# Automatic Face Recognition using Radial Basis Function Networks

Andrew Jonathan Howell

CSRP

September

ISSN

UNIVERSITY OF

Cognitive Science Research Papers

# Automatic Face Recognition using Radial Basis Function Networks

**Andrew Jonathan Howell**

## Summary

It is we

# Acknowledgements

First I would like to thank my supervisor Hilary Buxton for her constant advice support and inspiration throughout the thesis

My study for this work was made more enjoyable by the lively company and thought provoking environment within COGS for which I am very grateful

Special thanks must go to all at the Vision Group at Queen Mary and West eld College London for their invaluable discussion group but particularly to Shaogang Gong and Stephen

# Preface

Some parts of the work reported here appeared in the following publications

Howell A J Buxton H a Invariance in radial basis function neural networks in human face classi cation *Neural Processing Letters 2* b

Howell A J Buxton H b Receptive eld functions for face recognition In

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This thesis investigates the task of automatic recognition of human faces in dynamic environments

## 1.1 Computational Approaches

Many vision researchers following Marr believed that the ultimate product of any visual system was some type of *three-dimensional* reconstruction of its environment Although the lower levels of Marr s visual pipeline scheme were clearly defined the specific detail for higher level processes such as visual recognition of 3 D objects were quite vague This was mainly due to lack of evidence as the computational effort required to implement full object recognition schemes was not available at that time Once such systems using full 3 D models of objects were used to carry out useful recognition tasks it became clear that representations simpler than full 3 D reconstruction may be more appropriate and make the task more computationally tractable

The common tool of computer vision research is the video camera which will only ever give a *two-dimensional* capable of some high level mental manipulation and visualisation it may well be that our everyday visual processing of objects is done using simpler representations and reasoning

The issue of *invariance* has to be considered carefully for any task though this rarely needs to

Chapter describes the task of face recognition in unconstrained environments in detail and draws up speci c requirements to ful ll it Previous work on face recognition is then examined looking at computational models and psychological and psychophysical evidence about face recog nition in biological vision systems This is followed with discussion on how task requirements affect the suitability of techniques and a direct comparison of performance and generalisation in several approaches using the same face database using published research results and our own experimental data

Chapters to give details of the ve main experimental areas of research Chapter introduces our pose varying Sussex database and discusses methods for face representation normalisation and preprocessing techniques Variations in face images are also studied to analyse how they affect recognition performance with particular reference to the Euclidean distance measure for image comparisons The contribution of the radial basis function RBF network is also analysed and compared with related classi ers

Chapter explores the generalisation properties of the RBF network looking speci cally at pose scale and shift invariance This is important as it determines the accuracy of face segmenta tions required for data to be learnt or recognised

Chapter presents experimental work using a novel variant of the RBF network the Face Unit network which learns to identify one particular individual only This is useful for future applications as it gives an alternative parallel method of learning tasks which can then be used as additional evidence for identity

Chapter explains how the RBF network can be applied to image sequences The data used here was much taken from a much less constrained environment than the other face recognition databases so that the suitability of the proposed approach to real life applications could be assessed

Chapter explores the temporal learning abilities of the RBF network We focus on simple behaviours based on head rotation using a Time Delay variant of the network to give a fast and effective classi cation over time within image sequences

Chapter concludes the thesis summarising contributions to the eld of automatic face recog nition and discussing directions and issues for future work

In addition there are three appendices giving technical details to support the experimental

# Chapter 2

# Background

---

This chapter rst outlines our task requirements We then go on to survey general theories of object recognition including a review of psychological evidence and computational research within face recognition from the perspective of acquisition representation and reasoning The nal section will apply our proposed face recognition scheme to a standard database giving comparisons with published results for other approaches

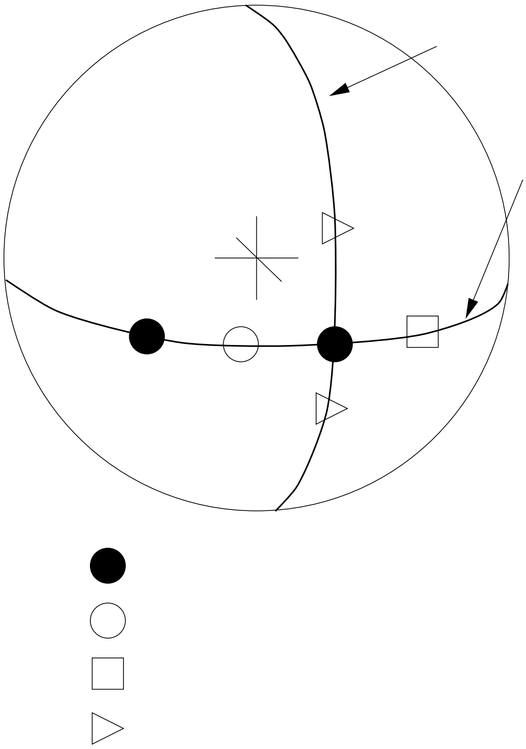
The particular face recognition task considered here concerns a known group of people in an indoor environment such as a domestic living room Within such a task it cannot be assumed that there will be clear frontal views of faces at all times Therefore it is important not to lose

would be required to monitor day to day events and allow some behavioural reasoning to help with ambiguous data This could allow expectations of who is likely to be present at a particular time of day and to assess the likelihood of encountering unknown people and conduct re learning of the database of distinct individuals known and unknown as required

## 2.1 Task Requirements

The requirements for a useful commercial face recognition and identity logging system for small groups of known individuals in busy unconstrained environments such as domestic living rooms or of ces can be split into groups there are *general requirements* that need to be satis ed by all parts of the system *acquisition requirements* concerned with monitoring and extraction of useful information *face recognition requirements*

- e Level of con dence in output available to allow discard of erratic or ambiguous data Note this should be able to reduce false positive results without creating a large pro



### 2.3.4 Face Detection and Segmentation

Detection of faces using speci c facial features will not be possible for our task due to the low resolution of the data see Section 1 An alternative is to use the whole face pattern as a holistic representation such as with eigenface information Turk Pentland ; Moghaddam Pentland A successful neural network face detector has been developed by Rowley et al which also used receptive elds to give some translation and scale invariance A bootstrapping algorithm is used to get around the problem of nding suitable non faces negative examples to train with by incorporating initial false positives as subsequent training data This use of only the most confusable near face examples rather than a potentially huge range from the whole spectrum of non faces can substantially reduce the size of training set required for good performance compared to earlier approaches

Face detection in image sequences is very much easier due to motion cues than for single images and can be integrated into tracking techniques Once a face been found in a frame temporal correlations greatly reduce the search space in subsequent frames for instance McKenna and Gong were able to combine motion detection by spatio temporal ltering with face detection with a neural network based on Rowley et al More recently they have been able to use colour to further reduce computation and give greater invariance to rotations in depth and partial occlusions McKenna et al a

Face detection not only includes nding a face in an image but also determines how much of the face and background is actually segmented for further testing The approach taken to face segmentation is important when assessing performance as transitory details such as hair style and background details if included in training data may be used as the most effective distinguishing detail For instance if one person stands next to a plant for a picture whilst another does not it is very much easier to check for the presence of the plant rather than to compare subtle facial details Some groups such as Craw et al ignore higher performance of experiments conducted with face images with hair included as this face representation is not seen as being suf ciently general for images taken over time and prefer to cite poorer results for hair free data There is some psychological evidence that person speci c details such as hair may be used by humans for unfamiliar face recognition Hancock et al however so the visual features that are used for recognition may well be dependent on the task

In contrast non person speci c details such as background are more obviously spurious for recognition Turk and Pentland acknowledged that the background surrounding the faces in their database was a signi cant part of the image data used to classify the faces Of course this must severely limit generalisation of such an approach when it is trained with data against one background and tested with data containing a different background

### 2.3.5 Normalisation and Vectorisation of Images

Once a face has been localised and segmented within an image the image itself must be standard ised or normalised prior to further processing to improve the ef ciency of matching Sometimes such normalisation is just an adjustment of grey level intensity values but here we are considering adjustments to the image shape This could be as simple as a rescaling to some standard size or as complex as remapping each pixel

The normalisation and vectorisation of an image are approximately similar processes Image normalisation is generally taken to be a process of adjusting to allow particular areas in different images to line up when any two images are matched together For example face images are very commonly normalised via af ne transform on the basis of the positioning of both eyes and some times mouth or nose position This can be taken further via the morphing the face texture on the basis of a larger number of standard facial landmark positions Dense correspondence is the ulti mate correspondence where all elements of the image vector correspond to pixel information from the same object feature in scene in other words the process creates a feature based representation from the pixel information in the most abstract meaning of feature

### 2.4.3 Principal Components Analysis and 'Eigenfaces'

Principal components analysis PCA is a simple statistical dimensionality reducing technique that has perhaps become the most popular for face recognition PCA via the Kahunen Loeve transform can extract the most statistically signi cant information

*Multi-layer Perceptrons and Associative Networks*

The Multi layer Perceptron MLP commonly trained using gradient descent with error back propagation is capable of good generalisation for dif cult problems but is notoriously dif cult to ensure global convergence under all training runs as the non linearity of the hidden units and the nature of the input output mapping lead to a large number of local minima and training times can typically be long Cottrell et al Fleming and Cottrell used multi layer networks with target output equal to input auto association in order to compress photographic images The network was trained on random patches of image The compressed signal could be taken from the hidden layer of units these values were effectively eigenvalues the eigenvectors called holons here being contained in the weight values between the unit layers and these values could in turn be put back in to decode or uncompress the original image as output values

Cottrell et al found that the non linear arrangement of their multi layer network did not actually improve the compression of images when compared to networks using linear units For this reason all following networks used for PCA such as Turk and Pentland for instance have used simpler linear associative networks However Valentin et al suggests that while linear associative networks and MLPs using back propagation which calculate PCA can be effective for single viewed classi cation tasks they may not be as effective as HyperBF networks Poggio Edelman ; Brunelli Poggio in a nonlinear mapping task for example the classi cation of people with varying head pose see Section

### 2.5.3 Hierarchical Neural Networks

The Cognitron Fukushima and Neocognitron Fukushima were biologically inspired o cel incr to train with small amounts of data due to the large numbers of layers the information has to passed through It is clear that D objects can be invariantly represented in such structures Rolls ; Wallis Rolls but at present the computational load precludes them from real time applications They would be suitable for a parallel process but the specialised hardware required would exclude them from task suitability this time through cost Task Requirement a

### 2.5.4 Radial Basis Function Networks

One can implicitly model a view based recognition task using linear combinations of D views Ullman Basri to represent any D view of an object A simpler approach is for the system to use view interpolation techniques Poggio Edelman ; Brunelli Poggio to learn the task explicitly Radial basis function RBF neural networks have been identi ed as valuable adaptive learning model by a wide range of researchers Moody Darken ; Broomhead Lowe ; Poggio Girosi b; Musavi et al ; Ahmad Tresp ; Bishop for such tasks Their main advantages are computational simplicity supported by well developed mathematical theory and robust generalisation powerful enough for real time real life tasks Pomerleau ; Rosenblum Davis They are seen as ideal for practical vision applications by Girosi as they are good at handling sparse high dimensional data and because they use approximation to handle noisy real life data The nonlinear decision boundaries of the RBF network make it better in general for function approximation than the hyperplanes created by the multi layer perceptron MLP with sigmoid units Poggio Girosi b and they provide a guaranteed globally optimal solution via simple linear optimisation The RBF network is a poor extrapolator compared to the MLP and this behaviour can give it useful low false positive rates in classi cation problems This is because its basis functions cover only small localised regions unlike sigmoidal basis functions which are nonzero over an in nitely large region of the input space

Regularisation Networks are based on mathematical regularisation theory and include RBF and HyperBF HBF networks in con gurations where the networks have an equal number of hidden units and training examples Girosi et al They can be seen as performing generalisation through non linear view approximation Bulthoff Edelman which has the advantage over linear interpolation linear combination of views Ullman Basri in that it is less affected by variation orthogonal to learnt variation see Figure The RBF network can be considered as a special case of the more general HBF network Poggio Girosi b

Once training examples have been collected as input output pairs that is with the target class attached to each image tasks can be simply learnt directly b very similar to those used in Brunelli and Poggio

Ahmad and Tresp trained a variety of nets to recognise stationary hand gestures from computer generated D polar coordinates of ngertips not actual images They achieved good generalisation in D orientation and their system was able to cope well even when much of the data was missing Their standard test data was best handled by a back propagation net but this performed badly with missing or uncertain noisy features suffering a serious fall off in performance as more

| Group | Technique | Images per Person | | | | | Processing Time | |
|---|---|---|---|---|---|---|---|---|
| | | | [illegible] | [illegible] | | | Training | Classification |
| Samaria | HMM | | | | | [illegible] | | |
| Harter | pseudo [illegible] D HMM | | | | | | | [illegible] min |
| Lawrence *et al.* | Eigenfaces | [illegible] | | [illegible] | | | | |
| | PCA MLP | | [illegible] | | | | | |
| | SOM MLP | | | | | [illegible] | | |
| | PCA CN | [illegible] | [illegible] | [illegible] | | [illegible] | | |
| | SOM CN | [illegible] | [illegible] | | [illegible] | | [illegible] hr | < [illegible] sec |
| Lin *et al.* | PDBNN | | | | | [illegible] | [illegible] min | < [illegible] sec |
| Lucas | *n* tuple | | [illegible] | | | | [illegible] sec | [illegible] sec |
| | cont *n* tuple | [illegible] | | [illegible] | [illegible] | | [illegible] sec | [illegible] sec |
| | NN | | | | | [illegible] | [illegible] sec | [illegible] sec |
| Howell Buxton | RBF before discard | | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] sec | [illegible] sec |
| | after discard | | [illegible] | | | | [illegible] sec | [illegible] sec |

Table [illegible] Test generalisation % correct and processs[illegible] i

pose and lter based preprocessing methods affect the distances between face identity classes In particular we will concentrate on how such distances are modi ed with pose variations as this is crucial for our task In addition the reasoning component of the RBF network will be analysed and compared with related classi cation methods

# Chapter 3

# Representations of Pose-Varying Faces

The previous chapter has shown the suitability of our proposed approach to the main task of face recognition using a computationally ef cient approach based on RBF networks with simple

We rst look at the fundamental similarity mechanism we use f strate practically how the resolution affects the entire database when used to train a variety of classi ers

### 3.1.2 Varying Face View

The pose view of the person is another factor besides resolution that affects the inter class dis tinction This is illustrated by Figure A in Appendix A which shows all Euclidean distances for six individual images at the × resolution from the Sussex Database three each from classes and using pose angles of ° frontal ° and a ° pro le As in Figures A and A all distances are shown on the graphs connected by lines according to class and the zero value can be seen where the image is compared to itself

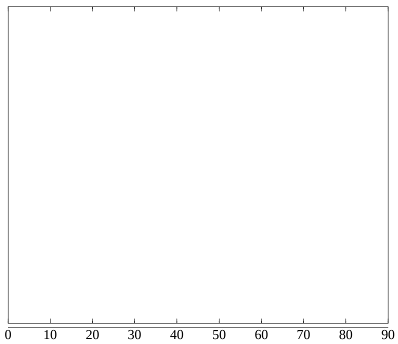
*Results*

The extreme pro le view ° is less distinct than the centre views in Figure A in Appendix A and this will add to the problem of lack of interpolative data when we come to use these images with the RBF network which largely relies on data interpolation Because of this we can expect that performance for the RBF networks using pro le information will be signi cantly lower than for the central views and also lower than for the frontal ° view where the intra class views remain distinct for a greater range of views

Intra class Euclidean distances have been shown to be less for some speci c images in the Sussex database at least than for inter class comparisons for small pose angle ranges This shows the potential of using such comparisons for recognition especially where training examples can be provided at regular pose intervals

Figure A shows some bias in intra inter class distinction for the frontal range ° over the pro le range ° This may help to explain experimental results in unfamiliar face recog nition such as O Toole et al where no advantage was found for views over frontal views instead both were equivalent and much faster to match than pro le views Bruce took such results as supporting the view that views were not serving as canonical represen tations for recognition and that full face and pro le view might be separately represented The mid pose views used in Figures A A A b i and ii all show that same class frontal views can quite often be discriminated from other class views simply on the basis of Euclidean distance

The nose centering technique we have employed here is not a rigorous mathematical vec torisation of the image such as used by Beymer but the hand alignment of facial features such as the left eye and nose over pose is shown to improve gen


0
10
20
30
40
50
60
70
80
90

| Resolution | % Correct | | |
|---|---|---|---|
| | NN WTA | NN Class based | PNN |
| [illegible] × | | | |

| Resolution | Centering | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|
| [illegible] × [illegible] | Nose | [illegible] | [illegible] | [illegible] |
| | Face | [illegible] | [illegible] | [illegible] |
| [illegible] × [illegible] | Nose | [illegible] | [illegible] | [illegible] |
| | Face | [illegible] | [illegible] | [illegible] |
| [illegible] × [illegible] | Nose | [illegible] | [illegible] | [illegible] |
| | Face | [illegible] | [illegible] | [illegible] |
| [illegible] × [illegible] | Nose | [illegible] | [illegible] | [illegible] |
| | Face | [illegible] | [illegible] | [illegible] |
| [illegible] × [illegible] | Nose | [illegible] | [illegible] | [illegible] |
| | Face | [illegible] | [illegible] | [illegible] |

Table [illegible] Test generalisation for [illegible] example RBF network using DoG preprocessed images at

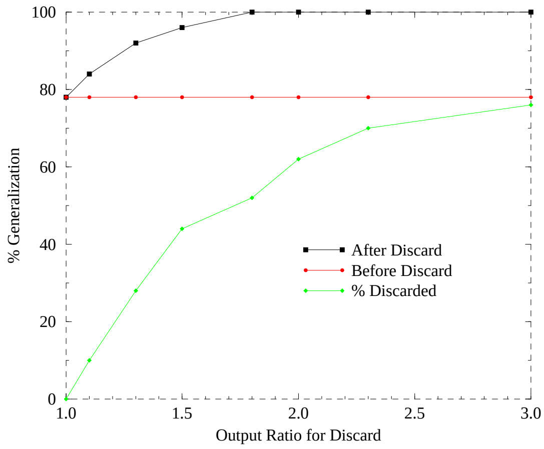



Figure Effect on test generalisation after discard of changing the low con dence threshold for RBF networks trained with DoG preprocessed × faces images from Sussex database The low con dence threshold is based on the ratio between highest and second highest output units and a value of has been found to be useful in practise

| Classi er | Distance Metric | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|
| NN WTA | City Block | [illegible] | | |
| | Euclidean | [illegible] | | |
| NN Class based | City Block | [illegible] | | |
| | Euclidean | [illegible] | | |
| PNN | City Block | [illegible] | | |
| | Euclidean | [illegible] | | |
| RBF | City Block | [illegible] | [illegible] | [illegible] |
| | Euclidean | [illegible] | [illegible] | [illegible] |

Table [illegible] Test generalisation for example classi ers using City Block and Euclidean distance measures trained with DoG preprocessed [illegible] × [illegible] faces images from Sussex database

In summary it was not as easy to distinguish pose classes as it was for identity classes This suggests that for this database at least images of different identity are further apart in Euclidean space than images of different pose We can say that the use of learning by examples distinguished by Euclidean distances is therefore especially appropriate for face recognition in the presence of large pose changes as the distances are affected more by identity than pose

### 3.2.6 Discussion

This section has presented generalisation performance from a variety of kernel based classi ers trained with the Sussex database These show that it is possible to distinguish face classes using simple classi ers moderately well even under fairly large pose ranges but that the con dence measure from the RBF network allows it to outperform the simpler methods

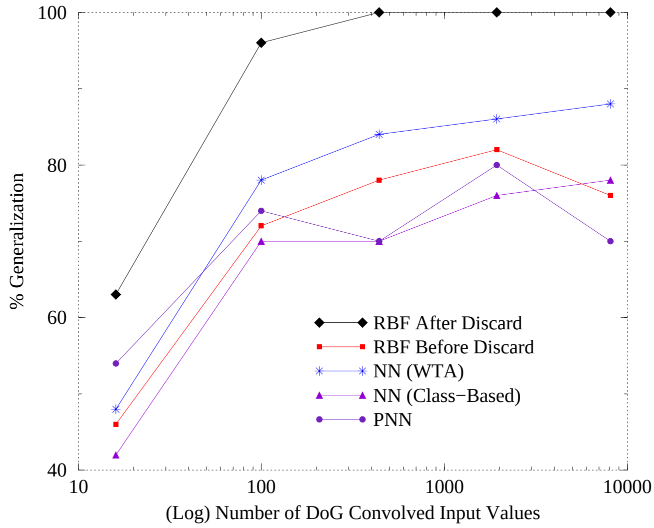
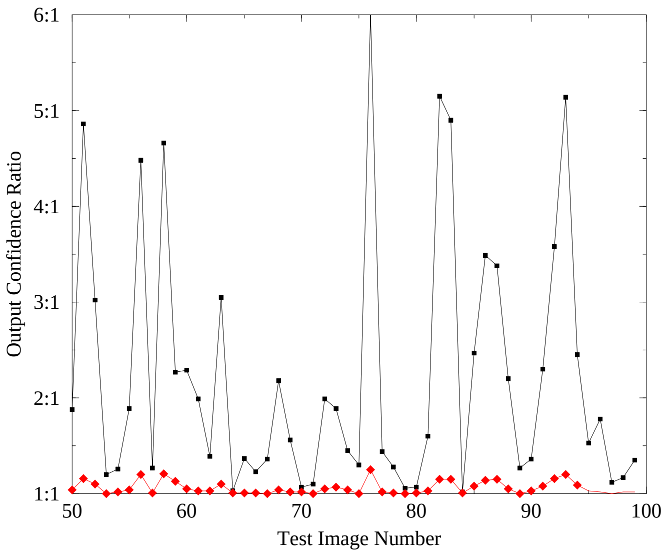
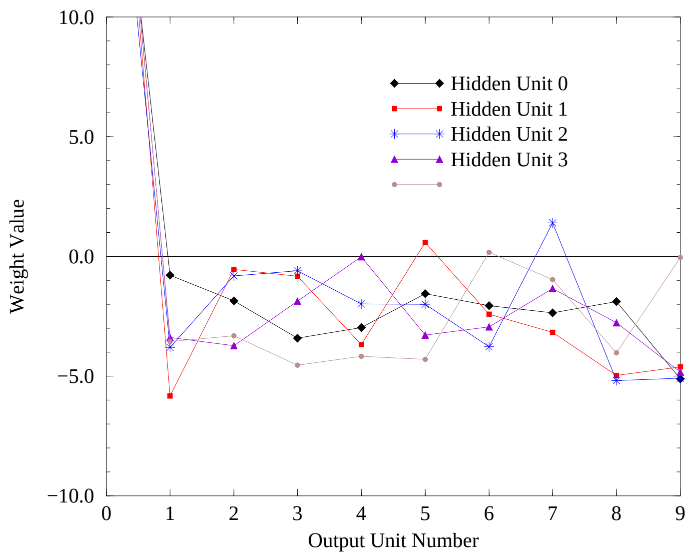



Figure Effect of changing the number of input data values on test generalisation for classi ers with ve training examples per class This number is varied via the original image resolution of face images from Sussex database before DoG preprocessing see Table C Appendix C for details
Discarding is only shown for RBF as the nearest neighbour NN classi ers and probabilistic neural network PNN do not provide enough differentiation between output units to enable a discard measure see Figure





10.0
5.0
0.0
−5.0
−10.0
0
1
2
3
4
5
6
7
8
9
Weight Value
Output Unit Number
Hidden Unit 0
Hidden Unit 1
Hidden Unit 2
Hidden Unit 3

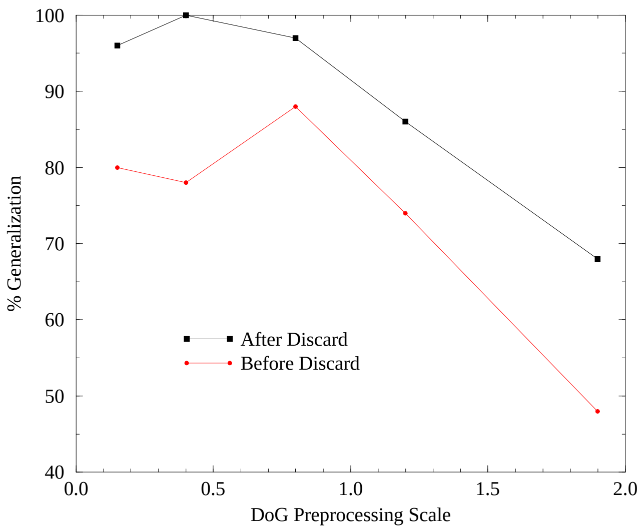



Figure Effect on test generalisation for RBF networks of changing DoG scale for the preprocessing of × faces images from Sussex database before and after discard Changes in DoG scale will affect the mask size and therefore the amount of data remaining after convolution see Figure C

| Number of Scales | Samples per Image | Thres holding | Grey Level Range | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|---|---|
| | | No | Full | | | |
| | | Yes | Full | | | |
| | | | Reduced | | | |
| | | Yes | Full | | | |

Table Test generalisation for example RBF networks using non thresholded gradient and thresholded zero crossings DoG preprocessing with one and four DoG scales

*Results*

Table shows the results with all these variations in the preprocessing stage Training with zero crossings thresholded data gave better generalisation compared to the non thresholded gradient data The use of multiple DoG scales gave a modest improvement in performance but required four times as much data than for one scale

The use of data with a reduced range of grey levels gave a great increased generalisation com pared to tests using the full range of grey levels but it is an *ad hoc* heuristic at present taking advantage of the constrained conditions of the Sussex database and it is unclear how to generalise such a technique to all lighting conditions

In summary varying a wide range of parameters in the DoG preprocessing did not seem to

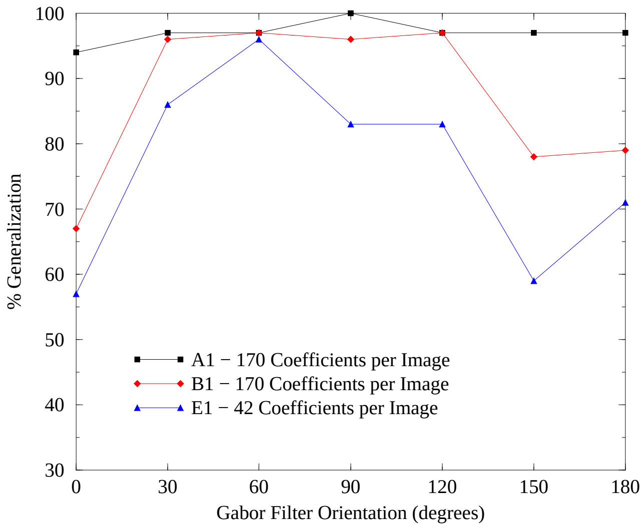



Figure . Effect of changing the angle of orientation in single orientation Gabor preprocessing on test generalisation after discard for RBF networks using × face images from the Sussex database see Section C Appendix C for details of sampling schemes

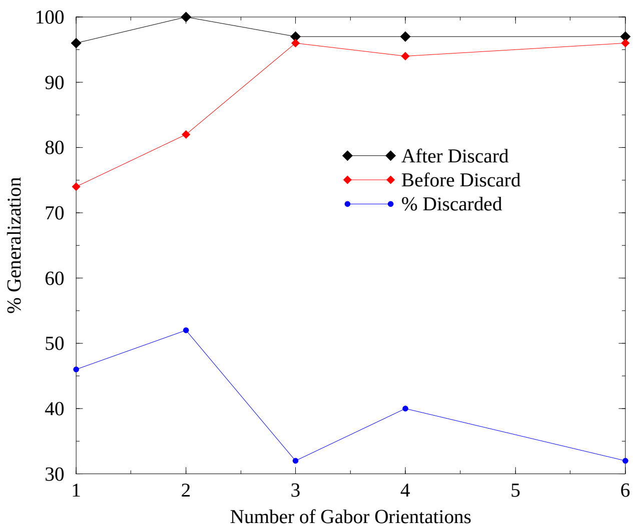



Figure . Effect of changing number of orientations on test generalisation and discard rates using Gabor Bx preprocessing of

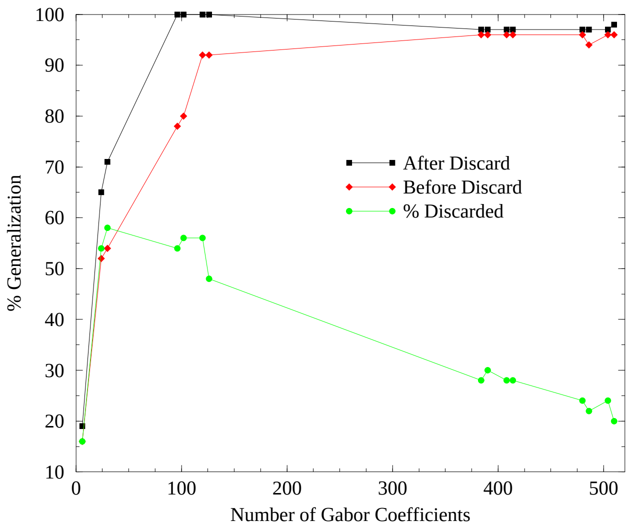



Figure Effect on test generalisation and discard rates of changing number of Gabor coef cients through selection of speci c scales see Table C Appendix C for details for A preprocessing of × faces images from Sussex database

decision puts constraints on the nature of the lter based preprocessing that can be performed as the number and extent determined by the lter mask size of sampling positions within the image will be restricted Thireo

Reassigning the Sussex database image classes in order to classify them in terms of speci c pose classes rather than identity classes met with less success than the other way around though a lower

# Chapter 4

# Invariance Properties of the RBF Network

This chapter explores the invariance characteristics of the RBF network looking at how tolerant it is to particular forms of image variation and how this is affected by the preprocessing of the input data It is important to know how robust our system is to the variation anticipated for the main task as this will determine the accuracy of face segmentation and preprocessing computational load required for data to be learnt or recognised

The experiments in the rst half of the chapter are designed to show how well the RBF network can learn identity and generalise to novel images with data w

## 4.1 Test Details

All the experiments in this chapter use the image person Sussex Database for details see Section A Appendix A This database has been designed to test recognition abilities for faces over a ° range of poses from frontal to pro le see Figure A for example A pixel based representation of the D image used as a D vector for input to the network will not provide any particular invariance to image variation by itself It will be the preprocessing and reasoning stages that provide the necessary invariance To compare and contrast the effects of preprocessing without a large number of results most of the tests will concentrate on two applications of the DoG and Gabor techniques discussed in Section of the previous chapter

**Single-scale Difference of Gaussians (DoG) filtering** This is performed as a convolution of the image with a D DoG lter mask of a single scale factor with thresholding to give binary zero crossing information Each processed image has samples corresponding to a $\times$ convolution of the original $\times$ image

**Gabor filtering** This is D Gabor wavelet analysis at four scales and three orientations termed A in Section Each processed image has coef cients corresponding to the outputs of the different scaled and oriented lters at different positions

## 4.2 Pose Invariance

Task Requirement d ii speci es an invariance to pose and so it is important to test our system to determine what limits it has in this respect In our potential environment the subjects are allowed unrestricted movement around the room and therefore will be visible at any pose angle towards the camera that is physiologically possible for the head around the vertical $y$ axis Obviously views such as the back of the head are not learnable in terms of identity especially as the requirements specify an invariance to hair style Task Requirement d iii

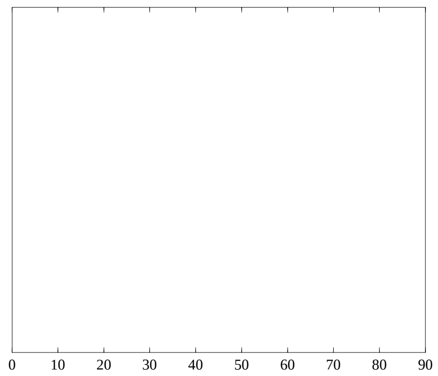
A useful system in an unrestricted environment should be expected to cope with the full range of views that contain facial information which is roughly $\pm$ ° where ° is the frontal view Such a wide pose range is in contrast to many face recognition systems which do not explicitly dealt with pose preferring to restrict data to face images with very slight pose variation typically $\pm$ ° which can be approximated as linear RBF networks in view of their interpolation properties should allow some pose invariance given suf ciently close examples for effective interpolation but the extent of this will need to be determined empirically

In this section we will be testing the RBF network for two types of pose invariance by training with two different arrangements of the data examples the rst searches for *inherent invariance* by training with unvaried images in other words one xed pose for all classes and testing with varied images only all the other poses not seen during training the second is looking for *learnt invariance* by training with explicit examples of pose variation

### 4.2.1 Inherent Pose Invariance

The pose invariance that we have termed inherent in this section is the generalisation obtained when the RBF network has been trained with images that have no pose variation that is they all come from one xed pose position and is then tested with images of different pose to that used for training

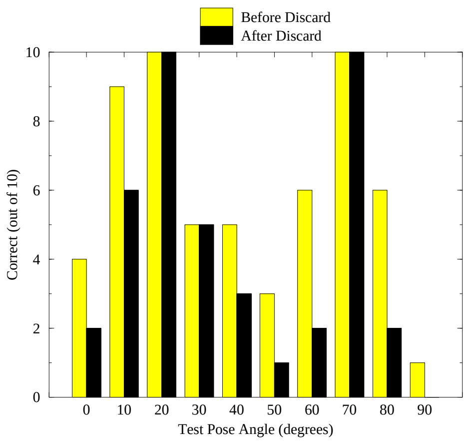
When testing for inherent pose invariance with the Sussex database where all images for each class have a different pose angle g

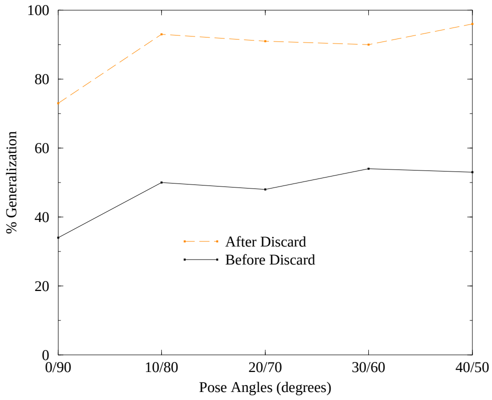

0
10
20
30
40
50
60
70
80
90

| Network | Number of Training examples | Number per Class | Number of Test images |
|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | | [illegible] |
| [illegible] | [illegible] | | [illegible] |

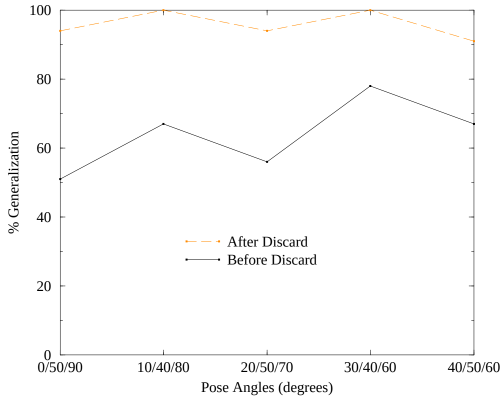
Table The four different types of interpolating RBF networks used to test learnt pose invari ance

| Pre processing | Network | Training Examples | Training Pose Angles ° | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|---|---|
| DoG | [illegible] | [illegible] | [illegible] | | [illegible] | |
| | [illegible] | [illegible] | [illegible] | [illegible] | | |
| | [illegible] | | [illegible] | [illegible] | [illegible] | [illegible] |


Before Discard
After Discard
Correct (out of 10)
Test Pose Angle (degrees)
0
2
4
6
8
10
0
10
20
30
40
50
60
70
80
90

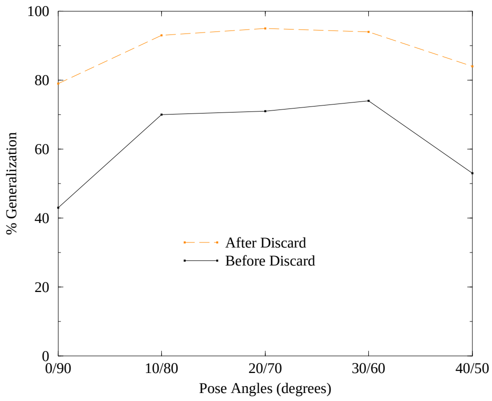



i average discard %

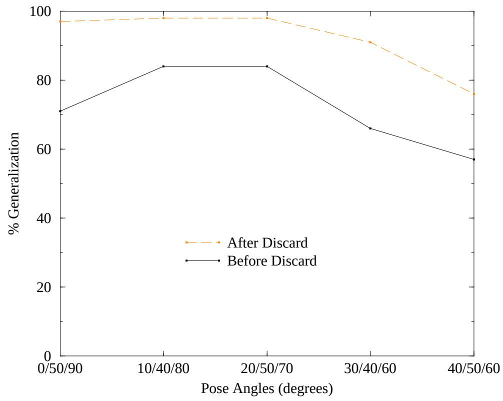



ii average discard %

a DoG preprocessing



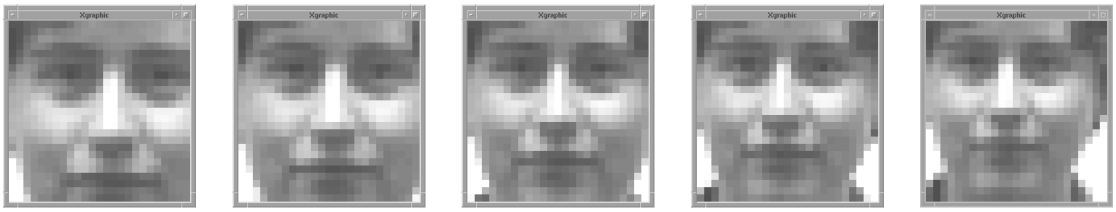

i average discard %




ii average discard %

b Gabor preprocessing

Figure Learnt Pose Invariance Test generalisation with trained with two images per class and three per class interpolating RBF networks varying over selections of pose angles from left to right widely to closely space intervals



a % × b % × c normal × d % × e % ×

Figure Example scaled versions of the original front view of one individual from the Sussex database used to test for scale invariance with relative size to the normal sampling area and size of window grabbed from in pixels

| Variation | Network | Pre processing | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|---|
| Shift | [illegible] | DoG | [illegible] | | [illegible] |
| | | Gabor | [illegible] | [illegible] | [illegible] |
| Scale | [illegible] | DoG | | [illegible] | [illegible] |
| | | Gabor | [illegible] | [illegible] | [illegible] |

Table Inherent Shift and Scale Invariance Effect on test generalisation for the RBF network of different variations in the dataset both before and after discarding of low con dence classi cations networks trained with all ten non varied versions of poses for each person and testing with varied versions training and test images

- A *shift-varying* data set with ve copies of each image one at the standard sampling window position and four others at the corners of a box where all $x$ $y$ positions were ± pixels from the centre see Figure
- A

| Variation | Network | Pre processing | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|---|
| Shift | [illegible] | DoG | [illegible] | | |
| | | Gabor | | [illegible] | |
| | [illegible] | DoG | [illegible] | | [illegible] |
| | | Gabor | | [illegible] | |
| Scale | [illegible] | DoG | | [illegible] | [illegible] |
| | | Gabor | [illegible] | | |
| | [illegible] | DoG | [illegible] | [illegible] | |
| | | Gabor | [illegible] | [illegible] | [illegible] |

Table Learnt Shift and Scale Invariance Effect on test generalisation for the RBF network of different variations in the dataset both before and after discarding of low con dence classi cations networks trained with all ve shift or scale varied versions of two or ve equally spaced poses for each person

% being even lower than random % The network using Gabor preprocessing was able to give a low level of useful generalisation

In contrast the scale varying test data appears to be much easier for the network to generalise to even without explicit training examples and a useful level of performance was obtained with both types of preprocessing As before the Gabor preprocessed training data was easier to learn and generalise with compared to the DoG preprocessed data and networks using the former were able to give a high level of generalisation performance even without discard

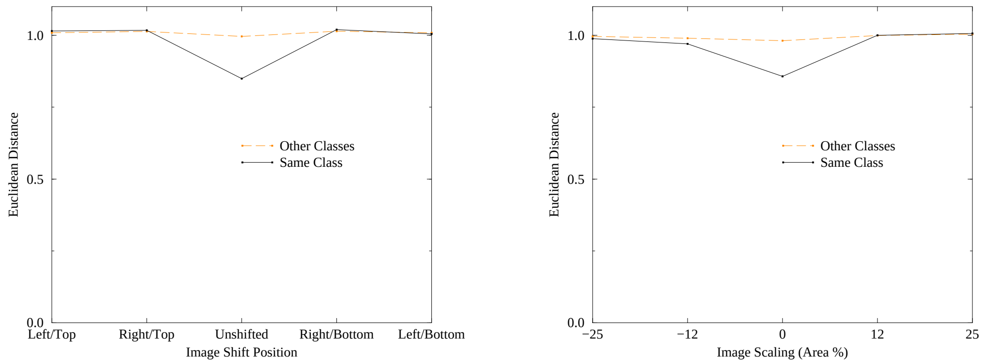
*Summary*

This section has shown that the RBF network has a signi cant inherent invariance to scale differ ences with the Gabor preprocessed face data from the Sussex database and a moderate invariance with the DoG preprocessed data In marked contrast the shifted images were very much harder for the network to generalise to with both preprocessing techniques

Figures a and b show that these differences primarily arise out of the choice of prepro cessing although the scale transformation also seems to alter the image vector less than the shift transformation This is shown by the other class line for the Gabor scaled images b ii being noticeably further away from the same class line than for the other combinations of transformation and preprocessing

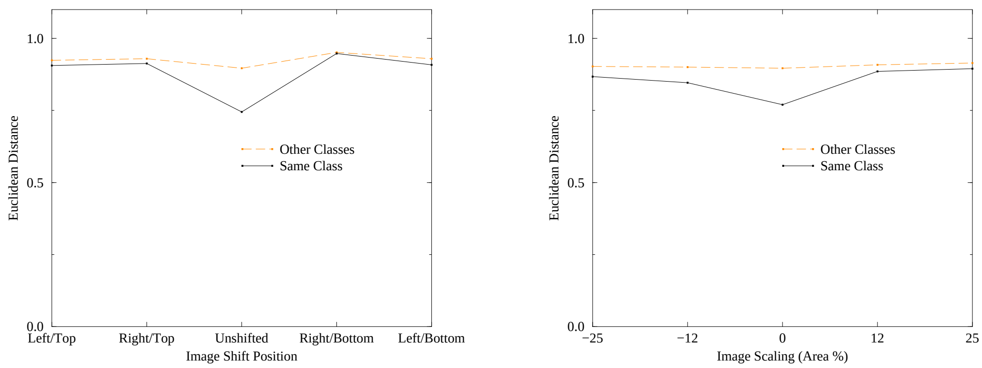
### 4.3.3 Learnt Shift and Scale Invariance

The experiments in this section test for learnt shift and scale invariance As before they use a xed selection of pose positions for training examples but this time use all ve versions varied unvaried of each original image This helps the network to learn about the shift and scale image variation during training and thus develop a learnt invariance The difference between the generalisation performance found in the previous section with inherent invariance and in the tests in this section will iex e



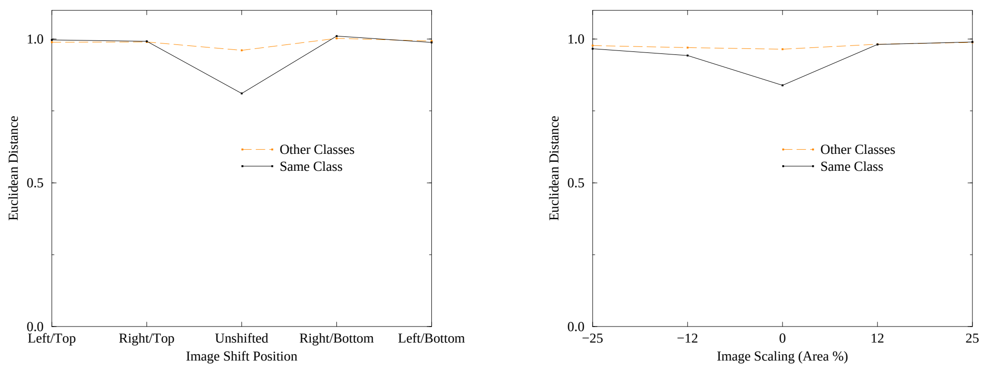

i ii

a Single scale normal DoG preprocessing




i ii

b Multi scale normal Gabor preprocessing




i ii

c Multi scale DoG preprocessing

Figure Euclidean distances for images from the Sussex database to same and other class images

## 4.4 General Discussion

Chapter showed how the representation used for input data can have a profound effect on the ability of the RBF network to generalise from a learnt task This chapter has developed these ideas to analyse how speci c variations in the image will affect such generalisation

The experiments in Section looked at invariance to head pose [illegible]

# Chapter 5

# Face Unit RBF Networks

This chapter introduces a different way of learning the face recognition task through the reorgani sation of the standard RBF networks into a group of smaller face recognition units each trained

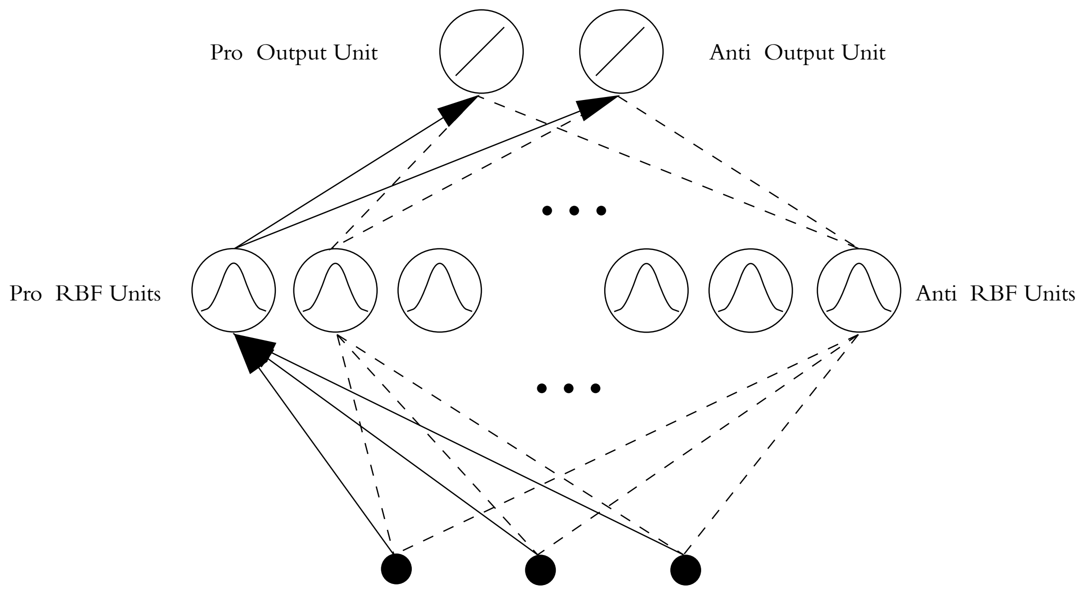



Figure General structure for a face unit RBF network Although there can be a varying number of and ratio between pro and anti hidden units there are always two output units for and against the class learnt by the network All hidden units are fully connected to both output units This can be compared with the standard RBF network model shown in Figure B Appendix B

per class only the face unit network has two output units one positive denoting yes for the current class and and one negative no for all other classes We use the term *pro* to denote hidden units or evidence for the class and *anti* for that against the class the negative evidence For each individual a face unit RBF network can be trained to discriminate between that person and others selected from the data set using this pro supporting and anti differentiating evidence for and against the individual The ratio between the two can be varied

Although this approach increases complexity as more networks need to be trained and and the training data needs to be manipulated differently for each face unit the splitting of the training for individual classes into separate networks gives a modular structure that can potentially support large numbers of classes since network size and computational load for weight calculations for the standard RBF model may become impractical as the number of classes increases

### 5.1.1 Selection of Negative Evidence

The fundamental process in the face unit network is the splitting of the training data into two halves class and non class The small size of the network is due to the limited amount of non class data used for training only those that are seen as har

| Network | Training Examples per Class | |
|---|---|---|
| | | |
| Standard RBF network | [illegible] | [illegible] |
| Single anti Face Unit Network | [illegible] | [illegible] |
| Double anti Face Unit Network | [illegible] | |

Table Numbers of hidden units used by different RBF networks for same task when using the Sussex database

**'Single anti' face unit network** This uses equal numbers of pro and anti hidden units

**'Double anti' face unit network** This uses two anti hidden units for every one pro

The double anti face unit network is closer than the single anti arrangement to the full standard RBF model in that it uses more negative than positive evidence It is included in the tests to show whether this additional information would give the network better discrimination from the negative classes than the single anti arrangement This characteristic will be more important as the number of classes in the dataset increases as the number of negative classes will become proportionately greater

We can compare the relative sizes of the face unit network and the standard RBF network The standard RBF network uses $cn$ hidden units where $c$ is the number of identity classes and $n$ is the number of training examples per class This gives $10n$ hidden units in total when using the Sussex database as shown in Figure The single anti face unit network has only two classes for training for and against a single person and a single anti hidden unit for every pro unit and therefore has $2n$ hidden units in total however many identity classes there are The double anti face unit network uses two anti hidden units for every one pro and therefore has $3n$ hidden units in all The outcome of this is that as $c$ the number of identity classes increases the face unit network required for a particular task will becomes much smaller relative to the standard RBF network needed for the same task

Once the number of examples is chosen we then use two different strategies for the selection of the anti evidence This gives two further types of network

**'Single best negative' (sbn) face unit networks** These use an average of all vector distances

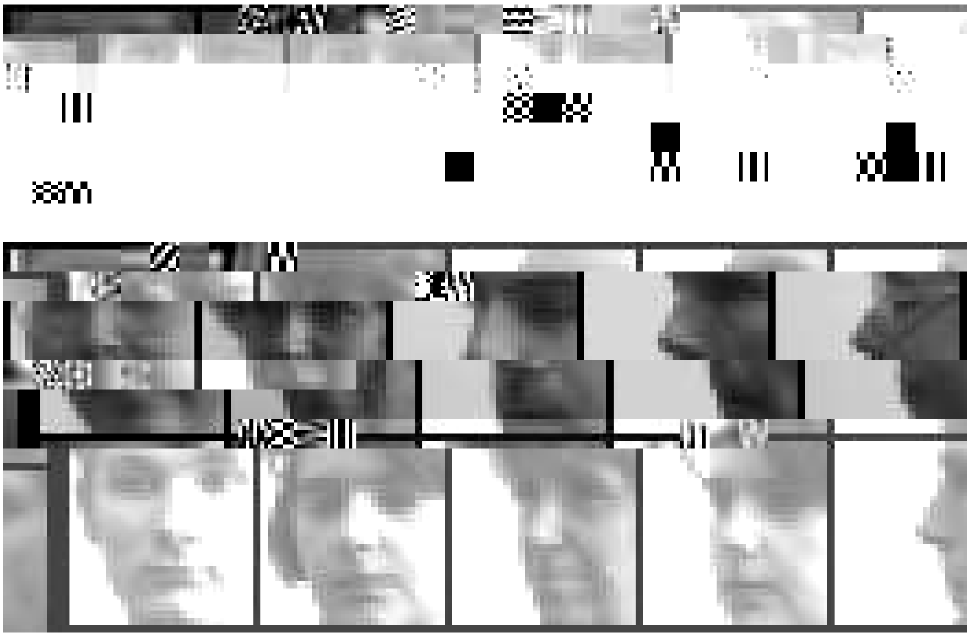


Figure  Example of the range of negative classes that can be selected during the training of a double anti multiple best negative *mbn* face unit RBF network
The top line shows the supporting pro evidence the middle and bottom lines the differentiating anti evidence middle line is the closest to the pro class bottom line the second closest

network sizes from to on the standard image Sussex database if these networks had been labelled in the standard train test form this would correspond to a range between and networks To give an optimal spread of the image data for training xed selections of pose angle were used for each size of network as used in Chapter see Table For instance the and networks used poses ° ° ° ° and ° where the pose range was ° frontal ° pro le

Figure shows how the images used for training were selected for a *mbn* face unit network in the experiment This illustrates not only how several anti classes are used in the *mbn* scheme but also how they are ranked for the double anti arrangement

### 5.1.4 Results

As in the previous chapter for clarity our tests use two standard preprocessing methods only the single scale DoG and the Gabor A with four scales and three orientations details in Section and Appendix C

Figure summarises the overall results for the various types of face unit networks with different pro anti ratios and different strategies for selection of anti images To simplify the information these graphs do not show the rates after discard but these gave a consistent improvement of about % over rates before discard for all networks

The face unit networks are essentially working in a two class classi cation problem so a random level of generalisation would be % Interestingly the double anti network arrangement did not appear to give radically better performance than the single anti except for the and example networks using Gabor preprocessed data This indicates that the selection of appropriate anti images is ef cient enough by itself to create a division in image space between the class and all others without requiring additional negative examples

Table shows speci c generalisation rates for the example and face unit networks before and after discard It can be seen here that the Gabor preprocessed data allowed the

| Variation | Pre processing | Network | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|---|
| Shift | DoG | [illegible] | [illegible] | [illegible] | |
| | | [illegible] | [illegible] | [illegible] | [illegible] |
| | Gabor | [illegible] | [illegible] | [illegible] | [illegible] |
| | | [illegible] | | [illegible] | [illegible] |
| Scale | DoG | [illegible] | [illegible] | [illegible] | |
| | | [illegible] | [illegible] | [illegible] | |
| | Gabor | [illegible] | | [illegible] | |
| | | [illegible] | [illegible] | [illegible] | [illegible] |

Table [illegible] Generalisation for pose example multiple best negative *mbn* face unit networks [illegible] and [illegible] with shift and scale varying data

RBF network to perform more ef ciently than the DoG preprocessed data both in lower discard rates and generalisation before and after discard

*Summary*

The *mbn* strategy for selecting anti evidence seemed slightly better than the *sbn* indicating that dealing with local at a pose level confusions was more ef cient that trying to identify one global class with which the main class should be contrasted

### 5.1.5 Shift and Scale-Varying Data

| Discard | |
|---|---|

| Discard Measure | | Initial % | % Discarded | % after Discard | Ratio after Discard |
|---|---|---|---|---|---|
| Standard RBF Network Only | | [illegible] | [illegible] | [illegible] | [illegible] |
| Cooperative Threshold | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | | [illegible] | [illegible] |
| | [illegible] | [illegible] | | [illegible] | [illegible] |
| | | [illegible] | | | [illegible] |
| | [illegible] | [illegible] | | [illegible] | [illegible] |
| | | [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | | | [illegible] |
| | | [illegible] | [illegible] | [illegible] | [illegible] |

a DoG preprocessing

| Discard Measure | Initial % | % | % after | Ratio after |
|---|---|---|---|---|

| Discard Measure | | Initial % | % Discarded | % after Discard | Ratio after Discard |
|---|---|---|---|---|---|
| Standard RBF Network Only | | | | | |
| Cooperative Threshold | | | | | |
| | | | | | |

a Shift varying data

| Discard Measure | | Initial % | % Discarded | % after Discard | Ratio after Discard |
|---|---|---|---|---|---|
| Standard RBF Network Only | | | | | |
| Cooperative Threshold | | | | | |
| | | | | | |

b Scale varying data

Table Generalisation and discard rates for different discard measures with shift and scale varying data Standard RBF Network Only is the result using a simple discard measure applied to the output of a standard RBF network by itself the Cooperative Threshold is a threshold value applied to the con dence rating arising from cooperating multi class standard RBF networks and double anti multiple best negative *mbn* face unit RBF networks using Gabor preprocessing

the cost of higher proportionhn refq cmBT R T n e Tn

the cthe v

simple one network threshold used previously

Different rating threshold levels can be used with the cooperative scheme to give either ***high confidence with high discard*** using a rating threshold of or ***moderate confidence with low discard*** rat

## 5.4 General Discussion

This chapter has presented experimental work using a novel variant of the RBF network model the face unit network which learns to distinguish a single class from a range of other classes This can be used either in groups one for each class or singly in conjunction with a multi class network to give greater reliability to classi cation

The most useful con guration of face unit RBF network overall seems to be the single anti multiple best negative *mbn* face unit network which selects the most useful anti evidence to match each pro example on a pose by pose basis

The standard RBF network will give similar positive and negative information about classes because of the fully interconnected hidden to output unit layer but the face unit network by concentrating only on distinguishing one class at a time allows the negative in uences of such non class connections to be more specialised indeed optimised to give the most effective one class against all others partitioning in image space

The modular approach presented in this chapter using face unit RBF networks to learn identity is especially attractive for the unconstrained recognition task as it allows the modi cation of the learned element of the system during use and can give a secondary classi cation decision which can either con rm or dispute the primary RBF network output

Although the face unit network allows ner control in the recognition process with the standard RBF network than can be provided by the latter alone it is not used for the next two chapters which deal with image sequences and the recognition of temporal patterns This is because it is felt that the results will be more understandable if the common baseline of system con guration from chapters and comprising of the standard RBF network with simple discard measure is maintained for this later experimental work

Chapters and and this chapter have explored the behaviour of the RBF network in the narrow context of training with the Sussex database The next chapter applies a more realistic test to the network using image sequences from a less tightly constrained environment

# Chapter 6

# Face Recognition using Image Sequences

---

This chapter presents experiments using the Radial Basis Function RBF network to tackle a more unconstrained face recognition problem using low resoluti we are computationally constrained to the inter frame period of the order of tens of milliseconds determined by the frame grabber and the localisation software Offsetting this limitation is the vast quantity of data from image sequences which suits any technique that can discard low con dence output to leave a high ratio of correct classi cations In the context of videos of people moving around a room where large numbers of images of each person in the environment will be produced and changes in the identities present will not be abrupt from one frame to the next even quite high discard ratios of [illegible] [illegible]% may be acceptable if the remaining output is of suf ciently high quality

## 6.1 Specification for Image Sequences

The image sequences used in the tests reported here are the result of collaboration with Stephen McKenna and Shaogang Gong at Queen Mary and West eld College QMW London who are researching real time face detection and tracking McKenna et al This work is still at a preliminary stage and many issues are still unresolved such as the nature of appropriate training data how constrained does it need to be and how automatic its original collection from the data

| Training Selection Interval | Train Test Frames | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] | |
| | [illegible] | | [illegible] | |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | | [illegible] | [illegible] |
| [illegible] | [illegible] | | | [illegible] |
| [illegible] | [illegible] | [illegible] | | |

| Training Selection Interval | Train Test Frames | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] | |
| [illegible] | [illegible] | | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |

a DoG preprocessing

| Training Selection Interval | Train Test Frames | Initial % | % Discarded | % After Discard |
|---|---|---|---|---|
| [illegible] | [illegible] | [illegible] | | [illegible] |
| | [illegible] | [illegible] | | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | |
| [illegible] | [illegible] | | [illegible] | [illegible] |

b Gabor preprocessing

Table [illegible] Effect of preprocessing methods on test generali

| Time Window | Initial % | % After Discard |
|---|---|---|
| | | |
| | | |
| | | |

unconstrained image sequences It is clear that the ability of the RBF networks to give a measure of con dence which allows temporal integration over image frames where the visual evidence is poor is essential for this development

Work is progressing together with colleagues at QMW in re ning the face detection scheme and automated on line learning of new classes of individual The next stage of development will integrate this re ned on line face detection and localisation with the trained RBF networks to cope with real time image sequences including the usual variations in illumination as well as position scale view and facial expression It is clear from the work of Bishop and others that using statistically based techniques is the key to good performance The RBF techniques are mathe matically well founded which gives a clear advantage in engineering a solution to our application problems

# Chapter 7

# Recognition of Simple Behaviours using Time-Delay RBF Networks

frames for each person Two schemes were devised to split the data up

- the *Alternate Frames* AF tests illustrated in Figure used alternate frames from each person so that training and test data contained all ten people and the window size range was frames and
- the *Alternate Person* AP tests illustrated in Figure which used all the frames from people for training and the other for testing The window size range for the AP tests was frames

It can be seen that the two types of selection process and the varying window sizes gave a wide range of numbers of sequences that could be used as data In addition to this the variety of data will be increased by the type of head rotation

*Head Rotation Classes*

Three classes were used for training the TDRBF network corresponding to three types of rotation present in the image sequence

**LR sequences** These simulate a left to right head rotation in a window within the ten frames for each person such as shown in Figure a Sequences were interleaved with each other to use all the frames for each person For example if the window size was the sequences used would be for the AF tests and and for the AP tests up to

**RL sequences** These are identical to LR except that the rotation is in the opposite direction from right to left as shown in Figure b so the frame numbers go from to For example if the window size was the sequences used would be for the AF tests and and for the AP tests down to

**Static sequences** These simulate a xed head position through time illustrat

a b c d

Figure [illegible] Example data sequences for Alternate Frame (AF) tests with a time window of three frames: a) LR Training Frames [illegible] and [illegible] b) RL training frames [illegible] and [illegible] c) Static training frame repeated d) LR Test Frames [illegible] and [illegible] of same person. Data from all [illegible] people in the Sussex database was used for both training and testing each taken from alternate frames.

| Window | Samples | Train Test | Initial % | % Discarded | % after Discard |
|---|---|---|---|---|---|
| | [illegible] | [illegible] [illegible] | [illegible] | | [illegible] |
| | [illegible] | [illegible] [illegible] | | | [illegible] |
| [illegible] | [illegible] | [illegible] [illegible] | [illegible] | | [illegible] |
| [illegible] | [illegible] | [illegible] [illegible] | [illegible] | | [illegible] |

a [illegible] Classes distinguishing LR and static sequences

| Window | Samples | Train Test | Initial % | % Discarded | % after Discard |
|---|---|---|---|---|---|
| | [illegible] | [illegible] [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] [illegible] | [illegible] | | [illegible] |
| [illegible] | [illegible] | [illegible] [illegible] | [illegible] | | [illegible] |
| [illegible] | [illegible] | [illegible] [illegible] | [illegible] | [illegible] | [illegible] |

b [illegible] Classes distinguishing LR RL and static sequences

Table [illegible] Effect of time window size on generalization rates for TDRBF network trained and tested on image sequences from alternate frames (AF testing). The test sequences contain alternate frames from those seen during training.

a b c d

Figure . Example data sequences for Alternate Person AP tests with a time window of frames a LR training frames and b RL training frames

Figure [illegible] The real image sequence from QMW used to test TDRBF networks trained on sequences from the Sussex database Note the wide variation in head position and gaze direction

| Time | Samples Window | Training Test | Integration Layer Size % Correct | | | | |
|---|---|---|---|---|---|---|---|
| | | | | [illegible] | | | |
| [illegible] | [illegible] | | | | | | |
| | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | | | [illegible] | [illegible] | |

### 7.3.2 Discussion

The issue of the *'time base'* of actions that is how fast or slow actions occur would have to be taken into account in any real life image sequences as any movement would occur at a variety of speeds

The tests presented in this chapter have shown that the TDRBF network has useful temporal recognition properties which could be of use in real life applications due to its rapid learning and operation in comparison to more general but slower recurrent networks

# Chapter 8

# Conclusion

The aim of this thesis has been to explore the practicalities of computer based face recognition in everyday environments such as living rooms or of ces This chapter summarises the main results and contributions from the thesis and outlines directions for future work

Chapter 1 described the task of face recognition in unconstrained environments in detail and drew up speci c requirements to ful ll it A review of general theories of object recognition and psychological evidence was then followed by a more detailed discussion of current approaches to face recognition with speci c emphasis on three aspects of the recognition process acquisition representation and reasoning This allowed us to establish a suitable approach using lter based network previously used for speech recognition to recognise simple image based behaviours based on head rotation We were able to show that such actions can be easily learnt and generalised to even with simple training methods

## 8.1 Contributions of the Thesis

There are four main contributions made by the thesis

*Gabor filter representation.* We have developed an ef cient sparse sampled Gabor lter repre sentation suitable for extraction from low resolution face images We have been able to establish that this can in contrast to a Difference of Gaussians representation provide some inherent scale invariance that is without the provision of explicit scale varying images which is not present for other variations such as translation lateral shift

*Radial basis function (RBF) network scheme for image sequences.* We have developed a fast RBF network scheme which has been shown to provide robust generalisation when used with pose varying face data in image sequences

*Face Unit RBF network scheme.* We have developed our own novel Face Unit RBF network model that can be used either alone to classify individuals in a known group one for each person or to accompany standard RBF network output for a cooperating classi cation

*Image-based Time-Delay RBF network scheme.* We have established the suitability of the Time Delay RBF model previously only used for speech recognition for image analysis We were able to show that the network could recognise simple head turning behaviours in image sequences in an extension to our previous static frame training methods

## 8.2 Discussion

Section devised four main areas of requirements for our target task Group general re quirements that need to be satis ed by all parts of the system Group acquisition requirements concerned with monitoring and extraction of useful information Group face recognition re quirements for the recognition stage and Group identity requirements which are concerned with how the recognition information is used As mentioned in that section those from Group are assumed to have been previously ful lled via existing technology and those from Group are the subject for future work

We believe the Group General Requirements a and b are addressed appropriately in our RBF network scheme with lter based preprocessing We have been able to show rapid preprocess ing training and classi cation in Section and robust generalisation of trained RBF networks to test image sequences containing signi cantly different examples of everyday lighting and pose variation in Section

For Group the Face Recognition Requirements the following requirements have been ful lled in the thesis with the sections where this was demonstrated shown in brackets a Fast learning and real time recognition of up to individuals Section ; b ability to work with low resolution face images Chapters and ; c i minor translation shift and scale in variance Section ; c ii moderate illumination invariance Sections and ; c iv background ldi eno Td b a d o Td i

Speci cally in terms of invariance more precise limits need to be established for the system for translation shift and scale variation and research needs to be done to establish limits on expression and lighting variation

As mentioned in Section the Group requirements were were not included in the frame work for the thesis and were to be addressed in future work This group of requirements was concerned with adapting the known group of individuals the system could recognise This is ob

# Bibliography


Ahmad S Tresp V Some solutions to the missing feature problem in vision In Hanson S J Cowan J D Giles C L Eds *Advances in Neural Information Processing Systems* Vol pp San Mateo CA Morgan Kaufmann

Akamatsu S Sasaki T Fukamachi H Masui N Suenaga Y An accurate and robust face identi cation scheme In *Proceedings of International Conference on Pattern Recognition* pp The Hague Netherlands

ATR Cognitive and representative models of visual image information Interaction of multi modal information relayed by face
Web page `h**p://www.hip.a*r.co.jp/depar*men*s/Dep*2/Repor*s_FP.h*ml`

Beymer D J Poggio T Image representations for visual learning *Science 272*

Bichsel M Ed *Proceedings of International Workshop on Automatic Face & Gesture Recognition* Zurich Switzerland University of Zurich

Bichsel M Pentland A Human face recognition and the face image set s topology *Computer Vision, Graphics, and Image Processing: Image Understanding 59*

Biederman I Recognition by components a theory of human image understanding *Psychological Review 94*

Bishop C M *Neural Networks for Pattern Recognition* Oxford University Press Oxford UK

Bossomaier T R J Ef cient image representation by Gabor functions an information theory approach In Kulikowski J J Dickinson C M Murray I J Eds *Seeing Contour and Colour* pp Pergamon Press Oxford UK

Bouattour H Fogelman Soulie F Viennet E Solving the human face recognition task

Craw I Costen N Kato T Robertson G Akamatsu S Automatic face recognition combining con guration and texture In Bichsel M Ed *Proceedings of International Workshop on Automatic Face & Gesture Recognition* pp Zurich Switzerland University of Zurich

Craw I Tock D Bennett A Finding face features In Sandini G Ed *Proceedings of European Conference on Computer Vision, Lecture Notes in Computer Science* Vol pp Santa Margherita Ligure Italy Springer Verlag

Daugman J G Complete discrete D Gabor transforms by neural networks for image analysis and compression *IEEE Transactions on Acoustics, Speech, & Signal Processing 36*

Diamond R Carey S Why faces are and are not special An effect of expertise *Journal of Experimental Psychology: General 115*

Dong D W Atick J J Statistics of natural time varying images *Network: Computation in Neural Systems 6*

Duc B Maitre G Fischer S Bigun J Person authentication by fusing face and speech information In *Proceedings of 1st International Conference on Audio & Video-based Biometric Person*

Kanade T Picture processing by computer complex and recognition of human faces Tech rep Department of Information Science Kyoto University Japan

Kirby M Sirovich L Application of the Karhunen Loeve procedure for the characterization of human faces *IEEE Transactions on Pattern Analysis & Machine Intelligence 12*

Kittler J Li Y P Matas J S

Lin S H Kung S Y Lin L J Face recognition detection by probabilistic decision based neural network *IEEE Transactions on Neural Networks 8*

Logothetis N K Pauls J Bulthoff H H Poggio T View dependent object recog nition by monkeys *Current Biology 4*

Lucas S M Face recognition with the continuous n tuple classi er In Clark A F Ed *Proceedings of British Machine Vision Conference* pp Colchester UK BMVA Press

Moghaddam B Pentland A Probabilistic visual learning for object detection In *Proceedings of International Conference on Computer Vision* pp Cambridge MA IEEE Computer Society Press

Moody J Darken C Learning with localized receptive elds In Touretzky D Hinton G Sejnowski T Eds *Proceedings of 1988 Connectionist Models Summer School* pp Pittsburgh PA Morgan Kaufmann

Moody J Darken C Fast learning in networks of locally tuned processing units *Neural Computation 1*

Moses Y Adini Y Ullman S Face recognition the problem of compensating for

Petkov N Biologically motivated computationally intensive approaches to image pattern recognition *Future Generation Computer Systems 11*

Petkov N Kruizinga P Lourens T Biologically motivated approach to face recog

Rolls E T Brain mechanisms for invariant visual recognition and learning *Behavioural Processes 33*

Rolls E T Learning mechanisms in the temporal lobe visual cortex *Behavioural Brain Research 66*

Rosenblum M Davis L S An improved radial basis function network for autonomous road following *IEEE Transactions on Neural Networks 7*

Rosenblum M Yacoob Y Davis L S Human emotion recognition from motion using a radial basis function network architecture *IEEE Transactions on Neural Networks 7*

Rowley H A Baluja S Kanade T Human face detection in visual scenes In Touretzky D S Mozer M C Hasselmo M E Eds *Advances in Neural Information Processing Systems*

Sung K K Poggio T Example based learning fo

Young D S Representing images for computer vision Tech rep CSRP School of Cognitive and Computing Sciences University of Sussex UK

Yuille A L Deformable templates for face recognition *Journal of Cognitive Neuroscience 3*

Yuille A L Hallinan P W Cohen D S Feature extraction from faces using deformable templates *International Journal of Computer Vision 8*

# Appendix A

# Face Database Information

---

This appendix gives speci c details on the face databases used for the experimental chapters Section A[illegible] contains Euclidean distance comparisons for the Sussex database

## A.1 The ORL Database

The Olivetti Research Laboratory ORL database of faces ORL has been used for the initial experiments in Section [illegible] It is valuable as there are a wide range of published face recognition results based on the database which can be used for comparison It contains [illegible] greyscale images of [illegible] people at a resolution of [illegible] × [illegible] see Figure A[illegible] Each individual is represented by [illegible] images and for some these have been taken at different times

Variations allowed in the image included lighting facial expressions such as open or closed eyes and smiling or not smiling and facial details such as glasses or no glasses All the images were taken against a plain background with tilt and rotation up to [illegible]° and scale variation up to [illegible]%

## A.2 The Sussex Database

The Sussex Database is designed to assess how the performance of a particular face recognition technique will be affected by signi cant pose variations It only contains data for ten people which is a relatively small number by current face database standards However the main purpose of the database is not to test how many individuals a recognition system can discriminate as there are

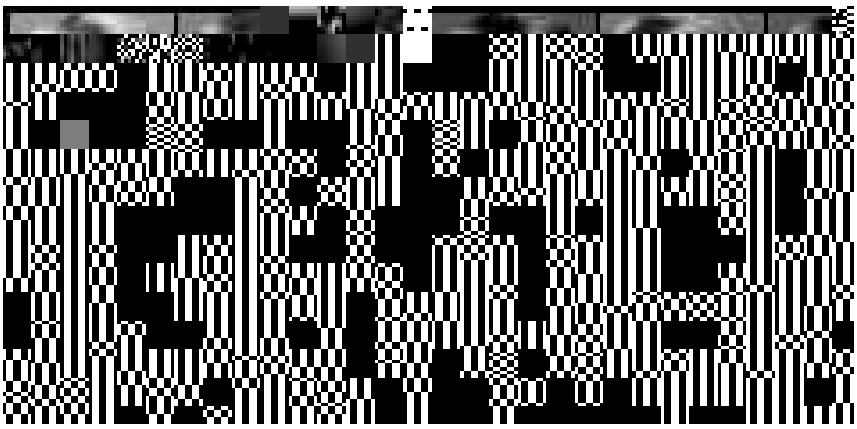


Figure A Set of [illegible] images for one person in ORL database illustrating moderate $x$ $y$ and $z$ axis rotation with expression and illumination variation

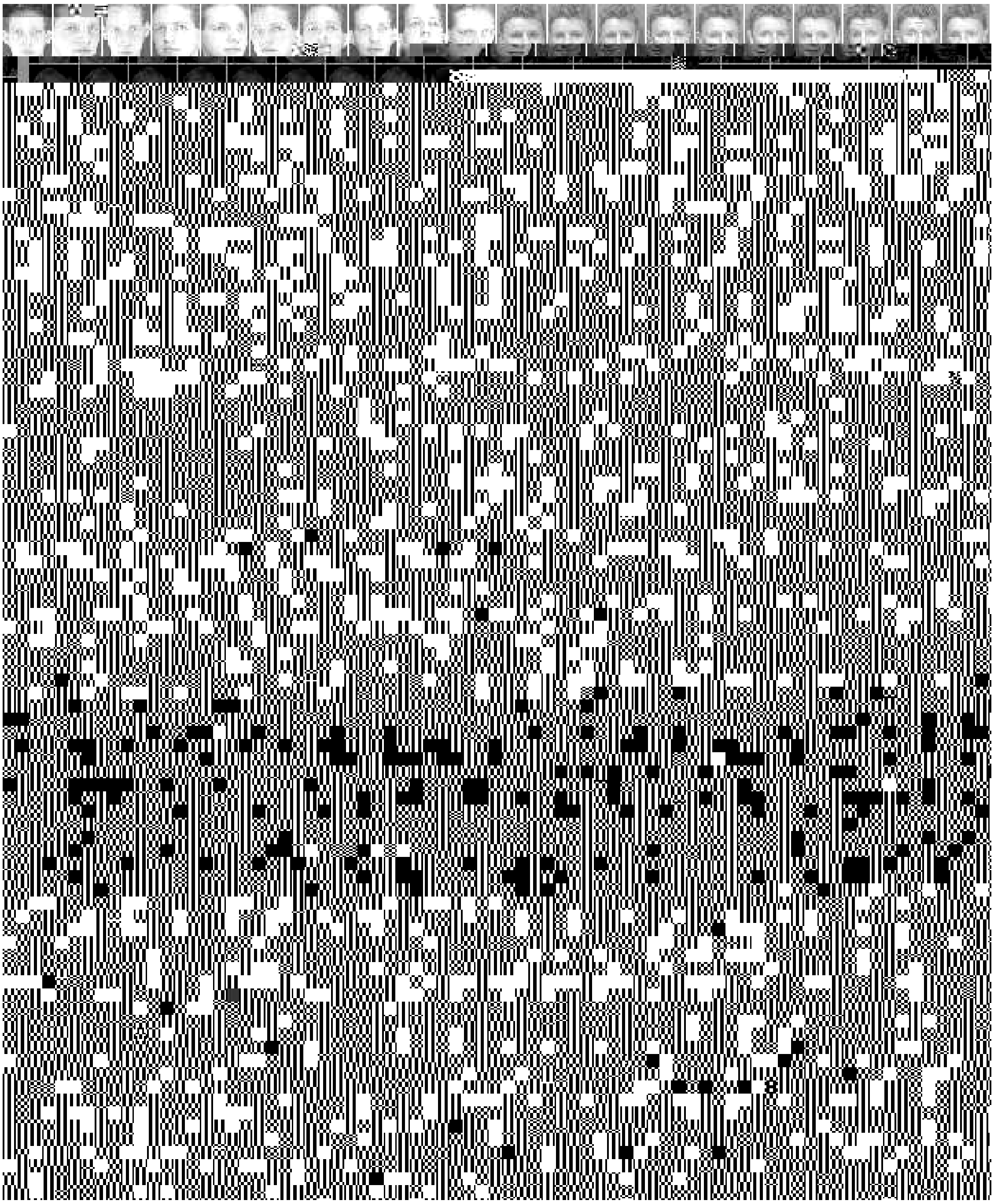
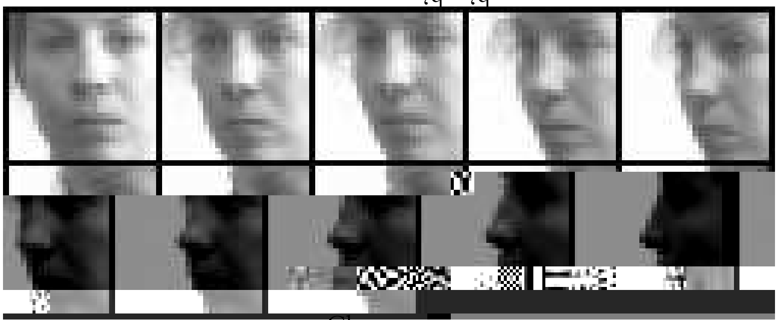


Figure A.1 The complete ORL Database

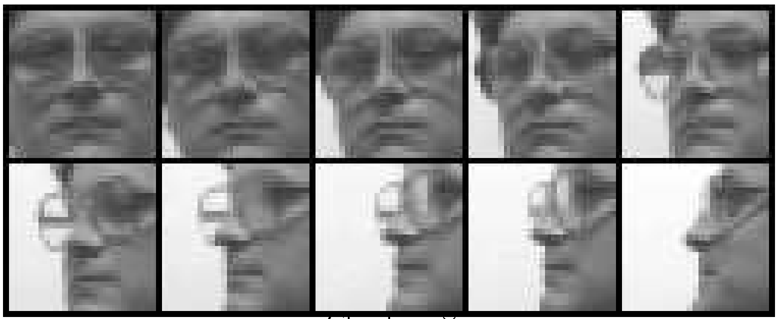
many publically available databases that could be used for this purpose The use of ten people is suf cient that the task is not trivial but not so large that computation is excessive

a Class [illegible] × [illegible]

b Class [illegible] × [illegible]

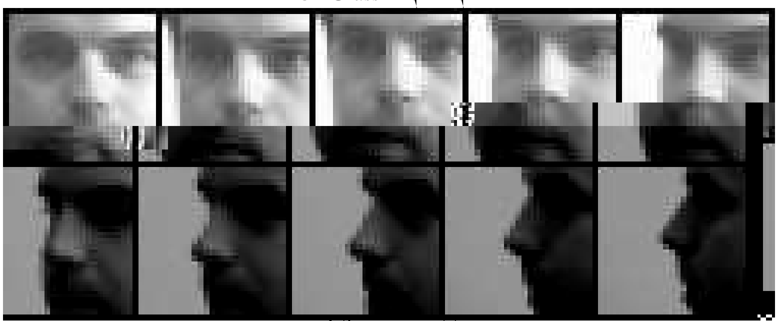


c Class [illegible] × [illegible]



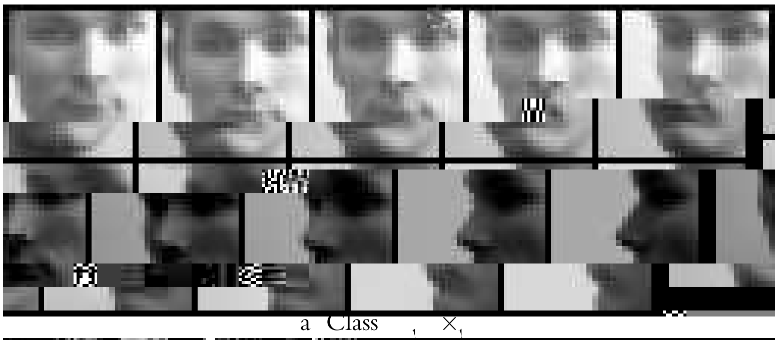
a Class , ×

b Class , ×



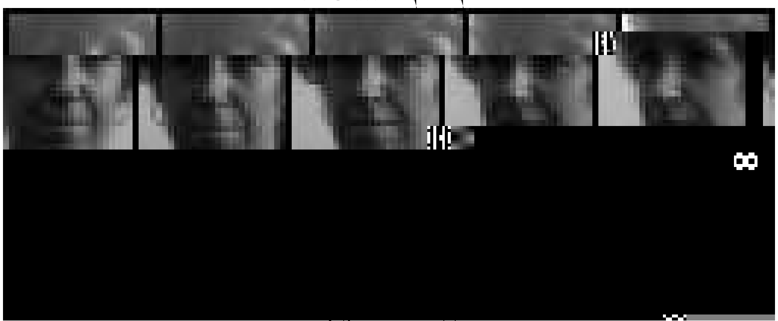
c Class , , ×

d Class , ×

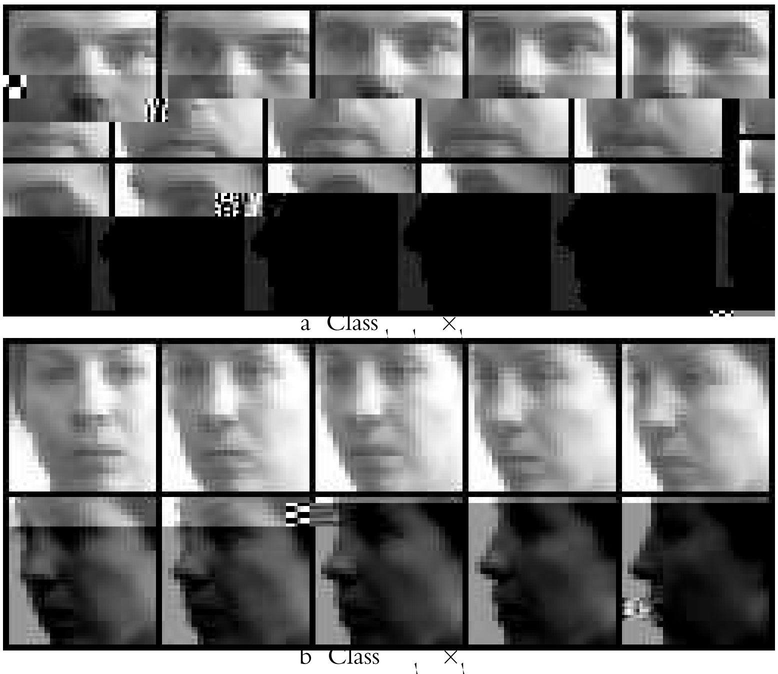


a Class ×

b Class ×



c Class ×

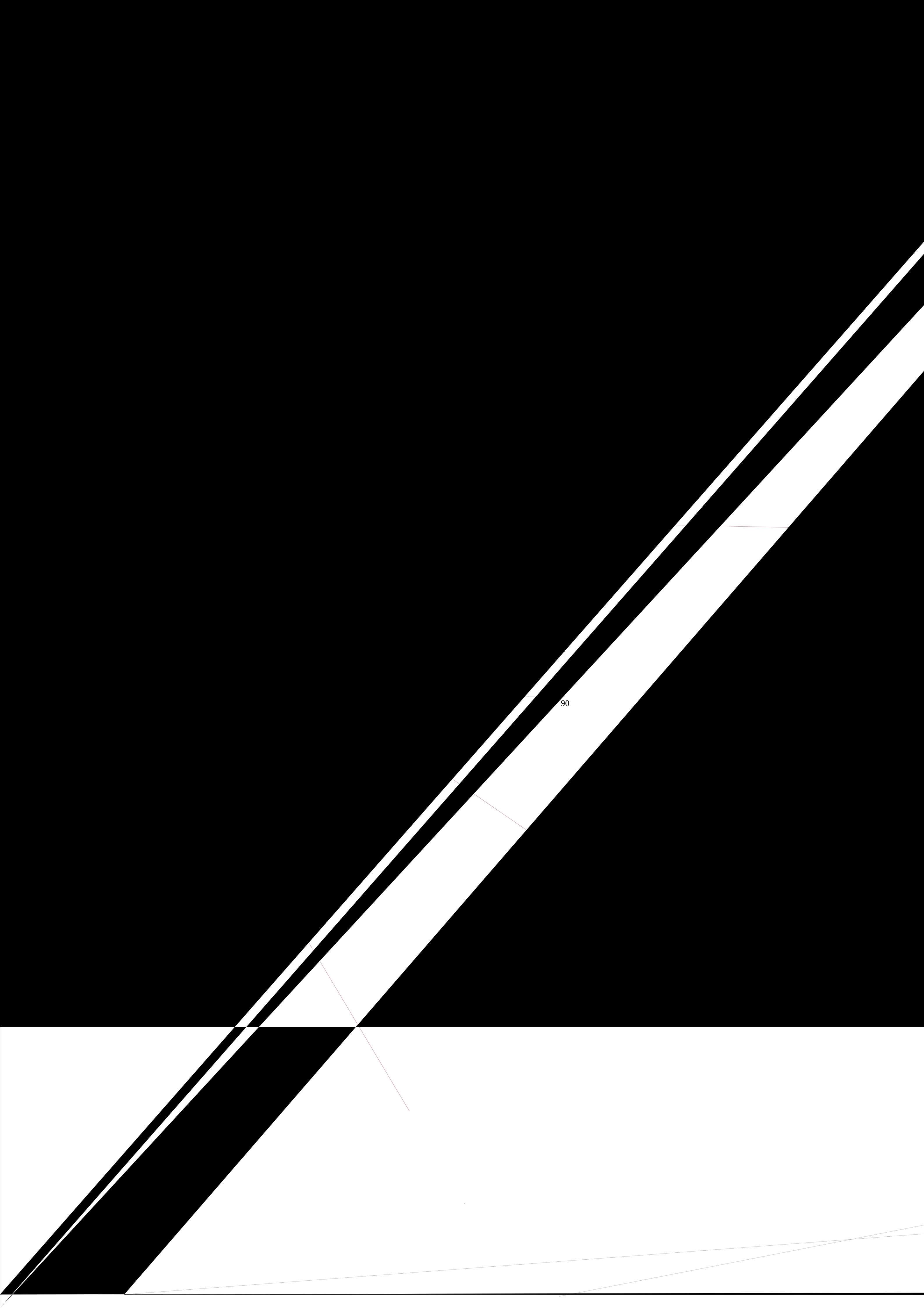


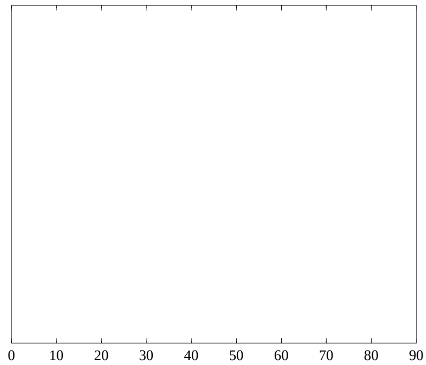
a Class , , ×,

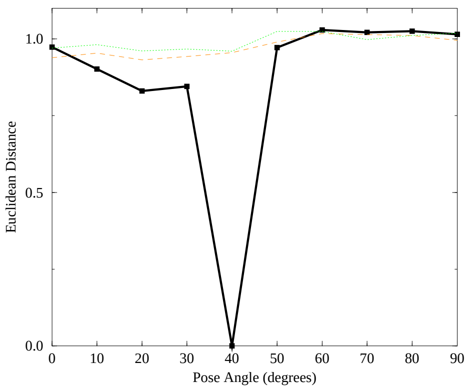
b Class × 

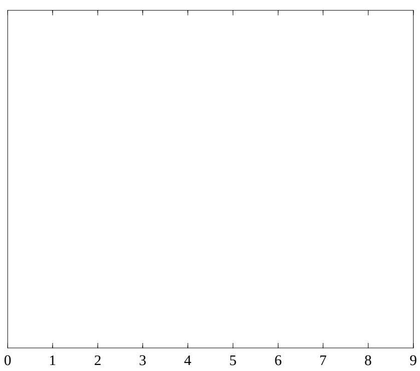
Figure A  As for Figure A  but using face centering rather than nose centering for localization of faces and only showing classes , and

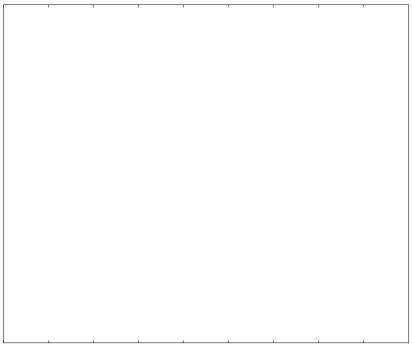
Note that this face centering technique only attempts to ll the image with as much surface area from the face as is possible A true pose free centering algorithm would use head mass for localization and the face area extracted would therefore contain the entire head


90


0
10
20
30
40
50
60
70
80
90

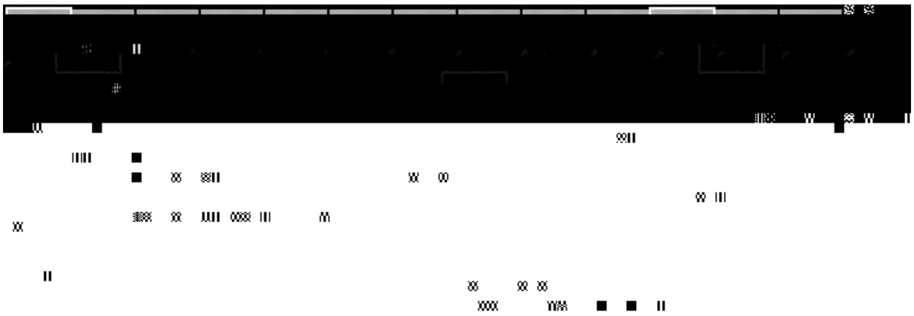

Euclidean Distance
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
90
Pose Angle (degrees)


0
1
2
3
4
5
6
7
8
9

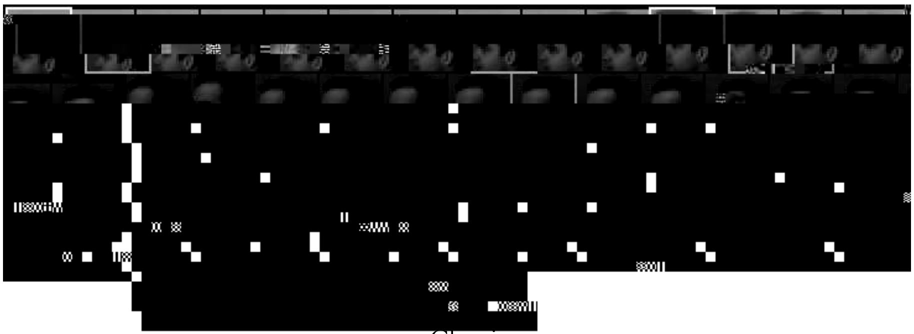


## A.3 QMW Image Sequences

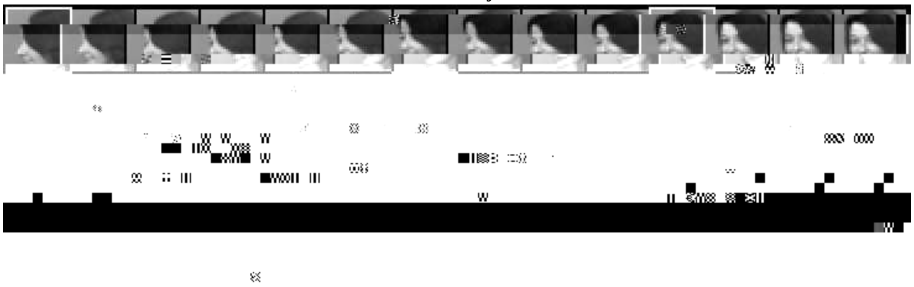
The image sequences used in the tests reported here are the re



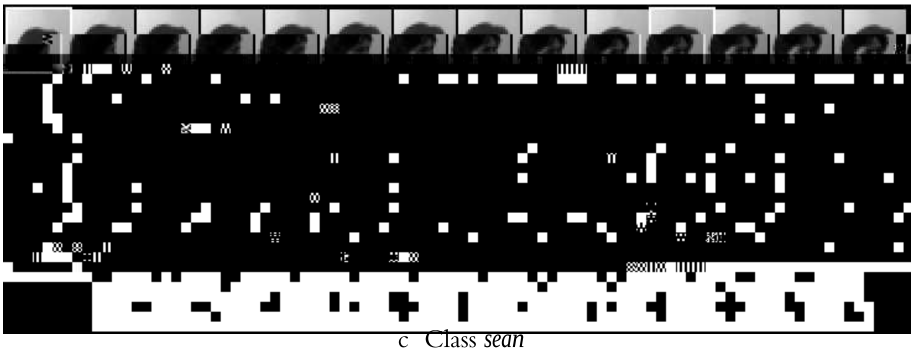
a Class



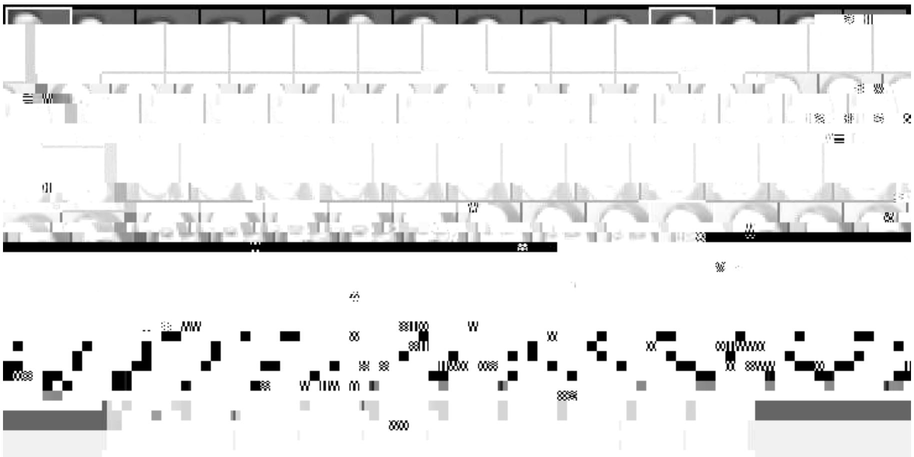
a Class *jon*



b Class *lorna*



c Class *sean*



d Class *steve*

Figure A As Figure A but the second four of the eight QMW sequences

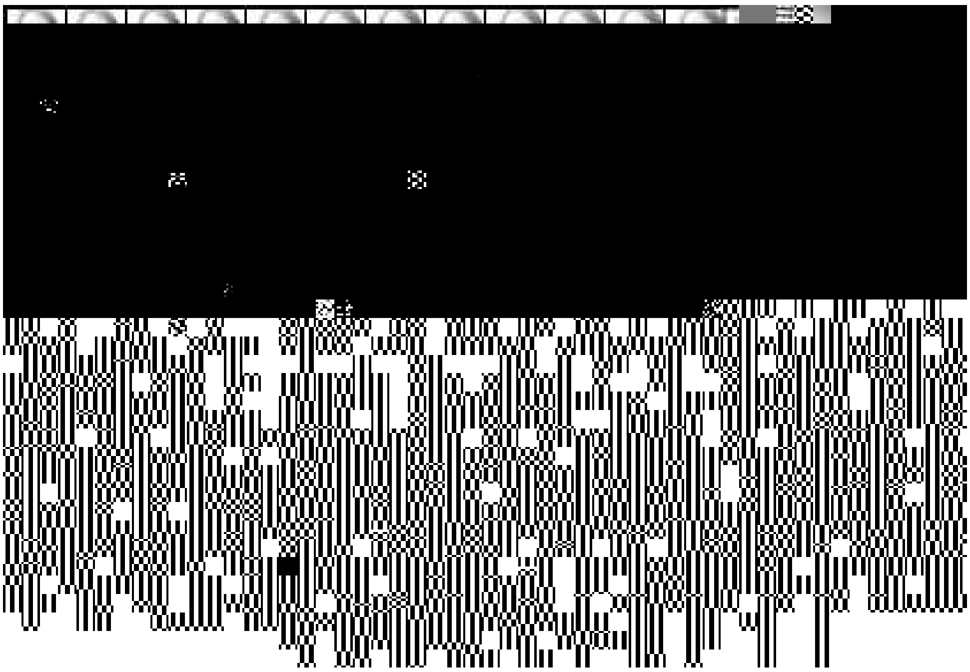


Figure A  A complete Secondary sequence for class *steve* after segmentation but before pre processing This shows the high level of lighting and pose variation which was designed to test the RBF network s generalization to conditions different to those used for training As only front view face detection has been implemented at this stage some non face frames are included and pro le views although segmented are incorrectly scaled

| Heuristic for σ Values | Initial % | % Discarded | % After Discard |
|---|---|---|---|
| Individual NN | | | |

### B.1.2 Hidden Unit Activations

The unnormalised output $u$ for hidden unit $h$ for a pattern $l$ uses a Gaussian function which can be expressed as

$$u_h^l = \exp[-\frac{(r_h^l)^2}{2\sigma_h^2}] \qquad \text{B.1}$$

where in this case $r$ is the Euclidean distance

$$r_h^l = d_E(\mathbf{j}^l, \mathbf{c}_h),$$

$$d_E(\mathbf{j}, \mathbf{c}) = \sqrt{\sum_{x=1}^{N} (j_x - c_x)^2}. \qquad \text{B.2}$$

This is the distance between the $N$ dimensional input vector $\mathbf{j}$ and hidden unit centre $\mathbf{c}$ Note

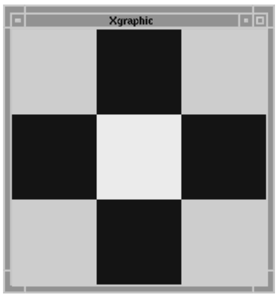
This is combined with two more xed parameters which control the speed of change $\eta$ the learning rate and $\gamma$ a momentum term to give the change in weight value $\Delta w_{ih}$

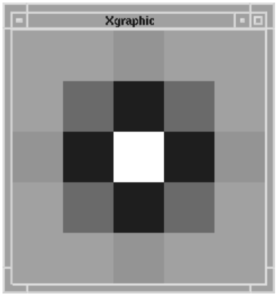
$$\Delta w_{ih}^{l} = \eta \delta_i^l \phi_h^l + \gamma \Delta w_{ih}^{l-1} \qquad \text{B}$$

# Appendix C

# Preprocessing Techniques

This appendix describes the speci c implementation of the two preprocessing techniques used



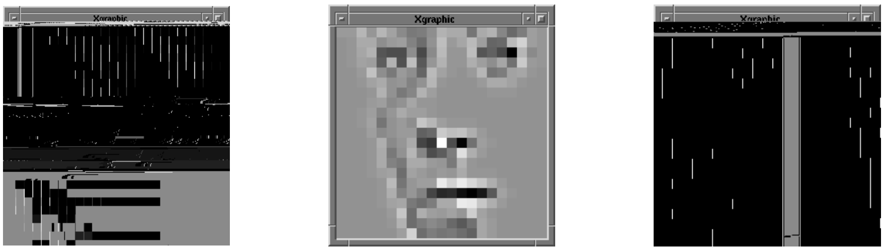

a Scale b Mask ×

| Original Resolution | DoG Scale | Convolved Resolution | Samples per Image |
|---|---|---|---|
| [illegible] × [illegible] | [illegible] | [illegible] × [illegible] | [illegible] |
| [illegible] × [illegible] | [illegible] | × | [illegible] |
| [illegible] × [illegible] | [illegible] | [illegible] × [illegible] | |
| [illegible] × [illegible] | [illegible] | [illegible] × [illegible] | [illegible] |
| [illegible] × [illegible] | [illegible] | × | [illegible] |

Table C [illegible] Resolutions of face data used from the Sussex database and the DoG preprocessing values for each image size

i ii iii

a full range of grey levels



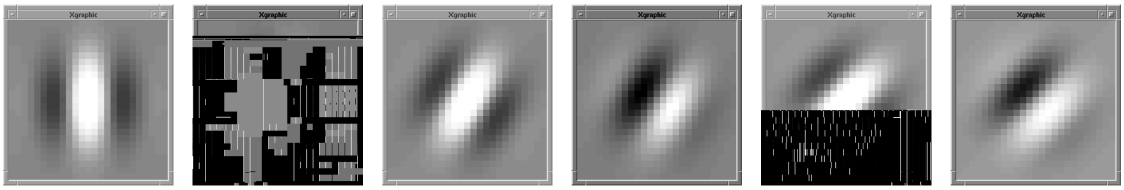
i ii iii

b reduced range of grey levels

Figure C [illegible] Effect of different ranges of grey levels for DoG preprocessing using a [illegible] × [illegible] image i before preprocessing ii after non thresholded DoG preprocessing iii after thresholded DoG preprocessing

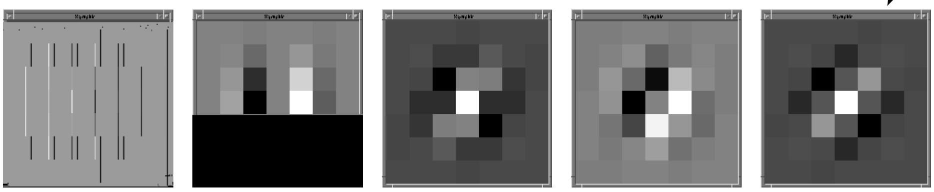
### C.1.2 Image Grey-Level Range

The range of grey levels present in the images can be reduced if it is considered that the areas of




i ii iii iv v vi

a $_{1} \times _{1}$ masks created from a Gabor function of period



i ii iii iv

| Scheme | Orientations | Scales | Over | Matrix | Coef cients |
|---|---|---|---|---|---|

| Number of Samples | Period | Mask Size |
|---|---|---|
| × × × × | • | × × × × • • × × |

a A sampling

| Number of Samples | Period | Mask Size |
|---|---|---|
| × × × × | | × × × × • • |

b B sampling

| Number of Samples | Period | Mask Size |
|---|---|---|
| × × × × | | × × • • × × |

c C sampling

| Number of Samples | Period | Mask Size |
|---|---|---|
| • • | | × × • • × × |

d D sampling

| Number of Samples | Period | Mask Size |
|---|---|---|
| × × × | • | × × × • • |

e E sampling

Table C Sampling and lter masks used for different Gabor preprocessing schemes

| Scale Combination | Sampling × | × | × | × | Coef cients |
|---|---|---|---|---|---|
| | • | | | | |
| | | • | | | |
| | • | • | | | |
| | | | • | | |
| | • | | • | | |
| | | • | • | | |
| E | • | • | • | | |
| | | | | • | |
| | • | | | • | |
| | | • | | • | |
| | • | • | | • | |
| | | | • | • | |
| | • | | • | • | |
| | | • | • | • | |
| A | • | • | • | • | |

Table C Numbers of coef cients for different A Gabor lter scale and sampling combinations The arrangement is equivalent to standard A sampling to E sampling See Table C a for details of lter masks at each sampling level