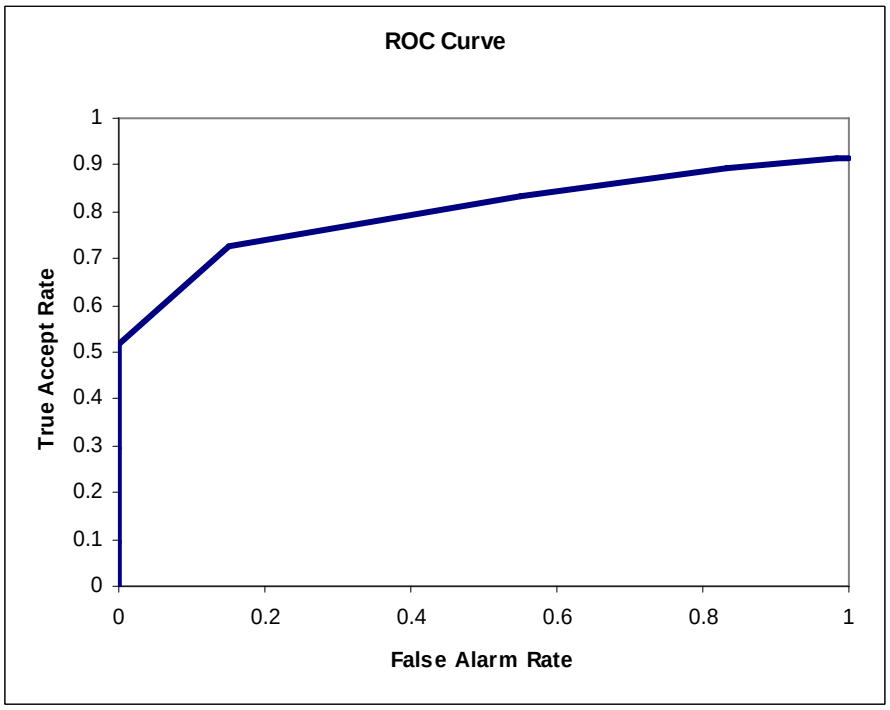

Figure AI.17 ROC curve for multimodal face-ear pre PCA version 1 experiment

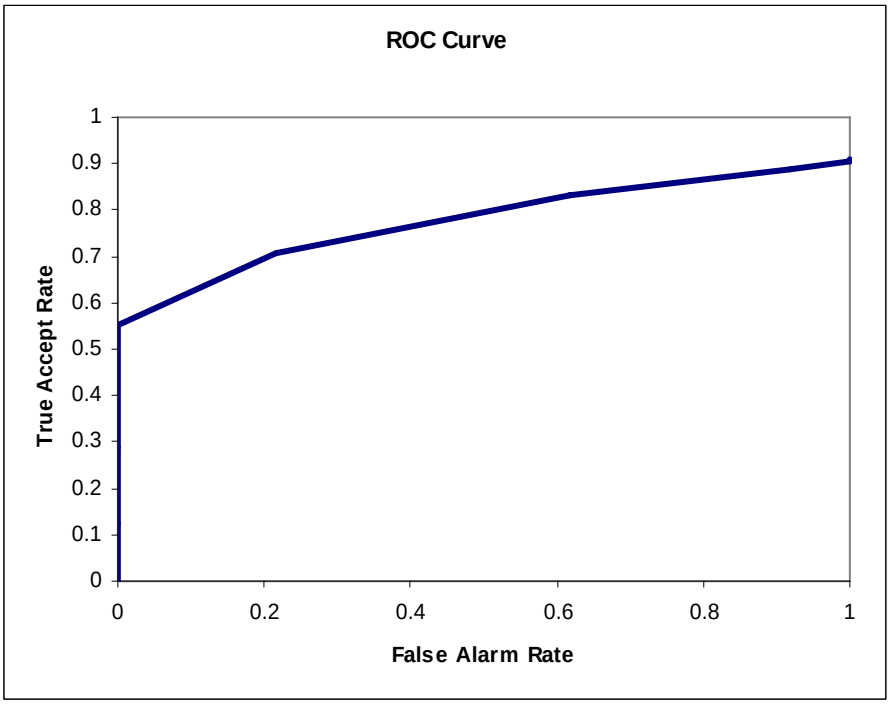

Figure AI.18 ROC curve for multimodal face-ear pre PCA version 2 experiment

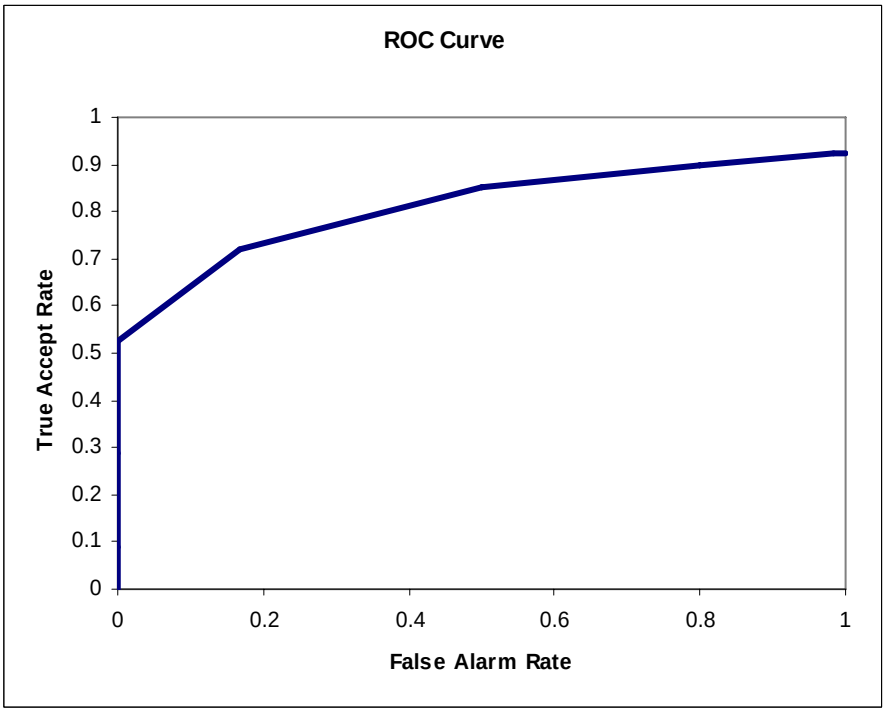

Figure AI.19 ROC curve for multimodal face-ear post PCA experiment

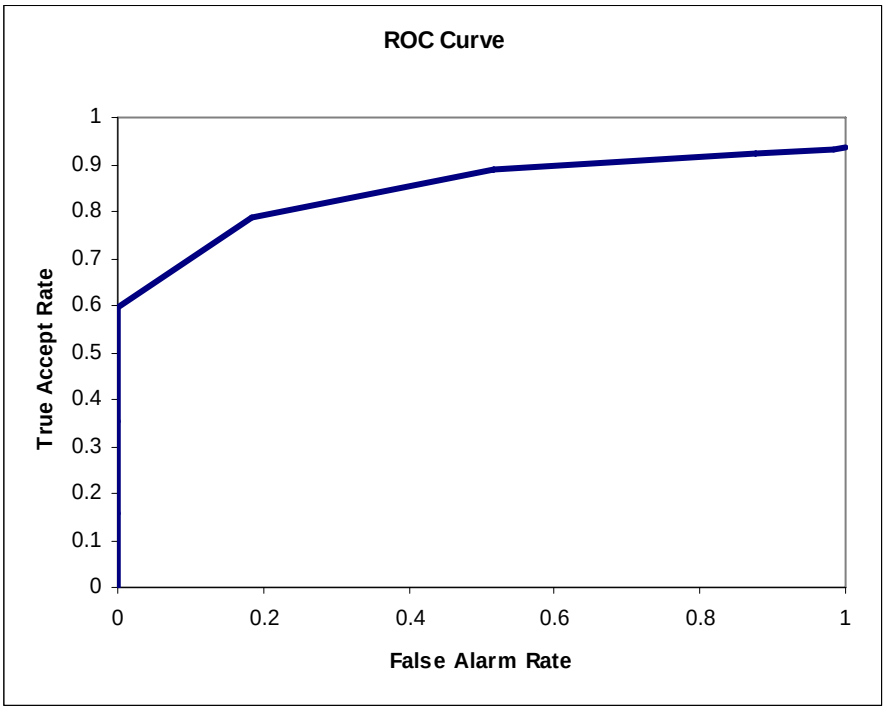

Figure AI.20 ROC curve for multimodal face-ear post PCA with permutation experiment

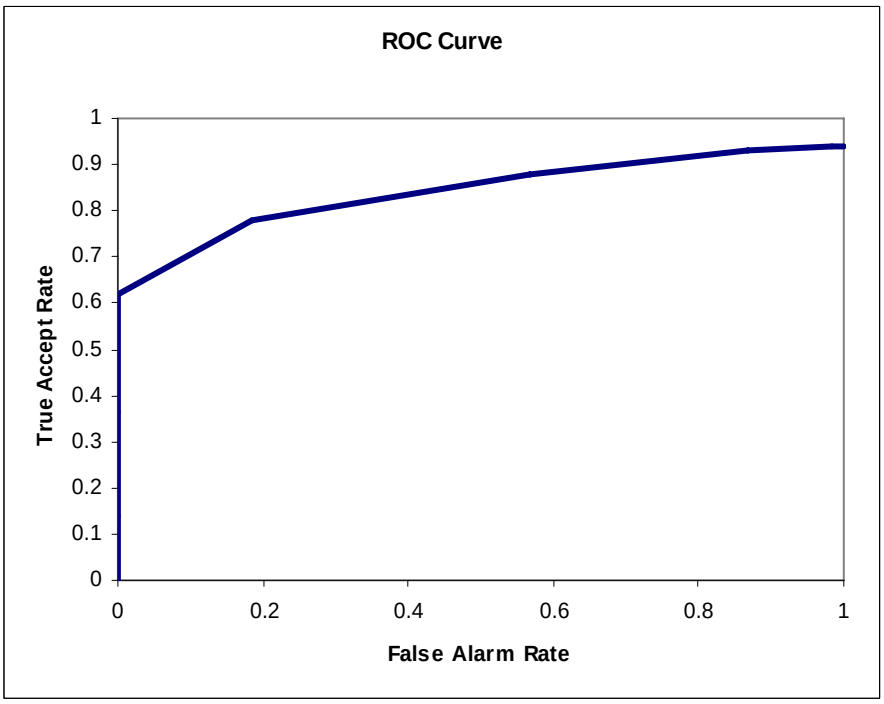

Figure AI.21 ROC curve for multimodal face-ear post PCA with permutation for Training set only experiment

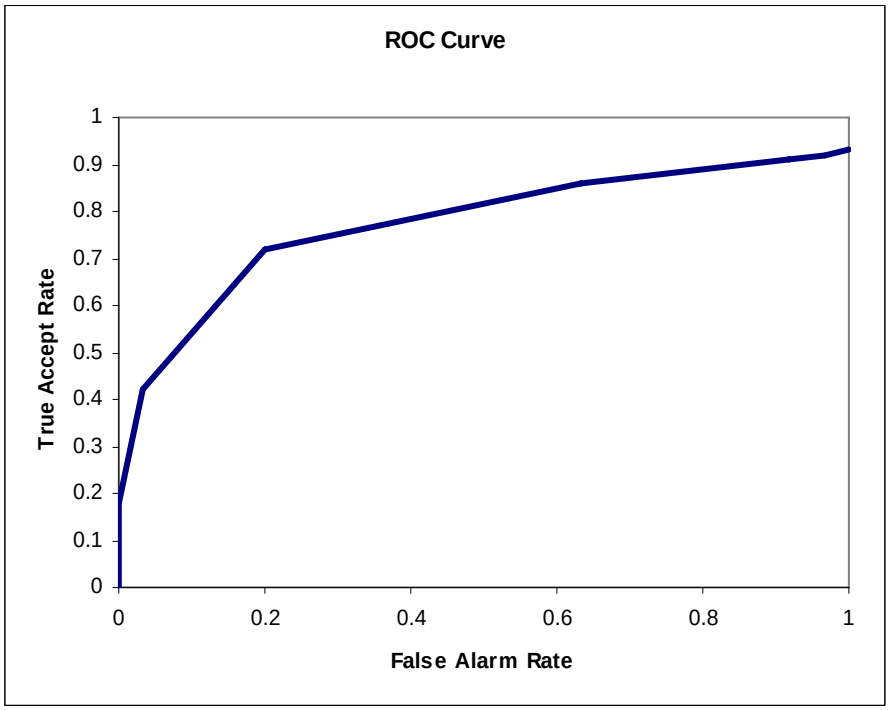

Figure AI.22 ROC curve for multimodal 1 circular segment and 4p segments pre PCA experiment

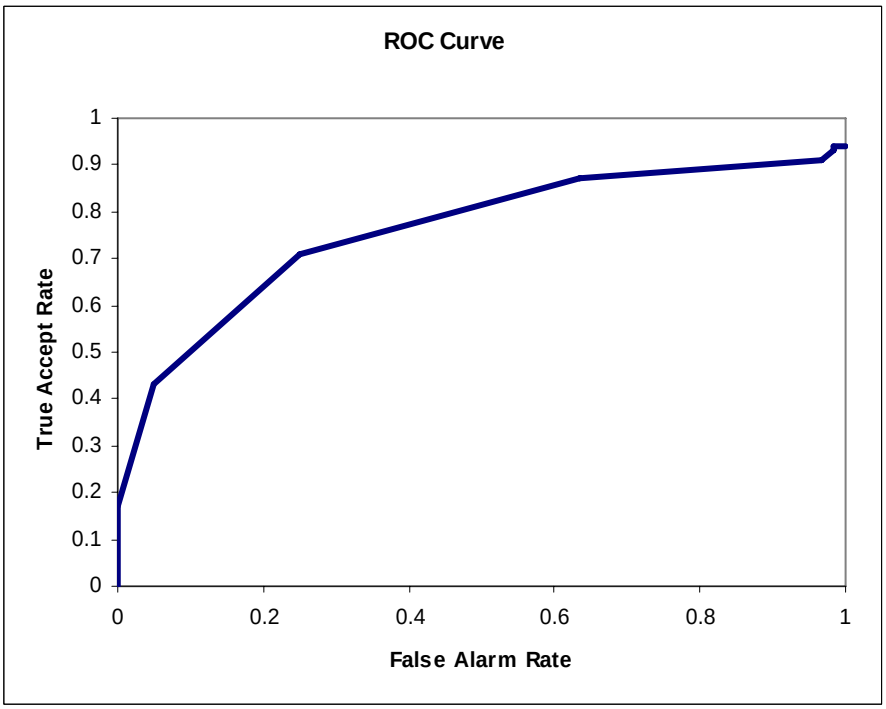

Figure AI.23 ROC curve for multimodal 1 circular segment and 4p segments post PCA experiment

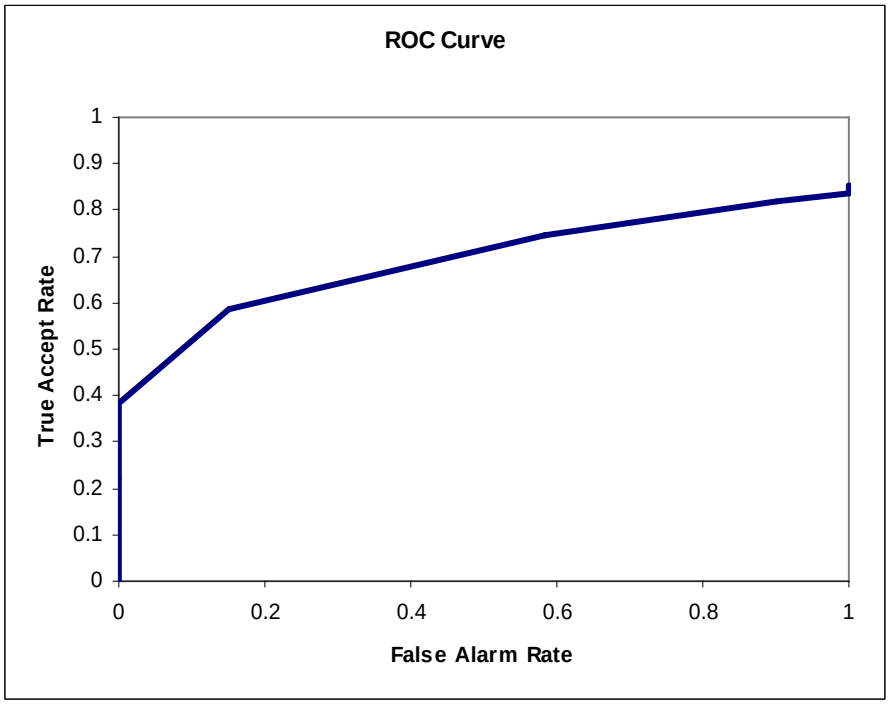

Figure AI.24 ROC curve for multimodal face and 1 circular segment pre PCA experiment

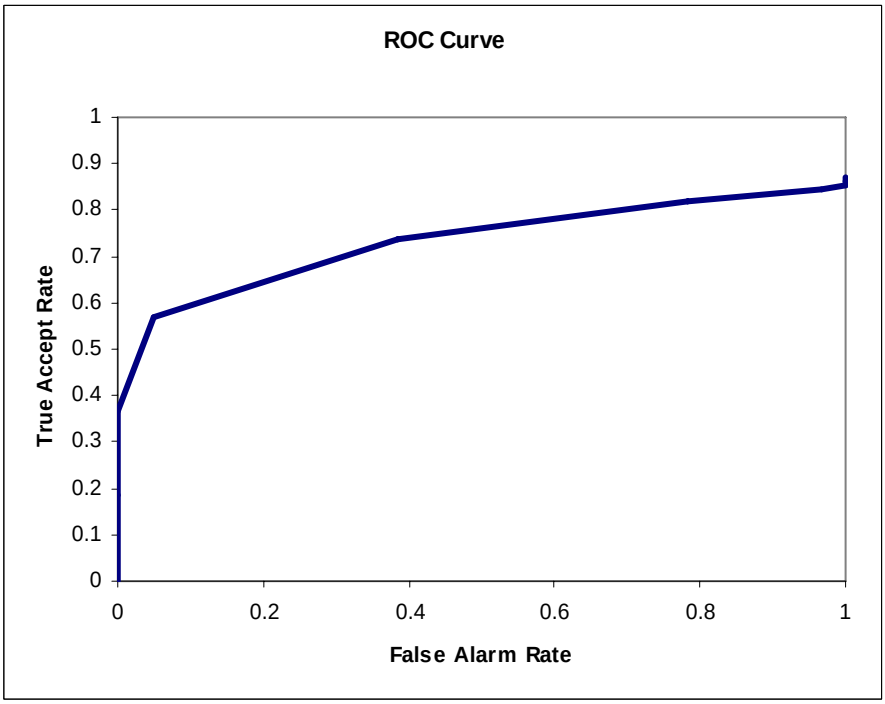

Figure AI.25 ROC curve for multimodal face and 1 circular segment post PCA experiment

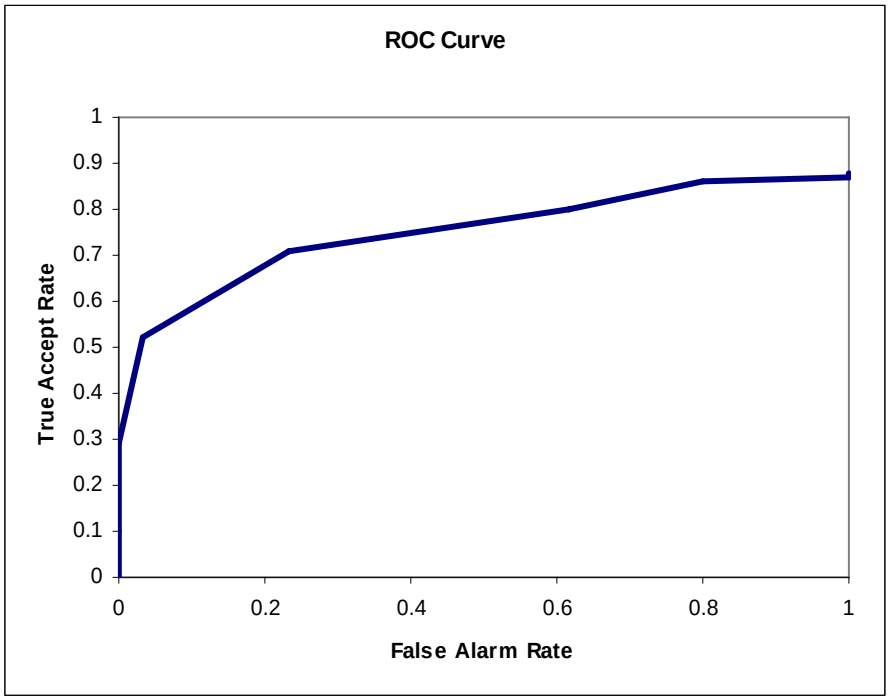

Figure AI.26 ROC curve for multimodal face and 2 combined segmentation methods pre PCA experiment

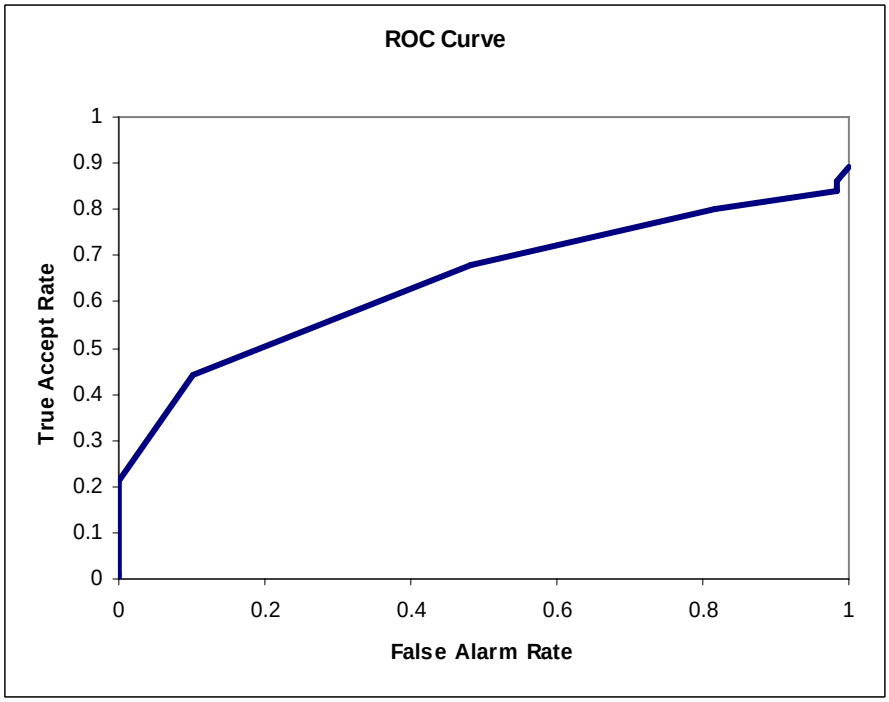

Figure AI.27 ROC curve for multimodal face and 2 combined segmentation methods post PCA experiment

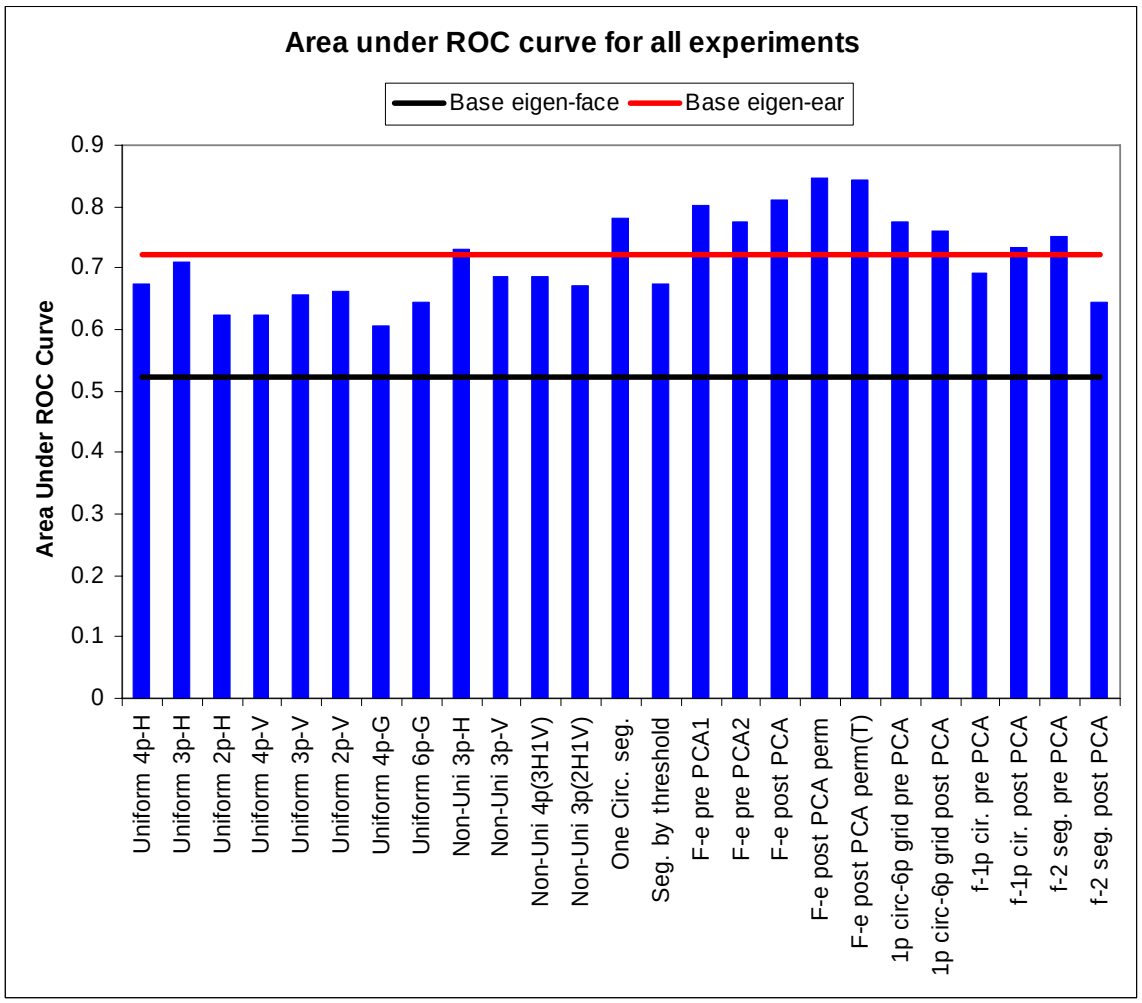

Figure AI.28 Area under the curve comparison for all experiments performed.

TableAI.1 Area under the curve for performed experiments

Experiment	Area Under ROC Curve
Base eigen-ear	0.7211
Base eigen-face	0.5227
Uniform 4p-H	0.674
Uniform 3p-H	0.7111
Uniform 2p-H	0.6233
Uniform 4p-V	0.6225
Uniform 3p-V	0.6557
Uniform 2p-V	0.6635
Uniform 4p-G	0.6046
Uniform 6p-G	0.646
Non-Uni 3p-H	0.7303
Non-Uni 3p-V	0.6863
Non-Uni 4p(3H1V)	0.6856
Non-Uni 3p(2H1V)	0.6701
One Circ. seg.	0.7801
Seg. by threshold	0.6756
F-e pre PCA1	0.8011
F-e pre PCA2	0.7757
F-e post PCA	0.8106
F-e post PCA perm	0.8473
F-e post PCA perm(T)	0.8428
1p circ-6p grid pre PCA	0.7747
1p circ-6p grid post PCA	0.7595
f-1p cir. pre PCA	0.6915
f-1p cir. post PCA	0.7328
f-2 seg. pre PCA	0.7511
f-2 seg. post PCA	0.6451

Appendix II

Cumulative Match Characteristic Curves

This appendix contains cumulative match curves for all experiments performed in our research. These curves are shown in figures AII.1 to AII.27, where the rank is defined as before and the recognition rate at a specific rank is defined as:

$$C(r) = \{p_j : rank(p_j) \leq r\} \quad \forall\ p_j \in P_g$$

We now define the CMC to be the identification rate as a function of *r*:

$$P_I(r) = \frac{|C(r)|}{|P_g|}$$

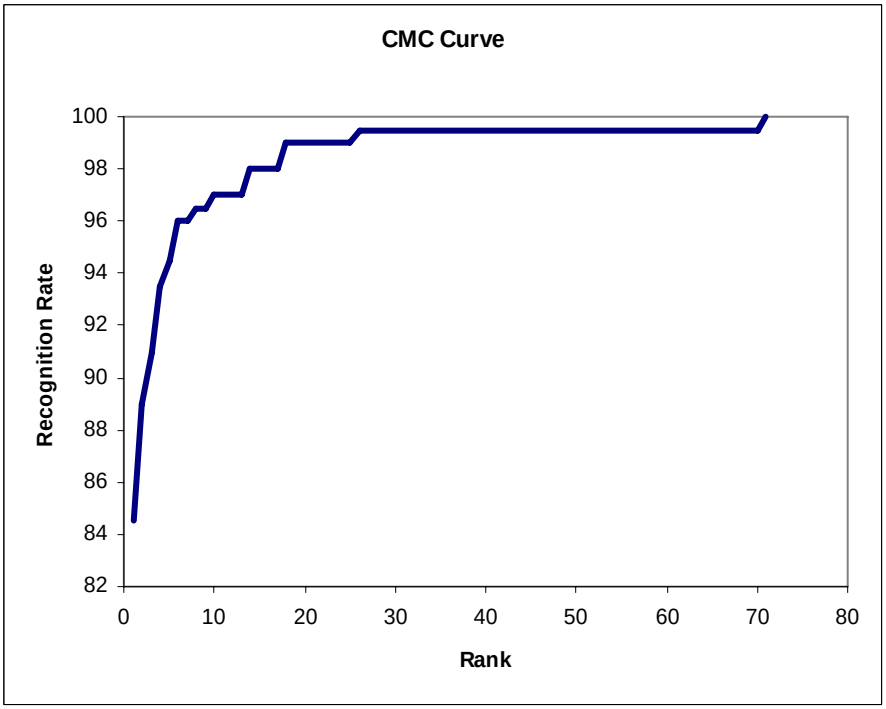

Figure AII.1 CMC curve for base eigen-ear experiment

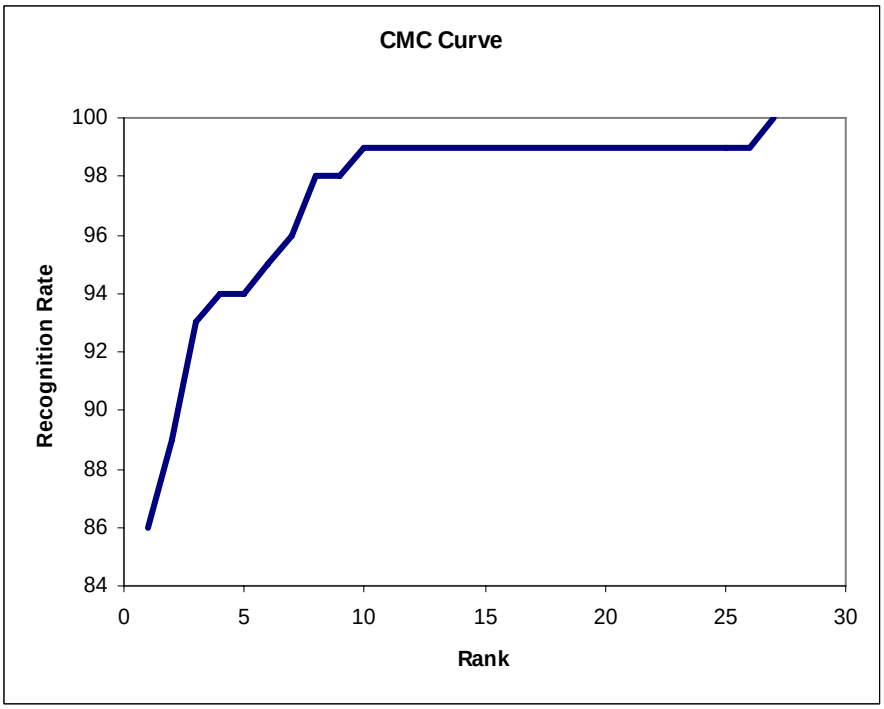

Figure AII.2 CMC curve for 4 uniform horizontal segments experiment

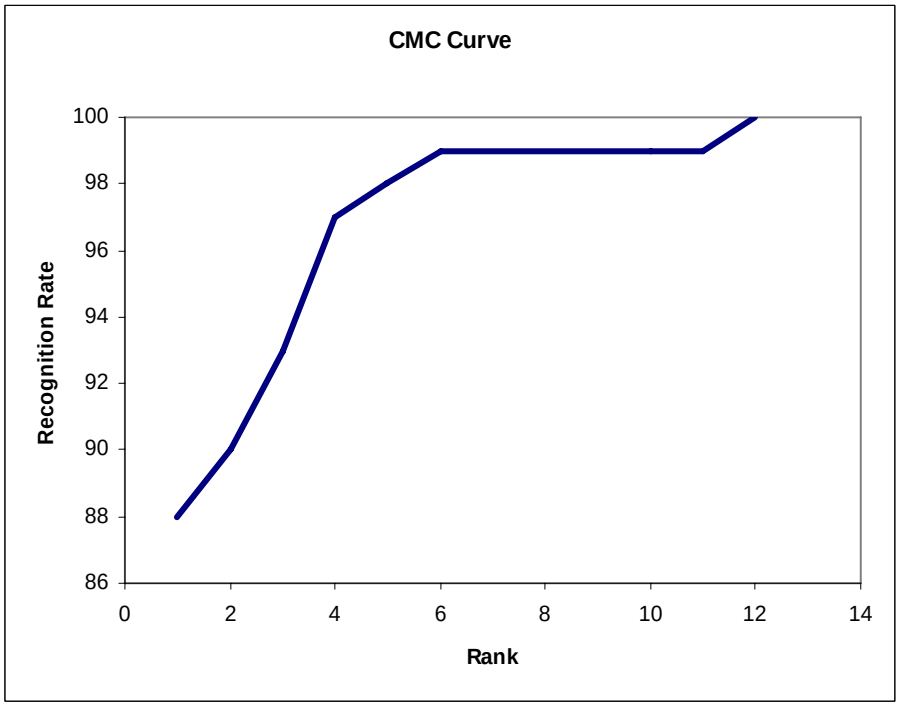

Figure AII.3 CMC curve for 3 uniform horizontal segments experiment

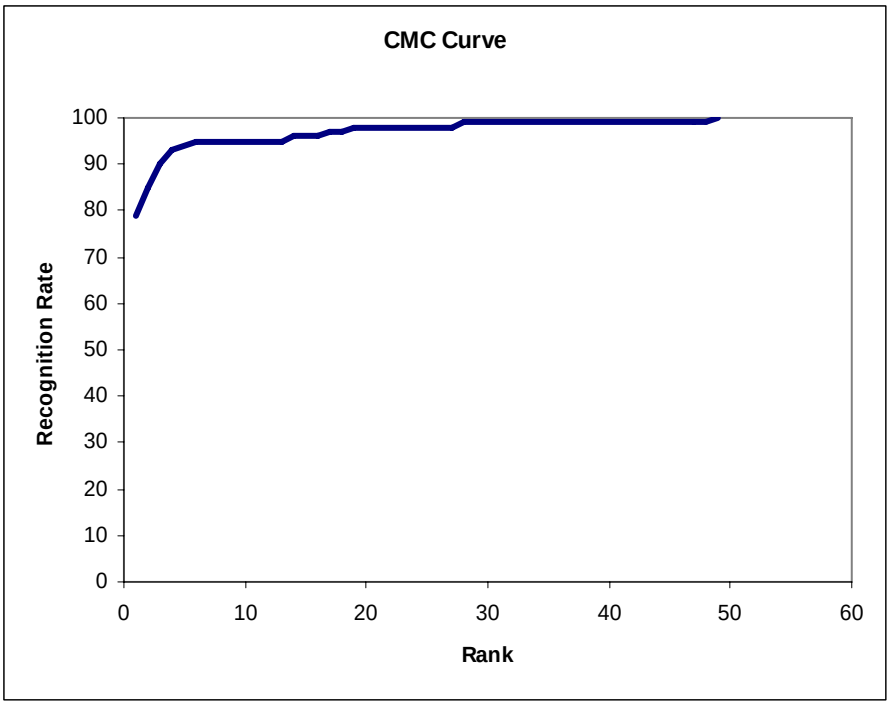

Figure AII.4 CMC curve for 2 uniform horizontal segments experiment

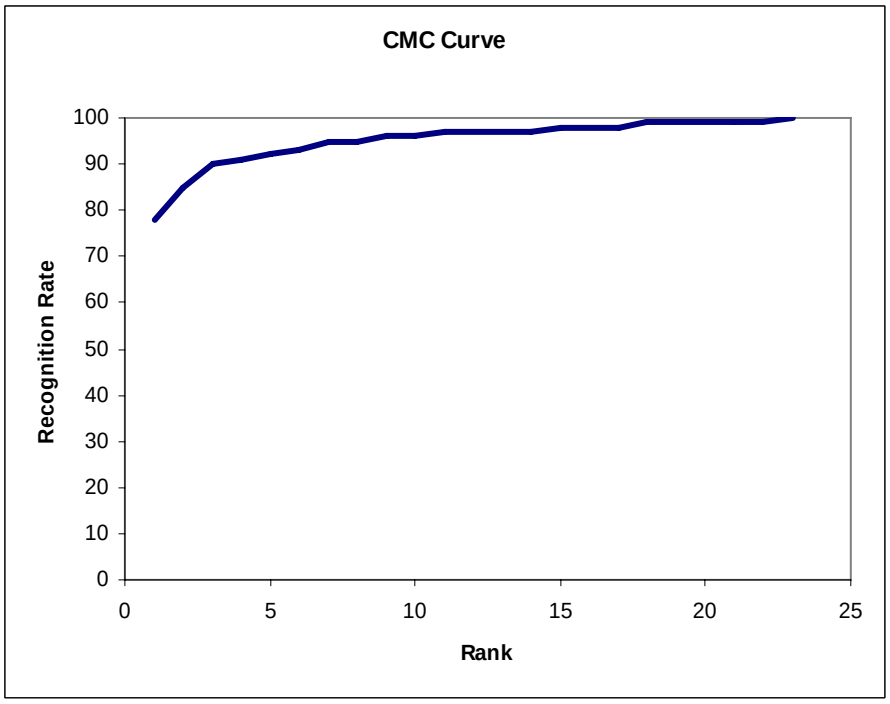

Figure AII.5 CMC curve for 4 uniform vertical segments experiment

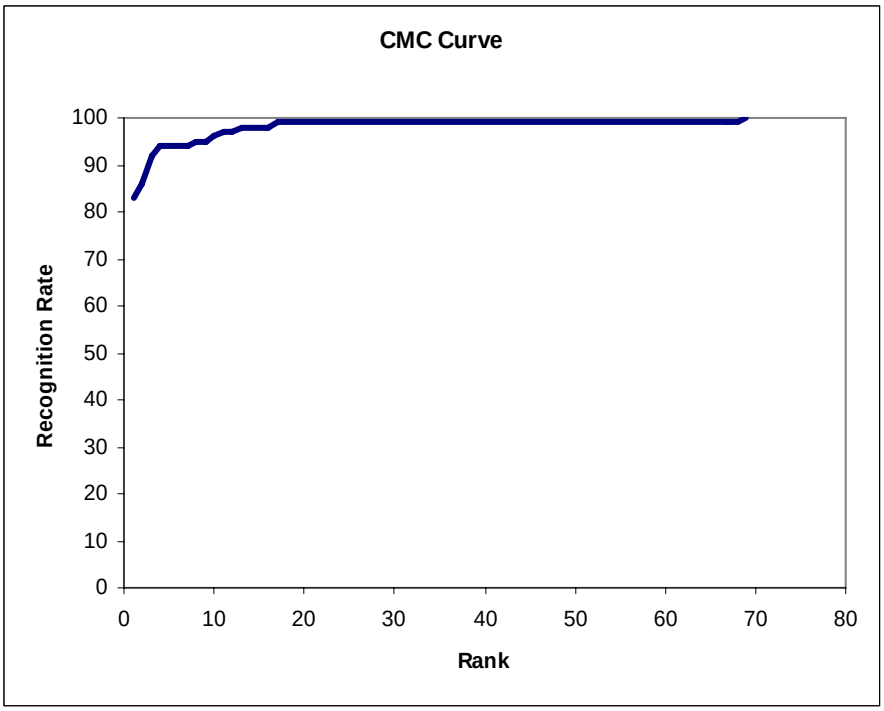

Figure AII.6 CMC curve for 3 uniform vertical segments experiment

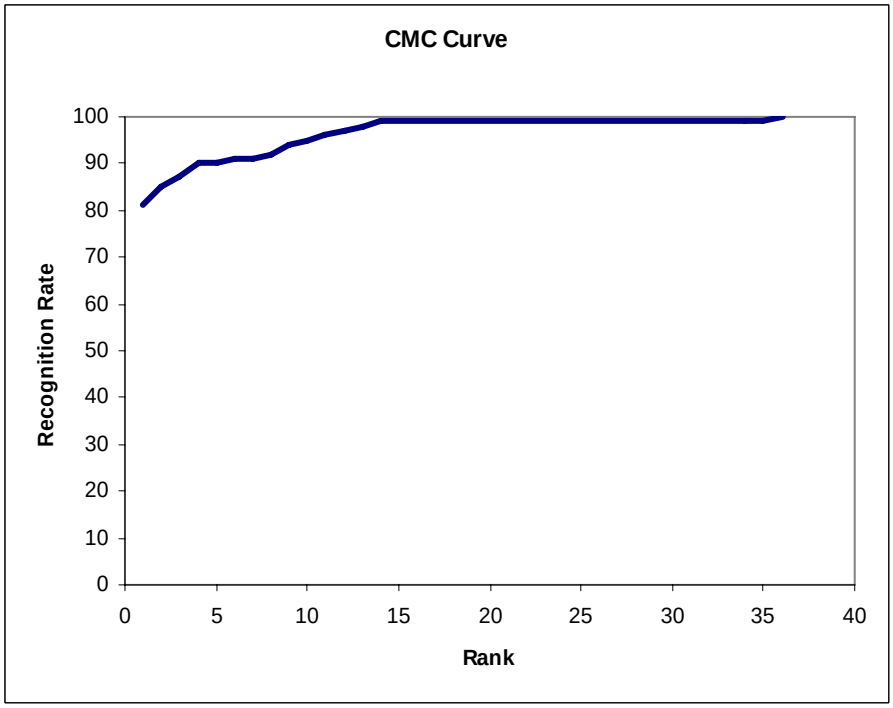

Figure AII.7 CMC curve for 2 uniform vertical segments experiment

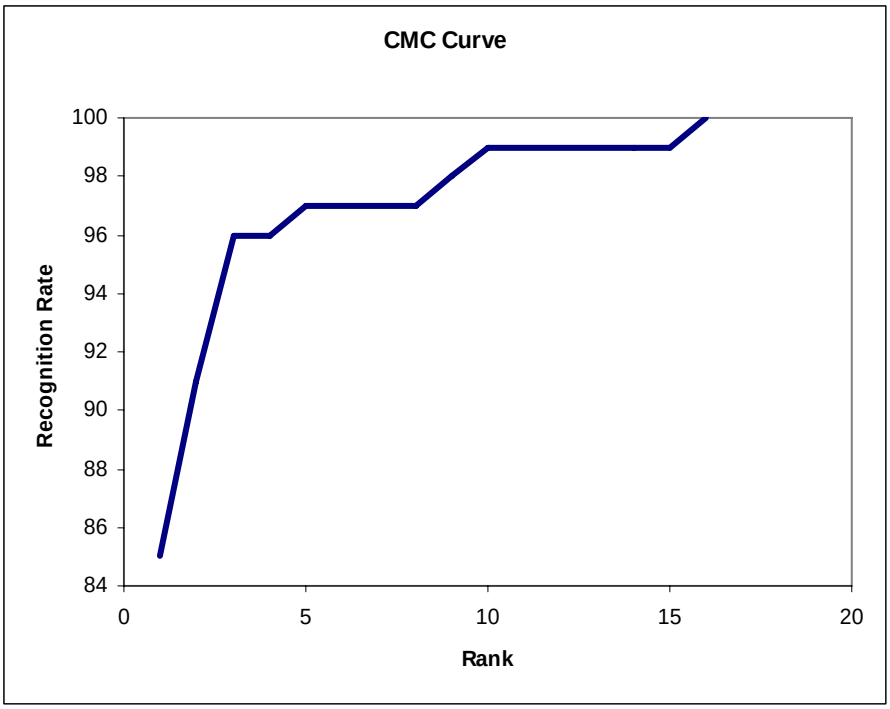

Figure AII.8 CMC curve for 4 uniform grid segments experiment

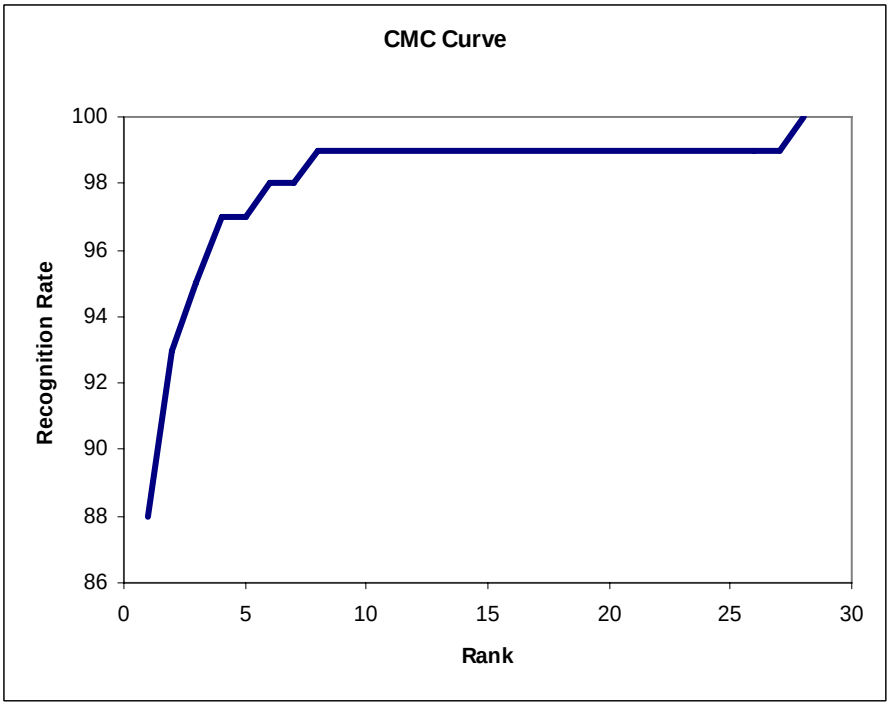

Figure AII.9 CMC curve for 6 uniform grid segments experiment

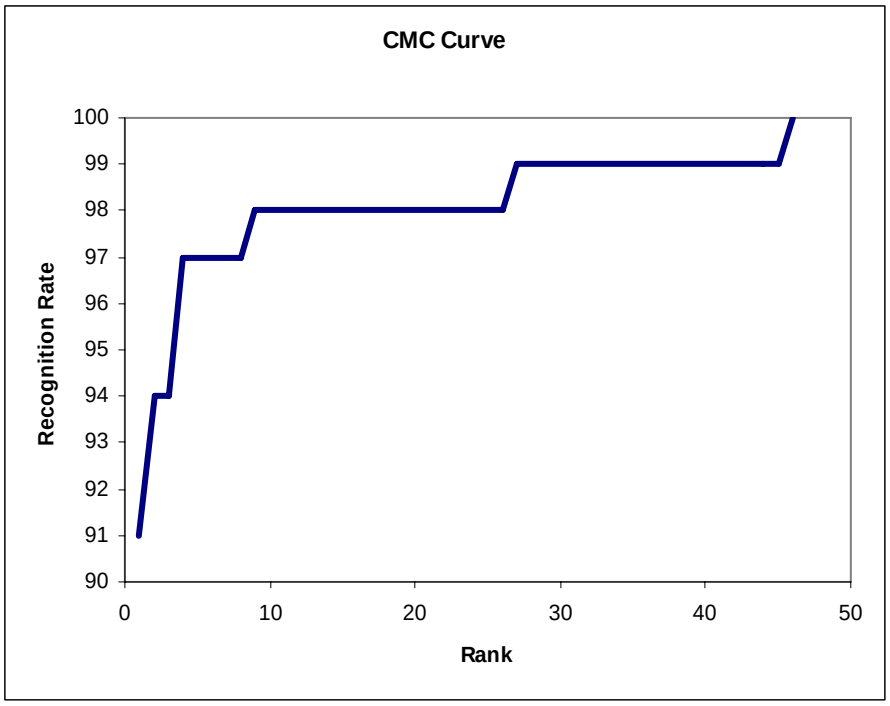

Figure AII.10 CMC curve for 3 non-uniform horizontal segments experiment

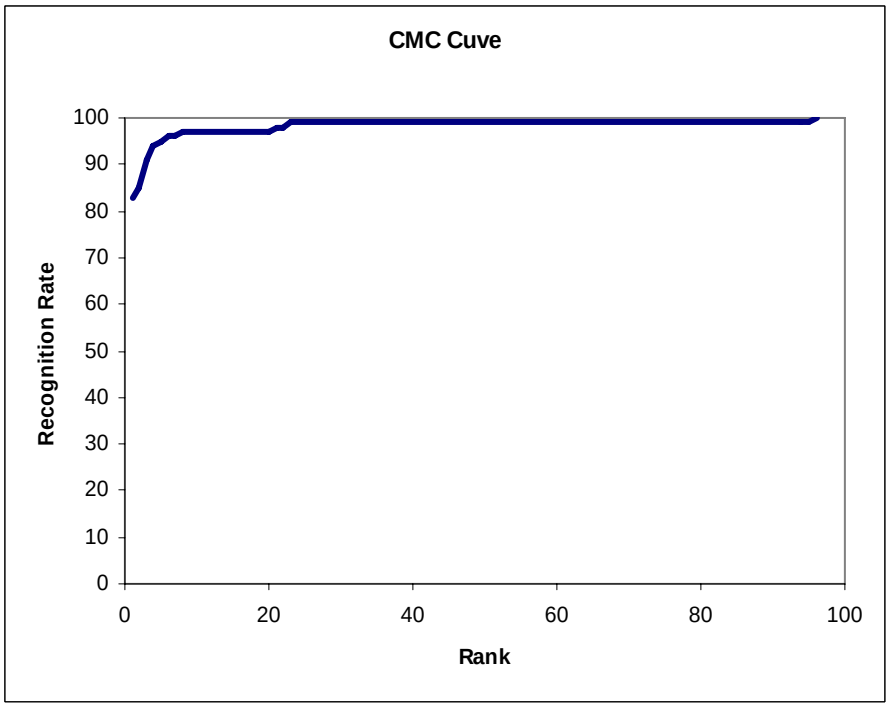

Figure AII.11 CMC curve for 3 non-uniform vertical segments experiment

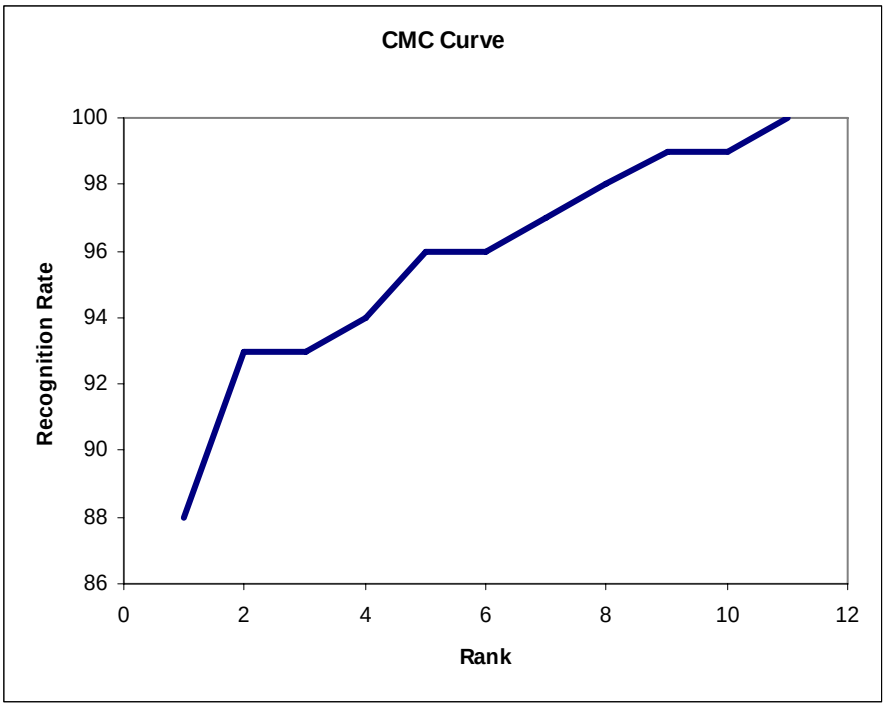

Figure AII.12 CMC curve for 4 (3 horizontal and 1 vertical) non-uniform segments experiment

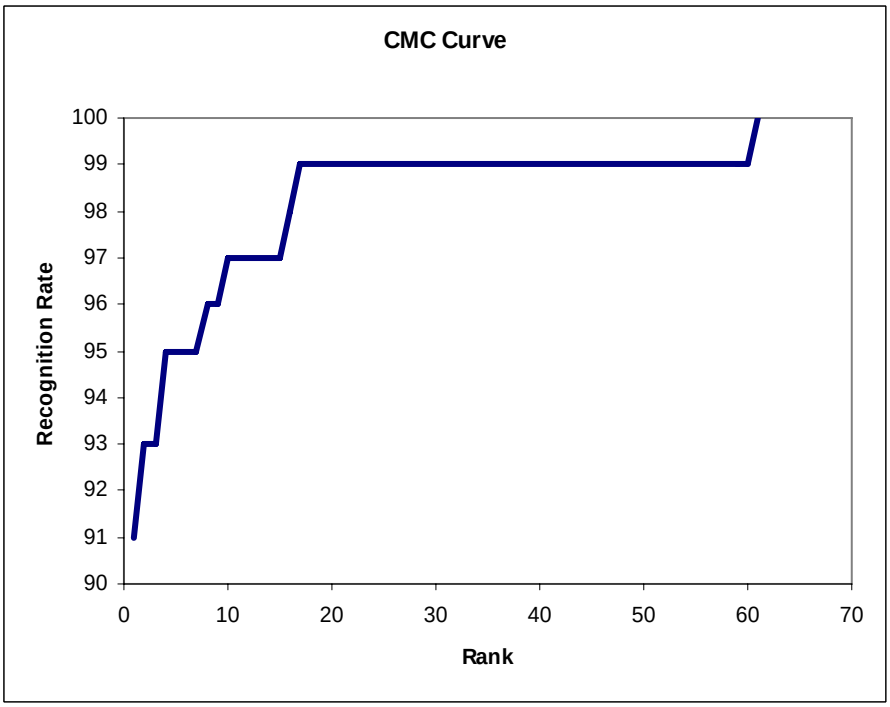

Figure AII.13 CMC curve for 3 (2 horizontal and 1 vertical) non-uniform segments experiment

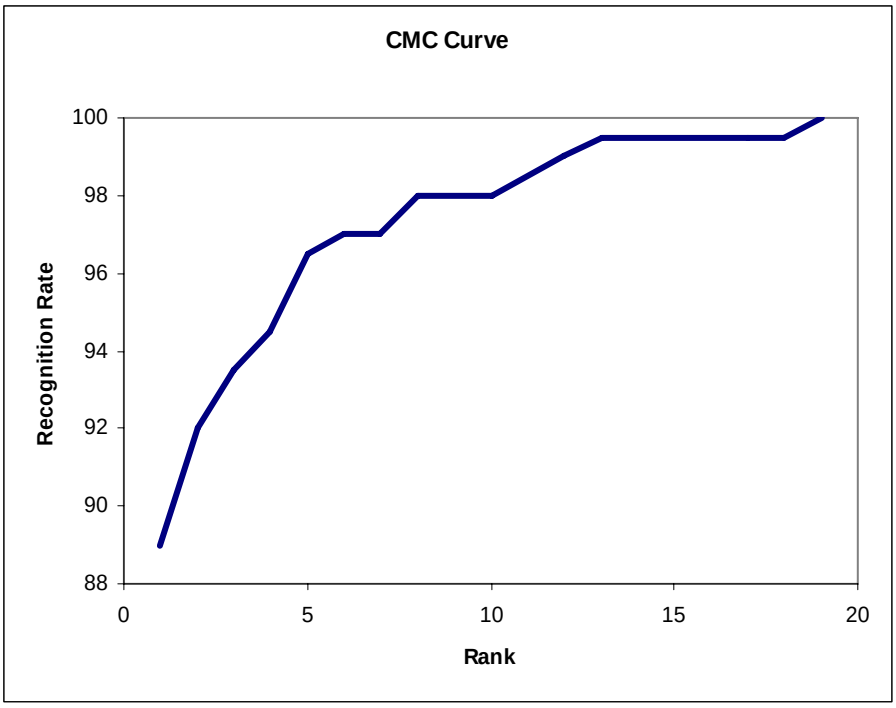

Figure AII.14 CMC curve for 1 circular segment experiment

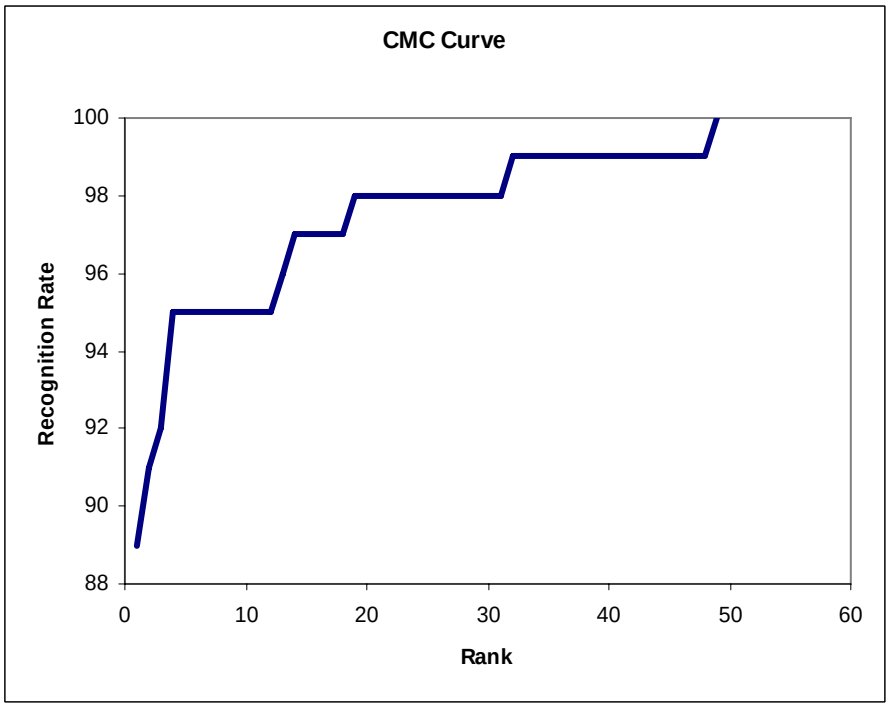

Figure AII.15 CMC curve for threshold segmentation experiment

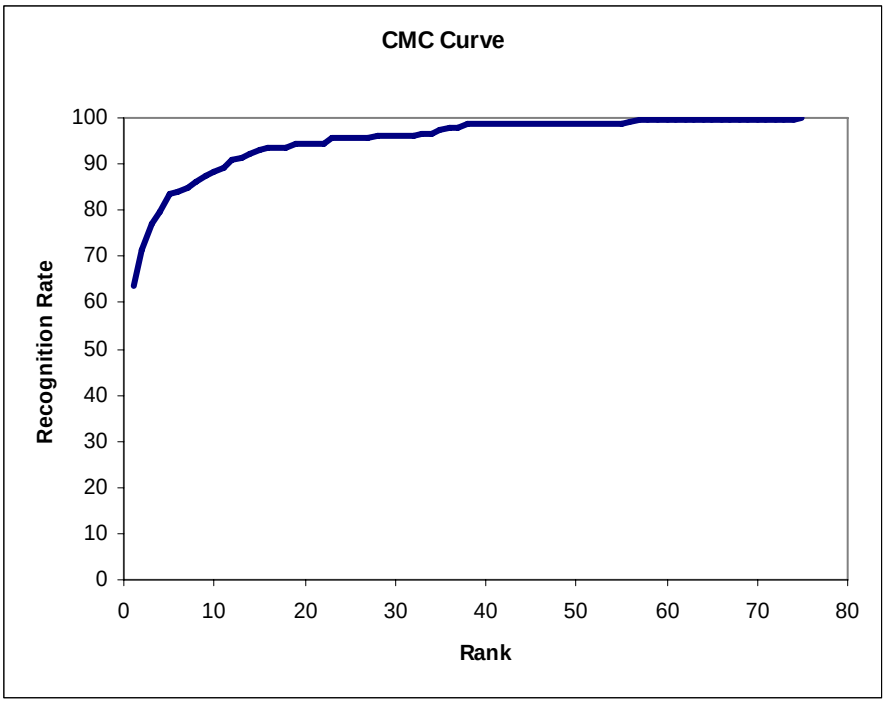

Figure AII.16 CMC curve base eigen-face experiment

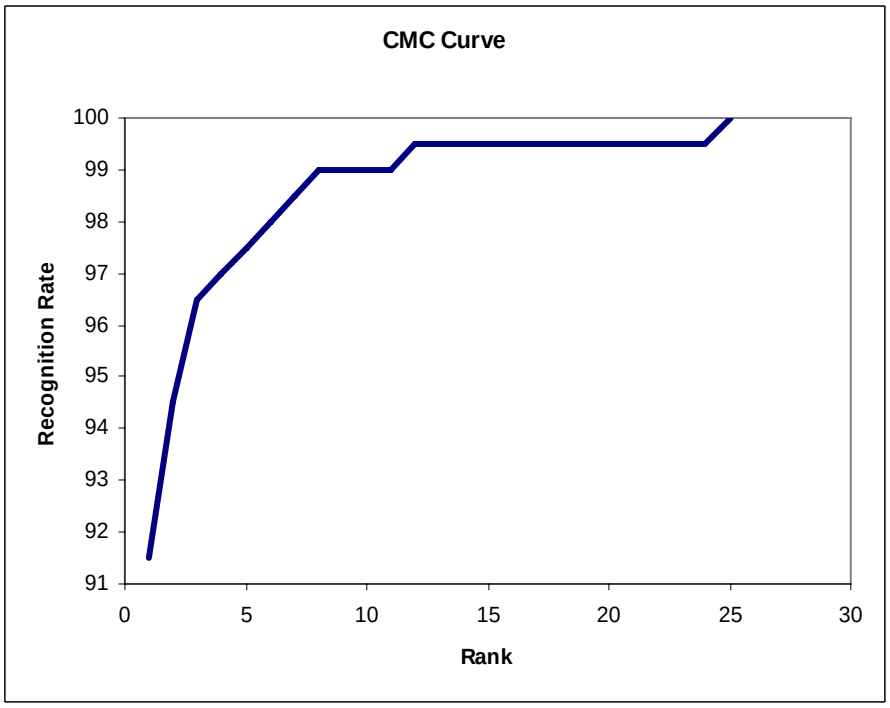

Figure AII.17 CMC curve for multimodal face-ear pre PCA version 1 experiment

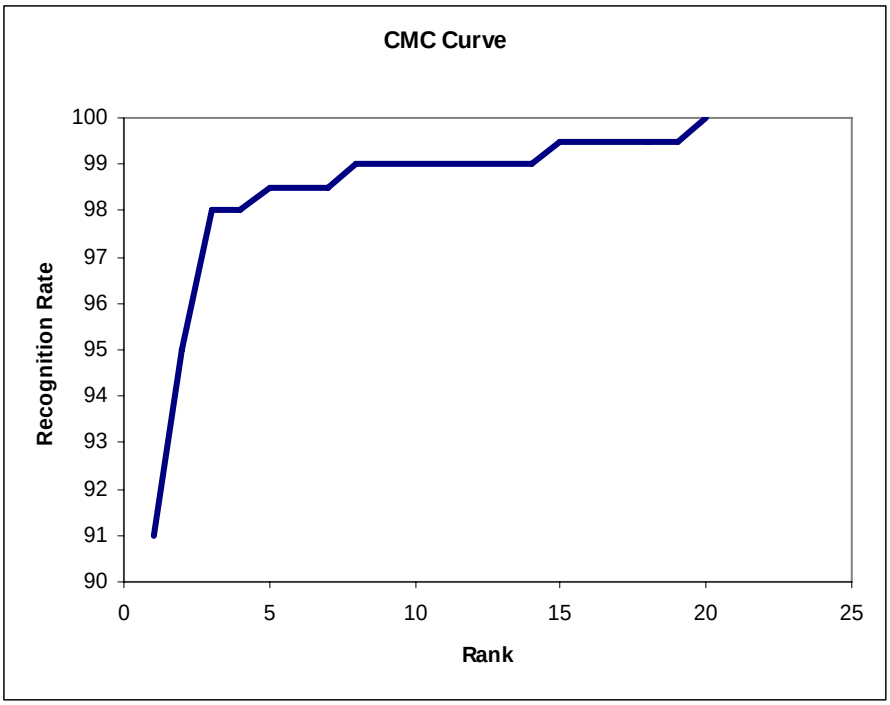

Figure AII.18 CMC curve for multimodal face-ear pre PCA version 2 experiment

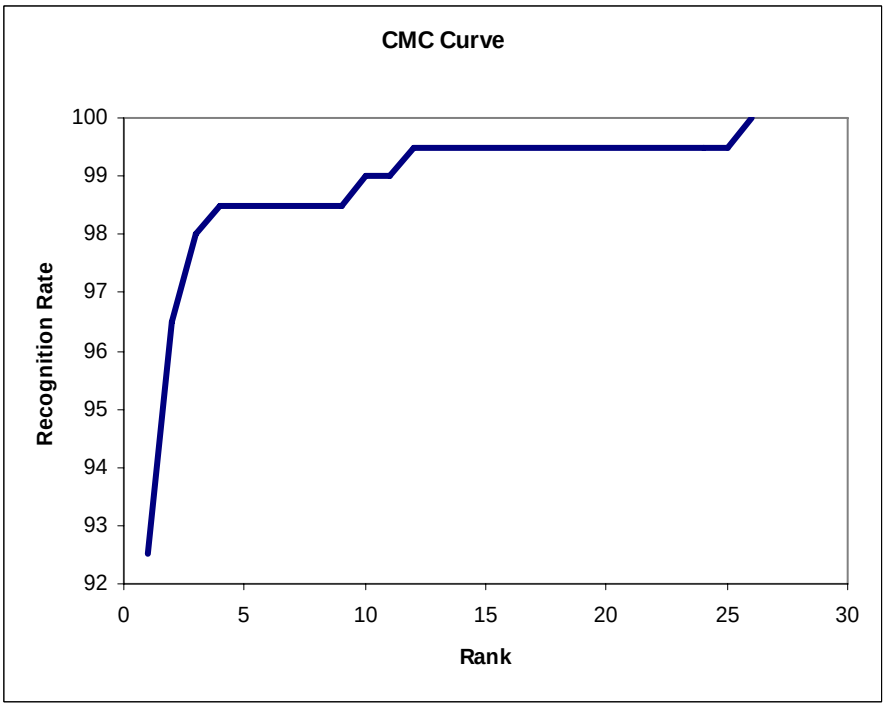

Figure AII.19 CMC curve for multimodal face-ear post PCA experiment

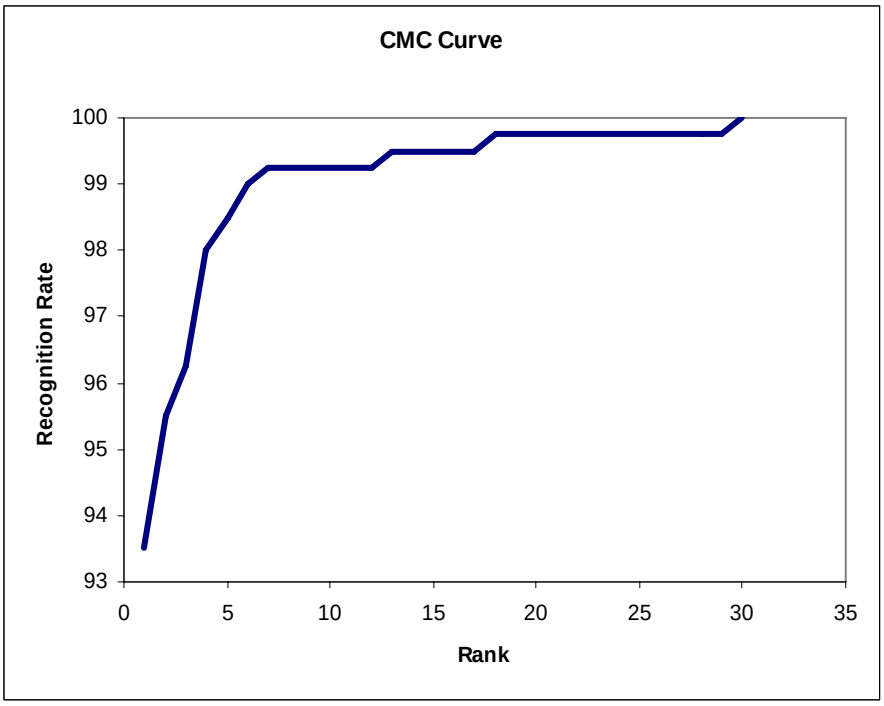

Figure AII.20 CMC curve for multimodal face-ear post PCA with permutation experiment

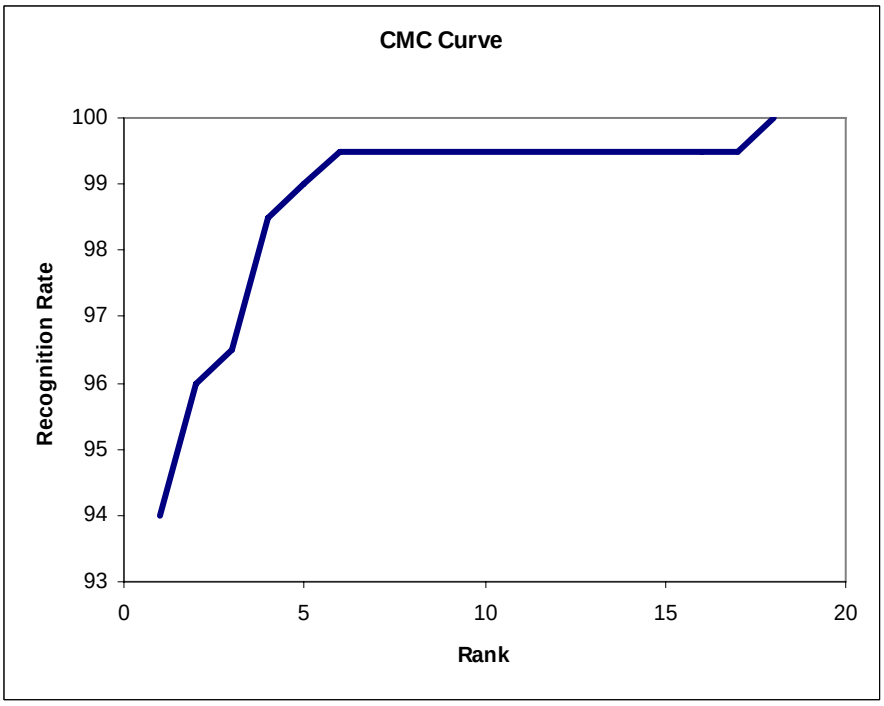

Figure AII.21 CMC curve for multimodal face-ear post PCA with permutation for Training set only experiment

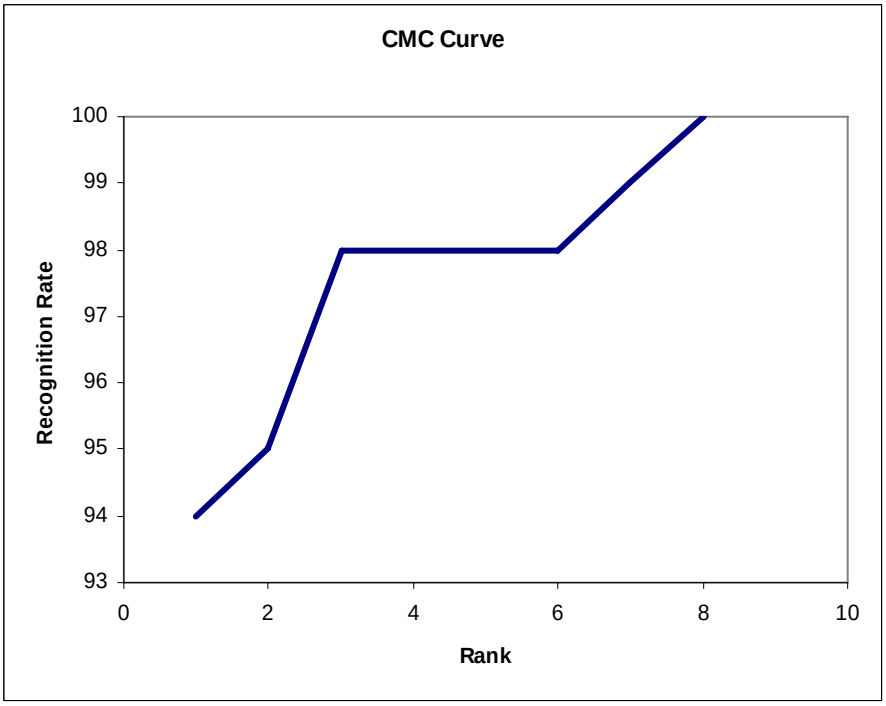

Figure AII.22 CMC curve for multimodal 1 circular segment and 4p segments pre PCA experiment

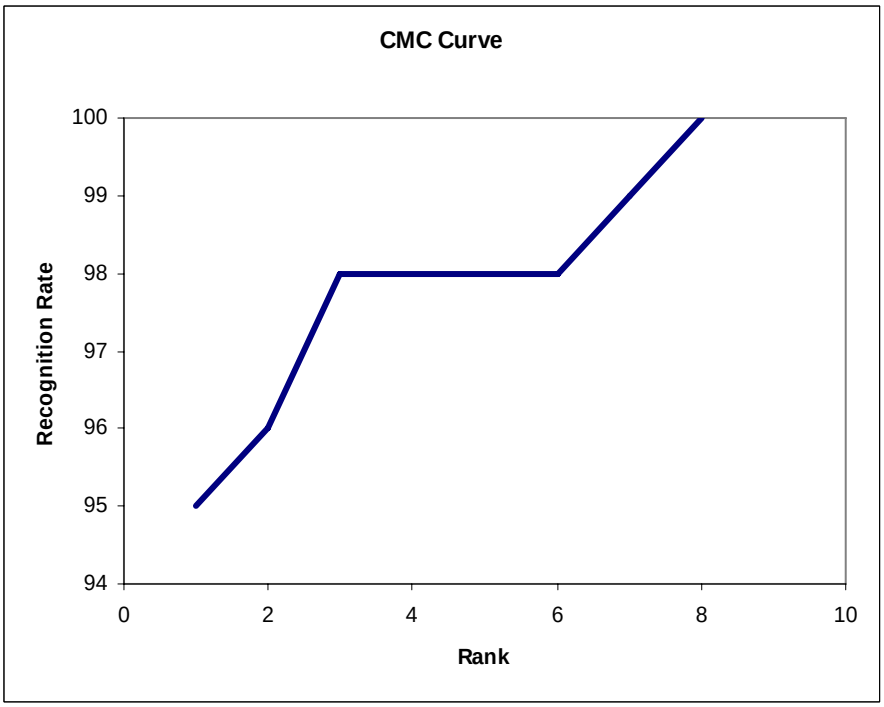

Figure AII.23 CMC curve for multimodal 1 circular segment and 4p segments post PCA experiment

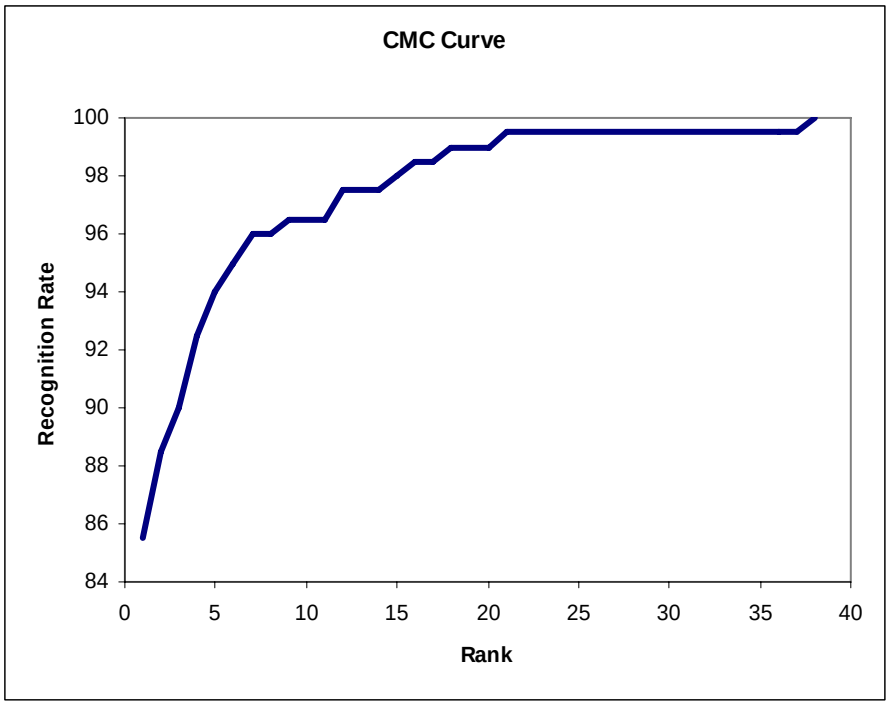

Figure AII.24 CMC curve for multimodal face and 1 circular segment pre PCA experiment

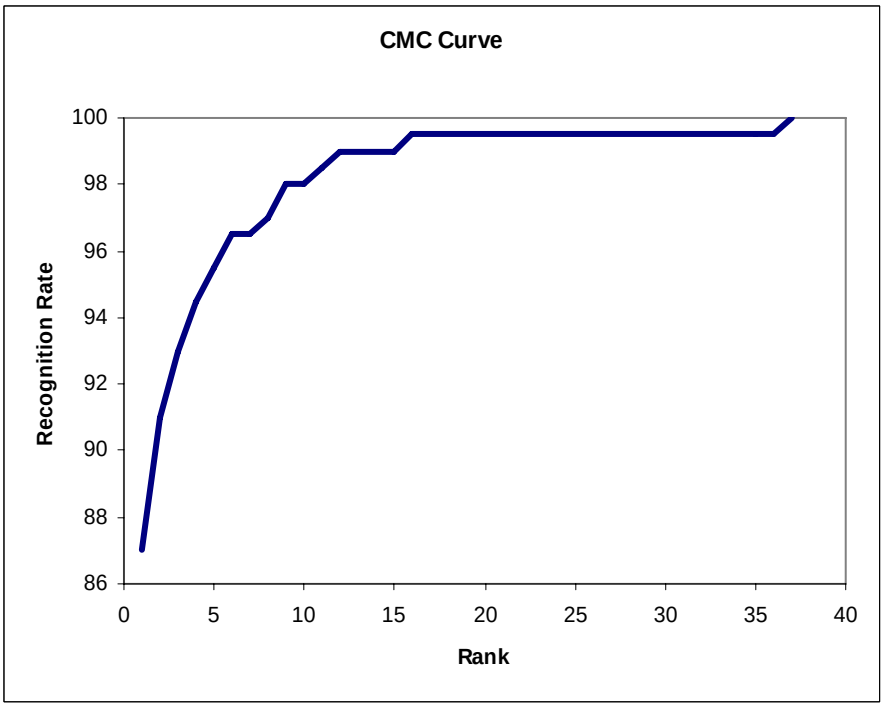

Figure AII.25 CMC curve for multimodal face and 1 circular segment post PCA experiment

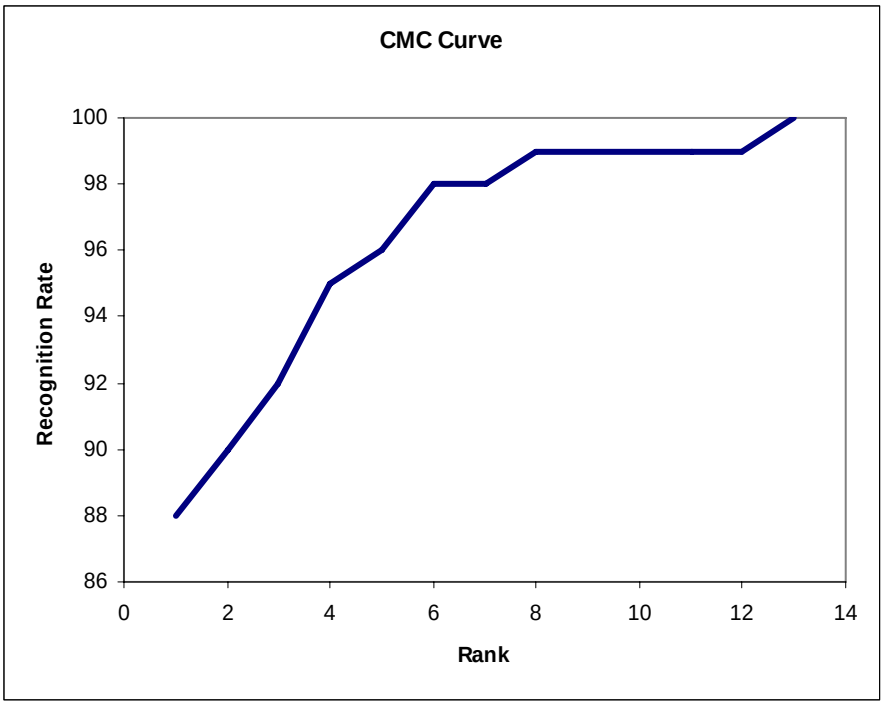

Figure AII.26 CMC curve for multimodal face and 2 combined segmentation methods pre PCA experiment

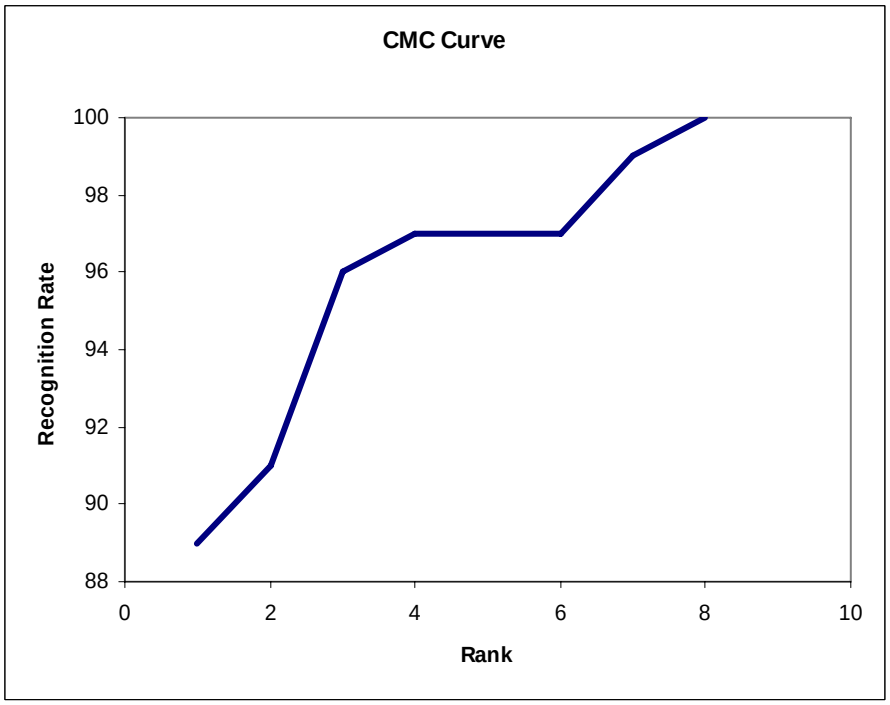

Figure AII.27 CMC curve for multimodal face and 2 combined segmentation methods post PCA experiment