Angular Feature Extraction and Ensemble Classification Methods for 2D, 2.5D and 3D Face Recognition

R.S.Smith

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

July 2008

Summary

It has been recognised that, within the context of face recognition, angular separation between centred feature vectors is a useful measure of dissimilarity. In this thesis we explore this observation in more detail and compare and contrast angular separation with the Euclidean, Manhattan and Mahalonobis distance metrics. This is applied to 2D, 2.5D and 3D face images and the investigation is done in conjunction with various feature extraction techniques such as local binary patterns (LBP) and linear discriminant analysis (LDA). We also employ error-correcting output code (ECOC) ensembles of support vector machines (SVMs) to project feature vectors non-linearly into a new and more discriminative feature space.

It is shown that, for both face verification and face recognition tasks, angular separation is a more discerning dissimilarity measure than the others. It is also shown that the effect of applying the feature extraction algorithms described above is to considerably sharpen and enhance the ability of all metrics, but in particular angular separation, to distinguish inter-personal from extra-personal face image differences.

A novel technique, known as angularisation, is introduced by which a data set that is well separated in the angular sense can be mapped into a new feature space in which other metrics are equally discriminative. This operation can be performed separately or it can be incorporated into an SVM kernel. The benefit of angularisation is that it allows strong classification methods to take advantage of angular separation without explicitly incorporating it into their construction. It is shown that the accuracy of ECOC ensembles can be improved in this way.

A further aspect of the research is to compare the effectiveness of the ECOC approach to constructing ensembles of SVM base classifiers with that of binary hierarchical classifiers (BHC). Experiments are performed which lead to the conclusion that, for face recognition problems, ECOC yields greater classification accuracy than the BHC method. This is attributed primarily to the fact that the size of the training set decreases along a path from the root node to a leaf node of the BHC tree and this leads to great difficulties in constructing accurate base classifiers at the lower nodes.

Key words: 2D, 2.5D and 3D Face Recognition, Distance Metrics, Angularisation, Linear Discriminant Analysis, Local Binary Patterns, Error Correcting Output Codes, Binary Hierarchical Classifiers, Support Vector Machines.

WWW: http://www.eps.surrey.ac.uk/

Acknowledgements

I would like to thank my supervisor, Prof. Josef Kittler, for providing advice and guidance throughout the course of this work. I am also grateful for the assistance given to me by many other members of the Centre for Vision, Speech and Signal Processing. In particular I would like to extend my thanks to Chi Ho Chan, Dr. Miroslav Hamouz, Prof. Adrian Hilton, Prof. John Illingworth, Dr. Joni Kamarainen, Dr. Norman Poh, Maria Ruiz and Dr. Rafael Tena.

Finally, special thanks go to my wife Jacqueline and daughter Olivia (not forgetting her pet hamster Oscar) for their patience and support over recent years.

Contents

0.1 Abbreviations

BCH	Bose-Chaudury-Hocquenghem. An error-correcting code algorithm based on the use of Galois fields.
BHC	Binary hierarchical classifier. An ensemble classifier that is based on a hierarchical decomposition of the set of target classes.
CMC	Cumulative match curve. Plots identification rate against rank.
DAG	Directed acyclic graph.
ECOC	Error-correcting output code classifier. An ensemble classifier that is based on repeated partitioning of the set of target classes into two families.
EER	Equal error rate. The value of FAR or FRR at a threshold for which these two values are equal.
FAR	False acceptance rate. The fraction of impostor claims that are wrongly accepted as valid by an authentication system.
FRR	False rejection rate. The fraction of valid client claims that are wrongly rejected by an authentication system. Equal to $1 - VR$.
HTER	Half-total error rate. The mean of FAR and FRR at the EER.
IR	Isotonic regression. A method of calibrating SVM outputs to probability values.
LBP	Local Binary Patterns. A texture analysis algorithm that is based on comparing pixel intensities with those of their neighbours at a given radial distance.
LDA	Linear discriminant analysis. An algorithm for reducing the dimensionality of a feature space in a way that increases the separation between classes. This is done by maximising the ratio of between-class scatter to within-class scatter of the training set.
NN	Nearest neighbour classification algorithm. An algorithm that assigns a probe vector to the class that has the nearest sample in training set.
PAV	Pair-adjacent violators algorithm for applying isotonic regression.
PCA	Principal components analysis. An algorithm for reducing the dimensionality of a feature space in a way that retains most of the total scatter of the training set.
RBF	Radial basis function. A function which is spherically symmetric about a point.
ROC	Receiver operating characteristics curve. Plots VR (or FRR) against FAR for a range of acceptance threshold settings.

SVM	Support vector machine. A 2-class classification method that is based on finding a maximum margin linear boundary between the training set samples. This may be done in the original feature space or after applying a non-linear projection into a higher dimensional feature space.
VR	Verification rate. The fraction of valid client claims that are correctly accepted by an authentication system. Equal to $1 - FRR$.

0.2 Mathematical Notation

$\mathcal{A}(\mathbf{x})$	the class label assigned to input column vector $\mathbf{x}$ by a classifier
B_{ang}	Bayes error estimate based on angular separation
B_{Euc}	Bayes error estimate based on Euclidean separation
B_{Man}	Bayes error estimate based on Manhattan separation
B_{Mah}	Bayes error estimate based on Mahalanobis separation
B_{rad}	Bayes error estimate based on radial separation
$\mathcal{B}_j$	j'th base classifier in an ensemble
C	number of target classes (i.e. person identities in biometric applications)
$\mathcal{C}(\mathbf{x})$	a discrete random variable that indicates the true class label associated with input column vector $\mathbf{x}$
$D_{ang}(\cdot,\cdot)$	distance measure based on angular separation
$D_{Euc}(\cdot,\cdot)$	distance measure based on Euclidean separation
$D_{Haus}(\cdot,\cdot)$	distance measure based on Hausdorff separation
$D_{Man}(\cdot,\cdot)$	distance measure based on Manhattan separation
$D_{Mah}(\cdot,\cdot)$	distance measure based on Mahalanobis separation
$D_{cos}(\cdot,\cdot)$	distance measure based on cosine separation
$D_{rad}(\cdot,\cdot)$	distance measure based on radial separation
$f(\mathbf{x})$	a 2-class discriminant function which outputs a "soft" value in the range $[-1,+1]$
$\mathcal{E}$	evaluation set of sample column vectors
G	cost parameter of an SVM
$I(\cdot)$	the indicator function which takes the value 0 or 1 depending on whether its argument is false or true
$J_F(\cdot)$	Fisher criterion function
L	number of input features; for greyscale images this is the number of pixels in a rectangular face image and for a 3D point cloud this is the number of points multiplied by 3 (since each point has x, y and z coordinates)
M	number of dimensions in a general feature space
N	number of vectors in a training set

$\mathcal{N}_k$	the k'th node of a BHC tree
P	number of base classifiers to be deployed in an ensemble
$\mathcal{Q}$	test set of column vectors
$\mathbf{S}$	sample covariance matrix
$\mathbf{S}_l$	sample class covariance matrix of the l'th class, with$1 \leq l \leq C$
$\mathbf{S_B}$	between-class covariance matrix
S_{cos}	similarity measure based on cosine separation
$\mathbf{S_W}$	within-class covariance matrix
$\mathcal{T}$	training set of sample column vectors
$\mathbf{T}$	an $L \times N$ matrix whose columns are training vectors
$\mathbf{U}$	a matrix whose columns are eigenvectors of a given matrix
$\mathbf{Z}$	code matrix for an ensemble
$\mathbf{Z}_i$	i'th row of matrix $\mathbf{Z}$
$\mathbf{Z}^j$	j'th column of matrix $\mathbf{Z}$
Z_{ij}	entry at the i'th row and j'th column of matrix $\mathbf{Z}$
$\mathbf{m}$	sample mean vector
$\mathbf{m}_l$	sample class mean of the l'th class, with$1 \leq l \leq C$
t_s	target class label for the s'th member of a training set where $t_s \in \{-1, +1\}$
$\mathbf{x}_s$	s'th column vector in a training set, with $1 \leq s \leq N$
$\mathbf{y}(\mathbf{x})$	column vector of output values from a classifier ensemble for a given input column vector $\mathbf{x}$
$y_j(\mathbf{x})$	output value from base classifier $\mathcal{B}_j$ for a given input vector $\mathbf{x}$
$\mathbf{\Lambda}$	a diagonal matrix of eigenvalues
Ω	family of all target classes (i.e. person identities)
Ω_j	family of classes that is recognised by the j'th base classifier in an ensemble
Ω_j^+	positive sub-family of classes for the j'th base classifier in an ensemble
Ω_j^-	negative sub-family of classes for the j'th base classifier in an ensemble
λ_p	an individual eigenvalue
σ	width parameter for an SVM with Gaussian kernel
ω_l	an individual class (i.e. person identity) label, with$1 \leq l \leq C$

Chapter 1

Introduction

1.1 The Need for Biometrics

In the day to day conduct of human economic, political and social affairs there has always been a need for accurate and reliable person recognition. Furthermore, its importance has risen considerably in recent decades with the advent of mass travel and communications technologies which have led to the globalised economy of today. Whilst these developments have brought many benefits, they have also created problems and challenges, for example in the shape of newly evolving threats from organised criminal and terrorist activities.

As one instance of such a problem, consider the relatively new phenomenon of identity theft. Here a criminal claims the identity of another person and engages in fraudulent activity, for example by raiding the victim's bank account or obtaining a loan in his or her name. This kind of crime is made possible only by the rather weak methods of person recognition that are currently widely used. These methods rely on the assumption that a person's identity can be equated with his or her knowledge of secret information such as passwords, account numbers, place of birth and so on. Whilst a well constructed and frequently changed password can be secure, it is known that many people do not follow best practice and often leave themselves open to their passwords being discovered, guessed or even extracted from them through "social engineering" techniques [22]. As far as other secret information is concerned, this tends to consist of details about a person's history or circumstances and can often be discovered by a determined fraudster.

For these reasons there is an increasing need, in modern society, for reliable and accurate methods of computer-based person recognition. Applications of these systems fall into two main categories according to their objective. In person *authentication* or *verification* a subject claims a certain identity and the task is to make a decision as to whether to accept or reject this claim. This is required whenever it is necessary to control access to an object such as a bank account, a secure building or a mobile phone. In the more difficult problem of person *identification* the aim is to decide an individual's identity out of a known population. This is useful, for example, in order to watch for wanted felons at an airport or to ensure that a person has not been entered more than once in a social security database. If it is assumed that the known subject list is exhaustive then this is referred to as *closed-set* identification and if it is allowed that the target individual may not be present in the list then this is referred to as *open-set* identification. Depending on the application, it is often useful to rank the list in decreasing order of likelihood of a match with the given individual. For open-set identification there is also the possibility of reaching the decision that the given subject is not a sufficiently close match with any member of the known population.

A major reason for the increasing importance of person recognition systems is that the rise of networks of computers and mass storage devices has given individuals much more power to access and manipulate sensitive information than has ever been the case previously. One instance of this is the fact that financial transactions, which used to be laboriously recorded and implemented by human beings, can now be carried out over a computer network almost instantaneously and without any manual intervention. Although the growth of computer power has exacerbated concerns over person recognition, this same phenomenon also makes possible a novel solution to the problem, namely making use of a person's unique biometric data as a means of identification.

By the term *biometric* data is meant any measurable and repeatable aspect of a person's physiological or behavioural characteristics that can be captured through a sensing device, processed by computer and compared with previously taken samples of similar data taken from the same person. The captured data is referred to as the *probe* and the stored data is referred to as *gallery* data. Gallery data is held in a database of *enrolled clients.* The main advantage of using biometric data over a knowledge of secret information is that the former cannot easily be faked and hence is a much more trustworthy method of confirming a person's identity. Nevertheless, it should be noted that biometric recognition systems are not a panacea and care must still be taken over issues such as protecting the integrity of the gallery database and ensuring that valid enrolment and probe data is collected.

A wide range of possible sources of biometric characteristics have been explored [9]. These include fingerprint, face, hand-geometry, signature, voice, lip motion, iris, retina, gait, keystroke, ear shape, body odour and DNA. Each method has its own advantages and disadvantages. Fingerprints, for example, are highly accurate but their collection requires the active participation of the subject and thus cannot be gathered covertly, as is necessary in some person recognition applications. Another problem is that fingerprints can be damaged or eroded due to physical or chemical activity. Although highly accurate (except when distinguishing between identical twins) and stable, the practical difficulties of gathering and analysing DNA samples mean that it has thus far been restricted to forensic science applications.

In practice no known biometric is capable of identifying subjects with perfect accuracy and reliability. This is because, for any given set of biometric data, there are likely to exist multiple individuals in the world who could give rise to similar data. For verification problems this means that there will be a tradeoff between the *false acceptance rate* (FAR) or *false positive* rate where claims are erroneously accepted and the *false rejection rate* (FRR) or *false negative* rate where claims are wrongly rejected. Similarly, in identification problems it will not normally be possible to identify all subjects with rank one accuracy. In practical applications the likelihood of error can be reduced by employing a fusion of two or more biometrics since the probability of a false positive occurring simultaneously on two independent biometric identifiers is the product of their individual probabilities. This is the reason, for example, why the proposed UK identity card scheme will require fingerprint or iris information as well as a digital face image to be stored on the card [21].

It should be noted that this kind of coincidental similarity between the biometric details of two different individuals is, in general, a much less serious problem than the deliberate faking of another's identity that is used in crimes such as identity theft. The reason for this is that it would be unfeasably difficult to find at random a person whose biometric data happens to be sufficiently close to that of a given target, even though this may be theoretically possible. Contrast this with the relative ease by which secret information about a person can be discovered. For this reason, the fact that biometric data is less than perfectly reliable in discriminating between different people does not preclude it from playing an important role in high security systems.

1.2 Face Biometrics

This thesis is concerned with the use of facial characteristics as a source of biometric information. Face data can be conveniently captured as a 2-dimensional colour or greyscale intensity image using a conventional camera. As a source of biometric data, facial images have a number of attractive properties. Firstly, if necessary they can be captured rapidly from a distance without the cooperation or even the knowledge of the target individual. This is important in applications such as detecting potential terrorist threats at airports or railway stations. Another advantage is that the ubiquitous closed-circuit television (CCTV) cameras in todays buildings and public spaces mean that this biometric data is already routinely collected in large amounts (albeit often with rather poor quality) and so, depending on the proposed application, may not require the deployment of further expensive infrastructure. There are also established databases of face images resulting, for example, from the long-established practice of storing "mug-shots" of convicted criminals. A further point to note is that the collection of face images, unlike some other biometrics, such as fingerprints or body odour, is already regarded as commonplace and socially acceptable and so its use tends to meet less resistance than other methods.

The main disadvantages of using facial images as a source of biometric data stem from the fact that large variations in the images can be caused by factors that have nothing to do with the subject's identity . These sources of unwanted noise include illumination variability, which can cause shadowing effects and specular reflections, pose variation, expression variation, occlusions, aging and changes in facial decorations such as glasses, jewellery, beards and hairstyle. A further problem is the need to detect and accurately register a face image so that corresponding pixels from two images of the same subject are representative of the same regions of the face. These issues, and possible solutions to them, are explored in more detail in chapters 2 and 3 and here we simply note that they represent considerable technical problems which, if not addressed, would greatly reduce the effectiveness of facial biometrics as a means of person recognition.

One recent development in the field of face recognition has been the increasing availability and falling price of cameras which are capable of capturing 3-dimensional shape information. This represents an additional modality which can be used, either alone, or, more commonly, in combination with 2D images to improve the accuracy of a face recognition system. For example, one way in which 3D information can be used is in correcting for problems of pose, illumination and expression variation in 2D images.

The architecture and data flow of a typical face recognition system, of the kind considered in this thesis, is shown in Fig. 1.1. During the enrolment stage a client image[1] is captured by a 2D colour or monochrome camera or by a 3D scanner. The resulting data is then subjected to geometric registration and normalisation so as to crop, rescale and translate it to a standard position and size; 2D intensity images are also photometrically normalised to mitigate the effect of shadowing and uneven lighting. The output from this normalisation process is a high-dimensional feature vector which is then passed through a feature extraction process. The aim of feature extraction is to reduce the number of dimensions to a more manageable level and also to increase the separation between different clients, thereby improving the ability of the system to discriminate between them. This lower dimensional feature vector thus constitutes a representation of the client in an abstract feature space and it is this that is stored in the gallery database for future reference. To improve the accuracy of the system it is possible to capture more than one gallery image per client, perhaps taken over a period

[1] For the sake of brevity, we use the general term "image" in this document to include 2D intensity images, 2.5D range images and full 3D scans. We only distinguish between these forms of biometric information in cases where the context makes the difference important.

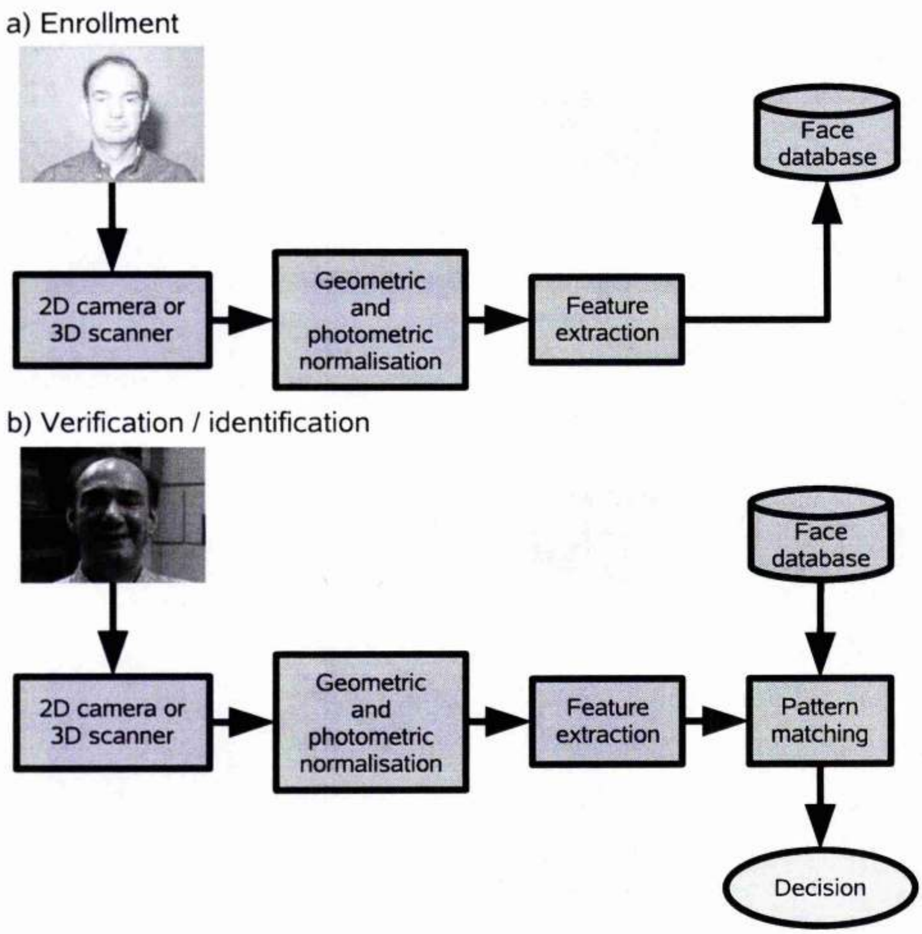

Figure 1.1: The architecture and data flow of a typical face recognition system. During the enrolment stage gallery image(s) are added to the face database. When the system is used to perform verification or identification, a probe image is compared with the gallery image(s) to reach a decision.

of time. Depending on the purpose of the face recognition system, gallery images may or may not be captured under controlled conditions, where the lighting is uniform, the face pose is frontal and the facial expression is neutral.

Having set up the gallery database, the face recognition system may then be used to perform verification or identification. Here a probe image is captured and subjected to the same normalisation and feature extraction processes that were applied to the gallery images. A pattern matching process then compares the probe with the gallery image(s) and makes a decision. For face *verification* this is a yes/no decision as to whether the probe image should be accepted as belonging to the claimed client. Ideally, this would equate to an indication of whether the identity of the person *is* that of the claimed client, however, for reasons discussed above, such perfect accuracy is not, in general, achievable. The required output for face *identification* applications depends on whether closed-set or open-set identification is being performed. In the former case, the output decision is either the most likely client identity (rank 1 identification) or a list of several clients in decreasing order of likelihood. For open set identification a yes/no indication is given as to whether the supplied probe is a sufficiently close match to any of the clients; if it is then information similar to closed-set identification is also output.

As with gallery images, depending on the purpose of the face recognition system, probe images may or may not be captured under controlled conditions. It is also possible for the conditions under which these two types of image are captured to be different. For example, the gallery images may be taken under the controlled conditions of police "mug-shots", but the probe images may be taken under the uncontrolled conditions of a CCTV camera in the street.

Although, for clarity, Fig. 1.1 shows only the capture of either a 2D intensity image or a 3D face scan, we also allow the possibility that both types of image of a given subject are captured and processed in parallel, thus allowing the information from the two modalities to be combined in order to achieve greater accuracy. This kind of simultaneous capture may be done either at the enrolment stage alone, or at both the enrolment and the probe stages, leading to different ways of using the captured data.

1.3 Contributions

In the previous sections of this chapter we have establised the need for good methods of person verification and identification using face biometrics. A summary of the main ways in which the thesis contributes to the realisation of this goal is as follows:

- It has been recognised that, when comparing centred feature vectors derived from two face images, the angular separation distance measure tends to be more discriminative than other commonly used metrics [34, 42, 52, 63, 65]. Here we perform extensive experiments on several face databases to show that angular separation tends to be better than Euclidean, Manhattan and Mahalanobis distance in distinguishing intra-class differences from inter-class differences. These experiments are conducted using a variety of image processing and feature extraction scenarios including homomorphic filtering, histogram equalisation, linear discriminant analysis (LDA) [8], local binary patterns (LBP) [2, 50] and ensembles of support vector machines (SVMs). In addition to 2D greyscale images, the experiments are extended to include 2.5D range image and 3D point cloud representations of the face data. The conclusion drawn from these experiments is that the superiority of angular separation is a general property of facial biometric data and is not, for example, limited to any particular representation or feature extraction technique.

- We show that, in all the above scenarios, the magnitude of centred feature vectors is of little use for discrimination between faces and is more a source of unwanted noise. It is suggested that the main reason for this is that radial movement in feature space, in contrast to angular movement, does not affect the arrangement and geometry of facial features and hence is of little relevance to the problem of distinguishing between the identities of two face images.

- On the wider question of what type of problem is suited to the use of angular separation, we show that its use is beneficial in the problem of detecting whether an image does or does not contain a vehicle image. Experiments with synthetic noise pattern images, however, show that angular separation is not universally superior and that in some cases it may be outperformed by other metrics.

- Many standard classification algorithms are not based, in the first instance, on the use of angular separation as a distance metric and hence must be modified in some way if they are to take advantage of it. We propose two broad approaches to achieving this objective. The first approach is completely general and consists in transforming the feature space into one for which non-angular metrics, such as Euclidean and Manhattan, are approximately in correspondence with angular separation in the original space. We refer to this process as *angularisation* and propose two implementations of the concept. The second approach is specific to ensembles of SVM classifiers and consists of defining kernel functions that incorporate angularisation into their formulation, thereby avoiding the need for a separate stage of feature space transformation. We propose two novel kernels and show that they possess the Mercer property.

- Experiments performed on 2D, 2.5D and 3D face data show that the application of a global angularisation transformation is successful in improving the discriminative capabilities of the Euclidean, Manhattan and Mahalanobis metrics, with the results from the first two becoming comparable to those from angular separation itself. Further experiments with ensembles of SVMs show that this leads to corresponding improvements in the performance of such classifiers when applied to practical face verification and identification problems. Finally, it is shown experimentally that the benefits of applying angularisation to face verification are largely insensitive to the details of how it is accomplished, as all four methods give similar results (one of the SVM kernels is, however, shown to have a small advantage over the other methods on the face identification task).

- Two possible architectures for constructing ensembles of SVMs for face recognition purposes are compared and contrasted. These are error-correcting output codes (ECOC) [18, 19, 33], in which the set of all face identities is repeatedly partitioned into two families of approximately equal size, and binary hierarchical classifiers (BHC) [38, 39, 49, 59] where the set of all face identities is recursively subdivided to construct a binary classification tree. It is shown that, for both methods, the Gaussian SVM kernel tends to outperform the linear kernel. ECOC can gain improved accuracy by deploying an increased number of SVM classifiers whereas BHC entails the use of a fixed number of classifiers, as dictated by the number of target classes. Different methods for calibrating the output values of SVM classifiers are examined and it is shown that ECOC performance, on both verification and identification tasks, is improved by the use of isotonic regression [85] or Platt's sigmoid fitting algorithm [56]. For the BHC architecture, however, only the sigmoid algorithm is beneficial for face verification and no calibration method is found to be beneficial for face identification. It is found that the best decoding procedure to use for ECOC is a nearest neighbour comparison with the gallery set [37] whilst that for BHC is to multiply soft outputs from the base classifiers on the path from root node to leaf node [39]. When partitioning class families, the Bose-Chaudury-Hocquenghem (BCH) algorithm [73] is found to give slightly better results than random assignment for ECOC, whilst for BHC the 2-means clustering algorithm improves on the use of a deterministic annealing algorithm [39].

- In terms of overall face recognition performance it is shown that ECOC significantly outperforms BHC. This is attributed to problems caused by the progressively smaller training set sizes associated with the lower nodes of the BHC tree, together with the fact that face data may not form a deeply nested natural hierarchy as required by the BHC algorithm.

- It is confirmed that, for 2D and 2.5D images, multi-scale LBP [16] is an excellent feature extraction technique and that the performance of this technique can be improved by the application of LDA, angularisation and ECOC.

- It is sometimes stated [11] that 2.5D images or 3D scans are more reliable for face recognition than 2D images because the first two modalities overcome problems of illumination variation. Here we make use of the FRGC corpus [54] as a good source of data for comparing these modalities. Evidence is presented which suggests that, under similar conditions, 3D is more reliable as a means of face recognition than

2D which in turn is more accurate than 2.5D. When the more advanced method of LBP feature extraction is applied to the latter two modalities, however, it is found that their performance is improved to the extent that 2D gives the greatest accuracy and whilst 2.5D gives comparable performance to 3D. A fusion of all three modalities is found to give greater accuracy than any single one.

1.4 Outline of the Thesis

The remainder of the thesis is structured as follows. Chapter 2 provides a review of the literature concerning face representations, with particular reference to angular methods. This is followed in chapter 3 by a more detailed description of the specific techniques that are used in the remainder of the document. Methods for applying angularisation to a set of feature vectors are described in chapter 4 and this is followed in chapter 5 by a description of the test data sets that are used in subsequent experiments. The first set of experiments is presented in chapter 6 and is concerned with relative effectiveness of different distance metrics and how their performance can be improved by angularisation. Chapter 7 then presents a detailed examination of the ECOC and BHC approaches to ensemble design and compares their usefulness in face recognition applications. Further remarks concerning these results are made in chapter 8 and the thesis concludes in chapter 9 with a summary of the conclusions to be drawn from this work.

Chapter 2

Literature Review

General surveys of the methods employed in 2D and 3D face recognition can be found in [11] and [86]. In this thesis the main emphasis is on the representations used to encode face information and the distance metrics that are used to measure the similarities and differences between instances of such information.

2.1 Psychological Studies of Face Representation

The study of different face representations is important to both machine-based and human face recognition. In the field of experimental psychology it has been recognised that a useful heuristic is to regard faces as being represented in memory as points in a multi-dimensional face space. Fig. 2.1 illustrates this concept together with the different types of face distortion which are used in the study of face recognition. It is known that the position of a face in the abstract face space is important for human face recognition, particularly in respect of whether the face lies in a region which has a high density of faces belonging to different people. Faces from the lower density regions tend to possess non-typical characteristics and are said to have a greater ***distinctiveness***; they are more memorable than other, more typical faces [60].

Two possible models for human face recognition have been proposed [78], namely *absolute-based coding* and *norm-based coding*. In both cases position in face space is

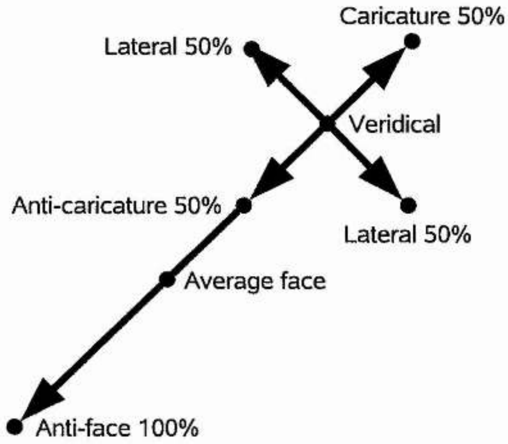

Figure 2.1: An illustration of different types of face distortion. The veridical is the true position of a face in a multi-dimensional face space and the other points represent different types of distortion which can be applied to the veridical.

important for face recognition; norm-based coding differs from absolute based coding, however, in that the relation between the probe face and the average face is also a significant factor. Which of these two models is applicable to the human face recognition system is an open question. Carey et al. [15] have found that speed of recognition of famous faces decreased in the order of caricature, veridical, anti-caricature then lateral distortions. The order of the first three is explained by decreasing levels of distinctiveness and fits either model. The fact that lateral distortions took longer to recognise than the anti-caricature, however, despite having the same distinctiveness as the veridical, is evidence that the axis through the average face and the veridical is a privileged direction in face space and thus supports the norm-based coding hypothesis. Rhodes et al. [60] were unable to reproduce these results, however, and found that speed of recognition decreased in the order caricature, veridical, lateral distortions then anti-caricature. This later result, which reverses the order of lateral distortions and the anti-caricature, indicates that only distinctiveness is important to human face recognition and is evidence for absolute-based coding.

It is shown experimentally in [6], by progressively morphing 3D head models, that there is a perceptual discontinuity when moving from the veridical side of the average face to the opposite, or anti-face, side. This means that, if one considers the linear trajectory between the caricature and the anti-face in Fig. 2.1, there is a greater perceptual distance between two faces that straddle the average face than there is between similarly spaced points elsewhere on the trajectory. In particular, the anti-face is perceptually highly dissimilar to the corresponding veridical. Another conclusion from this work is that, because the male and female average faces occupy different positions in feature space, reflecting through the general average face produces different results from reflecting through a gender-specific average face. For example, if a female face is reflected through the general average face the resulting anti-face shows more masculine characteristics than if the female face is reflected through the female average face. Similar reasoning shows that, for caricatures, the situation is reversed.

Another aspect of the human visual system is the question of what similarity measure is used by the brain to assess the closeness of one image to another. This is important for applications such as image database retrieval where it is desirable that machine matching mimics that of humans. It is shown in [64] that the Manhattan metric better captures notions of human similarity than the Euclidean metric when images are approximated as a collage of fragments taken from a code book (a method of image compression known as *vector quantisation*). In these experiments subjects were asked to rate the fidelity of the approximation when the closest matching fragments were chosen using either the Manhattan or the Euclidean metric. There was found to be a pronounced bias in favour of the former.

Santini and Jain [66] cast doubt upon whether the human similarity measure satisfies the axioms of a metric space at all. For example it is known that the perceived distance of images from themselves is not constant and it is generally acknowledged that the triangle inequality does not always hold. They go on to propose a new similarity measure, based on fuzzy logic, that satisfies a new set of axioms and which more closely replicates experimental findings in humans.

2.2 Dimensionality Reduction and Face Manifolds

One reason why face recognition is a challenging problem is that, considered as a vector, the dimensionality L of a face image is typically quite large, say of the order of 10,000. This makes direct comparisons between such images a computationally demanding process. It is clear, however, that in a realistic face image the pixels cannot be allowed to take on values which are completely independent of each other and it is known that the space of all possible face images, under varying conditions of illumination and pose, constitutes a low dimensional manifold which is embedded within the ambient space of all possible images. Various techniques have been applied to the problem of finding approximations to this manifold. These have the aim of discovering a lower dimensional representation in which important variations in facial images, particularly with regard to identity differences, are emphasised, whilst at the same time variations due to unwanted noise are reduced or eliminated. Another argument for dimensionality reduction is the *peaking phenomenon* which tends to lead to problems of over-training unless the training set size is substantially greater than the number of pattern dimensions.

Principal components analysis (PCA) [80], also known as the *Karhunen-Loeve expansion*, was introduced into the field of face recognition by Turk and Pentland [77] under the name *eigenfaces* and was the first really successful demonstration of machine-based face recognition algorithms [86]. The method is described in some detail in section 3.3 and summarised here. Given a training set $\mathcal{T}$ of face images the covariance matrix, or *total scatter matrix*, of the centred training set is diagonalised to find an orthonormal basis in which the different eigenvectors represent different modes of variation. If the eigenvectors are ordered by decreasing eigenvalue then the largest variance occurs in the directions of first few eigenvectors and so, by discarding the later eigenvectors, an approximate low dimensional linear subspace can be found that preserves most of the variance of $\mathcal{T}$. There is evidence [4, 82] that the first two eigenvectors simply code for differing lighting conditions and that better face recognition results are obtained when they are also discarded.

Although PCA represented an important advance in face recognition, the method suffers from the drawback that, being an unsupervised technique which is based entirely on the global variation between training images, it takes no account of class membership information in the training set and hence is not optimal for classification problems [4]. *Linear discriminant analysis* (LDA) [80] is a dimensionality reduction and feature extraction technique, first applied to face recognition by Belhumeur et al. [8], that aims to rectify these deficiencies. Details of the LDA algorithm are given in section 3.4 and here we give a short summary. Like PCA, LDA aims to find a low-dimensional linear subspace of the full image space in which face recognition is more readily carried out. The method consists of solving a generalised eigenvector problem to find a linear transformation of the original image space that maximises the ratio of the mean between-class scatter to that of the mean within-class scatter. The resulting eigenvectors, often referred to as *Fisherfaces*[1], form a (non-orthogonal) basis with respect to which discrimination between classes is enhanced.

The methods discussed so far are *linear* methods in the sense that they are aimed at finding a linear subspace that represents the manifold of realistic face images within the full image space. In practice, however, this manifold is only approximately linear and several attempts have been made to model the non-linearities with greater accuracy. Another way of looking at these ideas is to note that PCA and LDA are limited to finding second order statistical correlations in the data and that it may be beneficial to use techniques that incorporate higher order correlations. Kernel PCA and kernel LDA [25, 80] are two such approaches. They extend the methods of PCA and LDA by applying a non-linear kernel function to face image vectors in order to project them into a higher dimensional space where linear methods can then be applied (this technique, known as the "kernel trick", is discussed in more detail in section 3.6). Yang et al. [83] have made a comparative study of PCA and LDA, together with their kernelised variants, and have found that, for face identification experiments on the AT&T and Yale test databases, the best performing algorithm was kernel LDA followed by LDA then kernel PCA and finally PCA.

A number of methods have been explored for directly mapping the face manifold into a lower dimensional space based on the spatial relationships between the training images

[1]Named after R.A.Fisher who first proposed the use of such linear discriminant methods in statistical analysis.

rather than on their statistical properties. In practice these techniques tend to work best when a large amount of training data is available, for example when the input comes from a video stream so that multiple images are available for each subject with different pose angles, expressions and illumination conditions. *Multi-dimensional scaling* (MDS) [80] is the name given to a general technique for mapping a set of points in a high-dimensional feature space to a corresponding set of points in a lower dimensional feature space in such a way that the distance between pairs of points is preserved. One common use for this procedure is as a means visualising, in two or three dimensions, the similarities between, and clustering properties of, high dimensional data sets such as collections of face images.

The choice of distance metric used with MDS influences the resulting low-dimensional representation. When the Euclidean metric is used, MDS is, in fact, equivalent to PCA [76]. The *isomap* algorithm [76] uses approximate geodesic distance within the manifold, in conjunction with MDS, to obtain a more accurate low-dimensional representation of the relationship between points on the manifold. To calculate the geodesic distance, neighbouring points (defined either as the K nearest neighbours or as all points within a certain threshold distance) are assumed to lie on the manifold and, therefore, separated by their Euclidean distance; more distant points are handled by finding the shortest linking path that passes through successive neighbours and summing the local Euclidean distances along that path. Clearly, this approach depends for its success on the manifold of the underlying distribution being densely sampled by the training set. It is shown in [76] that isomap leads to a more accurate low-dimensional representation of the manifold than techniques such as PCA or Euclidean MDS. For example isomap correctly predicts the underlying 3D nature of a collection of face images of a single subject that are viewed and lit from different directions.

Local linear embedding (LLE) [62] is an algorithm that produces similar results to isomap but is computationally more efficient. In this approach each training point is first approximately reconstructed as a linear combination of its neighbours by solving a least squares optimisation problem. Once the sparse matrix of reconstruction weights is determined in this way, it is used to find a projection of the training set into a lower dimensional space that preserves global distances as measured in the original manifold.

Methods such as MDS, isomap and LLE are useful for visualising high-dimensional manifolds and for discovering their intrinsic dimensionality. The resulting maps are, however, only defined on the training points and it is not clear how they may be generalised to arbitrary data points, as is required in face recognition applications. The *locality preserving projection* (LPP) [27] algorithm addresses this problem by finding a set of vectors in the ambient image space that best preserve the local structure of the face manifold. These vectors, which are referred to as *Laplacianfaces,* are then used as the basis vectors of a linear subspace in a way similar to PCA and LDA. In this way, any image can be linearly projected into the LPP space. The method differs from PCA and LDA in that it emphasises the importance of local, rather than global structure. The face recognition experiments described in [27] on the PIE, Yale and MSRA databases show that Laplacianfaces outperforms both of these algorithms, particularly when a large number of training samples is available.

2.3 Distance Metrics and Similarity Measures

Having decided upon a feature space to represent facial information, the next issue that needs to be addressed is what similarity measure or distance metric to use within that feature space. Several candidate metrics have been studied in the face-recognition literature and in this section we present a summary of the main approaches (however we defer a discussion of the important class of angular similarity measures until section 2.4).

Given a pair of arbitrary M-dimensional column vectors $\mathbf{x} = [x_1, \dots x_M]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, \dots y_M]^{\mathrm{T}}$, the *Euclidean, L2* or *second-order Minkowski* distance [80] is defined as

$$D_{Euc}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{M} (x_i - y_i)^2}. \tag{2.1}$$

The *Manahattan, city-block, L1* or *first-order Minkowski* distance [80] is defined as

$$D_{Man}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} |x_i - y_i| \,. \tag{2.2}$$

The *Mahalanobis* distance [80] is defined as

$$D_{Mah}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}-\mathbf{y})}. \tag{2.3}$$

where $\mathbf{S}$ is the covariance matrix, or total scatter matrix, of the training set. In effect, Mahalanobis distance is similar to Euclidean distance (identical if $\mathbf{S}$ is the identity matrix) but it attaches greater weight to those directions in feature space where the training set distribution has low variance. In [42] it is established that use of the Mahalanobis distance metric is equivalent to the optimal Bayes decision rule, whereby probe vectors are assigned to the class with the maximum posterior probability, provided that all class conditional distributions are multivariate normal with equal prior probabilities and identical covariance matrices $\mathbf{S}$.

Another metric that has been successfully applied to face recognition is the *Hausdorff metric*. Given two point sets $A = \{a_i\}$,$B = \{b_i\}$ the Hausdorff distance between A and B is defined as

$$D_{Haus}(A,B) = \max\left(h(A,B), h(B,A)\right) \tag{2.4}$$

where

$$h(A,B) = \max_{a\in A}\min_{b\in B}\|a-b\|. \tag{2.5}$$

It thus represents the largest distance between any point of either set and the nearest point in the other set. This can be used as a means of comparing general binary images (i.e. images where each point is either black or white), with the sets A and B being taken to be the set of black points in each image [31]. In [24] this is applied to face recognition by first constructing edge maps of a face image, to convert them from greyscale to binary, and then using a Hausdorff metric to compare them. The closest match between two such edge maps is found by sliding one relative to the other until the Hausdorff distance is minimised. This approach has the advantage that it does not require the precise geometric alignment of images in order to compare them and thus circumvents the need for accurate landmarks.

In [81] a method is proposed whereby, starting with pre-defined "side" information about which pairs of training vectors are to be regarded as similar or dissimilar, a customised metric can be learned that respects this notion of similarity. The algorithm consists of finding a symmetric positive semi-definite matrix $\mathbf{A}$ such that the quadratic distance [80] between any pair of column vectors $\mathbf{x}$ and $\mathbf{y}$, defined as $\sqrt{(\mathbf{x}-\mathbf{y})^{\mathrm{T}} \mathbf{A} (\mathbf{x}-\mathbf{y})}$, tends to yield small values when the vectors are classified as similar and larger values otherwise. Experiments on a variety of data sets show that the performance of the k-means clustering algorithm [80] is improved by the use of this metric when compared with the Euclidean metric. In face recognition applications a natural definition of whether two vectors are similar would be whether they represent the same subject identity or not.

An alternative approach to constructing a similarity measure between face images is that of *Bayesian face recognition* [46, 47]. Here the differences between pairs of face images are divided into two classes, namely intra-personal differences Ω_I when both images are of the same subject and inter-personal differences Ω_E when the image identities are distinct. PCA is then applied separately to each of the two distributions and the resulting training sets are modelled as Gaussian distributions. Given arbitrary probe and gallery images with difference vector Δ, this model permits the calculation of the probability densities $\Pr(\Delta \mid \Omega_I)$ and $\Pr(\Delta \mid \Omega_E)$ that the differences result from intra-personal and inter-personal variations respectively. A similarity measure $\Pr(\Omega_I \mid \Delta)$ is then obtained, using Bayes rule:

$$\Pr(\Omega_I \mid \Delta) = \frac{\Pr(\Delta \mid \Omega_I)\Pr(\Omega_I)}{\Pr(\Delta \mid \Omega_I)\Pr(\Omega_I) + \Pr(\Delta \mid \Omega_E)\Pr(\Omega_E)}. \tag{2.6}$$

This is the maximum a posteriori probability (MAP) that the observed image differences can be explained by intra-personal variations; the priors $\Pr(\Omega_I)$ and $\Pr(\Omega_E)$ can be used to incorporate a priori domain knowledge or, in the absence of this, simply set to equal values. An alternative similarity measure, which is simpler but slightly less accurate, is to use the intra-class probability density function $\Pr(\Delta \mid \Omega_I)$ without regard to the inter-personal distribution.

2.4 Angular Methods

Classification techniques that are based on angular methods have been studied by a number of researchers. The most direct approach is to use *angular separation* between feature vectors as a distance pseudo-metric[2]. This is defined as

$$\begin{aligned} D_{ang}(\mathbf{x},\mathbf{y}) &= \arccos\left(\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right) \\ &= \theta(\mathbf{x},\mathbf{y}) \end{aligned} \tag{2.7}$$

where $\theta(\mathbf{x},\mathbf{y})$ is the angle between the vectors and is taken to be in the range $[0,\pi]$. A closely related pseudo-metric is the *cosine distance*

$$\begin{aligned} D_{cos}(\mathbf{x},\mathbf{y}) &= 1-\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \\ &= 1-\cos\theta(\mathbf{x},\mathbf{y}) \end{aligned} \tag{2.8}$$

where $D_{cos}(\cdot,\cdot)$ lies in the range $[0,2]$. When cosine distance is converted to a similarity measure in the range $[-1,1]$, with a value of 1 indicating maximum similarity, it is referred to as the *normalised correlation* or *cosine* similarity measure and is defined as

$$\begin{aligned} S_{cos}(\mathbf{x},\mathbf{y}) &= \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \\ &= 1-D_{cos}(\mathbf{x},\mathbf{y}). \end{aligned} \tag{2.9}$$

Note that, when applying these angular measures, it is assumed that the data is first centred by subtracting the training set mean vector.

In a study by Jonsson et al. [34] using the XM2VTS database, normalised correlation was compared with both Euclidean distance and a directly trained SVM as a means of generating face similarity scores. This was done using raw images (with or without photometric normalisation), PCA feature extraction and LDA feature extraction. On the face verification task it was found that LDA outperformed PCA and that normalised correlation gave substantially better results than Euclidean distance, but was slightly worse than the SVM. For face identification LDA was again markedly better than PCA but there was little difference in the performance of the three similarity measures. In a similar study by Sadeghi et al. [65], using the BANCA database to perform verification experiments under adverse illumination and pose conditions, it was again found that LDA outperformed PCA and that normalised correlation gave better results than the Euclidean metric. Under these adverse conditions, it was also found that normalised

[2]This is referred to as a pseudo-metric since it may take a value of zero on non-identical arguments.

correlation tended to outperform the SVM method unless the training set was large. For 3D face scans, a method is described in [63] for applying PCA to scans whose coordinates have been standardised by bringing them into dense correspondence with a reference model. In verification and identification experiments performed on subsets of the Face Recognition Grand Challenge (FRGC) corpus it was found that cosine distance performed better than Euclidean distance as a similarity measure.

For PCA-based nearest neighbour face recognition algorithms, the *whitened cosine* distance [4] is commonly used. For example, this is the baseline algorithm chosen for performance comparison on the FRGC experiment set [54]. The whitened cosine measure is similar to cosine distance but with an additional whitening transformation being first applied to ensure that the training set has unit variance in all directions. This is accomplished by pre-multiplying $\mathbf{x}_1$ and $\mathbf{x}_2$ in eqn. 2.8 by $\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{\mathrm{T}}$ where $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}} = \mathbf{S}$ is the eigen-decomposition of the sample covariance matrix $\mathbf{S}$. Moon and Philips [48] show, by face identification experiments on the FERET database, that whitened cosine outperforms several other metrics including Euclidean, Manhattan, Mahalanobis and cosine distance. Liu et al. [42] observe that minimising the whitened cosine distance is equivalent to first normalising feature vectors so that their whitened transforms have unit length and then applying the optimal multi-class Bayes decision rule under the assumptions that the resulting class-conditional distributions are multi-variate normal with equal prior probabilities and equal covariance matrices $\mathbf{S}$. Without the initial length normalisation step, the Bayes decision rule is equivalent to the inferior method of minimising the Mahalanobis distance. These assertions are supported by experiments on the FRGC database. The authors go on to propose two new similarity measures. The first of these, named the *probability reasoning model whitened cosine* (PWC) similarity measure, replaces Λ in the whitening transformation by a diagonal matrix whose i'th entry is the mean of the individual class variances in the i'th direction; in FRGC experiments it yields slightly better results than the standard whitened cosine measure. The second proposal, called the *within-class whitened cosine* (WWC) similarity measure, uses the mean within-class scatter matrix in place of the total scatter matrix $\mathbf{S}$ and achieves substantially better results on FRGC experiments.

An extensive comparison of different distance measures, used in conjunction with PCA feature extraction, is presented in [52]. These results are based on face verification and identification experiments conducted on a data base of 423 individuals taken from various publicly available data sets. They show that the overall best performance was obtained using a weighted angle similarity measure

$$S_{wt}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} w_i x_i y_i / \|\mathbf{x}\| \|\mathbf{y}\| \tag{2.10}$$

where the weighting factors w_i are given by $\sqrt{1/\lambda_i}$ with λ_i being the variance in the i'th direction. The whitened cosine distance also performed reasonably well, however recognition accuracy was shown to degrade as the number of PCA features was increased. The best non-angular measure was a whitened correlation-coefficient distance, which gave similar performance to whitened cosine.

The approach taken in [36] is to perform a separate PCA analysis for each individual in a training database and to use the resulting linear subspaces as approximations to the face manifold for each separate person. To be reliable, this requires the availability of a large amount of training data per subject, such as can be obtained from a video sequence. For any pair of subjects the principal angles between their respective subspaces is used as the basis for comparison, with the AdaBoost algorithm [20] being used to combine them into a similarity measure. Further improvements in recognition accuracy were obtained by using a weighted combination of these global linear manifold differences together with more local non-linear variations. The validity of this approach was confirmed by face identification experiments conducted using a database of video sequences of 100 individuals showing different illumination conditions, pose angles and expressions.

An interesting parallel with the success of angular similarity measures in the image domain is that of methods based on phase information in the Fourier domain. It has been observed [51] that, when considering the Fourier transform

$$F(\omega) = A(\omega) e^{j\theta(\omega)} \tag{2.11}$$

of a general multi-dimensional signal $f(\mathbf{x})$, many of the important features of the signal are embodied in the phase function $\theta(\omega)$, whilst the spectral magnitude function $A(\omega)$ is relatively uninformative. This means, for example, that if $A_1(\omega) e^{j\theta_1(\omega)}$ and $A_2(\omega) e^{j\theta_2(\omega)}$ are the spatial Fourier transforms of two 2-dimensional images $I_1(\mathbf{x})$ and

$I_2(\mathbf{x})$ respectively, then the inverse Fourier transform of $A_2(\omega)\,e^{j\theta_1(\omega)}$ will be much more similar to I_1 than to I_2. Similarly, the appearance of the image that is recovered from $A_1(\omega)\,e^{j\theta_2(\omega)}$ will be much closer to that of I_2 than to I_1.

The *eigenphases* method [67] is an approach to face recognition which exploits this property of Fourier transforms. It consists of applying PCA to the phase spectrums only of transformed images to establish a new face representation in which the spectral magnitude is discounted (note that performing PCA on the full Fourier transform achieves nothing as it produces eigenvectors that are trivially related, through the inverse Fourier transform, to those obtained by applying PCA in the original image domain). Experiments on the PIE database show that the eigenphases method is remarkably tolerant to problems of partial face occlusion and illumination variation, outperforming both Fisherfaces and eigenfaces under these adverse conditions.

Phase also plays an important role in *correlation pattern recognition* (CPR) [79]. This method uses the training images for each subject to design a filter which, when applied to the Fourier transform of a probe image of the same subject, produces a sharp peak in output for some displacement of the image. Impostor images, by contrast, produce no such sharp peak and hence the *peak to sidelobe ratio* (PSR), defined as $\frac{peak\,output-mean\,output}{standard\,deviation}$ can be used as a measure of similarity between a probe and a target identity. Like eigenphases, this algorithm has been shown to give good results on the PIE database and is robust to problems of occlusion and illumination variation. It also has the benefit that, because the probe image is tested at different displacements, it does not require the probe to be precisely registered. A potential problem with the method, however, is its high computational overhead. Because of this, an adaptation of CPR, known as *class-dependence feature analysis* (CFA), is described in [79] that sacrifices shift-invariance in order to achieve good illumination-invariant results at an acceptable computational cost; the experiments being conducted on the FRGC data set. The CFA algorithm consists of designing a filter for each of 222 subjects in a training database; these filters are then applied to gallery and probe images to produce 222-dimensional feature vectors and classification is performed using either cosine distance or SVMs.

2.5 Representations of 3D Face Scans

Although many of the ideas described in this chapter were originally designed with 2D intensity images in mind, they can often be readily adapted to 3D face scans provided the raw data is first expressed in a standard vectorised form. This means ensuring that the feature vectors derived from different face scans are of equal length and that corresponding components of the vectors have the same meaning in terms of the geometry of the face. The process of deriving such standard feature vectors should also allow for the correction of any defects in the 3D scan such as spurious holes or spikes.

One possible representation, an approach sometimes referred to as 2.5D, is to render the 3D scans as range images and then to treat them in a similar way to 2D greyscale intensity images. These range images consist of rectangular pixel grids, with the pixel brightness being proportional to the z-coordinate of the face surface rather than intensity of reflected light. The main advantage of this approach is that it allows many of the techniques that have been developed for 2D intensity images to be applied unchanged to 3D face recognition. For example Li et al. [41] have successfully achieved a fusion of 2D and 2.5D information by selecting AdaBoosted local binary pattern features (see section 3.5) from a pooled set of 2D greyscale and 2.5D range images. Some disadvantages of range images are that detailed information may be lost due to the need to discretise the 3D point-cloud data into a fixed array of pixels, absolute size information is lost due to the need to rescale and register images in a consistent way and it precludes the use of algorithms that are based on inherently 3D concepts such as shape or curvature.

An alternative to the range image approach, and one which allows for the application of a richer set of feature extraction and classification algorithms, is to retain the full 3D structure of the data, but to normalise it so that a fixed set of vertices is retained, with the vertices from different scans being in one to one correspondence with each other. This may be accomplished through the method of dense non-rigid 3D face registration which establishes a dense correspondence between a given 3D face scan and a 3D face model. In [5] this is done by representing a probe as the minimum error linear combination of a fixed basis of sample scans. A modified optical flow algorithm is used to perform matching between any pair of scans. The approach taken in [74] is to start with a generic 3D face model and to gradually morph it, using a thin-plate spline technique, until its shape is as close as possible to that of the probe scan. This is described in more detail in section 3.1.

A further benefit of using the method of dense 3D correspondence to create a standard representation of a probe image is that, once the correspondence has been established, the model can be manipulated in a number of ways to achieve desirable effects. For example, texture from a 2D image can be mapped onto the 3D model to produce a photo-realistic 3D reconstruction; this can then be rotated and relit to correct for a non-frontal pose or adverse illumination conditions [75, 7]. As another example, in [13] MDS is used to map 3D face scans to an expression invariant canonical representation that allows for higher face recognition accuracy in the presence of expression changes.

2.6 Summary

In this chapter we have looked at a number of aspects of the face recognition task, with particular reference to the representations that are used to encode face information and the distance measures that are used to assess the similarity or otherwise between different face images.

Face recognition is carried out effortlessly by humans and so it is worthwhile to investigate what is known about how the brain carries out this complex task. Psychological studies show that faces are represented in the brain as vectors in a multi-dimensional feature space. When assessing the similarity of two points in this feature space it appears that the Manhattan metric is closer to the human idea of similarity than the Euclidean metric; however there is evidence that the actual similarity measure used by the brain does not satisfy the axioms of a metric space at all and that other approaches, such as those based on fuzzy logic, may provide a more accurate model. Another area of active investigation is the question of whether relationship to the mean face plays any special role in the process of human face recognition (norm-based coding) or whether faces are identified solely on the basis of their position in feature space (absolute-based coding). Whilst this question is still open, it is known that distinctive faces, for example caricatures, are more easily recognised that those with more typical facial features and also that there is a large perceptual difference between the anti-face and the veridical.

Dimensionality reduction is an important aspect of most machine-based face recognition systems. It achieves the twin goal of reducing the computational overhead of processing face data whilst at the same time removing extraneous noise and thus focusing on information which is useful for discrimination purposes. PCA is a widely used unsupervised technique that accomplishes dimensionality reduction by finding the directions of maximal variance in feature space and eliminating the others. LDA is a supervised method that similarly finds a low dimensional subspace, but balances the need to maximise between-class scatter with the need to minimise within-class scatter. Several studies have shown that LDA is a more effective feature extraction technique than PCA for face recognition purposes. Both methods can be extended to take account of non-linearities in the data by using a kernel based approach.

Another way of achieving dimensionality reduction is to directly model the low-dimensional non-linear manifold that represents the space of possible face images within the ambient space of all images. MDS, isomap and LLE are three successful algorithms for performing this function, however their primary use is in visualising the relationships between complex data and they do not easily generalise to previously unseen data, as is required for face recognition tasks. The method of Laplacianfaces overcomes this problem by finding a linear subspace that approximates the local structure of the face manifold and thus allows any arbitrary face image to be mapped into the representation.

Having selected a reduced dimensionality representation for face images, it is then necessary to choose the method that will be used to assess the degree of similarity or dissimilarity between two face images. Several possibilities have been explored for this purpose; they include the Euclidean metric, the Manhattan metric, Mahalanobis distance and Hausdorff metric (between binary images). Alternatively, in Bayesian face recognition the distributions of intra-class and inter-class differences are modelled and a probabilistic similarity measure is used. A further option is to artificially construct a metric that, as far as possible, gives rise to small distances between training images belonging to the same person and larger distances otherwise.

Angular methods have proved to be of value in face recognition. It has been shown experimentally that similarity measures based on the angular separation between feature vectors tend to outperform non-angular metrics such as Euclidean, Manhattan and Mahalanobis. For example, good results have been obtained when LDA feature extraction is combined with the use of cosine distance (or the closely related normalised correlation similarity measure). When PCA is used it has been found beneficial to apply a pre-whitening step before using cosine distance to ensure that all dimensions are given equal emphasis. Alternatively, a weighted angular distance may be used to similar effect. If sufficient training data is available for each subject (as when the training images are extracted from video sequences) then it is feasible to model each face manifold as a separate linear subspace; in this case the principal angles between subspaces have been shown to lead to good discrimination between different face identities. In the Fourier domain, phase spectrum information has been shown to be more important than the magnitude spectrum in encoding the details of an image; a fact which is exploited in the eigenphases and CPR algorithms which both give good face recognition performance, particularly in the presence of occlusions and illumination variation.

In face recognition from 2D images, the raw data is presented as a rectangular array of pixel values (or three arrays in the case of colour images). For 3D scans, however, a choice of representations exists. One approach is that of 2.5D whereby range images are treated as though they were greyscale images. This approach allows established 2D techniques to be used but does not fully utilise the 3D shape information that is present in the data. An alternative representation, based on creating a standardised vector of 3D vertices in dense correspondence, provides a richer set of possibilities for manipulating the 3D information.

2.7 Conclusions from the Literature Review

In designing a machine-based system that will recognise faces from still images it is first necessary to choose the feature space representation that will be used to encode the captured facial data. Given such a representation, a choice must then be made as to the distance measure which will be used when deciding whether two feature vectors represent the same personal identity or not.

Several such distance measures have been proposed in the literature and it has been observed that angular separation tends to be particularly beneficial for face recognition applications. It is worthwhile, therefore, to examine this observation in more detail and to compare the effectiveness of angular separation with that of other commonly used metrics under a range of feature extraction scenarios. Some comparative studies have been previously reported in the literature, however they are usually carried out by measuring the accuracy of specific classification methods using different metrics. Whilst this is a important aspect of the research, there also a case for looking at the performance of the metrics in isolation so that it is not obscured by the details of any particular classifier design. Previous research in this area has also focused heavily on 2D facial images, with relatively little attention having been paid to the 2.5D and 3D modalities.

Much of the work to date on the relative performance of angular and other metrics has been carried out using techniques such as nearest neighbour classifiers in a PCA or LDA feature space. It is desirable to extend this work to include more sophisticated methods such as ensembles of SVMs and to investigate how such methods can be adapted so as to gain maximum benefit from the enhanced discrimination capabilities of angular separation metrics.

Chapter 3

Face Recognition Techniques

The previous chapter provided a survey of the current literature relating to face representations and metrics. In this chapter we describe in some detail several established face recognition algorithms and techniques; this list is not intended to be exhaustive, but is limited to those techniques which are referred to later in the experiments of chapters 6 to 8. With this in mind, a range of topics is covered, from methods of image normalisation through dimensionality reduction and feature extraction to the those used in the construction of ensemble classifiers.

3.1 2D and 3D Geometric Registration

The first step in applying face recognition algorithms is to geometrically normalise an image or 3D scan so that the facial features - nose, eyes, mouth etc. - occur at the same geometrical location in all cases. In this thesis we assume that a manual procedure is used to initially landmark the position of a small number of standard facial features; this allows the effect of classification algorithms to be studied independently of algorithms for performing automatic landmark location.

For 2D images just the (x, y) coordinates of the eye centres are required as a minimum. The image can then be translated, rotated, rescaled and cropped so as to map those coordinates to a standard location in the registered image. Further landmarks, such

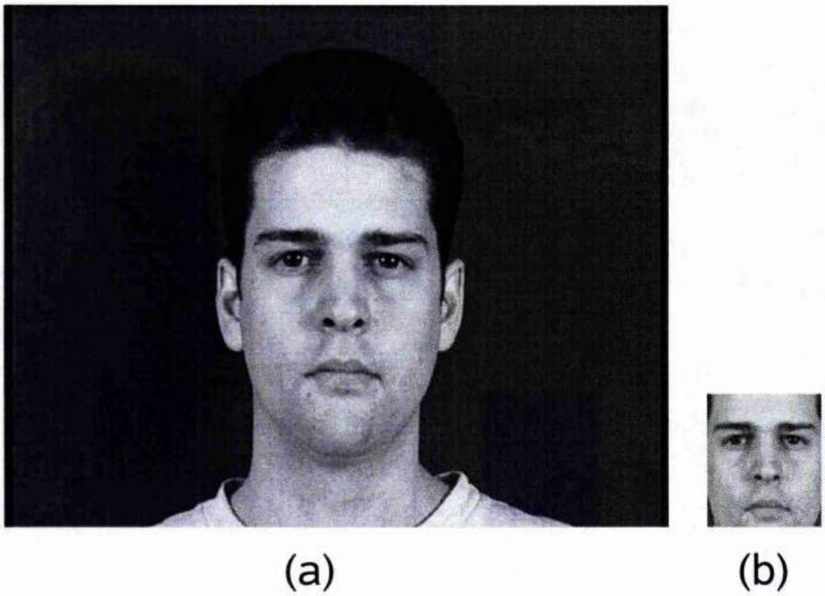

Figure 3.1: An example of 2D geometric normalisation. (a) the original image, (b) after rescaling and cropping.

as the tip of the nose, corners of the mouth or tip of the chin may also be used to achieve a closer geometric fit between images. An example of a 2D image before and after geometric registration is shown in Fig. 3.1.

As noted in section 2.5, there are various options for geometrically registering and representing 3D scans. Here we adopt the method of dense non-rigid 3D face registration using a morphable model, as described in [74]. The algorithm proceeds in three stages. Firstly a global mapping brings a pre-defined set of four landmarks (eye centres, nose tip and chin tip) into exact alignment with the generic model by using a thin-plate spline technique. Secondly local matching establishes a correspondence between each vertex of the probe and a vertex of the generic model. Finally, an energy minimisation stage fine-tunes the correspondences to minimise the RMS fitting error. Fig. 3.2 shows the result of applying this method of normalisation to a 3D face scan.

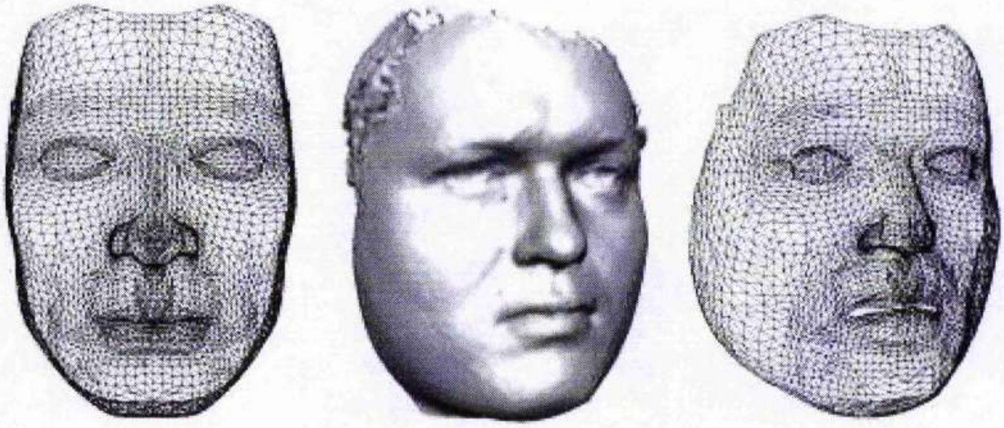

Figure 3.2: 3D normalisation. From left to right this shows the generic morphable model, an example of a 3D surface scan and the corresponding mesh after placing the scan into dense correspondence with the morphable model.

3.2 2D Photometric normalisation

As previously noted, the differences between a pair of two dimensional intensity images that are caused by factors such as lighting, pose angle and expression, can be greater than those caused by differences in the identity of the subjects. The purpose of photometric normalisation is to alleviate one of these sources of noise, namely illumination variability.

In general, the intensity of an image $I(x, y)$ at any point (x, y) can usefully be regarded as a product of two separate functions:

$$I(x, y) = R(x, y) L(x, y) \tag{3.1}$$

where $R(x, y)$ is the reflectance of the surface at (x, y) and $L(x, y)$ is the illuminance at that point. Of these two functions reflectance conveys important information about the object being viewed whereas illuminance is the chance result of lighting conditions and will often vary greatly between different images of the same object. *Homomorphic filtering* [53] is an algorithm that seeks to remove the effect of variations in $L(\cdot, \cdot)$, and hence to recover $R(\cdot, \cdot)$, by applying a high pass filter to the image. This is based on

Figure 3.3: Application of photometric normalisation to face images with varying degrees of illumination variation. (a) the original images, (b) after homomorphic filtering, (c) after homomorphic filtering and histogram equalisation.

the assumption that changes caused by illumination variability are characterised by a low spatial frequency whilst those due to the surface features and texture of the object have a high spatial frequency.

Variations in image contrast may be another problem that makes it difficult to reliably compare images of the same or different subjects. This may be caused by poor illumination or by differences in the sensitivity of the photographic equipment which was used to capture the images. *Histogram equalisation* [53] improves and standardises image contrast by spreading out the probability distribution of pixel intensities over a wider range. This is achieved by applying a monotonic transformation to pixel intensities with the property that the cumulative histogram of the transformed image is linear.

Homomorphic filtering can lead to a general darkening of the image and loss of contrast so it is beneficial to follow this by a histogram equalisation stage to restore or enhance the contrast. Fig. 3.3 shows the effect of applying homomorphic filtering and histogram equalisation to a 2D image.

Another successful and commonly used method of illumination correction is that proposed by Gross and Brajovic [23]. This approach uses insights gained from the psychology of human vision to estimate the illumination field $L(\cdot,\cdot)$ based on local values of the intensity $I(\cdot,\cdot)$; this estimate is then substituted into Eqn. 3.1 to obtain the reflectance field $R(\cdot,\cdot)$.

3.3 Principal Components Analysis

Given a training set $\mathcal{T} = \{\mathbf{x}_s : s = 1 \dots N\}$ consisting of N face image column vectors, the PCA algorithm consists of finding the eigenvectors (often referred to as *eigenfaces*) of the covariance matrix of the mean-subtracted training images and ranking them in decreasing order of eigenvalue. This gives rise to an orthonormal basis of eigenfaces where the first eigenface gives the direction of maximum variance or *scatter* within the training set and subsequent eigenfaces are associated with steadily decreasing levels of scatter. A probe image can be represented as a linear combination of these eigenfaces and, by choosing a cut-off point beyond which the basis vectors are ignored, a reduced dimension approximation to the probe image can be obtained.

More formally, the PCA approach is as follows. The sample covariance matrix of $\mathcal{T}$ is defined as an average outer product:

$$\mathbf{S} = \frac{1}{N}\sum_{p=1}^{N}(\mathbf{x}_s - \mathbf{m})(\mathbf{x}_s - \mathbf{m})^{\mathrm{T}} \tag{3.2}$$

where $\mathbf{m}$ is the sample mean column vector given by

$$\mathbf{m} = \frac{1}{N}\sum_{s=1}^{N}\mathbf{x}_s. \tag{3.3}$$

Hence the first step in the PCA algorithm is to find an orthonormal projection matrix $\mathbf{U} = [\mathbf{u}_1, \dots \mathbf{u}_L]$ that diagonalises $\mathbf{S}$ so that

$$\mathbf{SU} = \mathbf{U}\mathbf{\Lambda} \tag{3.4}$$

where L is the number of input dimensions and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. The columns $\mathbf{u}_q$ of $\mathbf{U}$ then constitute a new orthonormal basis of eigenfaces for the image space and we may assume, without loss of generality, that they are ordered so that their associated eigenvalues λ_q form a non-increasing sequence, that is:

$$q < r \Rightarrow \lambda_q \geq \lambda_r \tag{3.5}$$

for $1 \leq q, r \leq L$.

An important property of this transformation is that, with respect to the basis $\{\mathbf{u}_q\}$, the coordinates of the training vectors are decorrelated. Thus each $\mathbf{u}_q$ lies in a direction in which the total scatter between images, as measured over $\mathcal{T}$, is statistically independent of the scatter in other orthogonal directions. By virtue of Eqn. 3.5 the scatter is maximum for $\mathbf{u}_1$ and decreases as the index q increases. For any probe column vector $\mathbf{x}$, the vector $\mathbf{x}' = \mathbf{U}^{\mathrm{T}}(\mathbf{x} - \mathbf{m})$ is the projection of the mean-subtracted vector $\mathbf{x} - \mathbf{m}$ into the coordinate system $\{\mathbf{u}_q\}$ with the components being arranged in decreasing order of training set scatter. An approximation to $\mathbf{x}'$ may be obtained by discarding all but the first $M < L$ components to obtain the column vector $\mathbf{x}'' = [x'_1, \ldots, x'_M]^{\mathrm{T}}$. The value of M is chosen such that the RMS pixel-by-pixel error of the approximation is below a suitable threshold value. For face data sets it is found in practice that M can be chosen such that $M \ll L$ and so this procedure leads to the desired dimensionality reduction. The resulting linear subspace preserves most of the scatter of the training set and thus permits face recognition to be performed effectively within it.

The matrices $\mathbf{S}$ and $\mathbf{U}$ of Eqn. 3.4, being of size $L \times L$, are large matrices and thus present computational difficulties in solving the eigenvector problem. In fact the rank of $\mathbf{S}$ can be no larger than $N - 1$, and typically $N \ll L$, so most of the eigenvalues will be zero. This observation allows an alternative solution to the problem to be found as follows [77]. Firstly we note that the matrix $\mathbf{S}$ can be represented as

$$\mathbf{S} = \mathbf{T}\mathbf{T}^{\mathrm{T}} \tag{3.6}$$

where

$$\mathbf{T} = [\mathbf{x}_1 - \mathbf{m}, \ldots, \mathbf{x}_N - \mathbf{m}] \tag{3.7}$$

is the mean-subtracted training set expressed as an $L \times N$ matrix. If $\mathbf{v}$ is any solution to the eigenvalue equation

$$\mathbf{T}^{\mathrm{T}}\mathbf{T}\mathbf{v} = \lambda\mathbf{v} \tag{3.8}$$

with eigenvalue λ, then pre-multiplying by $\mathbf{T}$ gives $\mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{T}\mathbf{v} = \mathbf{S}\mathbf{T}\mathbf{v} = \lambda\mathbf{T}\mathbf{v}$, so $\mathbf{T}\mathbf{v}$ is also an eigenvector of $\mathbf{S}$. In this way the $N-1$ eigenvectors of $\mathbf{S}$ can be found as linear combinations of the training set vectors by solving Eqn. 3.8 and pre-multiplying each eigenvector solution by $\mathbf{T}$. Since the matrix $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is only of size $N \times N$, Eqn. 3.8 is a less computationally demanding problem to solve than Eqn. 3.4.

3.4 Linear Discriminant Analysis

The LDA algorithm operates as follows. Let $\Omega = \{\omega_1, \ldots, \omega_C\}$ be the complete set of class labels (personal identities) under consideration, these having prior probabilities $\Pr(\omega_l)$, and let the known class label of each training column vector $\mathbf{x}_s \in \mathcal{T}$ be represented by the variable $t_s \in \Omega$. The sample class covariance matrix for each ω_l is defined as

$$\mathbf{S}_l = \frac{1}{N_l} \sum_{m=1}^{N} I\left(t_m = \omega_l\right) \left(\mathbf{x}_m - \mathbf{m}_l\right) \left(\mathbf{x}_m - \mathbf{m}_l\right)^{\mathrm{T}} \tag{3.9}$$

where N_l is the number of representatives of class ω_l in $\mathcal{T}$, $I(\cdot)$ is the indicator function and $\mathbf{m}_l$ is the class mean column vector defined as

$$\mathbf{m}_l = \frac{1}{N_l} \sum_{m=1}^{N} I\left(t_m = \omega_l\right) \mathbf{x}_m. \tag{3.10}$$

Fisher's criterion is a scalar function $J_F : \mathbb{R}^L \mapsto \mathbb{R}$ that is defined, for an arbitrary column vector $\mathbf{u}$, as

$$J_F(\mathbf{u}) = \frac{\mathbf{u}^{\mathrm{T}}\mathbf{S}_{\mathbf{B}}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{S}_{\mathbf{W}}\mathbf{u}} \tag{3.11}$$

where

$$\mathbf{S}_{\mathbf{B}} = \sum_{l=1}^{C} \Pr(\omega_l) \left(\mathbf{m}_l - \mathbf{m}\right) \left(\mathbf{m}_l - \mathbf{m}\right)^{\mathrm{T}} \tag{3.12}$$

is known as the between-class covariance (or scatter) matrix and

$$\mathbf{S}_{\mathbf{W}} = \sum_{l=1}^{C} \Pr(\omega_l) \mathbf{S}_l \tag{3.13}$$

is known as the within-class covariance (or scatter) matrix. The value of $J_F(\mathbf{u})$ is a measure of the effectiveness of direction $\mathbf{u}$ in separating the classes from each other when all training vectors are projected onto this single dimension. The aim of the LDA method is to find a set of basis vectors $\{\mathbf{u}_q\}$ that maximise the values of $J_F(\mathbf{u}_q)$, subject to the constraint that $\mathbf{u}_q\mathbf{S_W}\mathbf{u}_r = \delta_{qr}$ which ensures that class-centralised vectors in the transformed space are uncorrelated. This leads to the generalised eigenvector problem

$$\mathbf{S_B}\mathbf{U} = \mathbf{S_W}\mathbf{U}\mathbf{\Lambda} \tag{3.14}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of generalised eigenvalues and $\mathbf{U} = [\mathbf{u}_1, \ldots \mathbf{u}_L]$ is the desired projection matrix. As with PCA it can be assumed that the generalised eigenvectors have been re-ordered so that the generalised eigenvalues form a non-increasing sequence.

If $\mathbf{S_W}$ is non-singular then Eqn. 3.14 can be solved by pre-multiplying by $\mathbf{S_W}^{-1}$ and using standard eigenanalysis methods to solve the resulting eigenvalue equation

$$\mathbf{S_W}^{-1}\mathbf{S_B}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}. \tag{3.15}$$

Note, however, that the matrix $\mathbf{S_W}^{-1}\mathbf{S_B}$ is not symmetric so the resulting eigenvectors, although they form a basis of the projected feature space, are not orthogonal. Also note that the rank of $\mathbf{S_B}$ is at most $C-1$ so the feature space will have at most $C-1$ dimensions.

For face recognition problems this approach cannot be applied directly because there is usually insufficient training data, relative to the dimensionality of the training images. This means that the rank of $\mathbf{S_W}$, which is at most $N-C$, is much less than the number of input features L. The solution proposed in [8] and also applied here, is to precede the LDA calculation by a PCA stage so as to transform to an intermediate space of dimensionality $N-C$. In this space $\mathbf{S_W}$ is non-singular and this allows Eqn. 3.15 to be solved.

Other variants of LDA have been proposed which take a different approach to solving this small sample size problem. For example, it has been observed [17, 84] that the null space[1] of the within-class scatter matrix is useful, from the point of view of discrimination between classes, because any variation within this space must be due solely to differences in face identity and not to extraneous factors such as changes in

[1] The null space of a matrix $\mathbf{A}$ is defined as the sub-space of vectors $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{0}$, that is the space of eigenvectors which have an associated eigenvalue of 0.

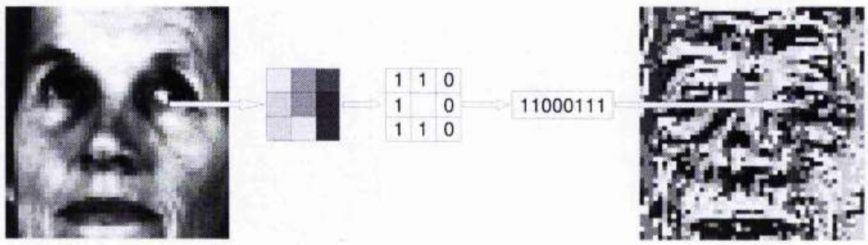

Figure 3.4: Local binary pattern image production. Each non-border pixel is mapped as shown.

lighting conditions. Chen [17] has proposed a method in which *only* the null space is used (provided the within-class scatter matrix is singular). The approach taken by Yu [84] is to first discard the null space of $\mathbf{S_B}$ (since this can contain no discriminatory information) and then to look for the most discriminating eigenvectors of a modified $\mathbf{S_W}$, giving priority to those with smaller (but not necessarily zero) eigenvalues.

Note that for 2-class problems the output feature space defined by Eqn. 3.15 is just 1 dimensional so the LDA procedure described above leads to a particularly simple form in which a single unit vector $\hat{\mathbf{u}}$, the *Fisher vector*, is obtained such that projection of the input feature vectors in the direction of $\hat{\mathbf{u}}$ leads to maximal discrimination between the two classes. In this case $\hat{\mathbf{u}}$ is given by the equation

$$\hat{\mathbf{u}} = \frac{\mathbf{S_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2)}{\left\|\mathbf{S_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\right\|}. \tag{3.16}$$

3.5 2D Local Binary Patterns

The *local binary pattern* (LBP) operator is a powerful 2D texture descriptor which was first used in face recognition by Ahonen et al [2]. As illustrated in Fig. 3.4, the method associates each interior pixel of an intensity image with a binary code number in the range 0-256. This code number is generated by taking the surrounding pixels and, working in a clockwise direction from the top left hand corner, assigning a bit value of 0 where the neighbouring pixel intensity is less than that of the central pixel and 1 otherwise. The concatenation of these bits produces an eight-digit binary code word which becomes the grey-scale value of the corresponding pixel in the transformed image. Fig. 3.4 shows a pixel being compared with its immediate neighbours. It is however also possible to compare a pixel with others which are separated by distances of two, three or more pixel widths, giving rise to a series of transformed images. Each such image is generated using a different radius for the circularly symmetric neighbourhood over which the LBP code is calculated. Another possible refinement is to obtain a finer angular resolution by using more than 8 bits in the code-word [50]. Note that the choice of the top left hand corner as a reference point is arbitrary and that different choices would lead to different LBP codes; valid comparisons can be made, however, provided that the same choice of reference point is made for all pixels in all images.

It is noted in [50] that in practice the majority of LBP codes consist of a concatenation of at most three consecutive sub-strings of 0s and 1s; this means that when the circular neighbourhood of the centre pixel is traversed, the result is either all 0s, all 1s or a starting point can be found which produces a sequence of 0s followed by a sequence of 1s. These codes are referred to as uniform patterns and, for an 8 bit code, there are 58 possible values. Uniform patterns are most useful for texture discrimination purposes as they represent local micro-features such as bright spots, flat spots and edges; non-uniform patterns tend to be a source of noise and can therefore usefully be mapped to a single common value.

In order to use LBP codes as a face comparison mechanism it is first necessary to subdivide a face image into a number of sub-windows and then compute the occurrence histograms of the LBP codes over these regions. These histograms can be combined to generate useful features, for example by concatenating them or by comparing corresponding histograms from two images.

Li [41] has used AdaBoost [20] to select the best histogram bins from a large number of candidate histograms, each corresponding to a different sub-window size and position. This technique has been applied to a fusion of 2D intensity and 2.5D range data. Chan [16] has obtained good results on 2D images by averaging the distances between corresponding sub-windows in a non-overlapping rectangular tiling. For each sub-window the histograms at multiple LBP scales are concatenated and LDA is used to reduce the dimensionality. Cosine distance is then used to compare corresponding sub-windows from two different images. In this thesis we adopt a modified form of this algorithm in which the LBP histograms from each non-overlapping sub-window are concatenated and a then a single LDA stage is performed.

3.6 Support Vector Machines

The *support vector machine* (SVM) concept is an increasingly popular tool for solving problems in pattern recognition. As a classification method, it falls into the category known as *discriminant analysis* [80]. Here the aim is not to directly model the class-conditional probability distributions $p(\mathbf{x} \mid \omega_i)$ of a problem, but rather to find a *decision boundary* that optimally separates two classes into different regions of feature space. Once such a boundary has been determined, a probe vector $\mathbf{x}$ can be assigned to a class by determining on which side of the boundary it falls. Compared with other discriminant analysis methods, such as neural networks, the SVM approach has the advantages that that it is based on sound theoretical principles, it produces a single globally optimal solution, it is less susceptible to "curse of dimensionality" problems and is thus less prone to overfitting.

An introductory tutorial on SVMs, which includes some example applications, can be found in [28] and a more detailed tutorial on their operation is given in Burges [14]. In this section we present without proof a brief overview of the main principles of SVMs.

The general aim in the SVM approach is to find a *discriminant function* $f : \mathbb{R}^L \mapsto \mathbb{R}$ that optimally separates two classes by mapping members of one class to positive values and members of the other class to negative values. The function is chosen from a restricted family of functions whose *capacity* is limited in accordance with the amount of available training data [28]. Here the term capacity comes from Vapnik-Chervonenkis theory and refers to the maximum number of points in a given feature space that can be partitioned into two classes in all possible ways. If the capacity is not limited in this way then any pair of classes can be fully separated. This leads to overfitting on the training set and consequently to poor generalisation performance.

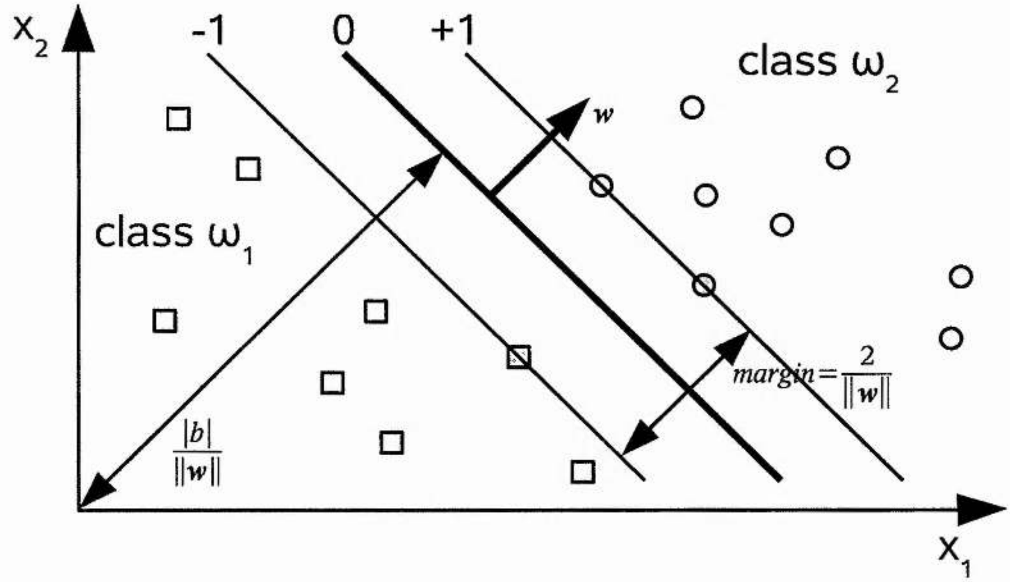

Figure 3.5: An example linear SVM for a 2-dimensional feature space in which the two classes ω_1 and ω_2 are linearly separable. The support vectors are shown as shaded.

In the simplest scenario for SVM application the two target classes ω_1 and ω_2 are *linearly separable*, that is there exist hyperplanes in the feature space such that all the training examples for ω_1 lie on one side and those for ω_2 lie on the other side. In this case it is reasonable to restrict the family of decision surfaces to the set of all possible hyperplanes and to look for the one which achieves optimal separation of the two classes. This give rise to the class of SVMs known as *linear SVMs*, an example of which is illustrated in Fig. 3.5.

The optimal separating hyperplane is the one for which the *margin* (that is the sum of the shortest distances from the hyperplane to the nearest examples of classes ω_1 and ω_2) is maximised. To solve this problem we must look for a minimal length vector $\mathbf{w}$ and scalar value b such that the set of constraints

$$f(\mathbf{x}_s) \geq +1 \; for \; t_s = +1 \tag{3.17}$$

$$f(\mathbf{x}_s) \leq -1 \; for \; t_s = -1 \tag{3.18}$$

is satisfied for $1 \leq i \leq N$. Here $\mathbf{x}_s \in \mathcal{T}$ is any training-set column vector and

$$f(\mathbf{x}_s) = \mathbf{x}_s^{\mathrm{T}}\mathbf{w} + b \tag{3.19}$$

is the discriminant function of the linear SVM; the values of $\|\mathbf{w}\|$ and b are scaled so that $f(\mathbf{x}_s) = \pm 1$ for members of ω_1 and ω_2 that are closest to the separating hyperplane. Such closest points are known as *support vectors* and they lie on the two parallel hyperplanes labelled $+1$ and -1 in Fig. 3.5; the optimal separating hyperplane, for which $f(\mathbf{x}) = 0$, is also shown in this diagram and labelled 0.

Using the method of Lagrange multipliers [4] to solve this optimisation problem leads to the primal form of the objective function

$$L_P = \frac{1}{2}\|\mathbf{w}^2\| - \sum_{s=1}^{N} \alpha_s t_s \left(\mathbf{x}_s^{\mathrm{T}}\mathbf{w} + b\right) + \sum_{s=1}^{N} \alpha_s \tag{3.20}$$

where $\alpha_s \geq 0$ are undetermined multipliers, one for each training point. L_P must be minimised with respect to $\mathbf{w}$ and b whilst simultaneously requiring that the its derivatives with respect to all the α_s vanish.

This is a convex quadratic programming problem and it can be solved [14] by maximising the dual form of the objective function

$$L_D = \sum_{s=1}^{N} \alpha_s - \frac{1}{2}\sum_{s,p=1}^{N} \alpha_s \alpha_p t_s t_p \mathbf{x}_s^{\mathrm{T}}\mathbf{x}_p. \tag{3.21}$$

subject to the constraints

$$\alpha_s \geq 0 \tag{3.22}$$

and

$$\sum_{s=1}^{N} \alpha_s t_s = 0 \tag{3.23}$$

This has a unique solution for $\{\alpha_s\}$ which can be used to obtain the required value for $\mathbf{w}$ through the equation

$$\mathbf{w} = \sum_{s=1}^{N} \alpha_s t_s \mathbf{x}_s. \tag{3.24}$$

The support vectors are those for which $\alpha_s > 0$ with the rest being zero. Hence, as expected, the value of $\mathbf{w}$ is determined only by the support vectors and other training points have no influence on it. Once $\mathbf{w}$ has been determined, the value of b can be

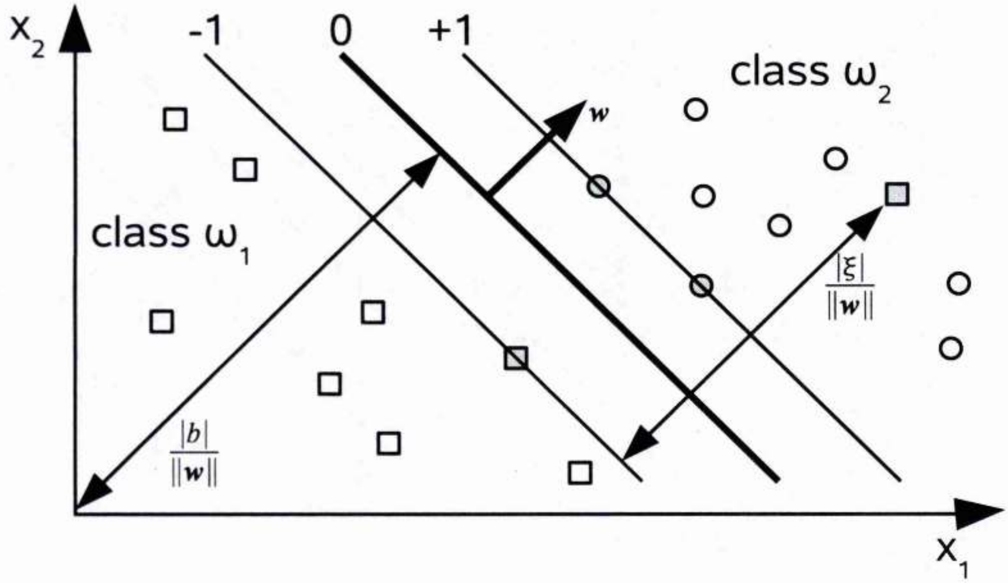

Figure 3.6: An example linear SVM for a 2-dimensional feature space in which the two classes ω_1 and ω_2 are not linearly separable. The support vectors are shown as shaded.

obtained by substitution into Eqn. 3.17 or 3.18 using any support vector, or, better, by averaging over all support vectors.

As illustrated in Fig. 3.6, to extend the linear SVM algorithm to the non-separable case it must be modified to allow some data points to occur on the wrong side of the separating hyperplane. This is done by introducing *slack variables* ξ_s, $s = 1 \dots N$ with

$$\xi_s \geq 0 \tag{3.25}$$

and relaxing the constraints 3.17 and 3.18 so that they become

$$f(\mathbf{x}_s) \geq +1 - \xi_s \; for \; t_s = +1 \tag{3.26}$$

$$f(\mathbf{x}_s) \leq -1 + \xi_s \; for \; t_s = -1. \tag{3.27}$$

The only difference that this makes to the linear SVM algorithm is that a user-defined *cost* parameter G must be introduced to place an upper bound on the values of α_s. Thus, the constraints of Eqn. 3.22 are replaced by

$$G \geq \alpha_s \geq 0. \tag{3.28}$$

This prevents the values of α_s growing without bound and allows a solution to be found. This solution gives rise to two types of support vector. The *margin vectors* are those for which $0 < \alpha_s < G$; they lie on the hyperplanes defined by $f(\mathbf{x}) = \pm 1$ and have $\xi_s = 0$. Only those support vectors for which $\alpha_s = G$ have non-zero values of the slack variables ξ_s. For these variables $\xi_s \geq 1$ implies that $f(\mathbf{x}_s)$ is in error because it has the incorrect sign for the data point $\mathbf{x}_s$, and $\xi_s < 1$ means that the point is classified correctly although it lies at a distance of less than $1/\|\mathbf{w}\|$ from the separating hyperplane.

If SVMs were restricted to linear decision boundaries then their usefulness would be severely limited. This is because, in practice, many datasets occupy non-linear manifolds in a feature space $\mathbb{R}^L$ and so cannot be adequately partitioned using linear boundaries. Fortunately the method described above for linear SVMs can be readily extended to the problem of finding non-linear decision boundaries by making use of the "kernel trick". This approach is based on the observation that it is only the scalar products of training vectors $\mathbf{x}_s^{\mathrm{T}}\mathbf{x}_p$ that appear in the dual form of the objective function L_D of Eqn. 3.21 and not their individual values. The scalar product term can be replaced, therefore, by a suitable kernel function $K(\mathbf{x}_s, \mathbf{x}_p)$ which represents the effect of using an associated non-linear function $\phi : \mathbb{R}^L \mapsto \mathbb{R}^S$ to project $\mathbf{x}_s$ and $\mathbf{x}_p$ into a higher dimensional feature space and then taking the scalar product within that space. The objective function to be maximised thus becomes

$$L_D = \sum_{s=1}^{N} \alpha_s - \frac{1}{2} \sum_{s,p=1}^{N} \alpha_s \alpha_p t_s t_p K(\mathbf{x}_s, \mathbf{x}_p) \tag{3.29}$$

where

$$K(\mathbf{x}_s, \mathbf{x}_p) = \phi(\mathbf{x}_s)^{\mathrm{T}} \phi(\mathbf{x}_p). \tag{3.30}$$

For any probe vector $\mathbf{x}$ the value of the discriminant function can be computed as

$$f(\mathbf{x}) = \sum_{s=1}^{N} \alpha_s t_s K(\mathbf{x}_s, \mathbf{x}) + b \tag{3.31}$$

and the value of b can be found as before by applying the constraint equations 3.25 to 3.28 to the margin vectors. Note that the non-linear transformation $\phi(\cdot)$ never has to be explicitly evaluated because it only appears in scalar products and these are determined by $K(\cdot,\cdot)$. This greatly increases the efficiency of calculations involving non-linear SVMs as it obviates the need to work with high dimensional vectors in the feature space $\mathbb{R}^S$.

Ideally, a kernel function will have the property that there exists an associated function $\phi(\cdot)$ and feature space dimension S such that Eqn. 3.30 is satisfied for any pair of input vectors. A necessary and sufficient condition for this is that the kernel satisfies *Mercer's condition*, which requires that

$$\int K(\mathbf{x},\mathbf{y})\, g(\mathbf{x})\, g(\mathbf{y})\, d\mathbf{x}d\mathbf{y} \geq 0 \tag{3.32}$$

for any function $g : \mathbb{R}^L \mapsto \mathbb{R}$ for which $\int g(\mathbf{x})^2 d\mathbf{x}$ is finite. If Mercer's condition is not satisfied then there may exist training sets for which the training process does not converge to a unique optimum solution. In practical applications, this may not be a problem for realistic data sets and kernels that do not satisfy Mercer's condition are sometimes found to be useful [10, 14].

A number of non-linear kernel functions have been explored for the application of SVMs to specific problems. In this thesis we make use of the *Gaussian kernel* for face recognition purposes and section 4 introduces two further SVM kernels that are optimised to make use of angular separation between feature vectors, rather than Euclidean separation. The Gaussian kernel is defined by the equation

$$\begin{aligned} K_{Gauss}(\mathbf{x},\mathbf{y}) &= exp\left(-\tfrac{1}{2\sigma^2}\|\mathbf{x}-\mathbf{y}\|^2\right) \\ &= exp\left(-\tfrac{1}{2\sigma^2}\left(D_{Euc}(\mathbf{x},\mathbf{y})\right)^2\right) \end{aligned} \tag{3.33}$$

where σ is an undetermined tuning parameter that determines the width of the kernel function. The value of σ, along with that of the cost parameter G, is typically determined by a method such as cross validation; this is discussed further in section 3.7. For the Gaussian kernel the associated function $\phi(\cdot)$ maps all vectors onto the unit sphere in an infinite dimensional space. Combining Eqns. 3.31 and 3.33 it can be seen that an SVM with Gaussian kernel is equivalent to a radial basis function (RBF) network representation of the discriminant function $f(\mathbf{x})$ but with the added advantage that the number and centres of the basis functions are determined automatically by the training data [14].

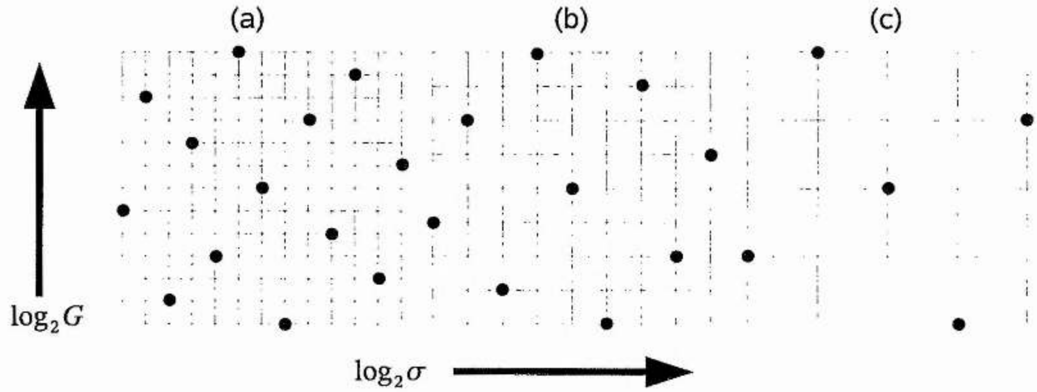

Figure 3.7: Examples of uniform design patterns with (a) 13, (b) 9 and (c) 5 sample points.

3.7 SVM Parameter Optimisation

One practical problem which arises when training SVMs is that of how to choose an optimal, or near-optimal, set of parameters which gives the lowest generalisation error for the given data set. For example, the $K_{Gauss}(\cdot,\cdot)$ kernel function of Eqn. 3.33 requires the selection of two parameters, namely the kernel width parameter σ and the cost parameter G. Such parameter selection can be performed by training a number of SVMs, with different combinations of parameter values, and noting which yields the lowest error. It is desirable to measure this error against an evaluation set $\mathcal{E}$ which is different from the training set $\mathcal{T}$ that is used in construction of the discriminant function of Eqn. 3.19, as otherwise overtraining on $\mathcal{T}$ may result.

A commonly used method for choosing SVM parameters is to perform an exhaustive grid search; that is to examine all combinations of parameter values at given intervals over a given range. For each set of values an SVM is trained on $\mathcal{T}$ and the classification error is measured on $\mathcal{E}$. Whilst this method produces the desired result it is computationally expensive as time is wasted examining parameter combinations that are close to those which are known to be sub-optimal. A gradient descent method can also be used, however this suffers from the disadvantage that it can fall into a local minimum.

In the method of *nested uniform design* [29] a two-stage search strategy is adopted. The first stage consists of sampling the search space using a uniform design pattern. This achieves the best coverage for a fixed number of sample points. Fig. 3.7 illustrates three such patterns with 13, 9 and 5 sample points. Note that each individual parameter value is tried exactly once and that the sample points are approximately uniformly distributed on a disc which is concentric with the centre of the search space. In the second stage, the best sample point from the first search is made the centre of a new search space whose width and height are half that of the first space; any duplicate points are not evaluated again. The number of sample points in the two stages may differ and it is shown in [29] that 13 followed by 9 samples (giving 21 SVM evaluations in all after removal of the duplicate point) produce near-optimal results on a variety of data sets.

3.8 SVM Calibration

Although SVM-based classification techniques do not attempt to directly model the class-conditional probability distributions of data within a feature space, it is shown later (see sections 7.1.3 and 7.2.3) that it is often desirable in practice to normalise the raw output from the discriminant function $f(\mathbf{x})$ of an SVM to obtain a value which is linearly related to the probabilities that the probe vector $\mathbf{x}$ belongs to the positive or negative target classes. This process is referred to as *calibration* and is an important factor in the use of SVMs as base classifiers in an ensemble classifier (see section 3.9) where the outputs from many disparate SVMs must be combined on an equal footing to produce an overall classification decision.

In this section we consider three possible methods for calibrating SVM output values. All three methods rely on using a representative training set $\mathcal{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_S\}$ to construct a calibration function $g : \mathbb{R} \mapsto [-1, 1]$ where

$$g(f(\mathbf{x})) = 2\widehat{\Pr}\left(\mathbf{x} \in \Omega^+\right) - 1 \tag{3.34}$$

and $\widehat{\Pr}(\mathbf{x} \in \Omega^+)$ is a probability estimate that vector $\mathbf{x}$ belongs to the positive set. A classifier is said to be well calibrated if $\widehat{\Pr}(\mathbf{x} \in \Omega^+)$ approaches the true probability $\Pr(\mathbf{x} \in \Omega^+)$ as $|\mathcal{S}|$ approaches infinity. In order to distinguish $\mathcal{S}$ from the training set $\mathcal{T}$ that was originally used to train the SVM, we refer to it as the *calibration set*. $\mathcal{S}$

may usefully be regarded as the union of two disjoint subsets, $\mathcal{S}^+$ and $\mathcal{S}^-$, consisting of training vectors which are members of Ω^+ and Ω^- respectively. It is desirable that $\mathcal{S}$ be different from $\mathcal{T}$ as otherwise overtraining on $\mathcal{T}$ may result and the probability estimates $g\left(f\left(\mathbf{x}\right)\right)$ may be unreliable. Note also that the nature of the SVM training algorithm, which is aimed at separating (most of) the positive examples from (most of) the negative examples should lead to the function g being monotonic in f.

The first method for classifier calibration to be considered is to use a *Gaussian mixture model* [80]. In this method $f\left(\mathcal{S}^+\right)$ and $f\left(\mathcal{S}^-\right)$ are modelled as Gaussian distributions with equal priors. Bayes rule is used to compute the probability value in Eqn. 3.34 as

$$\widehat{\Pr}\left(\mathbf{x} \in \Omega^+\right) = \frac{\frac{1}{s_+}\exp\left(-\frac{1}{2s_+^2}\left(f\left(\mathbf{x}\right) - \mathbf{m}_+\right)\right)}{\frac{1}{s_+}\exp\left(-\frac{1}{2s_+^2}\left(f\left(\mathbf{x}\right) - \mathbf{m}_+\right)\right) + \frac{1}{s_-}\exp\left(-\frac{1}{2s_-^2}\left(f\left(\mathbf{x}\right) - \mathbf{m}_-\right)\right)} \tag{3.35}$$

where s_+, s_-, $\mathbf{m}_+$ and $\mathbf{m}_-$ are the empirically derived standard deviation and mean values for the two distributions. Two disadvantages of this approach are that the assumption of Gaussian class-conditional probabilities may not be justified and also that it may lead to a non-monotonic calibration function.

In Platt's *sigmoid fitting* algorithm [56] monotonicity is enforced by assuming that the probability curve has a sigmoid form

$$\widehat{\Pr}\left(\mathbf{x} \in \Omega^+\right) = \frac{1}{1 + \exp\left(Af\left(\mathbf{x}\right) + B\right)} \tag{3.36}$$

where A and B are calibration parameters. To compute these parameters a maximum likelihood algorithm is used which minimises the negative log-likelihood error function

$$E\left(A, B\right) = -\sum_{\mathbf{x}_p \in S}\left[I\left(\mathbf{x}_p \in \Omega^+\right)\log\left(\widehat{\Pr}\left(\mathbf{x} \in \Omega^+\right)\right) + I\left(\mathbf{x}_p \in \Omega^-\right)\log\left(\widehat{\Pr}\left(\mathbf{x} \in \Omega^-\right)\right)\right]. \tag{3.37}$$

where $I\left(\cdot\right)$ is the indicator function.

The third approach to classifier calibration to be considered is that of *isotonic regression* (IR) by the *pair adjacent violators* (PAV) algorithm [85]. This is a non-parametric method which is akin to constructing a histogram of S with variable bin sizes, subject to the constraint that the histogram values are isotonic (i.e. monotonically non-decreasing). The operation of the algorithm is illustrated in Fig. 3.8. The calibration

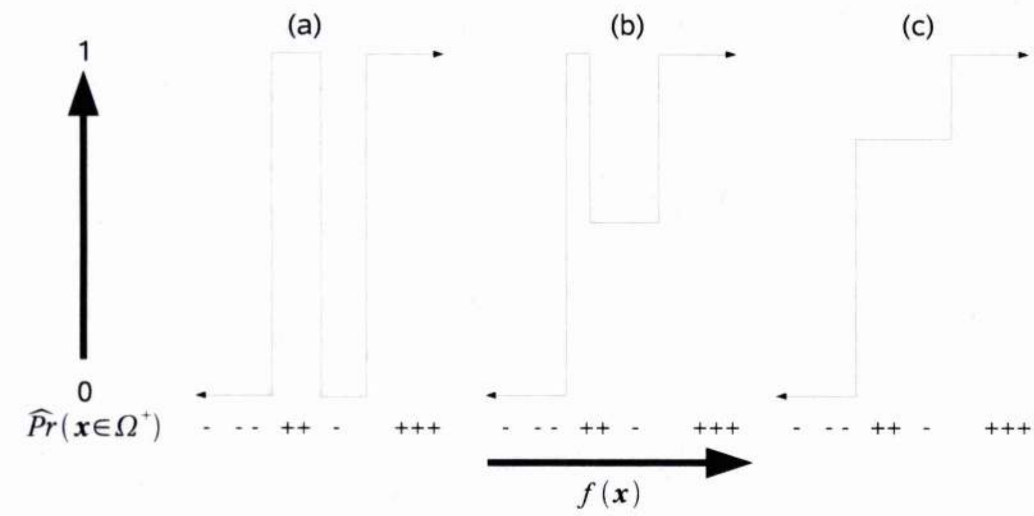

Figure 3.8: Operation of the isotonic regression by pair-adjacent violators algorithm. (a) is the initial configuration, (b) shows the result of averaging the first pair of violators and (c) the result of averaging the second (and final) pair.

set is first sorted into ascending values of $f(\mathbf{x}_p)$ and initial values of $\widehat{\Pr}(\mathbf{x}_p \in \Omega^+)$ are assigned as 0 or 1, depending on whether $\mathbf{x}_p$ belongs to Ω^- or Ω^+ respectively. The algorithm then repeatedly searches for pair-adjacent violators, that is pairs of values for which the calibration curve is non-isotonic, and replaces their calibrated values by the average of the current values. This process continues until $\widehat{\Pr}(\mathbf{x}_p \in \Omega^+)$ is fully isotonic. One advantage of isotonic regression over the sigmoid fitting algorithm is that it makes no assumptions about the form of the calibration curve (beyond the fact that it is monotonic) and thus can adapt to data sets for which the anti-symmetric sigmoid curve is not a good fit. The counter argument to this, however, is that isotonic regression does not include a method of regularisation and this could lead to overfitting on the calibration set with corresponding reduced accuracy on unseen data; this is likely to be particularly problematic when the calibration set is small.

3.9 Classifier Ensembles

Face recognition, in common with most real-world pattern recognition applications, is inherently a multi-class problem. Many successful classification techniques, however, such as SVMs and neural networks, are better suited to solving 2-class problems, or *dichotomies*. A fruitful approach to overcoming this mismatch between the needs of applications and the capabilities of classifiers has been to re-cast multi-class problems as a collection of 2-class sub-problems. A separate *base classifier* is trained to solve each sub-problem and the outputs from the ensemble of base classifiers are combined to produce an overall classification decision. These base classifiers are sometimes referred to as *dichotomisers* because they discriminate between just two classes.

Several possible architectures for constructing such ensembles have been described in the literature; these include *all-pairs* [32] in which $\frac{1}{2}C\,(C-1)$ base classifiers are trained to distinguish between each pair of target classes, *directed acyclic graph* (DAG) methods [32] in which a binary tree of $\frac{1}{2}C\,(C-1)$ base classifiers is constructed, *one-per-class* (OPC) [32] in which C base classifiers are trained to distinguish each class from the others, *error-correcting output codes* (ECOC) and *binary hierarchical classifiers* (BHC). In this thesis we evaluate the latter two methods with reference to face recognition.

Allwein [3] has proposed a general framework for describing these architectures; this consists of defining a $C \times P$ *code matrix* $\mathbf{Z}$ where P is the number of base classifiers to be deployed in the ensemble. Each row $\mathbf{Z}_i$ is associated with a single target class ω_i whilst each column $\mathbf{Z}^j$ is associated with a single base classifier $\mathcal{B}_j$. The entries Z_{ij} of the code matrix are fixed at either 0, +1 or -1. A value of 0 indicates that $\mathcal{B}_j$ is not involved in discriminating members of class ω_i and, as a consequence, is not trained with examples from class ω_i. The +1 and -1 entries in column $\mathbf{Z}^j$ define two families of target classes; these are the *positive set* $\Omega_j^+ = \{\omega_i \mid Z_{ij} = +1\}$ and the *negative set* $\Omega_j^- = \{\omega_i \mid Z_{ij} = -1\}$. Base classifier $\mathcal{B}_j$ is trained to distinguish between these two families by outputting a value of +1 and -1 respectively when presented with examples from the positive and negative families.

As base classifiers, any 2-class classification algorithm may be used, for example multi-layer perceptron neural networks or SVMs. In general a base classifier output $y_j\,(\mathbf{x})$

for an input vector $\mathbf{x}$ will not be exactly ± 1 but rather it will be a "soft" value which is positive (or zero) for Ω_j^+ and negative for Ω_j^-. If required, such soft values can be hardened into definite decisions by applying the $sign(\cdot)$ function. Alternatively, the soft base classifier outputs may be used directly; in this case, to facilitate a fair comparison between them, it may be necessary to calibrate the values so that they are related to the probabilities of class membership through Eqn. 3.34.

The operation of a classifier ensemble can be divided into two distinct stages. In the *encoding* stage each base classifier is applied to probe vector $\mathbf{x}$ to produce a vector of base classifier output values $\mathbf{y}(\mathbf{x})$ (where these may be the raw outputs or they may have been calibrated or discretised as described above). The *decoding* stage consists of applying some *decoding procedure* to $\mathbf{y}(\mathbf{x})$ in order to make a definite assignment of $\mathbf{x}$ to one of the target classes; here we denote this assignment by $\mathcal{A}(\mathbf{x}) \in \Omega$. As discussed in the following sections, each specific ensemble architecture tends to lead to its own set of decoding procedures with different advantages and disadvantages.

3.9.1 Error-Correcting Output Code Ensembles

Inspired by error-correcting codes from communications theory, the method of *error-correcting output code* (ECOC) ensembles [18, 19] is to repeatedly partition the complete family of target classes Ω into two sub-families Ω_j^- and Ω_j^+ and to construct a separate base classifier to handle each such partitioning. With this architecture, the code matrix contains no 0 entries because, in each column $\mathbf{Z}^j$ of $\mathbf{Z}$, every target class is assigned to one or the other of the families Ω_j^- and Ω_j^+. The only mandatory constraint on the code matrix is that each row $\mathbf{Z}_i$ of $\mathbf{Z}$ must be distinct from all other rows so that $\mathbf{Z}_i$ represents a unique set of base classifier target outputs that is specific to the associated target class ω_i. Beyond this requirement, there is flexibility in the choice of the number P of base classifiers to be used in any application of ECOC ; the minimum number is given by $\log_2 C$, however using more than this minimum is desirable as it introduces redundancy into the ensemble. This means that errors made by some classifiers can be compensated for by other classifiers which are not in error, leading to a correct classification decision by the ensemble as a whole. It is in this sense that the ensemble

Class	Classifier Id.							
	1	2	3	4	5	6	7	8
A	-1	-1	+1	-1	+1	+1	+1	-1
B	+1	+1	-1	-1	+1	-1	+1	-1
C	-1	-1	+1	+1	-1	-1	+1	+1
D	-1	-1	-1	+1	+1	-1	-1	+1
E	+1	+1	+1	-1	-1	+1	-1	-1
F	-1	+1	-1	+1	+1	+1	-1	-1

Table 3.1: An example ECOC code matrix for a 6-class classification task.

can be said to be error-correcting since it is analogous to transmitting a signal (class label) over a communications medium which adds noise (base classifier error) but from which the original signal can be recovered (ensemble decision). An example ECOC code matrix that uses eight base classifiers to discriminate between a set of six classes is shown in table 3.1.

In an ECOC ensemble, tolerance to base classifier errors is greatest when the rows of $\mathbf{Z}$ are chosen such that the minimum Hamming distance between any pair is as large as possible. Similarly diversity among base classifiers tends to be greatest when the columns of $\mathbf{Z}$ have maximal separation. In general, however, the construction of a code matrix that has both these properties is an NP-complete problem so an approximate method may need to be employed. The Bose-Chaudury-Hocquenghem algorithm [73] produces good row separation but may not give optimal column separation. Randomly generated code matrices, with equal probability of $+1$ or -1 in any position [33], tend to yield reasonable, though not optimal, row and column separation.

ECOC was introduced as a means of solving multi-class classification problems; that is, given an input feature vector $\mathbf{x}$, the base classifier outputs are assembled into an output vector $\mathbf{y}(\mathbf{x})$ and the objective is to make a decision as to which of the target classes is most probably indicated by $\mathbf{y}(\mathbf{x})$. The method proposed in [19] is to base the decision on the Hamming distance between the output vector and the class codewords. The base classifier outputs are first discretised to ± 1 so that they each make a definite decision

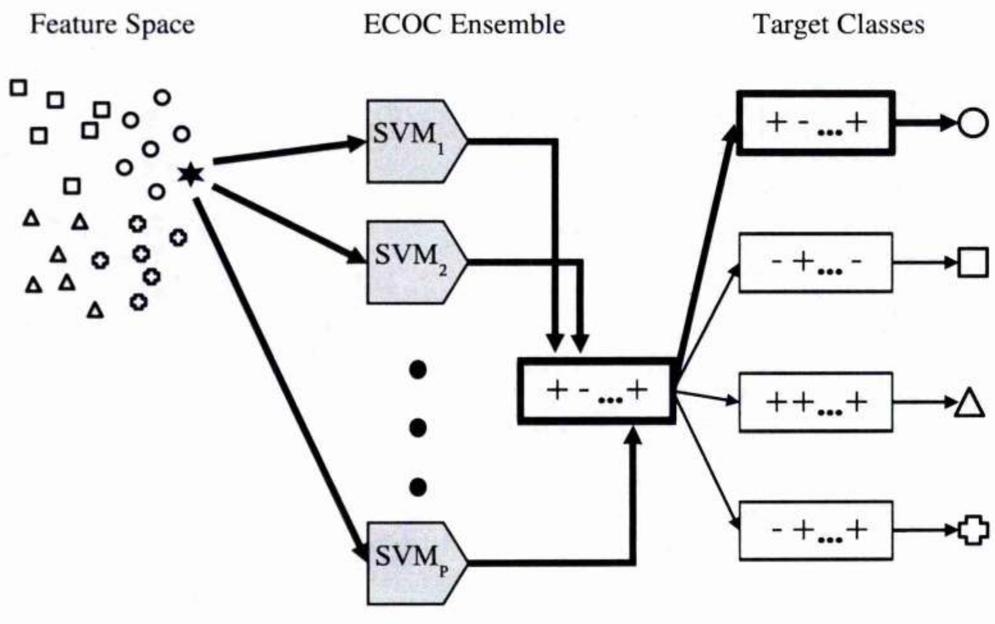

Figure 3.9: An illustration of the ECOC concept with the Hamming decoding procedure. Four classes are shown in the feature space; these are represented by circles, squares, triangles and crosses. The diagram shows a probe vector, indicated by a star, being correctly assigned to the circle class.

as to whether the probe vector $\mathbf{x}$ belongs to the positive or negative family of classes for that classifier. The class assignment $\mathcal{A}(\mathbf{x})$ is made to the class ω_i whose codeword Z_i is closest in Hamming distance to $\mathbf{y}(\mathbf{x})$. Fig. 3.9 illustrates the operation of an ECOC ensemble using SVM base classifiers and the Hamming decoding method.

The Hamming method is computationally efficient but it depends for its success on the majority of base classifiers making accurate individual decisions; it takes no account of different levels of confidence which these decisions may warrant. Other methods, such as using Manhattan distances between $\mathbf{y}(\mathbf{x})$ and $\{Z_i\}$, or probabilistic methods, have been proposed [72] to overcome these drawbacks.

An alternative to using the class codeword as an idealised class template is to compare $\mathbf{y}(\mathbf{x})$ with the actual outputs $\{\mathbf{y}(\mathbf{x}_j) \mid t_j = \omega_i\}$ obtained on the gallery set for the i'th class [37]. In this approach the ECOC ensemble is viewed as performing a transformation from one feature space into a new feature space, possibly of higher dimension, in which there is better separation between classes; this concept is illustrated in Fig. 3.10. Within this new feature space any classification technique, such as nearest neighbour or using the smallest average distance to class samples, can be used.

One disadvantage of the the second decoding method is that it requires the storage of the base classifier outputs for each member of the gallery set and hence can be computationally demanding. Against this, however, it does not assume anything about individual base classifier accuracy, but rather relies on base classifiers giving *consistent* soft outputs when presented with members of a given class.

Another issue to be considered is that of correllation between base classifier outputs. Ideally each base classifier would provide a statistically independent source of information about the class of a given probe vector. In practice however this ideal is unattainable and the errors made by base classifiers will tend to be correlated to some degree. The Hamming decoding method aims to use the error-correcting capability of the ensemble as a whole to counteract the effects of any such correlation between individual base classifiers. The transformational decoding method, however, does not take advantage of de-correlation caused by error-correcting behaviour and, again, rests on the assumption that base classifier outputs will be consistent for a given class, regardless of whether

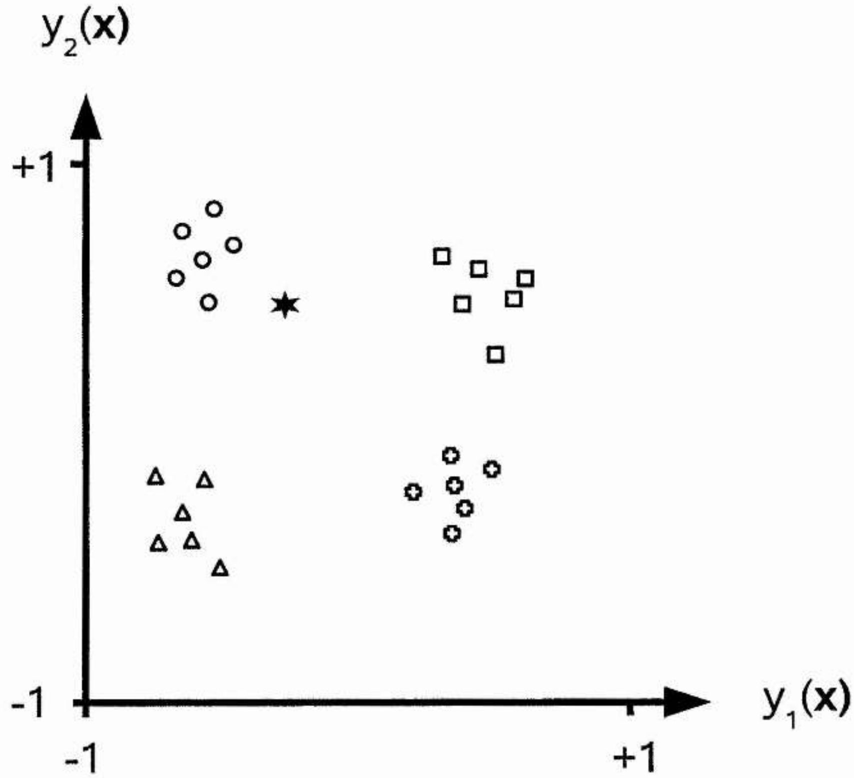

Figure 3.10: An illustration of ECOC viewed as a method for transforming the feature space. The original feature vector is x. As indicated by the star, this is transformed into a point in a two-dimensional space using two base classifiers which give rise to output values $y_1(\mathrm{x})$ and $y_2(\mathrm{x})$. Corresponding gallery set feature vectors for four classes are shown; these are represented by circles, squares, triangles and crosses.

they are correct in the Hamming sense.

3.9.2 Binary Hierarchical Classifier Ensembles

Another approach to constructing ensembles of 2-class classifiers to solve multi-class problems is that of Binary Hierarchical Classifiers (BHC). This method was first proposed as a solution to the ground cover classification problem in remotely sensed hyperspectral images [38, 39, 49]. In contrast to ECOC, in which no structure among classes is assumed, the BHC method proceeds by repeatedly partitioning families of classes into pairs of smaller sub-families where each such sub-family consists of classes which are more similar to each other than they are to classes in the other family of the pair. This process leads naturally to the construction of a binary tree where the root node contains all target classes and the leaf nodes contain just one class each. This tree is used to construct an ensemble classifier by training a 2-class base classifier for each internal node of the tree and combining the results from some or all of these classifiers to make an overall classification decision. Rajan [59] has shown that BHC can give comparable results to ECOC on problems other than satellite imaging, however these experiments were conducted on datasets with a relatively small number of classes (the maximum being 26) and features (the maximum being 64). The use of *random forests* [12] has also been proposed; in this method the results are combined from a number of trees constructed using different random subsets of features. Ham et. al. have shown [26] that this leads to an improvement in accuracy on hyper spectral data classification.

Fig. 3.11 shows a simple example of a BHC tree with just six classes, labelled A-F. Note that the tree will not, in general, be balanced and also that, for reference purposes, each node is assigned a unique identification number. For unbalanced trees, these identification numbers do not necessarily form a dense set. Expressed in binary and read from left to right (ignoring the initial 1 bit), a node identification number indicates the path down from the root node, with 0 signifying that a left branch is taken and 1 a right branch.

The code matrix for a BHC decomposition consists of C rows, since there are C classes in total, and $C-1$ columns representing the number of internal nodes of the tree.

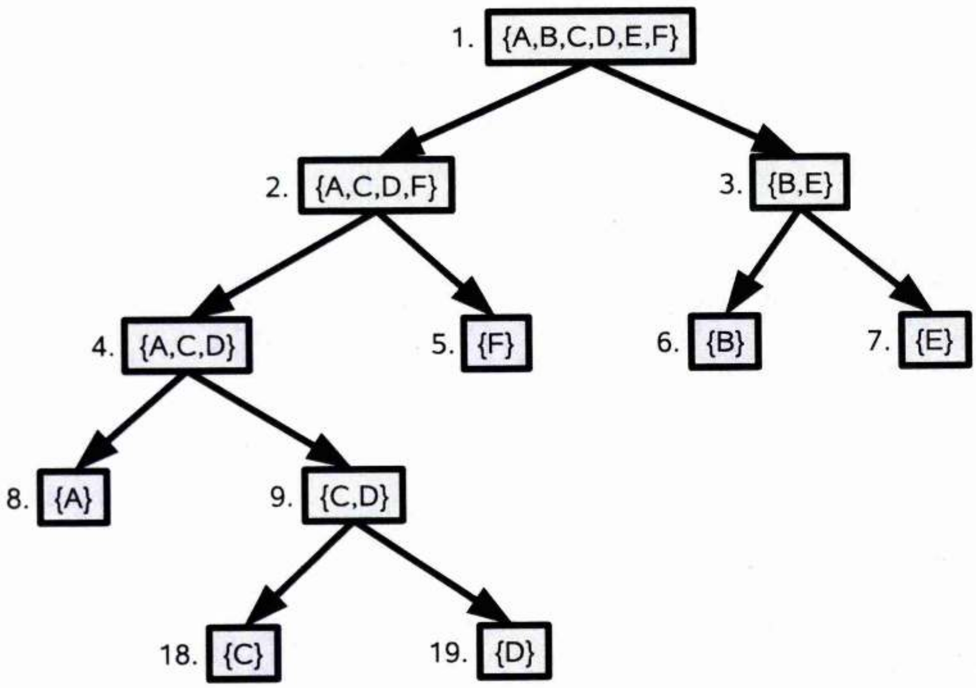

Figure 3.11: The BHC concept. Each node represents a set of classes. Numbers to the left of the nodes are their unique identification numbers.

	Classifier Id.				
Class	1	2	3	4	9
A	-1	-1	0	-1	0
B	+1	0	-1	0	0
C	-1	-1	0	+1	-1
D	-1	-1	0	+1	+1
E	+1	0	+1	0	0
F	-1	+1	0	0	0

Table 3.2: The code matrix for the example BHC tree shown in Fig. 3.11

Each column contains 0-valued entries for classes which are not on the path to the corresponding internal node, -1 for members of the negative sub-family for that node and $+1$ for members of the positive sub-family. An example code matrix for the BHC tree of Fig. 3.11 is shown in Table 3.2.

It is important to the success of the BHC method that, as far as possible, the two families of classes represented by each node be easily distinguishable from one another. This is a problem in 2-class clustering and, for this purpose, an algorithm based on deterministic annealing [61] is proposed in [39]. In section 7.2 we examine the performance of this algorithm on face data and compare it with the widely used 2-means algorithm [80] (i.e. the k-means algorithm with k, the target number of clusters, being set to 2). Many other clustering algorithms have been described in the literature [80], for example the neural gas algorithm [58] has proved successful in several applications.

The aim of the clustering algorithm described in [39] is to find a partitioning of Ω_j, the family of classes represented at the j'th node, into left and right sub-families Ω_j^- and Ω_j^+ that maximises the Fisher discriminant $J_F(\hat{\mathbf{u}})$ of Eqns. 3.11 and 3.16. It thus aims to achieve the twin purposes of finding a partition of Ω_j that gives good separation between the two sub-families of classes whilst simultaneously finding a projection vector $\hat{\mathbf{u}}$ that can be used to construct a base classifier. Since clustering is inherently an NP complete problem the method uses deterministic annealing [61] to find an approximate solution. In contrast to the 2-means approach, the algorithm does not initially make definite assignments of each class ω_i to either Ω_j^- or Ω_j^+, but rather associates each ω_i with *both* Ω_j^- and Ω_j^+ by means of the probability values $\Pr\left(\Omega_j^- \mid \omega_i\right)$ and $\Pr\left(\Omega_j^+ \mid \omega_i\right)$ (which sum to 1). Initially the value of just one of these probabilities is set to 1, thus making a definite assignment for one ω_i, whilst all the remaining probabilities are set to 0.5. As the algorithm proceeds, these probabilities diverge from their initial settings and are gradually hardened, as a temperature parameter is reduced, until definite assignments have been made for all ω_i.

Once clustering has been performed for the j'th node it is necessary to create a corresponding base classifier. The method proposed in [39] is to project all feature vectors on to the Fisher vector $\hat{\mathbf{u}}$ and to use a Gaussian mixture model in this one-dimensional space. Evidence presented in [59], however, shows that, for data other than hyperspectral images, greater accuracy can be achieved by training an SVM in the general feature space.

Two strategies are described in [39] for decoding the outputs $\mathbf{y}(\mathbf{x})$ from a BHC ensemble. In the BHC-hard decoding method the tree is descended, starting at the root node, and the j'th base classifier makes a definite decision as to which of the two families Ω_j^- or Ω_j^+ the vector $\mathbf{x}$ belongs. Depending on this decision, the left or right sub-node (labelled as node $2j$ or $2j+1$) respectively is examined next. This process is repeated until a terminal node $\mathcal{N}_k$ is reached; the probe vector is then assigned to the class $\mathcal{C}(\mathcal{N}_k)$ represented by that node. The main disadvantage of this method is that base classifier errors tend to be magnified because a classification error at any stage will lead to an overall misclassification (contrast this with the error-correcting approach of ECOC which aims to recover from a such errors). Unless the base classifiers are very accurate, therefore, the BHC-hard algorithm may lead to a large ensemble error.

In the BHC-soft decoding method each base classifier is assumed to be capable of generating posterior probabilities $\Pr\left(\mathcal{C}(\mathbf{x}) \in \Omega_j^-\right)$ and $\Pr\left(\mathcal{C}(\mathbf{x}) \in \Omega_j^+\right)$. For each leaf node $\mathcal{N}_k$ these probabilities are multiplied together, on the path from the root node to $\mathcal{N}_k$ to obtain a posterior probability that $\mathcal{C}(\mathbf{x})$ is equal to $\mathcal{C}(\mathcal{N}_k)$, the class represented by the leaf node. The probe vector is then assigned to the class with the largest posterior probability. Since this algorithm uses soft outputs it is more forgiving of base classifier errors than BHC-Hard, however ensemble accuracy still depends on the ability of the base classifiers to generate good probability estimates.

Note that for face verification, where it is only necessary to compare a probe image with a single claimed identity, a score can be computed by following the path from the root node to a single leaf node. For both decoding rules this is a computationally efficient $\mathcal{O}(\log C)$ process as it involves evaluating, on average, only $\log_2 C$ classifiers. For rank 1 identification problems the BHC-Hard decoding rule still has $\mathcal{O}(\log C)$ complexity, however the BHC-Soft decoding rule has $\mathcal{O}(C)$ complexity since it requires all $C - 1$ base classifiers to be evaluated so that the class with the highest probability can be found.

3.10 Summary

An essential first step in most face recognition systems is to register the images so that the facial features occur at the same geometrical location in all cases. In this thesis we assume the availability of manual landmarks for this purpose. When dealing with 3D face scans, we also adopt the method of dense non-rigid 3D face registration using a morphable model to obtain a standard representation of the source data.

A variety of techniques have been explored for carrying out illumination normalisation on 2D images and here we make use of homomorphic filtering and histogram equalisation.

PCA is an unsupervised technique which finds an orthonormal basis of eigenfaces, with the basis eigenfaces being arranged in decreasing order of training set scatter. Dimensionality reduction is accomplished, in a way which preserves most of the variance of the data distribution, by projecting each face image vector onto this basis and then discarding the later components which make only a small contribution to this variance. The exact set of components to discard is determined by requiring the mean square error to be below a pre-defined threshold.

LDA is a supervised algorithm which finds a discriminative (non-orthogonal) basis of Fisherfaces by maximising the ratio of between-class scatter to within-class scatter.

LDA is often preceded by a PCA stage in order to reduce the dimensionality of the input vectors and thus avoid singular within-class scatter matrices.

LBP is a powerful 2D and 2.5D texture description algorithm which has proved to be effective in face recognition applications. The approach used in this thesis is to concatenate LBP histograms at several scales from the different regions in a non-overlapping rectangular tiling of the image; LDA is then applied to reduce the dimensionality and improve class separation.

SVMs are a principled way of constructing classifiers for 2-class problems. The method consists of finding a hyperplane that maximises the margin of separation between the two classes. Non-linear decision surfaces can be accommodated by using a kernel function, such as the Gaussian kernel, to project the data into a higher dimensional feature space in which an appropriate linear decision boundary can be found. The method of nested uniform design is an efficient search strategy for finding near-optimal sets of SVM parameters. Gaussian mixture modelling, Platt's sigmoid fitting algorithm and isotonic regression are three ways in which SVM outputs can be calibrated so that they are linearly related to class membership probabilities.

Multi-class face classification problems can be handled by constructing ensembles of 2-class classifiers and then applying a suitable decoding procedure to the outputs from the ensemble so as to reach a classification decision. Two possible architectures for such ensembles are ECOC and BHC. The first of these consists of repeatedly partitioning the set of all face identities into two families of approximately equal size; in the second method the set of all face identities is recursively subdivided to construct a binary classification tree. Decoding of ECOC outputs may be done either by measuring the distance from the target codeword for a given class or by adopting a nearest neighbour approach based on the outputs obtained on the gallery set. BHC decoding may be accomplished by making a definite decision at each node until a leaf node is reached; alternatively, the product of the soft outputs from the root node to a given leaf node may be used as a similarity score for the class represented by that leaf node.

Chapter 4

Angularisation Methods

In this thesis we are primarily concerned with data sets for which angular separation is more discriminative than other metrics such as Euclidean separation. It has been noted in section 2.4, and is further shown by the experimental results of chapter 6, that a collection of mean-subtracted face images is an example of such a data set. A central theme of this thesis is that classifier performance can be improved by non-linearly mapping a data set of this kind into a new feature space in which the Euclidean or other metrics are placed on a par with angular separation. The generic name we give to this process is *angularisation* and in this section we propose two SVM kernel functions which incorporate angularisation into their construction. We begin by defining two general transformations which may be applied to any feature space and then show the Gaussian SVM kernel may be adapted to make use of these transformations without an explicit feature space transformation stage.

4.1 Explicit Implementations of Angularisation

The first proposed implementation of an explicit angularisation transformation consists of applying the vector transformation $ang : \mathbb{R}^M \mapsto \mathbb{R}^M$ which is defined as

$$ang\,(\mathbf{x}) = \left[90 - \frac{180}{\pi} \arccos \frac{x_i}{\|\mathbf{x}\|} \right]^{\mathrm{T}}, i = 1...M \tag{4.1}$$

Thus, if $\mathbf{y} = ang(\mathbf{x})$ then $y_i \in [-90, 90]$ is equal to $90 - \theta_i$ where $\theta_i \in [0, 180]$ is the principal angle in degrees between $\mathbf{x}$ and the ith coordinate axis of the original feature space. Note that $90 - \theta_i$ is used, rather than θ_i itself, in order to avoid introducing an unnecessary change of sign, or displacement of the data components. The $ang(\cdot)$ transformation collapses any ray emanating from the origin down to a single point; it preserves the angular separation between two vectors but not the Euclidean distance. It has the desirable property that points which are close together in the angular sense are mapped to points which are close together in the Euclidean sense and vice versa.

One disadvantage of $ang(\cdot)$, from an analytical point of view, is that a simple equation cannot express the precise relationship between the original angular separation of two vectors and their corresponding Euclidean separation after transformation. For this reason we also investigate an alternative method of achieving the same objective, which is to map feature vectors to the unit sphere by rescaling them to unit length. Accordingly, we introduce the vector transformation $sph : \mathbb{R}^M \mapsto \mathbb{R}^M$ where

$$sph(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{4.2}$$

This transformation has similar properties to $ang(\cdot)$ and also has the benefit that we can derive the equation

$$\|sph(\mathbf{x}) - sph(\mathbf{y})\| = D_{Euc}(sph(\mathbf{x}), sph(\mathbf{y})) = 2\sin\frac{1}{2}\theta(\mathbf{x}, \mathbf{y}) \tag{4.3}$$

which relates the angle between two feature vectors to their transformed Euclidean separation. The value of $D_{Euc}(sph(\mathbf{x}), sph(\mathbf{y}))$lies in the range $[0, 2]$; it takes the value 0 for parallel vectors and 2 for anti-parallel ones. By making use of the trigonometric identity $\cos\theta = 1 - 2\sin^2\frac{\theta}{2}$ it can also be shown that this distance is closely related to the cosine distance $D_{cos}(\cdot, \cdot)$ of Eqn. 2.8 as follows[1]:

$$D_{Euc}(sph(\mathbf{x}), sph(\mathbf{y})) = \sqrt{2D_{cos}(\mathbf{x}, \mathbf{y})}. \tag{4.4}$$

[1]This equation may appear to be suspect from a dimensional analysis point of view. The right hand side is dimensionless but the left hand side is a distance measurement. On closer inspection, however, the left hand side can also be seen to be dimensionless due to the presence of the $sph(\cdot)$ function. This means that any change of distance measurement units would leave the right hand side numerically unchanged.

4.2 Angularised SVM Kernels

The transformations $ang(\cdot)$ and $sph(\cdot)$ are applied as a separate stage of processing, independently of any particular classification technique. Our main objective in this research, however, is to investigate ways by which SVM classifiers can make use of angularisation implicitly without the need for a separate feature space transformation. We approach this problem by looking at how the Euclidean distance measure that is used in the formulation of the Gaussian kernel (see Eqn. 3.33) can be replaced by angle-based measures.

The first proposal for angularising the Gaussian kernel is modelled on the $ang(\cdot)$ transformation and it makes use of the $\arccos(\cdot,\cdot)^2$ function to measure the angular separation between vectors. The kernel function is defined as

$$\begin{aligned} K_{ang}(\mathbf{x},\mathbf{y}) &= \exp\left(-\tfrac{1}{2\sigma^2}\arccos\tfrac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right) \\ &= \exp\left(-\tfrac{1}{2\sigma^2}D_{ang}(\mathbf{x},\mathbf{y})\right). \end{aligned} \tag{4.5}$$

Note that, since $K_{ang}(\cdot,\cdot)$ is able to directly incorporate the angular separation between pairs of feature vectors without the need for a separate calculation of Euclidean distance, the effect of this approach is mathematically different to that obtained by applying the $ang(\cdot)$ transformation followed by a Gaussian kernel SVM.

The second proposal is based on rescaling feature vectors to unit length and the kernel function is defined as

$$\begin{aligned} K_{sph}(\mathbf{x},\mathbf{y}) &= \exp\left(-\tfrac{1}{2\sigma^2}\left\|\tfrac{\mathbf{x}}{\|\mathbf{x}\|}-\tfrac{\mathbf{y}}{\|\mathbf{y}\|}\right\|^2\right) \\ &= \exp\left(-\tfrac{1}{2\sigma_1^2}D_{cos}(\mathbf{x},\mathbf{y})\right) \end{aligned} \tag{4.6}$$

where $\sigma_1 = \sigma/\sqrt{2}$ is an undetermined tuning parameter and the second formulation follows from Eqn. 4.4. In effect this method incorporates the $sph(\cdot)$ function directly into the kernel and is, therefore, mathematically equivalent to applying the $sph(\cdot)$ transformation followed by an SVM with a Gaussian kernel.

[2] To be completely consistent with the Gaussian kernel the definition of the angular kernel would use $\arccos^2$ in its formulation rather than arccos. In section 6.3 we refer to this as the $K_{ang2}(\cdot,\cdot)$ kernel and show that the proposed $K_{ang}(\cdot,\cdot)$ definition yields slightly better performance, particularly on face identification problems.

4.3 Mercer Properties of SVM kernels

As noted in section 3.6, a desirable property of SVM kernels is that they satisfy the Mercer condition. This ensures that there exists a unique global optimum to which the SVM algorithm converges for all training sets. In this section we show that the two kernels $K_{ang}(\cdot,\cdot)$ and $K_{sph}(\cdot,\cdot)$ do possess the Mercer property.

First we state without proof some standard properties of Mercer kernels [68]. Let $K_1 : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ be a Mercer kernel, $p : \mathbb{R} \mapsto \mathbb{R}$ be a polynomial (or convergent infinite series) with positive coefficients, $\varphi : \mathbb{R}^M \mapsto \mathbb{R}^M$ and $a \in \mathbb{R}^+$. Then the following functions are all Mercer kernels:

1. $K(\mathbf{x},\mathbf{y}) = aK_1(\mathbf{x},\mathbf{y})$.
2. $K(\mathbf{x},\mathbf{y}) = K_1(\phi(\mathbf{x}),\phi(\mathbf{y}))$.
3. $K(\mathbf{x},\mathbf{y}) = p(K_1(\mathbf{x},\mathbf{y}))$.
4. $K(\mathbf{x},\mathbf{y}) = \exp(K_1(\mathbf{x},\mathbf{y}))$
5. $K_{Gauss}(\mathbf{x},\mathbf{y})$ of Eqn. 3.33.
6. $K(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$, i.e. the inner product of the two vectors is a Mercer kernel.

To prove that the kernel $K_{ang}(\cdot,\cdot)$ is a Mercer kernel we first note that the Taylor series expansion for the $\arcsin(\cdot)$ function

$$\arcsin\,\alpha = \alpha + \frac{1}{2}\frac{\alpha^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{\alpha^5}{5} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{\alpha^7}{7} + \dots$$

has only positive coefficients. It follows from rule 3 that, for any Mercer kernel $K_1(\mathbf{x},\mathbf{y})$, the function $\arcsin(K_1(\mathbf{x},\mathbf{y}))$ is also a Mercer kernel. By making use of the trigonometric identity

$$\arccos\,\alpha = \frac{\pi}{2} - \arcsin\,\alpha$$

it can be seen from Eqn. 4.5 that

$$K_{ang}(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\pi}{4\sigma^2}\right)\exp\left(\frac{1}{2\sigma^2}\arcsin\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}\right).$$

That this is a Mercer kernel follows from rules 1 and 4, noting that $\exp\left(-\frac{\pi}{4\sigma^2}\right) > 0$ and also that $\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$ is a Mercer kernel by rules 2 and 6 with $\phi(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$.

The result that $K_{sph}(\cdot,\cdot)$ of Eqn. 4.6 is a Mercer kernel follows from rules 5 and 2 with $\phi(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$.

4.4 Summary

It has been noted in section 2.4 that angular separation has been found to be a useful measure of dissimilarity in face recognition applications. In this chapter two angularised SVM kernel functions, named $K_{ang}(\cdot,\cdot)$ and $K_{sph}(\cdot,\cdot)$, have been defined that are based on adaptations of the Gaussian kernel. They replace the Euclidean distance measure, in the formulation of the Gaussian kernel, by alternative measures that are based on angular separation. It has been shown that both $K_{ang}(\cdot,\cdot)$ and $K_{sph}(\cdot,\cdot)$ are Mercer kernels.

In addition to these kernel functions, we have also presented two general methods, named $ang(\cdot)$ and $sph(\cdot)$, for mapping a general vector space non-linearly into a new space in such a way that angular separation between feature vectors in the original space is related to Euclidean separation in the new space. In computing $ang(\mathbf{x})$, each coordinate x_i is mapped to a new value determined by the angle between $\mathbf{x}$ and the i'th coordinate axis. The value of $sph(\mathbf{x})$ is simply the unit vector in the direction of $\mathbf{x}$. Use of these transformations allows any classification technique to be applied without it being explicitly based on angular separation.

In chapter 6, experimental results are presented that show that the application of angularisation to face-recognition problems can lead to improvements in classifier performance.

Chapter 5

Experimental Data

Chapters 6 to 8 present the results of experiments carried out using the face recognition techniques described in chapters 3 and 4. In this chapter we describe the data sets and protocols that were used in these experiments. With the exception of the synthetic data described below, these data sets are all publicly available. Any 2D colour images are converted to greyscale before use.

5.1 List of Data Sets

XM2VTS. This database is provided by the University of Surrey [44]. It contains 8 colour images at a resolution of 720 × 576 pixels for each of 295 subjects, taken under conditions of uniform illumination and neutral expression. Some examples of images from this data set are shown in Fig. 5.1.

We use an experimental protocol known as the Lausanne protocol, Configuration 1. This divides the database into a training set of 200 clients, an evaluation set of the same 200 clients plus 25 impostors and a test set of the same 200 clients plus 70 additional impostors. For each client, 3 images are used in the training set, 3 in the evaluation set and 2 in the test set. For each impostor all 8 images are used in the evaluation or test set as applicable. Verification tests, therefore, consist of 400 valid claims (one for each test-set client image) and 112,000 invalid claims (one per client per test-set

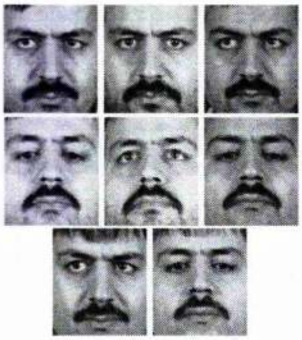

Figure 5.1: Example images from the XM2VTS corpus after conversion to greyscale, geometric normalisation and rescaling to 120 x 142 pixels. Training images are shown in the top row, evaluation images in the middle row and test images in the bottom row.

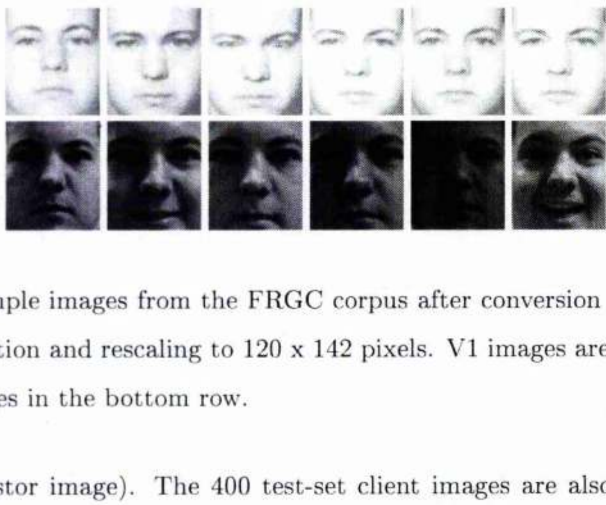

Figure 5.2: Example images from the FRGC corpus after conversion to greyscale, geometric normalisation and rescaling to 120 x 142 pixels. V1 images are shown in the top row and v2 images in the bottom row.

impostor image). The 400 test-set client images are also used to perform closed-set identification experiments.

FRGC experiment 3. The FRGC protocol [54] defines 6 experiments which are designed to test various face recognition scenarios. Of these, experiment 3 is concerned with the combination of 2D and 3D information. It is divided into experiments 3t, 3s and 3; these use, respectively, the texture channel (i.e. 2D images), the shape channel (i.e. 2.5D range images or 3D scans) and the fusion of both.

The supplied database is divided into two parts. The version 1 (v1) data consists of up to 8 colour images and scans of 270 subjects taken under uni-

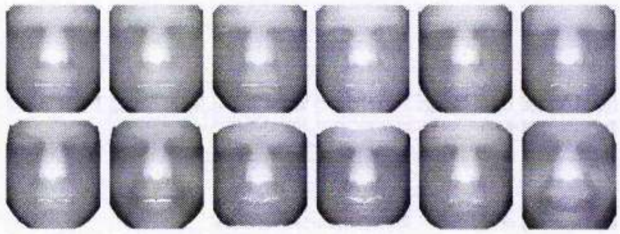

Figure 5.3: Example range images from the FRGC corpus after cropping to 120 x 142 pixels. V1 images are shown in the top row and v2 images in the bottom row.

form illumination in frontal pose and with neutral expressions. The version 2 (v2) data consists of up to 22 images and scans of each of 410 subjects taken under adverse illumination conditions in near frontal pose but with non-neutral expressions. The v2 images are split approximately evenly between the 270 v1 subjects and 140 new ones. The images and scans were captured almost simultaneously using a Minolta Vivid 900 range scanner at a resolution of 640×480. Some examples of 2D images are shown in Fig. 5.2 and examples of 2.5D range images are shown in Fig. 5.3.

A standard experimental protocol is defined in [54] whereby the FRGC v1 data is used only for training and tuning a face recognition algorithm; once training has been performed the images are discarded and are not used as gallery templates in the subsequent tests. Instead, verification performance is evaluated by comparing subsets of the v2 data against each other. These subsets are labelled I, II and III and represent increasingly difficult face verification problems due to the increasing time lapse between probe and gallery images. Because of the difference in expressions and lighting conditions between the v1 and v2 data, the FRGC protocol is designed primarily as a test of algorithms that correct for illumination and expression variation. A baseline algorithm is supplied as part of the FRGC Basic Experimentation Environment (BEE); this uses PCA feature extraction with a whitened cosine similarity measure. A brief summary of the results obtained by the participants in the 2005 FRGC competition is given in [55]

FRGCv2 2D images. Because this thesis is not targeted at investigating methods of illumination and expression correction, the full FRGC 2D and 3D data sets and protocols described above do not constitute an ideal test database. Instead, we wish to focus on the performance of classification algorithms and thus need to utilise a database in which the training set is representative of the same range of conditions as the test data. For this reason we define a new protocol in which the v1 data is discarded and the uncontrolled FRGCv2 images are divided randomly into five subsets in a manner similar to the Lausanne protocol for XM2VTS. These subsets consist of approximately 800 records each for the training client set, the evaluation client set, the evaluation impostor set, the test client set and the test impostor set. This leads to a set of 236 client identities, 115 evaluation impostor identities and 115 test impostor identities. Verification tests consist of 801 valid claims and 189,508 invalid claims. Identification tests are carried out using the 801 test client records. Full details of the protocol are given in appendix A.

FRGCv2 2.5D range images. To perform classification experiments using range images we similarly subdivide the FRGCv2 data. These scans are in one to one correspondence with the FRGCv2 2D images described in the previous paragraph and the same test protocol is used.

FRGCv2 3D scans. These are full 3D scans which again are in one to one correspondence with the FRGCv2 2D images and we again adopt the same test protocol. The use of identically structured FRGCv2 2D, 2.5D and 3D data sets allows comparisons to be made between the results of a face recognition algorithm for each of these modalities.

JAFFE 2D images. The JAFFE [43] database is utilised in order to ensure that the algorithms described in the thesis are tested on a wide variety of ethnic groups, including oriental faces. JAFFE contains 213 greyscale images of 10 Japanese female subjects. These images were captured under uniform illumination conditions but with seven different facial expressions per subject; these being happiness, sadness, surprise, anger, disgust, fear and neutral. Fig. 5.4 shows an example of each of these expressions.

Figure 5.4: Example images from the JAFFE database after geometric normalisation and rescaling to 120 x 142 pixels.. From left to right these show angry, disgusted, fearful, happy, sad, surprised and neutral expressions.

Figure 5.5: An example image from the synthetic data set. These images are used in experiments that contrast the behaviour of classification algorithms on face and non-face data.

A test protocol similar to that used for the FRGCv2 data set is employed. Accordingly, the data is partitioned into 57 training client images, 57 validation client images, 21 validation impostor images, 56 test client images and 22 test impostor images. This produces a set of 8 client identities, leaving 1 subject each for the evaluation and test impostor identities. Verification tests thus consist of 56 valid claims and 176 invalid claims. Identification tests make use of the 56 test client images. Full details of the protocol are given in appendix A.

Synthetic data. In order to be able to compare the behaviour of classification algorithms on both face and non-face data, a further synthetic image set is utilised in some of the experiments. This data set is modelled on the XM2VTS Configuration 1 Lausanne protocol in that it has the same number and arrangement of images. It differs, however, in that each image is an artificially generated noise pattern. Considered as a 2805 dimensional vector, the images belonging to each class are generated in such a way as to be normally distributed about the class centre, with the class centres being uniformly distributed through the available greyscale space. An example of a such a synthetic image is shown in Fig. 5.5.

Figure 5.6: Example images from the UIUC vehicle database. The top row shows non-car data and the bottom row shows car data.

UIUC vehicle images. Although the synthetic images described in the previous paragraph constitute an example of non-face data, it is also worthwhile to examine the behaviour of classification algorithms on real world images which, though they do not represent faces, nevertheless posses some structure and regularity in a similar way to face images. For this purpose we make use of the UIUC database of vehicle images [1]. This database was designed with multi-scale object detection algorithms in mind but here we are interested in the classification task and thus make use of just the training set of images; this results in an image database containing 550 car images, of approximately equal scale, and 500 non-car images. Some examples are shown in Fig. 5.6.

The experiments performed on this data set are aimed at solving the 2-class problem of determining whether a probe image does or does not contain a car. For this purpose we use a protocol that randomly partitions the data into training, evaluation and test sets of 350 images each, with approximately equal numbers of car and non-car images in each set. The concept of impostor images is not applicable to this data set. Full details of the protocol are given in appendix A.

5.2 Summary

The XM2VTS database provides a large set of 2D images captured under controlled illumination conditions with neutral expressions.

The FRGC Experiment 3 corpus is a large set of face image data which was simultaneously captured using 2D, 2.5D and 3D modalities. The original purpose of this data was to provide a testbed for illumination and expression correction algorithms. Such algorithms are not the main focus of this work, however, and for most of the experiments described in subsequent chapters we make use of a subset of the FRGC data. This subset is referred to as the version 2 data and consists of a large set of images and scans that were captured under uncontrolled illumination conditions and with varying facial expressions.

JAFFE is a fairly small database of 2D images of Japanese female faces which shows a fixed set of emotional expressions for each subject. It is included in order to test the algorithms described herein on oriental faces.

In addition to face images we also perform some experiments using randomly generated synthetic data and the UIUC vehicle image database. In the latter case a subset of the images are used as a source of examples of car and non-car images.

A publicly defined standard test protocol, known as the Lausanne protocol Configuration I, is adopted for XM2VTS experiments. In all other cases we define a protocol for the experiments which is modelled on the Lausanne protocol. In most cases this consists of subdividing the available images into approximately equally sized sets of training clients, evaluation clients, evaluation impostors, test clients and test impostors. The exception is the UIUC vehicle image database in which the concept of impostors does not apply so we subdivide the images into training clients, evaluation clients and test clients only.

Chapter 6

Angularisation Experiments

In this chapter we investigate experimentally the benefits of applying the process of angularisation that was introduced in chapter 4. Section 6.1 begins the investigation by looking at the power of each of five metrics, used with and without angularisation, to distinguish intra-class differences from inter-class differences. This is done for different feature extraction scenarios so as to gain an understanding of the interplay between the distance metric used within a feature space and the feature extraction techniques that are used to construct it. Section 6.2 then demonstrates the effect of applying these techniques in practical face verification and identification experiments and shows how performance can be improved by the correct combination of angularised feature extraction algorithms. Finally, section 6.3 shows that these techniques are largely insensitive to the choice of algorithm by which angularisation is achieved.

In carrying out these experiments the following methods were used except where stated otherwise. Photometric normalisation of 2D images was performed using homomorphic filtering and histogram equalisation. LDA was accomplished by first using PCA to reduce the number of dimensions to a more manageable level whilst retaining 98% of the total training set variation. LDA was then applied to reduce the number of dimensions further to a value of $C-1$, where C is the number of distinct face identities in the training set; this is the largest number of dimensions that can be supported without the within-class scatter matrix becoming singular. Angularisation was applied using and the $ang(\cdot)$ function (see Eqn. 4.1) to transform to a new feature space. To apply the LBP algorithm to 2D and 2.5D images, multi-scale 8-bit uniform pattern occurrence histograms were computed over a 4×4 tiling of each image, with radii varying from 1 to 10 pixels. These were then concatenated and LDA was applied to the resulting 9440 dimensional vectors. This method is similar to that described in [16] except that LDA is performed in a single stage rather than per sub-window, thereby avoiding the need to carry out a fusion operation on multiple scores. ECOC code matrices with 510 columns were produced by using the Bose-Chaudury-Hocquenghem algorithm to generate a square matrix and then selecting the best submatrix with the required number of rows out of 1000 random trials. ECOC was then used as a means of projecting the data into a 510-dimensional feature space, within which a distance metric could be used to compare feature vectors. SVMs with Gaussian radial basis function kernels, as defined in Eqn. 3.33, were used to implement the ECOC base classifiers. The method of nested uniform design, with 13 then 9 sample points, was used to optimise the parameters of each SVM on the appropriate evaluation set data. Isotonic regression by pair-adjacent violators was used to calibrate the SVM output scores. Verification and identification decisions were made by taking the mean distance between a probe vector in the final feature space and each member of the training set for the given client.

For 3D experiments the scans were first registered using the supplied manual landmarks and placed into dense correspondence with a morphable model as described in section 3.1. For each scan, the (x, y, z) coordinates of the 2762 vertices were concatenated to form an 8286 dimensional feature vector. LDA feature extraction, angularisation and the construction of ECOC ensembles was then carried out as described above.

6.1 Separation Performance

This section presents the results of an investigation into the effectiveness of five different distance measures in separating intra-class variations from inter-class variations. We refer to this as the *separation performance* of the metric and begin in section 6.1.1 by defining the means by which it is measured.

Four of the metrics used, namely Euclidean, Manhattan, Mahalanobis and angular distance, have been already been defined in Eqns. 2.1, 2.2, 2.3 and 2.7 respectively. To these we add the *radial distance* pseudometric:

$$D_{rad}(\mathbf{x}, \mathbf{y}) = |\|\mathbf{x}\| - \|\mathbf{y}\||, \tag{6.1}$$

which measures the difference in magnitude between feature vectors. Although radial distance is not expected to be as discriminative as the other measures, it is nevertheless instructive to investigate its properties in order to gain an insight into the relative importance of the magnitude of feature vectors versus their angular separation.

6.1.1 Measurement of Separation Performance

In order to isolate the performance of a distance measure from other factors, we require a simple numerical statistic that is independent of the details of any particular classifier design or test protocol. For this purpose we use an empirical estimate of the single-comparison Bayes error B_σ, for $\sigma \in \{Euc, Man, Mah, ang, rad\}$, that results when each member of the test set $\mathcal{Q}$ (including both clients and impostors) is compared with each member of the training set $\mathcal{T}$ using distance measure $D_\sigma(\cdot,\cdot)$. The comparisons are categorised, based on the known class identities, as either intra- or inter-class. The Bayes error statistic lies in the range 0 to 0.5 and represents the degree of overlap between the posterior distributions for the two categories. It can thus be regarded as a measure of the ability of the distance measure to discriminate between natural variations within known classes and those caused by impostor attacks.

To calculate the empirical Bayes error the set of feature vector pairs $\mathcal{T} \times \mathcal{Q}$ is partitioned into intra- and inter-class subsets and, for each subset, a frequency histogram of the distances between feature vectors is constructed. The required value is then given by

$$B_\sigma = \frac{1}{2}\sum_{i=1}^{P} \min(W_i, V_i) \tag{6.2}$$

where P is the total number of histogram bins (here fixed at 20) and W_i, V_i are respectively the fraction of the intra-class samples and inter-class samples which fall into the i'th histogram bin. Note that $\sum W_i = \sum V_i = 1$ and that we impose equal prior probabilities for the two distributions.

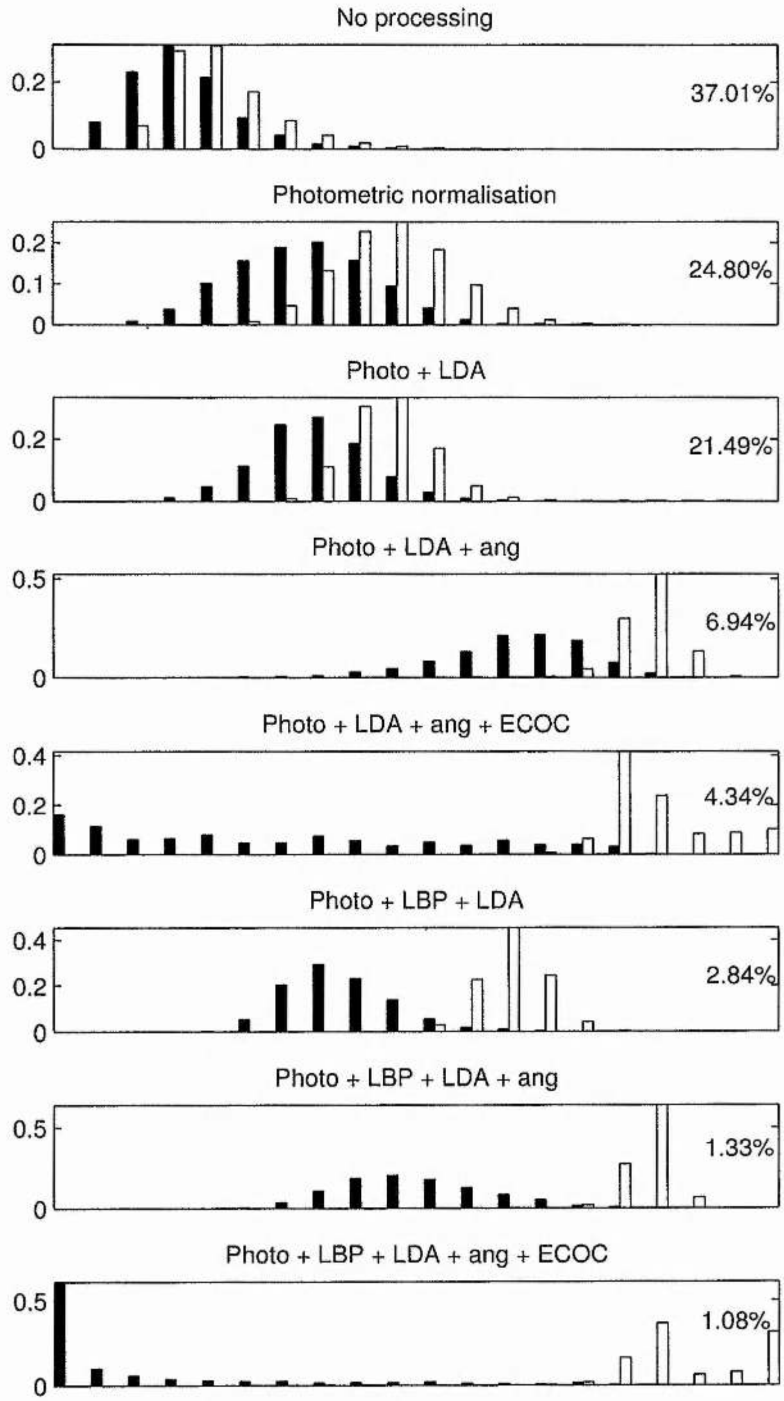

Figure 6.1: Example distributions of Euclidean distances between greyscale images under different feature extraction techniques. The intra-class distribution is shown in black and the inter-class distribution in white. Data is taken from the FRGCv2 2D corpus. Percentage Bayes errors are also shown.

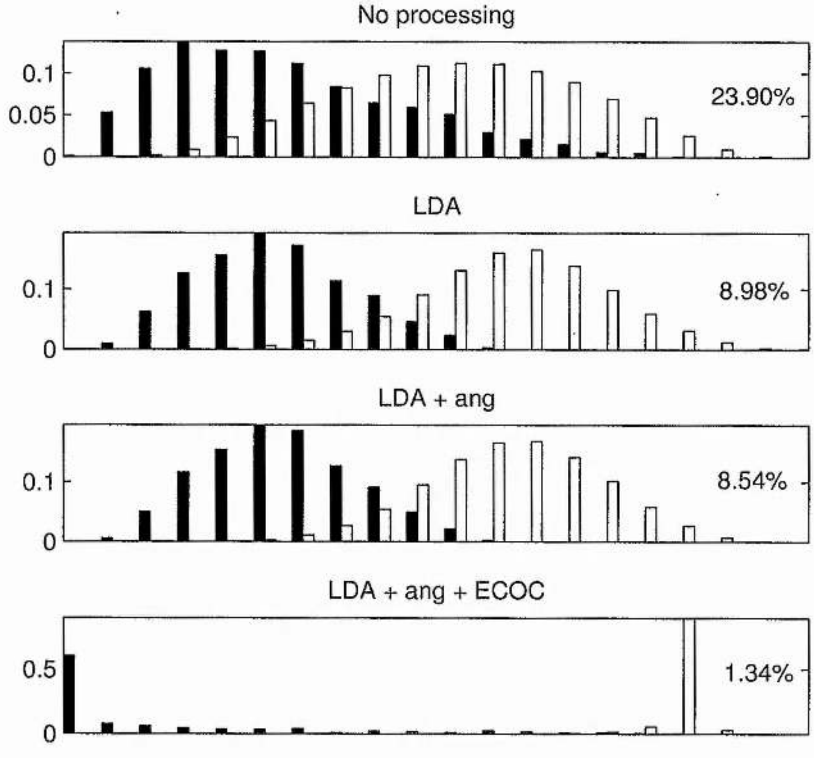

Figure 6.2: Example distributions of angular distances between 3D scans under different feature extraction techniques. The intra-class distribution is shown in black and the inter-class distribution in white. The data is taken from the FRGCv2 3D corpus. Percentage Bayes errors are also shown.

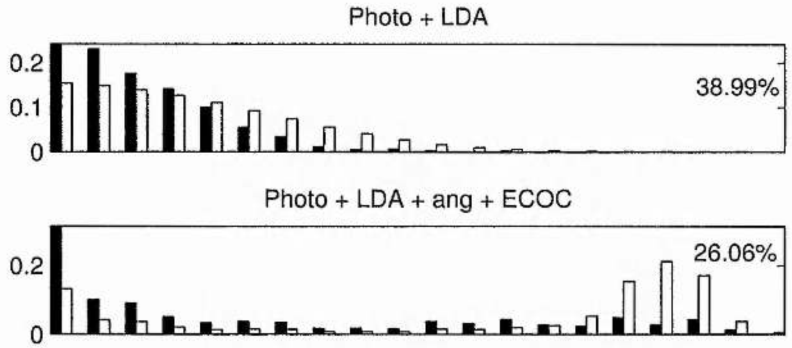

Figure 6.3: Example distributions of radial distances between images under different feature extraction techniques. The intra-class distribution is shown in black and the inter-class distribution in white. Distributions for other processing methods are similar to those for LDA and are not shown. Only ECOC leads to a qualitatively different pattern, as illustrated by the bottom graph. The data is taken from the XM2VTS corpus. Percentage Bayes errors are also shown.

Fig. 6.1 shows some example Euclidean distributions for 2D data and Fig. 6.2 shows some example angular distributions obtained on 3D data. These two figures illustrate the general pattern of the intra- and inter-class distributions which typify the distance metrics $D_{Euc}(\cdot,\cdot)$, $D_{Man}(\cdot,\cdot)$, $D_{Mah}(\cdot,\cdot)$ and $D_{ang}(\cdot,\cdot)$. The distributions fall naturally into two families. The non-ECOC distributions can be seen to be uni-modal and approximately Gaussian in shape. The effect of applying ECOC is to polarise the differences between intra- and inter-class so that the distributions, particularly those for intra-class, become much more skewed. In both cases the effect of the techniques described in this thesis is generally to increase the separation and reduce the overlap between intra- and inter-class distributions.

As shown in Fig. 6.3, the typical pattern for $D_{rad}(\cdot,\cdot)$ is somewhat different. The non-ECOC distributions have a common mode of zero, with the intra-class distribution being distinguished from the inter-class distribution only by the fact that the former is concentrated over a narrower region, with a higher modal value but lower variance. When ECOC is applied, the distributions do become better separated, however a significant proportion of inter-class scores still falls at or near to zero.

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
XM2VTS 2D							
None	20	19		10	33	17-108	33-149
	10	10		10	27	17-109	31-147
Photometric	13	11		8	31	20-101	35-127
	9	8		9	29	23-104	33-126
Photo + LDA	11	12	14	3.00	39	25-95	64-108
	2.62	3.68	6	2.99	35	25-95	64-108
Photo + LDA + ECOC	2.63	2.78		2.40	33	0-27	15-32
	2.18	2.08		1.98	26	0-91	36-101
Photo + LBP + LDA	5	5	8	2.33	39	29-93	65-107
	2.47	2.82	3.58	2.33	46	29-92	62-107
Photo + LBP + LDA + ECOC	1.42	1.44		1.44	26	0-87	33-103
	1.35	1.41		1.46	24	0-89	32-100

Table 6.1: Bayes errors and angle range statistics for the XM2VTS 2D face images. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

6.1.2 Separation Performance Experiments

Tables 6.1 to 6.5 show a list of empirical Bayes error estimates B_σ that were obtained on 2D, 2.5D and 3D face data bases when a range of feature extraction and enhancement techniques were applied. Fig. 6.4 also shows the same information in a more graphical form, however some of the effects are quite subtle and are not always obvious from the graphical information alone.

6.1.2.1 Angular and radial separation of face images

A number of observations can be made about the statistics shown in Fig. 6.4 and tables 6.1 to 6.5. Firstly, we note that within the context of face-recognition using

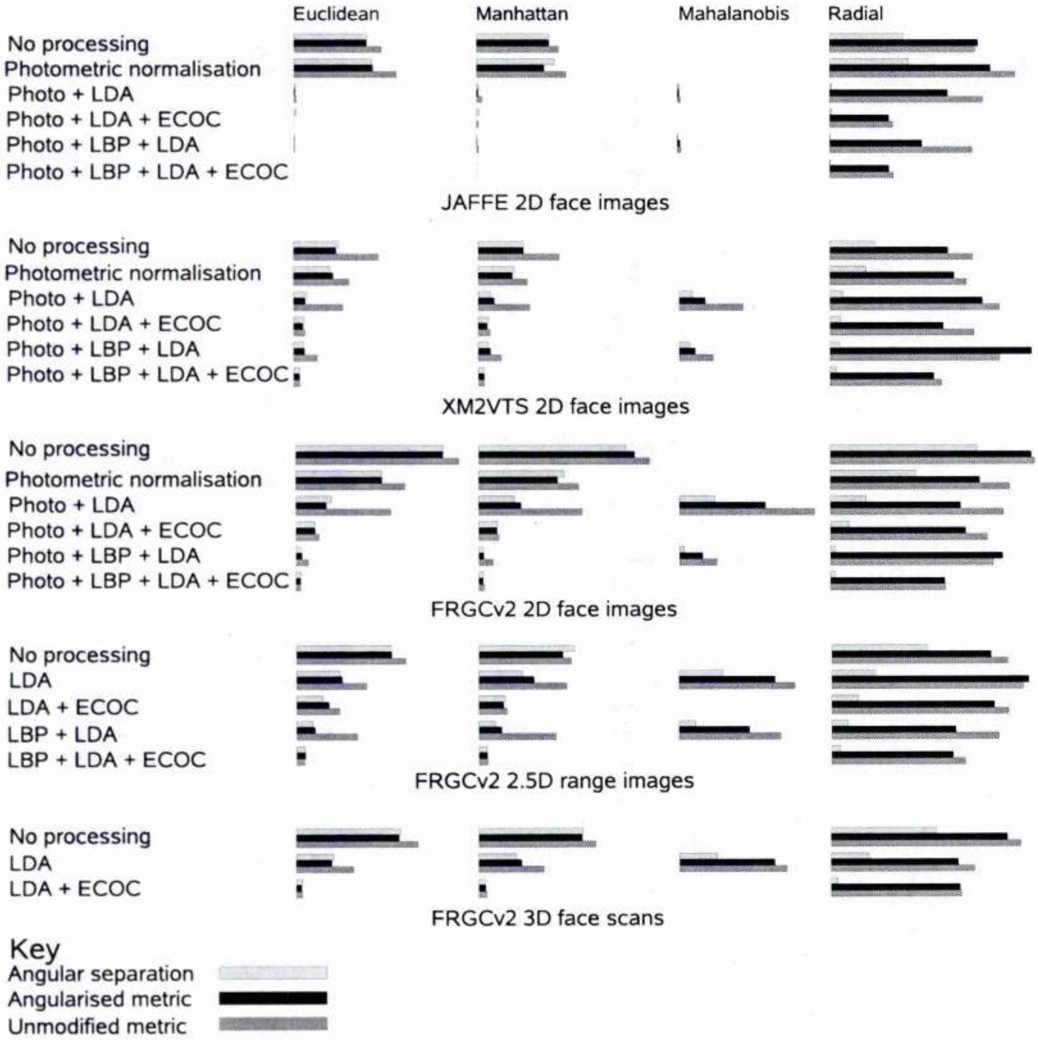

Figure 6.4: A comparison of the posterior Bayes errors made on face data using different metrics in different feature extraction scenarios. For each metric the error made by using the unmodified metric is compared with that obtained using the angularised version of the metric (achieved by application of the $ang(\cdot)$ transformation) and with the error resulting from using the angular separation metric itself. Mahalanobis distance is omitted when the covariance matrix is singular.

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
FRGCv2 2D							
None	37	39		33	47	9-148	5-158
	33	35		34	46	10-148	5-159
Photometric	25	23		19	41	20-111	35-130
	20	18		19	34	20-112	35-131
Photo + LDA	21	24	31	8	40	27-98	57-113
	7	10	20	8	30	27-99	57-114
Photo + LDA + ECOC	5	5		4	36	0-96	31-106
	4	4		4	31	0-96	28-103
Photo + LBP + LDA	2.84	3.24	9	1.07	37	26-89	61-108
	1.33	1.17	5	1.07	39	26-89	61-108
Photo + LBP + LDA + ECOC	1.02	1.37		1.14	26	0-90	21-104
	1.08	1.03		0.97	26	0-88	22-101

Table 6.2: Bayes errors and angle range statistics for the FRGCv2 2D face images. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
JAFFE 2D							
None	20	19		17	33	12-104	42-140
	17	17		17	34	12-104	42-138
Photometric	24	21		18	42	13-102	48-127
	18	16		18	37	13-103	47-126
Photo + LDA	0.51	1.25	0.56	0.42	35	2-45	43-143
	0.27	0.38	0.32	0.46	27	2-47	45-143
Photo + LDA + ECOC	0.00	0.35		0.52	14	0-78	27-101
	0.00	0.00		0.00	13	0-39	45-105
Photo + LBP + LDA	0.15	0.40	0.68	0.09	33	2-43	36-148
	0.09	0.11	0.52	0.11	21	2-43	35-149
Photo + LBP + LDA + ECOC	0.00	0.00		0.00	15	0-16	36-121
	0.00	0.00		0.00	13	0-21	56-100

Table 6.3: Bayes errors and angle range statistics for the JAFFE 2D face images. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
FRGCv2 2.5D							
None	25	21		21	39	18-112	24-132
	21	19		21	36	18-112	24-132
LDA	16	20	26	10	43	28-93	36-119
	10	12	21	10	44	28-92	35-120
LDA + ECOC	10	6		6	40	2-94	16-105
	7	6		6	37	0-97	7-102
LBP + LDA	14	17	23	3.69	38	18-96	45-117
	4.21	5.09	16	3.74	28	18-96	45-118
LBP + LDA + ECOC	1.91	2.03		1.87	30	0-91	24-104
	2.00	1.91		1.78	27	0-92	8-102

Table 6.4: Bayes errors and angle range statistics for the FRGCv2 2.5D range images. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
FRGCv2 3D							
None	28	27		24	43	6-146	10-162
	24	24		24	40	6-145	10-162
LDA	13	15	25	9	33	11-90	26-143
	8	10	22	9	29	10-90	22-147
LDA + ECOC	1.45	1.84		1.45	30	0-90	13-105
	1.27	1.58		1.34	29	0-93	11-106

Table 6.5: Bayes errors and angle range statistics for the FRGCv2 3D face scans. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

mean-subtracted images, the angular separation between feature vectors $D_{ang}(\cdot,\cdot)$ has much more discriminative power than the magnitude of the feature vectors $D_{rad}(\cdot,\cdot)$. Furthermore, the effectiveness of $D_{ang}(\cdot,\cdot)$ can be progressively increased by applying normalisation and feature extraction techniques, whereas the same is not true to anything like the same extent for $D_{rad}(\cdot,\cdot)$. As an example of this, consider the FRGCv2 2D data set where the angular Bayes error B_{ang} falls steadily from 33% on the raw data to 0.97% when photometric normalisation, LBP, LDA, $ang(\cdot)$ and ECOC are applied. Another example is the FRGCv2 3D data set where B_{ang} similarly falls from 24% to 1.34% with the application of LDA, $ang(\cdot)$ and ECOC. By contrast, the values for the radial Bayes error B_{rad} for the same data sets only fall from 47% to 26% and from 43% to 29% respectively. Whilst this does represent an improvement, it is not sufficiently great to make $D_{rad}(\cdot,\cdot)$ a useful measure of dissimilarity in itself. It thus appears that variability in the magnitude of feature vectors is at best only a weak source of discriminative information (the possibility of using $D_{rad}(\cdot,\cdot)$ in combination with $D_{ang}(\cdot,\cdot)$ is discussed in section 9.2).

An explanation as to why this should be the case is suggested by Fig. 6.5 which shows the effect of artificially rescaling a selection of mean-subtracted face images in the radial direction. Rescaling by a factor greater than one produces a higher contrast image with exaggerated facial features whilst rescaling by a factor of less than one produces a blander, lower contrast version of the image. Neither of these operations, however, lead to a change in the basic character of the face - the shape of the nose, distance between the eyes, the presence or absence of facial hair etc., nor do they produce a change in the spatial relationship between the facial features. This strongly suggests that, if one considers a ray emanating from the mean image in any particular direction, then all the images which lie on that ray will belong to a unique individual and will not, for example, change from one identity to another as the distance from the mean is increased.

Inspection of the final two columns of Tables 6.1 to 6.5, which show the extreme values found in the angular intra- and inter-class distributions, also supports this assertion. It can be seen that the lowest inter-class angular separations between unprocessed face images or scans range from 5° to 42°, with the lower figure occurring under the adverse conditions of illumination and expression variation of the FRGCv2 2D data (fur-

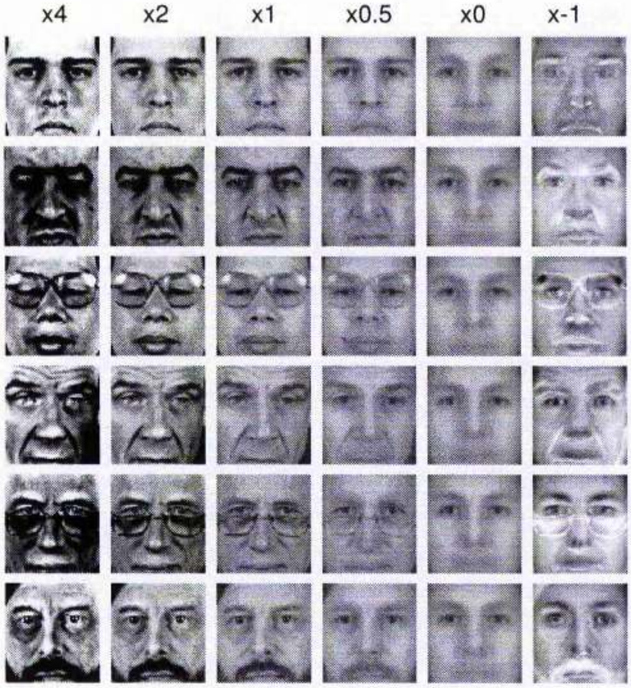

Figure 6.5: A sample of face images whose distance from the mean face has been rescaled by varying amounts. The original image (or veridical) is column x1, the mean face is column x0 and the anti-face is column x-1. Note that positive re-scaling does not change the basic character of the face and also that the anti-face does not resemble any real face.

thermore, as discussed below, the minimum angular separation increases when feature extraction and enhancement techniques are applied). This empirical evidence again suggests that, except under conditions of extreme adversity, such as very low or uneven illumination, the likelihood of finding two co-linear images of different subjects is small.

Although these considerations relate to the raw intensity image, as collected from the input sensor, columns B_{ang} and B_{rad} of Tables 6.1 to 6.5 indicate that they continue to apply even after post-processing by feature extraction techniques. Indeed, the main effect of these techniques, from a geometric point of view, is to greatly sharpen the angular separation between different individuals, whilst having a much smaller effect on the radial separation.

The final two columns of Tables 6.1 to 6.5 show that the effect of non-ECOC post-processing techniques is also to concentrate the angular distributions over much narrower ranges by increasing the minimum and reducing the maximum angular separation between images within each distribution. The process of angularisation does not have any significant effect on the angular distributions themselves, either in terms of the maximum spread of angular separations or on the B_{ang} statistics. This is to be expected since the objective of angularisation is to preserve the angles between feature vectors whilst at the same time bringing other distance metrics into line with angular separation. Angularisation does however appear to slightly improve the B_{rad} statistics.

Due to the non-linearity of the ECOC transformation, the effect on angular separation is somewhat different from other feature extraction methods as there is a strong tendency for the value to be mapped to either end of the intra- and inter-class distributions. This is because the nature of the SVM algorithm used in the ECOC base classifiers tends to force a definite class membership decision to be made, so that there is a bias towards the extreme values of ± 1 being output. Fig. 6.2 illustrates this effect on feature vectors extracted from 3D face scans but it applies equally well to 2D and 2.5D images. For the intra-class distribution the polarisation that occurs under ECOC means that many pairs of images are mapped to co-linear vectors whilst inter-class pairs become sharply clustered around a large angular value.

The artificially rescaled images of Fig. 6.5 suggest that, for 2D images, the point

Figure 6.6: Some examples of intra-class face image pairs with angular separation of more than 90° after photometric normalisation. Each image in the top row has the same subject identity as the corresponding image in the bottom row.

in feature space that represents the anti-face to any actual face image is unlikely to represent a realistic face. This is because the greyscale intensity is inverted with respect to the original image and this gives rise to artifacts such as white nostrils, lips, glasses and facial hair. We would thus expect the conical region surrounding any anti-face to be sparsely populated in the face database. This is supported by the fact that the largest angular separation to be found for the 2D data of Tables 6.1 to 6.2 is 158° (again, the worst-case example is furnished by the FRGCv2 2D images), indicating that the smallest angular separation between any real face and any anti-face is 22°.

It is perhaps a little surprising that two mean-subtracted image vectors of the same subject can be orthogonal to each other. Fig. 6.6 shows some examples of such pairs which are 90° or more apart. They suggest that the large angular separations can be explained by variations such as the presence or absence of glasses, whether the glasses are worn high or low on the nose, differences in facial expression and general misregistration errors.

6.1.2.2 Euclidean, Manhattan and Mahalanobis separation

The Euclidean, Manhattan and Mahalanobis distances between any pair of points in a feature space is influenced by both the angular and the radial separation between the points. It is to be expected, therefore, that in the absence of angularisation, the effectiveness of these two metrics in discriminating between intra-class and inter-class differences, falls somewhere between those of the angular and radial dissimilarity measures. Tables 6.1 to 6.5 and Fig. 6.4 show that this is indeed the case with, for example, the Bayes error on the XM2VTS data base after applying photometric normalisation and LDA being 11%, 12%, 14%, 3% and 39% respectively for the Euclidean, Manhattan, Mahalanobis, angular and radial dissimilarity measures. When LBP feature extraction is interposed before the LDA stage these figures fall to 5%, 5%, 8%, 2% and 39% but the relative ordering is preserved.

Tables 6.1 to 6.5 and Fig. 6.4 also clearly show that angularisation is effective in bringing the performance of $D_{Euc}(\cdot,\cdot)$ and $D_{Man}(\cdot,\cdot)$ into line with that of $D_{ang}(\cdot,\cdot)$, with the Bayes error for the former dissimilarity measures becoming similar to, and in some cases better than, that of the latter. An example of this can be seen in the figures for the FRGCv2 2D data (see Table 6.2) after application of photometric normalisation and LDA; the value of B_{ang} before and after angularisation is 8% whereas B_{Euc} and B_{Man} falls from 21% to 7% and 24% to 10% respectively. Again, the interposition of an LBP feature extraction stage greatly reduces the Bayes error rates but this does not alter the basic conclusion that the Euclidean and Manhattan metrics are further improved by angularisation, with reductions from 2.84% to 1.33% and 3.24% to 1.17% respectively being observed.

The Mahalanobis distance measure $D_{Mah}(\cdot,\cdot)$ is also improved by angularisation, with B_{Mah} falling, for example, from 0.68% to 0.52% on the JAFFE database using photometric normalisation together with LBP and LDA feature extraction. Generally speaking, however, the separation performance of $D_{Mah}(\cdot,\cdot)$ is significantly worse than that of $D_{Euc}(\cdot,\cdot)$ and $D_{Man}(\cdot,\cdot)$ and also suffers from the disadvantage that it cannot be calculated when the number of feature-space dimensions exceeds the number of training samples.

The purpose of applying angularisation is to allow stronger classifiers to be brought to bear on the face recognition problem and here we have used an ECOC ensemble of SVMs in that role. That the strategy can be highly effective is evidenced, for example, by the Euclidean Bayes error rate on the FRGCv2 3D data set which falls from 13% with just LDA to 1.27% after the further application of $ang(\cdot)$ and ECOC. Other data sets show a similar trend, for example B_{Man} on the XM2VTS database falls from 12%

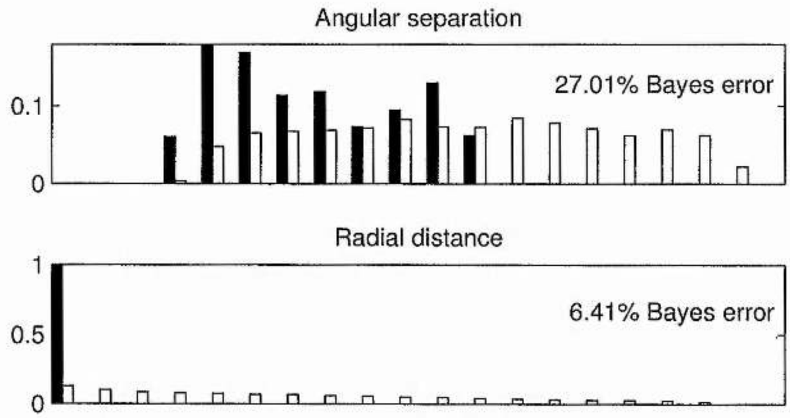

Figure 6.7: Intra-class (black) and inter-class (white) distributions for the synthetic data set.

Bayes Error Estimate (%)				Angle Range (degrees)	
B_{Euc}	B_{Man}	B_{ang}	B_{rad}	intra-class	inter-class
17	15	27	6	30-92	30-150

Table 6.6: Empirical error and angle range statistics that result when the training set for synthetic data is matched against the test set. Figures for Mahalanobis distance are omitted due to the singularity of the covariance matrix.

to 2.08% and the figure for JAFFE falls from 1.25% to 0%.

6.1.2.3 Non-face images

The question arises as to whether the effectiveness of $D_{ang}(\cdot,\cdot)$ as a measure of dissimilarity is a property of face data, or whether it is perhaps a general characteristic of high-dimensional data sets. It may be conjectured, for example, that any data set with sufficiently high dimensionality may be fitted to a hyper-ellipsoidal surface and that this would lead naturally to an angular dissimilarity measure. To investigate this an experiment similar to the raw data XM2VTS experiment was run, but using the synthetic image set described in chapter 5 instead of face images. The Bayes error estimates and intra/inter-class error distributions from this experiment are shown in Fig. 6.7 and Table 6.6 respectively. It can be seen that these are quite different in character

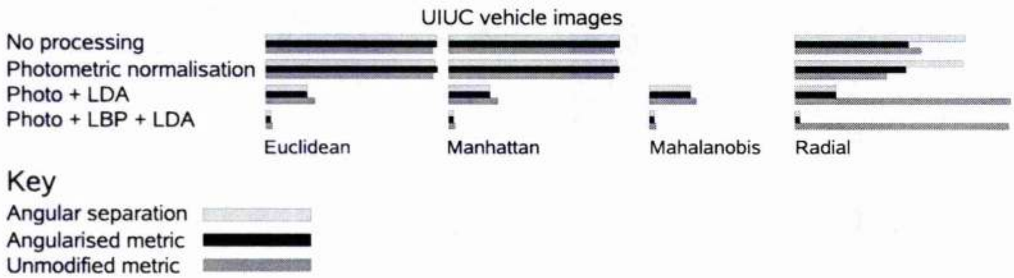

Figure 6.8: A comparison of the posterior Bayes errors made by using different metrics in different feature extraction scenarios when applied to the UIUC vehicle data. For each metric the error made by using the unmodified metric is compared with that obtained using the angularised version of the metric (achieved by application of the $ang(\cdot)$ transformation) and the error from the angular separation metric itself. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

to those obtained on face data (as shown in Figs. 6.2 and 6.3 together with Tables 6.1 to 6.5). In fact, for the synthetic data set, $D_{rad}(\cdot,\cdot)$, with a Bayes error of 6%, is much more discriminative than $D_{ang}(\cdot,\cdot)$ for which the value is 27%. This counter example shows that the superiority of angular over radial as a dissimilarity measure is indeed a property of the face data itself, rather than being merely due to the high dimensionality of the data.

To further investigate the application of these distance measures to non-face data the same techniques were applied to the problem of determining the presence or absence of a vehicle in the UIUC images (since this is 2-class problem and hence has only one way of partitioning the classes, ECOC cannot be usefully applied). The results of this investigation are shown graphically in Fig. 6.8 and detailed statistics are given in Table 6.7.

The vehicle data results fall into two distinct categories. On the raw, or photometrically normalised data, the Euclidean, Manhattan and angular metrics give very similar, and rather poor performance. Radial separation, however, yields the best separation performance, particularly after the application of photometric normalisation where the Bayes error of 21% is close to half that of angular separation. Under these circumstances

Processing Applied	Bayes Error Estimate (%)					Angle Range (degrees)	
	B_{Euc}	B_{Man}	B_{Mah}	B_{ang}	B_{rad}	intra-class	inter-class
UIUC vehicle images							
None	38	38		39	29	22-139	49-130
	39	39		39	26	22-138	49-13
Photometric normalisation	38	38		38	21	23-141	50-129
	39	39		38	25	23-140	50-128
Photo + LDA	11	11	11	9	49	0-180	0-180
	9	9	9	9	9	0-180	0-180
Photo + LBP + LDA	1.51	1.51	1.48	1.14	49	0-180	0-180
	1.14	1.14	1.14	1.14	1.14	0-180	0-180

Table 6.7: Bayes errors and angle range statistics for vehicle images. For each feature extraction scenario, the top and bottom rows show respectively the results obtained without and with application of the $ang(\cdot)$ transformation. Figures for Mahalanobis distance are omitted when the covariance matrix is singular.

angularisation is not beneficial.

When LDA is applied to the data, with or without LBP, the situation changes considerably and angular separation becomes much more discriminative than radial separation. Euclidean, Manhattan and Mahalanobis distances are also greatly improved and tend to give results that are close to, but slightly worse than angular separation. Now angularisation further improves the performance of these metrics, by up to 25%, and brings it up to the level of angular separation. This is illustrated by figure 6.9 which shows how angularisation greatly reduces the variance of the Euclidean distributions and the overlap between them. As expected angularisation has no effect on the angular separation metric itself.

It is also striking that, after the application of LDA and the $ang(\cdot)$ transformation, the separation performance of $D_{rad}(\cdot,\cdot)$ becomes identical to that of the other metrics. The reason for this is that LDA transforms to a feature space in which the vehicle and non-vehicle images form two diametrically opposite clusters which are tightly grouped

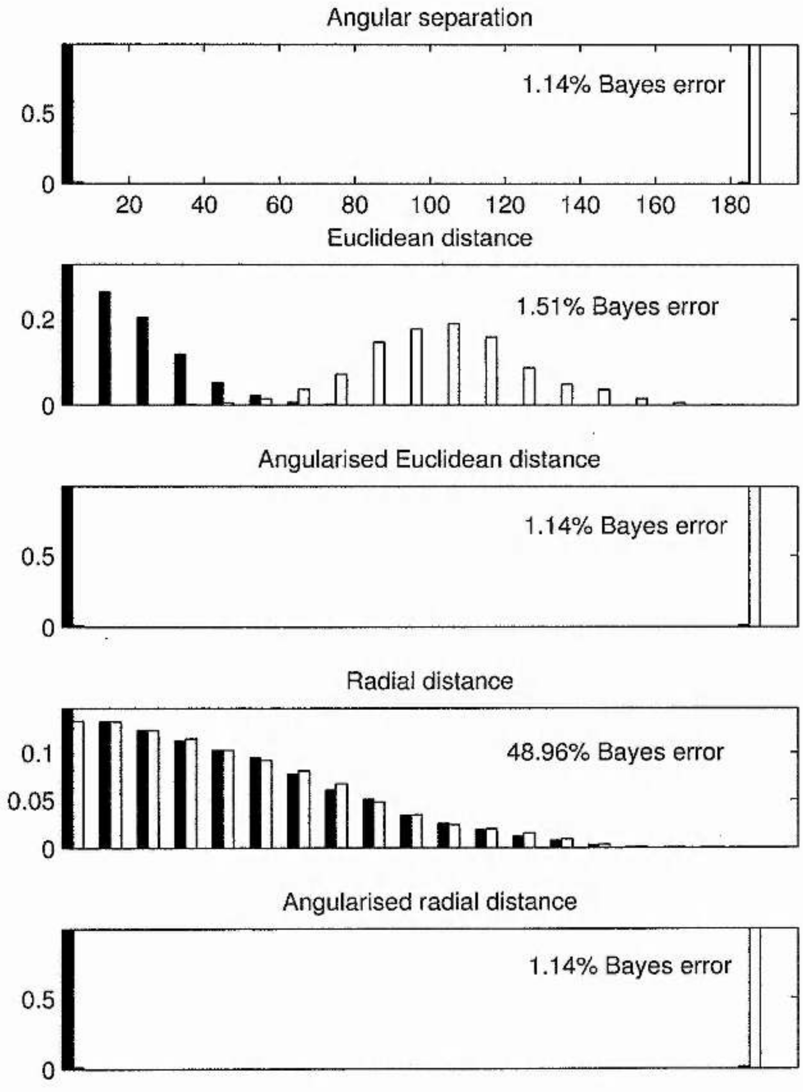

Figure 6.9: Intra-class (black) and inter-class (white) distributions for the UIUC vehicle data set after LBP + LDA feature extraction. Where applicable, angularisation was accomplished by application of the $ang(\cdot)$ transformation.

around a line passing through the origin. This is demonstrated by Fig. 6.9 which shows the angular, Euclidean and radial distributions for UIUC data with and without angularisation. Angular separation for intra-class variation peaks very sharply at $0°$ whilst that for inter-class variation peaks similarly at $180°$. In the absence of angularisation, distributions of radial distance follow a similar pattern to those observed with face data (see Fig. 6.3). Application of the $ang(\cdot)$ transformation, however, leads to the two clusters being projected into two different small regions of feature space and this in turn polarises the radial distance measure in the same way as angular separation.

It can thus be seen that when the vehicle image results are compared with those obtained on face data, there are some similarities and some differences. Nevertheless, the broad conclusions are that, for both face and vehicle images, the feature extraction techniques explored in this thesis of LBP, LDA and ECOC together lead to good angular separation and that the performance of non-angular metrics is improved by the process of angularisation.

6.2 Verification and Identification Experiments

The results of applying ECOC, with and without angularisation, to face recognition problems are shown in Fig. 6.10. For face verification this figure shows the *receiver operating characteristics* (ROC) curve and for closed-set face identification it shows the *cumulative match curve* (CMC). The former curve plots *verification rate* (VR) against *false acceptance rate* (FAR) as the acceptance threshold is varied. The latter curve plots the cumulative identification rate as the identification rank is increased. The term identification rank refers to the position of a face class in the list of all classes when they are arranged in decreasing order of likelihood of a match with a given probe image. Thus, rank 1 identification means that the true class is found to be the most likely class, or is in the first group of equally likely classes; rank 2 identification means that the true class is in second most likely group and so on. The cumulative match curve shows the results of summing these rank-n identification rates as the rank is increased.

As a baseline, nearest neighbour classification using a normalised correlation metric directly in the LDA space is also shown in Fig. 6.10. The processing applied in these

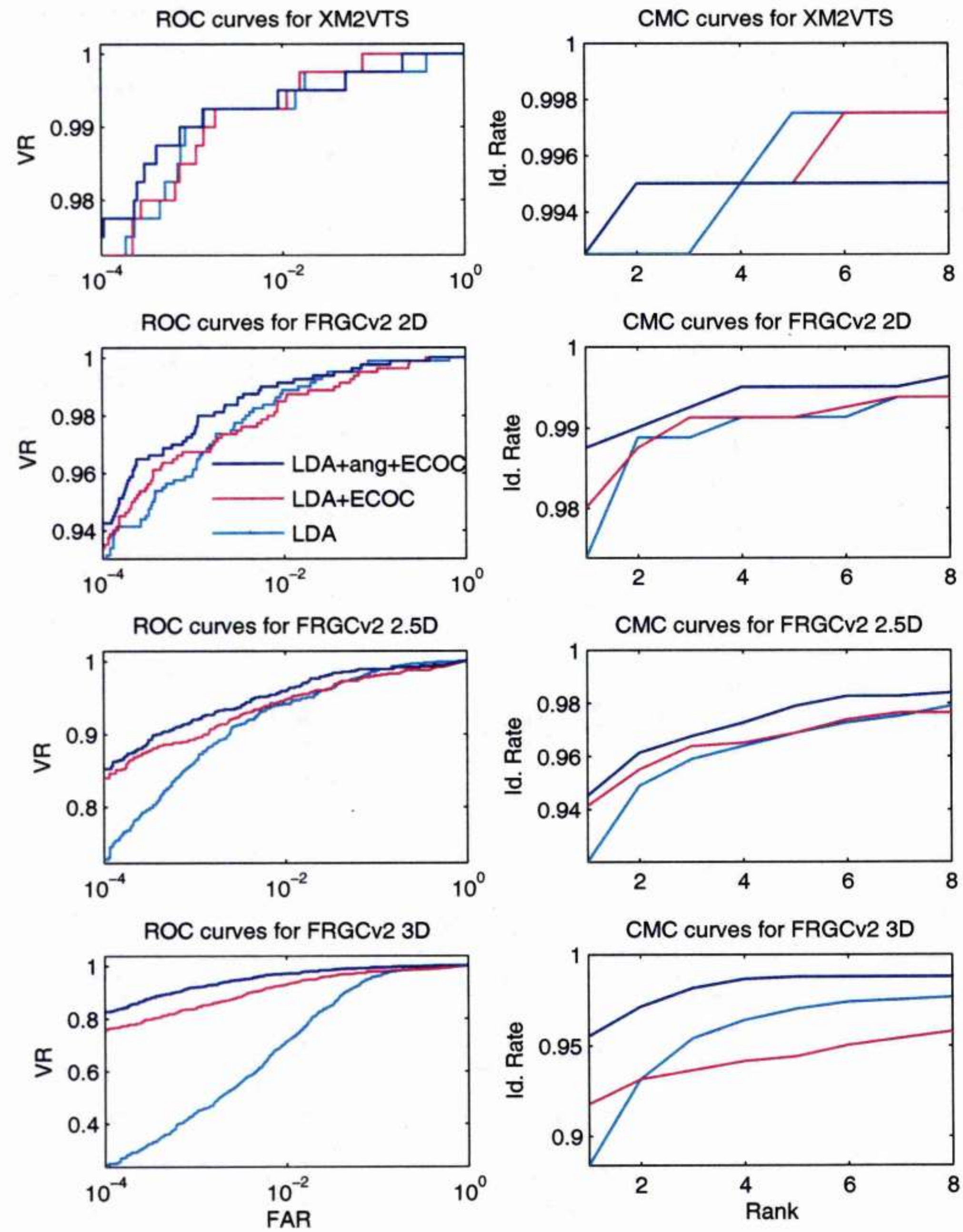

Figure 6.10: Best attainable ECOC verification and identification performance with and without angularisation. LBP feature extraction was first applied to 2D and 2.5D images. As a baseline, normalised correlation LDA results are also shown. JAFFE data is not shown as all methods gave 100% accuracy.

experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC with Gaussian kernel SVMs. This combination of techniques was selected as being the one that yields optimal results (a topic which is explored in more detail in chapter 7).

Note that there is some flexibility as to the order in which the above operations are applied, particularly with regard to the point at which the explicit angularisation step is performed. An important objective of this research, however, was to benchmark the performance of the SVM kernel functions $K_{ang}(\cdot, \cdot)$ and $K_{sph}(\cdot, \cdot)$ and compare it with explicit methods of angularisation when combined with a Gaussian kernel. In order to ensure a valid comparison, therefore, the $ang(\cdot)$ transformation always immediately precedes the point at which the ECOC ensemble of SVM base classifiers is trained.

It is apparent from Fig. 6.10 that the use of angularisation tends to have the desired effect of improving the performance of ECOC classifiers. Indeed, without angularisation, the performance of ECOC, particularly on the 2D images, is often comparable with that of LDA alone. After angularisation, however, ECOC tends to give the lowest error of the three methods; exceptions to this are the JAFFE data, where all methods gave 100% verification and identification accuracy, and the XM2VTS identification experiments where ECOC gave similar results with and without angularisation. The benefits of ECOC and angularised ECOC are particularly marked on the 3D verification task where a three-fold improvement in verification rate is observed at a FAR of 0.01%. The difference is less marked on the 3D identification task where angularised ECOC still significantly improves on LDA but, without angularisation, LDA outperforms ECOC (except for the rank 1 value).

6.3 Comparison of Angularisation Methods

Chapter 4 described four different approaches to incorporating angular dissimilarity measures into the construction of strong classifiers such as ECOC ensembles. In the first two methods either the $ang(\cdot)$ or $sph(\cdot)$ transformation is applied separately and this is followed by an ECOC ensemble of SVM base classifiers which use the $K_{Gauss}(\cdot, \cdot)$

kernel (see Eqn. 3.33). In the second two methods the ECOC ensemble is constructed directly from base classifiers which make use of the specialised SVM kernels $K_{ang}(\cdot,\cdot)$ or $K_{sph}(\cdot,\cdot)$ (as defined in equations 4.5 and 4.6 respectively). The discussion so far in this chapter has assumed that angularisation is applied by means of a general $ang(\cdot)$ transformation that is independent of the details of classifier construction. In this section we compare this transformation with other methods of angularisation.

Fig. 6.11 shows the verification and identification performance on the XM2VTS, FRGCv2 2D, 2.5D and 3D data sets. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the one of the four methods described above and ECOC with Gaussian kernel SVMs.

This figure suggests that, on the face verification task, there is no significant difference in performance between any of the proposed methods methods of angularisation, as all the curves lie very close to each other over the range of measurement. For face identification, all the performance curves are again very close to each other. There is, however, some evidence that, under the more adverse conditions of the FRGCv2 2D, 2.5D and 3D data, the $K_{ang}(\cdot,\cdot)$ kernel tends to give a slightly higher recognition rates. Note that the proposed version of $K_{ang}(\cdot,\cdot)$, as defined in Eqn. 4.5, tends to outperform $K_{ang2}(\cdot,\cdot)$ in which the $\arccos^2$ function is used. As expected, the results from the $sph(\cdot)$ and $K_{sph}(\cdot,\cdot)$ methods were identical apart from very small rounding errors.

In conclusion, these experiments imply that the improvement to be gained from angularisation is quite robust to the details of how it is accomplished. In the case of face identification using ensembles of SVMs, there is a slight preference for the use of the $K_{ang}(\cdot,\cdot)$ kernel.

6.4 Summary

This chapter has compared the performance of five distance metrics in separating intra-class from inter-class face differences. These are the Euclidean, Manhattan, Maha-

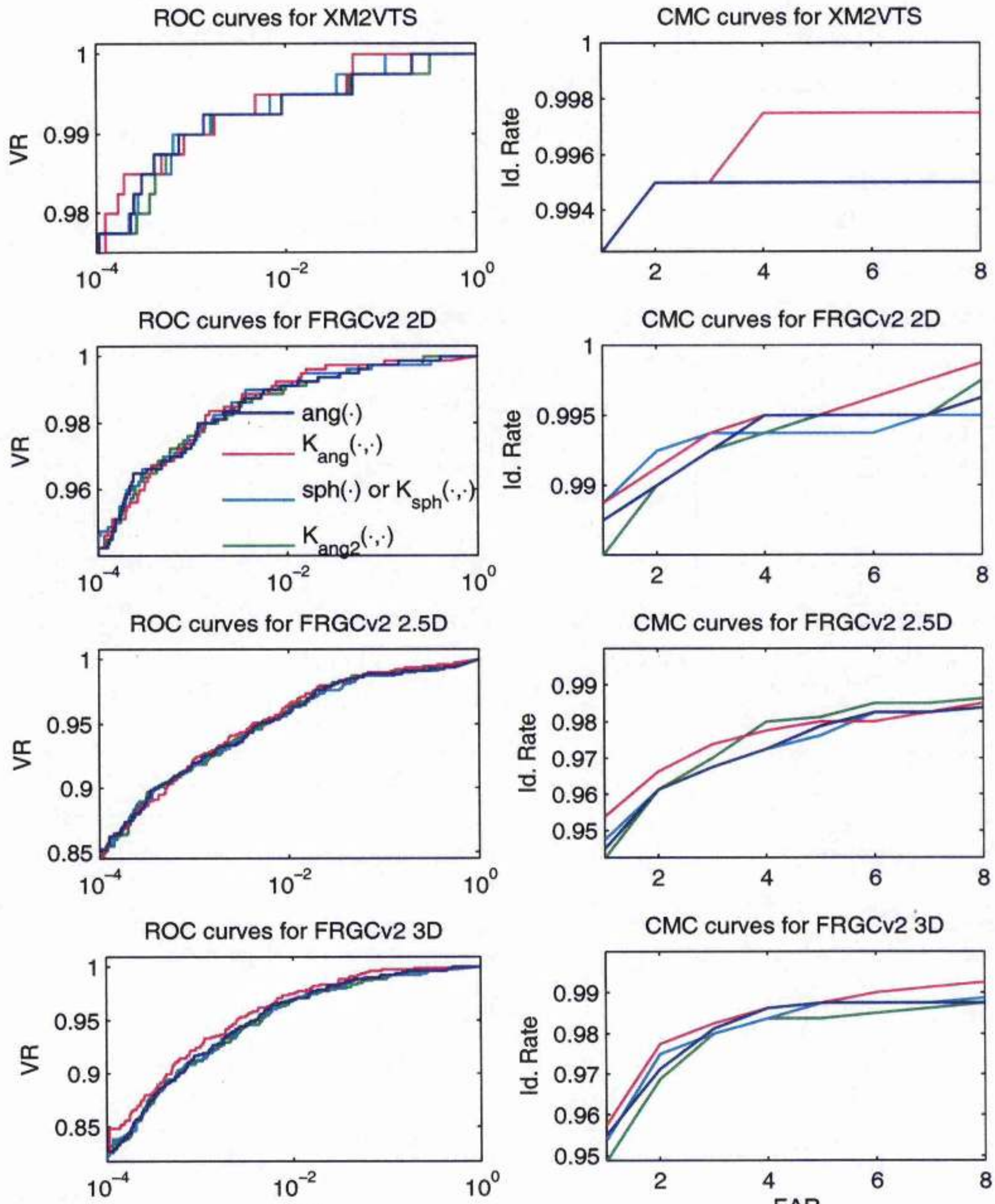

Figure 6.11: Verification and identification performance for different angularisation methods. Note that some curves are not always visible due to their being overlaid by others. $K_{ang2}(\cdot,\cdot)$ refers to a version of the angular kernel in which the $\arccos^2$ function is used.

lanobis, angular and radial distance metrics. To this end, the concept of the separation performance statistic was first defined as the overlap, or empirical Bayes error, that is measured between the intra-class and inter-class distributions for a given training and test set. It was shown experimentally that, when applied to face data, angular separation tends to have the best separation performance and radial separation the worst, with Euclidean, Manhattan and Mahalanobis being somewhere in between. It was further shown that the application of image enhancement and feature extraction techniques such as photometric normalisation, LBP, LDA and ECOC can considerably improve the separation performance of the first four of these metrics but that radial separation is not improved to any significant degree.

The application of angularisation was shown to be successful at bringing the separation performance of the Euclidean, Manhattan and, to some extent, the Mahalanobis metrics, into line with that of angular separation. The reasons for the success of angular, relative to radial, separation in face recognition were explored and it was shown, by experiments on synthetic data, that not all images have this property. Experiments with vehicle images, however, found that angularisation was again beneficial and imply that the method may be useful for a wider class of applications.

Angularisation allows classifiers to be constructed which, although not explicitly based on the angular separation metric, nevertheless capitalise on the improved discriminative capabilities of angular separation. It was demonstrated, using an ECOC ensemble of Gaussian kernel SVM classifiers, that the verification and identification accuracy of a general classifier can be enhanced by this method.

Finally, different methods of implementing angularisation, as defined in chapter 4, were compared. It was shown that all methods give broadly similar levels of accuracy, with the $K_{ang}(\cdot,\cdot)$ kernel having a slight advantage in face identification problems. These findings imply that the benefits of angularisation are quite robust to the details of how it is achieved.

Chapter 7

Ensemble Design for Face Recognition

Chapter 6 presented experimental results which demonstrated that, within the context of face recognition, the performance of a strong classifier can be enhanced by use of the angularisation process. This principle was illustrated by using an ECOC ensemble of SVM base classifiers.

In this chapter we look in more detail at the design choices that exist when constructing classifier ensembles. To this end we examine two possible approaches to their construction. These are the ECOC method, described in section 3.9.1, and the BHC algorithm which is described in section 3.9.2.

We begin in section 7.1 by looking at the ECOC approach in isolation and comparing the effectiveness of different SVM kernel and calibration algorithms and also looking at the effect of varying the ensemble size, using different code matrix design strategies and different decoding procedures. Section 7.2 then presents a similar discussion of BHC design considerations; it looks at the choice of SVM kernel and calibration algorithms, BHC decoding method and the clustering algorithm used to construct the BHC hierarchy. Finally, in section 7.3, the best results from ECOC are compared with those from BHC. Here experimental evidence is presented to show that, for face recognition purposes, ECOC is to be preferred to BHC and the reasons as to why this is the case are discussed.

7.1 ECOC Ensembles

In this section we study the performance characteristics of ECOC ensembles, as applied to face recognition problems.

7.1.1 Gaussian vs. linear SVM kernel

A comparison of the results of using Gaussian and linear SVM kernels as ECOC base classifiers is shown in Fig. 7.1. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC. For the Gaussian kernels, near optimal pairs of the parameters G and σ were found using the method of nested uniform design described in section 3.7. Linear kernels were found to be insensitive to the cost parameter and so fixed values of this parameter were used.

The implication of these experiments is that, under the relatively benign conditions represented by the XM2VTS data set, the linear kernel can perform as well as, or better than, the Gaussian Kernel. For example on XM2VTS the rank 3 identification rate achieved by the linear kernel is 100%, whereas the largest value attained by the Gaussian kernel is 99.5%. Under the more adverse conditions of the FRGCv2 2D, 2.5D and 3D data, however, the Gaussian kernel has a clear advantage over the linear one. This is particularly noticeable for 2D data set where illumination as well as expression variations contribute to the non-linearity of the decision surfaces. For the 2.5D and 3D data, where only expression variations exist, the difference in performance between the two kernels is less strong.

In summary, these experiments suggest that, under controlled conditions, the face manifolds can be accurately delineated by a series of hyperplanes and the linear kernel performs well; under more noisy conditions, however, this breaks down and the ability of the Gaussian kernel to model non-linear decision surfaces renders it more successful.

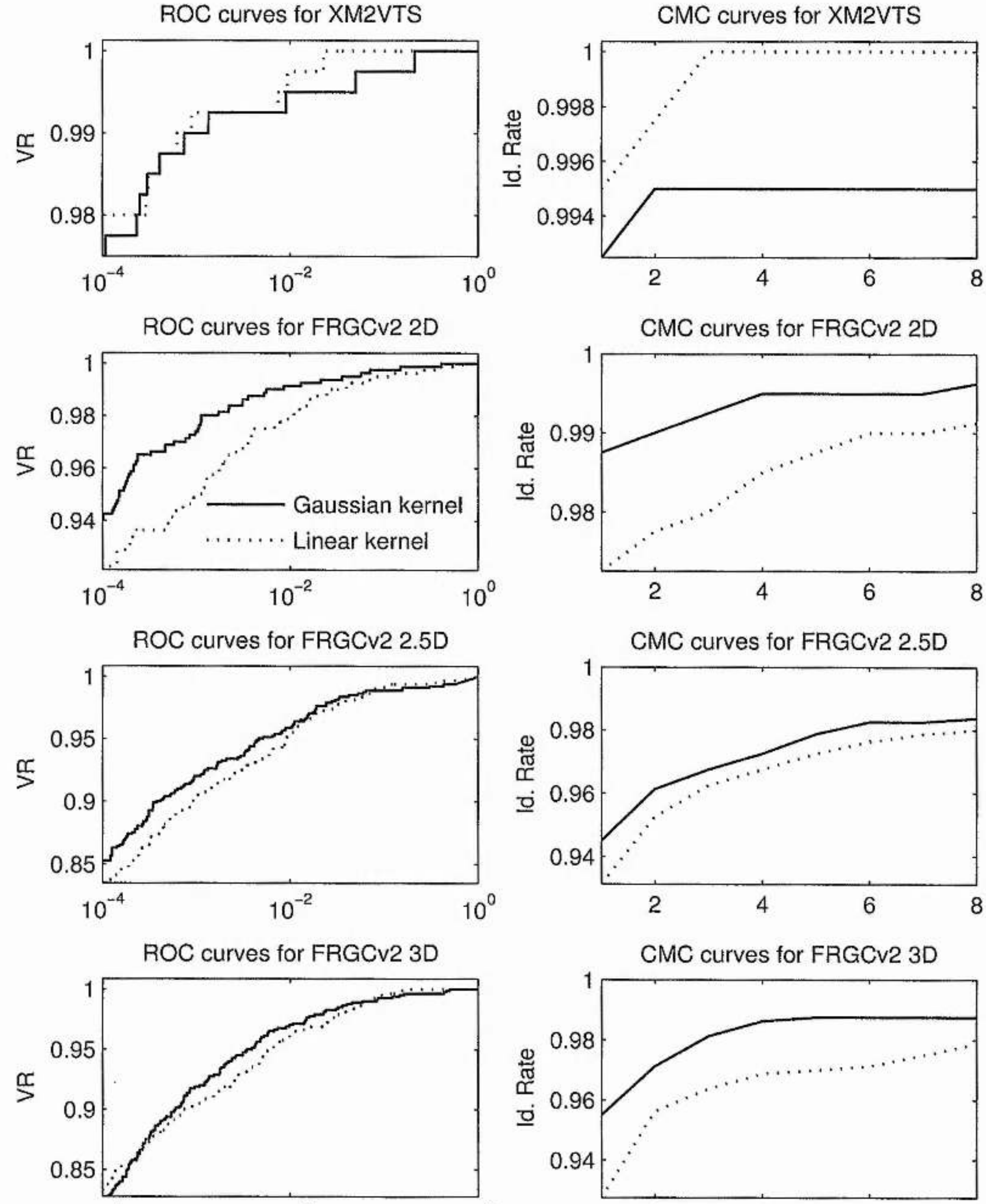

Figure 7.1: A comparison of Gaussian and linear kernels in ECOC SVM base classifiers.

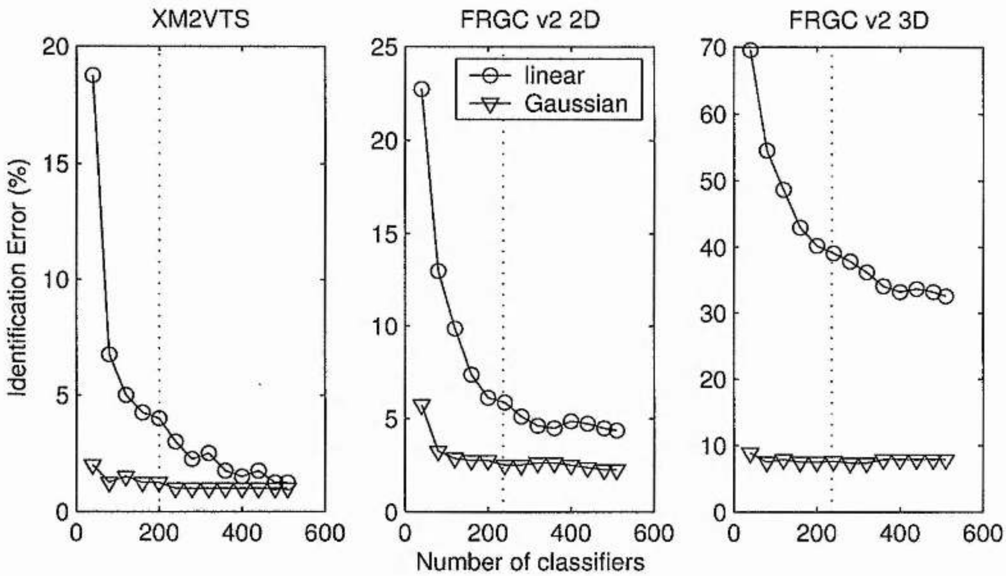

Figure 7.2: Variation in ECOC rank 1 closed-set identification error with different numbers of base classifiers. Values are shown for the Hamming decoding rule using Gaussian and linear base classifier SVM kernels. The dotted line indicates the total number of target classes.

7.1.2 Ensemble size

One advantage of the ECOC approach to constructing classifier ensembles is that the number of base classifiers can be varied to achieve a desired balance between accuracy and computational cost. In this section we examine how the performance characteristics of the ensemble are affected by the number of base classifiers used in the ensemble. Figs. 7.2 and 7.3 give a graphical representation of this information for identification and verification problems respectively, using both linear and Gaussian kernels. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC.

The general pattern, for both identification and verification problems, is that ensemble error reduces as the number of base classifiers increases until a point is reached where no appreciable improvement can be gained by the addition of more classifiers. The point

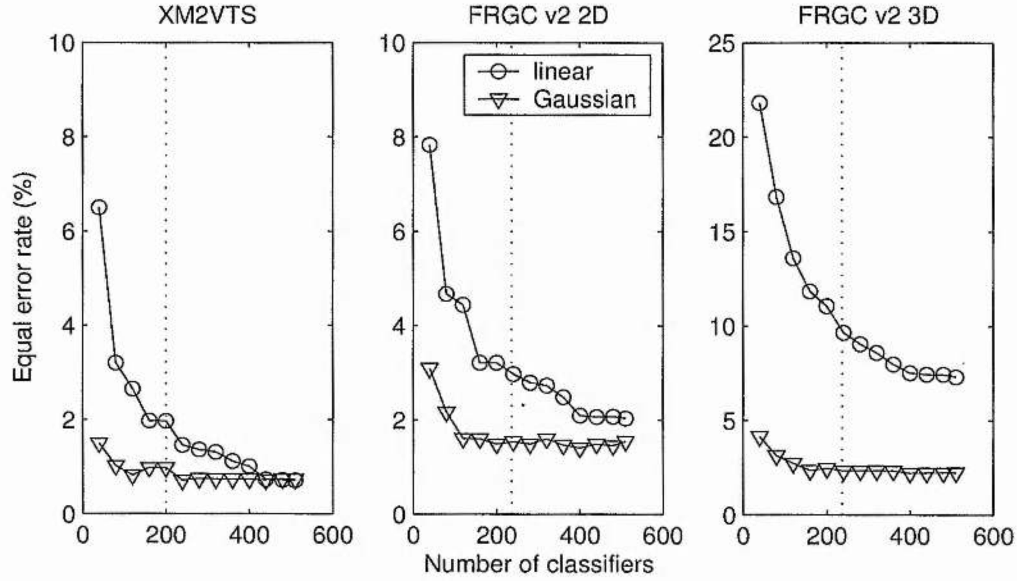

Figure 7.3: Variation in ECOC verification equal error rate with different numbers of base classifiers. Values are shown for the Hamming decoding rule using Gaussian and linear base classifier SVM kernels. The dotted line indicates the total number of target classes.

at which this happens varies depending on the data set and the type of SVM kernel in use. For Gaussian kernels it appears that about 1 to 1.5 times the number of target classes is sufficient to achieve optimal or near-optimal accuracy. For linear classifiers, however, about 2 to 2.5 times the number of target classes is needed before the error rate begins to stabilise.

These graphs also reinforce the general conclusion stated in section 7.1.1 that the Gaussian kernel is a better fit for face recognition problems than the linear kernel. Using Gaussian classifiers, reasonable accuracy can be obtained with a relatively small number of base classifiers; for linear classifiers however, a small ensemble size leads to unacceptably high error rates. Furthermore, on the FRGCv2 2D and 3D data, even with a large number of base classifiers, the performance of the linear kernel is clearly inferior to that of the Gaussian one. On XM2VTS data the results are different as, given a sufficient number of base classifiers, the performance of the linear kernel converges to that of the

Gaussian one. We suggest that this is due to the fact that the XM2VTS data set, being much less noisy than the FRGC data, is more readily separated using linear decision boundaries.

7.1.3 SVM calibration procedures

As noted in section 3.8, when combining the outputs from several SVM base classifiers, it is beneficial to calibrate the SVM output values so that they are linearly related to the probability of membership of the positive and negative target sets. This ensures that base classifier outputs can be combined with one another on an equal basis. The results of applying four different calibration procedures are shown in Fig. 7.4. These are: no calibration, a Gaussian mixture model, Platt's sigmoid fitting algorithm and isotonic regression. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC with Gaussian kernels.

The benefits of applying calibration are immediately apparent from Fig. 7.4 since in all cases the curve for no calibration lies below the other curves, indicating lower verification and identification accuracy. For 2D and 2.5D data the sigmoid and isotonic regression methods give very similar results with no clear advantage to either. Gaussian mixture modelling appears not to capture the underlying probability distributions with the same level of accuracy and this leads to slightly worse performance than the other two methods. For 3D data all three calibration methods give similar results on face verification, but with the sigmoid and Gaussian methods showing a slight advantage over isotonic regression on rank 2 and rank 3 face identification.

An illustration of the effect of applying the three calibration methods to a typical base classifier is shown in Fig. 7.5. This shows how the raw output from the SVM is mapped to a calibrated value which more accurately reflects the probability of membership of the positive and negative classes. It can be seen that effect of a calibration function is to both shift the position of the score of equal probability (i.e. the score which gives a calibrated output of zero) and to magnify the effect of small deviations from this score. Due to the discrete nature of the isotonic regression algorithm, the resulting calibration

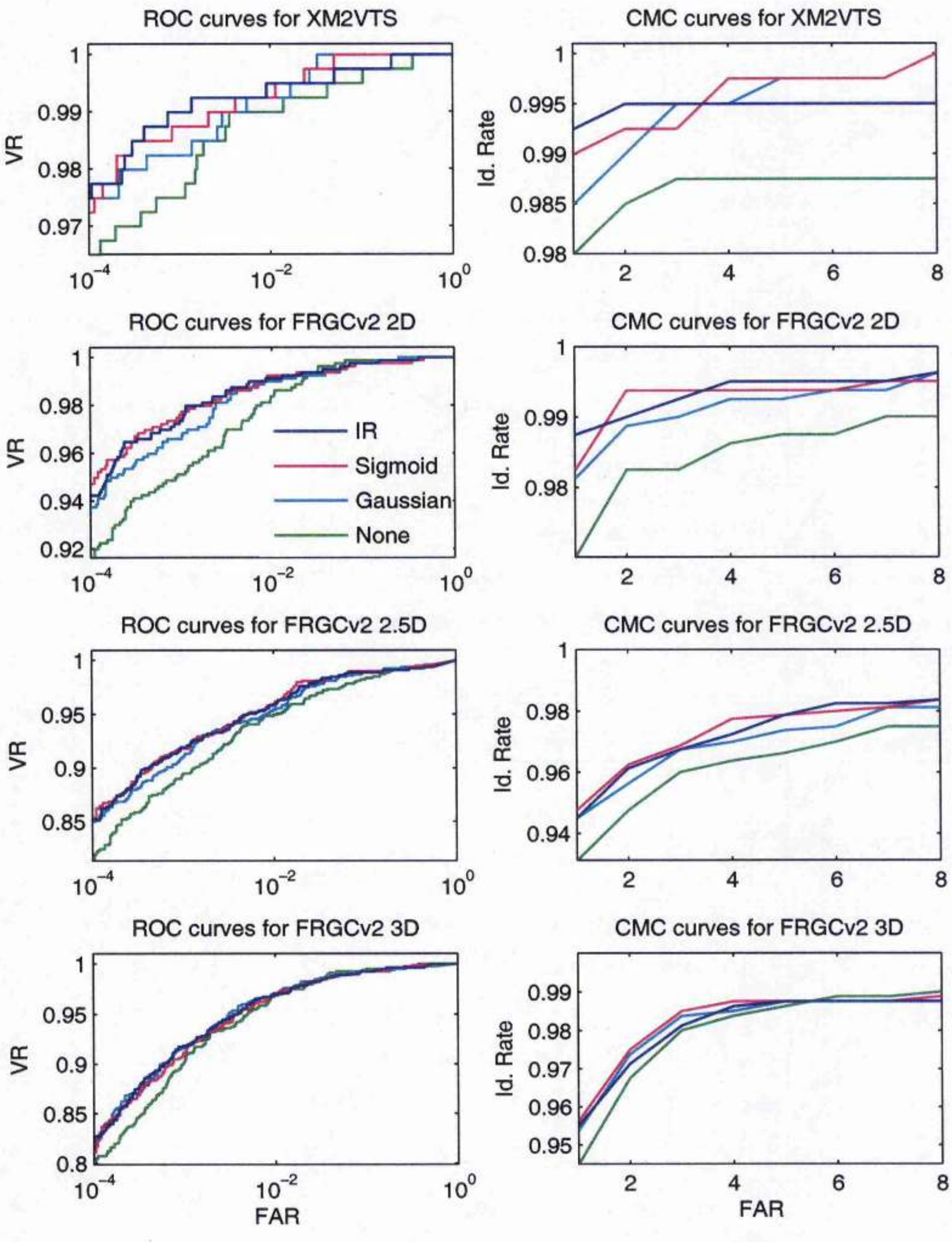

Figure 7.4: A comparison of different methods for calibrating ECOC SVM base clasifiers

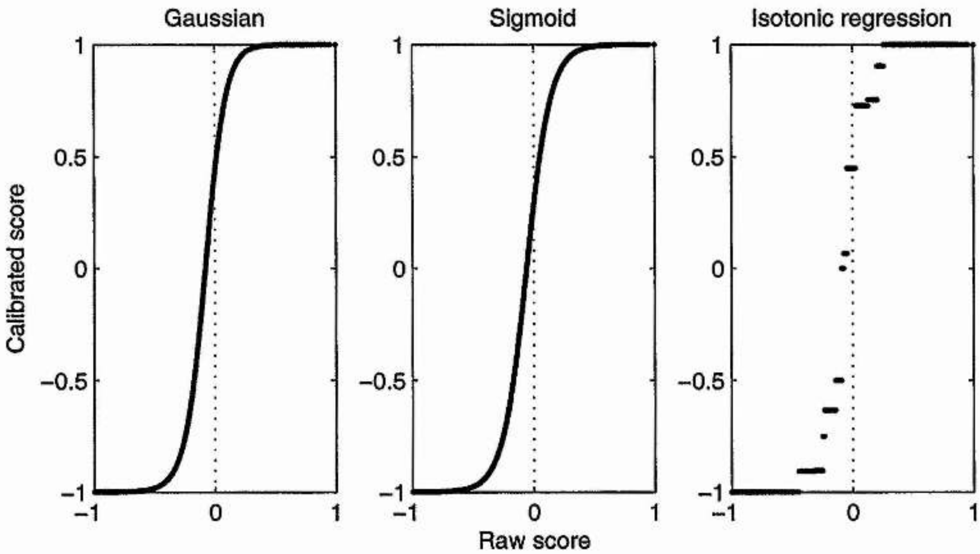

Figure 7.5: ECOC calibration graphs for a typical base classifier. The raw score is the value output by the SVM base classifier and the calibrated score is the modified value which is a more accurate reflection of class membership probability.

curve is discontinuous, whereas those generated by the Gaussian mixture modelling and sigmoid curve fitting methods are smooth.

7.1.4 Decoding methods

Section 3.9.1 describes two different approaches to using the output vector $\mathbf{y}(\mathbf{x})$ from an ECOC ensemble to generate a class distance score. The first of these is to take the codeword Z_i as a template for class ω_i and to measure the distance between $\mathbf{y}(\mathbf{x})$ and Z_i. In the second approach the training data images under the ECOC transformation are used as class templates and a nearest neighbour strategy is employed.

In Fig. 7.6 these methods are compared. In producing these graphs the processing applied was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC with Gaussian SVMs. For the nearest neighour method both the training and evaluation sets were used as a source of class examples and the Euclidean

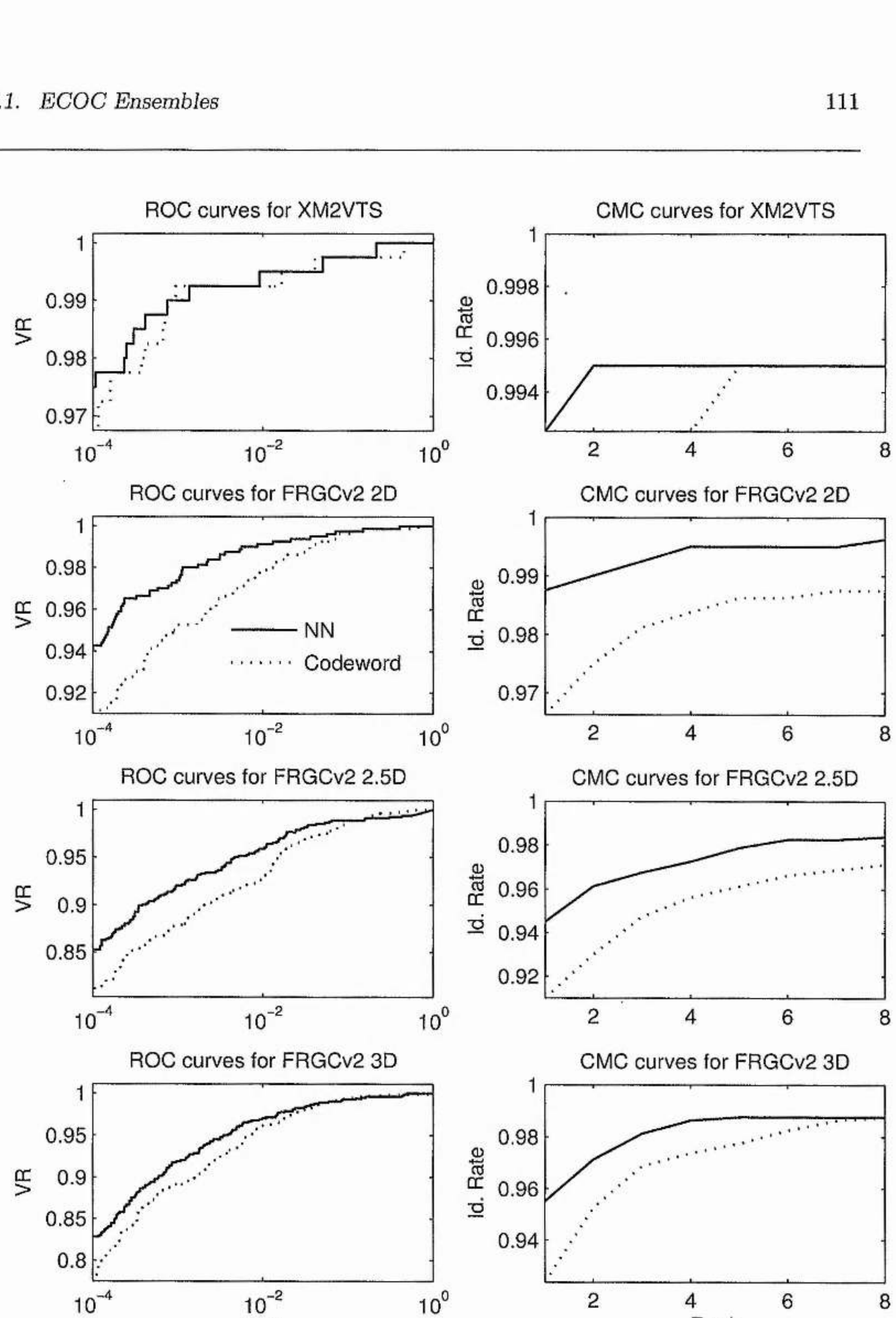

Figure 7.6: A comparison of different ECOC decoding procedures. 'NN' is the nearest neighbour method, averaged over class templates. 'Codeword' measures the Manhattan distance between probe and class codeword.

metric was used to measure the distance between a probe vector and each example; these distances were then averaged to obtain an overall class distance score. For the codeword method the Manhattan distance between probe and class codeword was used.

It can be seen from Fig. 7.6 that, in all cases, the nearest neighbour method outperforms the class codeword method. The difference is particularly marked on the FRGCv2 2D data where the codeword method produces a verification error at an FAR of 0.01% that is 50% greater and a rank 1 identification error that is 3 times greater. This suggests that, for noisy data of this kind, it is more reliable to base classification decisions on the actual training set classifier outputs, rather than the target outputs as represented by the codeword. The differences between the two approaches are least pronounced under the controlled conditions of the XM2VTS data whilst the moderately noisy 2.5D and 3D data give intermediate values.

7.1.5 Bose-Chaudury-Hocquenghem vs. random code matrix

When constructing ECOC code matrices it is desireable for the pairs of rows and columns to be maximally separated in terms of Hamming distance (see section 3.9.1). Fig. 7.7 compares the effect of using two different methods of constructing an ECOC code matrix. In the first of these the Bose-Chaudury-Hocquenghem algorithm is used to construct a rectangular binary code matrix with optimal row separation. A sub-matrix with the required number of rows is then constructed by choosing the one with the best column separation from 1000 random trials. In the second approach 1000 random binary matrices are generated and the one with the best row and column separation is chosen. These two methods are comparable in the sense that they consume similar amounts of CPU resource. In producing these graphs the processing applied was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC with Gaussian SVMs.

It can be seen from Fig. 7.7 that that there is in fact very little difference in the performance of the two methods of code matrix construction. Such differences as do exist, however, generally favour the Bose-Chaudury-Hocquenghem approach.

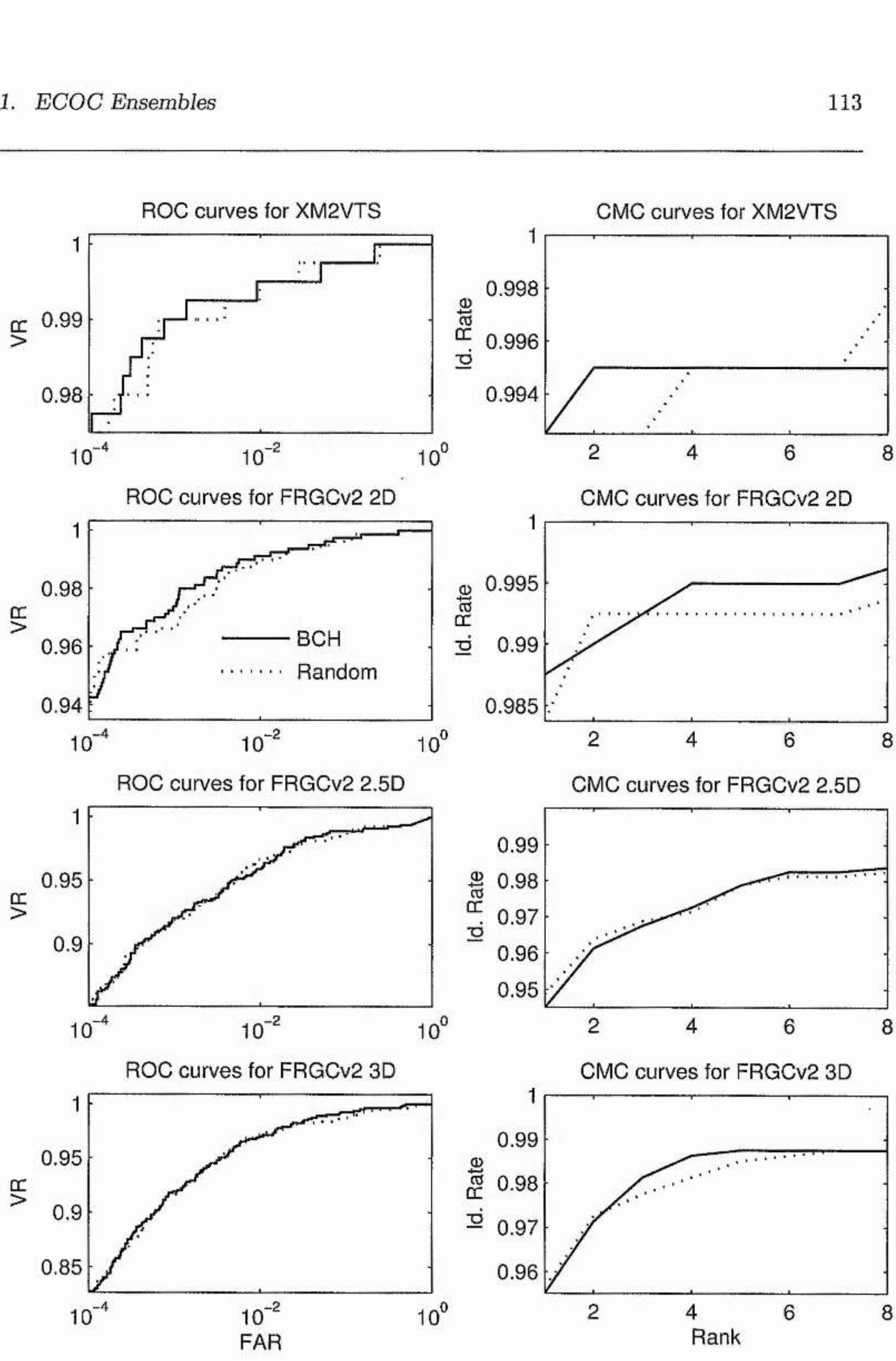

Figure 7.7: A comparison of ECOC code matrix generation methods. These are the Bose-Chaudury-Hocquenghem algorithm with random sub-selection of rows and the method of randomly generated matrices.

7.2 BHC Ensembles

In this section we undertake an experimental investigation for BHC ensembles which is similar to that for ECOC described in section 7.1. The aim is to find the combination of decoding rules, feature extraction procedures and base classifier algorithms that leads to the optimal BHC performance.

7.2.1 Decoding methods

Section 3.9.2 describes two different approaches to using the output from a BHC ensemble to generate a distance measure between an input vector $\mathbf{x}$ and an arbitrary class ω_i. The first of these methods is to descend the tree, from the root node to the terminal node corresponding to class ω_i, and at each stage j take a hardened decision as to whether $\mathbf{x}$ is most probably indicative of the left branch Ω_j^- or the right branch Ω_j^+. If all the decisions take the branch which leads to ω_i then $\mathbf{x}$ is deemed to belong to class ω_i and the distance is 0. If any branch fails to meet this condition, however, then the distance between $\mathbf{x}$ and the target class ω_i is 1. This method thus leads to a discrete distance measure which takes values from the set $\{0, 1\}$. The second method is to regard the soft outputs from the base classifiers as measures of the probability that $\mathbf{x}$ belongs to Ω_j^- or $\Omega_j^+$and to multiply these values on the path from the root node to the terminal node for ω_i. This is then subtracted from 1 to obtain a probabilistic distance measure in the range $[0, 1]$.

In Fig. 7.8 these methods are compared on the XM2VTS and FRGCv2 data sets. In producing these graphs the processing applied was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and BHC with Gaussian SVMs.

The main point to note from Fig. 7.8 is the unsuitability of the hard decoding method for face recognition applications. In face verification tasks it allows only two effective threshold settings; when the verification threshold is in the range $[0, 1)$ then a single pair of VR and FAR values is produced which corresponds to the set of probe images which exactly match the target class. When the threshold is set to 1 all claims are

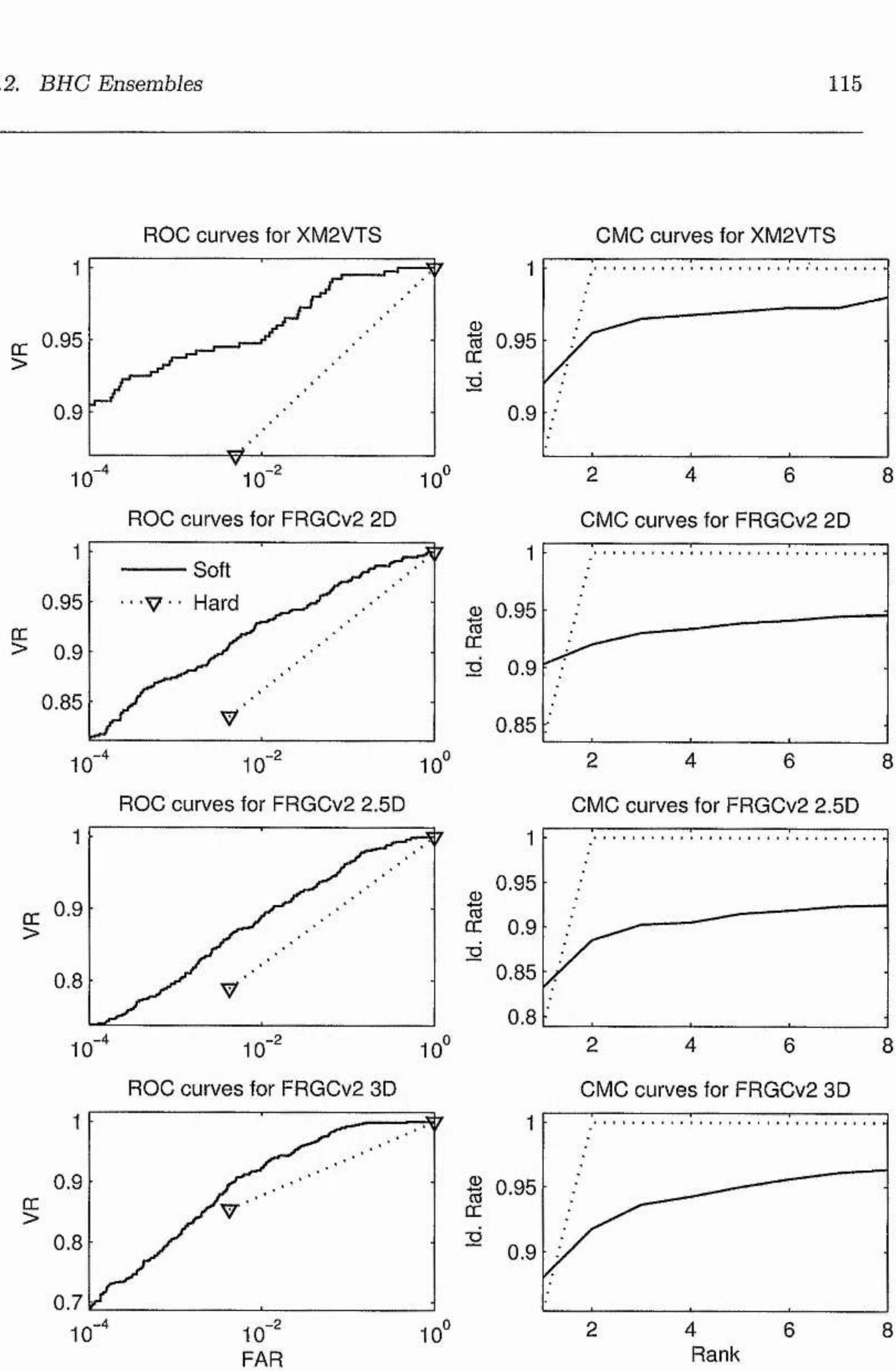

Figure 7.8: A comparison of BHC hard and soft decoding procedures.

accepted so the values of VR and FAR are both 1; a situation which is of no practical value. By contrast, the soft decision method allows for a gradual transition from low VR and FAR values to high values, thereby allowing the threshold to be set to a level which gives an acceptable tradeoff between false acceptance and false rejection errors.

In the case of face identification, the hard decoding method gives lower rank 1 recognition rates than soft decoding on all the data sets examined. Hard decoding appears to give better performance on rank 2 rates and above but again this is of no practical use since it is achieved by equating every member of the face database and thus does not produce a small enough subset of candidate matches. As with verification, therefore, the graduated performance curve of soft decoding is more suitable for face identification applications.

7.2.2 Gaussian vs. linear SVM kernel

A comparison of the results of using Gaussian and linear SVM kernels as BHC base classifiers is shown in Fig. 7.9. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and a BHC ensemble of SVMs. For the Gaussian kernels, near optimal pairs of gamma and cost parameters were found using the method of nested uniform design described in section 3.7. Linear kernels were found to be insensitive to the cost parameter and so fixed values were used.

It is apparent from these experiments that, for 2D, 2.5D and 3D data, the Gaussian kernel tends to perform better than the linear kernel on both verification and closed-set identification tasks. The difference is particularly marked on the 2D data. For 2.5D data the linear kernel is only slightly less accurate on face verification and actually gives a slightly better rank 1 recognition rate with identical values for higher rankings. A similar picture is presented for 3D data where the rank 1 identification rates are equal but the linear kernel slightly lags the Gaussian kernel for higher identification rankings.

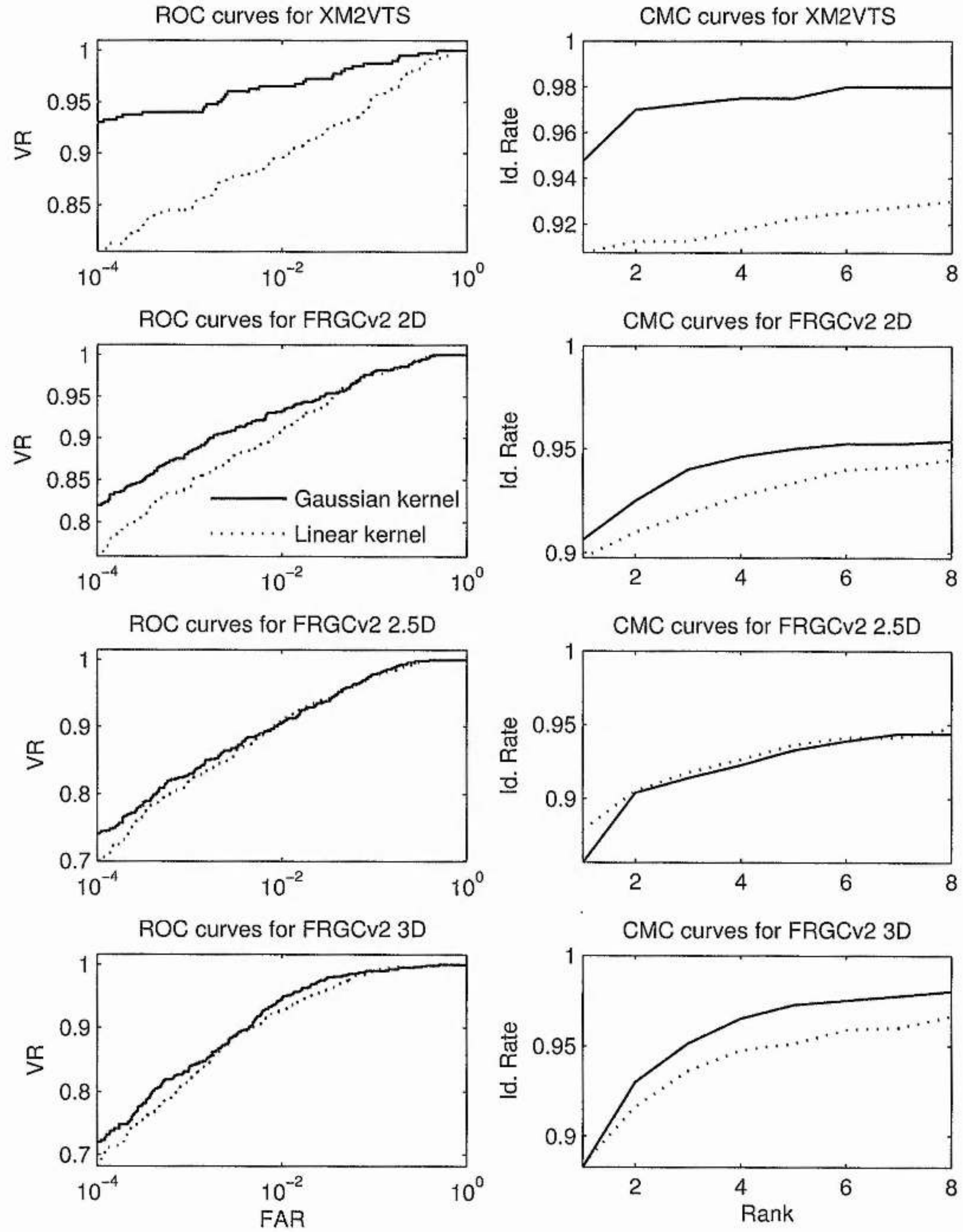

Figure 7.9: A comparison of the use of Gaussian and linear kernels in BHC SVM base classifiers.

7.2.3 SVM calibration procedure

As noted in section 3.8, when combining the outputs from several SVM base classifiers, it is generally beneficial to calibrate the SVM output values so that they are linearly related to the probability of membership of the positive and negative target sets. This ensures that base classifier outputs can be combined with one another on an equal footing. The results of applying four different calibration procedures are shown in Fig. 7.10. These are: no calibration, a Gaussian mixture model, Platt's sigmoid fitting algorithm and isotonic regression. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and BHC with Gaussian kernels.

It is apparent from Fig. 7.10 that the benefits of applying calibration are less clear cut for BHC than for ECOC (see section 7.1.3) where all three calibration methods performed better than no calibration. By contrast, for BHC verification, only the sigmoid method consistently outperforms no calibration whilst the Gaussian and isotonic regression methods give lower accuracy, particularly at low FAR values. For face identification, the uncalibrated rank 1 identification rate is equal to or better than the others in all experiments. Isotonic regression gives better results for the higher rankings but tends to give a lower rank 1 recognition rate. The sigmoid calibration method gives an identification performance which is similar the uncalibrated curve. Gaussian mixture modelling gives the worst performance of all the methods.

It is suggested that the reason for the different behaviour of these calibration methods, when compared with ECOC, is the great disparity in the size of training sets available for calibration, depending on the position of a node in the BHC tree. This is illustrated by Fig. 7.11 which shows the calibration curves for three different nodes. The first of these is the root node, for which a large number - typically hundreds - of examples is available. Secondly an intermediate node is shown which has a few dozen training examples. Finally, the bottom row shows the calibration curves for a leaf node which has only around five to ten training examples; in this case, therefore, a continuous curve cannot be shown.

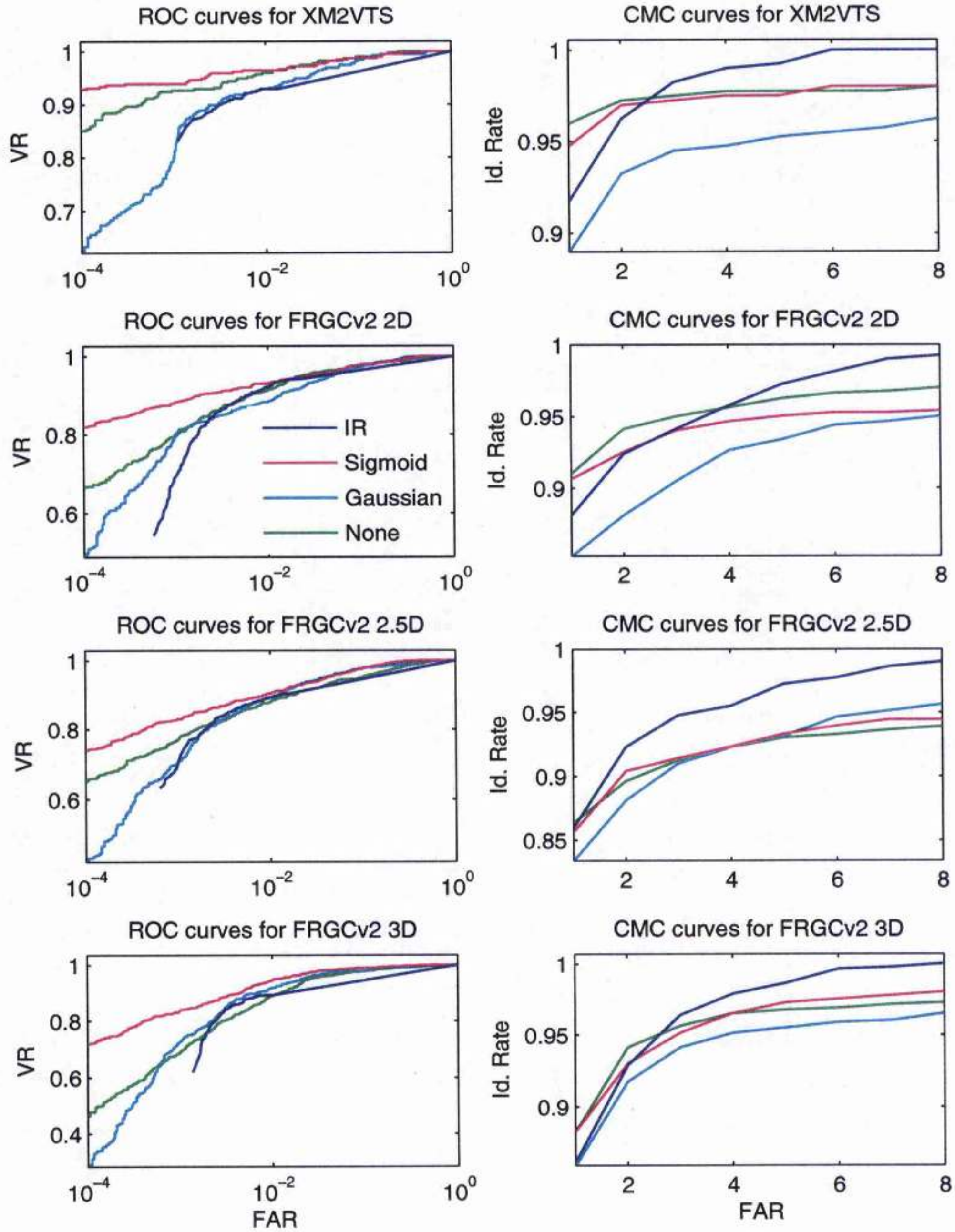

Figure 7.10: A comparison of different methods for calibrating BHC SVM base classifiers

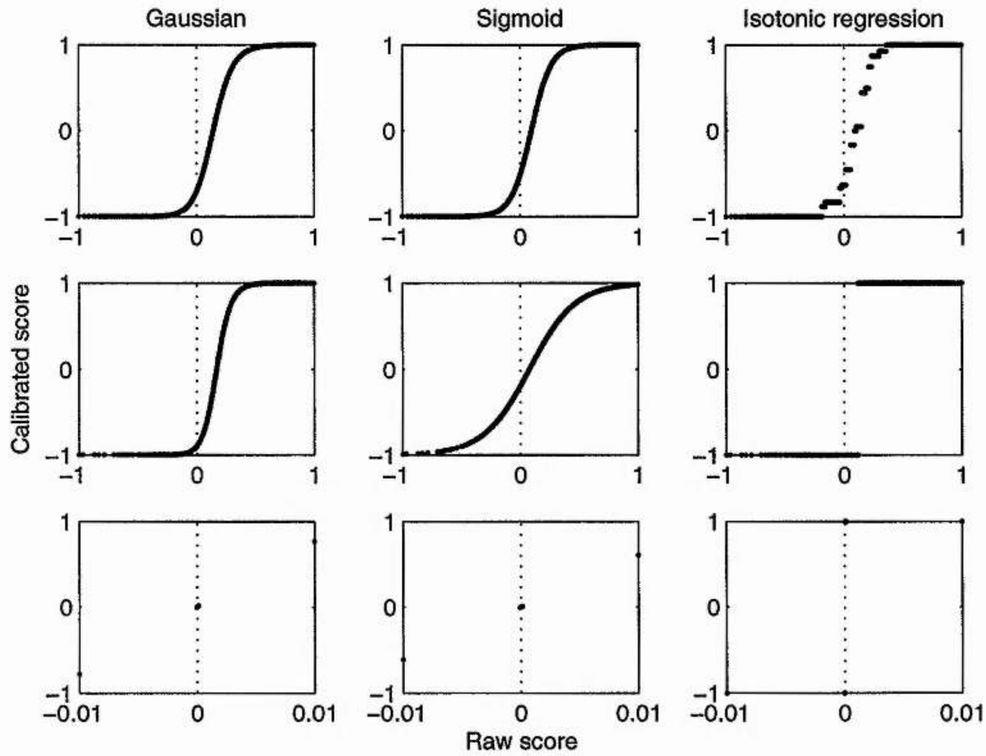

Figure 7.11: BHC calibration graphs for different base classifiers. The top, middle and bottom rows show respectively calibration graphs for the root node, an intermediate node and a leaf node. The raw score is the value output by the SVM discriminant function and the calibrated score is the modified value which is a more accurate reflection of class membership probability.

The top row of Fig. 7.11 is similar to the situation for ECOC base classifiers and there is sufficient data for all three methods to work effectively. In the second row it can be seen that there is insufficient data for isotonic regression to be trained correctly since it has become a simple step function which, in effect, shifts the position of equal probability, but otherwise simply serves to convert soft outputs into hardened decisions. This behaviour is continued in the bottom row but in this case the sample size is insufficent for any of the methods to produce a realistic calibration function. The success of the sigmoid method in the experiments of Fig. 7.10, however, suggests that it is the most successful in producing reasonable calibrated outputs; this may be attributed to the fact that it is the most heavily regularised of the three methods. These findings also illustrate the overfitting problems, alluded to in section 3.8, which can occur with isotonic regression when the callibration set is too small.

7.2.4 Clustering algorithm

An important aspect of the implementation of a BHC ensemble classifier is the clustering algorithm that is used at each node in the tree to partition a family of classes into a pair of sub-families. The method proposed in [39], which is described in section 7.2, is based on the use of a deterministic annealing algorithm; this gave good results for the problem of classifying hyper-spectral satellite image pixels into different types of ground cover. In this section we contrast the performance of deterministic annealing with that of the widely used 2-means clustering algorithm.

One aspect of this performance is whether the clustering algorithm produces balanced sub-families in the sense that, on average, the two clusters are of approximately equal size. This is a desirable property since it leads to approximately equal numbers of training samples for each target class. Fig. 7.12 shows the relationship between the cluster sizes produced by the two algorithms on the XM2VTS data set and it can be seen that, from the point of view of cluster sizes, the deterministic annealing method is to be preferred to 2-means. In the former case the two clusters tend to be of approximately equal size whilst in the latter case there is often a large discrepancy between them, with one cluster being typically much larger than the other.

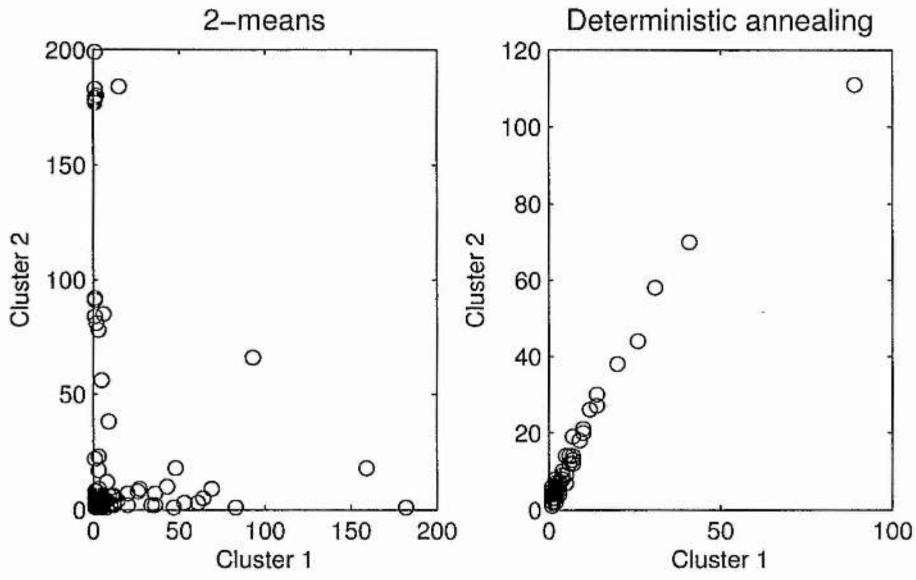

Figure 7.12: Scatter plot of cluster sizes produced by two different BHC clustering algorithms on the XM2VTS data set.

The distributions of base classifier errors for these two clustering algorithms, as measured on the XM2VTS test set, are shown in Fig. 7.13. It can be seen from this that both algorithms give rise to similar distributions, with the majority of classifiers being highly accurate, but with a long tail of base classifiers giving higher error rates. Notwithstanding the unbalanced cluster sizes, the 2-means algorithm appears to be slightly better than deterministic annealing in terms of base classifier error as it produces a higher number of base classifiers with zero error (75% vs. 65%) and also gives a lower mean base classifier error rate (2.8% vs. 3.6%). Against this, however, the range of error rates is higher for 2-means (0-50% vs. 0-25%), indicating that that the base classifiers may be a little more erratic in their generalisation performance.

These observations are reinforced by Fig. 7.14 which plots base classifier error as a function of $|\Omega_j|$, the size of family of classes which is partitioned by the j'th base classifier. This figure indicates that, for both clustering algorithms, the variance in base classifier error increases as $|\Omega_j|$ decreases; for small $|\Omega_j|$ the majority of base classifiers give a reasonable performance on the test set, but the number of poor performers becomes larger as $|\Omega_j|$ approaches zero. The effect is more pronounced for the 2-means

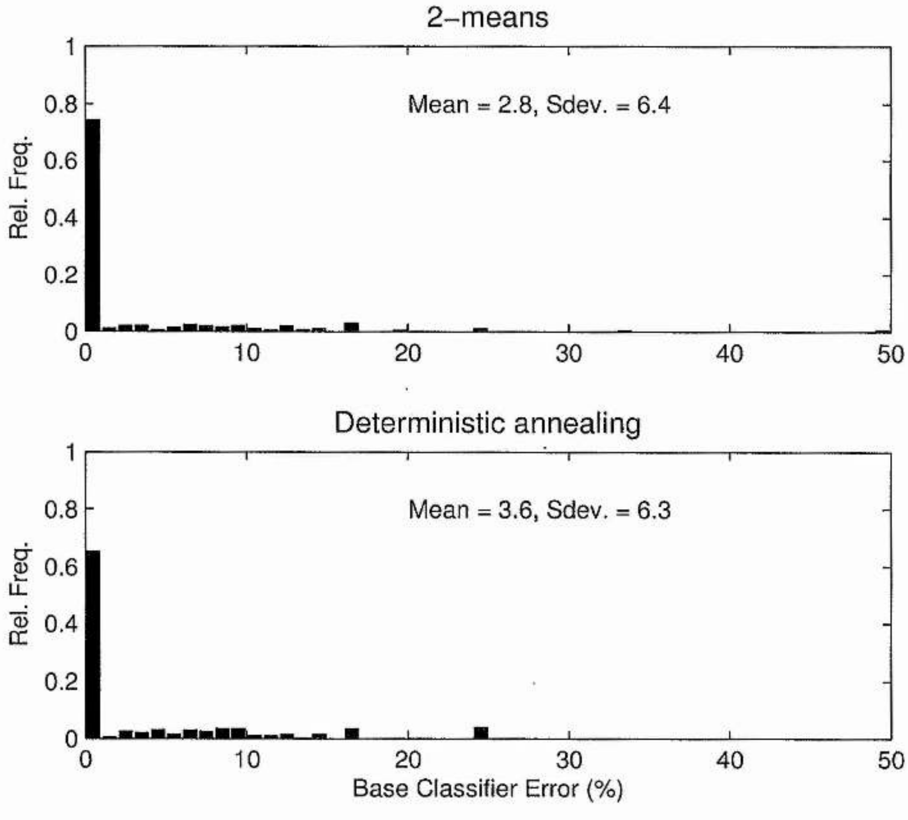

Figure 7.13: Distribution of XM2VTS test-set base classifier error rates from two different BHC clustering algorithms.

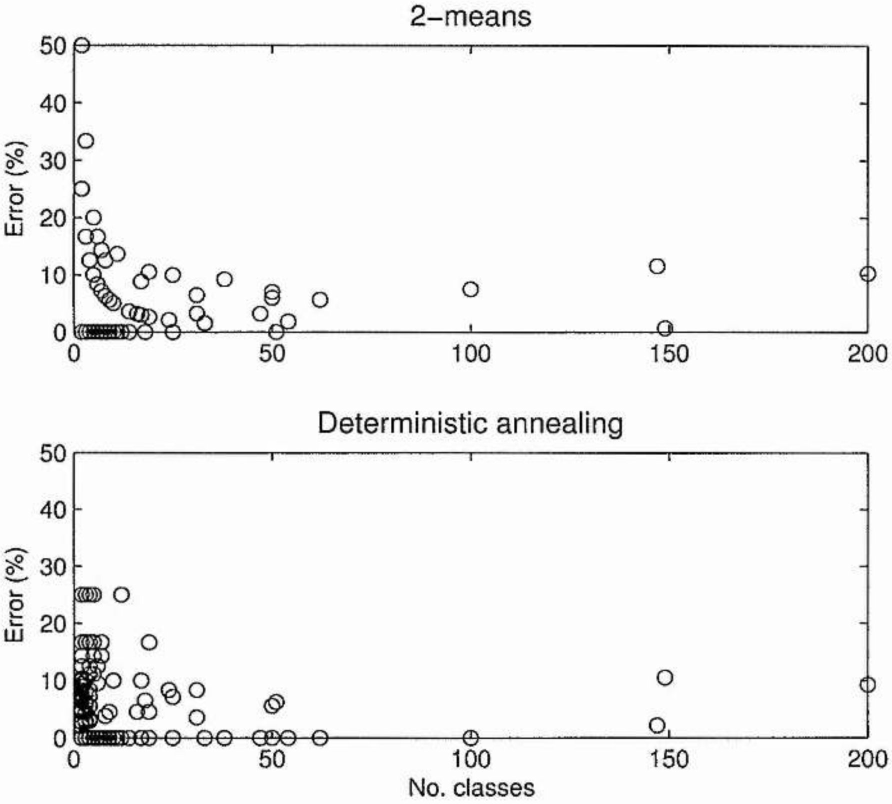

Figure 7.14: Scatter plot showing the variation in XM2VTS test-set base classifer error rates from two different BHC clustering algorithms as a function of the number of target classes.

algorithm than for deterministic annealing, perhaps as a result of the unbalanced cluster sizes. Note that the base classifier errror at the root node of the BHC tree (for which $|\Omega_j|$ is maximum) is around 10% in both cases. This figure limits the performance of the ensemble, particularly when the BHC-hard decision procedure is used, as errors accumulate on the path from the root node to a terminal node.

Finally, in order to determine which of the two clustering algorithms gives better ensemble accuracy in face recognition applications, they were compared using the XM2VTS and FRGCv2 data sets. The results of this comparison are shown in Fig. 7.15. It can be seen from this figure that, for both face verification and identification, the 2-means

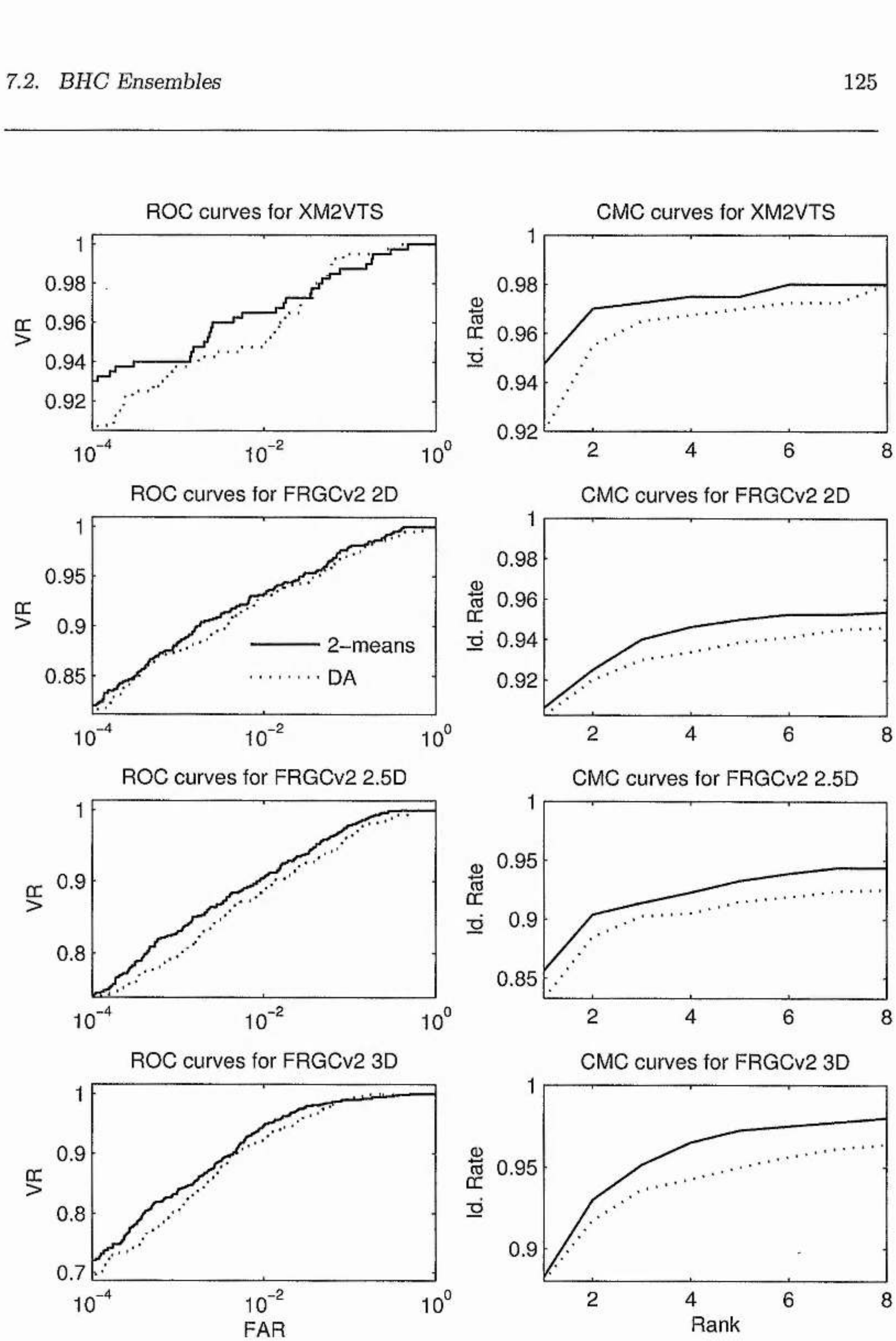

Figure 7.15: A comparison of the 2-means and the deterministic annealing (DA) clustering algorithms for BHC ensembles.

algorithm is consistently better than deterministic annealing. On the FRGC data sets the difference is quite small but it becomes more marked on the less noisy XM2VTS data.

The conclusion to be drawn from these experiments is that, despite the unbalanced cluster sizes, the 2-means clustering algorithm gives greater ensemble accuracy and is to be preferred to deterministic annealing for face recognition purposes. One possible reason for this is that in face recognition problems, unlike hyper-spectral remote sensing, the target classes do not form a natural hierarchy. In the latter problem, for example, pixels can be characterised as vegetation, rocks or water; vegetation can then be subdivided into upland and wetland and these in turn can be classified as trees and grasses and so on. These natural groupings can be exploited by the BHC algorithm. For face recognition applications, however, such deeply nested natural groupings do not exist so that the subdivisions found by the clustering algorithms will tend to be rather arbitrary. This problem is exacerbated by the much larger number of face classes (i.e. one per face identity) leading to hundreds or thousands of target classes, rather than the dozen or so classes that exist in the land classification problem.

Another aspect of the deterministic annealing, but not the 2-means, clustering algorithm is that it requires the input vector dimensions at each stage to be be reduced down to $C_j - 1$, where C_j is the number of target classes at the j'th node. In the case of terminal nodes, for example, this means that the input vector must be reduced to a single dimension. For satellite images, the special features of the problem allow this to be achieved by merging adjacent spectral bands as these tend to be highly correlated. For face recognition applications, however, this approach is not applicable and an alternative, perhaps less successful, solution must be found. In these experiments we make use of the fact that LDA dimensions are ranked in decreasing order of discriminative power and discard all but the first $C_j - 1$ dimensions at each node.

7.3 ECOC vs. BHC

Sections 7.1 and 7.2 have examined in some detail the design considerations to be taken into account when constructing ECOC and BHC ensemble classifiers for face recognition applications. In each case a set of techniques has been proposed that leads to optimal performance. In this section we compare these two ensemble architectures and comment on their relative merits.

Fig. 7.16 shows the results of applying the ECOC and BHC algorithms to face verification and identification problems. The processing applied in these experiments was photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D data), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and either ECOC or BHC with Gaussian kernels. For ECOC, code matrices with 510 columns were constructed using the Bose-Chaudury-Hocquenghem method, the base-classifier outputs were callibrated using isotonic regression and the nearest neighbour decision rule was employed. For BHC, clustering was performed using the 2-means algorithm, the base-classifier outputs were calibrated using the sigmoid algorithm and the soft decoding rule was used.

It is immediately apparent from this figure that ECOC performs better than BHC in all the experimental scenarios examined. The differences are most pronounced on the 2D experiments where, for example, rank 1 recognition error rates from BHC are up to 8 times greater than those from ECOC and false rejection rates at a FAR of 0.01% are up to 3 times greater. The differences on 2.5D and 3D data are not quite so great but are nevertheless significant.

The previous sections have touched upon a number of reasons as to why the ECOC architecture should be more successful than BHC for face recognition purposes and these are summarised here. One of the main differences is that the training task for all ECOC base classifiers is approximately equivalent in the sense that there exists a large training set that is evenly balanced between the positive and negative class families. For BHC ensembles, by contrast, both the size of the training set and the number of target classes is halved at each step from a node to its left or right child node. For the lower nodes of the tree this leads to a reduced training set size and can also result in unbalanced sizes for the positive and negative training sets. This in turn can lead to problems of poor generalisation performance by the base classifiers and inaccuracies when attempting to calibrate the base classifier output values so as to convert them to

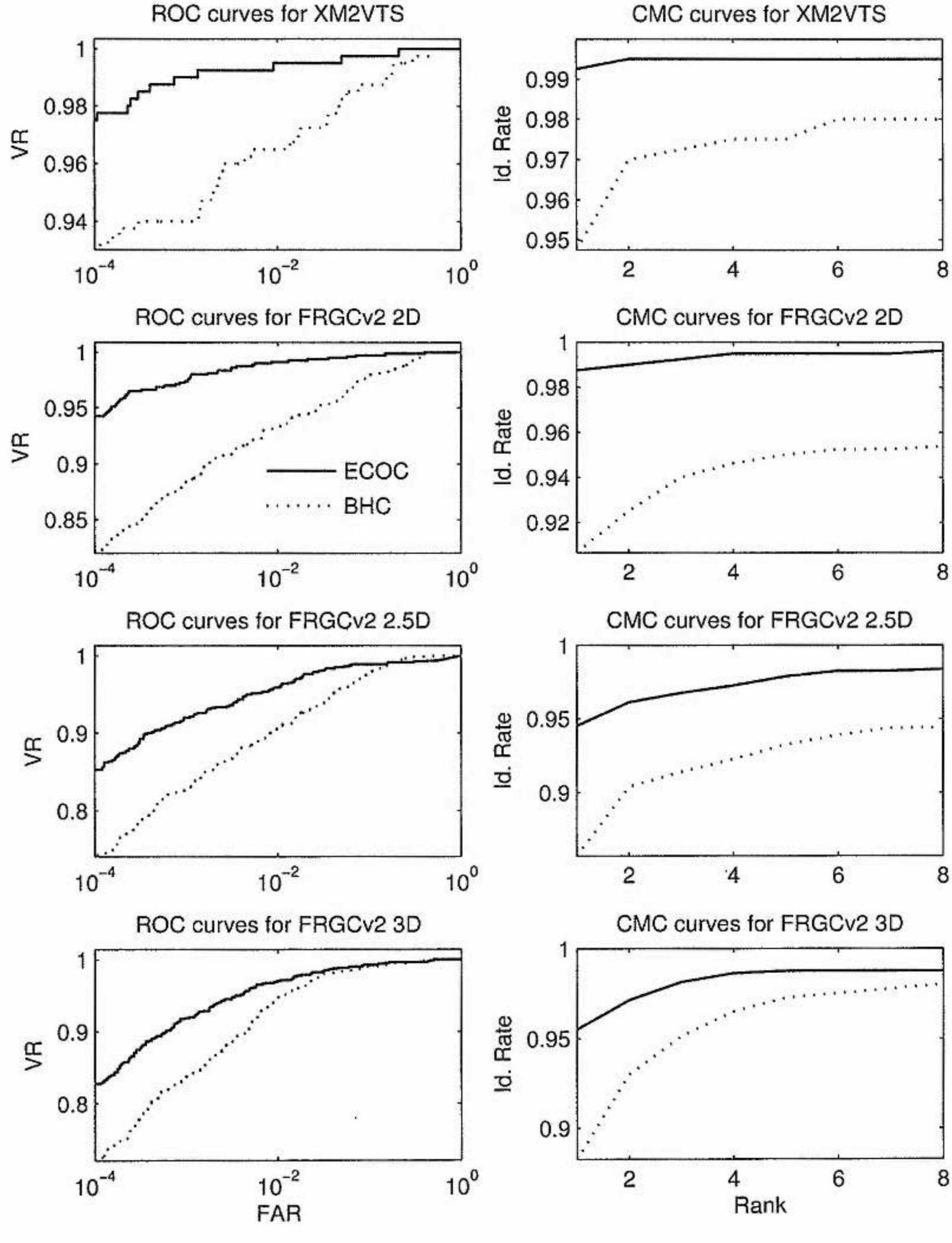

Figure 7.16: A comparison of ECOC and BHC ensemble classifiers.

probability estimates.

A further problem with BHC, when compared with its use for hyperspectral satellite image analysis, is that the deterministic annealing clustering algorithm is less successful on face images. This is thought to be due to the fact that face images do not naturally form deep hierarchies and also the problems caused by the need to reduce the number of input dimensions in order to handle small training set sizes.

Finally, it should be noted that the number of base classifiers in an ECOC ensemble is variable and can be increased as necessary to achieve a lower classification error. In a BHC ensemble, by contrast, the number of base classifiers is fixed at $C - 1$, where C is the total number of target classes, and this places a limit on ensemble performance.

7.4 Summary

A number of factors need to be considered when constructing ensembles of SVM classifiers for face recognition applications and these choices can have a significant impact on the accuracy of the resulting classification decisions. In chapter 6 it was shown that angularisation, whether performed as a pre-processing stage or incorporated into the SVM kernel function, tends to improve ensemble performance. This chapter has looked at other aspects of the design of ECOC and BHC ensembles and has presented evidence for a number of conclusions.

Firstly it can be stated that the Gaussian kernel tends to outperform the linear kernel on both face verification and face identification tasks. The benefit of the Gaussian kernel is most pronounced on 2D images, especially when the face images are captured under uncontrolled conditions, and less so on 2.5D and 3D data. For ECOC ensembles, where the number of base classifiers can be varied, the Gaussian kernel is also less sensitive to the number of classifiers used in the ensemble and reasonable results can be achieved with a smaller number of base classifiers.

When combining base classifier outputs it is generally beneficial to apply a calibration algorithm. Three such algorithms have been examined, namely Gaussian mixture modelling, isotonic regression and Platt's sigmoid fitting algorithm. The results of these experiments differ between ECOC and BHC ensembles. For ECOC the isotonic regression and sigmoid methods give approximately equal results which are generally better than those obtained from the Gaussian method. All three methods improve on no calibration. The picture that emerges when base classifier calibration is applied to BHC ensembles is more complex. For face verification, only the sigmoid algorithm shows a consistent improvement over uncalibrated results, particularly at low FAR values. The Gaussian and isotonic regression methods are actually counter productive. For face identification, however, the sigmoid method gives only comparable, and in some cases slightly worse results, than no calibration. Isotonic regression tends to give better results at the higher rankings, but at the cost of lower rank 1 recognition rates. It is probable that the reason for this more variable performance, when compared with ECOC, is that a large range of training set sizes is encountered at different levels of the BHC tree and this means that there is insufficient data to construct accurate calibration curves for the lower nodes.

The decoding algorithm used can also make a difference to ensemble accuracy. With the ECOC architecture one possibility is to measure the distance of the vector of output values from the codeword template for each class. For face recognition purposes, however, it appears that this method is not optimal and better results can be achieved using a nearest neighbour approach. It has been shown that the BHC-hard decoding method is unsuitable for face recognition applications since, for face verification, it does not allow sufficient flexibility in adjusting the tradeoff between false acceptance and false rejection errors and, for face identification, it does not allow for the extraction of a small group of candidate faces. Both these objections are overcome by using the BHC-soft decoding method.

Two methods for generating ECOC code matrices have been examined, namely randomly generated matrices and sub-selection of rows from a Bose-Chaudury-Hocquenghem matrix. The latter method yields better row separation and this appears to lead to marginally better ensemble performance.

Evidence has been presented to show that, for BHC ensembles, the deterministic annealing clustering algorithm is less effective for face recognition than the 2-means algorithm.

It is likely that this is a reflection of the fact that face images do not form deep natural hierarchies and also of the difficulty in progressively reducing the number of feature-space dimensions without sacrificing classification accuracy.

Finally, a comparison between the best attainable ECOC and BHC performance has shown that ECOC is greatly superior to BHC when applied to face recognition problems. This is attributed to the fact that ECOC allows more classifiers to be deployed to achieve greater accuracy and also the fact that ECOC produces large balanced training sets for each base classifier. By contrast, BHC necessitates a fixed number of classifiers and the training sets become smaller at the lower nodes of the tree. The latter fact can lead to problems of poor generalisation performance and difficulties in calibrating the outputs from some base classifiers. For face recognition applications BHC also suffers from the difficulties alluded to above, namely the non-hierarchical nature of the data and the lack of an effective method for drastically reducing the number of feature-space dimensions.

Chapter 8

Further Remarks

The previous two chapters have described respectively the advantages of applying the technique of angularisation in face recognition and the design decisions to be considered when constructing an ensemble classifier to solve face recognition problems. In this chapter we make a number of further observations which do not fall within the scope of those two themes.

8.1 The Benefits of LBP Feature Extraction

In the experiments described in the preceding chapters, multi-scale LBP feature extraction has been applied to 2D and 2.5D images as a pre-processing step prior to dimensionality reduction by LDA. In this section we show why this is advantageous.

Some examples of 2D and 2.5D face images using different LBP radii are shown in Fig. 8.1. A comparison of the verification and identification results on the XM2VTS, FRGCv2 2D and 2.5D data sets with and without LBP is shown in Fig. 8.2. In both cases photometric normalisation (for 2D images) and LDA feature reduction was applied followed by an ECOC classifier using Gaussian kernel SVM base classifiers. The difference is that in one case LDA was applied to a concatenation of multi-scale LBP histograms and in the other case LDA was applied directly to the input data.

It is immediately apparent from Fig. 8.2 that the ability of LBP pre-processing to

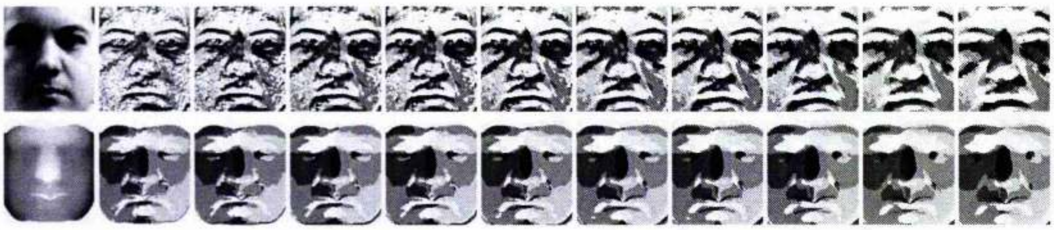

Figure 8.1: Examples of LBP processing applied to 2D and 2.5D images. The top row shows 2D images with the original (photometrically normalised) image on the left. The bottom row shows 2.5D range images with the original again shown on the left. In both cases LBP radii from 1 to 10 are shown.

extract discriminative texture features leads to a significant improvement in face recognition accuracy. For face verification the benefit is greatest at low FAR values; for example, at 0.01% FAR verification accuracy on the FRGCv2 2.5D data is four times greater when LBP is used. Rank 1 identification rates also show a marked improvement with, for example, a factor of eight improvement on the FRGCv2 2D images. The benefit of LBP is weakest, although still signficant, on the XM2VTS images, indicating perhaps that LBP is most useful under the less controlled conditions of the FRGC data.

8.2 Comparison of Image Modalities

In the preceeding chapters we have shown the results of various experiments using the 2D, 2.5D and 3D modalities of the FRGCv2 data set. It of interest to compare the best results that were achieved on these different modalities, using the techniques described in this thesis, and to look at the effect of carrying out a fusion of all three of them.

Fig. 8.3 shows the verification and identification curves for the 2D, 2.5D and 3D FRGCv2 data. These curves were generated using photometric normalisation (for 2D images), LBP feature extraction (for 2D and 2.5D images), LDA feature extraction, angularisation through the $ang(\cdot)$ transformation and ECOC with Gaussian SVMs. The figure also shows the results of performing a fusion of 2D, 2.5D and 3D data; for comparison the 2D and 2.5D results are also shown without LBP feature extraction.

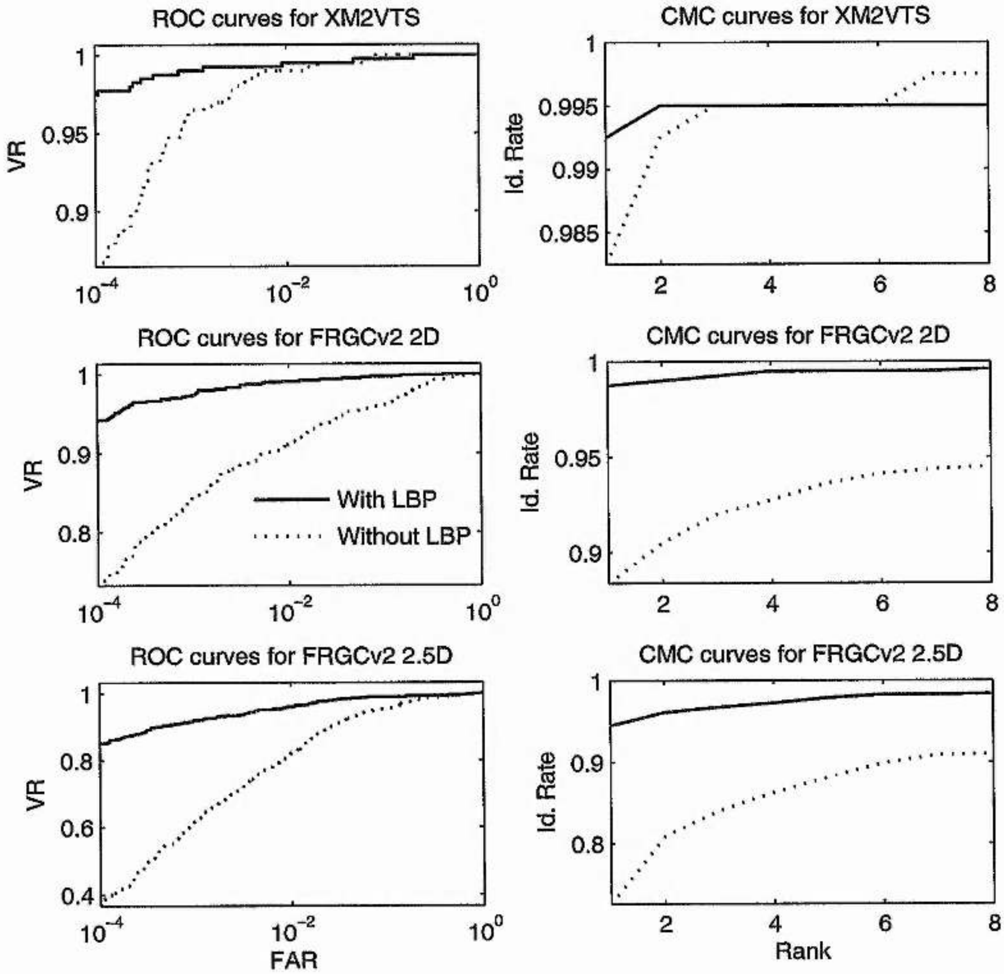

Figure 8.2: ECOC classification on 2D and 2.5D images with and without LBP preprocessing.

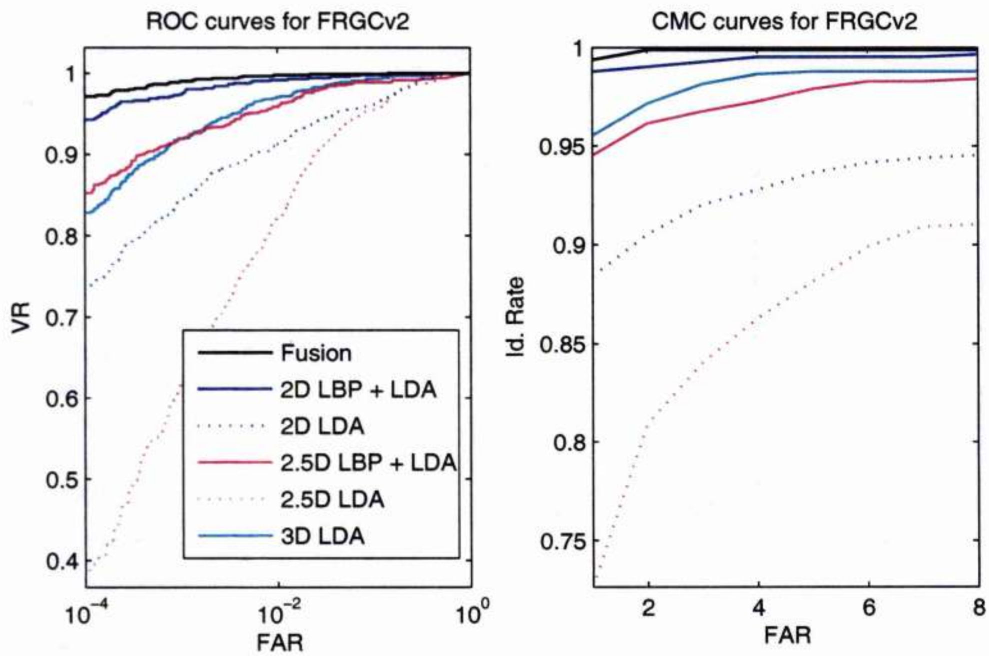

Figure 8.3: A comparison of different face image modalities on the FRGCv2 data sets. Note that $ang(\cdot)$ and ECOC is applied in all cases.

Inspection of Fig. 8.3 shows that, in the absence of LBP feature extraction, the 3D modality is considerably more discriminative than 2D which itself is significantly better than 2.5D. The application of LBP feature extraction to the latter two modalities, however, improves their performance to the extent that 2D is now more discriminative than 2.5D and 3D; the latter two modalities are now comparable in performance for face verification, whilst 3D still retains a small advantage on the identification task.

Also shown in Fig. 8.3 is the result of carrying out a fusion of all three modalities and it can be seen that this yields an even greater level of accuracy than 2D, with a verification rate of 97.0% at 0.01% FAR and a rank 1 recognition rate of 99.4%. The method adopted here is to apply decision level fusion by averaging the output scores from each of the three modalities. Although this is a simple approach it has been shown to give good results in face recognition applications [30, 35].

8.3 Benchmark Results

Table 8.1 shows the half total error rates (HTER) obtained using the XM2VTS Lausanne configuration I protocol when the validation set is used to select a threshold of equal false acceptance and false rejection error. These results bear out the claim that LDA, ECOC and angularised ECOC lead to a progressive improvement in verification accuracy and also that LBP feature extraction greatly improves on LDA alone. They also show that the angularised ECOC method, when used together with LBP and LDA, compares favourably with the best reported result in the ICB 2006 competition [45][1].

Table 8.2 shows the equal error rate (EER) values obtained on the FRGC face verification experiments 3t, 3s and 3. As noted in chapter 5, the FRGC training set was captured under conditions of uniform illumination and neutral expression whilst the

[1] The winning entry came from the Chinese Academy of Sciences. In their method images were first photometrically normalised using region-based histogram equalisation. Gabor filters at 5 scales and 8 orientations were then applied to produce feature vectors with 40 times the dimensionality of the original images. These high dimensional feature sets were then adaptively divided into sub-groups and LDA was used to train a classifier on each sub-group. The final classification decisions were made by combining the scores from these individual classifiers.

Processing applied	Evaluation set FAR	Evaluation set FRR	Evaluation set HTER	Test set FAR	Test set FRR	Test set HTER
XM2VTS 2D, neutral-neutral						
Photo + LDA	4.00	4.00	4.00	4.12	1.50	2.81
Photo + LDA + ECOC	1.24	6.67	3.95	1.37	6.00	3.68
Photo + LDA + ang + ECOC	2.44	2.33	2.39	2.50	2.00	2.25
Photo + LBP + LDA	0.99	1.00	1.00	0.96	0.75	0.86
Photo + LBP + LDA + ECOC	0.37	0.33	0.35	0.29	1.00	0.65
Photo + LBP + LDA + ang + ECOC	0.22	0.33	0.28	0.20	0.75	**0.47**
ICB 06 Best	0.80	0.80	0.80	0.96	0.00	0.48

Table 8.1: Percentage half-total error rates obtained on the XM2VTS database using the Lausanne configuation I protocol.

test set was obtained under adverse illumination and expression conditions. In this thesis we are not primarily concerned with the problem of correcting for illumination and expression variability in the test set versus the training set and these results reflect that fact. It was found on experiment 3t that the lowest verification error was obtained by disregarding the training set altogether and generating a similarity measure between probe and gallery images by taking the Manhattan distance between concatenated LBP histograms taken over 4×4 tilings of the images. Methods that rely on supervised learning to establish classifier parameters suffer from the problem of being overtrained on a non-representative data set and thus give lower performance. For this reason, the use of LDA and ECOC lead to a significant drop in performance. The interposition of an angularisation stage is able to bring the performance back somewhat, but only to a level that is comparable with the baseline algorithm. The BEE baseline algorithm itself suffers from the problem of overtraining and yields an EER value that is 45% worse than the untrained LBP histogram comparison method.

When the 2.5D modality is used in experiment 3s the problem of illumination variability is removed but noise due to expression variation in the test set remains an issue. Table 8.2 shows that, in this case, the use of supervised training is beneficial. This can be seen by comparing the results from the untrained LBP histogram comparison method with those obtained when LBP is followed by LDA. As with 2D data, however, no further improvement is to be gained by the additional application of ECOC techniques.

Experiment 3s with 3D scans gives much better results than those obtained using either 2D or 2.5D data. Furthermore, progressive improvements are obtained by the application of LDA, ECOC and angularisation. The performance of the latter algorithm improves on the BEE baseline figure for 3D by 27% and it is also 35% lower than the best figure obtained on the 2D data. These observations lead to the conclusions that not only is the 3D modality unaffected by illumination variation in the test set, but it is also more robust to the problem of expression variability than is the 2.5D modality.

The benefits of using a fusion of 2D and 3D information are also evident from Table 8.2, which shows the effect of combining the best performing algorithms for 2D and 3D data by averaging their respective similarity scores. It can be seen that the combined error rate is 28% less than that based on 3D alone and is 53% less than that based on 2D alone.

8.4 Summary

In this chapter experimental evidence has been presented to show that, for 2D and 2.5D face images, the application of multi-scale LBP feature extraction leads to a significant improvement in accuracy on both verification and identification problems. It is noteworthy that, when using similar algorithms (i.e. LDA, $ang(\cdot)$ and ECOC) the 3D modality gives greater accuracy than 2D which, in turn, is more accurate than 2.5D. When an LBP feature extraction stage is applied to the 2D and 2.5D modalities, however, their performance is improved so that 2.5D and 3D now give comparable results (with 3D being slightly more accurate on the face identification task) and both modalities are significantly outperformed by 2D.

A comparison with the best reported XM2VTS results show that the face verification algorithms described here are capable of delivering state of the art performance. When the standard FRGC Experiment 3 protocol is applied, however, the algorithms are adversely affected by the fact that they are not optimised to correct for illumination

Processing applied	Sub-experiment		
	I	II	III
Experiment 3t (2D greyscale images)			
Photo + LBP	**9.30**	**11.95**	**13.98**
Photo + LBP + LDA	13.98	16.07	17.90
Photo + LBP + LDA + ECOC	39.16	40.68	42.72
Photo + LBP + LDA + ang + ECOC	19.97	22.08	24.10
BEE Baseline	17.03	21.08	24.80
Experiment 3s (2.5D range images)			
LBP	18.68	20.81	22.74
LBP + LDA	**11.06**	**12.21**	**12.98**
LBP + LDA + ECOC	11.06	12.21	12.98
LBP + LDA + ang + ECOC	11.10	12.20	13.00
BEE Baseline	13.54	14.70	15.38
Experiment 3s (3D scans)			
LDA	9.93	11.20	11.90
LDA + ECOC	8.93	9.68	10.24
LDA + ang + ECOC	**6.09**	**6.78**	**7.18**
BEE Baseline	8.33	8.80	9.30
Experiment 3 (Fusion of shape and texture channels)			
(Photo + LBP) and (LDA + ang + ECOC)	**4.37**	**4.94**	**5.25**
BEE Baseline	7.05	7.61	8.19

Table 8.2: Percentage EER results on FRGC experiment 3.

and expression variability in the test set. Nevertheless, it has been demonstrated that the 3D modality is less affected by these problems than either the 2D or 2.5D modalities.

A fusion of image modalities has been shown to improve face recognition performance beyond that attainable by any single one of them.

Chapter 9

Conclusions and Further Work

9.1 Summary and Conclusions

This thesis began by establishing the need for accurate methods of person authentication and identification using face biometrics. In this context, drawing on the results of experiments performed on databases of 2D greyscale images, 2.5D range images and 3D facial scans, we summarise below the main contributions of this thesis and the conclusions to be drawn from this work:

- The separation performance experiments of section 6.1.2 show that, for centred face images, angular separation is a more discriminative measure of the dissimilarity between two images than either Euclidean, Manhattan or Mahalanobis distance. This was previously known to be the case for 2D images and is the reason for the success of methods such as PCA with whitened cosine and LDA with normalised correlation. This work shows that it is also true for 2.5D range images and, perhaps surprisingly, for densely registered 3D scans. The experiments also show that angular separation performs better than the other metrics over a wide range of feature extraction scenarios. These include using the raw data, photometrically normalised data (for 2D images), and the feature vectors that result from the application of combinations LBP (for 2D and 2.5D images), LDA and ECOC. The implication of this is that good angular separation is a general property of face data and is not, among the methods described in this thesis, limited to any particular representation or feature extraction technique.

- The separation performance experiments of section 6.1.2 also show that the magnitude of feature vectors is only a weak source of information for discrimination between faces. The main reason for this is that radial movement in feature space, in contrast to angular movement, does not affect the arrangement and geometry of facial features and hence is of little relevance to the problem of distinguishing between the identities of two face images.

- In chapter 4 the concept of angularisation was introduced and we proposed two novel Mercer kernel functions, known as $K_{ang}(\cdot,\cdot)$ and $K_{sph}(\cdot,\cdot)$, by which angularisation can be incorporated into the design of an SVM classifier. We also discussed two transformations, referred to as $ang(\cdot)$ and $sph(\cdot)$, by which a general feature space may be mapped into one in which non-angular metrics such as Euclidean and Manhattan are approximately in correspondence with angular separation in the original space. The separation performance experiments of section 6.1.2 show that angularisation is successful in improving the performance of the Euclidean, Manhattan and Mahalanobis metrics, with the results from the first two becoming comparable to angular separation. The experiments of section 6.3 also show that the results of applying angularisation are largely insensitive to the details of how it is accomplished since all four of the proposed methods give broadly similar results. There is, however, shown to be a slight advantage to using the $K_{ang}(\cdot,\cdot)$ kernel in face identification applications.

- The vehicle image experiments of section 6.1.2.3 suggest that the above three observations apply to a wider range of image types than just face images. However experiments with synthetic data show that they are not universally applicable to all images. It is, therefore, currently an open question as to what are the characteristics of images and data sets which make angular separation more discriminative than other commonly used distance metrics. This point is discussed further in section 9.2 below.

- The main objective in applying angularisation is to allow strong classifiers to be trained which, though they do not explicitly incorporate angular separation into their construction, can nevertheless benefit from its improved performance. For this purpose ECOC ensembles of Gaussian SVMs have been used, with the $ang(\cdot)$ transformation being applied as a pre-processing stage. These experiments are described in section 6.2 and they show that angularisation does indeed improve verification and identification accuracy on 2D, 2.5D and 3D images, with the improvement being greatest on the noisier FRGC data.

- Chapter 7 presented a study of two approaches to constructing ensembles of SVMs for face recognition purposes, namely ECOC and BHC. It was shown that, for both methods, the Gaussian SVM kernel tends to outperform the linear kernel. ECOC can gain improved accuracy by deploying an increased number of base classifiers, whereas BHC entails the use of a fixed number of classifiers as dictated by the number of target classes. Base classifier calibration, either by isotonic regression or Platt's sigmoid fitting algorithm, improves ECOC performance on both verification and identification tasks. For the BHC algorithm, however, only the sigmoid algorithm was beneficial for face verification and no calibration method was found to be beneficial for face identification. The best decoding procedures were found to be nearest neighbour for ECOC and the soft combination rule for BHC. When partitioning class families, the Bose-Chaudury-Hocquenghem algorithm was found to give slightly better results than random assignment for ECOC whilst for BHC the 2-means clustering algorithm improved on deterministic annealing.

- As shown in section 7.3, in terms of overall ensemble classification performance ECOC was significantly better than BHC. This is attributed to problems caused by the progressively smaller training set sizes associated with the lower nodes of the BHC tree, together with the fact that face data may not form a deeply nested natural hierarchy as required by the BHC algorithm.

- In section 8.1 it was confirmed that, for 2D and 2.5D images, multi-scale LBP is an excellent feature extraction method. It was also shown in section 6.2 that the performance of this technique can be improved by the application of angularisation and ECOC.

- It is sometimes stated [11] that 2.5D images or 3D scans are more reliable for face recognition than 2D images because the first two modalities overcome problems of illumination variation. The FRGC corpus is a good source of data for comparing these modalities since all images are captured at the same time and thus with the same pose angles and facial expressions. The evidence of the experiments described in section 8.2 suggests that, under similar conditions, 3D is more reliable as a means of face recognition than 2D which in turn is more accurate than 2.5D. When the more advanced method of LBP feature extraction was applied to the latter two modalities, however, it was found that 2D gave the greatest accuracy and that 2.5D gave comparable performance to 3D using just LDA (with slightly worse performance on the face identification task). A fusion of all three modalities gave greater accuracy than any single one.

- It has been shown that the methods described in this thesis are capable of delivering state-of-the art classification results. For example in section 8.3 it was shown that the perfomance of these methods on the XM2VTS database under the Lausanne Configuration I protocol was slightly better than that of the winning entry in the ICB 2006 competition[45]. These results were obtained using multi-scale LBP, LDA, angularisation and an ECOC ensemble of Gaussian SVMs, with isotonic regression being used to calibrate the SVM outputs. The winning entry from the ICB 2006 competition employed Gabor filters to create a high-dimensional feature space and then trained multiple LDA-based classifiers on different feature subsets within that space.

  It has also been shown that good results can also be obtained on the more challenging FRGCv2 data set in which the images are made harder to identify by the presence of illumination and expression variations. For example, figure 8.3 in section 8.2 shows that a fusion of the 2D, 2.5D and 3D modalities yields a verification rate of 98.6% at an FAR 0.1%, together with a rank 1 recognition rate of 99.5%. For the reasons discussed in chapter 5 these experiments make use of a new protocol (see appendix A) and hence cannot be directly compared with other published FRGC results; they can, however, be taken as an indication of the kind of performance which can be achieved through these methods. As a rough guide for comparison, it is stated in [55] that the best verification rate obtained in the 2006 FRGC experiment 3 fusion competition was 97.0% at an FAR of 0.1% but no details are given of how this result was obtained. Kakadiaris et. al. [35] has since obtained a verification rate of 97.3% at an FAR of 0.1% on the 3D modality alone by using a fusion of Haar and pyramid wavelet transforms.

9.2 Further Work

The work described in this thesis could be extended in a number of ways:

- As stated in section 9.1, it is an open question as to the circumstances under which the angular separation metric outperforms others such as Euclidean, Manhattan and Mahalanobis distance. This determines when it is beneficial to apply angularisation and when it is counterproductive. Angularisation has been shown to be of value for face recognition using three different representations and under a range of feature extraction scenarios. It has also been shown to be of use when dealing with vehicle images, although the benefits were not as pronounced and they did not become apparent until after the application of LDA (with or without LBP). It seems a reasonable hypothesis that such angular methods will be of value when dealing with data, such as face images and scans, that have a regular and repeatable structure. The high dimensionality of the input data may also be a factor. It would be useful to explore these ideas further with the aim of obtaining a more detailed characterisation of the circumstances under which angular methods are and are not beneficial in solving classification problems.
- It is clear from section 8.2 that the reason why the 2D and 3D modalities are able to equal or exceed the face recognition performance of 3D data is that the method of LBP feature extraction is available in the first two cases. In the absence of LBP processing, 3D has been shown to be a much more discriminative modality than either 2D or 2.5D. A potentially fruitful research direction, therefore, would be to investigate methods by which the concept of LBP can be generalised to apply to densely registered 3D face models. One possible way of achieving this would be

to consider the tangent plane at a vertex and assign the values 0 or 1 to each of the neighbouring vertices depending on whether they lie above or below (or on) that plane. These bit values could then be concatenated into a binary word as for greyscale LBP.

- In the separation performance experiments of section 6.1.2 it was noted that radial separation does tend to show some improvement with the application of feature extraction methods. ECOC in particular appears to be somewhat successful in separating intra-class and inter-class differences into two distinct groups. Whilst this improvement is not sufficient to allow radial separation in itself to be used as a classification metric, it may be that a fusion of radial separation with angular separation would lead to an improvement over the use of angular separation alone. Whether this is the case will depend on the degree to which the angular and radial pseudo-metrics act as independent sources of dissimilarity information.

- Fig. 6.6 of section 6.1.2.1 illustrates the kind of differences between images that occur when two images of the same subject have the maximum observed angular separation of 90°. This work could be extended by looking at different angular ranges from 0°up to 90° and attempting to characterise the type of facial difference that arises within each range. The aim of this research would be to determine the textural differences that are encoded by different bands of angular separation.

- Throughout this thesis, when angularisation has been applied as a separate feature space transformation through the $ang(\cdot)$ and $sph(\cdot)$ functions, this transformation has been delayed until just before ECOC has been applied. This approach has been adopted in order to facilitate a comparison of explicit methods of angularisation with the proposed SVM kernel functions. It would, however, be possible to move the point at which explicit angularisation is applied to a different position in the sequence of operations. In particular it would be instructive to investigate the effects of applying angularisation *before* LDA is performed so as to maximise class separation within an angularised, rather than the original feature space. It would also be worthwhile to investigate other dimensionality reduction techniques within the angularised space, for example the method of locality preserving projections

(LPP) that is described in section 2.2.

- The results shown in section 8.3 that were obtained using the standard FRGC Experiment 3 protocol were less satisfactory than those obtained by partitioning the FRGCv2 data. This may be attributed to the fact that in the former case the training set (i.e. the FRGCv1 data) was obtained under ideal controlled conditions whereas in the latter case the training set was more representative of the test data. Although homomorphic filtering and histogram equalisation were employed in these experiments, the evidence of section 8.3 is that these methods are inadequate to deal with the kind of major mismatch between training and probe data that is illustrated by Fig. 5.2.

  A number of possible approaches to improving performance on the standard experiment suggest themselves. The first method would be to transform the test images so as to correct, as far as possible, for the problems of illumination and expression variation and thus render them more similar to the training images. For example, the illumination correction method of Gross and Brajovic [23] is a promising approach to correcting for illumination variation in 2D images whilst that of Bronstein et al [13] could be used to address the problem of expression variation in 3D scans. A second possibility would be to augment the training set by a suitably large set of non-FRGC images which would be more representative of the FRGC test set. A practical difficulty here, however, would be that of obtaining sufficiently large numbers of 2.5D and 3D scans. A third improvement would be to make better use of the available colour information by treating each of the three colour channels as separate 2D images and fusing the resulting classification outputs.

Appendix A

Experimental Protocols

Chapter 5 gives a brief description of the various test databases that were used in the experiments of chapters 6 to 8. In this appendix we list in detail the constituents of the training, evaluation and test client sets together with (where applicable) the evaluation and test impostor sets for each of the FRGC, JAFFE and UIUC data sets. Information about the Lausanne Configuration I protocol for XM2VTS is supplied as part of the package available from the University of Surrey [44] and is not repeated here.

The verification and identification tests are conducted as follows. The set of valid verification claims is obtained by claiming the true client identity for each member of the test client set. The set of invalid verification claims is constructed by taking each of the test impostor images and claiming it to have the identity of each client in turn. For identification tests each test client image is ranked in order of decreasing likelihood of a match with each client identity.

A.1 FRGCv2 Database

Each image or scan in the FRGCv2 database (see Chapter 5) is assigned a unique name of the form *subject*d*session* where *subject* is a 5 digit code that uniquely identifies the subject of the image and *session* is a 1-3 digit code that distinguishes the image from all other images of the same subject. For 3D scans (including range images) the *session* code is an even number. For 2D images the *session* code is an odd number which is 1 greater than that of the corresponding 3D scan. In the following sections the constituents of the training, evaluation and test client sets, together with the evaluation and test impostor sets, are listed. Only the names of the 3D scans are shown; those of the 2D images can be obtained from the 3D names by adding 1 to each.

A.1.1 Training Client Set

802 images:

02463d546,02463d552,02463d558,02463d652,02463d658,02463d664,02463d670,04203d436,04203d442,04203d448,04203d536,04203d542,
04203d550,04203d556,04217d403,04217d409,04217d455,04217d461,04219d417,04221d431,04221d437,04221d443,04221d543,04221d549,
04221d555,04233d394,04233d400,04233d500,04233d506,04233d512,04237d141,04237d149,04237d155,04261d299,04261d329,04265d261,
04265d267,04265d339,04265d345,04273d248,04273d254,04273d290,04284d53,04284d61,04286d267,04286d273,04286d367,04286d373,
04286d379,04298d67,04298d73,04300d222,04300d228,04301d240,04301d246,04301d252,04301d258,04301d353,04301d359,04309d163,
04309d169,04309d175,04309d249,04313d56,04313d62,04319d190,04319d196,04319d266,04319d272,04319d278,04320d274,04320d280,
04320d346,04324d276,04324d282,04324d288,04324d348,04324d356,04334d304,04334d310,04334d410,04334d416,04334d422,04334d428,
04338d80,04339d226,04339d290,04339d296,04341d155,04341d191,04343d321,04343d327,04343d333,04343d427,04343d433,04343d439,
04347d291,04347d297,04347d303,04347d389,04347d395,04347d401,04350d258,04350d264,04350d322,04350d328,04361d175,04361d181,
04365d320,04365d328,04369d250,04370d225,04370d231,04370d295,04370d301,04373d54,04373d60,04378d203,04378d209,04378d231,
04381d110,04382d170,04382d176,04385d323,04385d329,04385d335,04385d431,04385d437,04385d443,04385d449,04388d285,04388d291,
04388d297,04388d373,04388d379,04388d385,04395d196,04395d202,04395d265,04395d273,04397d336,04397d342,04397d348,04397d448,
04397d454,04397d460,04400d294,04400d300,04400d306,04400d380,04400d386,04400d392,04406d90,04409d137,04410d180,04411d190,
04418d285,04418d291,04418d297,04418d386,04418d392,04418d398,04419d250,04419d256,04419d262,04419d322,04419d328,04423d190,
04423d196,04423d274,04428d241,04428d247,04429d335,04429d343,04429d443,04429d449,04429d455,04429d461,04433d184,04434d154,
04434d191,04435d342,04435d348,04435d354,04436d308,04436d314,04436d360,04436d366,04440d121,04440d91,04440d97,04446d271,
04446d277,04446d365,04446d371,04446d379,04446d385,04449d173,04449d245,04449d251,04449d257,04449d263,04456d271,04456d277,
04456d283,04456d353,04460d262,04460d326,04460d332,04460d338,04461d295,04461d301,04461d307,04461d407,04461d413,04461d419,
04471d263,04471d269,04471d285,04472d224,04472d230,04472d316,04472d324,04472d330,04475d114,04475d120,04475d128,04476d120,
04476d126,04476d214,04476d220,04476d226,04479d222,04479d262,04482d304,04482d310,04482d316,04482d410,04482d416,04482d422,
04484d187,04484d193,04485d286,04485d292,04485d298,04485d398,04485d404,04485d410,04493d202,04493d226,04495d309,04495d315,
04495d321,04495d421,04495d427,04495d433,04495d439,04502d56,04502d62,04505d220,04505d226,04505d232,04505d326,04505d332,
04508d79,04508d85,04509d276,04509d282,04509d288,04509d388,04509d394,04509d400,04512d324,04512d330,04512d428,04512d436,
04512d442,04512d448,04514d320,04514d326,04514d332,04514d432,04514d438,04514d444,04514d450,04531d285,04531d291,04531d297,
04531d391,04531d397,04535d213,04535d219,04535d259,04542d112,04542d118,04542d194,04542d200,04546d75,04553d236,04553d264,
04556d307,04556d313,04556d319,04556d411,04556d417,04556d423,04560d265,04560d271,04560d277,04560d376,04560d384,04560d390,
04560d396,04568d91,04575d294,04575d300,04575d396,04575d402,04575d410,04577d282,04577d288,04577d294,04577d300,04577d346,
04579d260,04580d293,04580d299,04580d305,04580d405,04580d411,04580d417,04580d423,04587d110,04587d116,04588d129,04588d135,
04588d231,04588d237,04588d243,04588d249,04593d196,04593d202,04593d208,04593d262,04593d268,04596d166,04596d174,04596d180,
04596d80,04596d86,04598d251,04598d257,04598d263,04598d357,04598d363,04603d135,04603d141,04603d147,04603d247,04603d253,
04603d259,04603d265,04606d178,04606d184,04609d102,04609d195,04609d201,04609d207,04609d94,04612d63,04613d176,04613d182,
04618d160,04622d232,04622d238,04622d244,04622d324,04622d332,04622d338,04626d235,04626d241,04626d247,04626d347,04626d353,
04626d359,04626d365,04629d140,04629d146,04629d152,04629d246,04629d254,04629d260,04633d180,04633d186,04633d192,04633d288,
04633d294,04633d300,04633d306,04637d196,04638d193,04641d175,04641d181,04641d245,04641d251,04644d200,04644d206,04644d212,
04644d258,04644d264,04650d148,04650d154,04652d156,04652d162,04662d123,04662d131,04667d194,04667d200,04667d206,04667d302,
04667d308,04667d314,04667d320,04673d184,04673d192,04673d277,04673d283,04673d289,04676d163,04681d149,04681d155,04681d177,
04682d120,04682d126,04682d132,04682d218,04682d224,04682d232,04683d231,04683d237,04683d243,04683d339,04683d345,04683d351,
04683d357,04688d40,04689d24,04689d30,04689d96,04691d124,04691d130,04691d138,04691d52,04695d68,04695d74,04695d80,
04696d38,04696d44,04697d182,04697d188,04697d78,04697d84,04697d90,04700d20,04701d155,04701d161,04701d167,04701d70,
04701d76,04704d18,04705d42,04707d52,04707d62,04708d163,04708d171,04708d177,04708d54,04708d60,04711d149,04711d155,
04711d161,04711d167,04711d49,04714d186,04714d192,04714d200,04714d78,04714d84,04714d90,04714d96,04717d45,04718d18,
04719d177,04719d183,04719d191,04719d80,04719d86,04719d92,04722d44,04722d50,04724d136,04724d142,04724d148,04724d42,

04724d48,04728d42,04728d94,04729d24,04729d54,04730d128,04730d134,04730d56,04730d62,04731d133,04731d139,04731d35,
04731d43,04733d34,04734d184,04734d190,04734d196,04734d80,04734d86,04737d40,04737d68,04742d78,04742d84,04742d90,
04743d134,04743d140,04743d146,04743d48,04743d54,04745d168,04745d174,04745d180,04745d76,04745d82,04745d88,04748d111,
04748d117,04748d123,04750d52,04750d58,04751d46,04753d18,04754d176,04754d182,04754d190,04754d196,04754d76,04754d82,
04754d88,04757d159,04757d165,04757d73,04757d79,04757d85,04758d129,04758d135,04758d61,04758d67,04760d168,04760d174,
04760d180,04760d80,04763d152,04763d158,04763d164,04763d66,04763d72,04766d28,04767d102,04767d108,04767d40,04770d24,
04770d48,04772d40,04773d181,04773d189,04773d195,04773d80,04773d86,04773d92,04774d158,04774d164,04774d70,04774d76,
04774d82,04777d186,04777d192,04777d200,04777d72,04777d78,04777d84,04778d48,04778d54,04779d50,04779d56,04782d195,
04782d201,04782d207,04782d213,04782d81,04782d87,04782d93,04785d68,04785d74,04785d80,04787d12,04787d54,04790d100,
04790d106,04790d46,04790d98,04792d45,04792d67,04796d142,04796d148,04796d74,04796d80,04799d36,04801d60,04801d66,
04801d72,04802d20,04803d166,04803d172,04803d178,04803d74,04803d80,04803d86,04806d40,04808d32,04809d50,04809d56,
04811d112,04811d118,04811d50,04813d138,04813d144,04813d150,04813d156,04813d46,04815d194,04815d200,04815d206,04815d212,
04815d82,04815d88,04815d94,04820d38,04821d104,04821d110,04821d44,04824d54,04824d60,04827d122,04827d128,04827d44,
04829d110,04829d116,04829d42,04829d48,04830d144,04830d80,04830d86,04830d92,04833d150,04833d156,04833d162,04833d168,
04833d52,04836d45,04836d51,04838d154,04838d160,04838d166,04838d50,04838d56,04839d180,04839d186,04839d78,04839d84,
04839d90,04842d158,04842d164,04842d172,04842d178,04842d58,04842d64,04843d155,04843d163,04843d75,04843d81,04846d144,
04846d74,04846d80,04848d138,04848d144,04848d150,04848d156,04848d46,04848d52,04850d34,04850d70,04851d160,04851d166,
04851d172,04851d178,04851d58,04852d60,04852d66,04852d72,04855d88,04855d94,04856d102,04856d88,04856d94,04857d72,
04857d78,04857d84,04859d28,04863d62,04865d102,04865d108,04865d92,04865d98,04868d76,04868d82,04869d52,04869d58,
04869d66,04870d88,04870d94,04872d42,04872d48,04873d47,04874d28,04876d70,04876d76,04880d35,04880d41,04881d104,
04881d110,04881d94,04883d100,04883d88,04883d94,04887d72,04887d78,04887d84,04888d38,04888d44,04889d57,04892d104,
04892d90,04892d98,04893d70,04893d76,04894d68,04894d76,04898d60,04898d66,04899d80,04899d86,04899d92,04900d104,
04900d92,04903d30,04903d36,04905d58,04905d64,04907d80,04907d86,04910d65,04910d73,04914d46,04915d48,04915d54,
04915d60,04917d104,04917d92,04917d98,04921d46,04922d62,04922d70,04923d48,04923d54,04923d60,04925d34,04927d38,
04927d44,04929d44,04929d50,04932d38,04932d44,04934d54,04936d100,04936d106,04936d92,04936d98

A.1.2 Evaluation Client Set

801 images:

02463d548,02463d554,02463d560,02463d654,02463d660,02463d666,02463d672,04203d438,04203d444,04203d450,04203d538,04203d546,
04203d552,04217d399,04217d405,04217d411,04217d457,04217d463,04219d419,04221d433,04221d439,04221d445,04221d545,04221d551,
04233d390,04233d396,04233d402,04233d502,04233d508,04233d514,04237d143,04237d151,04237d157,04261d301,04261d331,04265d263,
04265d269,04265d341,04265d347,04273d250,04273d256,04273d292,04284d55,04286d263,04286d269,04286d275,04286d369,04286d375,
04286d381,04298d69,04300d218,04300d224,04300d258,04301d242,04301d248,04301d254,04301d349,04301d355,04301d361,04309d165,
04309d171,04309d245,04309d251,04313d58,04319d186,04319d192,04319d198,04319d268,04319d274,04320d270,04320d276,04320d340,
04320d348,04324d278,04324d284,04324d290,04324d350,04334d300,04334d306,04334d312,04334d412,04334d418,04334d424,04334d430,
04338d82,04339d228,04339d292,04339d298,04341d157,04341d193,04343d323,04343d329,04343d335,04343d429,04343d435,04343d441,
04347d293,04347d299,04347d305,04347d391,04347d397,04347d405,04350d260,04350d266,04350d324,04350d330,04361d177,04361d195,
04365d322,04369d246,04369d252,04370d227,04370d233,04370d297,04370d303,04373d56,04373d62,04378d205,04378d211,04381d106,
04381d112,04382d172,04382d190,04385d325,04385d331,04385d337,04385d433,04385d439,04385d445,04385d451,04388d287,04388d293,
04388d299,04388d375,04388d381,04395d192,04395d198,04395d204,04395d267,04397d332,04397d338,04397d344,04397d444,04397d450,
04397d456,04397d462,04400d296,04400d302,04400d376,04400d382,04400d388,04406d86,04406d92,04409d139,04410d182,04411d192,
04418d287,04418d293,04418d299,04418d388,04418d394,04418d400,04419d252,04419d258,04419d318,04419d324,04419d330,04423d192,
04423d198,04423d276,04428d243,04429d331,04429d339,04429d345,04429d445,04429d451,04429d457,04433d180,04433d186,04434d156,
04435d338,04435d344,04435d350,04435d372,04436d310,04436d316,04436d362,04436d368,04440d123,04440d93,04440d99,04446d273,
04446d279,04446d367,04446d373,04446d381,04446d387,04449d175,04449d247,04449d253,04449d259,04456d267,04456d273,04456d279,
04456d349,04456d357,04460d264,04460d328,04460d334,04461d291,04461d297,04461d303,04461d403,04461d409,04461d415,04461d421,
04471d265,04471d271,04472d220,04472d226,04472d232,04472d318,04472d326,04472d332,04475d116,04475d124,04475d130,04476d122,
04476d128,04476d216,04476d222,04476d228,04479d224,04479d264,04482d306,04482d312,04482d406,04482d412,04482d418,04482d424,
04484d189,04485d282,04485d288,04485d294,04485d394,04485d400,04485d406,04485d412,04493d204,04493d228,04495d311,04495d317,
04495d323,04495d423,04495d429,04495d435,04495d441,04502d68,04505d216,04505d222,04505d228,04505d322,04505d328,04505d334,
04508d81,04508d87,04509d278,04509d284,04509d384,04509d390,04509d396,04512d320,04512d326,04512d332,04512d430,04512d438,
04512d444,04512d450,04514d322,04514d328,04514d334,04514d434,04514d440,04514d446,04514d452,04531d287,04531d293,04531d387,
04531d393,04531d399,04535d215,04535d221,04535d261,04542d114,04542d190,04542d196,04546d71,04553d232,04553d238,04553d266,

04556d309,04556d315,04556d321,04556d413,04556d419,04556d425,04560d267,04560d273,04560d279,04560d378,04560d386,04560d392,
04560d398,04568d93,04575d296,04575d302,04575d398,04575d404,04575d412,04577d284,04577d290,04577d296,04577d342,04577d348,
04579d262,04580d295,04580d301,04580d307,04580d407,04580d413,04580d419,04580d425,04587d112,04587d118,04588d131,04588d137,
04588d233,04588d239,04588d245,04593d192,04593d198,04593d204,04593d210,04593d264,04593d270,04596d168,04596d176,04596d182,
04596d82,04596d88,04598d253,04598d259,04598d353,04598d359,04598d365,04603d137,04603d143,04603d149,04603d249,04603d255,
04603d261,04603d267,04606d180,04606d186,04609d191,04609d197,04609d203,04609d209,04609d96,04612d65,04613d178,04613d184,
04618d162,04622d234,04622d240,04622d246,04622d326,04622d334,04626d231,04626d237,04626d243,04626d343,04626d349,04626d355,
04626d361,04629d136,04629d142,04629d148,04629d154,04629d248,04629d256,04629d262,04633d182,04633d188,04633d194,04633d290,
04633d296,04633d302,04633d308,04637d198,04638d195,04641d177,04641d241,04641d247,04641d253,04644d202,04644d208,04644d254,
04644d260,04650d144,04650d150,04652d152,04652d158,04662d119,04662d125,04662d133,04667d196,04667d202,04667d208,04667d304,
04667d310,04667d316,04667d322,04673d188,04673d273,04673d279,04673d285,04676d159,04681d145,04681d151,04681d157,04681d179,
04682d122,04682d128,04682d214,04682d220,04682d226,04682d234,04683d233,04683d239,04683d245,04683d341,04683d347,04683d353,
04688d36,04688d42,04689d26,04689d92,04689d98,04691d126,04691d134,04691d48,04695d100,04695d70,04695d76,04695d98,
04696d40,04697d178,04697d184,04697d190,04697d80,04697d86,04697d92,04700d22,04701d157,04701d163,04701d66,04701d72,
04701d78,04704d20,04705d44,04707d54,04708d159,04708d165,04708d173,04708d179,04708d56,04708d62,04711d151,04711d157,
04711d163,04711d169,04711d51,04714d188,04714d196,04714d202,04714d80,04714d86,04714d92,04717d41,04717d47,04718d20,
04719d179,04719d185,04719d193,04719d82,04719d88,04719d94,04722d46,04722d62,04724d138,04724d144,04724d150,04724d44,
04724d50,04728d44,04728d96,04729d50,04729d56,04730d130,04730d136,04730d58,04730d64,04731d135,04731d141,04731d37,
04733d30,04733d48,04734d186,04734d192,04734d198,04734d82,04737d36,04737d42,04737d70,04742d80,04742d86,04742d92,
04743d136,04743d142,04743d148,04743d50,04743d56,04745d170,04745d176,04745d72,04745d78,04745d84,04748d107,04748d113,
04748d119,04750d48,04750d54,04750d60,04751d48,04753d20,04754d178,04754d184,04754d192,04754d72,04754d78,04754d84,
04757d153,04757d161,04757d167,04757d75,04757d81,04758d125,04758d131,04758d137,04758d63,04758d69,04760d170,04760d176,
04760d76,04760d82,04763d154,04763d160,04763d166,04763d68,04766d24,04766d30,04767d104,04767d36,04767d96,04770d26,
04772d06,04772d42,04773d183,04773d191,04773d197,04773d82,04773d88,04773d94,04774d160,04774d66,04774d72,04774d78,
04777d182,04777d188,04777d194,04777d202,04777d74,04777d80,04777d86,04778d50,04778d56,04779d52,04782d191,04782d197,
04782d203,04782d209,04782d77,04782d83,04782d89,04782d95,04785d70,04785d76,04785d82,04787d50,04787d56,04790d102,
04790d42,04790d48,04792d41,04792d47,04796d138,04796d144,04796d150,04796d76,04796d82,04799d38,04801d62,04801d68,
04801d80,04802d33,04803d168,04803d174,04803d180,04803d76,04803d82,04806d36,04806d42,04808d34,04809d52,04811d108,
04811d114,04811d120,04811d52,04813d140,04813d146,04813d152,04813d42,04813d48,04815d196,04815d202,04815d208,04815d78,
04815d84,04815d90,04815d96,04820d40,04821d106,04821d112,04821d46,04824d56,04827d118,04827d124,04827d130,04827d46,
04829d112,04829d118,04829d44,04830d140,04830d146,04830d82,04830d88,04830d94,04833d152,04833d158,04833d164,04833d48,
04833d54,04836d47,04836d53,04838d156,04838d162,04838d168,04838d52,04839d176,04839d182,04839d188,04839d80,04839d86,
04839d92,04842d160,04842d168,04842d174,04842d54,04842d60,04843d151,04843d159,04843d71,04843d77,04846d140,04846d146,
04846d76,04846d82,04848d140,04848d146,04848d152,04848d42,04848d48,04850d30,04850d66,04850d72,04851d162,04851d168,
04851d174,04851d54,04851d60,04852d62,04852d68,04855d84,04855d90,04855d96,04856d84,04856d90,04856d96,04857d74,
04857d80,04859d24,04863d58,04863d64,04865d104,04865d110,04865d94,04868d72,04868d78,04868d84,04869d54,04869d62,
04870d84,04870d90,04870d96,04872d44,04873d41,04874d24,04876d66,04876d72,04876d78,04880d37,04881d100,04881d106,
04881d90,04881d96,04883d84,04883d90,04883d96,04887d74,04887d80,04887d86,04888d40,04889d51,04892d100,04892d106,
04892d92,04893d66,04893d72,04894d64,04894d70,04894d78,04898d62,04898d68,04899d82,04899d88,04900d100,04900d88,
04900d96,04903d32,04905d54,04905d60,04907d76,04907d82,04907d88,04910d67,04914d42,04914d48,04915d50,04915d56,
04917d100,04917d88,04917d94,04921d40,04921d48,04922d64,04922d72,04923d50,04923d56,04925d30,04925d36,04927d40,
04927d46,04929d46,04929d52,04932d40,04934d50,04934d58,04936d102,04936d88,04936d94

A.1.3 Evaluation Impostor Set

800 images:

04200d74,04202d438,04202d440,04202d442,04202d444,04202d446,04202d448,04202d450,04202d452,04202d454,04202d456,04202d552,
04202d554,04202d556,04202d558,04202d560,04202d562,04202d564,04202d566,04202d568,04202d570,04202d572,04202d574,04226d357,
04226d359,04226d361,04236d154,04236d156,04236d158,04236d160,04239d378,04239d380,04239d382,04239d384,04239d386,04239d388,
04239d390,04239d480,04239d482,04239d484,04239d486,04239d488,04239d490,04239d492,04239d494,04239d496,04239d498,04288d252,
04288d254,04288d256,04288d258,04288d260,04288d292,04288d294,04288d296,04288d298,04288d300,04299d191,04299d193,04311d226,
04311d228,04311d230,04311d232,04311d234,04311d236,04311d280,04311d282,04311d284,04314d53,04314d61,04321d108,04321d110,
04321d112,04321d114,04321d116,04321d118,04329d100,04329d102,04329d104,04329d106,04329d108,04329d110,04329d112,04336d291,
04336d293,04336d295,04336d297,04336d299,04336d301,04336d303,04336d393,04336d395,04336d397,04336d401,04336d403,04336d405,

04336d407,04336d409,04349d312,04349d314,04349d316,04349d318,04349d320,04349d322,04349d324,04349d326,04349d328,04349d330,
04349d406,04349d408,04349d410,04349d412,04349d414,04349d416,04349d418,04374d211,04374d213,04379d280,04379d282,04379d284,
04379d286,04379d288,04379d290,04379d292,04379d294,04379d296,04379d357,04379d359,04379d361,04379d363,04379d365,04379d367,
04394d295,04394d297,04394d299,04394d301,04394d303,04394d305,04394d307,04394d309,04394d395,04394d397,04394d399,04394d401,
04394d403,04394d405,04394d407,04394d409,04394d411,04394d413,04394d415,04394d417,04407d261,04407d263,04407d265,04407d269,
04407d271,04407d273,04407d315,04407d317,04407d319,04407d321,04412d107,04422d238,04422d240,04422d242,04422d244,04424d219,
04424d221,04424d223,04424d225,04424d227,04424d229,04424d231,04424d233,04430d271,04430d273,04430d275,04430d277,04430d279,
04430d281,04430d367,04430d369,04430d371,04430d373,04430d375,04430d377,04430d379,04430d381,04430d383,04430d385,04447d125,
04447d127,04447d129,04447d131,04447d133,04447d135,04447d137,04447d157,04447d159,04453d285,04453d287,04453d291,04453d293,
04453d295,04453d297,04453d357,04453d359,04453d361,04453d363,04453d365,04453d367,04470d289,04470d291,04470d293,04470d295,
04470d297,04470d299,04470d301,04470d303,04470d305,04470d397,04470d399,04470d401,04470d403,04470d405,04470d407,04470d409,
04470d411,04470d413,04470d415,04470d417,04470d419,04481d287,04481d289,04481d291,04481d293,04481d295,04481d297,04481d299,
04481d301,04481d395,04481d397,04481d399,04481d401,04481d405,04481d407,04481d409,04481d411,04481d413,04496d240,04496d242,
04496d244,04496d246,04496d248,04496d250,04496d252,04496d254,04496d288,04496d290,04504d105,04504d85,04507d291,04507d293,
04507d295,04507d297,04507d299,04507d301,04507d303,04507d305,04507d307,04507d309,04507d402,04507d404,04507d406,04507d408,
04507d410,04507d414,04507d418,04507d420,04507d422,04507d424,04519d204,04519d206,04519d208,04519d210,04519d212,04519d232,
04519d234,04519d236,04537d320,04537d322,04537d324,04537d326,04537d328,04537d330,04537d332,04537d410,04537d412,04537d414,
04537d416,04537d418,04537d420,04540d253,04540d255,04557d329,04557d331,04557d333,04557d335,04557d337,04557d339,04557d341,
04557d343,04557d345,04557d347,04557d443,04557d445,04557d447,04557d449,04557d451,04557d453,04557d455,04557d457,04557d459,
04557d461,04557d463,04578d22,04578d24,04585d184,04585d186,04585d188,04585d190,04585d192,04585d194,04585d266,04585d268,
04585d270,04585d272,04585d274,04585d276,04585d278,04585d280,04585d282,04595d143,04595d145,04595d147,04595d149,04595d151,
04595d153,04595d87,04595d89,04595d91,04595d93,04595d95,04595d99,04602d104,04602d106,04604d96,04604d98,04605d239,
04605d241,04605d243,04605d245,04605d247,04605d249,04605d251,04605d253,04605d255,04605d279,04605d281,04605d283,04615d178,
04615d180,04615d182,04615d184,04615d186,04615d188,04615d190,04615d192,04615d194,04615d196,04615d198,04615d200,04615d78,
04615d80,04615d82,04615d84,04615d86,04631d166,04631d168,04631d170,04631d172,04631d174,04631d176,04631d178,04631d180,
04631d182,04631d254,04631d256,04631d258,04631d260,04631d264,04631d268,04631d270,04635d47,04635d49,04643d20,04646d41,
04647d160,04647d170,04651d128,04651d130,04657d154,04657d156,04664d143,04664d145,04664d147,04664d149,04664d151,04664d153,
04664d155,04664d187,04664d189,04664d191,04675d247,04675d249,04675d251,04675d253,04675d255,04675d339,04675d341,04675d343,
04675d345,04675d347,04675d349,04675d351,04675d353,04684d226,04684d228,04684d230,04684d232,04684d234,04684d236,04684d238,
04684d240,04684d242,04684d302,04684d304,04684d306,04684d308,04684d310,04693d48,04693d50,04693d52,04693d54,04693d56,
04693d58,04693d70,04694d06,04699d106,04699d108,04699d110,04699d112,04699d114,04699d42,04699d44,04699d46,04699d48,
04703d112,04703d114,04703d116,04703d118,04703d120,04703d122,04703d42,04703d44,04703d46,04710d06,04712d101,04712d103,
04712d105,04712d47,04712d49,04712d51,04712d53,04712d97,04712d99,04713d06,04715d12,04715d14,04715d74,04715d76,
04715d78,04715d80,04715d82,04721d42,04721d44,04721d46,04721d48,04721d80,04721d82,04721d84,04726d132,04726d134,
04726d136,04726d138,04726d140,04726d142,04726d144,04726d146,04726d148,04726d150,04726d152,04726d42,04726d44,04726d46,
04732d06,04732d14,04735d30,04735d32,04735d34,04735d36,04735d38,04738d42,04738d44,04738d46,04738d48,04738d50,
04738d92,04738d94,04738d96,04738d98,04746d42,04746d44,04746d46,04746d86,04746d88,04747d100,04747d102,04747d104,
04747d106,04747d108,04747d110,04747d112,04747d36,04747d38,04747d40,04752d06,04755d18,04755d20,04756d111,04756d113,
04756d115,04756d117,04756d67,04756d69,04756d71,04756d73,04756d75,04756d77,04756d79,04756d81,04761d42,04761d44,
04761d46,04761d48,04764d24,04764d26,04764d28,04765d156,04765d158,04765d160,04765d162,04765d164,04765d166,04765d168,
04765d170,04765d172,04765d54,04765d56,04765d58,04765d60,04765d62,04769d06,04771d06,04775d138,04775d140,04775d142,
04775d72,04775d74,04775d76,04775d78,04775d80,04775d82,04775d84,04775d86,04775d88,04775d90,04783d36,04783d38,
04783d74,04783d76,04784d36,04784d38,04784d40,04784d42,04784d66,04784d68,04784d70,04788d12,04788d14,04789d12,
04789d14,04791d12,04793d12,04794d24,04794d38,04795d36,04795d38,04797d144,04797d146,04797d148,04797d150,04797d152,
04797d154,04797d156,04797d158,04797d160,04797d162,04797d164,04797d42,04797d44,04797d46,04797d48,04797d50,04797d52,
04797d54,04812d42,04812d44,04812d72,04814d18,04814d20,04814d22,04816d06,04816d60,04816d62,04816d66,04816d68,
04816d70,04819d12,04819d14,04822d144,04822d146,04822d148,04822d150,04822d152,04822d154,04822d156,04822d158,04822d160,
04822d162,04822d48,04822d50,04822d52,04822d54,04826d24,04828d06,04832d140,04832d142,04832d144,04832d146,04832d148,
04832d150,04832d152,04832d78,04832d80,04832d82,04832d84,04832d86,04832d88,04832d90,04832d92,04834d06,04837d06,
04841d140,04841d142,04841d144,04841d146,04841d148,04841d150,04841d152,04841d48,04841d50,04841d52,04844d06,04845d06,
04847d160,04847d162,04847d164,04847d166,04847d168,04847d170,04847d172,04847d174,04847d176,04847d178,04847d54,04847d56,
04847d58,04847d60,04847d62,04847d64,04854d140,04854d142,04854d144,04854d146,04854d148,04854d150,04854d152,04854d48,
04854d50,04854d52,04854d54,04854d56,04860d06,04862d06,04866d100,04866d102,04866d104,04866d84,04866d86,04866d88,
04866d90,04866d92,04866d94,04866d96,04866d98,04875d06,04877d18,04877d20,04877d22,04879d06,04882d64,04882d66,
04882d68,04882d70,04882d72,04882d74,04885d46,04885d48,04885d50,04885d52,04885d54,04890d72,04890d74,04890d76,
04890d78,04890d80,04890d82,04890d84,04895d30,04895d32,04895d34,04895d36,04895d38,04897d06,04901d72,04901d74,
04901d76,04901d78,04901d80,04901d82,04901d84,04906d06,04908d42,04908d44,04908d46,04908d48,04912d06,04916d100,
04916d82,04916d84,04916d86,04916d88,04916d90,04916d92,04916d94,04916d96,04916d98,04926d12,04926d14,04928d100,
04928d102,04928d104,04928d106,04928d90,04928d92,04928d94,04928d96,04928d98

A.1.4 Test Client Set

801 images:

02463d550,02463d556,02463d562,02463d656,02463d662,02463d668,02463d674,04203d440,04203d446,04203d452,04203d540,04203d548,
04203d554,04217d401,04217d407,04217d413,04217d459,04219d415,04221d429,04221d435,04221d441,04221d541,04221d547,04221d553,
04233d392,04233d398,04233d498,04233d504,04233d510,04237d139,04237d145,04237d153,04261d297,04261d303,04261d333,04265d265,
04265d337,04265d343,04273d246,04273d252,04273d288,04273d294,04284d57,04286d265,04286d271,04286d277,04286d371,04286d377,
04286d383,04298d71,04300d220,04300d226,04300d260,04301d244,04301d250,04301d256,04301d351,04301d357,04309d161,04309d167,
04309d173,04309d247,04309d253,04313d60,04319d188,04319d194,04319d264,04319d270,04319d276,04320d272,04320d278,04320d342,
04320d350,04324d280,04324d286,04324d346,04324d354,04334d302,04334d308,04334d314,04334d414,04334d420,04334d426,04334d432,
04338d90,04339d230,04339d294,04339d300,04341d189,04343d319,04343d325,04343d331,04343d337,04343d431,04343d437,04347d289,
04347d295,04347d301,04347d387,04347d393,04347d399,04347d407,04350d262,04350d268,04350d326,04350d332,04361d179,04361d197,
04365d324,04369d248,04370d223,04370d229,04370d235,04370d299,04370d305,04373d58,04378d201,04378d207,04378d229,04381d108,
04381d114,04382d174,04382d192,04385d327,04385d333,04385d339,04385d435,04385d441,04385d447,04388d283,04388d289,04388d295,
04388d301,04388d377,04388d383,04395d194,04395d200,04395d206,04395d269,04397d334,04397d340,04397d346,04397d446,04397d452,
04397d458,04397d464,04400d298,04400d304,04400d378,04400d384,04400d390,04406d88,04406d94,04409d165,04410d184,04411d194,
04418d289,04418d295,04418d301,04418d390,04418d396,04418d402,04419d254,04419d260,04419d320,04419d326,04423d188,04423d194,
04423d272,04423d278,04428d245,04429d333,04429d341,04429d347,04429d447,04429d453,04429d459,04433d182,04434d152,04434d189,
04435d340,04435d346,04435d352,04435d374,04436d312,04436d318,04436d364,04440d101,04440d125,04440d95,04446d269,04446d275,
04446d281,04446d369,04446d377,04446d383,04449d171,04449d177,04449d249,04449d255,04449d261,04456d269,04456d275,04456d281,
04456d351,04460d260,04460d266,04460d330,04460d336,04461d293,04461d299,04461d305,04461d405,04461d411,04461d417,04461d423,
04471d267,04471d273,04472d222,04472d228,04472d314,04472d320,04472d326,04472d334,04475d118,04475d126,04476d118,04476d124,
04476d212,04476d218,04476d224,04476d230,04479d226,04479d266,04482d308,04482d314,04482d408,04482d414,04482d420,04484d185,
04484d191,04485d284,04485d290,04485d296,04485d396,04485d402,04485d408,04485d414,04493d224,04495d307,04495d313,04495d319,
04495d325,04495d425,04495d431,04495d437,04502d54,04502d60,04505d218,04505d224,04505d230,04505d324,04505d330,04505d336,
04508d83,04509d274,04509d280,04509d286,04509d386,04509d392,04509d398,04512d322,04512d328,04512d334,04512d434,04512d440,
04512d446,04514d318,04514d324,04514d330,04514d336,04514d436,04514d442,04514d448,04531d283,04531d289,04531d295,04531d389,
04531d395,04531d401,04535d217,04535d223,04535d263,04542d116,04542d192,04542d198,04546d73,04553d234,04553d262,04556d305,
04556d311,04556d317,04556d409,04556d415,04556d421,04556d427,04560d269,04560d275,04560d281,04560d382,04560d388,04560d394,
04568d89,04568d95,04575d298,04575d394,04575d400,04575d406,04575d414,04577d286,04577d292,04577d298,04577d344,04579d258,
04580d291,04580d297,04580d303,04580d309,04580d409,04580d415,04580d421,04580d427,04587d114,04587d54,04588d133,04588d139,
04588d235,04588d241,04588d247,04593d194,04593d200,04593d206,04593d260,04593d266,04593d272,04596d170,04596d178,04596d78,
04596d84,04596d90,04598d255,04598d261,04598d355,04598d361,04603d133,04603d139,04603d145,04603d151,04603d251,04603d257,
04603d263,04606d176,04606d182,04609d100,04609d193,04609d199,04609d205,04609d92,04609d98,04612d67,04613d180,04613d188,
04618d164,04622d236,04622d242,04622d248,04622d330,04622d336,04626d233,04626d239,04626d245,04626d345,04626d351,04626d357,
04626d363,04629d138,04629d144,04629d150,04629d244,04629d250,04629d258,04633d178,04633d184,04633d190,04633d196,04633d292,
04633d298,04633d304,04637d194,04638d191,04641d173,04641d179,04641d243,04641d249,04644d198,04644d204,04644d210,04644d256,
04644d262,04650d146,04650d152,04652d154,04652d160,04662d121,04662d127,04662d135,04667d198,04667d204,04667d210,04667d306,
04667d312,04667d318,04667d324,04673d190,04673d275,04673d281,04673d287,04676d161,04681d147,04681d153,04681d159,04682d118,
04682d124,04682d130,04682d216,04682d222,04682d228,04683d229,04683d235,04683d241,04683d247,04683d343,04683d349,04683d355,
04688d36,04689d100,04689d28,04689d94,04691d122,04691d128,04691d136,04691d50,04695d66,04695d72,04695d78,04696d36,
04696d42,04697d180,04697d186,04697d192,04697d82,04697d88,04700d18,04701d153,04701d159,04701d165,04701d68,04701d74,
04701d80,04704d22,04705d46,04707d60,04708d161,04708d167,04708d175,04708d181,04708d58,04708d64,04711d153,04711d159,
04711d165,04711d47,04711d53,04714d190,04714d198,04714d204,04714d82,04714d88,04714d94,04717d43,04717d49,04718d22,
04719d181,04719d189,04719d78,04719d84,04719d90,04722d42,04722d48,04724d134,04724d140,04724d146,04724d152,04724d46,
04728d100,04728d92,04728d98,04729d52,04730d126,04730d132,04730d54,04730d60,04731d131,04731d137,04731d143,04731d39,
04733d32,04734d182,04734d188,04734d194,04734d78,04734d84,04737d38,04737d44,04742d106,04742d82,04742d88,04742d94,
04743d138,04743d144,04743d150,04743d52,04745d166,04745d172,04745d178,04745d74,04745d80,04745d86,04748d109,04748d115,
04748d121,04750d50,04750d56,04751d40,04751d60,04753d22,04754d180,04754d188,04754d194,04754d74,04754d80,04754d86,
04757d155,04757d163,04757d71,04757d77,04757d83,04758d127,04758d133,04758d59,04758d65,04760d166,04760d172,04760d178,
04760d78,04760d84,04763d156,04763d162,04763d168,04763d70,04766d26,04767d100,04767d106,04767d38,04767d98,04770d46,
04772d38,04773d179,04773d185,04773d193,04773d78,04773d84,04773d90,04773d96,04774d162,04774d68,04774d74,04774d80,
04777d184,04777d190,04777d196,04777d204,04777d76,04777d82,04777d88,04778d52,04779d48,04779d54,04782d193,04782d199,
04782d205,04782d211,04782d79,04782d85,04782d91,04785d66,04785d72,04785d78,04785d94,04787d52,04787d58,04790d104,
04790d44,04790d96,04792d43,04792d65,04796d140,04796d146,04796d72,04796d78,04796d84,04799d40,04801d64,04801d70,
04802d18,04803d164,04803d170,04803d176,04803d72,04803d78,04803d84,04806d38,04808d30,04809d48,04809d54,04811d110,
04811d116,04811d48,04811d54,04813d142,04813d148,04813d154,04813d44,04815d192,04815d198,04815d204,04815d210,04815d80,

04815d86,04815d92,04820d36,04821d102,04821d108,04821d42,04821d48,04824d58,04827d120,04827d126,04827d42,04829d108,
04829d114,04829d120,04829d46,04830d142,04830d78,04830d84,04830d90,04830d96,04833d154,04833d160,04833d166,04833d50,
04836d43,04836d49,04838d152,04838d158,04838d164,04838d48,04838d54,04839d178,04839d184,04839d192,04839d82,04839d88,
04839d94,04842d162,04842d170,04842d176,04842d56,04842d62,04843d153,04843d161,04843d73,04843d79,04846d142,04846d72,
04846d78,04848d136,04848d142,04848d148,04848d154,04848d44,04848d50,04850d32,04850d68,04851d158,04851d164,04851d170,
04851d176,04851d56,04851d62,04852d64,04852d70,04855d86,04855d92,04856d100,04856d86,04856d92,04856d98,04857d76,
04857d82,04859d26,04863d60,04865d100,04865d106,04865d90,04865d96,04868d74,04868d80,04868d86,04869d56,04869d64,
04870d86,04870d92,04870d98,04872d46,04873d43,04874d26,04876d68,04876d74,04876d80,04880d39,04881d102,04881d108,
04881d92,04881d98,04883d86,04883d92,04883d98,04887d76,04887d82,04888d36,04888d42,04889d53,04892d102,04892d88,
04892d96,04893d68,04893d74,04894d66,04894d72,04894d80,04898d64,04899d78,04899d84,04899d90,04900d102,04900d90,
04900d98,04903d34,04905d56,04905d62,04907d78,04907d84,04907d90,04910d69,04914d44,04914d50,04915d52,04915d58,
04917d102,04917d90,04917d96,04921d42,04921d50,04922d68,04922d74,04923d52,04923d58,04925d32,04927d36,04927d42,
04929d42,04929d48,04932d36,04932d42,04934d52,04934d60,04936d104,04936d90,04936d96

A.1.5 Test Impostor Set

803 images:

04201d368,04201d370,04201d372,04201d374,04201d376,04201d378,04201d434,04201d436,04201d438,04201d442,04201d444,04213d280,
04213d282,04213d338,04213d340,04213d344,04213d348,04214d155,04222d391,04222d393,04222d395,04222d397,04225d291,04225d293,
04225d295,04225d297,04225d299,04225d301,04225d303,04225d305,04225d307,04225d396,04225d398,04225d400,04225d402,04225d404,
04225d408,04267d149,04274d176,04279d283,04279d285,04279d287,04279d289,04287d45,04287d47,04287d49,04287d51,04287d53,
04297d261,04297d263,04297d265,04297d267,04297d269,04297d305,04297d307,04297d309,04297d311,04302d116,04302d142,04302d144,
04302d146,04302d148,04312d207,04312d209,04312d211,04312d215,04312d217,04312d219,04312d221,04312d223,04312d225,04322d130,
04322d132,04322d134,04322d136,04327d290,04327d292,04327d294,04327d296,04327d298,04327d300,04327d392,04327d394,04327d396,
04327d398,04327d400,04327d402,04327d404,04327d406,04327d408,04327d410,04344d245,04344d247,04344d249,04344d251,04344d253,
04344d335,04344d337,04344d339,04344d341,04344d345,04344d347,04344d349,04344d351,04344d353,04351d100,04351d102,04351d104,
04351d106,04351d108,04351d96,04351d98,04366d82,04372d269,04372d271,04372d273,04372d275,04372d277,04372d331,04372d333,
04372d335,04372d337,04372d341,04372d343,04386d159,04386d161,04386d163,04386d165,04387d322,04387d324,04387d326,04387d328,
04387d330,04387d332,04387d334,04387d336,04387d425,04387d427,04387d429,04387d431,04387d433,04387d435,04387d437,04387d439,
04387d441,04387d443,04404d209,04404d211,04404d213,04404d217,04404d219,04404d223,04408d266,04408d268,04408d270,04408d272,
04408d274,04408d276,04408d278,04408d280,04408d282,04408d360,04408d362,04408d364,04408d366,04408d368,04408d370,04408d372,
04408d374,04427d264,04427d266,04427d268,04427d270,04427d272,04427d274,04427d276,04427d278,04427d280,04427d362,04427d364,
04427d366,04427d368,04427d372,04427d374,04427d376,04427d378,04427d380,04444d206,04444d208,04444d210,04444d212,04444d214,
04444d240,04444d242,04444d244,04444d246,04451d243,04451d245,04451d247,04451d249,04451d251,04451d253,04451d255,04451d311,
04451d313,04451d315,04451d317,04451d319,04451d321,04459d74,04463d201,04463d203,04463d205,04463d207,04463d255,04463d257,
04463d259,04463d261,04473d183,04473d185,04473d187,04473d189,04473d191,04473d193,04473d195,04473d197,04473d199,04473d241,
04473d243,04473d245,04473d247,04477d103,04477d105,04477d107,04477d109,04477d111,04477d113,04477d115,04477d117,04477d159,
04477d161,04488d280,04488d282,04488d284,04488d286,04488d288,04488d290,04488d292,04488d384,04488d386,04488d388,04488d390,
04488d392,04488d394,04488d396,04488d398,04488d400,04488d402,04488d404,04503d51,04506d194,04506d196,04506d198,04506d200,
04506d202,04506d226,04506d228,04511d176,04511d178,04511d240,04511d242,04511d244,04511d246,04511d248,04511d250,04513d299,
04513d301,04513d303,04513d305,04513d307,04513d309,04513d311,04529d101,04530d313,04530d315,04530d317,04530d319,04530d321,
04530d323,04530d325,04530d327,04530d329,04530d331,04530d425,04530d427,04530d429,04530d431,04530d433,04530d437,04530d439,
04530d441,04530d443,04530d445,04530d447,04554d71,04559d310,04559d312,04559d314,04559d316,04559d318,04559d320,04559d322,
04569d280,04569d282,04569d284,04569d286,04569d288,04569d290,04569d292,04569d374,04569d376,04569d380,04569d382,04569d384,
04569d386,04572d138,04581d192,04581d194,04581d196,04581d198,04581d200,04581d202,04581d204,04581d206,04581d247,04581d249,
04581d251,04581d253,04589d238,04589d240,04589d242,04589d244,04589d246,04589d248,04589d250,04589d268,04589d270,04597d105,
04597d107,04597d121,04597d123,04600d243,04600d245,04600d247,04600d249,04600d251,04600d253,04600d255,04600d257,04600d259,
04600d351,04600d353,04600d355,04600d357,04600d359,04600d361,04600d363,04600d367,04600d369,04600d371,04608d76,04608d84,
04619d161,04619d163,04619d165,04619d167,04621d89,04624d114,04628d219,04628d221,04628d223,04628d225,04628d227,04628d229,
04628d231,04628d233,04628d275,04628d277,04628d279,04632d132,04632d134,04632d136,04632d138,04632d140,04632d142,04632d144,
04632d230,04632d232,04632d234,04632d236,04632d238,04632d240,04632d242,04632d246,04642d46,04642d48,04645d87,04645d89,
04645d91,04645d93,04645d95,04653d36,04661d167,04661d169,04661d171,04661d173,04661d175,04661d177,04669d47,04670d161,
04670d163,04670d165,04670d167,04670d169,04670d171,04670d173,04670d175,04670d177,04670d179,04670d269,04670d271,04670d273,
04670d275,04670d277,04670d279,04670d281,04670d283,04670d285,04685d49,04687d111,04687d113,04687d115,04687d117,04687d131,

04687d133,04690d42,04690d44,04690d46,04692d136,04692d138,04692d140,04692d142,04692d144,04692d146,04692d148,04692d42,
04692d44,04692d46,04698d80,04698d82,04698d84,04698d86,04698d88,04698d90,04698d92,04698d94,04702d126,04702d128,
04702d130,04702d132,04702d134,04702d136,04702d138,04702d48,04702d50,04702d52,04709d153,04709d155,04709d157,04709d159,
04709d161,04709d163,04709d165,04709d167,04709d169,04709d54,04709d56,04709d58,04709d60,04709d62,04716d24,04716d26,
04720d36,04720d38,04720d70,04720d72,04720d74,04720d76,04723d06,04725d24,04725d26,04725d28,04725d54,04727d162,
04727d164,04727d166,04727d168,04727d170,04727d172,04727d174,04727d176,04727d178,04727d180,04727d66,04727d68,04727d70,
04727d72,04727d74,04727d76,04727d78,04736d48,04736d50,04736d52,04736d70,04739d06,04740d12,04740d14,04741d06,
04744d142,04744d144,04744d146,04744d148,04744d152,04744d154,04744d156,04744d158,04744d48,04744d50,04744d52,04744d54,
04749d174,04749d176,04749d178,04749d180,04749d182,04749d184,04749d186,04749d188;04749d190,04749d192,04749d72,04749d74,
04749d76,04749d78,04749d80,04749d82,04749d84,04749d86,04749d88,04749d90,04759d29,04759d31,04759d35,04762d101,
04762d103,04762d41,04762d43,04762d93,04762d95,04762d97,04762d99,04768d144,04768d146,04768d148,04768d150,04768d152,
04768d154,04768d156,04768d158,04768d66,04768d68,04768d70,04768d72,04768d74,04768d76,04776d103,04776d105,04776d107,
04776d30,04776d32,04776d34,04776d36,04776d38,04776d95,04776d97,04776d99,04780d100,04780d102,04780d104,04780d106,
04780d108,04780d110,04780d42,04780d44,04780d46,04786d142,04786d144,04786d146,04786d148,04786d150,04786d152,04786d154,
04786d156,04786d158,04786d160,04786d48,04786d50,04786d52,04786d54,04786d56,04798d18,04798d78,04798d80,04798d82,
04798d84,04798d86,04800d12,04804d06,04805d56,04805d58,04805d60,04805d62,04805d64,04807d06,04810d119,04810d121,
04810d123,04810d129,04810d131,04810d133,04810d135,04810d42,04810d44,04810d46,04810d48,04810d50,04817d06,04818d36,
04818d38,04818d40,04818d42,04818d44,04818d46,04818d54,04823d134,04823d136,04823d138,04823d140,04823d142,04823d48,
04823d50,04823d52,04823d54,04823d56,04825d12,04825d14,04831d152,04831d154,04831d156,04831d158,04831d160,04831d162,
04831d164,04831d166,04831d168,04831d170,04831d172,04831d48,04831d50,04831d52,04831d54,04831d56,04835d06,04840d148,
04840d150,04840d152,04840d154,04840d156,04840d158,04840d160,04840d162,04840d164,04840d166,04840d168,04840d53,04840d55,
04840d57,04840d59,04840d63,04849d126,04849d128,04849d130,04849d132,04849d134,04849d136,04849d138,04849d140,04849d48,
04849d50,04849d52,04849d54,04853d146,04853d148,04853d150,04853d152,04853d156,04853d158,04853d160,04853d162,04853d164,
04853d48,04853d50,04853d52,04858d06,04861d06,04864d06,04867d72,04867d74,04867d76,04867d80,04867d82,04867d86,
04871d58,04871d60,04871d62,04871d64,04871d66,04871d68,04878d76,04878d80,04878d82,04878d84,04878d86,04878d88,
04878d90,04878d92,04878d94,04884d53,04884d55,04884d57,04884d59,04884d61,04886d06,04891d75,04891d77,04891d79,
04891d81,04891d85,04891d87,04891d89,04896d100,04896d102,04896d88,04896d90,04896d92,04896d94,04896d96,04896d98,
04902d06,04904d36,04904d38,04904d40,04904d42,04904d44,04909d06,04911d65,04911d67,04911d69,04911d73,04911d75,
04911d77,04911d79,04911d81,04918d12,04918d14,04919d06,04920d06,04924d66,04924d68,04924d70,04924d72,04924d74,
04924d76,04924d78,04931d06,04933d54,04933d56,04933d58,04933d60,04933d62,04933d64,04933d66,04933d68

A.2 JAFFE Database

Each image in the JAFFE database (see Chapter 5) is assigned a unique name of the form *subject.expression.session* where *subject* is a 2-letter code that uniquely identifies the subject of the image, *expression* is a 2-letter code followed by a version number digit that signifies the facial expression, and *session* is a 1-3 digit code that distinguishes the image from all other images of the same subject.

A.2.1 Training Client Set

57 images:

KA.AN1.39,KA.DI1.42,KA.FE1.45,KA.FE4.48,KA.HA3.31,KA.NE2.27,KA.SA2.34,KA.SU2.37,KL.AN2.168,KL.DI2.171,KL.FE1.174,KL.HA1.158,
KL.NE1.155,KL.SA1.161,KL.SU1.164,KM.AN1.17,KM.DI1.20,KM.FE2.24,KM.HA2.5,KM.NE1.1,KM.SA1.9,KM.SA5.13,KM.SU3.16,KR.AN3.85,
KR.DI3.88,KR.FE3.91,KR.NE1.71,KR.SA1.77,KR.SU1.80,MK.AN1.125,MK.DI1.128,MK.FE2.131,MK.HA1.116,MK.NE1.113,MK.SA1.119,MK.SU1.122,
NA.AN1.211,NA.DI1.214,NA.FE1.217,NA.HA1.202,NA.NE1.199,NA.SA1.205,NA.SU1.208,NM.AN1.104,NM.DI1.107,NM.FE2.111,NM.HA2.96,NM.NE2.93,
NM.SA2.99,NM.SU2.102,TM.AN2.191,TM.DI2.194,TM.FE2.197,TM.HA2.181,TM.NE2.178,TM.SA2.185,TM.SU2.188

A.2.2 Evaluation Client Set

57 images:

KA.AN2.40,KA.DI2.43,KA.FE2.46,KA.HA1.29,KA.HA4.32,KA.NE3.28,KA.SA3.35,KA.SU3.38,KL.AN3.169,KL.DI3.172,KL.FE2.175,
KL.HA2.159,KL.NE2.156,KL.SA2.162,KL.SU2.165,KM.AN2.18,KM.DI3.22,KM.FE3.25,KM.HA3.6,KM.NE2.2,KM.SA2.10,KM.SU1.14,
KR.AN1.83,KR.DI1.86,KR.FE1.89,KR.HA1.74,KR.NE2.72,KR.SA2.78,KR.SU2.81,MK.AN2.126,MK.DI2.129,MK.FE2.132,MK.HA2.117,
MK.NE2.114,MK.SA2.120,MK.SU2.123,NA.AN2.212,NA.DI2.215,NA.FE2.218,NA.HA2.203,NA.NE2.200,NA.SA2.206,NA.SU2.209,NM.AN2.105,
NM.DI3.109,NM.FE3.112,NM.HA3.97,NM.NE3.94,NM.SA3.100,NM.SU3.103,TM.AN3.192,TM.DI3.195,TM.FE3.198,TM.HA3.182,TM.NE3.179,
TM.SA3.186,TM.SU3.189

A.2.3 Evaluation Impostor Set

21 images:

UY.AN1.146,UY.AN2.147,UY.AN3.148,UY.DI1.149,UY.DI2.150,UY.DI3.151,UY.FE1.152,UY.FE2.153,UY.FE3.154,UY.HA1.137,UY.HA2.138,
UY.HA3.139,UY.NE1.134,UY.NE2.135,UY.NE3.136,UY.SA1.140,UY.SA2.141,UY.SA3.142,UY.SU1.143,UY.SU2.144,UY.SU3.145

A.2.4 Test Client Set

56 images:

KA.AN3.41,KA.DI3.44,KA.FE3.47,KA.HA2.30,KA.NE1.26,KA.SA1.33,KA.SU1.36,KL.AN1.167,KL.DI1.170,KL.DI4.173,KL.FE3.176,
KL.HA3.160,KL.NE3.157,KL.SA3.163,KL.SU3.166,KM.AN3.19,KM.FE1.23,KM.HA1.4,KM.HA4.7,KM.NE3.3,KM.SA3.11,KM.SU2.15,
KR.AN2.84,KR.DI2.87,KR.FE2.90,KR.HA2.75,KR.NE3.73,KR.SA3.79,KR.SU3.82,MK.AN3.127,MK.DI3.130,MK.FE3.133,MK.HA3.118,
MK.NE3.115,MK.SA3.121,MK.SU3.124,NA.AN3.213,NA.DI3.216,NA.FE3.219,NA.HA3.204,NA.NE3.201,NA.SA3.207,NA.SU3.210,NM.AN3.106,
NM.FE1.110,NM.HA1.95,NM.NE1.92,NM.SA1.98,NM.SU1.101,TM.AN1.190,TM.DI1.193,TM.FE1.196,TM.HA1.180,TM.NE1.177,TM.SA1.184,
TM.SU1.187

A.2.5 Test Impostor Set

22 images:

YM.AN1.61,YM.AN2.62,YM.AN3.63,YM.DI1.64,YM.DI2.65,YM.DI3.66,YM.FE1.67,YM.FE2.68,YM.FE3.69,YM.FE4.70,YM.HA1.52,YM.HA2.53,
YM.HA3.54,YM.NE1.49,YM.NE2.50,YM.NE3.51,YM.SA1.55,YM.SA2.56,YM.SA3.57,YM.SU1.58,YM.SU2.59,YM.SU3.60

A.3 UIUC Vehicle Database

Each image in (training set of) the UIUC vehicle database (see Chapter 5) is assigned a unique name of the form neg-*id* or pos-*id* where the prefix indicates whether the image is a negative example (i.e. does not contain a car) or a positive example (i.e. contains a car) respectively and *id* is a unique identifier within the set of negative or positive examples. As explained in Chapter 5, only "client" images are defined for this data set.

A.3.1 Training Client Set

350 images:

```
neg-0,neg-100,neg-103,neg-106,neg-109,neg-111,neg-114,neg-117,neg-12,neg-122,neg-125,neg-128,neg-130,neg-133,
neg-136,neg-139,neg-141,neg-144,neg-147,neg-15,neg-152,neg-155,neg-158,neg-160,neg-163,neg-166,neg-169,neg-171,
neg-174,neg-177,neg-18,neg-182,neg-185,neg-188,neg-190,neg-193,neg-196,neg-199,neg-200,neg-203,neg-206,neg-209,
neg-211,neg-214,neg-217,neg-22,neg-222,neg-225,neg-228,neg-230,neg-233,neg-236,neg-239,neg-241,neg-244,neg-247,
neg-25,neg-252,neg-255,neg-258,neg-260,neg-263,neg-266,neg-269,neg-271,neg-274,neg-277,neg-28,neg-282,neg-285,
neg-288,neg-290,neg-293,neg-296,neg-299,neg-300,neg-303,neg-306,neg-309,neg-311,neg-314,neg-317,neg-32,neg-322,
neg-325,neg-328,neg-330,neg-333,neg-336,neg-339,neg-341,neg-344,neg-347,neg-35,neg-352,neg-355,neg-358,neg-360,
neg-363,neg-366,neg-369,neg-371,neg-374,neg-377,neg-38,neg-382,neg-385,neg-388,neg-390,neg-393,neg-396,neg-399,
neg-400,neg-403,neg-406,neg-409,neg-411,neg-414,neg-417,neg-42,neg-422,neg-425,neg-428,neg-430,neg-433,neg-436,
neg-439,neg-441,neg-444,neg-447,neg-45,neg-452,neg-455,neg-458,neg-460,neg-463,neg-466,neg-469,neg-471,neg-474,
neg-477,neg-48,neg-482,neg-485,neg-488,neg-490,neg-493,neg-496,neg-499,neg-51,neg-54,neg-57,neg-6,neg-62,
neg-65,neg-68,neg-70,neg-73,neg-76,neg-79,neg-81,neg-84,neg-87,neg-9,neg-92,neg-95,neg-98,pos-1,
pos-101,pos-104,pos-107,pos-11,pos-112,pos-115,pos-118,pos-120,pos-123,pos-126,pos-129,pos-131,pos-134,pos-137,
pos-14,pos-142,pos-145,pos-148,pos-150,pos-153,pos-156,pos-159,pos-161,pos-164,pos-167,pos-17,pos-172,pos-175,
pos-178,pos-180,pos-183,pos-186,pos-189,pos-191,pos-194,pos-197,pos-2,pos-201,pos-204,pos-207,pos-21,pos-212,
pos-215,pos-218,pos-220,pos-223,pos-226,pos-229,pos-231,pos-234,pos-237,pos-24,pos-242,pos-245,pos-248,pos-250,
pos-253,pos-256,pos-259,pos-261,pos-264,pos-267,pos-27,pos-272,pos-275,pos-278,pos-280,pos-283,pos-286,pos-289,
pos-291,pos-294,pos-297,pos-3,pos-301,pos-304,pos-307,pos-31,pos-312,pos-315,pos-318,pos-320,pos-323,pos-326,
pos-329,pos-331,pos-334,pos-337,pos-34,pos-342,pos-345,pos-348,pos-350,pos-353,pos-356,pos-359,pos-361,pos-364,
pos-367,pos-37,pos-372,pos-375,pos-378,pos-380,pos-383,pos-386,pos-389,pos-391,pos-394,pos-397,pos-4,pos-401,
pos-404,pos-407,pos-41,pos-412,pos-415,pos-418,pos-420,pos-423,pos-426,pos-429,pos-431,pos-434,pos-437,pos-44,
pos-442,pos-445,pos-448,pos-450,pos-453,pos-456,pos-459,pos-461,pos-464,pos-467,pos-47,pos-472,pos-475,pos-478,
pos-480,pos-483,pos-486,pos-489,pos-491,pos-494,pos-497,pos-5,pos-501,pos-504,pos-507,pos-51,pos-512,pos-515,
pos-518,pos-520,pos-523,pos-526,pos-529,pos-531,pos-534,pos-537,pos-54,pos-542,pos-545,pos-548,pos-56,pos-59,
pos-61,pos-64,pos-67,pos-7,pos-72,pos-75,pos-78,pos-80,pos-83,pos-86,pos-89,pos-91,pos-94,pos-97
```

A.3.2 Evaluation Client Set

350 images:

```
neg-1,neg-101,neg-104,neg-107,neg-11,neg-112,neg-115,neg-118,neg-120,neg-123,neg-126,neg-129,neg-131,neg-134,
neg-137,neg-14,neg-142,neg-145,neg-148,neg-150,neg-153,neg-156,neg-159,neg-161,neg-164,neg-167,neg-17,neg-172,
neg-175,neg-178,neg-180,neg-183,neg-186,neg-189,neg-191,neg-194,neg-197,neg-2,neg-201,neg-204,neg-207,neg-21,
neg-212,neg-215,neg-218,neg-220,neg-223,neg-226,neg-229,neg-231,neg-234,neg-237,neg-24,neg-242,neg-245,neg-248,
neg-250,neg-253,neg-256,neg-259,neg-261,neg-264,neg-267,neg-27,neg-272,neg-275,neg-278,neg-280,neg-283,neg-286,
```

neg-289,neg-291,neg-294,neg-297,neg-3,neg-301,neg-304,neg-307,neg-31,neg-312,neg-315,neg-318,neg-320,neg-323,
neg-326,neg-329,neg-331,neg-334,neg-337,neg-34,neg-342,neg-345,neg-348,neg-350,neg-353,neg-356,neg-359,neg-361,
neg-364,neg-367,neg-37,neg-372,neg-375,neg-378,neg-380,neg-383,neg-386,neg-389,neg-391,neg-394,neg-397,neg-4,
neg-401,neg-404,neg-407,neg-41,neg-412,neg-415,neg-418,neg-420,neg-423,neg-426,neg-429,neg-431,neg-434,neg-437,
neg-44,neg-442,neg-445,neg-448,neg-450,neg-453,neg-456,neg-459,neg-461,neg-464,neg-467,neg-47,neg-472,neg-475,
neg-478,neg-480,neg-483,neg-486,neg-489,neg-491,neg-494,neg-497,neg-5,neg-52,neg-55,neg-58,neg-60,neg-63,
neg-66,neg-69,neg-71,neg-74,neg-77,neg-8,neg-82,neg-85,neg-88,neg-90,neg-93,neg-96,neg-99,pos-10,
pos-102,pos-105,pos-108,pos-110,pos-113,pos-116,pos-119,pos-121,pos-124,pos-127,pos-13,pos-132,pos-135,pos-138,
pos-140,pos-143,pos-146,pos-149,pos-151,pos-154,pos-157,pos-16,pos-162,pos-165,pos-168,pos-170,pos-173,pos-176,
pos-179,pos-181,pos-184,pos-187,pos-19,pos-192,pos-195,pos-198,pos-20,pos-202,pos-205,pos-208,pos-210,pos-213,
pos-216,pos-219,pos-221,pos-224,pos-227,pos-23,pos-232,pos-235,pos-238,pos-240,pos-243,pos-246,pos-249,pos-251,
pos-254,pos-257,pos-26,pos-262,pos-265,pos-268,pos-270,pos-273,pos-276,pos-279,pos-281,pos-284,pos-287,pos-29,
pos-292,pos-295,pos-298,pos-30,pos-302,pos-305,pos-308,pos-310,pos-313,pos-316,pos-319,pos-321,pos-324,pos-327,
pos-33,pos-332,pos-335,pos-338,pos-340,pos-343,pos-346,pos-349,pos-351,pos-354,pos-357,pos-36,pos-362,pos-365,
pos-368,pos-370,pos-373,pos-376,pos-379,pos-381,pos-384,pos-387,pos-39,pos-392,pos-395,pos-398,pos-40,pos-402,
pos-405,pos-408,pos-410,pos-413,pos-416,pos-419,pos-421,pos-424,pos-427,pos-43,pos-432,pos-435,pos-438,pos-440,
pos-443,pos-446,pos-449,pos-451,pos-454,pos-457,pos-46,pos-462,pos-465,pos-468,pos-470,pos-473,pos-476,pos-479,
pos-481,pos-484,pos-487,pos-49,pos-492,pos-495,pos-498,pos-50,pos-502,pos-505,pos-508,pos-510,pos-513,pos-516,
pos-519,pos-521,pos-524,pos-527,pos-53,pos-532,pos-535,pos-538,pos-540,pos-543,pos-546,pos-549,pos-57,pos-6,
pos-62,pos-65,pos-68,pos-70,pos-73,pos-76,pos-79,pos-81,pos-84,pos-87,pos-9,pos-92,pos-95,pos-98

A.3.3 Test Client Set

350 images:

neg-10,neg-102,neg-105,neg-108,neg-110,neg-113,neg-116,neg-119,neg-121,neg-124,neg-127,neg-13,neg-132,neg-135,
neg-138,neg-140,neg-143,neg-146,neg-149,neg-151,neg-154,neg-157,neg-16,neg-162,neg-165,neg-168,neg-170,neg-173,
neg-176,neg-179,neg-181,neg-184,neg-187,neg-19,neg-192,neg-195,neg-198,neg-20,neg-202,neg-205,neg-208,neg-210,
neg-213,neg-216,neg-219,neg-221,neg-224,neg-227,neg-23,neg-232,neg-235,neg-238,neg-240,neg-243,neg-246,neg-249,
neg-251,neg-254,neg-257,neg-26,neg-262,neg-265,neg-268,neg-270,neg-273,neg-276,neg-279,neg-281,neg-284,neg-287,
neg-29,neg-292,neg-295,neg-298,neg-30,neg-302,neg-305,neg-308,neg-310,neg-313,neg-316,neg-319,neg-321,neg-324,
neg-327,neg-33,neg-332,neg-335,neg-338,neg-340,neg-343,neg-346,neg-349,neg-351,neg-354,neg-357,neg-36,neg-362,
neg-365,neg-368,neg-370,neg-373,neg-376,neg-379,neg-381,neg-384,neg-387,neg-39,neg-392,neg-395,neg-398,neg-40,
neg-402,neg-405,neg-408,neg-410,neg-413,neg-416,neg-419,neg-421,neg-424,neg-427,neg-43,neg-432,neg-435,neg-438,
neg-440,neg-443,neg-446,neg-449,neg-451,neg-454,neg-457,neg-46,neg-462,neg-465,neg-468,neg-470,neg-473,neg-476,
neg-479,neg-481,neg-484,neg-487,neg-49,neg-492,neg-495,neg-498,neg-50,neg-53,neg-56,neg-59,neg-61,neg-64,
neg-67,neg-7,neg-72,neg-75,neg-78,neg-80,neg-83,neg-86,neg-89,neg-91,neg-94,neg-97,pos-0,pos-100,
pos-103,pos-106,pos-109,pos-111,pos-114,pos-117,pos-12,pos-122,pos-125,pos-128,pos-130,pos-133,pos-136,pos-139,
pos-141,pos-144,pos-147,pos-15,pos-152,pos-155,pos-158,pos-160,pos-163,pos-166,pos-169,pos-171,pos-174,pos-177,
pos-18,pos-182,pos-185,pos-188,pos-190,pos-193,pos-196,pos-199,pos-200,pos-203,pos-206,pos-209,pos-211,pos-214,
pos-217,pos-22,pos-222,pos-225,pos-228,pos-230,pos-233,pos-236,pos-239,pos-241,pos-244,pos-247,pos-25,pos-252,
pos-255,pos-258,pos-260,pos-263,pos-266,pos-269,pos-271,pos-274,pos-277,pos-28,pos-282,pos-285,pos-288,pos-290,
pos-293,pos-296,pos-299,pos-300,pos-303,pos-306,pos-309,pos-311,pos-314,pos-317,pos-32,pos-322,pos-325,pos-328,
pos-330,pos-333,pos-336,pos-339,pos-341,pos-344,pos-347,pos-35,pos-352,pos-355,pos-358,pos-360,pos-363,pos-366,
pos-369,pos-371,pos-374,pos-377,pos-38,pos-382,pos-385,pos-388,pos-390,pos-393,pos-396,pos-399,pos-400,pos-403,
pos-406,pos-409,pos-411,pos-414,pos-417,pos-42,pos-422,pos-425,pos-428,pos-430,pos-433,pos-436,pos-439,pos-441,
pos-444,pos-447,pos-45,pos-452,pos-455,pos-458,pos-460,pos-463,pos-466,pos-469,pos-471,pos-474,pos-477,pos-48,
pos-482,pos-485,pos-488,pos-490,pos-493,pos-496,pos-499,pos-500,pos-503,pos-506,pos-509,pos-511,pos-514,pos-517,
pos-52,pos-522,pos-525,pos-528,pos-530,pos-533,pos-536,pos-539,pos-541,pos-544,pos-547,pos-55,pos-58,pos-60,
pos-63,pos-66,pos-69,pos-71,pos-74,pos-77,pos-8,pos-82,pos-85,pos-88,pos-90,pos-93,pos-96,pos-99

Bibliography

[1] Agarwal S, Awan A, Roth D. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 26, Issue 11, pp:1475 - 1490, Nov. 2004.

[2] Ahonen T, Hadid A, Pietikainen M. Face recognition with local binary patterns. In *Proceedings of the European Conference on Computer Vision*, pages 469-481, Prague, 2004.

[3] Allwein EL, Schapire RE, Singer Y. Reducing multiclass to binary: A unifying approach for margin classifiers. *Proc. 17th International Conf. on Machine Learning*, pages 9-16. Morgan Kaufmann, San Francisco, CA, 2000.

[4] Bishop MC. *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995.

[5] Blantz V, Vetter T. A Morphable Model for the Synthesis of 3D Faces. SIGGRAPH pp. 187-194, 1999.

[6] Blantz V, O'Toole AJ, Vetter T, Wild HA. On the other side of the mean: The perception of dissimilarity in human faces. Perception, volume 29, pp 885-891, 2000.

[7] Blantz V, Vetter T. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 25, No. 9, pp. 1063-1074. 2003

[8] Belhumeur V, Hespanda J, Kiregeman D. Eigenfaces vs. Fisherfaces: Recognition

Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 19, No. 7, pp. 711-720. July 1997.

[9] Bolle MR, Connell JH, Pankanti S, Ratha NK, Senior AW. Guide to Biometrics. Springer Verlag. 2004.

[10] Boughorbel S, Tarel J-P, Fleuret F. Non-Mercer Kernel for SVM Object Recognition. *British Machine Vision Conference*, Sptember 7-9, 2004.

[11] Bowyer KW, Chang K, Flynn JP. A survey of 3D and multi-modal 3D+2D face recognition. *Notre Dame Department of Computer Science and Engineering Technical Report*, January 2004.

[12] Breiman L. Random Forests. *Journal of Machine Learning*, Vol. 45, No. 1, pp. 5-32, Springer, October 2001.

[13] Bronstein AM, Bronstein MM, Kimmel R. Three-dimensional face recognition. *International Journal of Computer Vision*, Vol. 64, issue 1, pp. 5-30, August 2005.

[14] Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. *Knowledge Discovery and Data Mining*, 2(2), 1998.

[15] Carey S. Becoming a face expert. *Philos Trans R Soc Lond Series B,* 335: 95-103, 1992.

[16] Chan CH, Kittler J, Messer K. Multi-Scale Local Binary Pattern Histograms for Face Recognition. *Proc. 2nd International Conference on Biometrics*, Springer, pp. 809-818, Aug, 2007.

[17] Chen LF, Ko MT, Lin JC, Yu GJ. A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition, Vol. 33, Issue 10, pp.1713-1726, 2000.

[18] Dietterich TG, Bakiri G. Error-correcting output codes: A general method for improving multiclass inductive learning programs. *Proceedings of the Ninth National Conference on Artificial Intelligence* (AAAI-91) (pp. 572-577). Anaheim, CA: AAAI Press, 1991.

[19] Dietterich TG, Bakiri G. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research* 2: 263-286, 1995.

[20] Freund Y, Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, Vol 55, pp. 119-139. 1997.

[21] Goodwin B. Biometric card scheme will thwart identity theft, says Blunkett. *Computer Weekly*, 27th April 2004.

[22] Grant I. Social engineering attacks on the rise. *Computer Weekly*, 27th November 2007.

[23] Gross R, Brajovic V. An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. *4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp: 10-18, June 2003.

[24] Guo B, Lam K-M, Lin K-H, Siu W-C. Human face recognition based on spatially weighted Hausdorff distance. *Pattern Recognition Letters*, Volume 24, Number 1, pp. 499-507(9), January 2003.

[25] Haykin S. *Neural Networks, A Comprehensive Foundation*, second edition, Prentice Hall, New York, 1999.

[26] Ham J, Chen Y, Crawford MM, Ghosh J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 3, pp. 492- 501, March 2005.

[27] He X, Yan S, Hu Y, Niyogi P, Zhang H-J. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 27 No. 3, pp. 328- 340, March 2005.

[28] Hearst MA. Support Vector Machines. *IEEE Intelligent Systems, Trends & Controversies*.13(4): 18-28,1998

[29] Huang CM, Lee YJ. Automatic Model Selection for Support Vector Machines. Technical Report, Laboratory of Data Mining, Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipai, Taiwan. 2005.

[30] Husken M, Brauckmann M, Gehlen S, Von der Malsburg C. Strategies and Benefits of Fusion of 2D and 3D Face Recognition. Proc. 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 3, pp 174-181, 2005.

[31] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 15, Issue 9, pp:850 - 863, Sept. 1993.

[32] Hsu C-W, Lin C-J. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, pp. 415-425, Mar 2002.

[33] James G. Majority Vote Classifiers: Theory and Applications. PhD Dissertation, Stanford University, 1998.

[34] Jonsson K, Matas J, Kittler J, Li YP. Learning support vectors for face verification and recognition. Proceedings *Automatic Face and Gesture Recognition, 2000.* pp:208 - 213.

[35] Kakadiaris IA, Passalis G, Toderici G, Murtuza N, Theoharis T. "3D Face Recognition". *Proc. British Machine Vision Conference*, pp. 200-208, Edinburgh, United Kingdom, September 4-7 2006.

[36] Kim T-K, Arandjelovic O, and Cipolla R. Learning over Sets Using Boosted Manifold Principal Angles (BoMPA). In *Proc. of British Machine Vision Conference*, 2:779-788, 2005.

[37] Kittler J, Ghaderi R, Windeatt T, Matas J. Face Verification via Error Correcting Output Codes, British Machine Vision Conference, BMVC01, Birmingham, UK, pp 593-602, Sept 2001.

[38] Kumar S. Modular Learning Through Output Space Decomposition. *Ph.D. Dissertation, Department of Electrical and Computer Engineering, The University of Texas at Austin*, December 2000.

[39] Kumar S, Ghosh J, Crawford MM. Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis. *Pattern Analysis and Applications, spl. Issue on Fusion of Multiple Classifiers*. Vol. 5, No. 2, pp. 210-220. 2002.

[40] Lee JH, Lin CJ. Automatic model selection for support vector machines. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, 2000

[41] Li SZ, Zhao C, Ao M, Lei Z. Learning to Fuse 3D+2D Based Face Recognition at Both Feature and Decision Levels. AMFG pages 44-54, 2005.

[42] Liu C. The Bayes Decision Rule Induced Similarity Measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 29, Issue: 6 pp: 1086-1090, June 2007.

[43] Lyons MJ, Akamatsu S, Kamachi M, Gyoba J. Coding Facial Expressions with Gabor Wavelets. Proceedings, *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, April 14-16 1998.

[44] Messer K, Matas J, Kittler J, Luettin J and Maitre G. XM2VTSDB: The Extended M2VTS Database. In *Proc. of AVBPA'99*, pp. 72-77, 1999. http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/

[45] Messer K, Kittler J, Short J, Heusch G, Cardinaux F, Marce S, Rodriguez Y, Shan S, Su Y, Gao W and Chen X. Performance Characterisation of Face Recognition Algorithms and Their Sensitivity to Severe Illumination Changes. *Proceedings ICB 2006*, Hong Kong, China, January 5-7, 2006, pp. 1 - 11.

[46] Moghaddam B, Pentland A. Beyond Euclidean Eigenspaces: Bayesian Matching for Visual Recognition. *Face Recognition: From Theories to Applications*, Springer-Verlag, Berlin, 1998.

[47] Moghaddam B, Jebara T, Pentland AP. Bayesian Face Recognition. *Pattern Recognition*, Vol. 33, No. 11, pps. 1771-1782, November, 2000.

[48] Moon H, Phillips PJ. Computational and Performance Aspects of PCA-Based Face Recognition Algorithms. *Perception*, vol. 30, pp. 303-321, 2001.

[49] Morgan JT, Henneguelle A, Ham J, Ghosh J, Crawford MM. Adaptive Feature Spaces for Land Cover Classification with Limited Ground Truth Data. *Intl. Jl. of Pattern Recognition and Artificial Intelligence*, Spl. Issue on Fusion of Multiple Classifiers, J. Kittler and F. Roli (Eds.), 18(5), pp. 777-800, Aug 2004.

[50] Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 24, Issue 7, Page(s): 971-987, 2002.

[51] Oppenheim AV, Lim JS. The importance of phase in signals. *Proc. IEEE*, vol 69, No. 5, pp. 529-541, May 1981.

[52] Perlibakas V. Distance measures for PCA-based face recognition. Pattern Recognition Letters, Volume 25 , Issue 6, Pages: 711 - 724, April 2004.

[53] Petrou M, Bosdogianni P. Image Processing: The Fundamentals, John Wiley, 2000.

[54] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Jin Chang, Hoffman K, Marques J, Jaesik Min, Worek W. Overview of the face recognition grand challenge. In *Computer Vision and Pattern Recognition* (CVPR), 2005. page(s): 947- 954 vol. 1.

[55] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Worek W. Preliminary Face Recognition Grand Challenge Results. In: *Automatic Face and Gesture Recognition*, 2006. FGR 2006, page(s): 15 - 24.

[56] Platt J. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. *In Advances in Large Margin Classifiers*, pp. 61-74, A.J. Smola, P. Bartlett, B. Scholkopf, D. Schuurmans, eds., MIT Press, 1999.

[57] Poh N. Kittler J, Smith R, Tena JR. A Method for Estimating Authentication Performance over Time, with Applications to Face Biometrics. *Proc. 12th Ibero-american Congress on Pattern Recognition (CIARP)*, LNCS 4756, Springer, pp 360-369, 2007.

[58] Qin AK, Suganthan PN. Robust growing neural gas algorithm with application in cluster analysis. *Neural Networks*, Volume 17, Issues 8-9, pp. 1135-1148, October-November 2004.

[59] Rajan S, Ghosh J. An Empirical Comparison of Hierarchical vs. Two-Level Approaches to Multiclass Problems. In F. Roli, J. Kittler and T. Windeatt editors, *Multiple Classifier Systems 2004*, LNCS 3077, Springer, pp. 283-292, 2004.

[60] Rhodes G, Carey S, Byatt G, Proffitt F. Coding spatial variations in faces and simple shapes: a test of two models. *Vision Research*, 38:2307-232 1, 1998.

[61] Rose K. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *IEEE Trans. On Information theory* Vol. 86, No. 11, pp. 2210-2239, Nov. 1998.

[62] Roweis ST, Saul LK. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, vol. 290, pp. 2323-2326, Dec. 2000.

[63] Russ T, Boehnen C, Peters T. 3D Face Recognition Using 3D Alignment for PCA. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, Page(s):1391 - 1398, 2006.

[64] Russell R, Sinha P. Perceptually-based Comparison of Image Similarity Metrics. *Massachusetts Institute of Technology AI Memo 2001-014*, July 2001

[65] Sadeghi M, Kittler J, Kostin A, Messer K. A Comparative Study of Automatic Face Verification Algorithms on the BANCA Database. Proceedings AVBPA 2003, Guildford, UK, June 9-11, 2003, pp. 1 - 11.

[66] Santini S, Jain R. Similarity Measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume 21, Issue 9, pp:871 - 883 , Sept. 1999.

[67] Savvides M, Kumar BVKV, Khosla PK. Eigenphases vs eigenfacesPattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on Pattern Recognition. Volume 3, pp:810 - 813, Aug. 2004.

[68] Shawe-Taylor J, Christianini N. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.

[69] Short J, Kittler J, Messer K. A comparison of photometric normalisation algorithms for face verification. *Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp:254-259 , 2004.

[70] Smith RS, Kittler J, Hamouz M, Illingworth J. Face Recognition Using Angular LDA and SVM Ensembles. *Proc. 18th IEEE Int. Conf. on Pattern Recognition (ICPR)*, pp. 1008-1012, 2006.

[71] Smith RS, Kittler J. Angularisation for 2D and 3D Face Recognition. *IEEE Transactions on Information Forensics and Security.* Submitted.

[72] Smith RS, Windeatt T. Decoding Rules for ECOC. *Proc. 6th Int. Workshop Multiple Classifier Systems* (MCS), LNCS 3541, Springer, pp 53-63, 2005.

[73] Sudan M. Coding theory: tutorial & survey. *Proceedings. 42nd IEEE Symposium on Foundations of Computer Science* pp. 36 - 53, 2001.

[74] Tena JR, Hamouz M, Hilton A, Illingworth J. A Validated Method for Dense Nonrigid 3D Face Registration. *IEEE Int. Conf. on Video and Signal Based Surveillance (AVSS)*, p. 81, 2006.

[75] Tena JR, Smith RS, Hamouz M, Kittler J, Hilton A, Illingworth I. 2D Face Pose Normalisation Using a 3D Morphable Model. *Proc. IEEE Int. Conf. on Advanced Video and Signal based Surveillance (AVSS)*, pp 51-56, 2007.

[76] Tenenbaum JB, de Silva V, Langford JC. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, vol. 290, pp. 2319-2323. Dec. 2000.

[77] Turk M, Pentland A. Eigenfaces for recognition. J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[78] Valentine T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quart J Exp Psychol*; 43A:161-204, 1991.

[79] Vijaya Kumar BVK, Savvides M, Xie C. Correlation Pattern Recognition for Face Recognition Proceedings of the IEEE Volume 94, Issue 11, pp:1963 - 1976 , Nov. 2006.

[80] Webb A. *Statistical Pattern Recognition*, second edition, Wiley, Chichester, 2002.

[81] Xing EP, Ng AY, Jordan MI, Russell S. Distance metric learning, with application to clustering with side-information. *Advances in NIPS*, vol. 16, pp: 521-628, 2002.

[82] Yambor W, Draper B, Beveridge R. Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures. *2nd Workshop on Empirical Evaluation in Computer Vision*, Dublin, Ireland, July 2000.

[83] Yang MH. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods. *Proc. Fifth International Conf. Automatic Face and Gesture Recognition*, pp. 215-220, May 2002.

[84] Yu H,Yang J. A Direct LDA Algorithm for High-Dimensional Data-with Application to Face Recognition. Pattern Recognition, Vol. 34, pp. 2067-2070, 2001.

[85] Zadrozny B, Elkan C. Transforming classifier scores into accurate multiclass probability estimates. In: *Knowledge Discovery and Data Mining*, pp. 694-699, 2002

[86] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. *ACM Computing Surveys* (CSUR) Volume 35 , Issue 4 Pages: 399 - 458. December 2003.