Receptive Field Functions for Face Recognition

A. Jonathan Howell and Hilary Buxton

UNIVERSITY OF

SUSSEX

AT BRIGHTON

Cognitive Science
Research Papers

Receptive Field Functions for Face Recognition

A. Jonathan Howell and Hilary Buxton

School of Cognitive and Computing Sciences,

University of Sussex, Falmer, Brighton BN1 9QH, UK

{jonh,hilaryb}@cogs.susx.ac.uk

Abstract

Preprocessing of face images was performed to mimic the effects of
receptive field functions found at various stages of the human vision sys-
tem. These were then used as input representations to Radial Basis Func-
tion (RBF) networks that learnt to classify and generalise over different
views for a standard face recognition task. Two main organisations of
the RBF networks (standard and face unit) and two main types of pre-
processing (Difference of Gaussian filtering and Gabor wavelet analysis)
were compared. Quantitative and qualitative differences in these schemes
are described and conclusions drawn about the best approach for our face
recognition problem using low resolution images.

1 Introduction

Face recognition has been the subject of a great deal of research in computer
vision and work on biologically-motivated approaches has begun to deliver real
solutions. One of the main problems is dimensionality reduction to remove
much of the redundant information in the original images. There are many
possibilities for effectively representing this data, including principal component
analysis, Gabor filters and various isodensity map or feature extraction schemes.
A well known example is Turk & Pentland (1991) which is widely acknowledged
as a practical approach. More recent work has improved on this (Pentland et
al. 1994, Petkov et al. 1993, Rao & Ballard 1995) and has removed some of
the restrictions on the range of scales and orientations required by the original
eigenface scheme. In particular, it seems that appropriate preprocessing of
input representations for a face recognition scheme can overcome the problems
of lighting variation and multiple scales. Other sources of variation such a view
and expression still remain.

In our work (Howell & Buxton 1995a, Howell & Buxton 1995b) we use
an adaptive learning component based on RBF networks to tackle the uncon-
strained face recognition problem. We want our face recognition scheme to

generalise over a wide range of conditions to capture the essential similarities of a given face. In this paper, we are concentrating on the issues of finding an effective input representation for our networks. In particular, we contrast the use of Difference of Gaussian filtering and Gabor wavelet analysis at a range of scales. One way of thinking about these input representations and mapping them onto our RBF networks is to use the analogy with visual neurons. The receptive field of such a neuron is the area of the visual field (image) where the stimulus can influence its response. For the different classes of these neurons, a receptive field function $f(x, y)$ can be defined. For example, retinal ganglion cellas and lateral geniculate cells early in the visual processing have receptive fields which can be implemented as Difference of Gaussian filters (Marr & Hildreth 1980). Later, the receptive fields of the simple cells in the primary visual cortex are oriented and have characteristic spatial frequencies. Daugman (1988) proposed that these could be modelled as complex 2-D Gabor filters. Petkov et al. (1993) successfully implemented a face recognition scheme based on Gabor wavelet input representations to imitate the human vision system. The question we want to ask here is whether these later stages of processing make more information explicit than the earlier DoG filters for our face recognition task.

2   The RBF Network Model

The RBF network is a two-layer, hybrid learning network (Moody & Darken 1988), with a supervised layer from the hidden to the output nodes, and an unsupervised layer, from the input to the hidden, where individual radial Gaussian functions for each hidden unit simulate the effect of overlapping and locally tuned receptive fields. This gives an activation that is related to the relative proximity of the test data to the training data, allowing a direct measure of confidence in the output of the network for a particular pattern. In addition, if the pattern is more than slightly different to those trained, very low (or no) output will occur.

For the following tests, two types of network were used: a 'standard' RBF model and a 'face unit' RBF model. The standard network is trained with all possible classes from the data with a 'winner-takes-all' output strategy, whilst the 'face unit' network produces a positive signal only for the particular person it is trained to recognise. For each individual, a 'face unit' RBF network can be trained to discriminate between that person and others selected from the data set, using 'pro' and 'anti' evidence for and against the individual. Details can be found in Howell & Buxton (1995b). Although this second approach increases complexity, 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 training times for the 'standard' model quickly become impractical as the number of classes increases.

Grey-Levels	Fixed %	Fixed % with 1.5 Discard	Epochs
Full	88	95	1148381
Reduced	92	100	335524

(a)

Grey-Levels	Ave. %	Ave. % with 1.5 Discard	Ave. Epochs
Full	94	97	72040
Reduced	93	96	7586

(b)

Table 1: Training with no preprocessing, with full and reduced grey-levels (a) Standard 50/50 RBF Network (b) 6+12 Face Unit RBF Network

3   Form of Test Data

Lighting and location for the training and test face images in these initial studies has been kept fairly constant to simplify the problem. For each individual to be classified, ten images of the head and shoulders were taken in ten different positions in 10° steps from face-on to profile of the left side, 90° in all. This gave a data set of 100 8-bit grey-scale 384×287 images from ten individuals.

A 100×100-pixel 'window' was located manually in each image centred on the tip of the person's nose, so that visible features on profiles, for instance, should be in roughly similar locations to face-on. This 'window' region was sub-sampled to a variety of resolutions for testing. Full details are given in Howell & Buxton (1995a). The resolution of the images is represented as '$n \times n$', a resolution of 25×25 being used for the work reported here. The ratio of training and test images used is represented as 'train/test', eg '20/80', where 100 images were in the data set and 20 were used for training and 80 for test. The 'face unit' network size is denoted by '$p + a$', where p is the number of 'pro' hidden units, and a is the number of 'anti' hidden units. Tests were made on a range of network sizes from 1+1 to 6+12 (which are effectively 2/98 and 18/82 networks).

4   Training with No Pre-Processing

To gauge the effectiveness of later preprocessing methods, the networks were initially trained simply with the original grey-level information, giving 625 pixel values per image. A fixed discard threshold of 1.5 was used[1].

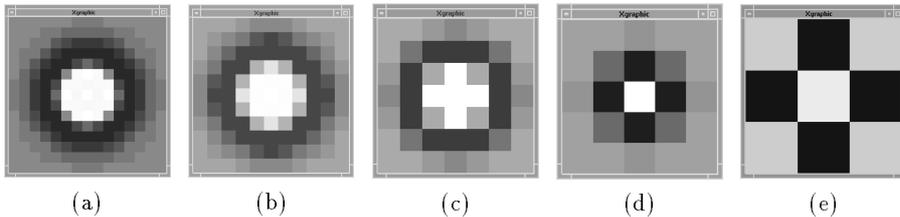[1] An optimal value for this parameter has not yet been determined.

Figure 1: Masks created from various DoG scales (with mask sizes): (a) 1.6 (15×15) (b) 1.2 (11×11) (c) 0.8 (7×7) (d) 0.4 (5×5) (e) 0.15 (3×3)

In summary, good generalisation performance was obtained, although the training times were unacceptably long. The reduction of the range of grey-levels gave very much shorter convergence times, though still very slow.

5 Difference of Gaussians (DoG) Pre-Processing

Where there is a change of intensity in an image, peaks or troughs are found in the first derivative of the intensity, and zero-crossings in the second derivative. To isolate the latter, Marr & Hildreth (1980) suggested the $\nabla^2 G$, or *Laplacian of the Gaussian*, operator, which can be closely apprimated by a *Difference of Gaussians* (DoG) operator, constructed from two Gaussians G of the form:

$$G(x, y) = \frac{1}{\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \tag{1}$$

where the space constants σ have a ratio of 1:1.6. The DoG masks were constructed using the POPVISION CONVOLVE_DOG_2D routines. Figure 5 shows these masks at various scale values, whilst Figure 2 shows the result of their convolution with an image at a fixed resolution.

5.1 DoG Gradients *vs.* 'Zero-Crossings'

With a typical, grey level image, such as Figure 3(a), DoG convolution will give continuously-valued[2] gradient information, as shown in Figure 3(b). Where these values change from one sign to the other is the 'zero-crossing' point; if the values are thresholded at 0 into either 0 (for negative) and 1 (for positive), the boundaries between black and white are the zero-crossings for the image, as shown in Figure 3(c). To test how useful it was to explicitly concentrate only on this boundary point, preprocessing was carried out with and without this thresholding stage. This produced both gradient DoG (non-binarised) and zero-crossings DoG (binarised) information.

True zero-crossings, ie lines following this zero contour around the gradients, were found to be almost the same as the binarised data, due to the low resolution

---

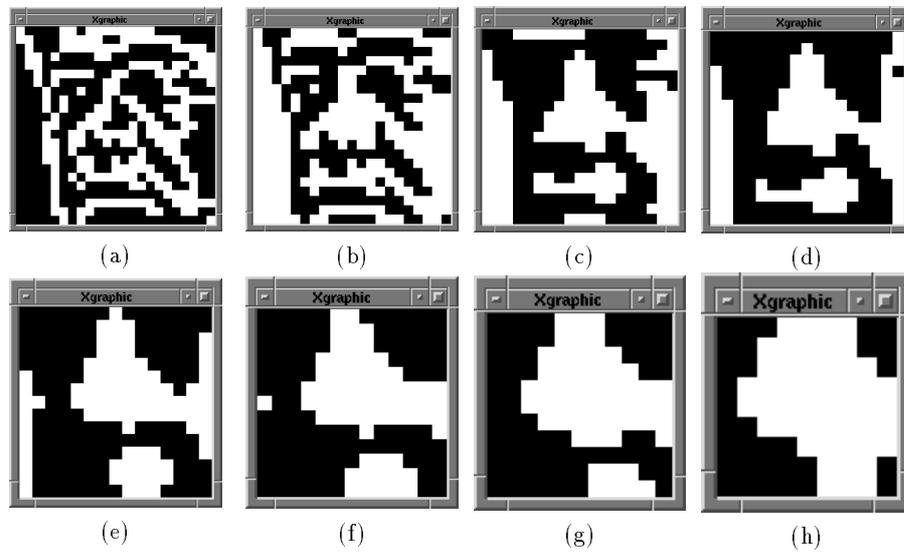[2]*ie.* with both positive and negative gradient values

4

Figure 2: DoG scales applied to 25×25 image (with convolved image sizes): (a) 0.15 (23×23) (b) 0.4 (21×21) (c) 0.8 (19×19) (d) 1.0 (17×17) (e) 1.2 (15x15) (f) 1.4 (13x13) (g) 1.6 (11x11) (h) 1.9 (9x9)
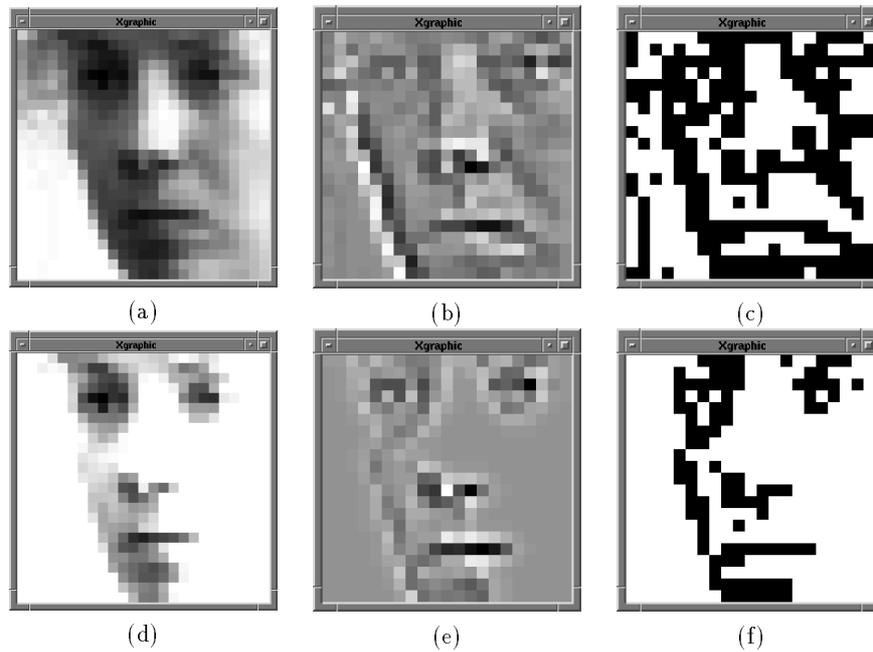

Figure 3: Effect of reducing range of grey-levels on 25×25 image (a) full range of grey-levels (b) after non-thresholded DoG (c) after thresholded DoG (d) reduced range of grey-levels (e) after non-thresholded DoG (f) after thresholded DoG

Scale	Thres-holding	Grey-Levels	Fixed %	Fixed % with 1.5 Discard	Epochs
0.4	No	Full	50	59	54515
0.4	No	Reduced	68	86	35801
0.4	Yes	Full	72	90	13311
0.4	Yes	Reduced	86	100	27463
0.15, 0.4, 0.8, 1.6	Yes	Reduced	78	90	19119

(a)

| Scale	Thres-holding	Grey-Levels	Ave. %	Ave. % with 1.5 Discard	Ave. Epochs
0.4	No	Full	74	79	4362
0.4	No	Reduced	82	86	2430
0.4	Yes	Full	81	87	544
0.4	Yes	Reduced	89	93	884
0.15, 0.4, 0.8, 1.6	Yes	Reduced	96	97	916

(b)

Table 2: Non-Thresholded (Gradient) and Thresholded (Zero-Crossings) DoG Preprocessing, with one and four DoG scales, giving 441 and 1452 samples per image respectively (a) Standard 50/50 RBF Network (b) 6+12 Face Unit RBF Network

used. For this reason, tests were not done on this type of data. It should also be noted that bare zero-crossings lines are not as informative as the binarised data, as they have lost the sign of the gradient value, as the latter makes explicit whether the change is from positive to negative or the reverse.

5.2 Summary of DoG Results

Training with 'zero-crossings' thresholded data gave better generalisation and faster convergence when compared with tests using the un-thresholded 'gradient' data (see Table 3). Similarly, the use of data with reduced grey-levels gave better generalisation compared to tests using the full range of grey-levels, though at a higher computation cost shown by the slower convergence. The use of multiple DoG scales also improved performance, but required four times as much data than for one scale.

The scale parameter is shown to have a clear effect on generalisation and convergence rates, as shown by Figures 4 and 5.
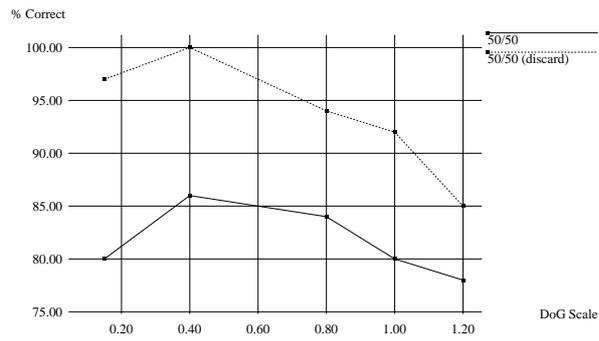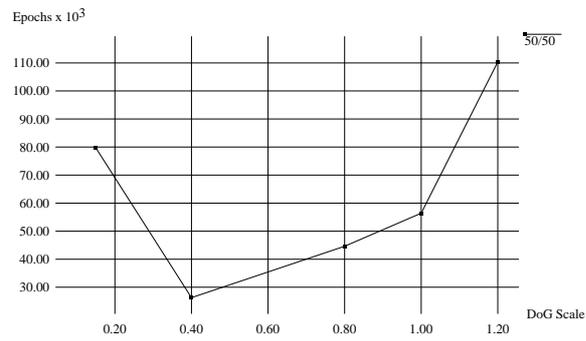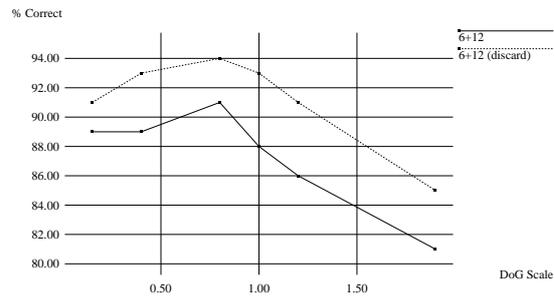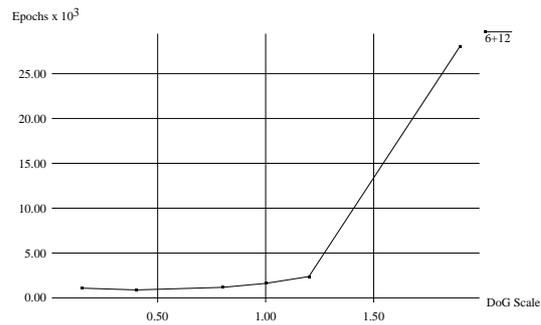
Figure 4: Effect of varying the scale in DoG pre-processing (a) on test general-isation (b) on training epochs with 50/50 RBF networks

Figure 5: Effect of varying the scale in DOG pre-processing (a) on test general-isation (b) on training epochs with 6+12 RBF 'face unit' networks
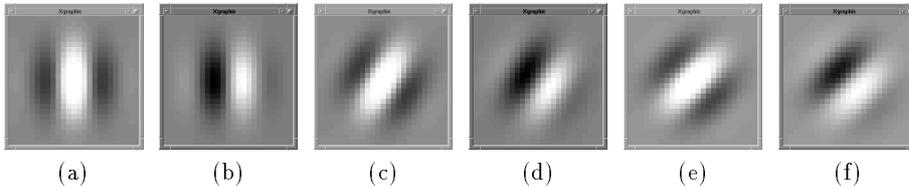
Figure 6: 25×25 masks created from Gabor filter of period 13: (a) 0° real (b) 0° imaginary (c) 30° real (d) 30° imaginary (e) 45° real (f) 45° imaginary
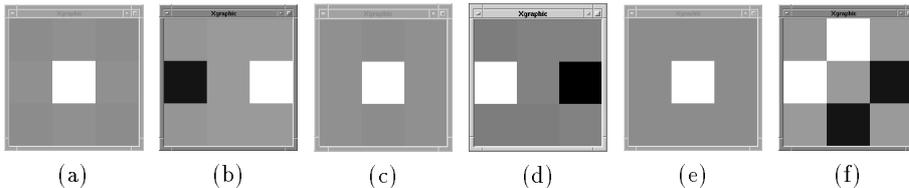

Figure 7: 3×3 masks created from Gabor filter of period 1: (a) 0° real (b) 0° imaginary (c) 30° real (d) 30° imaginary (e) 45° real (f) 45° imaginary

6   Gabor Pre-Processing

6.1   Gabor scales and orientations

We have selected Gabor filters (Daugman 1988) as an alternative preprocesing method, as it provides oriented information, which, we hope, will provide input information for the network in a more useful form than the previous methods. One disadvantage of isolated orientation-specific value is that if a full convolution of the image is carried out, more values are output than input (as there is a data value for each pixel for each orientation required). In addition, there are sine and cosine components of the Gabor filter, which doubles the number of coefficients produced.

The Gabor masks were constructed using the POPVISION GABORMASK routines, using three parameters: σ for width, p the period of the harmonic component ($\sigma = p/(2\sqrt{2})$), and o the orientation of the mask.

The real (cosine) component, C, of the Gabor mask is calculated as:

$$C(x, y) = N \exp(-\frac{r^2}{2\sigma^2})cos(x'\omega), \qquad (2)$$

where $r^2 = x^2 + y^2$, $x' = x\ cos(o) + y\ sin(o)$, and $\omega = (2\pi)/p$, and N is a real normalisation constant. The imaginary (sine) component, S, is:

$$S(x, y) = N \exp(-\frac{r^2}{2\sigma^2})sin(x'\omega). \qquad (3)$$

9

Scheme	Orien- tations (degrees)	Scales	Over- lapping	Matrix	Coeffic- ients Per Image
A1	0	4	No	Square	170
A2	0, 180	4	No	Square	340
A3	0, 120, 240	4	No	Square	510
A3X	60, 180, 300	4	No	Square	510
A3S	30, 150, 270	4	No	Square	510
A4	0, 90 180, 270	4	No	Square	680
A6	30, 90, 150 210, 270, 330	4	No	Square	1020
B3	0, 120, 240	4	Most	Square	510
C3	0, 120, 240	4	Less	Square	510
D3	0, 120, 240	3	No	Circular	420

Table 3: Types of Gabor sampling schemes tested, with filter orientations and number of coefficients sampled per image

6.2 Gabor Sampling Schemes

In order to reduce the number of coefficients calculated for each image, sparse sampling schemes were constructed, with a range of scales. The 'A' square matrix sampling scheme which had the least amount of overlap on sampling points, proved to be the most successful arrangement. Others were tested which used large amounts of overlap on the sampling receptive fields, or circular sets of sampling points; Table 3 summarises the different sampling schemes used. Tables 4(a) and (b) show the sampling arrangements for the 'A' and 'B' square matrix sampling schemes, with Figures 8(a) and (b) showing how these masks were positioned to cover the image area. Note that the 'A' scheme only covers 24×24 at the 8×8 scale and the some overlap was needed to fit the 2×2 and 4×4 scales.

The 'C' square matrix sampling scheme (Table 4(c) and Figure 8(c)) was devised after the 'B' scheme performed poorly. The scales used were intended to retain fine detail from the original image.

Table 4(d) and Figure 8(d) show similar details for the 'D' circular matrix sampling scheme. Note that due to the fairly coarse alignment to pixel boundaries in the low resolution 25×25 image area, some masks placements do not coincide with the exact mathematical position.

Number of Samples	Period	Mask Size
1×1	13	25×25
2×2	7	13×13
4×4	3	7×7
8×8	1	3×3

(a)

Number of Samples	Period	Mask Size
1×1	13	25×25
2×2	10	19×19
4×4	5	11×11
8×8	3	7×7

(b)

Number of Samples	Period	Mask Size
1×1	13	25×25
2×2	9	17×17
4×4	4	9×9
8×8	2	5×5

(c)

Number of Samples	Period	Mask Size
7	4	9×9
7	3	7×7
19	2	5×5
37	1	3×3

(d)

Table 4: Masks used for the Gabor schemes: (a) 'A' (b) 'B' (c) 'C' (d) 'D'
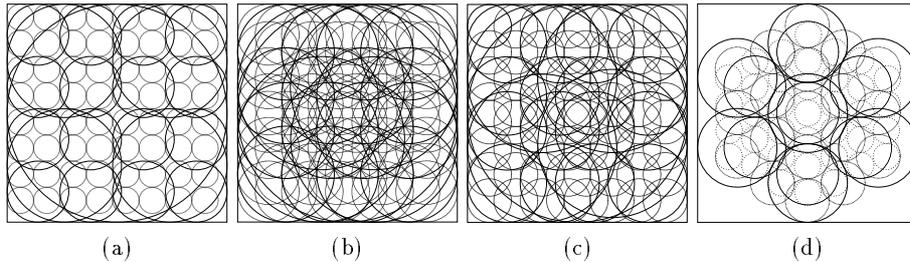

(a)   (b)   (c)   (d)

Figure 8: Sampling positions for Gabor sampling schemes: (a) 'A', least overlap, square matrix (b) 'B' most overlap, square matrix (c) 'C' less overlap, square matrix (d) 'D' least overlap, circular matrix

Scheme	Coefficients per Image	Fixed %	Fixed % with 1.5 Discard	Epochs
A3	510	92	95	35752
A3R	510	92	100	40016
Non-Thresholded A3	510	82	90	330035
A3 (Sine mask only)	255	86	95	30806
A3 (Cosine mask only)	255	46	72	5555197
B3	510	86	97	28832
C3	510	84	92	24122
D3	420	82	94	40532

Table 5: Gabor Preprocessing for standard 50/50 RBF Network (all schemes used full range of grey-levels except A3R)

7 Gabor Results

Binarisation of the coefficients was found to increase test generalisation, and dramatically reduce training times, but the advantage shown in using reduced grey-levels for DoG preprocessing was *not* seen for Gabor preprocessing (see Table 6). No advantage was found for using more than three orientation angles for the training data (see Figures 8 and 9).

The coarse nature of the masks at the 3×3 resolution is illustrated by Figure 8, where the real (cosine) masks at different orientations look very similar. Table 6 shows the effect of that when the individual masks were separated: the data set with only sine coefficients performs as well as the joint dataset, whilst the data set with only sine coefficients does not perform well.

The importance of the choice of orientation angles used for Gabor preprocessing was demonstrated in the improvement in performance of tests with two and four orientations when the angles used were modified. This may be due to the A2 and A4 having 180° offsets, unlike the A3, where no angle was at 180° to another. The bar shape of the Gabor mask may mean that the information extracted from 2 masks perpedicular to (180° apart from) each other may not gather information that is radically different from each other.

The standard RBF network seems to be more sensitive to the orientations used in Gabor pre-processing than the face unit network (see Figures 9 and 10), but the predicted advantage of overlapping receptive fields, as used for the B3 tests, over non-overlapping, eg A3, was not demonstrated. The C3 scheme, which uses smaller scale masks than B3, performs slightly better than the B3, but still well below A3.

Scheme	Coefficients per Image	Ave. %	Ave. % with 1.5 Discard	Ave. Epochs
A3	510	96	98	654
A3R	510	95	98	755
Non-Thresholded A3	510	91	95	8288
A3 (Sine mask only)	255	94	99	368
A3 (Cosine mask only)	255	83	88	18980
B3	510	88	100	698
C3	510	92	96	556
D3	420	89	93	881

Table 6: Gabor Preprocessing for 6+12 Face Unit RBF Network (all schemes used full range of grey-levels except A3R)
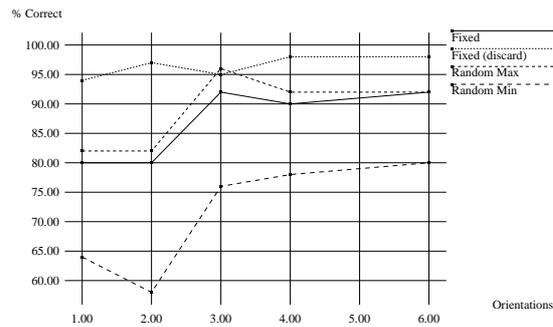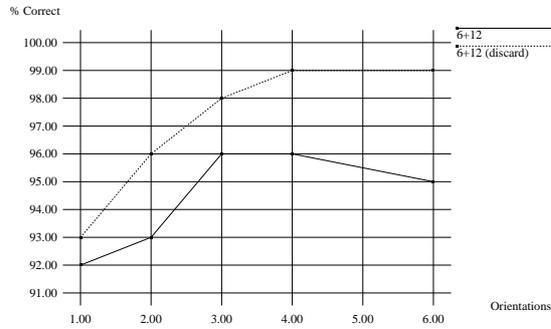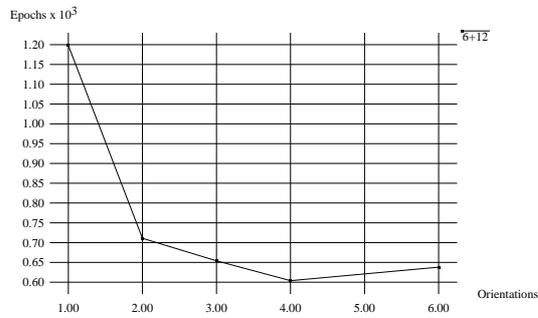

(a)
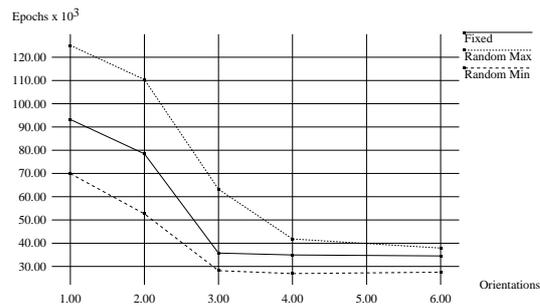
(b)

Figure 9: Effect of varying the number of orientations in Gabor pre-processing (a) on test generalisation (b) on training epochs for 50/50 RBF networks

13

% Correct

Orientations

(a)

Epochs x 10^3

Orientations

(b)

Figure 10: Effect of varying the number of orientations in Gabor pre-processing (a) on test generalisation (b) on training epochs for 6+12 RBF 'face unit' networks
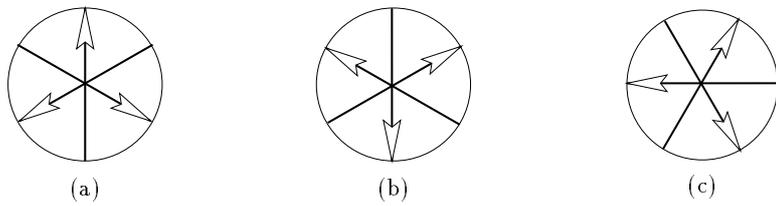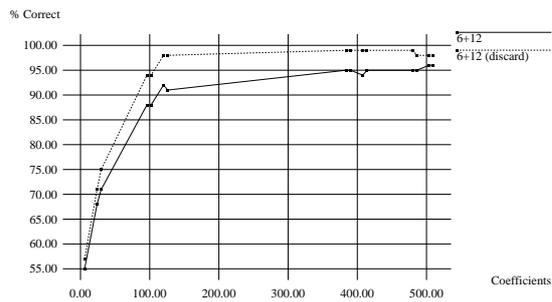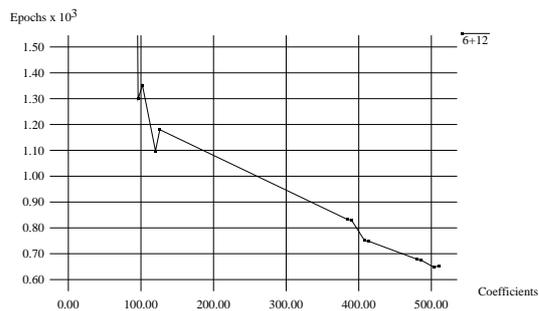

(a)　　　　　(b)　　　　　(c)

Figure 11: Angles used for three-orientation tests: (a) A3 (b) A3X (c) A3S

14

(a)

(b)

Figure 12: Effect of specific arrangements of scales in Gabor pre-processing (a) on test generalisation (b) on training epochs for 6+12 RBF 'face unit' networks, with total number of coefficients for the combination of scales

Sampling	1×1	2×2	4×4	8×8	
Mask Size	25×25	13×13	7×7	3×3	
Combination					Coefficients
1	•				6
2		•			24
21	•	•			30
4			•		96
41	•		•		102
42		•	•		120
421	•	•	•		126
8				•	384
81	•			•	390
82		•		•	408
821	•	•		•	414
84			•	•	480
841	•		•	•	486
842		•	•	•	504
8421	•	•	•	•	510

Table 7: Sampling and mask sizes for Gabor scale combinations

8 Effect of Scales in Gabor Pre-Processing

Tests with individual scales (see Table 8) were made to investigate the effect of individual scales on the overall performance of the network. Identification of redundant scales could significantly reduce the number of coefficients, and therefore the computation required,

The results (see Figure 12) show that quite dramatic savings can be made in the amount of information sampled from the images without a large loss of test generalisation or impractical increase in epochs for training convergence. For example, the A3-421 scheme uses only 126 coefficients compared to the 510 used for the standard A3 (and thus using only 25% of the original data), and, for the face units tests at least, shows minimal loss of performance. In addition, the effect of the individual Gabor scales was not shown to be additive, eg the A3-8 and A3-421 schemes perform similarly.

9 Conclusion/Future Work

In summary, the RBF face unit organisation together with the Gabor wavelet preprocessing give the best combination of results on our face recognition task. The 'face unit' organisation gives a flexible, scaleable architecture that performs at a high level in terms of classification, generalisation, and speed of training. It is also highly modular so more face units can be added as required. The Gabor wavelet preprocessing is successful as it allows the number of coeeficients at

16

different scales and orientations to be closely tailored to the task at hand. The DoG preprocessing, on the other hand, results in an image-like representation with as many coeeficients as there are pixels. In our future work, we will extend the face unit RBF scheme and look at the problem of tracking faces in image sequences. We will also extend the Gabor preprocessing scheme to the space-time case.

Acknowledgments

References

Daugman, J. G. (1988), 'Complete discrete 2-D gabor transforms by neural networks for image analysis and compression', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**(7), 1169–1179.

Howell, A. J. & Buxton, H. (1995*a*), 'Invariance in radial basis function neural networks in human face classification', *Neural Processing Letters* **2**(3), 26–30.

Howell, A. J. & Buxton, H. (1995*b*), A scaleable approach to face identification, *in* 'Proceedings of International Conference on Artificial Neural Networks (ICANN'95)', Vol. 2, EC2 & Cie, Paris, pp. 257–262.

Marr, D. & Hildreth, E. (1980), 'Theory of edge detection', *Proc. R. Soc. London* **B207**, 187–217.

Moody, J. & Darken, C. (1988), Learning with localized receptive fields, *in* D. Touretzky, G. Hinton & T. Sejnowski, eds, 'Proceedings of the 1988 Connectionist Models Summer School', Morgan Kaufmann, pp. 133–143.

Pentland, A., Moghaddam, B. & Starner, T. (1994), View-based and modular eigenspaces for face recognition, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', pp. 84–91.

Petkov, N., Kruizinga, P. & Lourens, T. (1993), Biologically motivated approach to face recognition, *in* 'Proceeding of International Workshop on Artificial Neural Networks', pp. 68–77.

Rao, R. P. N. & Ballard, D. H. (1995), Natural basis functions and topographic memory for face recognition, *in* 'Proceeding of International Joint Conference on Articial Intelligence (IJCAI'95)', Montréal, Canada, pp. 10–17.

Turk, M. & Pentland, A. (1991), 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience* **3**(1), 71–86.