

FACE RECOGNITION WITH EIGENFACES – A DETAILED STUDY

by

NADEEM VAWDA

Submitted in fulfilment of the academic
requirements for the degree of Master
of Science in the School of Mathematics,
Statistics and Computer Science,
University of KwaZulu-Natal, Durban

April 2012

As the candidate's supervisor, I have approved this dissertation for submission.

Signed: Name: Date:

EXAMINER'S COPY

ABSTRACT

With human society becoming increasingly computerised, the use of biometrics to automatically establish the identity of an individual is of great interest in a wide variety of applications. Facial appearance is an appealing biometric, on account of its relatively non-intrusive nature. As such, automated face recognition systems have been the subject of much research in recent years.

This dissertation describes the development of a fully automatic face recognition system, and provides an analysis of its performance under various different operating conditions, in comparison with results published in prior literature. In addition to giving a detailed description of the mathematical underpinnings of the techniques used by the system, we discuss the practical considerations involved in implementing the described techniques.

The system presented here uses the eigenface approach to representing facial features. A number of different recognition techniques have been implemented and evaluated. These include a number of variants of the original eigenface technique proposed by Turk and Pentland, as well as a related technique based on the probabilistic approach of Moghaddam et al.

Due to the wide range of datasets used to evaluate face recognition systems in the literature, it is difficult to reliably compare the performance of different systems. The system described here has been tested with datasets encompassing a wide range of different conditions, allowing us to draw conclusions about how the characteristics of the test data affect the results that are obtained.

The performance of this system is comparable to other eigenface-based systems documented in the literature, achieving success rates in the region of 85% for large datasets under controlled conditions. However, performance was observed to degrade significantly when testing with more free-form images; in particular, the effects of ageing on facial appearance were noted to cause problems for the system. This suggests that the matter of ageing is still a fruitful direction for further research.

PREFACE

The experimental work described in this dissertation was carried out in the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, from January 2010 to April 2012, under the supervision of Professor Jules-Raymond Tapamo, and the co-supervision of Professor Fulufhelo Nelwamondo of the Council for Scientific and Industrial Research.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

DECLARATION – PLAGIARISM

I, Nadeem Vawda, declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a. Their words have been re-written but the general information attributed to them has been referenced
 - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References section.

Signed:

CONTENTS

1	Introduction	1
1.1	Motivation	1
1.2	Problem Description	2
1.3	Objectives	2
1.4	Contributions	3
1.5	Document Outline	3
2	Literature Review	4
2.1	Biometrics	4
2.2	General Techniques for Face Recognition	5
2.3	Ageing and Face Recognition	8
2.4	Face Recognition with Eigenfaces	9
2.5	Evaluation of Face Recognition Systems	10
2.6	Conclusion	11
3	Methods and Techniques	12
3.1	Image Preprocessing	12
3.1.1	Face Localisation	12
3.1.2	Illumination Compensation	13
3.2	Structure of the Recognition Process	13
3.3	Classical Eigenface Recognition	13
3.3.1	Training	14
3.3.2	Feature Extraction	15
3.3.3	Identification	16
3.4	Probabilistic Eigenface Recognition	18
3.4.1	Training	19
3.4.2	Probability Estimation	20
3.4.3	Identification	24
3.5	Conclusion	25
4	Implementation Notes	26
4.1	Library Use	26
4.2	Preprocessing Pipeline	26

4.3	Numerical Precision Concerns	29
4.3.1	Maximum Likelihood Formulation	29
4.3.2	Maximum <i>A Posteriori</i> Formulation	29
5	Experimental Results	31
5.1	Datasets	31
5.1.1	FERET	31
5.1.2	MORPH	32
5.1.3	Training Data Organisation	33
5.2	Experiment 1 – Face Localisation and Masking	33
5.3	Experiment 2 – Different Distance Functions	34
5.4	Experiment 3 – Subspace Dimensionality	35
5.5	Experiment 4 – Dataset Size	37
5.6	Experiment 5 – Alternate Datasets	38
5.6.1	Experiment 5(a) – FERET “Duplicate I” subset	39
5.6.2	Experiment 5(b) – MORPH Album 2	40
5.6.3	Experiment 5(c) – MORPH Album 1	41
5.7	Comparison With Other Studies	42
5.7.1	Eigenfaces	42
5.7.2	Other Techniques	46
6	Conclusion	48
6.1	Directions for Future Work	49
	References	50

LIST OF FIGURES

3.1	Example eigenfaces ($N = 272$)	15
3.2	Mean intrapersonal and extrapersonal differences	18
4.1	The preprocessing pipeline	27
4.2	Preprocessing of an example image	28
5.1	Exp. 3: Accuracy vs. eigenvector selection threshold	36
5.2	Exp. 3: CMC curves ($v = 0.99$)	37
5.3	Exp. 4: Accuracy vs. dataset size	38
5.4	Exp. 5(a): CMC curves (FERET “Duplicate I” subset)	39
5.5	Exp. 5(b): CMC curves (MORPH album 2)	41
5.6	Exp. 5(c): CMC curves (MORPH album 1)	42

LIST OF TABLES

5.1	Results for exp. 1: face localisation and masking	34
5.2	Results for exp. 2: distance functions	34
5.3	Results for exp. 3: subspace dimensionality	36
5.4	Results for exp. 4: dataset size	37
5.5	Results for exp. 5(a): FERET “Duplicate I” subset	39
5.6	Results for exp. 5(b): MORPH album 2	40
5.7	Results for exp. 5(c): MORPH album 1	41

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Prof. Jules-Raymond Tapamo, for his guidance during the course of my research. He has been a constant source of good advice in all aspects of this endeavour, and his energy and passion for computer science are truly inspirational.

I would like to thank my co-supervisor, Prof. Fulufhelo Nelwamondo, for his feedback and support.

I also thank my parents Cati and Yousuf, and my sister Leila, for their support and encouragement throughout my studies.

Finally, I wish to thank the Council for Scientific and Industrial Research's Modelling and Digital Science unit for funding my research.

Portions of this research use the FERET database of facial images collected under the FERET program, sponsored by the DoD Counterdrug Technology Development Program Office (Phillips et al., 1998, 2000).

Portions of this research use the UNCW Craniofacial Morphology Database (Ricanek and Tesafaye, 2006).

CHAPTER 1

INTRODUCTION

1.1 Motivation

Modern society has come to rely on computers for many essential services. As a result, security concerns in computing are of great importance. Amongst such security concerns is the question of how to determine the identity of an individual in an automated fashion. Biometric identification techniques are the subject of much interest, having a wide range of applications in this regard.

By making use of intrinsic traits of the individual in question, biometrics can be used to replace or augment classical knowledge-based and token-based authentication schemes to improve security and convenience. Applications to which this is relevant include physical access control, border control, and the prevention of identity fraud. Additionally, biometrics enable the development of applications that are otherwise infeasible, such as assisting law enforcement agencies in identifying criminals, and finding missing persons.

Facial appearance is a biometric modality that is the subject of a great deal of research. While use of the face modality poses greater challenges than other commonly-used modalities such as the fingerprint and iris, it has the advantage of being extremely non-intrusive – ideally, a face recognition system can obtain sufficient data for identification without any action on the part of the subject. This is in contrast to, for example, a fingerprint system, where the subject’s active participation is required to obtain fingerprints.

While much progress has been made in this field over the past two decades, the development of a fully general automatic face recognition system is still some way away from being a solved problem. In particular, current state-of-the-art recognition techniques perform well when using inputs captured under controlled conditions, but their performance tends to degrade dramatically when exposed to certain kinds of variations present in more free-form environments. Thus, it is important that further research be done with the aim of developing a robust, generally-applicable face recognition system.

1.2 Problem Description

Face recognition is the problem of determining the identity of an individual from an image of the subject's face.

A technique for face recognition can generally be described in terms of the following components:

- A representation of the facial features that are useful for recognition.
- A feature extraction algorithm, which extracts features (in this representation) from a bitmap image of a subject's face.
- A similarity measure, which quantifies the likelihood that two sets of features were derived from images of the same individual.

Within this framework, the recognition process then consists of extracting features from the input image, and comparing this output to the stored features for each individual known to the system.

The problem of interest, then, is that of finding a suitable representation of facial appearance, and a similarity measure for comparing these features.

While prior research has led to the development of a number of proposed solutions to this problem (each with some degree of success), there still remains much space to improve on the performance attained by current systems, particularly with regard to operating conditions less strictly controlled than those typically used for evaluation. Improvements can be made both by devising entirely new techniques, and by making incremental improvements to existing ones.

1.3 Objectives

The aim of this study is to provide a detailed comparative analysis of several face recognition techniques based on the widely-used eigenface representation.

The primary focus of the study is the implementation of a recognition system (based on the techniques in question), and the evaluation of its performance in relation to other results published in the literature. In addition to presenting a survey of the literature, we discuss the challenges involved in accurately comparing the obtained results with the findings of other studies. A significant concern in this regard is the choice of dataset used for testing, and the impact of the various characteristics of the dataset on the results obtained.

Additionally, we aim to present a comprehensive treatment of both the mathematical underpinnings of the evaluated techniques, and the practical considerations involved in their implementation. It is intended that this part of the study should serve as an implementer's handbook, providing the necessary information for someone with minimal prior knowledge of face recognition techniques to implement a recognition system from scratch.

1.4 Contributions

This study makes the following contributions to the body of knowledge in the field of face recognition.

Firstly, it presents an in-depth experimental evaluation of eigenface-based face recognition techniques, on a number of datasets with a variety of different characteristics. A key aspect of the resulting analysis is the discussion of how various aspects of experiment design affect the results for a given recognition system being evaluated. The focus here is on the effects of factors such as dataset size, the presence/absence of *a priori* face location data, and the characteristics of the images making up the dataset (illumination, age differences, etc.). This provides a better basis for comparing results between studies in which these factors differ significantly.

Secondly, it reviews various techniques that have been proposed for face recognition in the literature, and discusses the associated experimental results in comparison with the results of our own experiments. This comparison is informed by the previously-mentioned analysis of experimental design considerations.

Finally, it provides a detailed, coherent treatment of the techniques that have been implemented. This is presented in an accessible manner, requiring minimal prior familiarity with face recognition techniques. In addition to describing the mathematical formulation of the eigenface techniques, we include notes on the practical issues that were encountered in the implementation of these techniques.

1.5 Document Outline

The rest of this document is organised as follows. Chapter 2 provides a survey of prior work in the field of face recognition. Chapter 3 is a detailed presentation of the algorithms used for face recognition in this study, as well as associated image preprocessing techniques. Chapter 4 discusses the implementation of the face recognition system. Chapter 5 describes the experiments performed, and discusses the results obtained. Finally, Chapter 6 draws conclusions and outlines possible directions for future work.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a survey of prior research relevant to the topic of face recognition, covering both the wide variety of techniques that have been investigated, and the performance of systems based on these techniques. We first discuss the general field of biometric research, to put the case of face recognition into context, and then move on to the discussion of specific techniques for face recognition. The family of techniques based on the eigenface representation receives particular attention, being the subject of much research, and the focus of this study.

2.1 Biometrics

In general, the subject of biometrics is concerned with the study of techniques for determining or verifying the identity of a human subject, based on some intrinsic traits of the individual in question (Ross and Jain, 2007). This is in contrast with knowledge-based identification schemes (such as passwords) and token-based schemes (such as ID cards), where the basis for identification is not inherently tied to the subject.

Within the general context of biometrics, there are a wide range of traits that can be used for identification. These can be divided into two categories: physiological and behavioural (Jain, 2007). Physiological modalities make direct use of anatomical characteristics of the subject, while behavioural modalities rely on recognising distinctive patterns in the way the subject performs some action. Some of the more widely-used physiological traits are the fingerprint, face and iris. Among behavioural biometrics, the signature and voice are two popular modalities.

While accuracy of identification is the most obviously important criterion for judging a biometric's utility, there are other considerations that should also be taken into account. One issue of particular significance is that of intrusiveness – a system that can operate without any active participation from the subject has advantages over one that requires action on the subject's part. This is evident in that a non-intrusive system is more convenient for users, but also in that such a system can be used in applications (e.g. surveillance) where a more intrusive system simply would not be feasible.

Indeed, in some cases accuracy and non-intrusiveness are attributes that come into

conflict, and in choosing a biometric one must make a trade-off between the two. For instance, fingerprint recognition can easily yield extremely accurate results, but requires the user to come into physical contact with a scanning device. In contrast, face recognition at present achieves a considerably lower level of accuracy, but requires much less user participation – the subject merely needs to face in the direction of a camera.

2.2 General Techniques for Face Recognition

Investigations into automated face recognition date back to the 1960s, with the semi-automatic recognition system developed by Bledsoe (1966), which was extended to a fully-automated system by Kanade (1977). These early systems attempted to characterise the face solely in terms of the distances between various key points. However, the performance of systems based on this simplistic approach was found to be lacking, even over the small datasets that were common at the time.

Techniques for face recognition can generally be grouped into two classes: *structural* and *holistic* (Zhao et al., 2003). Structural methods are based on the identification of facial landmarks (for example, the tip of the nose), and the extraction of local features pertaining to the appearance of these landmarks, and the distances between them. Holistic methods, on the other hand, use the raw face image as a whole as their input, without first identifying facial landmarks. In this scheme, the earliest approaches to face recognition can be seen as simplistic structural techniques.

While little research on face recognition was done during the 1980s (Zhao et al., 2003), the 1990s saw the development of a number of more successful techniques in both classes. Motivated by the experiments of Sirovich and Kirby (1987) in using Principal Component Analysis (PCA) for efficient representation of face images, Turk and Pentland (1991) proposed the (holistic) PCA-based eigenfaces technique. Their system achieved 100% accuracy on a dataset of 16 individuals. The authoritative literature survey of Zhao et al. (2003) cites this as the first truly successful automated recognition system. Eigenface-based approaches have seen a high level of popularity, and have formed the basis of numerous subsequent systems.

Another approach that has proven successful is the Dynamic Link Architecture (DLA), presented by Lades et al. (1993). DLA is a feature-based (i.e. structural) method based on elastic graph matching. While DLA is presented as a general-purpose object recognition technique, Lades et al. have successfully applied it to the face recognition problem, and it has been used as the basis for other face recognition systems. Lades et al. (1993) report accuracies of between 80 and 90%, achieved on a dataset of 87 individuals.

In the DLA approach, a face is represented by a labelled graph, with nodes corresponding to points on the face. Nodes are annotated with local Gabor features, and each edge is labelled with the distance between the landmarks corresponding to its endpoints. A cost function is defined to quantify how different two graphs are. The graph matching process makes deformations to the shape of the graph and small changes to the positions

of individual nodes, so as to minimise the cost between the input graph and the stored graph. Recognition is carried out by performing graph matching on the input face graph with respect to each of the stored graphs, and selecting the stored graph with the lowest cost after matching.

It should be noted that the DLA technique's recognition process is quite computationally intensive; Lades et al. (1993) report that it took between two and five seconds to compare a 128×128 pixel image to a single stored graph.

Elastic Bunch Graph Matching was proposed by Wiskott et al. (1997) as an extension of DLA. In EBG, rather than organising graph nodes in a rectangular grid, the nodes are located at key facial landmarks such as the nose and eyes. Stored graphs are combined into groups, called bunch graphs, allowing for more efficient graph extraction, and yielding a better model of potential variations in facial appearance. The cost function used for graph matching is refined to make use of the phase component of Gabor features, allowing for more accurate localisation of facial landmarks.

This system was found to give similar results to the earlier DLA system, with improved performance in the case of moderate pose variation (22°). It should be noted, though, that the EBG system must be trained before it can construct face graphs automatically, typically requiring 70 or so faces with landmark positions extracted manually. In contrast, the DLA approach can extract graphs automatically with no training, since the layout of graph nodes is not dependent on the input image.

Related to the previously-mentioned eigenfaces representation is the Independent Component Analysis (ICA) technique (Bartlett and Sejnowski, 1997). ICA-based systems are typically similar in structure to the eigenfaces system of Turk and Pentland (1991), relying on a subspace projection followed by nearest-neighbour matching of the resulting feature vectors. The key difference lies in the method used to construct the subspace onto which the inputs are projected.

Draper et al. (2003) present two architectures for applying the ICA technique to the problem of face recognition. Architecture I is structured to produce statistically independent basis images, yielding localised features (in contrast with the global features produced by PCA). Architecture II, on the other hand, constructs basis images such that the projection coefficients are independent (rather than the basis images themselves). This results in global features conceptually similar to PCA-based systems, unlike architecture I.

Draper et al.'s study compares the performance of systems based on each of these architectures to an eigenface-based system, using the various image sets found in the FERET dataset (Phillips et al., 1998). ICA architecture II was found to consistently give the best performance of the three systems, particularly on image sets containing unfavourable lighting conditions, or significant time lapses between the capture of probe and reference images. The eigenface-based system came in second place, with the ICA architecture I system having the poorest recognition rate.

Also related to the eigenfaces representation is the Fisherfaces technique due to Belhumeur et al. (1997). As with eigenface- and ICA-based techniques, the crux of the

system is a subspace projection operation, which is followed by nearest-neighbour matching. In this case, the subspace is constructed using Fisher’s Linear Discriminant (FLD; also referred to as Linear Discriminant Analysis, or LDA), rather than PCA or ICA.

Belhumeur et al. observe that while the basis produced by PCA is statistically optimal for image reconstruction, it is not optimal for recognition, since PCA acts to maximise the variance of the training dataset in the subspace, even though not all of this variance is relevant to the recognition problem. In contrast, the Fisherfaces technique attempts to distinguish variations due to differences in identity from those due to other factors such as variations in illumination and expression, and maximise the ratio of the former to the latter in the resulting subspace. In order to achieve this, the dataset used to train the system must contain multiple images of each subject, so that it can learn the variations between same-subject images as well as those between different-subject images.

In their study, Belhumeur et al. compare the performance of the Fisherfaces technique to that of the eigenfaces method. Under optimal conditions, both systems were able to achieve 100% accuracy (on a dataset of 30 images). However, on a dataset containing significant expression and lighting variations (comprising 65 images), the Fisherfaces system maintained an accuracy of 95%, while the eigenfaces system performed much more poorly at 68% accuracy.

A more recently-developed technique is that of sparse representation-based classification (SRC), due to Wright et al. (2009). The SRC technique represents the subspace projection of an input image as a linear combination of the projections of the stored reference images. This linear combination can be computed efficiently through ℓ^1 minimisation, using standard linear programming techniques.

Unlike most subspace-based techniques, SRC does not use a nearest-neighbour classifier. Instead, it considers all stored features for a given subject together, and computes an aggregate residual for that subject (rather than doing a separate comparison against each stored feature vector). For the technique to yield meaningful results, it is necessary to store multiple reference images of each subject.

Wright et al. evaluate the SRC technique using a number of different feature extraction methods, including classical features such as eigenfaces and Fisherfaces, as well as two types of unconventional features (down-sampling and “randomfaces”). They compare its performance with a nearest-neighbour classifier, an SVM-based classifier, and the “nearest subspace” technique of Ho et al. (2003).

The classifiers were evaluated on two datasets – one containing 38 individuals, and the other 100 individuals. The SRC and SVM classifiers were found to give consistently better results than the nearest-neighbour and nearest-subspace classifiers, with SRC giving slightly better results on the first dataset, while SVM had a small advantage on the second dataset. Additional experiments were conducted to study the effects of occlusion and random noise on recognition performance. The SRC classifier was found to be substantially more robust to such image corruption than the other systems.

2.3 Ageing and Face Recognition

One particular challenge in the field of face recognition is the problem of ageing. Over time, an individual's face can change significantly in appearance. As a result, many recognition techniques perform poorly when the reference image of the subject is of a substantially different age from the image to be identified. While this matter has received relatively little attention within the field, there has been some fruitful research done in recent years.

One such study is that of Ramanathan and Chellappa (2006), who present a face verification system based on a probabilistic eigenface framework (refer to section 2.4 for discussion).

Singh et al. (2007) propose a preprocessing transformation for improving the performance of existing recognition techniques. The transformation is applied to each stored reference image before it is compared to the probe image, in order to synthesise a version of the reference image at the age that the subject is in the probe image.

The transformation works by identifying corner points in the reference and probe images, correlating them, and then warping the reference image so that the two sets of points occur at the same positions. This has the effect of transforming the general shape of the face in the reference image to match the probe image more closely.

The age transformation was found to improve recognition accuracy dramatically, with increases of between 10 and 30 percentage points on the different datasets used. However, the overall performance of the system was still poor, failing to achieve more than 50% accuracy on any of the datasets.

Another approach to the ageing problem is the 3D ageing model of Park et al. (2010). In this technique, as in the previous system, ageing compensation is treated as a preprocessing step, with images being age-normalised before recognition is performed. Here, both the shape and texture of the face are taken into account, allowing for a more accurate model of the ageing process (particularly for older subjects, where the changes that occur are predominantly textural).

Changes in face shape are handled by constructing a three-dimensional model of each image in the database. For a given probe, the shape model is expressed as a weighted sum of models of other images of subjects at the same age. Then, an analogous weighted sum is constructed using models of subjects at the desired age, resulting in an age-adjusted shape model of the probe. Age-adjusted textural data is derived in a similar fashion, using a PCA-based representation applied to a pose-normalised version of the probe image. From these two components the system then synthesises an age-adjusted rendition of the probe image for use by the underlying recognition mechanism.

The technique was evaluated using a commercial face recognition system. It was found to improve recognition rates by around 10 percentage points on each of the three datasets used (FG-NET, MORPH and BROWNS). Particularly noteworthy was the recognition rate of approximately 66% that was achieved on the MORPH album 1 dataset (containing 625 subjects).

2.4 Face Recognition with Eigenfaces

The success of Turk and Pentland’s eigenface system has led to much interest and further work on this approach to the recognition problem.

Flynn et al. (2003) analysed the effects of time variance on eigenface recognition performance. The study found that while the system was able to achieve the high recognition rates observed in prior literature when all the images of a given individual are taken in the same session (and thus, on the same day). However, it was found that when using images from several different sessions for each individual, performance deteriorated significantly, even though the time between sessions was only on the order of weeks. While such time scales are too short to observe significant ageing effects, this deterioration could be attributed to cosmetic variations such as change in hairstyle.

Yambor et al. (2002) have evaluated several different distance metrics for identifying the “nearest” matching gallery image in the eigenspace. The study found that the Mahalanobis metric performs better than the conventional L_2 (Euclidean) metric and the angle metric when a large fraction of the eigenvectors are used to construct the eigenspace, while performance of the three metrics is equivalent when using a small number of eigenvectors.

Several techniques for eigenvector selection are also discussed, including Kirby’s energy and stretch functions. In many earlier eigenface-based systems, the fraction of eigenvectors used to construct the eigenspace was determined empirically. These techniques provide systematic ways to decide how many eigenvectors to use, independent of performance on a particular dataset.

Perlibakas (2004) investigates a large number of different distance metrics. As in Yambor et al.’s study, the Mahalanobis metric was found to give best performance. Here, however, the normalised L_2 metric gave comparable results.

One significant modification to the basic eigenface technique is the Bayesian eigenface method developed by Moghaddam et al. (1996). In this approach, instead of using a conventional distance metric, a probabilistic similarity measure is used to determine which gallery image is most similar to the input image. This modification is presented using the conventional intensity-difference representation, but also in conjunction with with an optical flow-based representation, and with a deformable 3D surface representation. The system was evaluated on a dataset of 76 individuals, using images taken from the FERET database. The deformable 3D surface representation achieved a recognition rate of 92%, while the intensity-difference representation reached an accuracy of 89%. The optical flow representation was less performant at 87%.

The probabilistic eigenimage framework has also proven useful in the related field of object detection. Moghaddam and Pentland (1997) present a method for detecting objects of a particular class, based on the probabilistic recognition technique. The discussion includes demonstrations of the method’s application to the problems of face detection and hand detection.

While the probabilistic recognition technique has proven significantly more accurate

than the classical eigenface approach, it has the disadvantage of being substantially more computationally intensive, since the system must compute a difference image for every (probe, gallery image) pair, rather than extracting features from each probe and gallery image in isolation and then comparing just these lower-dimensional features.

Moghaddam et al. (2000) present an optimised formulation of the probabilistic recognition technique that allows for independent calculation of the eigenface coefficients for each input image, thereby greatly reducing the running time required for recognition. A system using this formulation was found to give significantly improved performance over those based on other techniques in the FERET evaluation (Phillips et al., 2000), where it achieved a recognition rate of approximately 95% on a dataset of 1196 individuals.

The probabilistic approach has also been applied to the related problem of face verification. Ramanathan and Chellappa (2006) present an identity verification system based on the probabilistic eigenfaces technique, using the formulation of Moghaddam and Pentland (1997). This system was found to achieve promising results, particularly in its ability to handle age differences between the input and reference images. Also presented are techniques for normalisation of pose and illumination.

2.5 Evaluation of Face Recognition Systems

In assessing the performance of a face recognition technique, it is important to be able to reliably compare the results obtained with those for other techniques. However, in general this is not a trivial matter, due to factors such as differences in the details of the evaluation procedure, and in the size and composition of the dataset used.

To resolve this issue, the FERET program (Phillips et al., 2000) developed a standardised methodology for evaluating face recognition systems for purposes of comparison. Key points in the design of the evaluation protocol are the use of a common dataset in evaluating all techniques under consideration, and the requirement that training must be completed prior to the test. The latter stipulation ensures that the system cannot tailor its face representation to suit the specific dataset in use, instead requiring it to provide a more general recognition mechanism.

Related to the matter of using a common dataset, the FERET program also compiled and published a large body of face images known as the FERET database. This collection of images has been widely used in subsequent studies evaluating the performance of face recognition algorithms (including several of those previously mentioned), in addition to its use in the tests undertaken as part of the FERET program itself. The database includes a number of subsets containing images of different characteristics, such as same-session alternate-expression image pairs, same-session pairs taken under different illumination conditions, and pairs taken six months apart. This allows tests to analyse how such varied conditions affect recognition accuracy.

2.6 Conclusion

This chapter has presented a survey of the large body of prior research that has been conducted in the field of face recognition. One particularly prominent class of recognition techniques is that of subspace projection. This class encompasses the Fisherfaces and ICA techniques, the more sophisticated SRC technique, as well the eigenfaces family of techniques that are the focus of this study. Other noteworthy techniques include the DLA and EBGM systems, which make use of local features. A significant challenge faced by contemporary face recognition systems is that of ageing, which causes current techniques to perform poorly. One approach to solving this problem is to apply a preprocessing step before the actual recognition algorithm, in order to normalise the apparent age of the input image.

CHAPTER 3

METHODS AND TECHNIQUES

This chapter discusses the techniques used by the facial recognition system implemented in this study, with a focus on the mathematical underpinnings of these techniques. We begin by detailing the preprocessing techniques that are applied to improve the quality of input images before they are received by the recognition system, and thereafter describe the recognition algorithms themselves. This latter part of the chapter first describes the basic structure of the recognition process, then introduces the basic eigenfaces technique, and finally presents a related technique based on probabilistic similarity measures.

3.1 Image Preprocessing

3.1.1 Face Localisation

When considering an image containing a face, it is often the case that, in addition to the actual face, the image contains a non-negligible amount of background detail. This can introduce noise into the features extracted from the image, and consequently reduce the accuracy of the recognition system. Additionally, the eigenface recognition technique is known to perform poorly in the presence of variations in face size (Turk and Pentland, 1991).

To overcome these problems, it is necessary to extract the face itself from the image, excluding as much of the surrounding background as possible. To this end, we have used an object detector developed by Viola and Jones (2004).

The detector makes use of Haar-like rectangular features. An intermediate representation called the “integral image” allows these features to be extracted extremely efficiently. The features are used as input to a collection of classifiers, each consisting of a network of simple perceptrons.

The classifiers are organised into a cascade, where each stage of the cascade rejects those sub-windows for which it deems not to contain a face. Only those sub-windows giving a positive result for the current stage are considered by the next stage of the cascade. This approach is significantly more efficient than using a single monolithic classifier, as it typically allows the majority of sub-windows to be rejected after evaluating only part of the cascade.

The classifiers are trained using a modified version of the AdaBoost learning algorithm, which selects the features used as input to the classifier as well as the weight and threshold making up the classifier itself.

3.1.2 Illumination Compensation

Variations in lighting level are known to decrease the effectiveness of face recognition systems. Histogram equalisation is a simple technique that has been found to compensate for moderate variations in lighting level (Bourlai et al., 2009).

Histogram equalisation is a transformation that can be performed on the intensity levels of a greyscale image to enhance contrast in the image. This is achieved by mapping each grey level in the source image onto a destination grey level which is determined by the proportion of pixels in the source image whose grey levels are less than or equal to the source grey level in question. This has the effect of spreading the grey levels in the image over a wider range, thus enhancing contrast.

For an image of size $R \times C$, with grey levels $0, 1, 2, \dots, L-1$, the mapping for a grey level k is given by:

$$T(k) = \frac{(L-1)}{RC} \sum_{j=0}^k n_j \quad (3.1)$$

where n_j denotes the number of pixels in the source image that have grey level j (Gonzalez and Woods, 2008).

3.2 Structure of the Recognition Process

The execution of the recognition system can be divided into two phases: training and recognition. The training phase consists of preparatory calculations specific to the particular recognition technique being used. The recognition phase then involves actually ascertaining the identity of an unknown image. The inputs to the recognition phase are:

- A set of reference images (one per individual known to the system), termed the *gallery*
- An image of unknown identity, termed the *probe*

The recognition algorithm then estimates the likelihood of each possible identity for the probe image. As its output, it produces a ranked list indicating the most likely identity, second most likely identity, and so forth.

3.3 Classical Eigenface Recognition

The central idea underlying the eigenface technique is that it is possible to express any given input image (of a fixed size) as a linear combination of a set of basis images chosen

ahead of time. By using PCA, it is possible to construct a basis for which a small subset of the basis images can capture most of the variance between face images. Such a subset can still provide a good representation of facial features, while having a much lower dimensionality than the full basis. This dimensionality reduction procedure then makes the technique computationally feasible; computing a complete basis for the image space would be quite impractical.

The feature space spanned by the chosen subset of basis images is dubbed “face space”, since this subspace (ideally) corresponds to the subset of images that are of human faces.

The technique described in this section is largely the same as the one presented by Turk and Pentland (1991), with the following three differences.

Firstly, for simplicity, this treatment does not make provision for detecting if the input image is of an individual not known to the system.

Secondly, in Turk and Pentland (1991), the dimensionality of the face space is chosen based on unspecified heuristics. Our discussion includes a systematic method for determining an appropriate size for the face space, as described in Yambor et al. (2002).

Finally, the Turk and Pentland study used the Euclidean distance metric as the dissimilarity measure for comparing feature vectors. In addition to the Euclidean distance, we discuss a number of alternative functions that have been proposed for this purpose in later works on the matter such as Yambor et al. (2002) and Perlibakas (2004).

3.3.1 Training

During the training process, the system takes as input a set of N preprocessed face images, each of dimensions $r \times c$. Each image is represented as a (column) vector \mathbf{t}_i of dimensionality $D = rc$.

The PCA technique proceeds as follows:

Firstly, the mean and covariance of the training data are determined:

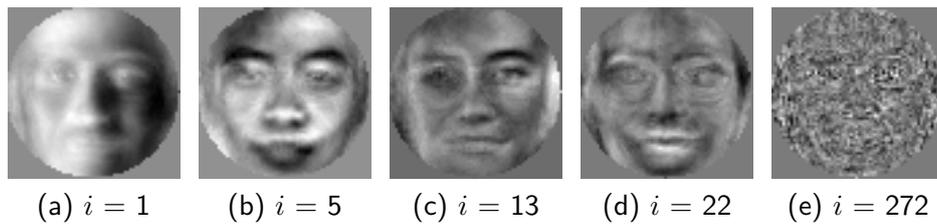
$$\bar{\mathbf{t}} = \frac{1}{N} \sum_{i=1}^N \mathbf{t}_i \quad (3.2)$$

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N (\mathbf{t}_i - \bar{\mathbf{t}}) (\mathbf{t}_i - \bar{\mathbf{t}})^T \quad (3.3)$$

We then solve the eigenvector problem:

$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (3.4)$$

yielding N solutions $(\mathbf{u}_i, \lambda_i)$. The solutions are sorted such that λ_1 is the largest eigenvalue, and λ_N is the smallest. The vectors \mathbf{u}_i are also referred to as *eigenfaces*, since (for low values of i) they appear face-like when considered as images. Figure 3.1 shows some example eigenfaces.

Figure 3.1: Example eigenfaces ($N = 272$)

It can be shown that the vectors $\{\mathbf{u}_i\}$ define a N -dimensional subspace of the original D -dimensional image space, such that each vector \mathbf{u}_i is in the direction of largest variance in the training set (while being orthogonal to the preceding vectors $\mathbf{u}_1 \dots \mathbf{u}_{i-1}$), and the variance in this direction is equal to λ_i . For proof of this result, see Bishop (2006).

While the above procedure allows for a substantial dimensionality reduction, it is possible to further reduce the size of the basis without adversely affecting its power to distinguish between the faces of different individuals. Since the eigenvectors corresponding to the lowest eigenvalues typically encode very little variance between facial images (Sirovich and Kirby, 1987), one might surmise that the data encoded by these eigenvectors is of marginal utility in distinguishing between individuals. This hypothesis is supported by the findings of Yambor et al. (2002), in which only the first 100 or so (out of 500) eigenvectors were observed to be significant to recognition, with the remaining eigenvectors having a marginal (or even negative) impact on accuracy.

To accomplish this additional reduction, we define a parameter $v \in [0, 1]$ indicating what fraction of variance (in the training data) the truncated basis should be capable of representing. The truncated basis then consists of the eigenvectors $\{\mathbf{u}_i\}_{i=1}^M$ where M is defined as:

$$M = \min \left\{ j \mid \sum_{i=1}^j \lambda_i \geq v \sum_{i=1}^N \lambda_i \right\} \quad (3.5)$$

The value of v is chosen empirically, with 0.95 and 0.99 being typical values.

The M -dimensional feature space spanned by the selected eigenvectors is the face space, into which input images are projected during feature extraction.

The final output of the training process consists of two components: the average face vector $\bar{\mathbf{t}}$, the eigenvectors $\{\mathbf{u}_i\}_{i=1}^M$ (the truncated basis), and the eigenvalues $\{\lambda_i\}_{i=1}^M$.

Algorithm 1 summarises the training process in pseudocode.

3.3.2 Feature Extraction

For the basic eigenface technique, feature extraction is simply a matter of projecting the input image into face space, using the basis computed during training.

Algorithm 1 Training algorithm for classical eigenface system

Input: Training images $\{\mathbf{t}_i\}$, variance fraction v

Output: Mean image $\bar{\mathbf{t}}$, eigenvectors $\{\mathbf{u}_i\}$, eigenvalues $\{\lambda_i\}$

- 1: $\bar{\mathbf{t}} \leftarrow \text{mean}(t_1, t_2, \dots, t_N)$
 - 2: $S \leftarrow \text{covar}(t_1, t_2, \dots, t_N)$
 - 3: $\{(u_i, \lambda_i)\}_{i=1}^N \leftarrow \text{solve}(Su = \lambda u)$
 - 4: $M \leftarrow \text{truncate_basis}(v; \lambda_1, \lambda_2, \dots, \lambda_N)$
 - 5: **return** $(\bar{\mathbf{t}}, \{(u_i, \lambda_i)\}_{i=1}^M)$
-

The feature vector \mathbf{y} for an image \mathbf{x} is computed as:

$$\mathbf{y} = \mathbf{U}(\mathbf{x} - \bar{\mathbf{t}}) \quad (3.6)$$

where \mathbf{U} is the $M \times D$ matrix whose rows are the eigenvectors $\{\mathbf{u}_i^T\}$.

3.3.3 Identification

Given a probe image \mathbf{x}^P and a gallery of reference images $\{\mathbf{x}_k\}$, we estimate the likelihood of each candidate identity by comparing \mathbf{y}^P (its projection into face space) with each gallery image's projection \mathbf{y}_k . This is done using a distance function $d(\mathbf{a}, \mathbf{b})$, which quantifies the degree of difference between two feature vectors \mathbf{a} and \mathbf{b} .

The most likely identity k is taken to be that for which the quantity $d(\mathbf{y}_k, \mathbf{y}^P)$ is the minimum. The second most likely identity is that for which this quantity has the second-smallest value, and so on.

Algorithm 2 provides pseudocode summarising the recognition process.

Algorithm 2 Recognition algorithm for classical eigenface system

Input: Probe \mathbf{x}^P , gallery $\{\mathbf{x}_k\}$, mean image $\bar{\mathbf{t}}$, eigenvectors $\{\mathbf{u}_i\}$, eigenvalues $\{\lambda_i\}$

Output: Most likely identity k for \mathbf{x}^P

- 1: $y^P \leftarrow \text{project}(x^P)$
 - 2: **for** each image x_k in gallery **do**
 - 3: $y_k \leftarrow \text{project}(x_k)$
 - 4: $\text{ranked}[k].\text{dist} \leftarrow d(y_k, y^P)$
 - 5: $\text{ranked}[k].\text{id} \leftarrow k$
 - 6: **end for**
 - 7: sort ranked in ascending order of dist
 - 8: **return** $\text{ranked}[1].\text{id}$
-

There are a number of different functions that can be used as distance measures. The remainder of this section presents those used in this study.

In the discussion that follows, \mathbf{a} and \mathbf{b} are column vectors (each of dimensionality M), with a_i denoting the i^{th} entry of \mathbf{a} and b_i denoting the i^{th} entry of \mathbf{b} .

The simplest distance functions are the Manhattan (L_1) and Euclidean (L_2) distances, defined in (3.7) and (3.8) respectively (Duda et al., 2001).

$$d_1(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^M |a_i - b_i| \quad (3.7)$$

$$d_2(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^M (a_i - b_i)^2} \quad (3.8)$$

A more sophisticated distance measure is the Mahalanobis distance (Bishop, 2006), which normalises the contribution of each dimension to the total distance by taking into account the covariance of the distribution from which the input vectors are drawn. The Mahalanobis distance is defined as follows:

$$d_M(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a} - \mathbf{b})^T \boldsymbol{\Sigma}^{-1} (\mathbf{a} - \mathbf{b})} \quad (3.9)$$

where $\boldsymbol{\Sigma}$ is the aforementioned covariance matrix.

In the specific case where \mathbf{a} and \mathbf{b} are eigenface feature vectors, it is possible to obtain a simpler formulation that can be computed more efficiently. Since the eigenvectors constituting the face space are orthogonal, the covariance matrix $\boldsymbol{\Sigma}$ will be diagonal. As stated earlier, the variance in each dimension is equal to the corresponding eigenvalue λ_i , giving us the values on the diagonal. From this, we can express the distance as:

$$d_M(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^M \frac{(a_i - b_i)^2}{\lambda_i}} \quad (3.10)$$

In addition to these three well-known distance metrics, the study evaluates two distance functions put forward specifically for the purpose of eigenface-based face recognition.

First, there is an alternative formulation of the Mahalanobis distance proposed by Yambor et al. (2002):

$$d_Y(\mathbf{a}, \mathbf{b}) = - \sum_{i=1}^M \frac{a_i b_i}{\sqrt{\lambda_i}} \quad (3.11)$$

It should be noted that the function defined in (3.11) is not equivalent to the canonical Mahalanobis distance defined by (3.9), as can be seen by examining each function's behaviour when one of the parameters is the zero vector.

Additionally, it lacks certain properties conventionally associated with distance metrics. For example, it admits negative distances, and does not satisfy the property:

$$d_Y(\mathbf{a}, \mathbf{b}) = 0 \iff \mathbf{a} = \mathbf{b} \quad (3.12)$$

However, this does not necessarily mean that it is unsuitable as a distance function for the purpose of comparing eigenface feature vectors.

The other function is the “modified SSE-based distance” proposed by Perlibakas (2004):

$$d_P(\mathbf{a}, \mathbf{b}) = \frac{\sum_{i=1}^M (a_i - b_i)^2}{\left(\sum_{i=1}^M a_i^2\right) \left(\sum_{i=1}^M b_i^2\right)} = \frac{\|\mathbf{a} - \mathbf{b}\|^2}{\|\mathbf{a}\|^2 \|\mathbf{b}\|^2} \quad (3.13)$$

It should be noted that this function is undefined when either \mathbf{a} or \mathbf{b} is the zero vector (which would occur if an input image matched the average face $\bar{\mathbf{t}}$ exactly). While this is not a case that one can expect to arise frequently, it is a possibility that would need to be taken into account in the development of a robust face recognition system for use in real-world applications.

3.4 Probabilistic Eigenface Recognition

This section details how the eigenfaces representation can be adapted to make use of probabilistic similarity measures. Here, instead of extracting features from each image in isolation, the system considers the differences between pairs of images. Such a difference can belong to one of two classes – the *intrapersonal* class (for a difference between two images of the same individual), or the *extrapersonal* class (for a difference between images of two different individuals). The system operates by estimating the probability that a given difference belongs to the intrapersonal class.

Rather than making use of a single face space, the probabilistic scheme considers two “face difference spaces” – one for the intrapersonal class, and another for the extrapersonal class. Figure 3.2 shows the mean differences for the intrapersonal and extrapersonal classes, for the training data used in experiment 3 of chapter 5.

The technique detailed in this section is based on the one presented by Moghaddam et al. (1996), with the following three differences.

Firstly, it makes direct use of image intensity differences, rather than employing the deformable 3D mesh representation used by Moghaddam et al.

Secondly, as an alternative to the “maximum *a posteriori*” classification rule, we incorporate the “maximum likelihood” rule used by Moghaddam and Pentland (1997).

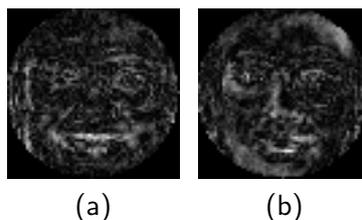


Figure 3.2: Mean image differences for the (a) intrapersonal and (b) extrapersonal classes.

Finally, as with the non-probabilistic technique, this treatment includes a systematic method for determining appropriate dimensionalities for the face difference spaces (Yamorb et al., 2002).

3.4.1 Training

The input for the training process consists of two sets of image pairs: the intrapersonal set T_I (of size N_I), and the extrapersonal set T_E (of size N_E). Each pair in T_I is composed of two distinct images $\mathbf{t}_{i1}^{(I)}$ and $\mathbf{t}_{i2}^{(I)}$ of the same individual, while each pair in T_E is composed of two images $\mathbf{t}_{i1}^{(E)}$ and $\mathbf{t}_{i2}^{(E)}$ of different individuals. As before, each $\mathbf{t}_{ij}^{(?)}$ is a vector of dimensionality D .

The training process begins with the construction of a set of image differences $\{\boldsymbol{\tau}_i^{(I)}\}$ from the pairs in T_I . Each $\boldsymbol{\tau}_i^{(I)}$ is calculated as:

$$\boldsymbol{\tau}_i^{(I)} = \mathbf{t}_{i1}^{(I)} - \mathbf{t}_{i2}^{(I)} \quad (3.14)$$

We then apply the PCA technique described in section 3.3 to these differences. The first step is to calculate the mean intrapersonal difference vector and associated covariance matrix:

$$\bar{\boldsymbol{\tau}}^{(I)} = \frac{1}{N_I} \sum_{i=1}^{N_I} \boldsymbol{\tau}_i^{(I)} \quad (3.15)$$

$$\mathbf{S}^{(I)} = \frac{1}{N_I} \sum_{i=1}^{N_I} (\boldsymbol{\tau}_i^{(I)} - \bar{\boldsymbol{\tau}}^{(I)}) (\boldsymbol{\tau}_i^{(I)} - \bar{\boldsymbol{\tau}}^{(I)})^T \quad (3.16)$$

Then, proceeding in a similar manner as in section 3.3, we solve the following eigen-vector problem:

$$\mathbf{S}^{(I)} \mathbf{u}_i^{(I)} = \lambda_i^{(I)} \mathbf{u}_i^{(I)} \quad (3.17)$$

yielding N_I solutions $(\mathbf{u}_i^{(I)}, \lambda_i^{(I)})$.

Finally, we compute a truncated basis by retaining the M_I eigenvectors corresponding to the largest eigenvalues, where M_I is defined as:

$$M_I = \min \left\{ j \left| \sum_{i=1}^j \lambda_i^{(I)} \geq v \sum_{i=1}^{N_I} \lambda_i^{(I)} \right. \right\} \quad (3.18)$$

In this equation (as in the non-probabilistic technique), v is an empirically-chosen value in the range $[0, 1]$. This value specifies what fraction of the variance in the training data the truncated basis will be capable of representing.

The end result of this process is the mean intrapersonal difference $\bar{\boldsymbol{\tau}}^{(I)}$, the eigenvectors $\{\mathbf{u}_i^{(I)}\}_{i=1}^{M_I}$ spanning the intrapersonal face difference space, and the eigenvalues $\{\lambda_i^{(I)}\}_{i=1}^{M_I}$.

The process is then repeated with the pairs from T_E , yielding the mean extrapersonal difference $\bar{\tau}^{(E)}$, eigenvectors $\{\mathbf{u}_i^{(E)}\}_{i=1}^{M_E}$, and eigenvalues $\{\lambda_i^{(E)}\}_{i=1}^{M_E}$.

Algorithms 3 and 4 provide pseudocode summarising the training process.

Algorithm 3 Training algorithm for probabilistic eigenface system (intrapersonal half)

Input: Training images $\{(\mathbf{t}_{i1}^{(I)}, \mathbf{t}_{i2}^{(I)})\}$, variance fraction v

Output: Mean delta $\bar{\tau}^{(I)}$, eigenvectors $\{\mathbf{u}_i^{(I)}\}$, eigenvalues $\{\lambda_i^{(I)}\}$, mean eigenvalue ρ_I

```

1: for  $i = 1$  to  $N_I$  do
2:    $\tau_i^{(I)} \leftarrow t_{i1}^{(I)} - t_{i2}^{(I)}$ 
3: end for
4:  $\bar{\tau}^{(I)} \leftarrow \text{mean}(\tau_1^{(I)}, \tau_2^{(I)}, \dots, \tau_{N_I}^{(I)})$ 
5:  $S^{(I)} \leftarrow \text{covar}(\tau_1^{(I)}, \tau_2^{(I)}, \dots, \tau_{N_I}^{(I)})$ 
6:  $\{(u_i^{(I)}, \lambda_i^{(I)})\}_{i=1}^{N_I} \leftarrow \text{solve}(S^{(I)}u = \lambda u)$ 
7:  $M_I \leftarrow \text{truncate\_basis}(v; \lambda_1^{(I)}, \lambda_2^{(I)}, \dots, \lambda_{N_I}^{(I)})$ 
8:  $\{\lambda_i^{(I)}\}_{i=N_I+1}^D \leftarrow \text{extrapolate}(\lambda_1^{(I)}, \lambda_2^{(I)}, \dots, \lambda_{N_I}^{(I)})$ 
9:  $\rho_I \leftarrow \text{mean}(\lambda_{M_I+1}^{(I)}, \lambda_{M_I+2}^{(I)}, \dots, \lambda_D^{(I)})$ 
10: return  $(\bar{\tau}^{(I)}, \{(u_i^{(I)}, \lambda_i^{(I)})\}_{i=1}^{M_I}, \rho_I)$ 

```

Algorithm 4 Training algorithm for probabilistic eigenface system (extrapersonal half)

Input: Training images $\{(\mathbf{t}_{i1}^{(E)}, \mathbf{t}_{i2}^{(E)})\}$, variance fraction v

Output: Mean delta $\bar{\tau}^{(E)}$, eigenvectors $\{\mathbf{u}_i^{(E)}\}$, eigenvalues $\{\lambda_i^{(E)}\}$, mean eigenvalue ρ_E

```

1: for  $i = 1$  to  $N_E$  do
2:    $\tau_i^{(E)} \leftarrow t_{i1}^{(E)} - t_{i2}^{(E)}$ 
3: end for
4:  $\bar{\tau}^{(E)} \leftarrow \text{mean}(\tau_1^{(E)}, \tau_2^{(E)}, \dots, \tau_{N_E}^{(E)})$ 
5:  $S^{(E)} \leftarrow \text{covar}(\tau_1^{(E)}, \tau_2^{(E)}, \dots, \tau_{N_E}^{(E)})$ 
6:  $\{(u_i^{(E)}, \lambda_i^{(E)})\}_{i=1}^{N_E} \leftarrow \text{solve}(S^{(E)}u = \lambda u)$ 
7:  $M_E \leftarrow \text{truncate\_basis}(v; \lambda_1^{(E)}, \lambda_2^{(E)}, \dots, \lambda_{N_E}^{(E)})$ 
8:  $\{\lambda_i^{(E)}\}_{i=N_E+1}^D \leftarrow \text{extrapolate}(\lambda_1^{(E)}, \lambda_2^{(E)}, \dots, \lambda_{N_E}^{(E)})$ 
9:  $\rho_E \leftarrow \text{mean}(\lambda_{M_E+1}^{(E)}, \lambda_{M_E+2}^{(E)}, \dots, \lambda_D^{(E)})$ 
10: return  $(\bar{\tau}^{(E)}, \{(u_i^{(E)}, \lambda_i^{(E)})\}_{i=1}^{M_E}, \rho_E)$ 

```

3.4.2 Probability Estimation

Given two face images \mathbf{f} and \mathbf{g} , we seek to estimate the probability that the images depict the same individual, or equivalently that the difference $\delta = \mathbf{f} - \mathbf{g}$ belongs to the

intrapersonal face difference space.

This *a posteriori* probability can be calculated using Bayes' theorem (Bishop, 2006):

$$P(\Omega_I|\boldsymbol{\delta}) = \frac{P(\boldsymbol{\delta}|\Omega_I)P(\Omega_I)}{P(\boldsymbol{\delta}|\Omega_I)P(\Omega_I) + P(\boldsymbol{\delta}|\Omega_E)P(\Omega_E)} \quad (3.19)$$

where Ω_I denotes the intrapersonal face difference space, and Ω_E correspondingly denotes the extrapersonal space.

The probabilities $P(\Omega_I)$ and $P(\Omega_E)$ must be chosen *a priori*. Choosing to set $P(\Omega_I) = P(\Omega_E) = 0.5$ allows the computation to be simplified to the following:

$$P(\Omega_I|\boldsymbol{\delta}) = \frac{P(\boldsymbol{\delta}|\Omega_I)}{P(\boldsymbol{\delta}|\Omega_I) + P(\boldsymbol{\delta}|\Omega_E)} \quad (3.20)$$

Under the assumption that Ω_I follows a Gaussian distribution, the likelihood of encountering a given image difference is

$$P(\boldsymbol{\delta}|\Omega_I) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{\delta} - \bar{\boldsymbol{\delta}}_I)^T \boldsymbol{\Sigma}_I^{-1} (\boldsymbol{\delta} - \bar{\boldsymbol{\delta}}_I)\right)}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_I|^{\frac{1}{2}}} \quad (3.21)$$

where $\bar{\boldsymbol{\delta}}_I$ is the mean of Ω_I , and $\boldsymbol{\Sigma}_I$ is the associated covariance matrix.

Assuming that $\boldsymbol{\Sigma}_I$ is not singular, it can be diagonalised using the D -dimensional orthonormal basis formed by its eigenvectors:

$$\boldsymbol{\Sigma}_I = \boldsymbol{\Phi}_I \boldsymbol{\Lambda}_I \boldsymbol{\Phi}_I^T \quad (3.22)$$

where $\boldsymbol{\Phi}_I$ is the (orthogonal) matrix whose rows are the eigenvectors of $\boldsymbol{\Sigma}_I$, and $\boldsymbol{\Lambda}_I$ is the diagonal matrix composed of the corresponding eigenvalues.

Adopting the shorthand $\tilde{\boldsymbol{\delta}}_I = \boldsymbol{\delta} - \bar{\boldsymbol{\delta}}_I$, the product in the numerator of (3.21) can be simplified as follows:

$$\begin{aligned} \tilde{\boldsymbol{\delta}}_I^T \boldsymbol{\Sigma}_I^{-1} \tilde{\boldsymbol{\delta}}_I &= \tilde{\boldsymbol{\delta}}_I^T \left(\boldsymbol{\Phi}_I \boldsymbol{\Lambda}_I \boldsymbol{\Phi}_I^T \right)^{-1} \tilde{\boldsymbol{\delta}}_I \\ &= \tilde{\boldsymbol{\delta}}_I^T \left(\boldsymbol{\Phi}_I^T \boldsymbol{\Lambda}_I^{-1} \boldsymbol{\Phi}_I \right) \tilde{\boldsymbol{\delta}}_I \\ &= \left(\tilde{\boldsymbol{\delta}}_I^T \boldsymbol{\Phi}_I^T \right) \boldsymbol{\Lambda}_I^{-1} \left(\boldsymbol{\Phi}_I \tilde{\boldsymbol{\delta}}_I \right) \\ &= \mathbf{v}^T \boldsymbol{\Lambda}_I^{-1} \mathbf{v} \end{aligned} \quad (3.23)$$

where $\mathbf{v} = \boldsymbol{\Phi}_I \tilde{\boldsymbol{\delta}}_I$ is the projection of $\boldsymbol{\delta}$ onto the eigenvector basis. Since $\boldsymbol{\Lambda}_I$ is diagonal, the above expression can be simplified further:

$$\mathbf{v}^T \boldsymbol{\Lambda}_I^{-1} \mathbf{v} = \sum_{i=1}^D \frac{v_i^2}{\lambda_i^{(I)}} \quad (3.24)$$

The fact that $\mathbf{\Lambda}_I$ is diagonal also allows the denominator of (3.21) to be simplified:

$$|\mathbf{\Sigma}_I| = |\mathbf{\Lambda}_I| = \prod_{i=1}^D \lambda_i^{(I)} \quad (3.25)$$

Thus (3.21) can be reformulated as:

$$P(\boldsymbol{\delta}|\Omega_I) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^D \frac{v_i^2}{\lambda_i^{(I)}}\right)}{(2\pi)^{\frac{D}{2}} \prod_{i=1}^D \left(\lambda_i^{(I)}\right)^{\frac{1}{2}}} \quad (3.26)$$

It must be noted that the discussion thus far implicitly assumes the presence of a full D -dimensional basis for the face difference space. However, in practice it is computationally infeasible to construct such a basis, so it is necessary to develop an approximation that can be evaluated using only the truncated M_I -dimensional basis. We begin by observing that $P(\boldsymbol{\delta}|\Omega_I)$ can be expressed as the product of two independent Gaussian densities:

$$\begin{aligned} P(\boldsymbol{\delta}|\Omega_I) &= P_F(\boldsymbol{\delta}|\Omega_I)P_{\bar{F}}(\boldsymbol{\delta}|\Omega_I) \\ &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M_I} \frac{v_i^2}{\lambda_i^{(I)}}\right)}{(2\pi)^{\frac{M_I}{2}} \prod_{i=1}^{M_I} \left(\lambda_i^{(I)}\right)^{\frac{1}{2}}} \right] \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=M_I+1}^D \frac{v_i^2}{\lambda_i^{(I)}}\right)}{(2\pi)^{\frac{D-M_I}{2}} \prod_{i=M_I+1}^D \left(\lambda_i^{(I)}\right)^{\frac{1}{2}}} \right] \end{aligned} \quad (3.27)$$

Here, $P_F(\boldsymbol{\delta}|\Omega_I)$ is the marginal density for the subspace spanned by the highest-ranked eigenvectors $\{\mathbf{u}_i\}_{i=1}^{M_I}$, and $P_{\bar{F}}(\boldsymbol{\delta}|\Omega_I)$ is the marginal density for the complementary subspace spanned by the remaining eigenvectors.

$P_F(\boldsymbol{\delta}|\Omega_I)$ can be calculated directly, but an approximation for $P_{\bar{F}}(\boldsymbol{\delta}|\Omega_I)$ is required. To this end, we substitute a single value ρ_I in place of the unknown eigenvalues:

$$\hat{P}_{\bar{F}}(\boldsymbol{\delta}|\Omega_I) = \frac{\exp\left(-\frac{1}{2}\sum_{i=M_I+1}^D \frac{v_i^2}{\rho_I}\right)}{(2\pi\rho_I)^{\frac{D-M_I}{2}}} \quad (3.28)$$

The optimal value for ρ_I is the mean of the unknown eigenvalues (Moghaddam and Pentland, 1997):

$$\rho_I = \frac{\sum_{i=M_I+1}^D \lambda_i^{(I)}}{D - M_I} \quad (3.29)$$

Since the eigenvalues in question are unknown, it is not possible to evaluate this mean exactly. However, the unknown eigenvalues can be estimated by fitting a function

to the known eigenvalues, and then extrapolating the fitted function.

Prior literature proposes various types of functions for this purpose. Moghaddam and Pentland (1997) suggest the use of a function of the form $\frac{1}{f(i)}$, while the formulation presented by Ramanathan and Chellappa (2006) makes use of a cubic spline.

In the experiments conducted in this study (see chapter 5), it was found that a good fit for the observed eigenvalues was obtained by using a function of the form

$$\lambda(i) = \frac{a}{i} \quad (3.30)$$

While (3.28) provides a simpler expression for $\hat{P}_{\bar{F}}(\boldsymbol{\delta}|\Omega_I)$, it still depends on the lower-rank eigenvectors for calculating the PCA coefficients $\{v_i\}_{i>M_I}$. This can be remedied by making use of the fact that $\boldsymbol{\Phi}_I$ is an orthogonal matrix, and therefore acts as an isometry:

$$\|\tilde{\boldsymbol{\delta}}_I\|^2 = \|\mathbf{v}\|^2 = \sum_{i=1}^D v_i^2 \quad (3.31)$$

From this, we can calculate a quantity known as the *residual reconstruction error*:

$$\epsilon_I^2(\boldsymbol{\delta}) = \sum_{i=M_I+1}^D v_i^2 = \|\tilde{\boldsymbol{\delta}}_I\|^2 - \sum_{i=1}^{M_I} v_i^2 \quad (3.32)$$

Equation (3.28) can then be reformulated as follows:

$$\hat{P}_{\bar{F}}(\boldsymbol{\delta}|\Omega_I) = \frac{\exp\left(-\frac{\epsilon_I^2(\boldsymbol{\delta})}{2\rho_I}\right)}{(2\pi\rho_I)^{\frac{D-M_I}{2}}} \quad (3.33)$$

Note that this formulation depends only on the first M_I eigenvalues and eigenvectors, so it can be evaluated using the truncated basis constructed during training. We can thus express the estimated overall probability in a form that can be computed efficiently:

$$\hat{P}(\boldsymbol{\delta}|\Omega_I) = P_F(\boldsymbol{\delta}|\Omega_I)\hat{P}_{\bar{F}}(\boldsymbol{\delta}|\Omega_I) = \frac{\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{M_I} \frac{v_i^2}{\lambda_i^{(I)}} + \frac{\epsilon_I^2(\boldsymbol{\delta})}{\rho_I}\right]\right)}{(2\pi)^{\frac{D}{2}} \rho_I^{\frac{D-M_I}{2}} \prod_{i=1}^{M_I} (\lambda_i^{(I)})^{\frac{1}{2}}} \quad (3.34)$$

The extrapersonal likelihood estimate $\hat{P}(\boldsymbol{\delta}|\Omega_E)$ can then be calculated similarly:

$$\hat{P}(\boldsymbol{\delta}|\Omega_E) = \frac{\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{M_E} \frac{w_i^2}{\lambda_i^{(E)}} + \frac{\epsilon_E^2(\boldsymbol{\delta})}{\rho_E}\right]\right)}{(2\pi)^{\frac{D}{2}} \rho_E^{\frac{D-M_E}{2}} \prod_{i=1}^{M_E} (\lambda_i^{(E)})^{\frac{1}{2}}} \quad (3.35)$$

using $\mathbf{w} = \boldsymbol{\Phi}_E \tilde{\boldsymbol{\delta}}_E$ in place of $\mathbf{v} = \boldsymbol{\Phi}_I \tilde{\boldsymbol{\delta}}_I$.

Finally, the intra- and extrapersonal likelihoods can be combined to yield an estimate for the *a posteriori* probability $P(\Omega_I|\boldsymbol{\delta})$ using (3.20).

3.4.3 Identification

Based on the mathematical framework laid out in the preceding subsections, this subsection describes two identification schemes: the maximum *a posteriori* (MAP) scheme, which makes use of the *a posteriori* probability $P(\Omega_I|\boldsymbol{\delta})$, and the maximum likelihood (ML) scheme, which relies on the image difference likelihood $P(\boldsymbol{\delta}|\Omega_I)$.

The MAP scheme proceeds in a similar manner to the classical eigenfaces scheme. Given a probe image \mathbf{x}^P and a gallery of reference images $\{\mathbf{x}_k\}$, we compute the difference $\boldsymbol{\delta}_k = \mathbf{x}^P - \mathbf{x}_k$ for each \mathbf{x}_k , and rank the candidate identities by their estimated *a posteriori* probability, $\hat{P}(\Omega_I|\boldsymbol{\delta}_k)$. The most likely identity k is then taken to be that for which the probability is highest, the second most likely identity being that with the second-highest probability, and so forth.

For each image difference $\boldsymbol{\delta}_k$, the *a posteriori* probability is calculated as follows.

First, the difference is projected into the intrapersonal difference space:

$$\tilde{\boldsymbol{\delta}}_I = \boldsymbol{\delta}_k - \bar{\boldsymbol{\tau}}^{(I)} \quad (3.36)$$

$$\mathbf{v} = \boldsymbol{\Phi}_I \tilde{\boldsymbol{\delta}}_I \quad (3.37)$$

Here, $\boldsymbol{\Phi}_I$ is the $M_I \times D$ matrix whose rows are the (transposed) eigenvectors $\{\mathbf{u}_i^{(I)}\}$.

The results of this calculation are then used to determine the intrapersonal likelihood $\hat{P}(\boldsymbol{\delta}|\Omega_I)$ using (3.34).

Then, a similar process is followed for the extrapersonal difference space:

$$\tilde{\boldsymbol{\delta}}_E = \boldsymbol{\delta}_k - \bar{\boldsymbol{\tau}}^{(E)} \quad (3.38)$$

$$\mathbf{w} = \boldsymbol{\Phi}_E \tilde{\boldsymbol{\delta}}_E \quad (3.39)$$

after which the extrapersonal likelihood $\hat{P}(\boldsymbol{\delta}|\Omega_E)$ is determined, using (3.35).

Finally, the *a posteriori* probability estimate $\hat{P}(\Omega_I|\boldsymbol{\delta}_k)$ is calculated, as described in (3.20).

The ML scheme is largely identical to the MAP scheme. The sole difference is that candidate identities are ranked using the estimated image difference likelihood $\hat{P}(\boldsymbol{\delta}_k|\Omega_I)$, rather than the *a posteriori* probability. This provides a coarser approximation of the actual probability that a given image pair belongs to the intrapersonal class Ω_I . However, it has the advantage of being less computationally intensive than the MAP scheme, since it is not necessary to evaluate $\hat{P}(\boldsymbol{\delta}_k|\Omega_E)$.

Algorithm 5 summarises the identification process in pseudocode.

Algorithm 5 Recognition algorithm for probabilistic eigenface system

Input: Probe \mathbf{x}^P , gallery $\{\mathbf{x}_k\}$, mean deltas $\bar{\tau}^{(I)}$ and $\bar{\tau}^{(E)}$, eigenvectors $\{\mathbf{u}_i^{(I)}\}$ and $\{\mathbf{u}_i^{(E)}\}$, eigenvalues $\{\lambda_i^{(I)}\}$ and $\{\lambda_i^{(E)}\}$, mean eigenvalue estimates ρ_I and ρ_E
Output: Most likely identity k for \mathbf{x}^P

```

1: for each image  $x_k$  in gallery do
2:    $\delta_k \leftarrow x^P - x_k$ 
3:    $P_I \leftarrow \text{likelihood}(\delta_k; \bar{\tau}^{(I)}, \{u_i^{(I)}\}_{i=1}^{M_I}, \{\lambda_i^{(I)}\}_{i=1}^{M_I}, \rho_I)$  // eq. (3.34)
4:    $P_E \leftarrow \text{likelihood}(\delta_k; \bar{\tau}^{(E)}, \{u_i^{(E)}\}_{i=1}^{M_E}, \{\lambda_i^{(E)}\}_{i=1}^{M_E}, \rho_E)$  // eq. (3.35)
5:   if using MAP scheme then
6:      $\text{ranked}[k].\text{probability} \leftarrow P_I / (P_I + P_E)$ 
7:   else
8:      $\text{ranked}[k].\text{probability} \leftarrow P_I$ 
9:   end if
10:   $\text{ranked}[k].\text{id} \leftarrow k$ 
11: end for
12: sort  $\text{ranked}$  in descending order of probability
13: return  $\text{ranked}[1].\text{id}$ 

```

3.5 Conclusion

This chapter has outlined the general structure of the recognition system used in this study, and presented detailed mathematical and algorithmic descriptions of the techniques used – classical eigenface recognition (Turk and Pentland, 1991) and probabilistic eigenface recognition (Moghaddam et al., 1996). Also included are descriptions of various ancillary preprocessing techniques used by the system.

CHAPTER 4

IMPLEMENTATION NOTES

This chapter provides a brief discussion of various matters concerning the implementation of the algorithms described in chapter 3.

4.1 Library Use

The face recognition system implemented in this study makes extensive use of the OpenCV computer vision library (website <http://opencv.willowgarage.com/>). This library is used most significantly for the efficient computation of PCA eigenvectors and eigenvalues (using the technique described in Bishop (2006)), but also for face localisation, as well as less complex tasks such as image resizing and histogram equalisation.

Concerning face localisation, the library provides an implementation of the Viola-Jones face detector (discussed in section 3.1.1), including data describing the features for a pre-trained classifier. For this reason, the system does not need a facility for training its own face detector.

4.2 Preprocessing Pipeline

As noted earlier, each input image presented to the system undergoes a series of preprocessing transformations to enhance image quality, before it is used for training or recognition. The details of these transformations are described in Figure 4.1. Figure 4.2 illustrates the process, showing an example image after each stage of the pipeline.

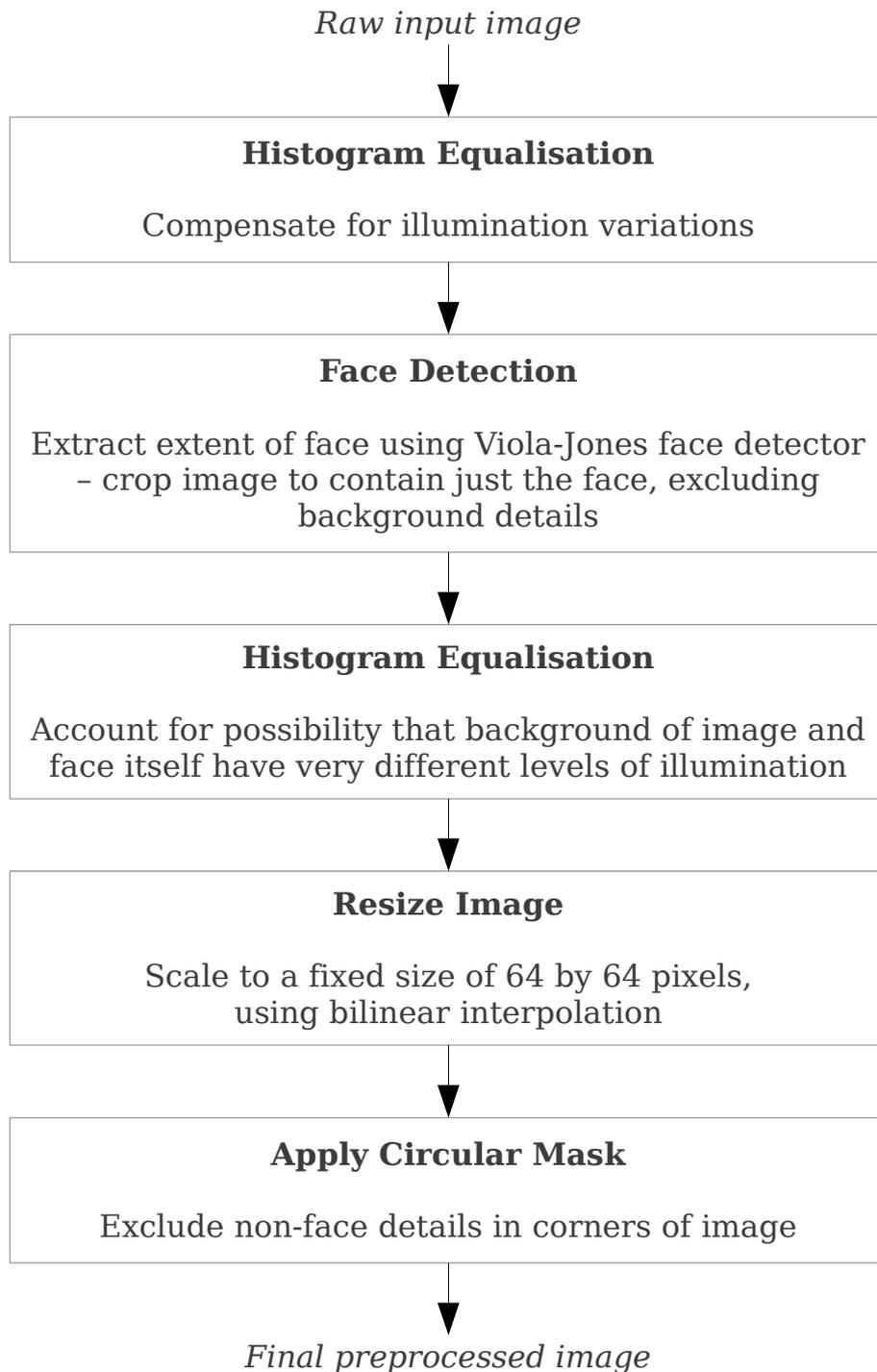


Figure 4.1: The preprocessing pipeline

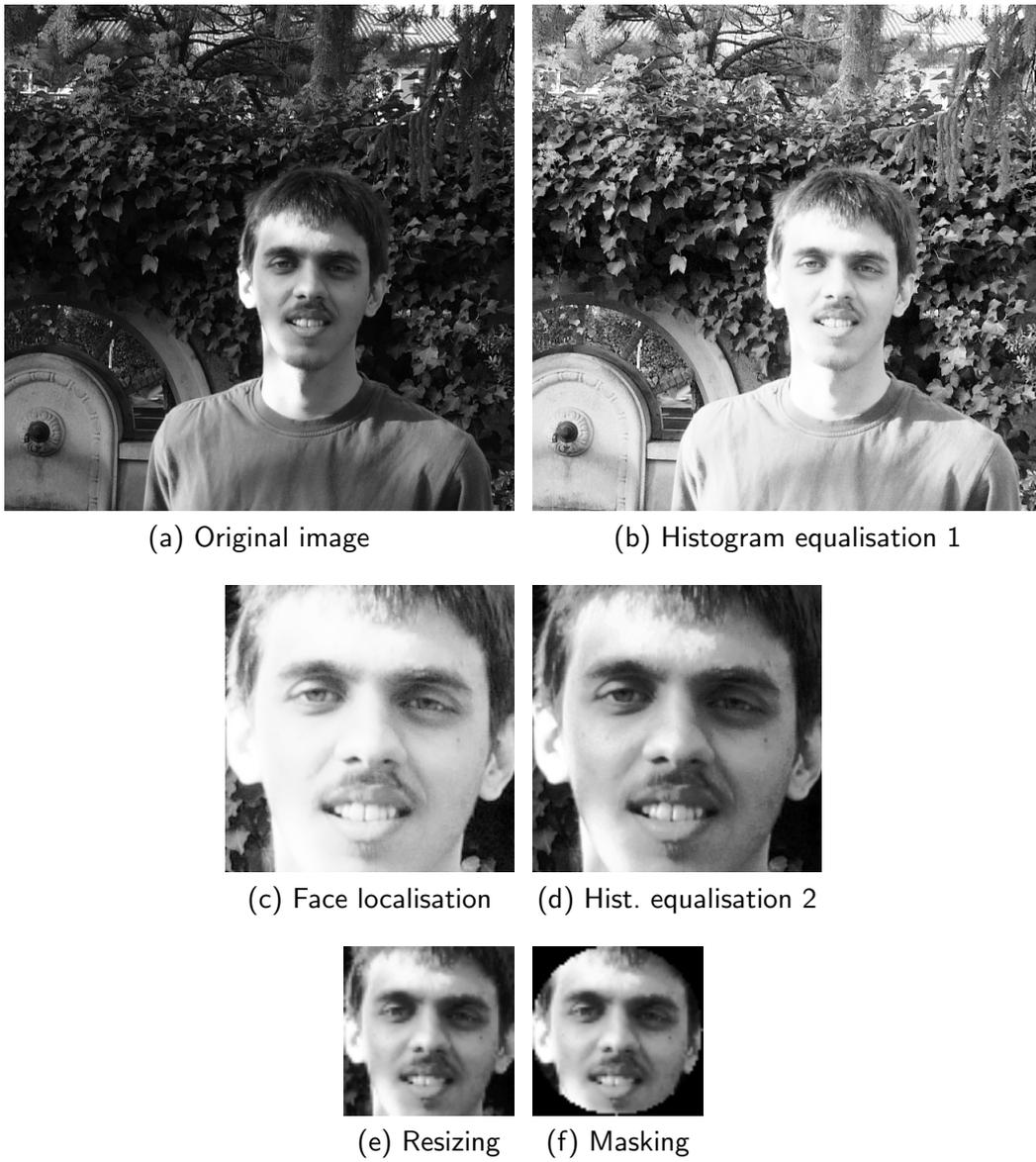


Figure 4.2: Operation of the preprocessing pipeline on an example image

4.3 Numerical Precision Concerns

In the implementation of the probabilistic technique described in section 3.4, it was found that for certain inputs, the probabilities $\hat{P}(\boldsymbol{\delta}|\Omega_I)$ and $\hat{P}(\boldsymbol{\delta}|\Omega_E)$ are too small to be accurately represented using an IEEE-754 64-bit binary floating point data type.

In order to resolve this problem, it is necessary to employ a modified formulation of the identification schemes discussed in subsection 3.4.3.

Rather than ranking the candidate identities for an image using the relevant probability directly, we employ a quantity which will give the same ordering of identities in the absence of precision limitations, and is more amenable to accurate computation.

4.3.1 Maximum Likelihood Formulation

For the maximum likelihood scheme, we can simply use the logarithm of the likelihood:

$$L_I = \log \hat{P}(\boldsymbol{\delta}|\Omega_I) \quad (4.1)$$

Since the logarithm function is strictly increasing, ranking candidate identities using this quantity will produce the same ordering as using the likelihood itself.

From (3.34), the calculation can be simplified as follows:

$$\begin{aligned} L_I &= \log \left\{ \frac{\exp \left(-\frac{1}{2} \left[\sum_{i=1}^{M_I} \frac{v_i^2}{\lambda_i^{(I)}} + \frac{\epsilon_I^2(\boldsymbol{\delta})}{\rho_I} \right] \right)}{(2\pi)^{\frac{D}{2}} \rho_I^{\frac{D-M_I}{2}} \prod_{i=1}^{M_I} (\lambda_i^{(I)})^{\frac{1}{2}}} \right\} \\ &= -\frac{1}{2} \left(\sum_{i=1}^{M_I} \frac{v_i^2}{\lambda_i^{(I)}} + \frac{\epsilon_I^2(\boldsymbol{\delta})}{\rho_I} \right) - \log \left([2\pi]^{\frac{D}{2}} \rho_I^{\frac{D-M_I}{2}} \prod_{i=1}^{M_I} [\lambda_i^{(I)}]^{\frac{1}{2}} \right) \\ &= -\frac{1}{2} \left(\sum_{i=1}^{M_I} \frac{v_i^2}{\lambda_i^{(I)}} + \frac{\epsilon_I^2(\boldsymbol{\delta})}{\rho_I} + D \log 2\pi + [D - M_I] \log \rho_I + \sum_{i=1}^{M_I} \log [\lambda_i^{(I)}] \right) \quad (4.2) \end{aligned}$$

By avoiding exponentiation altogether, this formulation avoids producing problematically small quantities (either in the final output, or in any intermediate results), thereby allowing for improved accuracy.

4.3.2 Maximum *A Posteriori* Formulation

With the maximum *a posteriori* scheme, the approach required is somewhat more subtle. Based on (3.20), we can derive the following:

$$\begin{aligned}
\hat{P}(\Omega_I|\boldsymbol{\delta}) &= \frac{1}{1 + \frac{\hat{P}(\boldsymbol{\delta}|\Omega_E)}{\hat{P}(\boldsymbol{\delta}|\Omega_I)}} \\
&= \frac{1}{1 + \exp\left(\log \hat{P}(\boldsymbol{\delta}|\Omega_E) - \log \hat{P}(\boldsymbol{\delta}|\Omega_I)\right)} \\
&= \frac{1}{1 + \exp(L_E - L_I)} \tag{4.3}
\end{aligned}$$

with L_I as in (4.1), and L_E defined analogously.

From this, it can be shown that $(L_I - L_E)$ is a suitable alternative ranking criterion, guaranteed to produce the same ordering of candidate identities as using the *a posteriori* probability. This follows from the fact that the function

$$f(x) = \frac{1}{1 + \exp(-x)}$$

is strictly increasing. Proof of this fact is given below.

Lemma 1. *The following function is strictly increasing:*

$$f(x) = \frac{1}{1 + \exp(-x)}$$

Proof.

$$\begin{aligned}
x_1 > x_2 &\Rightarrow -x_1 < -x_2 \\
&\Rightarrow \exp(-x_1) < \exp(-x_2) \\
&\Rightarrow 1 + \exp(-x_1) < 1 + \exp(-x_2) \\
&\Rightarrow \frac{1}{1 + \exp(-x_1)} > \frac{1}{1 + \exp(-x_2)} \\
&\Rightarrow f(x_1) > f(x_2)
\end{aligned}$$

□

CHAPTER 5

EXPERIMENTAL RESULTS

This chapter presents the experimental results that have been obtained using systems implementing the techniques described in chapter 3. We begin with a description of the datasets used in the experiments, followed by the results of each experiment and an analysis of these results. Finally, our results are compared with those reported elsewhere in the literature.

In this chapter, the following terminology will be used: “Classical” refers to the classical eigenface technique described in section 3.3, whereas “Probabilistic” refers to the probabilistic technique detailed in section 3.4. “Probabilistic (MAP)” indicates the variant using the maximum *a posteriori* identification scheme, and “Probabilistic (ML)” indicates the variant using the maximum likelihood scheme (see section 3.4.3).

5.1 Datasets

The experiments described in this section make use of images taken from the FERET database (Phillips et al., 1998) and the MORPH database (Ricanek and Tesafaye, 2006). The FERET database was chosen to facilitate comparison of results with other studies in the literature, since it has been used in a number of such prior studies. The MORPH database was chosen as a more challenging dataset containing significant ageing effects (in contrast with the shorter timespan and controlled conditions of the FERET images).

5.1.1 FERET

For the FERET database, four subsets were used: a training subset, the gallery subset, and two probe subsets. The training subset was constructed from the “Training CD” image set, by filtering out non-frontal images and images that are also present in any of the other three subsets. This resulted in a set of 272 frontal images. While the remaining three subsets are from the original greyscale version of the FERET database, the training subset was assembled using the larger colour images from the Color FERET database.

The gallery subset is the same gallery set used in the FERET tests. It is a collection of 1196 neutral-expression frontal images, each depicting a different subject. All of the images in the two probe sets are of subjects present in the gallery.

The first probe subset is the FERET “fab” probe set, comprising 1195 frontal alternative-expression images. For each subject, the image(s) in this set were taken seconds after the corresponding image in the gallery. This probe subset (along with the FERET gallery) is used in experiments 1 through 4.

The other probe subset is the FERET “Duplicate I” probe set, comprising 722 frontal images. Here, for each subject, each image was captured between 0 and 1031 days after the corresponding gallery image. This probe subset (along with the FERET gallery) is used in experiment 5(a).

5.1.2 MORPH

The primary purpose of the MORPH database, as noted above, is to capture ageing effects. To this end, it contains images of individuals over an extended period of time. The time differences for same-subject image pairs range from less than a year to 69 years.

The MORPH database has less inherent structure than the FERET database. Rather than being divided into specific training, gallery and probe subsets, it is split into two “albums”, containing images of two disjoint sets of subjects. Album 1 consists of 631 subjects with a total of 1690 images, while the much larger album 2 contains 13 673 subjects with a total of 55 608 images.

Before the database could be used for testing, it was necessary to remove a number of problematic images. These consisted largely of non-face images and duplicate faces. Others were face images where a significant portion of the face is obscured by bandages, and images framed in such a way that only part of the face is visible, as well as images taken from a profile or half-profile (rather than frontal) view, and images containing invasive compression artefacts or colour distortion. The duplicate faces were cases where one or more image(s) of a subject appear to have been replaced by exact copies of some image of a different subject. The non-face images were primarily images consisting of the text “Photo Unavailable”, with the remainder being images of tattoos (typically on the subject’s arm), and blank images. A total of 406 such images were removed (1 from album 1, and 405 from album 2), of which 252 were non-face images, and 83 were duplicate faces.

For the purposes of these experiments, the remainder of the dataset was divided into a training subset, and nine pairs of gallery/probe subsets.

The training subset consists of the 335 images of the first¹ 100 individuals in album 2.

The rest of album 2 was used to construct eight pairs of gallery/probe subsets for testing. Each such pair is made up of the next 1000 subjects, with the first image of each subject going to the gallery subset, and the second image going to the probe subset. Subjects with only one image in the database were discarded. Each probe and gallery subset thus contains exactly 1000 images. These album 2 subsets are used in experiment 5(b).

¹References made to ordering of images assume lexicographic ordering based on file name, and not numerical ordering based on subject ID. However, since each file name begins with the (zero-padded) subject ID, images are still grouped by subject.

The final pair of testing subsets was constructed from album 1. The gallery again consists of the first image of each subject, while the second image of each subject serves as a probe. The resulting gallery contains 631 images, while the probe subset is of size 624. The album 1 testing data is used in experiment 5(c).

5.1.3 Training Data Organisation

In the preceding discussion, data for training is treated as a monolithic, unstructured collection of face images. However, while this is sufficient for use with the classical system, the probabilistic eigenfaces technique requires its training input to consist of two distinct sets, one made up of same-subject image pairs, and the other of different-subject image pairs.

To generate appropriately-organised training data from a raw training subset, the following procedure is used.

Intrapersonal pairs: For each subject S with multiple images, select two different images of S at random.

Extrapolational pairs: For each subject S_1 , randomly select a different subject S_2 , and then pair a randomly-selected image of S_1 with a randomly selected image of S_2 .

For the FERET training subset (used in experiments 1 through 5(a)), this yields 135 intrapersonal pairs and 135 extrapolational pairs. For the MORPH training subset (used in experiments 5(b) and 5(c)), it results in 100 intrapersonal pairs and 100 extrapolational pairs.

5.2 Experiment 1 – Face Localisation and Masking

Preprocessing of input images can have a substantial impact on the quality of the input seen by the recognition system. In this experiment, we consider the effect of two preprocessing steps on system performance: face localisation, and masking. The classical system used the Euclidean distance measure, and all three systems used a fractional variance threshold of $v = 0.99$. The results are shown in Table 5.1.

From the results, it can be seen that the face localisation step improves accuracy dramatically in all cases. This is as expected, since it removes a substantial amount of extraneous (non-facial) detail present in the raw images, allowing the algorithms to operate specifically on facial features.

For the masking filter, the situation is less clear-cut. In the cases where localisation is used, masking appears to have an insignificant effect on accuracy – increasing it very slightly in two cases, but decreasing it slightly in the other. This can be explained by the observation that the face localisation procedure removes the vast majority of background details, typically leaving only a small amount of these details in the corner of the cropped image. Additionally, most images in the dataset have a monochrome background, so the

Table 5.1: Impact of face localisation and masking (experiment 1)

System	Localisation?	Masking?	Accuracy (%)
Classical	no	no	62.18
	no	yes	56.65
	yes	no	77.24
	yes	yes	77.49
Probabilistic (MAP)	no	no	59.92
	no	yes	54.23
	yes	no	84.27
	yes	yes	84.02
Probabilistic (ML)	no	no	59.50
	no	yes	53.64
	yes	no	83.60
	yes	yes	83.77

mask will in some cases cause the non-facial area of the image to become *less* uniform. Of course, this analysis does not apply to real-world applications in which the system has no control over the image backdrop; in such cases masking may still be advantageous.

In the degenerate case where masking is used in the absence of face localisation, accuracy is observed to decrease markedly. This is likely due to the fact that, depending on the positioning of the face within the larger image, the mask will occlude different portions of the subject, including parts of the face in some cases.

Based on these results, all subsequent experiments use face localisation, but not masking.

5.3 Experiment 2 – Different Distance Functions

The classical eigenface system uses a distance function to evaluate the similarity of a pair of feature vectors. This experiment assesses the effectiveness of various different distance functions. As in experiment 1, a fractional variance threshold of $v = 0.99$ was used. The results are shown in Table 5.2.

Here we see that the Manhattan distance gives the best accuracy, followed by the Euclidean distance, with the Mahalanobis distance giving the poorest performance.

Table 5.2: Performance of different distance functions (experiment 2)

Distance Function	Accuracy (%)
Manhattan	78.49
Euclidean	77.24
Mahalanobis	72.22
Yambor pseudo-Mahalanobis	75.48
Perlibakas modified SSE	73.97

The weaker performance of the Mahalanobis distance (and its variant) may be explained by the following observation: In the Manhattan and Euclidean metrics, the coefficient of each principal component in the feature vector has an equally-weighted contribution to the distance, while in the Mahalanobis distance, the contribution is weighted in inverse proportion to the corresponding eigenvalue. This means that principal components with larger eigenvalues are penalised, while those with smaller eigenvalues are emphasised. However, the principal components associated with larger eigenvalues account for a greater fraction of the variance between face images, and can therefore be expected to be of greater relevance in distinguishing between images of different individuals. As a result, the Mahalanobis distance favours irrelevant inputs over useful ones.

Based on these results, the classical system uses the Manhattan distance in all subsequent experiments.

5.4 Experiment 3 – Subspace Dimensionality

This experiment investigates the effect of varying the dimensionalities of the subspaces used in representing facial features. In our system, subspace dimensionality is determined by a parameter v specifying a fraction of the variance seen in the training dataset, as detailed in sections 3.3.1 and 3.4.1.

The results are shown in Table 5.3 and Figure 5.1. Additionally, Figure 5.2 gives the Cumulative Match Characteristic (CMC) curves for the three systems for $v = 0.99$. Eigenvector counts in Table 5.3 are given as follows: M for the classical system, $M_I + M_E$ for the MAP probabilistic system, and M_I for the ML probabilistic system.

One would expect the use of a larger number of eigenvectors to improve accuracy, since the system should then be better able to represent the differences between images. In the case of the classical system, this expectation is correct, with performance increasing alongside the number of eigenvectors up to $v = 0.99$. Accuracy levels off for $v = 1.00$, suggesting that the last 46 eigenvectors do not encode information that is useful for recognition.

However, the results for the two probabilistic systems do not follow the same pattern. Here, accuracy increases by a small margin from $v = 0.50$ up to $v = 0.75$, and fluctuates within a very narrow range from 0.75 up to 1.00. This indicates that the probabilistic techniques require a substantially smaller number of eigenvectors to be effective (compared to the classical technique), with a larger fraction of the eigenvectors encoding irrelevant information in these cases.

In comparing the overall performance of the three systems, we see that the two probabilistic systems give similar performance, with the MAP variant generally outperforming the ML variant by a small margin (on the order of 1% or less). The classical system is uniformly less accurate, generally by a margin of 5–10%. These patterns are consistent with expectations – the probabilistic classifiers are trained in such a way as to learn the characteristics of the intrapersonal class (and extrapersonal class in the MAP case), while

Table 5.3: Effect of eigenvector selection threshold on accuracy (experiment 3)

System	Threshold	Eigenvectors	Accuracy (%)
Classical	0.50	6	45.69
	0.75	28	71.05
	0.90	88	76.90
	0.95	141	77.66
	0.99	226	78.49
	1.00	272	78.49
Probabilistic (MAP)	0.50	23 + 5	82.59
	0.75	56 + 20	84.02
	0.90	91 + 49	84.35
	0.95	108 + 70	84.18
	0.99	127 + 106	84.27
	1.00	134 + 134	83.93
Probabilistic (ML)	0.50	23	82.59
	0.75	56	83.68
	0.90	91	83.18
	0.95	108	83.35
	0.99	127	83.60
	1.00	134	83.51

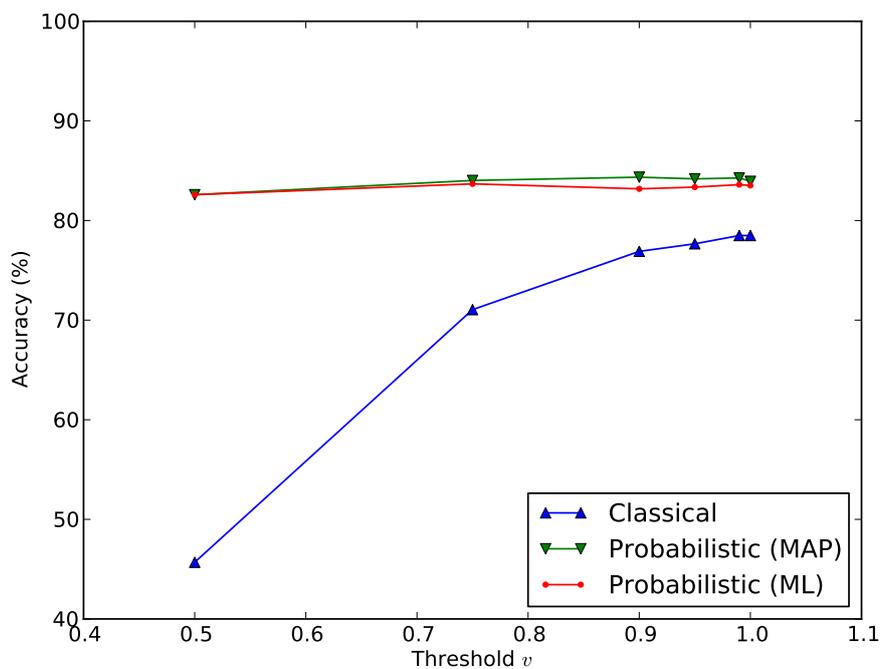


Figure 5.1: Effect of eigenvector selection threshold on accuracy (experiment 3)

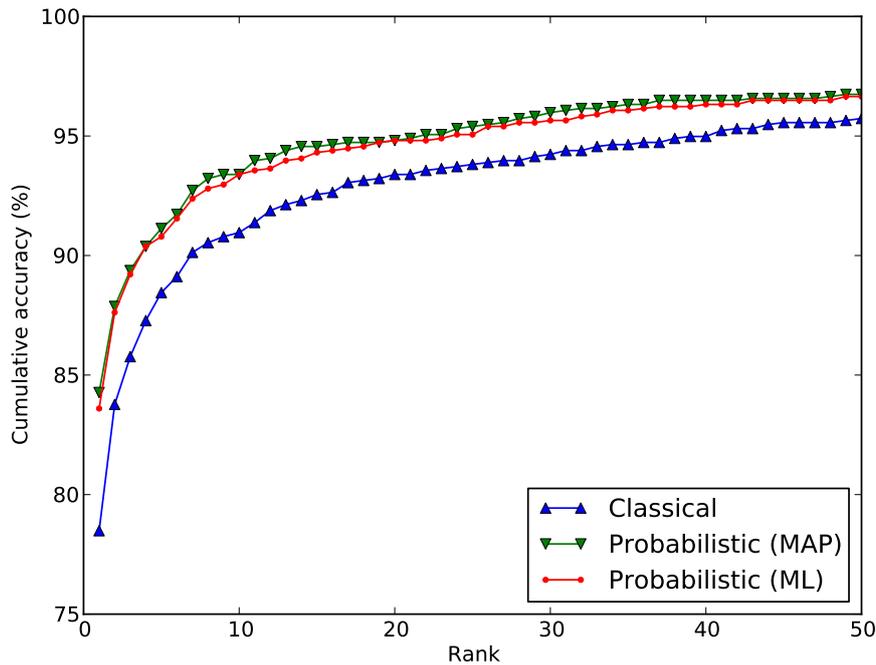


Figure 5.2: CMC curves for experiment 3 where $v = 0.99$

the classical technique operates as a more naïve image similarity measure.

5.5 Experiment 4 – Dataset Size

Experiments in prior literature have used a numerous different datasets, of a wide range of sizes. This experiment examines the relationship between dataset size and recognition accuracy. To this end, the probe and gallery datasets were split up into subsets – first, 2 subsets of 500 individuals each, then 4 subsets of 250 individuals, 7 subsets of 150 individuals, and finally 15 subsets of 75 individuals. For each size, recognition was performed independently on each subset of that size. As in experiments 1 and 2, a fractional variance threshold of $v = 0.99$ was used. Table 5.4 shows the average accuracy for each size of subset, along with the relevant full-dataset results from experiment 3. Figure 5.3 presents this data graphically.

For all three algorithms, the results show a clear trend of accuracy decreasing as

Table 5.4: Effect of dataset size on accuracy (experiment 4)

System	Dataset Size				
	1196	500	250	150	75
Classical	78.49	83.18	84.28	85.12	86.12
Probabilistic (MAP)	84.27	87.98	88.88	89.41	89.50
Probabilistic (ML)	83.60	87.59	88.69	89.32	89.23

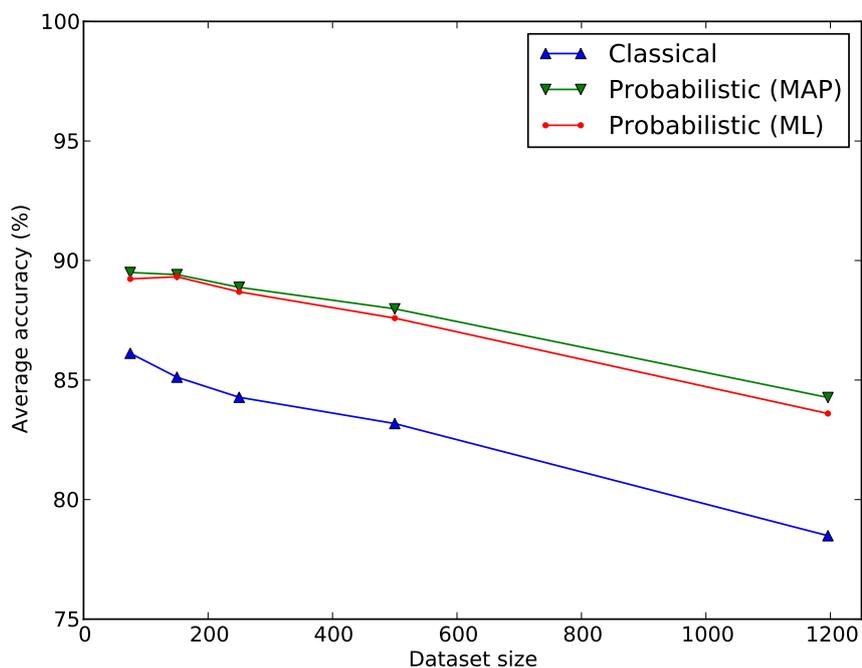


Figure 5.3: Effect of dataset size on accuracy (experiment 4)

dataset size increases. This supports the unsurprising conclusion that it is easier to identify an individual from a small set of candidates than from a larger set.

5.6 Experiment 5 – Alternate Datasets

Experiments 1 through 4 have made use of the “fafb” probe subset of the FERET database, which (along with the FERET gallery subset) was taken under controlled conditions, with the majority of variation between images of a subject being due to changes in facial expression. However, in real-world applications, facial recognition systems are exposed to many other types of intrapersonal variations, so these tests alone are not sufficient to judge the performance of the techniques under consideration.

To gain a better understanding of how these techniques perform under a wider range of conditions, this experiment evaluates them using more challenging datasets. Experiment 5(a) uses an alternate probe subset from the FERET database, while experiments 5(b) and 5(c) make use of the MORPH database.

The parameters used here are the same as those for experiment 4. The eigenvector selection threshold is $v = 0.99$, the classical system uses the Manhattan distance, and the preprocessing pipeline performs face localisation but not circular masking.

Table 5.5: Results for FERET “Duplicate I” subset (experiment 5(a))

System	Accuracy (%)
Classical	26.45
Probabilistic (MAP)	29.50
Probabilistic (ML)	29.50

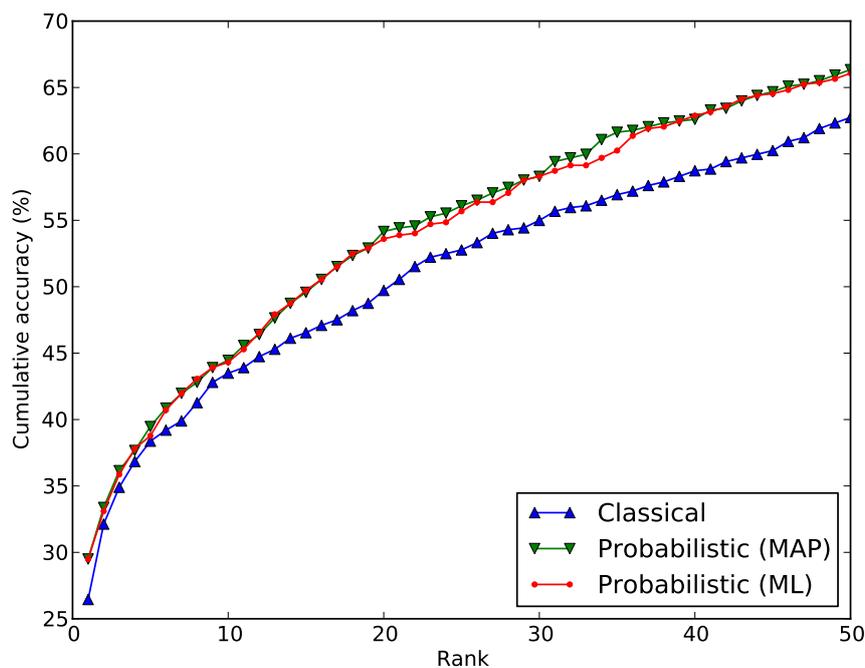


Figure 5.4: CMC curves for FERET “Duplicate I” subset (experiment 5(a))

5.6.1 Experiment 5(a) – FERET “Duplicate I” subset

This experiment uses the “Duplicate I” probe subset from the FERET database, along with the same gallery and training subsets used in the preceding experiments. In addition to the facial expression variations present in the fafb probes, the Duplicate I probes exhibit significant differences in lighting conditions and hairstyle when compared to the corresponding gallery image. Also present are differences in presence of facial hair, glasses and jewellery.

The results of the experiment are shown in Table 5.5. In addition to the rank-1 recognition rates given in the table, Figure 5.4 shows the Cumulative Match Characteristic (CMC) curves for the three systems.

Here we observe a dramatic decrease in performance compared to the preceding experiments, for all three systems. As in previous experiments, the probabilistic systems outperform the classical system, with the two probabilistic variants having similar performance.

The primary cause of this performance drop is most likely the large differences in il-

illumination present in the Duplicate I probe set. Neither the classical nor the probabilistic eigenface techniques have any explicitly-designed provision to compensate for such differences, and the preprocessing step used for this purpose is rather primitive. Additionally, it is worth noting that the training dataset lacks such illumination variations. It is reasonable to anticipate that the use of a more diverse training dataset could yield somewhat improved results (though presumably still significantly poorer than those in exp. 1–4).

5.6.2 Experiment 5(b) – MORPH Album 2

While the FERET Duplicate I subset contains a wider variety of conditions than the fafb subset, it contains only a very limited degree of age variation. For any given subject, the oldest and newest images were captured less than three years apart; as a result, the dataset is of little use in studying the changes in facial appearance that are caused by the ageing process.

In contrast, the MORPH database provides a much wider range of ages, with some subjects having been photographed over a range of as much as 69 years. This experiment makes use of the MORPH training subset, and the eight probe and gallery subsets derived from MORPH album 2. In the images making up these subsets, the effects of ageing are the primary source of variation in each subject’s appearance. The images are generally well-illuminated, and show little variation in pose. In contrast with the two FERET subsets, the MORPH data exhibits a smaller variety of facial expressions; most images show the subject with a neutral expression.

For this experiment, eight separate tests were run on the eight probe and gallery subsets. Average rank-1 recognition rates are given in Table 5.6, and Figure 5.5 shows the associated CMC curves.

As expected, we see another large drop in accuracy. This is consistent with reports in the literature that face recognition systems typically perform more poorly in the presence of ageing effects.

It is interesting to note that while the ML probabilistic system outperforms the classical system (as in the FERET-based experiments), the MAP probabilistic system is the weakest performer in this test.

This could be explained by the possibility that a single Gaussian distribution provides an inadequate model of the extrapersonal class Ω_E for the degree of variation seen in this dataset. Indeed, Moghaddam and Pentland (1997) and Ramanathan and Chellappa (2006) both propose more sophisticated probabilistic recognisers that model the extrap-

Table 5.6: Results for MORPH album 2 (experiment 5(b))

System	Average Accuracy (%)
Classical	17.82
Probabilistic (MAP)	16.62
Probabilistic (ML)	18.04

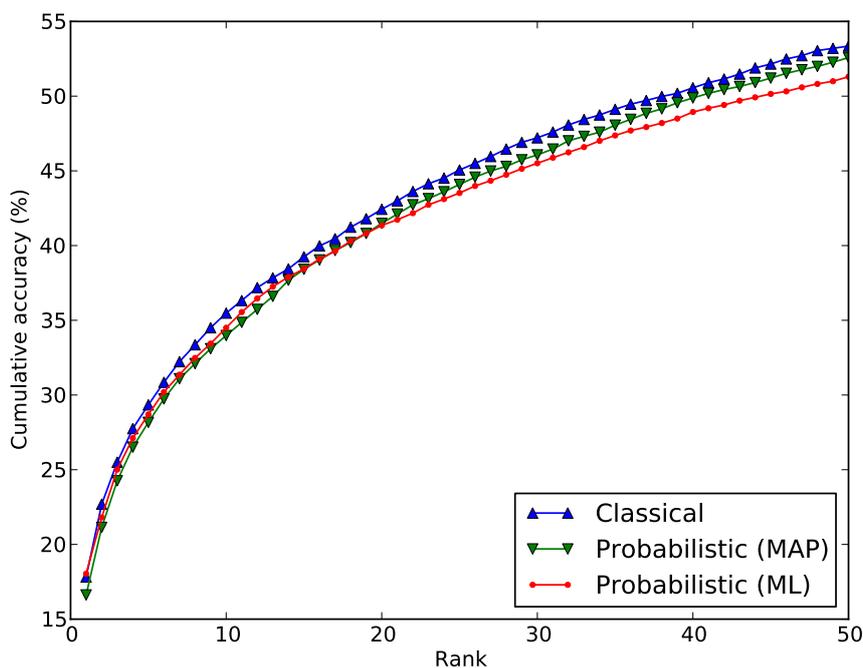


Figure 5.5: CMC curves for MORPH album 2 (experiment 5(b))

ersonal class (or its equivalent) as a mixture of Gaussians. Such a technique may yield better performance in this case.

5.6.3 Experiment 5(c) – MORPH Album 1

This experiment evaluates the three recognition systems on album 1 of the MORPH database. As with album 2, the images in album 1 show a wide range of intrapersonal age variations. While album 1 is smaller than the subsets of album 2 used in experiment 5(b) (containing 631 individuals as opposed to 1000), it still poses a challenge for recognition due to the generally lower quality of the images. The album 1 images appear to largely consist of analogue photographs that have been digitised long after being captured; in many cases the images show signs of physical degradation such as fading and crease marks. In other cases, the subjects are over- or under-illuminated.

The results of this experiment are shown in Table 5.7 (rank-1 recognition rates) and Figure 5.6 (CMC curves).

Here we see a further drop in accuracy (compared to experiment 5(b)). While the

Table 5.7: Results for MORPH album 1 (experiment 5(c))

System	Accuracy (%)
Classical	8.17
Probabilistic (MAP)	8.01
Probabilistic (ML)	8.81

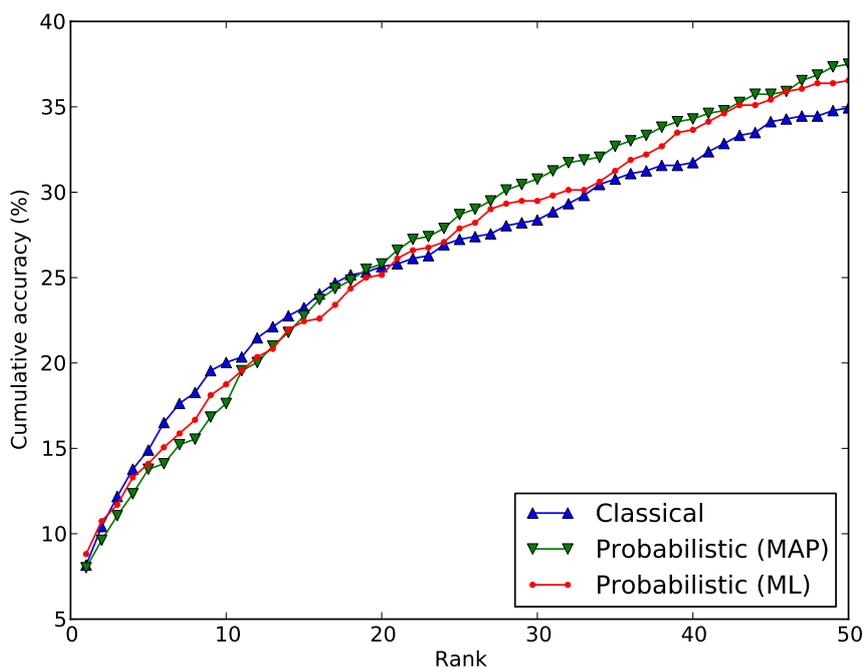


Figure 5.6: CMC curves for MORPH album 1 (experiment 5(c))

poor performance in this case can again largely be attributed to ageing effects, the fact that accuracy is lower than in 5(b) indicates that the image quality concerns discussed above also pose a significant challenge. This could possibly be counteracted by making use of more sophisticated preprocessing techniques. Additionally, it is worth noting that the training subset used in this experiment is derived from MORPH album 2, which has much less in the way of lighting variation. It should be possible to improve performance by training the recognition systems on a dataset that is more diverse in this regard.

In this experiment, as in 5(b), the ML probabilistic system achieves the highest accuracy, with the MAP system again having the poorest performance, and the classical system falling in the middle.

5.7 Comparison With Other Studies

In this section, we compare the results obtained in the experiments described above with those reported elsewhere in the literature. We begin by focusing on techniques that make use of eigenface-based representations, and then move on to other techniques.

5.7.1 Eigenfaces

In the study that first proposed the classical eigenfaces technique, Turk and Pentland (1991) evaluated the technique on a dataset containing images of 16 individuals under a wide range of controlled conditions. In the case of lighting variations between gallery

and probe images, the system achieved a recognition rate of 96%. This would seem to contradict the results of experiment 5(a), but it should be noted that the size of the dataset used here was almost two orders of magnitude larger (containing 1196 individuals rather than 16). Based on the results of experiment 4, it is reasonable to conclude that the difference between these results is due to the different sizes of the datasets used.

In their study proposing the MAP probabilistic eigenfaces technique, Moghaddam et al. (1996) evaluated the technique (along with two variants not tested here) on a 76-individual subset of the FERET database. The MAP technique achieved a recognition rate of 89.5% on this dataset, in comparison with the 84% accuracy attained by the baseline system using the classical eigenfaces technique. These results are consistent with those seen in experiment 4; the probabilistic technique achieved the same 89.5% accuracy, while our classical system gave 86.12% accuracy. The difference in performance for the classical technique can be attributed to the different distances used – Moghaddam et al.’s implementation used the Euclidean distance, which has been shown to give poorer performance than the Manhattan distance.

The FERET test (Phillips et al., 2000) evaluated various face recognition systems on the FERET database using a standardised testing protocol. Included among the tested systems were three eigenface-based systems: a classical eigenface system using the Manhattan distance (“baseline EF”), a classical system using the Euclidean distance (“MIT ’95”), and a MAP probabilistic system (“MIT ’96”). The majority of the tests were conducted in the “partially automatic” mode, wherein the recognition system is provided with face location data for each input image, rather than requiring the system to perform automatic face localisation (as is the case in the “fully automatic” mode).

In the fully automatic testing, the MIT ’96 system achieved an accuracy of 87% on the “fafb” probe set, and an accuracy of 50% on the Duplicate I probe set. For the fafb case, the accuracy is close to our system’s performance (84% in experiment 3). Here, the difference can be attributed to differences in the size and composition in the training dataset. For the Duplicate I case, however, our system gives substantially worse performance (30% in experiment 5(a)). This may be due to differences between the face localisation techniques used. The MIT system uses an eigenimage-based object detector (Moghaddam and Pentland, 1997), which may have different performance characteristics from the Haar-based cascaded classifier used in our system. In particular, the illumination variations present in the Duplicate I images may be responsible for this drop in accuracy, since such variations are not present in the fafb images.

In the partially automatic tests, the MIT ’96 system achieved accuracies of 95% and 57% on the fafb and Duplicate I subsets, respectively. The MIT ’95 system achieved accuracies of 83% and 34%, while the baseline EF system’s results were 79% and 41%. The baseline EF system’s fafb result matches the accuracy of our Manhattan-based classical recogniser in experiment 2, but the corresponding Duplicate I result is significantly higher than the 26% seen in experiment 5(a). Based on the difference between the fully- and partially-automatic results for the MIT ’96 system, we can conclude that this discrepancy

is due to the partially automatic nature of the FERET test, with our face detection subsystem not giving perfect results. A similar observation applies to the MIT '95 system, which yielded better results in both cases.

Distance Functions

The relative performance of different distance functions for the classical eigenfaces technique has been investigated in several studies, with various outcomes.

Yambor et al. (2002) have used the FERET database to test the Manhattan and Euclidean distances as well as an alternative formulation of the Mahalanobis distance (described in section 3.3.3). With the Duplicate I subset, this pseudo-Mahalanobis function provides the best performance by a significant margin, while the Manhattan distance performs best on the “fafb” subset. In both cases, the Euclidean distance performs relatively poorly.

For the fafb subset, the results for the Manhattan distance (77%) and pseudo-Mahalanobis distance (74%) are reasonably close to the corresponding numbers in our experiment 2, while the Euclidean distance gives weaker performance at 72%. On the Duplicate I subset, the Manhattan distance achieves 35% accuracy, compared to the 26.45% accuracy obtained by our system in experiment 5(a). This discrepancy can be attributed to two factors.

Firstly, Yambor et al.’s system was trained on a set of 500 images taken from the gallery set, rather than a completely distinct set of images (as per the FERET protocol). Having overlap between the training and gallery sets may result in a subspace more suited to the recognition of those specific images in the training set. Additionally, having a larger set of training images is likely to improve system performance, given that the classical system in experiment 3 was consistently observed to perform better with the use of larger subspaces.

Secondly, as noted earlier, the automatic face localisation part of our system may perform poorly on the Duplicate I images. Assuming that Yambor et al.’s study used a partially-automatic system (like the bulk of the FERET tests), this would explain the performance difference.

In addition to this comparison of distances, the effect of subspace size on accuracy (for the Manhattan distance) was investigated. For values of the threshold v from 0.43 to 0.92, larger subspaces were found to correspond to increase performance substantially, while performance decreased slightly for values of v from 0.92 to 1.00. This is generally consistent with the results of experiment 3, except in that our experiment found accuracy to level off more slowly, increasing up to $v = 0.99$. This is again probably due to the difference in training dataset size – with a smaller dataset, there will be less total variance, so a basis with an equal fraction of variance will contain a smaller number of eigenvectors which may limit the system’s ability to distinguish between similar-looking individuals.

Draper et al. (2003) have evaluated various distances for the classical eigenface sys-

tem and compared their performance to that of an ICA-based system. Results on the FERET fafb dataset were generally similar to those of Yambor et al. (2002) – the Manhattan distance was found to perform best, with the Mahalanobis distance outperforming the Euclidean distance. Compared to our experiment 2, the results for the Manhattan and Mahalanobis distances were 2–3% higher, while the Euclidean distance again gave 5% lower accuracy. For the FERET Duplicate I dataset, the Manhattan distance achieved a recognition rate of 40%, again significantly higher than that seen in our experiment 5(a).

Since the test conditions of Draper et al.’s study were very similar to those used by Yambor et al., the similarity of the two sets of results is expected. Likewise, in considering the differences between these results and our own, the same explanations are applicable. A larger training dataset (which is also a subset of the gallery) can be expected to yield superior performance, while the partially automatic nature of the system may be responsible for the larger difference in accuracy on the Duplicate I dataset.

As with Yambor et al. (2002), experiments were also performed to examine the relationship between subspace size (number of basis vectors) and recognition accuracy. For the Manhattan and Euclidean distances, accuracy was found to increase for larger subspaces, uniformly across the range of sizes tested. This is consistent with the results of experiment 3. However, it is interesting to note that for the Mahalanobis distance, performance stayed relatively constant over this range on the Duplicate I dataset, and saw a substantial decrease on the fafb dataset. This may be related to the observation in section 5.3 that the Mahalanobis distance emphasises dimensions corresponding to lower-ranked eigenvalues at the expense of those associated with higher-ranked eigenvalues.

Perlibakas (2004) conducted a study on the effectiveness of a large number of distance functions, using a non-standard dataset of 423 subjects. Each distance function was evaluated with various subspace sizes. The distances evaluated include the Manhattan and Euclidean distances, as well as Yambor et al.’s pseudo-Mahalanobis formulation and a novel function referred to as the “modified SSE-based distance”.

For all subspace sizes, the Euclidean and Manhattan distances gave very similar accuracies, with the Manhattan outperforming the Euclidean by a small margin (less than 1%) in each case. The modified SSE-based distance consistently performed more poorly than either of the conventional metrics, by a margin of less than 2%. The pseudo-Mahalanobis function was found to perform significantly better than any of the other three distances, in some cases by as much as 5%.

The general pattern of performance between these distances is generally consistent with our experiment 2, with the exception of the pseudo-Mahalanobis case. However, the individual results differ noticeably even in the case of similarly-sized subspaces (254, compared to 226 in experiment 2), being around 5% higher for the Manhattan distance, 10% higher for the modified SSE-based distance, and 13% higher for the pseudo-Mahalanobis. This appears to be due to the large difference in dataset size between the two experiments. The results of our experiment 4 support this hypothesis, with the Manhattan distance achieving an average accuracy within 1% of Perlibakas’s corresponding result

when using a database size of 500 subjects.

Returning to the discrepancy in the results for the pseudo-Mahalanobis distance, it is possible that it is due to differences in dataset composition. This is uncertain, however, since the dataset in question is composed of an unspecified mixture of images drawn from nine different databases, with different degrees of variation in illumination, pose, etc. As such, the characteristics of the actual images used is not at all evident. However, this hypothesis is supported by the fact that, as noted previously, our results for the pseudo-Mahalanobis distance agree well with those of Yambor et al. (2002).

5.7.2 Other Techniques

As mentioned in section 5.7.1, Draper et al. (2003) have evaluated ICA-based recognition techniques in comparison with the classical eigenface approach. Two ICA-based system architectures were tested, each using two different distance measures. Architecture II with the cosine distance was found to give the best results of the four configurations, outperforming the best eigenface system on all datasets used.

In comparison with the results from our experiments, this system outperformed all three systems on the FERET Duplicate I subset (experiment 5(a)), with an accuracy of 48.48%. On the fafb subset (experiment 3), performance was superior to the classical eigenface system, but slightly weaker than the two probabilistic systems, at 82.26%.

Lades et al. (1993) proposed the DLA approach to object recognition, and evaluated it on a dataset containing 88 individuals. The results obtained were 88% for a dataset containing small (15°) pose variation, and 84% for a dataset containing expression variation. It should be noted, though, that these datasets include non-face images, and much of the error is due to faces being incorrectly discarded as non-faces. If the non-facial cases are discarded, these results improve to 98% and 97%, respectively. These results are much higher than those for similarly-sized datasets in our experiment 4, the best being 89.50% for the MAP probabilistic system. However, it must be noted that the use of different datasets may be responsible for a part of this large difference in accuracies.

Wiskott et al. (1997) proposed EBGGM, a refinement of the DLA technique. On a 250-subject subset of the FERET fafb dataset, the system achieved 98% accuracy, a large improvement over the 85-90% accuracies seen for equal-sized datasets in our experiment 4.

The FERET evaluation included a system based on the EBGGM technique. The USC system gave very strong performance in the partially-automatic tests, achieving 95% accuracy on the fafb subset, and 59% accuracy on the Duplicate I subset. For the fully-automatic tests, the USC system achieved similarly high performance, attaining 94% accuracy in the fafb case, and 58% accuracy in the Duplicate I case. These results are dramatically higher than the results for the best-performing system in our experiments 3 and 5(a) – the MAP probabilistic system’s results being 84% and 30% respectively.

Another approach that was found to perform well in the FERET evaluation was that of Linear Discriminant Analysis (LDA), a family of techniques including the Fisherfaces

technique of Belhumeur et al. (1997). Three partially-automatic LDA-based systems were evaluated. The MSU system's accuracies on the fafb and Duplicate I subsets were 89% and 33%, respectively. The UMD '96 system achieved accuracies of 84% and 31% respectively, while the results for the UMD '97 system were 96% and 47% respectively. For the fafb subset, these results are all substantially better than those for the baseline eigenface system (79%), while for the Duplicate I subset, only UMD '97 outperforms the baseline system. For the MSU and UMD '97 systems, these results are significantly higher than those seen in our own experiments. However, it is worth noting that the partially-automatic nature of these systems will have contributed to their higher performance.

Wright et al. (2009) proposed the SRC technique, and evaluated it in comparison with a number of other subspace-based approaches to recognition. Each system was evaluated on the 38-subject Extended Yale B database, and a 100-subject subset of the AR database. The best configuration of the SRC approach achieved 98% accuracy on the Extended Yale B dataset, and 95% on the AR dataset. In comparison, a classical eigenface recognition system using the Euclidean distance gave accuracies of 88% and 81% respectively. This suggests that the SRC approach is superior to the classical eigenface technique, but meaningful comparison with the probabilistic eigenface technique is difficult in the absence of results on a common dataset.

Park et al. (2010) proposed an ageing modelling technique to compensate for age variations between images. This technique was tested in conjunction with the commercial FaceVACS face recognition software. Tests were conducted using album 1 of the MORPH database, along with the FG-NET database, and the novel "BROWNS" dataset. On MORPH album 1, an accuracy of 77% was achieved when using the ageing simulation technique, compared to 66% for the FaceVACS system without ageing simulation. These results compare favourably with the accuracies of 8–9% obtained by our eigenface systems in experiment 5(c).

CHAPTER 6

CONCLUSION

In this study, we have implemented and evaluated face recognition systems using the classical and probabilistic eigenface techniques. These techniques have been evaluated on a variety of different datasets. The results obtained were found to be broadly consistent with those presented elsewhere in the literature.

In general, the eigenface-based techniques were found to perform reasonably well when dealing with images captured under carefully-controlled conditions. In this context, the probabilistic technique was found to consistently outperform the classical technique, generally by a margin of around 5%. However, both techniques' accuracies were found to decrease markedly when operating on a dataset containing a range of different lighting conditions. Similarly, very poor performance was observed in the case where subjects had aged significantly between image capture sessions. In this case, the ML formulation of the probabilistic technique continued to give higher accuracy than the classical technique, while the MAP formulation performed worse.

The use of face localisation to eliminate background details was found to dramatically improve the performance of both systems. However, the technique used for face localisation is not perfect, and improvements in this area can be expected to further enhance system accuracy.

In evaluating the classical technique, various different distance metrics were considered. Of these, the Manhattan distance was found to give the best performance, outperforming the Euclidean and Mahalanobis distances, as well as the unconventional distance functions of Yambor et al. (2002) and Perlibakas (2004). This finding is in agreement with other results in the literature that make use of the same dataset, but may not generalise to other datasets.

Another parameter that was investigated is the dimensionality of the subspace(s) used for feature representation. For the classical technique, the use of a larger subspace was found to improve performance significantly, up to a size sufficient to express 99% of the variance seen in the training data. However, the probabilistic system appears to be much less sensitive to this parameter – even when using very small subspaces (with a variance fraction of only 50%), the resulting accuracy was within 1.5% of the overall best result obtained in that experiment.

Tests were also conducted to investigate the influence of dataset size on system accuracy. It was found that, for a given recognition system, using a gallery containing a larger number of subjects caused performance to decrease. This information proved useful in comparing our results with those presented elsewhere in the literature, due to the wide variety of datasets that have been used for evaluating face recognition systems.

6.1 Directions for Future Work

One possible direction for future work is to remedy the implementation shortcomings of the systems used in this study. At present, the techniques used for preprocessing are rather rudimentary, so performance can be expected to improve if more sophisticated methods of face localisation and illumination normalisation are incorporated. Additionally, the datasets used for training are currently small, and each is rather homogeneous in terms of image conditions. It is expected that the use of a larger, more diverse dataset will improve system accuracy (particularly on more challenging test datasets).

Another more challenging matter for further research is that of ageing. Clearly, an unaugmented eigenface recognition system is not able to handle the challenge of recognising individuals across age differences of multiple years. To deal with this problem, ageing models such as that of Park et al. (2010) merit investigation. Such a mechanism can be incorporated into an existing system, preprocessing an input image before comparison with each gallery image.

Finally, it is worth investigating other face recognition techniques for purposes of comparison, as there are a number of non-eigenface-based approaches that have shown promising results. In the context of subspace-based recognition techniques, the SRC approach of Wright et al. (2009) is of particular interest. It would be interesting to observe the results obtained with this technique on larger and more challenging datasets.

REFERENCES

- M. S. Bartlett and T. J. Sejnowski. Viewpoint invariant face recognition using independent component analysis and attractor networks. *Advances in Neural Information Processing Systems*, 9:817–823, 1997.
- P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- W. W. Bledsoe. The model method in facial recognition. Technical Report PRI:15, Panoramic Research Inc., Palo Alto, CA, Aug. 1966.
- T. Bourlai, A. Ross, and A. Jain. On matching digital face images against scanned passport photos. In *Proc. First IEEE Intl. Conf. on Biometrics, Identity and Security (BI&S), Tampa, USA*, Sept. 2009.
- B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge. Recognizing faces with pca and ica. *Computer Vision and Understanding*, 91(1–2):115–137, 2003.
- R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, second edition, 2001.
- P. J. Flynn, K. W. Bowyer, and P. J. Phillips. Assessment of time dependency in face recognition: An initial study. In *Proc. 4th Intl. Conf. on Audio- and Video-based Biometric Person Identification*, Lecture Notes in Computer Science, pages 44–51, 2003.
- R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson Education, third edition, 2008.
- J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 11–18, 2003.
- A. K. Jain. Biometric recognition. *Nature*, 449:38–40, 6 September 2007.
- T. Kanade. *Computer Recognition of Human Faces*, volume 47 of *Interdisciplinary Systems Research*. Birkhäuser, 1977.

-
- M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, Mar. 1993.
- B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, July 1997.
- B. Moghaddam, C. Nastar, and A. Pentland. A bayesian similarity measure for direct image matching. In *Proc. Intl. Conf. on Pattern Recognition, Vienna, Austria*, volume 2, pages 350–358, Aug. 1996.
- B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33:1771–1782, 2000.
- U. Park, Y. Tong, and A. Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):947–954, May 2010.
- V. Perlibakas. Distance measures for pca-based face recognition. *Pattern Recognition Letters*, 25:711–724, 2004.
- P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
- P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, Oct. 2000.
- N. Ramanathan and R. Chellappa. Face recognition across age progression. *IEEE Transactions on Image Processing*, 15(11):3349–3361, Nov. 2006.
- K. Ricanek, Jr. and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Proc. IEEE 7th Intl. Conf. on Automatic Face and Gesture Recognition, Southampton, UK*, pages 341–345, Apr. 2006.
- A. Ross and A. K. Jain. Human recognition using biometrics: An overview. *Annals of Telecommunications*, 62(1):11–35, January/February 2007.
- R. Singh, M. Vatsa, A. Noore, and S. K. Singh. Age transformation for improving face recognition performance. *Lecture Notes in Computer Science*, 4815:576–583, 2007.
- L. Sirovich and M. Kirby. A low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, Mar. 1987.
- M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition, Maui, Hawaii*, pages 586–591, 1991.

- P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
- J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, Feb. 2009.
- W. S. Yambor, B. A. Draper, and J. R. Beveridge. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. In H. Christensen and P. Phillips, editors, *Empirical Evaluation Methods in Computer Vision*, chapter 3, pages 39–60. World Scientific Publishing, 2002.
- W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, Dec. 2003.