Towards building a Practical Face Recognition System

by

Wasiuddin A.S.M. Wahid

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Bachelor of Science in Electrical Engineering and Computer Science

and

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1997

Author .....
Department of Electrical Engineering and Computer Science
May 31, 1997

Certified by .....
Alex P. Pentland
Toshiba Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Theses

Towards building a Practical Face Recognition System

by

Wasiuddin A.S.M. Wahid

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 1997, in partial fulfillment of the
requirements for the degrees of
Bachelor of Science in Electrical Engineering and Computer Science
and
Master of Engineering in Electrical Engineering and Computer Science

Abstract

This thesis analyzes the various modeling techniques for face recognition that are available to us within the eigenface framework and experiments with different methods that can be used to match faces using eigenfaces. It presents a probabilistic approach to matching faces and demonstrates it's superiority over other methods. It also carries out comprehensive parameter exploration experiments that determine the optimal parameter values for recognition. Finally it lays down some foundation for future work for improvement in face detection wherein a similar Baysian framework can be implemented for detection.

Thesis Supervisor: Alex P. Pentland
Title: Toshiba Professor, Media Arts and Sciences

Acknowledgments

First of all, I would like to thank my parents for all their love, support and encouragement. My sister for being there for me always.

I would like to thank Sandy(Professor Alex Pentland) for giving me the opportunity to work in VISMOD and for allowing me to write a thesis under him. He was always very understanding especially during the last month which was a particularly difficult month for me. He was always available to talk and his advice has always been invaluable.

I would like to thank Baback Moghaddam for all his advice. Most of my work was based on Baback's work and without him, I probably would not have been able to write this thesis.

Thanks to Nuria for helping me to proofread the thesis at the very last moment. Thanks to Lee for constantly urging me to write this thesis(he even bribed me with a wonderful rubber ball).

Special thanks to Jeffrey Levine and Thomas Slowe for keeping me awake at the Lab late at night. Same goes to Kofi Fynn, Muayyad Quabbaj and Gillian Elcock for keeping me going when things looked bad.

Thanks to my academic advisor Professor James Roberge. Many thanks to Aaron Bobick and B.K.P. Horn for two very useful classes. Thanks to Anne Hunter for all the incredible amount of help.

I will take this opportunity to thank people who have helped me not only on my thesis but also throughout my general MIT career.

I want to thank everyone in VISMOD for a great time here, Thanks Sumit for talking about my thesis and helping me go through all the stress with life. Thanks Tony for inspiring me with the guitar. Thanks Chris for all the help Friday night. Thanks Karen for Kyoto. Thanks Caraway for being so understanding. Thanks Ifung for being around. Kate for being so nice. Yuri for being the way he is. Thanks to Thad for all the advice. Thomas Minka for always answering any questions I had. Ken for all his perspectives on life. Thanks Erik. Thank you Dave for tolerating my weirdness. Flavia. Deb. All the Pentlandians. Thanks to Martin Szummer, Kris Popat for all the intellectually stimulating discussions.

A very very special thank you to all my friends here without whom I wouldn't have been able to make it through MIT: Dudzai, who's always been around. Dragan who's always been an inspiration. Kofi and Gillian: 170 is nearly as thick as blood. Thank you so much Lila and Muayyad - for everything. Abhijit for all the inspiring "adda". Jeff for keeping me sane. Tom for cheering me up whenever I felt down. Katya for being the wonderful person she is. Alo for all her help. Yun who will always continue to awe me. Nuria, Sumit, Tony, Lee, for everything they have done for me. Arif Bhai for always talking to me. Thanks Sharon for being so nice. Lastly but not the least, thank you KT for being the very special person that you are.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, there has been considerable progress in the problem of face detection and recognition. The best results have been obtained for 2D view-based techniques based on either template matching or matching using *eigenfaces* - a template matching method using the Karhunen-Loeve transformation of a set of face pictures.

The idea of using eigenfaces was motivated by a technique developed by Sirovich and Kirby [16] for efficiently representing pictures of faces using principal component analysis(PCA). The scheme was later extended by Turk and Pentland [18] to the problem of automatic face recognition. Moghaddam and Pentland [11] devised an unsupervised method for visual learning based on density estimation in high-dimensional spaces using an eigenspace decomposition. This allowed them to improve head detection and allow a maximum likelihood technique for detecting heads, thus increasing recognition accuracy.

In this thesis, we further extend the system by incorporating a Bayesian similarity measure for image matching. We analyze different modeling techniques that are available to us within the eigenface framework. We also carry out numerous experiments to determine the effect of different parameter values on recognition. Finally we integrate all these techniques to build a practical face recognition system that can be used to recognize faces in large databases such as the U.S. Army Research Laboratory's FERET database.

1.1 Organization of this thesis

This thesis is structured in the following format:

Chapter 2 introduces the problem of face recognition and detection using eigenfaces. It discusses the previous work done at the Media Lab on this subject. It talks about some of the problems with the previous system at the Media Lab and how this thesis addresses them.

Chapter 3 discusses different experiments carried out to determine the optimal parameters required to build a reliable face recognition system using eigenfaces.

Chapter 4 studies different modeling techniques that are available to us within the eigenface framework. It provides a detailed study of the performance of each approach and discusses their relative merits

Chapter 5 presents our final system and its performance on the ARL FERET tests administered in September 1996. It compares the performance of the final system with the system used for the FERET I tests held in March 1995.

Chapter 6 discusses the future direction of this work and the limitations of this thesis.

Chapter 2

Background

This chapter describes previous work in face recognition and detection. It describes the system that was developed in the Media Lab and discusses some of the problems that it had which we tried to address in this thesis.

2.1 What are Eigenfaces?

The idea of using eigenfaces was motivated by a technique developed by Sirovich and Kirby [16] for efficiently representing pictures of faces using principal component analysis(PCA). Starting with an ensemble of original face images, they calculated a best coordinate system for image compression, where each coordinate is actually an image that they termed an *eigenpicture*. They argued that any collection of face images can be approximately reconstructed by storing a small collection of weights for each face and a small set of standard pictures (the eigenpictures). The weights describing each face are found by projecting the face image onto each eigenpicture.

The scheme was later extended by Turk and Pentland [18] to the problem of automatic face recognition. They reasoned that if a multitude of face images can be reconstructed using a small collection of eigenpictures and weights, then an efficient way to learn and recognize faces would be to build up the characteristic features by experience over time and recognize particular faces by comparing the feature weights with the weights associated with a known individual. Each individual, therefore would

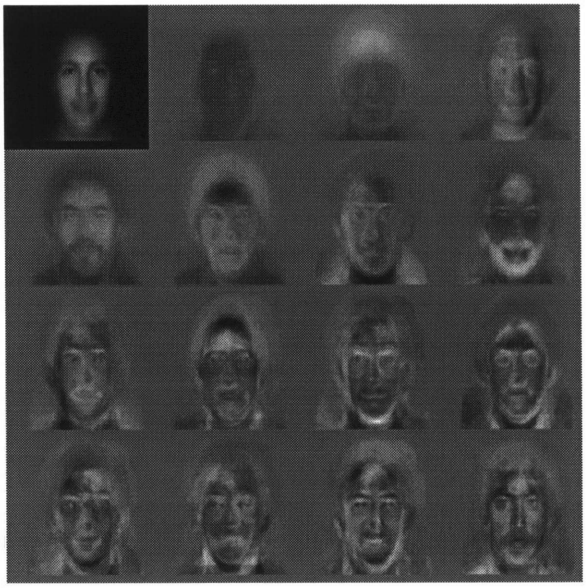


Figure 2-1: Some of the first few eigenfaces computed from a large database

be characterized by the small set of feature or eigenpicture weights needed to describe them.

In the language of information theory, we want to extract the relevant information in a face image, encode it as efficiently as possible and compare one face encoding with a database of models encoded similarly. A simple approach to extracting the information contained in an image of a face is to somehow capture the variations in a collection of face images, independent of any judgment of features and compare individual face images.

In mathematical terms, we wish to find the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of face images, treating an image as a point (or vector) in a very high dimensional space. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images.

These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image location contributes more or less to each eigenvector, so we can display the eigenvector as a sort of ghostly face which we call an *eigenface*. Some of the eigenfaces are shown in the figure 2-1.

Each individual face can be represented exactly in terms of a linear combination

of these eigenfaces, however normally the set of faces is approximated using only the "best" eigenfaces - those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images.

2.2 Calculating Eigenfaces

Given a set of m-by-n images, $\{I^t\}_{t=1}^{t=N_T}$, we can form a training set of vectors $\{\mathbf{x}^t\}$, where $\mathbf{x} \in \Re^{N=mn}$, by lexicographic ordering of the pixel elements of each image $\mathbf{I}^t$. The basis functions in a Karhunen-Loeve Transform(KLT) [6] are obtained by solving the eigenvalue problem

$$\Lambda = \Phi^T \Sigma \Phi \tag{2.1}$$

where Σ is the covariance matrix of the data, Φ is the eigenvector matrix of Σ and Λ is the corresponding diagonal matrix of eigenvalues. In PCA, a partial KLT is performed to identify the largest-eigenvalue eigenvectors and obtain a principal component feature vector $\mathbf{y} = \Phi_M^T \tilde{\mathbf{x}}$ where $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$ is the mean-normalized image vector and Φ_M is a submatrix of Φ containing the principal eigenvectors. PCA can be seen as a linear transformation which extracts a lower-dimensional subspace of the KL basis corresponding to the maximal eigenvalues.

2.3 Detecting Heads in an image

Moghaddam and Pentland [11] devised a method for estimating probability densities in a high-dimensional space using eigenspace decomposition. Their method is outlined in the appendix. The following section is also from that paper.

The density estimate $\hat{P}(\mathbf{x}|\Omega)$ can be used to form probabilistic *saliency* maps for target detection. This done by computing the likelihood estimate at each spatial location in the input image

$$S(i,j;\Omega) = \hat{P}(\mathbf{x}^{ij}|\Omega) \tag{2.2}$$

where $\mathbf{x}^{ij}$ is the observation vector obtained by vectorizing the local subimage centered at location (i, j) in the input image. The maximum likelihood estimate of the spatial position of the target is then obtained by

$$(i^*, j^*) = argmax_{i,j} S(i, j; \Omega) \tag{2.3}$$

Thus each test region in the image is projected onto the eigenspace and then tested to see if it is a head or not. This method is repeated for different scales and the head with the highest maximum likelihood score is selected.

Once the head has been selected, a similar search is carried out for the eyes, nose and mouth and all the scores from the searches are combined to provide a better estimate of the location of the head.

2.4 Recognizing Faces using Eigenfaces

The approach to face recognition involves the following initialization operations:

1. Acquire an initial set of face images (the training set).

2. Calculate the eigenfaces from the training set, keeping only the M images that correspond to the highest eigenvalues. These M images define the the *face space.*

3. Calculate the corresponding distribution in M-dimensional weight space for each known individual by projecting their face images onto the "face space". So each face image is now converted to a set of coefficients. In the original system, there were 100 eigenfaces and hence there were 100 coefficients.

Having initialized the system, the following steps are then used for recognizing face images.

1. Project each input image (which is a face already normalized for scale and contrast) on the M eigenfaces and determine the projection coefficients.

2. If it is a face, classify the weight pattern as either a known person or as unknown. This is done by a nearest neighbor method, by finding the Euclidean distance between the input face coefficients and all the other projected known faces in face space.

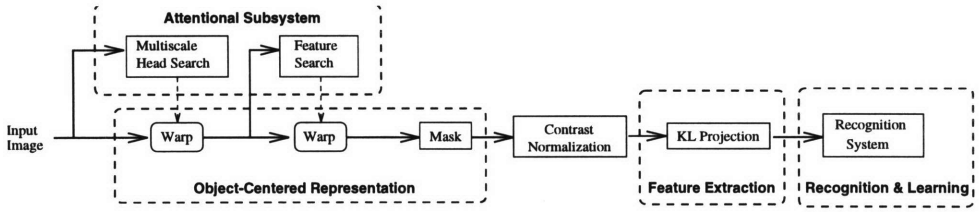


Figure 2-2: The face processing system.

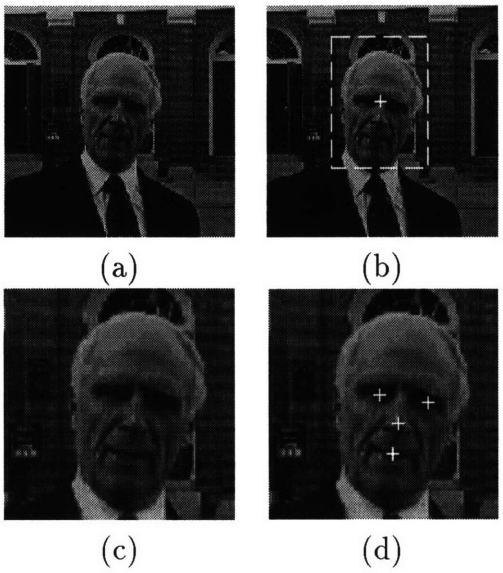


Figure 2-3: (a) original image, (b) position and scale estimate, (c) normalized head image, (d) position of facial features.

2.5 The Media Lab Face Detection and Recognition System

This section is from [11]. The block diagram of the system is shown in figure 2-2 which consists of a two-stage object detection and alignment stage, a contrast normalization stage, and a feature extraction stage whose output is used for both recognition and coding.

Figure 2-3 shows the operation of the detection and alignment stage on a natural test image containing a human face.

The first step in this process is illustrated in Figure 2-3(b) where the ML (maximum likelihood) estimate of the position and scale of the face are indicated by the

(a) (b) (c)

Figure 2-4: (a) aligned face, (b) eigenspace reconstruction (85 bytes)(c) JPEG reconstruction (530 bytes).

Figure 2-5: The first 8 eigenfaces.

cross-hairs and bounding box. Once these regions have been identified, the estimated scale and position are used to normalize for translation and scale, yielding a standard "head-in-the-box" format image (Figure 2-3(c)). A second feature detection stage operates at this fixed scale to estimate the position of 4 facial features: the left and right eyes, the tip of the nose and the center of the mouth(Figure 2-3(d)). Once the facial features have been detected, the face image is warped to align the geometry and shape of the face with that of a canonical model. Then the facial region is extracted (by applying a fixed mask) and subsequently normalized for contrast. The geometrically aligned and normalized image (shown in Figure 2-4(a)) is then projected onto a custom set of eigenfaces to obtain a feature vector which is then used for recognition purposes as well as facial image coding.

Figure 2-4 shows the normalized facial image extracted from Figure 2-3(d), its reconstruction using a 100-dimensional eigenspace representation (requiring only 85 bytes to encode) and a comparable non-parametric reconstruction obtained using a standard transform-coding approach for image compression (requiring 530 bytes to encode). This example illustrates that the eigenface representation used for recognition is also an effective *model-based* representation for data compression. The first 8 eigenfaces used for this representation are shown in Figure 2-5.

Figure 2-6: Photobook: FERET face database.

Figure 2-6 shows the results of a similarity search in an image database tool called Photobook [14]. Each face in the database was automatically detected and aligned by the face processing system in Figure 2-2. The normalized faces were then projected onto a 100-dimensional eigenspace. The image in the upper left is the one searched on and the remainder are the ranked nearest neighbors in the FERET database. The top three matches in this case are images of the same person taken a month apart and at different scales. The recognition accuracy (defined as the percent correct rank-one matches) on a database of 155 individuals is 99% [10].

2.6 Problems with the Media Lab face recognition system

The results of the Media Lab face recognition system in two categories of the FERETI tests are summarized in the tables 2.1 and 2.2. These results suggested that the performance of the Media Lab system could be improved substantially. (It should be mentioned here that the tests were adminstered in March 1995 and was taken in November 1995 by Rockefeller).

There are a number of ways that the Media Lab face recognition system can be improved. (A brief description of the Rockefeller and USC methods are given later in this chapter).

Table 2.1: FA vs FB results on the FERET I tests

Institution	Recognition Rate
Rockefeller (November 1995)	96
USC (March 1995)	92
MIT Media Lab (March 1995)	88

Table 2.2: Duplicate Scores on the FERET I tests

Institution	Recognition Rate
Rockefeller (November 1995)	62
USC (March 1995)	58
MIT Media Lab (March 1995)	40

1. The head detection scheme can be improved by incorporating a training system that learns negative examples of heads and uses that within a Bayesian framework to detect heads more accurately.

2. The present face matching system uses a nearest neighbor approach with a Euclidean distance metric within the face space. This has been criticized in the past.

Critics have argued that the eigenface approach takes into consideration the *most expressive features* for classifying faces. This according to them is inferior to the approach using *most discriminating eigenfeatures.* Basically what this means is that in the Euclidean distance metric, you weigh all the projection coefficients equally. This is not the best approach since some eigenfaces capture the most representative axis among faces but these eigenfaces are not necessarily the best ones for classification.

3. Other important and often ignored aspect of all face recognition systems is the effect of different parameters on recognition. Parameters like the number of eigenfaces used, the effect of masks, the effect of using different regions of the faces and the effect of different databases for training. Tweaking of these parameters can help build a practical and very robust face recognition from the already existing one.

2.7 Other work on Face Recognition

There has been an incredible amount of work done in face recognition in recent years. In this section, we try to give a very brief description of some of the work done in this field (though it is impossible to mention all the work).

The Rockefeller system developed at the Laboratory of Computational Neuroscience has developed factorial coding into a mathematical theory for visual processing. They report that their model has been found to account quantitatively for much of the observed spatio-temporal and color coding properties of neurons. In addition they claim that the theory has been used to derive "learning algorithms" which allow a computer to generate factorial codes for any ensemble of complex images such as faces. [1]

The University of Southern California presents a method for recognizing objects (faces) on the basis of just one stored view, inspite of rotation in depth. It is not based on the construction of a three-dimensional model for the object. This achieved with the help of a simple assumption about the transformation of local feature vectors with rotation in depth. The parameters of this transformation are learned on training examples. This was the abstract in a paper by the developers of the system, Christroph Malsburg and Thomas Maurer [8].

John Weng also worked on a discriminant eigenfeatures for recognition, which uses a version of the Fisher discriminant for classifying between clusters of different people [17].

Among other significant work include that of Tom Poggio at MIT and Rama Challappa at University of Maryland.

2.8 Goals of this thesis

This thesis tries to address the problems discussed in the previous section. It tries to build a practical and robust face recognition system for improved accuracy.

It studies different modeling techniques for matching faces within the eigenface

framework and tries to come up with different methods that can appropriately weigh the different eigenfaces for better matching. It ultimately comes up with a Bayesian approach to matching that converts the face recognition problem into a classical binary classification problem where faces are characterized as a known person or an unknown person according to the *a posterior* probabilities associated with clustering of the database into *intra* and *extra* personal groups.

It also carries out comprehensive parameter exploration experiments that help determine optimal parameter values for recognition.

Finally it lays down some foundation for future work for improvements in face detection wherein a similar Bayesian framework can be implemented for detection. In such a system, positive and negative examples of faces can be incorporated into the training system and the system can automatically learn and classify heads and non-heads more accurately.

Chapter 3

Modeling Parameters

In recent years, there has been considerable progress made in the problems of face detection and recognition using eigenfaces. However there has not been too much effort spent on general parameter exploration. There are a lot of ways that the problem of face detection and recognition can be modeled within the eigenface framework. This chapter attempts to explore certain parameters such as what regions of the face are essential for recognition, how many eigenfaces are optimal for recognition, how to incorporate negative examples into training and how to improve performance on duplicate images. We will try to integrate all these modeling parameters and build a system that provides increase performance.

3.1 Number of Eigenfaces

An important parameter for recognition is the number of eigenfaces used for representation of the faces. The original eigenface technique of Turk and Pentland [18] used 7-10 eigenfaces. More recent work by Pentland *et al* [10] used 100 eigenfaces. Although Kirby and Sorvich [16] and others realized the importance of the number of eigenfaces on the performance of recognition, there has been no systematic effort to investigate exactly how many eigenfaces are optimal for recognition.

One reason for using fewer eigenfaces is computational efficiency, however an even more important reason is *generalization.* A classifier based on all the information

within a training set is inevitably able to to use random correlations specific to that training set to improve classification; e.g., if all the males in the database had beards, it is relatively easy to discriminate males from females for that training set, but such a rule does not work in general. By restricting ourselves to the large variance eigenvectors, we experimentally find much better generalization to other datasets. This is well-known in applications such as image coding and neural networks [7].

In the next section, we describe a few experiments that will help us determine the optimal number of eigenfaces needed for recognition.

3.1.1 Varying the number of Eigenfaces

For these experiments, we used a 1274 face database (which we called the bank). This database included several images of the same person under different pose, lighting condition and facial expression. The database was divided into two parts - the half bank and the reverse half bank. In all the experiments, we describe in this chapter, we use the intra/extra modeling method (also discussed in this thesis).

Our first experiment was to train on the full bank and then use that training information to test on the full bank itself. We tried this for different number of eigenfaces ranging from 50 to 500 eigenfaces at regular intervals of 25. The results of our experiments are summarized in the figure 3-1 and in table 3.1.

These results show that increasing the number of eigenfaces increases recognition when you test on the training set. There are however a few points that one should notice while analyzing the results: We can see that recognition increases with the number of eigenfaces but only to a certain point. Thus the performance using 200 eigenfaces are the same as the performance using 500 eigenfaces. Therefore there is a limit above which using more eigenfaces does not help recognition. The very high recognition obtained in this test must be considered with caution because we are testing on the training set. Issues of generalization and over-fitting must be considered.

A much better experiment is to train on the half-bank database, test on the reverse half-bank database and vice versa and analyze the results. The results are

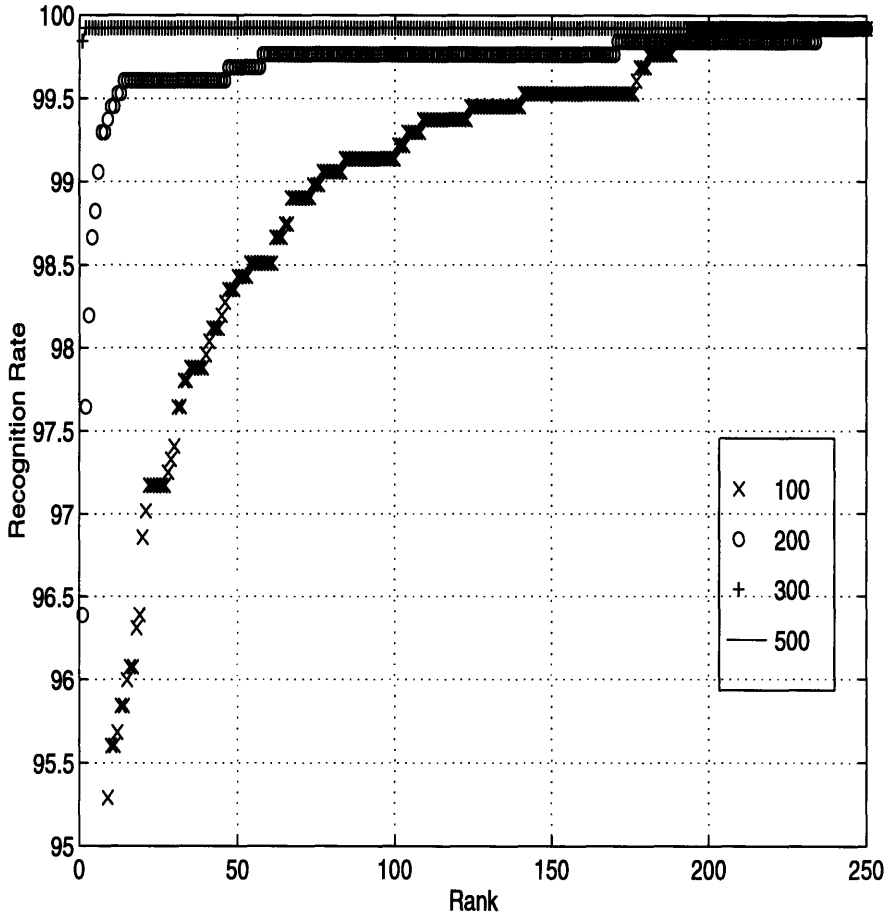


Figure 3-1: Testing on Training set while varying the number of eigenfaces used

Table 3.1: Recognition rates for different numbers of eigenfaces used. Testing on training data

Number of Eigenfaces Used	Recognition Rate
75	93.87 (1196/1274)
100	98.12 (1250/1274)
125	98.98 (1261/1274)
150	99.61 (1269/1274)
175	99.69 (1270/1274)
200	99.84 (1272/1274)
225	99.84 (1272/1274)
250	99.84 (1272/1274)
300	99.84 (1272/1274)
400	99.84 (1272/1274)
500	99.84 (1272/1274)

Table 3.2: Training on reverse half bank and testing on half bank while varying the number of eigenfaces

Number of Eigenfaces Used	Recognition Rate
100	93.89 (599/638)
125	94.20 (601/638)
150	93.10 (594/638)
175	92.63 (591/638)
200	92.00 (587/638)
300	90.28 (576/638)
500	76.65 (489/638)

summarized in the following two figures 3-2 and 3-3 and in the tables 3.2 and 3.3:

These experiments give us very interesting insights. We can see that the recognition rate increases with the number of eigenfaces used but only till a certain point. After that point, the recognition rate falls pretty dramatically. For our database, the recognition rate seems to be best when we use 125 eigenfaces.

It is interesting to compare the performance of the 500 eigenface method in the two tests. It performs admirably on the first test but fails in the next tests. This supports our speculation on over-fitting and generalization to different databases.

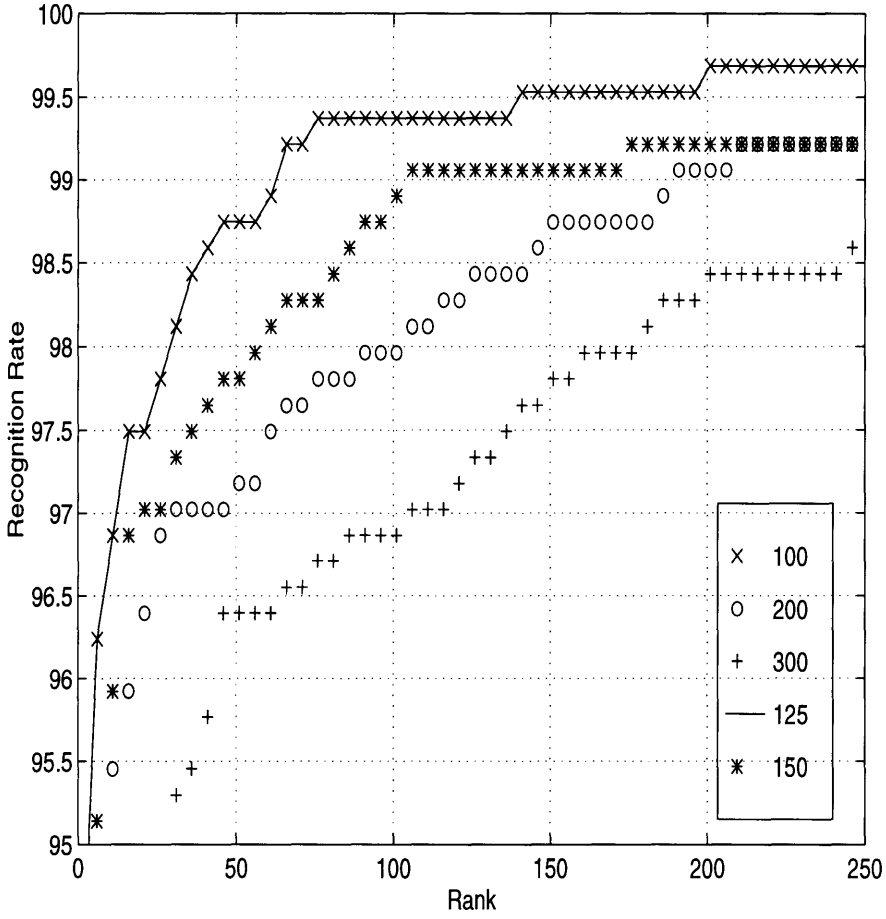


Figure 3-2: Training on reverse half bank and testing on half bank for different numbers of eigenfaces

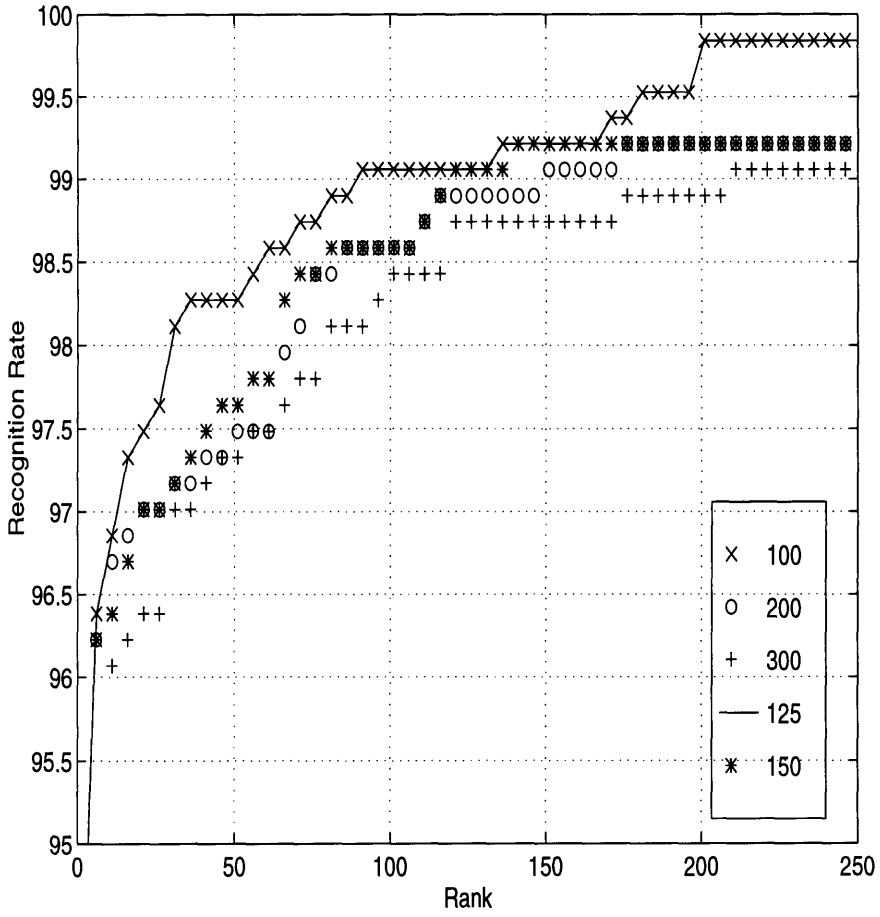


Figure 3-3: Training on half bank and testing on reverse half bank for different numbers of eigenfaces

Table 3.3: Training on half bank and testing on reverse half bank while varying the number of eigenfaces

Number of Eigenfaces Used	Recognition Rate
100	93.71 (596/636)
125	93.87 (597/636)
150	94.02 (598/636)
175	92.76 (590/636)
200	92.77 (590/636)
300	91.04 (579/636)
400	88.21 (561/636)
500	80.19 (510/636)

3.2 Training with a larger set

One of the experiments we carried out was increasing our training from 300 images to 2000 images. So instead of forming the original eigenfaces from 300 images, we calculate the eigenfaces from a training set of 2000 images.

The results of increasing the training set are shown in figure 3-4. The results are as one would expect. We can see that the 2000 Bank eigenfaces show a substantial improvement in recognition using both the Euclidean distance metric as well as the Intra/Extra modeling method. The Intra/Extra modeling method in particular shows a dramatic improvement. This is because the 2000-bank eigenfaces span more of the relevant variations among faces then the 300-bank eigenfaces and this leads to improved performance.

The first twenty 2000-bank eigenfaces are shown in figure 3-5:

3.3 Investigating Which Regions of the Face are Important for Recognition

We carried out a few experiments to determine which regions of a face are important for face recognition. We designed different masks and tested them for recognition. The results of our experiments are summarized in the figure 3-6. These experiments

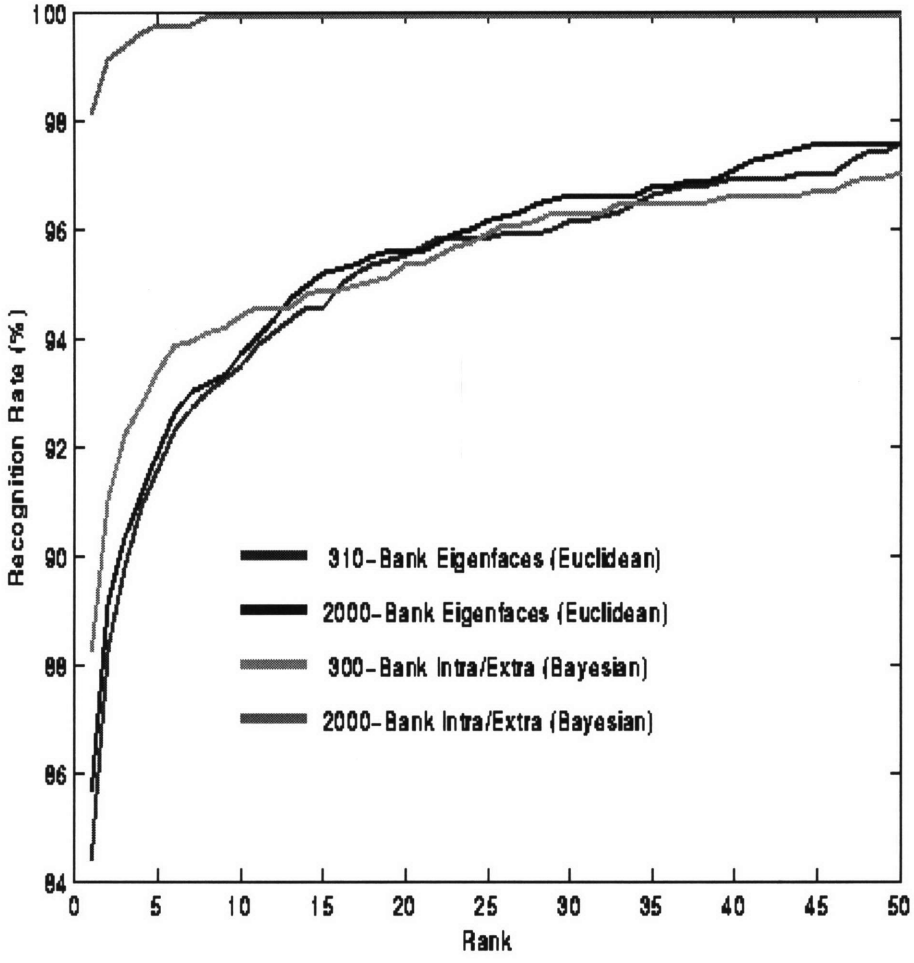


Figure 3-4: Performance of the training with 300-bank and training with 2000-bank compared

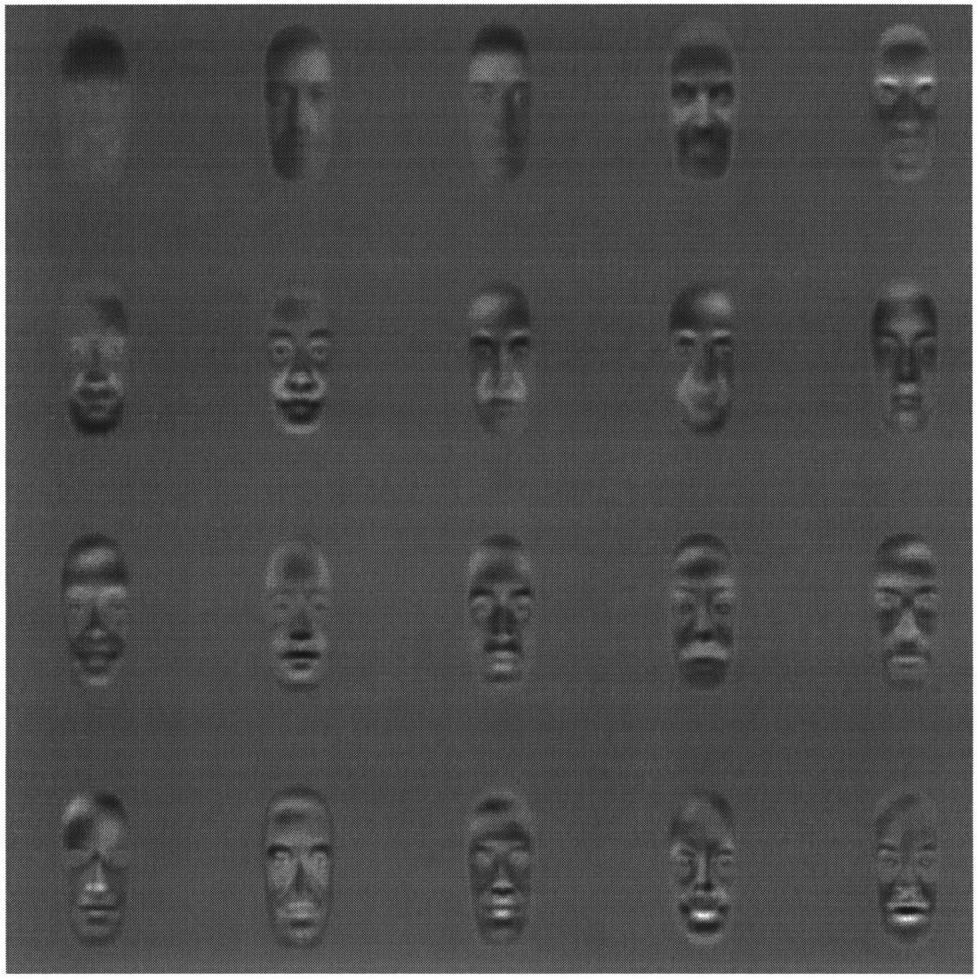


Figure 3-5: The first 20 eigenfaces built from a 2000 face database

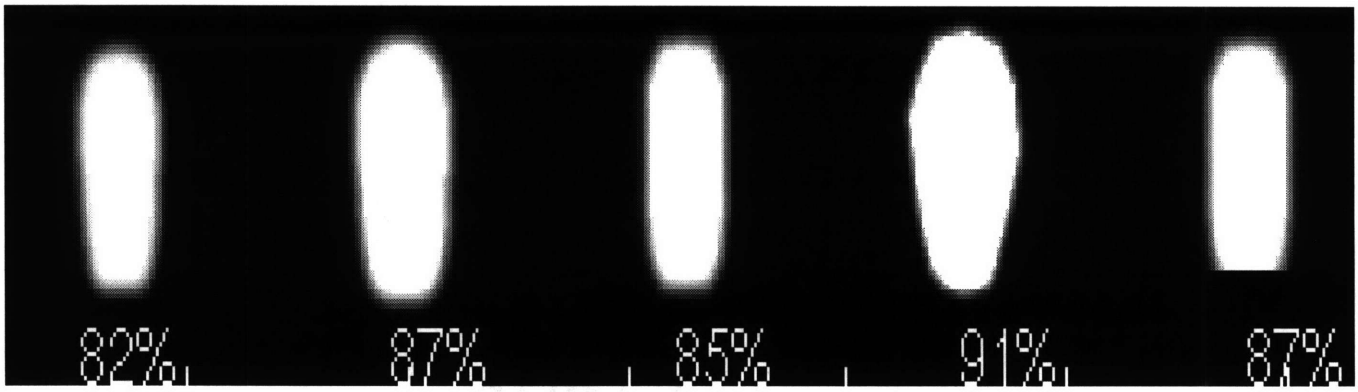


Figure 3-6: Different Masks and their performances

Figure 3-7: An image containing only the eyes

were carried out using the 300-bank eigenfaces and using the Euclidean distance metric.

The results suggests that a mask that allows some of the hair in an image performs better than the a tight-mask which is what was originally used in our system. This result is very suggestive and more experiments are needed to make further claims.

3.3.1 Using Only the Eyes

It is generally believed that over a period of many years, the eyes undergo very little change. We carried out an experiment to test how our system recognizes faces given just the eyes of a person. An example of the images that we fed into the system is given in figure 3-7.

The recognition results obtained using just the eyes is 91.05% which is remarkable considering how much less visual information the eye contains compared to the entire face. What would be interesting to test would be the performance of using only eyes on images taken years apart. Most recognition systems using the entire face fare very poorly on such images (referred to as *dupes*). This might be due to lack of sufficient training examples which represent the changes in a person's face over a period of time. However, if eyes do undergo very little change over time, then maybe recognition on

dupes can be improved substantially using only the eyes.

3.4 Summary

In this chapter, we presented numerous experiments that explore some of the parameters that affect our recognition system. We found out that the best recognition results are obtained with a 125-dimensional eigenspace. We found out that a large training sample gives better results and investigate which regions of the face are important for recognition.

Chapter 4

Recognition

Once every face has been projected onto facespace, we are ready to "recognize" or match them against each other. For every test image (which we will call the "probe") we match it against each face in our database (which we will call the "gallery") scoring it with an appropriate function and thresholding the end scores to actually recognize the image.

Each face is projected onto the 125 dimensional facespace. So at this stage our faces are transformed to 125 floats which are simply the projection coefficients.

There are a number of ways one can match different faces. In this chapter, some of the techniques for feature extraction and image matching are analyzed. The first few sections describe the different approaches while the latter sections describe a few experiments and analyze the performance of the different methods.

4.1 Euclidean distance metric

One can find the *euclidean* distance between the probe and all the gallery images and find the nearest neighbor to the test image. Thus if p and g are vectors representing the probe and gallery points, in a n-dimensional space, the Euclidean distance $d_E(x, y)$ between the two points is:

$$d_E(x,y) = |p-q| = [\sum_{i=1}^{n}(x_i - y_i)^2]^{1/2} \tag{4.1}$$

The probe is recognized by choosing the gallery image that is nearest to it.

This is perhaps one of the simplest methods for matching images. The probability of error for this nearest neighbor decision rule is bounded above by twice the Bayes probability of error [2] if there is an infinite number of samples. This simple measure of similarity is often used because it does not require a satisfactory estimation of the distribution function, which in many cases is impractical in a high dimensional space.

The obvious disadvantage with the Euclidean distance metric is that it does not take into consideration the underlying probability distribution of the faces. It thus attaches equal weights to all the projection coefficients.

4.2 Using Intra/Extra Personal Variations

The nearest-neighbour rule based on a Euclidean distance metric to match probe and gallery images ignores the underlying probability distribution and so cannot be expected to perform as well as a method that uses a Bayesian framework to classify faces.

The problem can be formulated in a Bayesian fashion as follows: Two distinct and mutually exclusive classes are defined: Ω_I representing the *intrapersonal* variations between multiple images of the same individual (e.g. with different expressions, poses and lighting conditions) and $Omega_E$ representing the *extrapersonal* variations among two different individuals.

Once we have characterized the probability distributions of the two different classes, we can try to classify a face as being in the intra or extra class according to some decision boundary.

Both classes are assumed to be Gaussian distributed and the estimates of the likelihood functions $P(\tilde{U}|\Omega_I)$ and $P(\tilde{U}|\Omega_E)$ are obtained for any given test face vector $\tilde{U}$

One difficulty with this approach is that face image vectors are very high-dimensional

with $\tilde{U} \in \Re^N$ where $N = O(10^3)$. Thus we typically lack sufficient training data to compute reliable 2nd-order statistics for the likelihood densities (i.e. singular covariance matrices will result). Even if we were able to estimate these statistics, the computational cost of evaluating the likelihoods is formidable. Also the computation would be inefficient since the *intrinsic* dimensionality or major degrees of freedom of $\tilde{U}$ for each class is most likely much smaller than N.

As explained in the beginning of this chapter, at this point all the faces have been projected onto the principal components or eigenfaces, so at this point we are dealing with with 125-dimensional vectors so we can get quite accurate 2nd-order statistics since our training set comprises about 2000 faces.

Once we have defined these mutually exclusive clusters of intra and extrapersonal variations, we can try to classify a face as belonging to one of those clusters.

4.2.1 Quadratic Classifiers

We will use a Bayesian likelihood ratio test for classification. The decision rule is of the form:

$$\ell(\tilde{y}) = \frac{p_{y|\omega_1}(\tilde{y}|\omega_1)}{p_{y|\omega_2}(\tilde{y}|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \lambda \tag{4.2}$$

Lets say that the vector y from the ith class has a mean m_i and a covariance K_i and are characterized by a Gaussian density, then we can rewrite equation 4.2 as

$$\ell(\tilde{y}) = \sqrt{\frac{|K1|}{|K2|}} exp[\frac{1}{2}(\tilde{y}-m_1)^T {K_1}^{-1}(\tilde{y}-m_1) + \frac{1}{2}(\tilde{y}-m_2)^T {K_2}^{-1}(\tilde{y}-m_2)] \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \lambda \tag{4.3}$$

which can be written as:

$$h(\tilde{y}) = (\tilde{y}-m_1)^T {K_1}^{-1}(\tilde{y}-m_1) - (\tilde{y}-m_2)^T {K_2}^{-1}(\tilde{y}-m_2) + ln\frac{|K_1|}{|K_2|} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} T \tag{4.4}$$

which can be expressed as follows:

$$h(\tilde{y}) = (\tilde{y})^T A\tilde{y} + b^T y + c \overset{\omega_1}{\underset{\omega_2}{\gtrless}} T \tag{4.5}$$

where

$$A = {K_1}^{-1} - {K_2}^{-1} \tag{4.6}$$

$$b = 2({K_2}^{-1}m_2 - {K_1}^{-1}m_1) \tag{4.7}$$

$$c = ({m_1}^T {K_1}^{-1} m_1 - {m_2}^T {K_2}^{-1} m_2 + ln\frac{|K_1|}{|K_2|}) \tag{4.8}$$

The decision rule 4.5 is called the Gaussian or quadratic classifier and it essentially the Mahalanobis distances to the mean of each class using the class covariance matrix and compares them to a threshold.

One can assume (incorrectly) that the class covariance matrices are diagonal and calculate the mahalanobis distance metric as follows

$$(x' - m^T)\Lambda^{-1}(x' - m') = C \tag{4.9}$$

However this decision boundary does not take into account the relative orientations of the two different clusters.

4.2.2 Linear Classifiers

If we assume that $K_1 = K_2 = K$, then the matrix A of equation 4.5 becomes equal to 0 and the 4.5 reduces to

$$h(\tilde{y}) = b^T + C \overset{\omega_1}{\underset{\omega_2}{\gtrless}} T \tag{4.10}$$

where

$$b = 2K^{-1}(m_2 - m_1) \tag{4.11}$$

$$c = {m_1}^T K^{-1} m_1 - {m_2}^T K^{-1} m_2 \tag{4.12}$$

This decision boundary is a linear hyperplane.

4.2.3 Intra/Extra Eigenvectors

Thus we can also classify the faces into the intra/extra clusters by first performing a separate PCA on each class, projecting each test faces on the intra/extra eigenvectors and using a MAP classification rule.

We can use an appropriately weighted sum of the *distance in intra/extra space* as a suitable score to classify the faces.

We then have:

$$\hat{P}(\tilde{U}|\Omega) = [\frac{exp(-\frac{1}{2}\sum_{i=1}^{M}\frac{(y_i)^2}{\lambda_i})}{(2\pi)^{(\frac{M}{2})}\prod_{i=1}^{M}(\lambda_i)^{(\frac{1}{2})}}] \tag{4.13}$$

where $P_F(\tilde{U})$ is the true marginal density in F. Here F is either the intra or the extra space. Since we can estimate the marginal densities, we can define the similarity score between a pair of images directly in terms of the intrapersonal *a posteriori* probability as given by Bayes rule:

$$P(\Omega_I|\tilde{U}) = \frac{P(\tilde{U}|\Omega_I)P(\Omega_I)}{P(\tilde{U}|\Omega_I)P(\Omega_I) + P(\tilde{U}|\Omega_E)P(\Omega_E)} \tag{4.14}$$

This method is equivalent to the quadratic classifier described earlier. This Bayesian formulation casts a face recognition task into a classical binary pattern classification problem which can then be solved using the maximum *a posteriori* (MAP) rule - i.e. two facial images are determined to belong to the same individual if $P(\Omega_I|\tilde{U}) > P(\Omega_E|\tilde{U})$.

An efficient density method was proposed by Moghaddam & Pentland [12]. They break down the vector space $\Re^N$ into two complementary subspaces using an eigenspace decomposition. A low dimensional estimate of the the probability distribution using only the first M principal components (where $M << N$) obtained by Principal Components Analysis (PCA) [4]. This is shown in Figure 4-1 which shows the orthogonal decomposition of the vector space $\Re^N$ to the principal subspace F comprising the

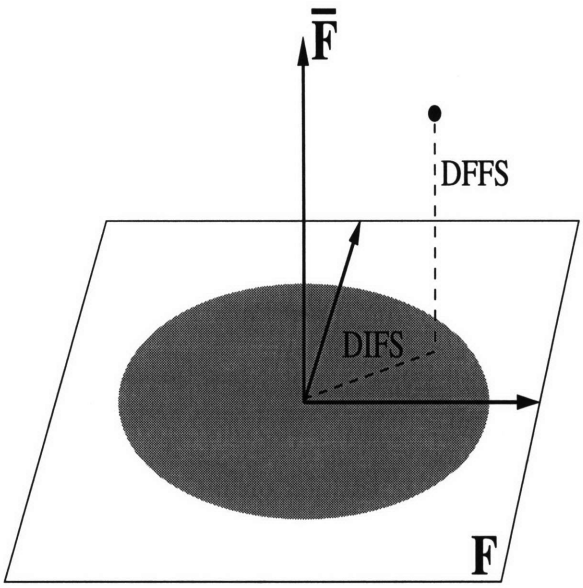


Figure 4-1: Decomposition of $\Re^N$ into the principal subspace F and its orthogonal complement $\bar{F}$ for a Gaussian density

M principal components and the orthogonal complement $\bar{F}$. The component of the orthogonal subspace $\bar{F}$ is the so-called "distance-from-feature-space"(DFFS), which is equivalent to the PCA residual error. The component of $\tilde{U}$ which lies *in* the feature space F is refereed to as the "distance-in-feature-space" and is a *Mahalanobis* distance for Gaussian densities.

As derived in [11] and in the Appendix A, the complete likelihood estimate can be written as the product of two independent marginal Gaussian densities or equivalently as an appropriately weighted sum of the DIFS and DFFS.

$$\begin{aligned} \hat{P}(\tilde{\mathbf{U}}|\Omega) &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\cdot\left[\frac{\exp\left(-\frac{\epsilon^2(\tilde{\mathbf{U}})}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \\ &= P_F(\tilde{\mathbf{U}}|\Omega)\ \hat{P}_{\bar{F}}(\tilde{\mathbf{U}}|\Omega) \end{aligned} \tag{4.15}$$

where $P_F(\tilde{U}|\Omega)$ is the true marginal density in F and $(\hat{P})_F(\tilde{U}|\Omega)$ estimated marginal density in the orthogonal complement $\bar{F}$, y_i are the principal components(eigenvectors) and $\epsilon^2(\tilde{U})$ is the residual (or DFFS). The value of ρ is bound by λ_{M+1} as discussed

in [11]

4.3 Performance of different methods

In this section we discuss the performance of the different methods described earlier on our Media Lab database. The database known as bank consists of a total of 1273 faces, including multiple images of the same person with different lighting, pose and expressions. The database was divided into two parts, half bank and reverse half bank respectively. The system was trained on one part and tested on the other. The other test carried out was training and testing on the full bank. The results obtained for the various methods are summarized in 4.1:

Table 4.1: Performance of the different methods on different databases

Method Used	Performance on half bank	Performance on reverse half bank	Performance on full bank
Euclidean	89.18(569)	87.26(555)	86.1(1096)
Linear Classifier	0.3(2)	1.0(6)	4(51)
Quadratic Classifier using only diagonal elements of covariance matrix	88.56(565)	88.68(564)	88.06(1121)
Quadratic Classifier	94.2(601)	93.87(597)	98.98(1261)
Intra/Extra Eigenvectors	94.2(601)	93.87(597)	98.98(1261)

4.4 Analysis

In this section, we will analyze the performance of the different methods on the database.

4.4.1 Euclidean

The performance of this method is very impressive considering its simplicity. It however does not take into account the underlying probability distribution of the images

and weights all the coefficients equally and this is the reason why its performance is not as good as some of the other methods described.

4.4.2 Linear Classifier

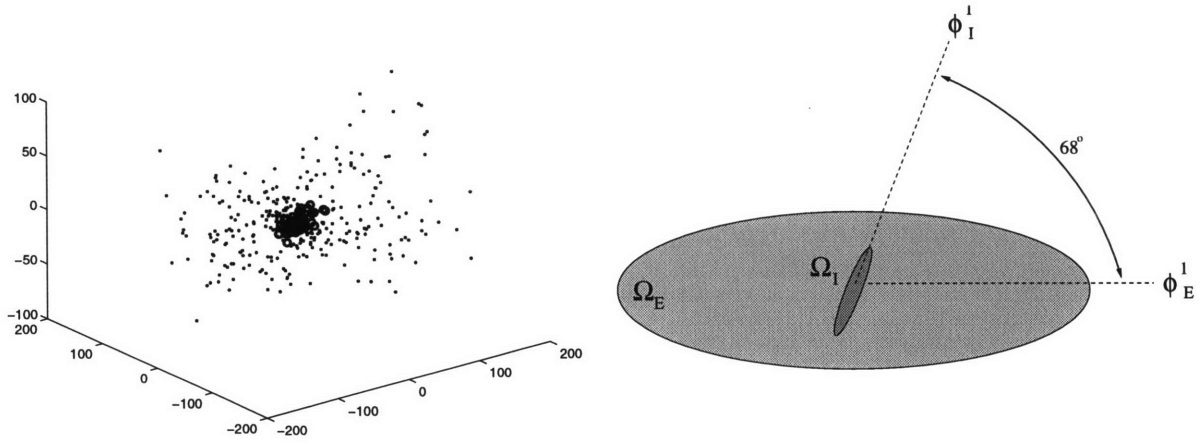
This method posts some incredibly low scores. The simple explanation for the poor performance is that the intra/extra clusters are not linearly separable.

If we plot the first three principal components for the intra Ω_I class versus the extra class Ω_E as shown in figure 4-2, we see that they are not linearly separable. Thus simple linear discriminant techniques cannot be used with any degree of reliability. The proper decision boundary is inherently nonlinear (quadratic in fact) and is best defined in terms of the *a posterior* probabilities.

Visual interpretation of figure 4-2 is misleading since we are dealing with low dimensional hyper-ellipsoids which are intersecting near the origin of a very high-dimensional space. The key distinguishing factor between the two distributions is their relative orientations. In figure 4-2, we calculate the angle between the major axes of the two hyper-ellipsoids by computing the dot product between their respective first eigenvectors. The angle between them was found to be 68 deg, implying that that the orientations are indeed different.

4.4.3 Quadratic Classifier Using only Diagonal Elements of Covariance Matrix

The performance of the quadratic classifier using only the diagonal elements of the covariance matrix is better than the Euclidean distance metric. This is because it weights the different coefficients by the variance along that direction. However it still does not take into account the relative orientations of the clusters and ignores the cross correlation terms. Thus it does not perform as well as the Quadratic Classifier using the full covariance matrix.



(a) (b)

Figure 4-2: (a) distribution of the two classes in the first 3 principal components (circles for Ω_I, dots for Ω_E) and (b) schematic representation of the two distributions showing orientation difference between the corresponding principal eigenvectors.

4.4.4 Quadratic Classifier Using the Full Covariance Matrix

This method has the best performance of all the methods used. This Bayesian formulation casts the recognition problem into a classical binary pattern classification problem which is then solved by using the maximum *a posterior* (MAP) rule.

4.4.5 Intra/Extra Eigenvectors

This method performs as well as the Quadratic classifier with the full covariance matrix method. The two methods are actually equivalent. Both methods characterize the distribution of the intra and extra classes and use the same classification technique to match images.

The intra/extra eigenvectors are diagonal however the quadratic classifier is less computational intensive.

It is interesting to view the intra/extra eigenfaces in the original facespace. This allows us to visualize what features are prominent for the intra class and what features are prominent for extrapersonal variations. The first 15 intra and extra eigenfaces are shown in the figures 4-3 and 4-4 respectively.

We can see that most of the variations among the intrapersonal group is around

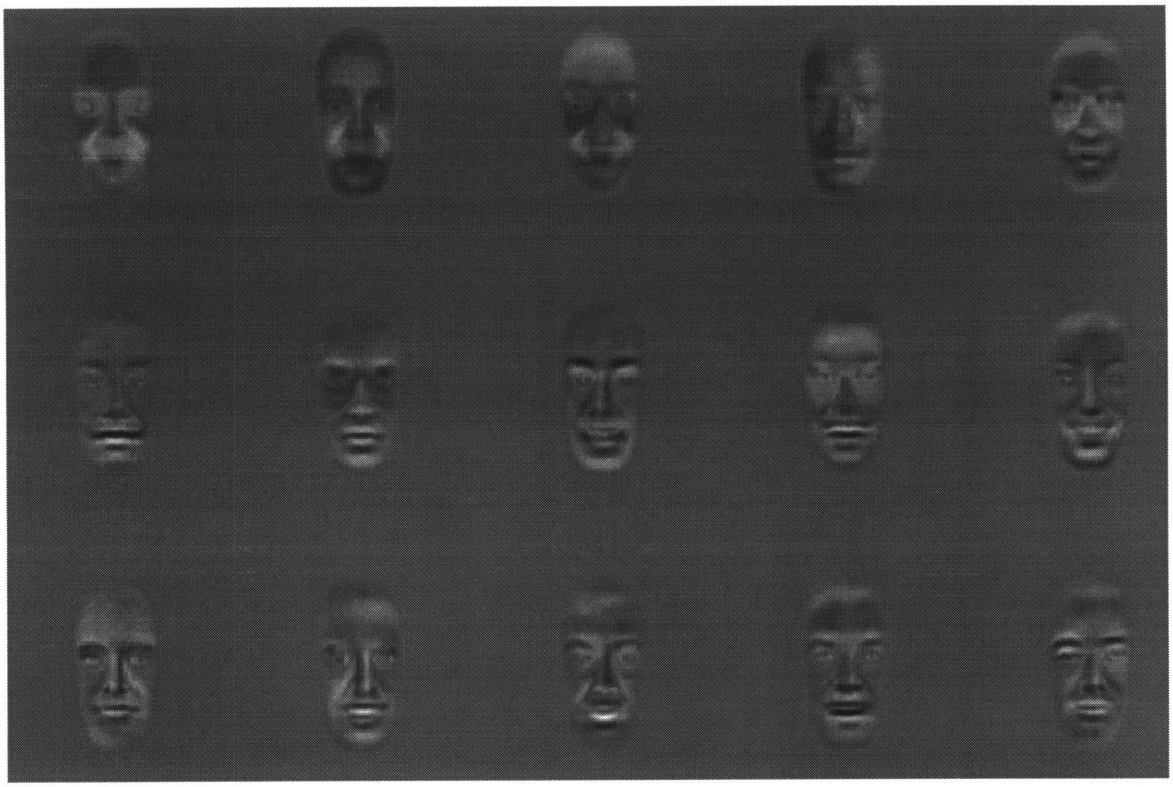


Figure 4-3: The first 15 intra eigenvectors viewed in the original facespace

the mouth region. This makes sense because most images of the same person are similar except for differences in facial expression, pose and glasses. For the extra personal group, the differences are more varied and this can be seen by studying the extrapersonal eigenfaces.

4.5 Summary

In this chapter, we presented several ways of matching the face coefficients for recognition. We found that a Bayesian approach to classification outperformed other methods of matching faces. This work is very similar to that of Pentland *et al* [9]. However this is different in that uses just the projection coefficients, instead of the deformable intensity surfaces $((x, y, I(x, y))$ that was used before.

One idea that we did not have time to explore is to use only the intra-eigenvectors in the original facespace. So instead of having a two-sided test, we have a one sided test. The recognition results might not be as accurate (though it should be pretty

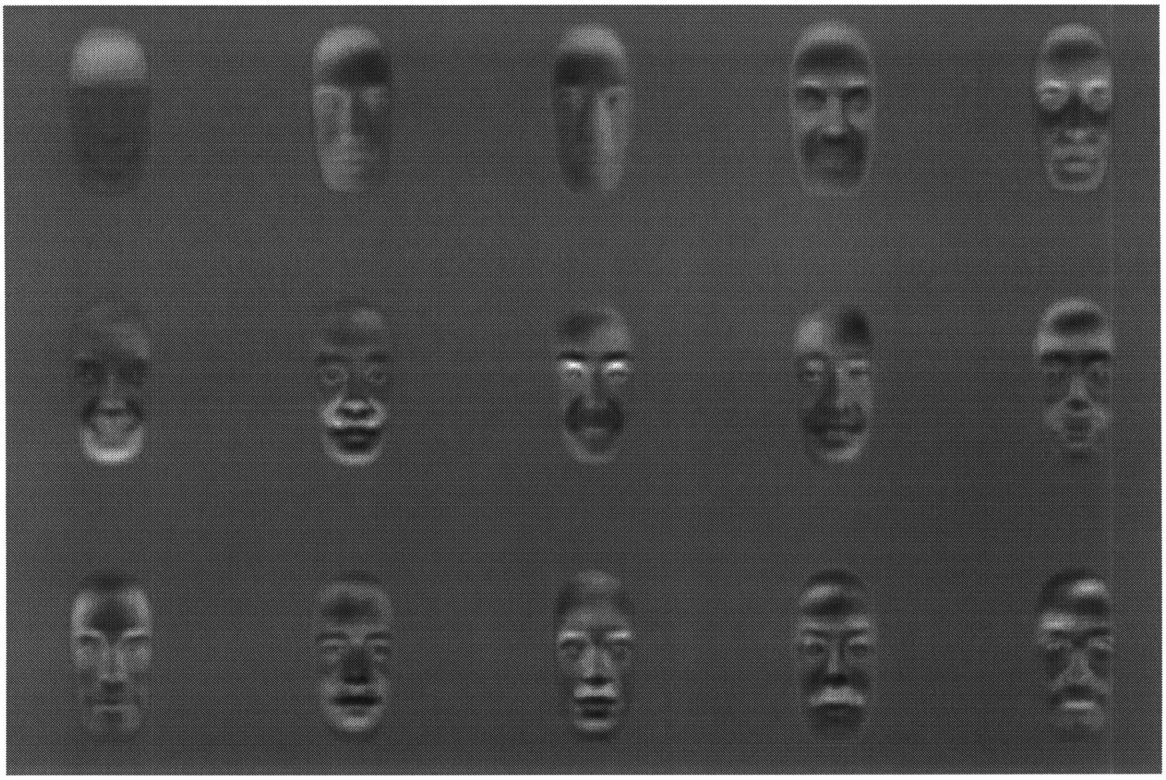


Figure 4-4: The first 15 extra eigenvectors viewed in the original facespace

close), but the computation is cut into half.

Another interesting test would be to carry out the intra/extra modeling in the original image space instead of in face space and see if the differences in performance.

Chapter 5

The FERET tests

As part of the Face Recognition Technology (FERET) program, the U.S. Army Research Laboratory (ARL) conducted a series of supervised government tests and evaluations of automatic face recognition algorithms. The goal of the tests was to provide an independent method of evaluating algorithms and assessing the state of the art in automatic face recognition systems.

This chapter provides a brief summary of the performance of our system in the FERET II tests compared against other face recognition systems available. We describe our final system, analyze its performance and compare its performance against our older system which participated in the FERET I tests.

5.1 THE FERET Database

Images of a person were acquired in sets of 5 to 11 images, collected under relatively unconstrained conditions. Two frontal views were taken (**fa** and **fb**); a different facial for the second view. The images were collected at different locations so there is some variation in illumination from one session to another. There is also variation in scale and pose. The test uses both *gallery* and *probe* images. The gallery is the set of known individuals while the set of unknown faces to be tests are called the probes. A duplicate image (called *dupe*) is defined as an image of a person whose gallery image was taken at a different date from the probe.

(a) (b)

Figure 5-1: Examples of FERET frontal-view image pairs used for (a) the Gallery set (training) and (b) the Probe set (testing).

In the September 1996 FERET test, the gallery set contained 3323 images while the probe set contained 3816 images. The probe set consisted of all the images in the gallery set plus rotated images and digitally modified images. The digitally modified images had variations in illumination and scale. To obtain a copy of the official government report on the FERET program and test results [15], contact Jonathan Phillips at jphillip@nvl.army.mil

An example of the faces in the FERET is given in figure 5-1.

5.2 Our Final System

Our final system for the September 1996 FERET test used a 125 dimensional eigenspace and used intra/extra eigenface modeling for the matching method. It was trained on the 2000-bank database at the Media Lab. It was not trained with *marginal* data nor was it trained with duplicate images. The head detection scheme was the same one that was used in the previous system (described in the second chapter).

5.3 Performance on the FERET tests

The performance of our system on the FERET I test is presented in the tables 5.1 and 5.2. The March 1995 Media Lab method used the Euclidean distance metric to match faces in a 100 dimensional facespace while the August 1996 Media Lab method (our current method) used intra/extra modeling in a 125 dimensional facespace.

Table 5.1: FA vs FB results on the FERET I tests

Institution	Recognition Rate
MIT Media Lab (August 1996)	96
Rockefeller (November 1995)	96
USC (March 1995)	92
MIT Media Lab (March 1995)	88

Table 5.2: Duplicate Scores on the FERET I tests

Institution	Recognition Rate
MIT Media Lab (August 1996)	69
Rockefeller (November 1995)	62
USC (March 1995)	58
MIT Media Lab (March 1995)	40

One should note the dramatic improvement of the August 1996 Media Lab performance over the March 1995. The improvement for the duplicate images is an amazing 29% !

The results for the FERET II tests which were administered in September 1996 is summarized in the figures 5-2, 5-3, 5-4 and in the table 5.3

These tests show the superiority of the intra/extra modeling technique over the previous Euclidean distance metric. It also tests our recognition system very extensively under more or less practical applications.

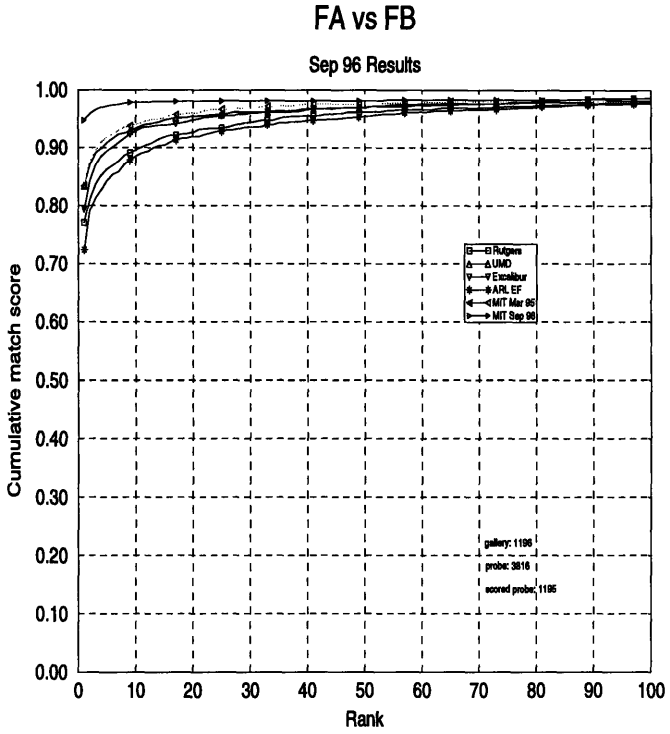


Figure 5-2: FERET II scores for FA vs FB

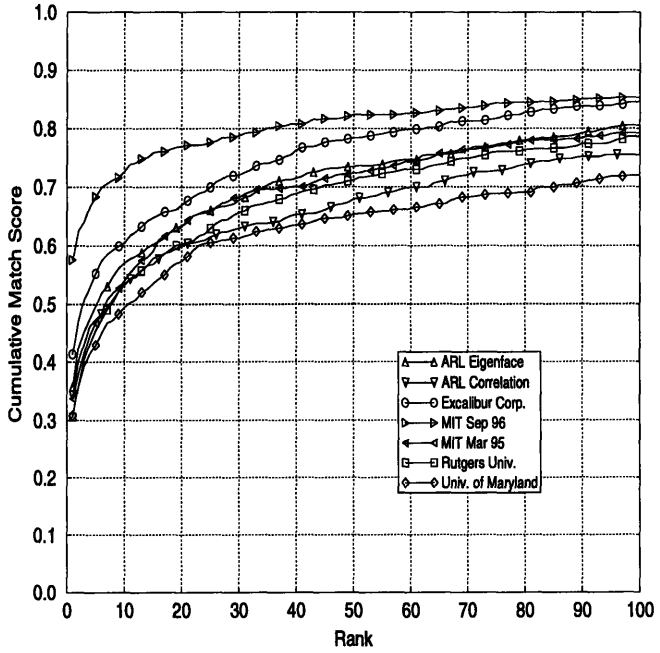


Figure 5-3: FERET II scores for duplicate images taken weeks or months apart

Table 5.3: Variations in performance over 5 different galleries of fixed size(200) on duplicate probes. Algorithms are order by performance (1 to 7). The order is by percentage of probes correctly identified (rank 1). Also included in the table is average rank 1 performance for all algorithms and number of probes scored

Algorithm	gallery 1	gallery 2	gallery 3	gallery 4	gallery 5
ARL Eigenface	6	6	3	2	5
ARL Correlation	7	4	4	4	6
Excalibur Corp.	2	3	2	3	1
MIT Sep 96	1	1	1	1	1
MIT Mar 95	4	2	5	6	7
Rutgers Univ.	3	4	7	5	4
Univ. of Maryland	4	6	6	7	1
Average	0.220	0.587	0.626	0.512	0.653
Number of Probes Scored	143	64	194	277	44

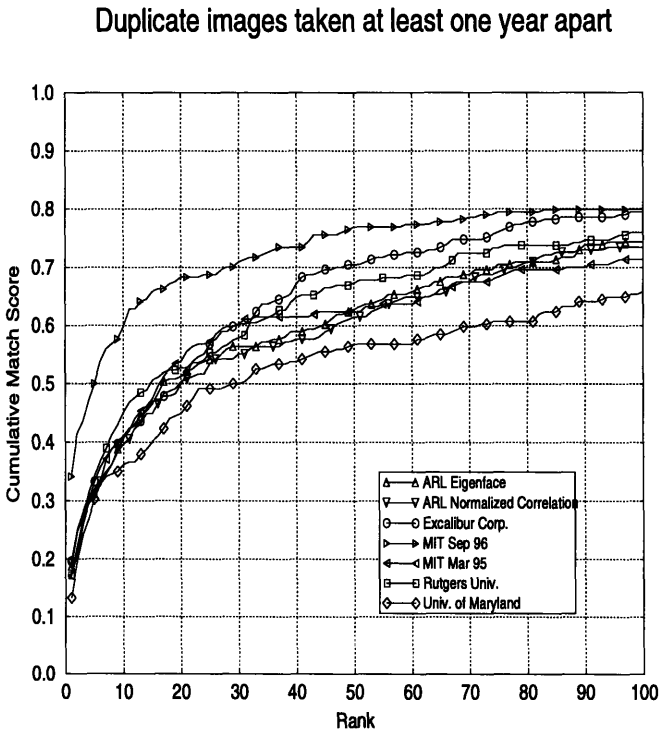


Figure 5-4: FERET II scores for duplicate images taken at least a year apart

Chapter 6

Conclusion and Future Directions

With these preliminary experiments, the advantages of the Bayesian framework for matching faces are clear. We have addressed the problems with our previous recognition system at the Media Lab and improved our recognition results. The biggest performance increase was on duplicate images where we topped our previous mark by a 30% margin. We have generated optimal values for various parameters and have demonstrated the superiority of our system over other existing systems through thorough testing on the FERET database.

However there are a lot of further improvements that can be made to the face recognition system.

6.1 Head Detection: The need for a real time system

The system we have is not a real time system. For a lot of practical applications, there is the need for a real time system that can identify and recognize a face instantly. Examples of applications include security clearance and even entertainment. The bottleneck in our system is the multiscale face detection module which takes a couple of seconds to produce the best detection.

The system can be speeded by very easily by integrating it with a real time face

detection and tracking system. Such a system developed by Oliver and Pentland[13] called LAFTER already exists in the Media lab. This is a real-time system for finding and tracking a human face and mouth. It has been tested on hundreds of users and demonstrated to be extremely reliable and accurate. We can feed the input the head images obtained by the LAFTER system into our feature detection module and then in to our recognition system. This will provide not only real time recognition but also allow us to accurately recognize faces in situations which are less restrictive than mug-shots. We can also integrate the system with the STIVE [19] system in the Media Lab's "Smart Desks" project. This system tracks heads and hands in 3-D at a very casual environment. Another useful application would be to use a wearable computer to track and locate a head [5].

We can also incorporate negative examples into our face detection scheme. At present, we estimate the probability distribution of the cluster of heads and use a maximum likelihood score to find the best head. This is a one-sided test though and we can imagine getting better accuracy using a two-sided test where we distinguish with positive examples and negative examples of training heads.

6.2 Improving performance on Duplicate Images and Rotated Images

The performance of most face recognition systems on duplicate images is dismal. This might be due to a lack of good training data which means that we do not have examples of the changes in someone's face over a number of years. Getting training data would be a good place to start. It will also be interesting to see which facial regions undergo the least change over the years (as we mentioned earlier, there is widespread speculation that the eyes undergo very little change with age).

The performance of most systems on rotated images (half, quarter, and profile views) also leaves a lot to be desired. The best performance on profile views is about 40% and that means that a lot of work needs to be done in this region. This problem probably needs more advanced representation of faces. 3-D models of heads might be

appropriate where the 3-D can be reconstructed from images and compared to known individuals. There is a lot of work being done at the moment on 3-D pose estimation and reconstruction [3].

To me, the perfect face recognition system will be one that can be used in a very casual environment (for example your living room) and can detect and recognize faces in a crowded environment instantly. With all the work that is being done today, this is not an impossibility and we might reach that goal very soon.

Appendix A

Probabilistic Visual Learning for Object Detection

This appendix is derived from [11]. It describes a maximum likelihood estimation framework for visual search and target detection.

We begin by consider a Gaussian density whose mean $\bar{x}$ and covariance Σ have been robustly estimated from a training set x^t. The likelihood of an input pattern $\mathbf{x}$ is given by

$$P(\mathbf{x}|\Omega) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right]}{(2\pi)^{N/2}\ |\Sigma|^{1/2}} \tag{A.1}$$

The sufficient statistic for characterizing this likelihood is the *Mahalanobis* distance

$$d(\mathbf{x}) \ = \ \tilde{\mathbf{x}}^T\Sigma^{-1}\tilde{\mathbf{x}} \tag{A.2}$$

where $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$.

Using the eigenvectors and eigenvalues of Σ, we can rewrite Σ^{-1} in the diagonalized form

$$\begin{aligned} d(\mathbf{x}) &= \tilde{\mathbf{x}}^T\Sigma^{-1}\tilde{\mathbf{x}} \\ &= \tilde{\mathbf{x}}^T\left[\Phi\Lambda^{-1}\Phi^T\right]\tilde{\mathbf{x}} \\ &= \mathbf{y}^T\Lambda^{-1}\mathbf{y} \end{aligned} \tag{A.3}$$

Here $\mathbf{y} = \Phi^T\tilde{\mathbf{x}}$ are the new variables obtained by change of coordinates in a KLT.

Because of the diagonalized form, the Mahalanobis distance can be expressed as

$$d(\mathbf{x}) = \sum_{i=1}^{N} \frac{y_i^2}{\lambda_i} \tag{A.4}$$

An estimator for $d(\mathbf{x})$ using only the M principal components is:

$$\begin{aligned} \hat{d}(\mathbf{x}) &= \sum_{i=1}^{M} \frac{y_i^2}{bda_i} + \frac{1}{\rho}\left[\sum_{i=M+1}^{N} y_i^2\right] \\ &= \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i} + \frac{1}{\rho}\,\epsilon^2(\mathbf{x}) \end{aligned} \tag{A.5}$$

where the term $\epsilon^2(\mathbf{x})$ is the DFFS and can be computed as :

$$\epsilon^2(\mathbf{x}) = \sum_{i=M+1}^{N} y_i^2 = ||\tilde{\mathbf{x}}||^2 - \sum_{i=1}^{M} y_i^2 \tag{A.6}$$

Thus we can write the liklihood estimate based on $d(\mathbf{x})$ as the product of two marginal and independent Gaussian densities:

$$\begin{aligned} \hat{P}(\mathbf{x}|\Omega) &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right] \cdot \left[\frac{\exp\left(-\frac{\epsilon^2(\mathbf{x})}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \\ &= P_F(\mathbf{x}|\Omega)\ \hat{P}_{\bar{F}}(\mathbf{x}|\Omega) \end{aligned} \tag{A.7}$$

where $P_F(\mathbf{x}|\Omega)$ is the true marginal density in F-space and $\hat{P}_{\bar{F}}(\mathbf{x}|\Omega)$ is the estimated marginal density in the orthogonal complement $\bar{F}$-space. The optimal value of ρ can now be determined by minimizing a suitable cost function $J(\rho)$.

Bibliography

[1] Joseph Atick, Paul Griffin, and Norman Redlich. Face recognition from live video for real-world applications – now. *Advanced Imaging Magazine*, 1995.

[2] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. IEEE Transactions of Information Theory, 1995.

[3] Tony Jebara and Alex Pentland. Parametrized structure from motion for 3d adaptive feedback tracking of faces. *Submitted to IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

[4] I.T. Jolliffe. *Principal Component Analysis.* Springer-Verlag, New York, 1986.

[5] Jeffrey Levine. Real-time target and pose recognition for 3-d graphical overla. Master's thesis, Massachusetts Institute of Technolog, Department of Electrical Engineering and Computer Science, June 1997.

[6] M.M. Loeve. *Probability Theory.* Van Nostrand, Princeton, 1955.

[7] Fleming M. and G. Cottrell. Categorization of faces using unsupervised feature extraction. IJCNN-90, 1990.

[8] T. Maurer and C. Malsbur. Single-view based recognition of faces rotated in depth. Zurich, 1995. Proc. Int. Workshop on Automatic Face and Gesture Recognition.

[9] Baback Moghaddam, Chahab Nastar, and Alex Pentland. Bayesian face recognition using deformable intensity surfaces. *IEEE Conference on Computer Vision and Pattern Recognition*, 1996.

[10] Baback Moghaddam and Alex Pentland. View-based and modular eigenspaces for face recognition. Washington, D.C., 1994. Computer Vision and Pattern Recogniton.

[11] Baback Moghaddam and Alex Pentland. Probabilistic visual learning for object detection. Cambridge, MA, October 1995. 5th International Conference on Computer Vision.

[12] Baback Moghaddam and Alex Pentland. A subspace method for maximum likelihood target detection. Washington, D.C., October 1995. IEEE International Conference of Image Processing.

[13] Nuria Oliver and Alex Pentland. Lafter: Lips and face real time tracker. Technical report, Massachusetts Institute of Technology, 1997.

[14] A.P. Pentland, R Picard, and S. Sclaroff. Photobook: tools for content-based manipulation. Storage and Retrieval of Image and Video DatabasesII, 1994.

[15] P. Jonathan Phillips, Hyeonjoon Moon, Patrick Rauss, and Syed A. Rizvi. The feret septmber 1996 database and evaluation procedure. Crans-Montana, Switzerland, March 1997. First International Conference on Audio and Video-Based Biometric Person Authentication.

[16] L Sorvich and M Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, pages 519–524, 1987.

[17] Daniel Swets and John(Juyang) Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

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

[19] Christopher Wren and Alex Pentland. Dynamic modeling of human motion. Technical report, Massachusetts Institute of Technology, 1997.