Improving Generalisation in Radial Basis Function Networks for Face Recognition

A. Jonathan Howell and Hilary Buxton

CSRP 416

April 1996

ISSN 1350-3162

UNIVERSITY OF SUSSEX AT BRIGHTON

Cognitive Science Research Papers

Improving Generalisation in Radial Basis Function Networks 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

April 1996

Abstract

This paper presents experiments using an adaptive learning component based on Radial Basis Function (RBF) networks to tackle the unconstrained face recognition problem using low resolution video information. Firstly, we performed preprocessing of face images to mimic the effects of receptive field functions found at various stages of the human vision system. These were then used as input representations to RBF networks that learnt to classify and generalise over different views for a standard face recognition task. Two main types of preprocessing (Difference of Gaussian filtering and Gabor wavelet analysis) are compared. Secondly we provide an alternative, 'face unit' RBF network model that is suitable for large-scale implementations by decomposition of the network, which avoids the unmanagability of neural networks above a certain size. It uses small, individual networks for each class and allows the addition of new data to the database without complete re-training of the system. Finally, we show the 2-D shift, scale and y-axis rotation invariance properties of the standard RBF network. Quantitative and qualitative differences in these schemes are described and conclusions drawn about the best approach for real applications to address the face recognition problem using low resolution images.

Introduction

The human face poses several severe tests for any visual system: the high degree of similarity between different faces, the extent to which expressions and hair can alter the face, and the large number of angles from which a face can be viewed in common situations. A face recognition system must be robust with respect to this variability and generalise over a wide range of conditions to capture the essential similarities for a given human face. It is only recently that work on biologically-motivated, statistical approaches to face recognition has begun to deliver real solutions. One of the main problems that these approaches tackle is dimensionality reduction to remove much of the redundant information in the original images. There are many possibilities for such representations of the data, including principal component analysis, Gabor filters and various isodensity map or feature extraction schemes. A well known example is the work of Turk & Pentland (1991), on the 'eigenface' approach, which is widely acknowledged to be useful for practical application. However, the need for representations at a range of scales and orientations causes extra complexity and updating the average eigenface (used for localisation) when new faces are added to the dataset are problems for this scheme. These difficulties have been overcome to some extent in later work by various researchers (Pentland et al. 1994, Petkov et al. 1993, Rao & Ballard 1995). 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 as face orientation, expression, occlusion etc. still remain.

In our work we use an adaptive learning component based on RBF networks to tackle the unconstrained 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. The RBF network has been identified as valuable model by a wide range of researchers (Moody & Darken 1988, Poggio & Girosi 1990*b*, Girosi 1992, Musavi et al. 1992, Ahmad & Tresp 1993, Bishop 1995). Its main characteristics are first, its computational simplicity (only one layer involved in supervised training which gives fast convergence), and second, its description by a well-developed mathematical theory (resulting in statistical robustness). RBFs are seen as ideal for practical vision applications by (Girosi 1992) as they are good at handling sparse, high-dimensional data (common in images), and because they use approximation which is better than interpolation for handling noisy, real-life data. RBF networks are claimed to be more accurate than those based on Back-Propagation (BP), and they provide a guaranteed, globally optimal solution via simple, linear optimisation. An RBF interpolating classifier (Edelman et al. 1992), was effective and gave performance error of only 5–9% on generalisation under changes of orientation, scale and lighting. This compares favourably with other state of the art systems such as the Turk & Pentland scheme. RBF techniques should be well suited to the face recognition task and may find second-order (relative distance) differences that can generalise well rather than first-order (absolute distance) information.

Cognitive studies of the way human faces are perceived (for example (Bruce 1988)) can contribute to the design of systems that automate this kind of visual processing. There is support for having 'face recognition units' (FRUs) for recognising familiar faces (Bruce & Young 1986, Bruce 1988, Bruce et al. 1995). This idea is partly captured by the standard RBF techniques described next where the first layer of the network maps the inputs with a hidden unit devoted to each view of the face to be classified. The second layer is then trained to combine the views so that a single output unit corresponds to the individual person. We have taken this idea further and have developed a 'face unit' network model, which allows rapid network training and classification of examples of views of the person to be recognised. These face units give high performance and also alleviate the problem of adding new data to an existing trained network. We are use the various views of the person to be recognised together with selected confusable views of other people as the negative evidence for the network. Our face units have just 2 outputs corresponding to 'yes' or 'no' decisions for the individual. This is in contrast with Edelman et al. (1992) who did not use such negative evidence in their study. We show that this system organisation allows flexible scaling up which could be exploited in real-life applications.

The RBF Network Model

The RBF network is a two-layer, hybrid learning network (Moody & Darken 1988, Moody & Darken 1989), with a supervised layer from the hidden to the output units, and an unsupervised layer, from the input to the hidden units, where individual radial Gaussian functions for each hidden unit simulate the effect of overlapping and locally tuned receptive fields. They use the vector norm distance, $|\mathbf{i}-\mathbf{c}|$, equivalent to $\sum_{x=1}^{N}(i_x - c_x)^2$, between the N-dimensional input vector $\mathbf{i}$ and hidden unit centre $\mathbf{c}$ (N being the number of input units). The output value can be seen to approach a maximum when $\mathbf{i}$ becomes most similar to $\mathbf{c}$. The input vectors are unit-normalised.

Each hidden unit has an associated σ (sigma) 'width' value which defines the nature and scope of the unit's receptive field response[1]. 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.

The output o for hidden unit h (for a pattern l) can be expressed as:

$$o_h(l) = \exp[-\frac{|\mathbf{i}(l) - \mathbf{c}_h|^2}{2\sigma_h^2}], \tag{1}$$

the hidden layer output being unit-normalised, as suggested by (Hertz et al. 1991). For output unit i, the output is:

$$o_i(l) = \sum_h w_{ih} o_h(l). \tag{2}$$

Whilst the weights w_{ih} can be adjusted using the Widrow-Hoff (Widrow & Hoff 1960) delta learning rule, the single layer of linear output units permits a

[1] It is equivalent to the standard deviation of the width of the Gaussian response, so larger values allow more points to be included.

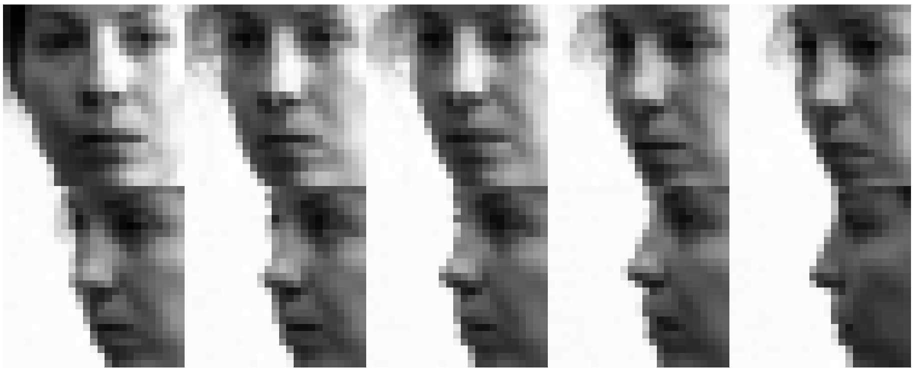

Figure 1: Entire 10-image range (rotating around the y-axis) for one person before preprocessing

matrix pseudo-inverse method (Poggio & Girosi 1990*a*) for their exact calculation. The latter approach allows almost instantaneous 'training' of the network, regardless of size[2]. The RBF network's success in approximating non-linear multidimensional functions is dependent on sufficient hidden units being used and the suitability of the centres' distribution over the input vector space (Chen et al. 1991).

'Face Unit' RBF Model

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 (1995*c*). 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.

[2] A network of 250 hidden units and 10 outputs, *ie.*2500 parameters, which required several hours of Sparc 20 processing time for gradient descent can be computed in a small fraction of a second.

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 (see Figure 1), 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 subsampled to a variety of resolutions for testing. Full details are given in Howell & Buxton (1995*a*). 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).

Pre-processing Methods

Although the RBF network was able to learn the dataset without preprocessing, *ie.*on pure grey-level values (Howell & Buxton 1995*b*), the authors see preprocessing of the images as a valid and important intermediate step, highlighting relevant parts of the information, and adding an essential invariance to illumination (Marr & Hildreth 1980).

Two main techniques are used for the preprocessing of the images: Difference of Gaussian (DoG) 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 cells 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. Our earlier studies (see Howell & Buxton (1995*b*)) showed that these later stages of processing make information more explicit for our face recognition task than the earlier DoG filters.

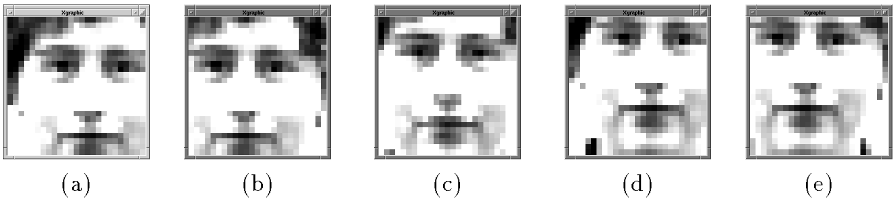

Figure 2: **Shift-varying** data for the 'face on' view of one individual: (a) top left (b) top right (c) normal view (d) bottom left (e) bottom right

The experiments presented here concentrate on two specific applications of these techniques:

- DoG convolution with a scale factor of 0.4, with a reduced range of grey-levels. The sampled values were thresholded to give zero-crossings information. A 25×25 image gave 21×21 convolved values, *ie.*441 samples per image.
- Gabor 'A3' sampling (for details, see Howell & Buxton (1995*b*)), with a full range of grey-levels. Data was sampled at four non-overlapping scales from 8×8 to 1×1 and three orientations (0°, 120°, 240°) with sine and cosine components. A 25×25 image gave 510 coefficients per image.

Generalization Over Views (y-axis Rotation) by the RBF Network

Fixed selections of images used for training to keep the experiments as constrained as possible. Table 1 shows both the standard and face unit RBF network models able to generalise very well over the different views with either the DoG or Gabor preprocessing method.

(a)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	88	28	100
Gabor	94	30	100

(b)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	92	35	95
Gabor	95	25	100

Table 1: Effect of pre-processing methods on **original** dataset: (a) Standard 50/50 RBF Network (b) 6+12 Face Unit RBF Network

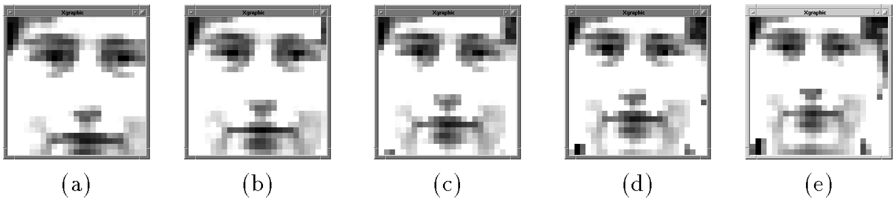

Figure 3: **Scale-varying** data for the ‘face on’ view of one individual: (a) +25% (uses 111×111 window) (b) +12.5% (107×107) (c) normal view (100×100) (b) –12.5% (94×94) (b) –25% (87×87)

Shift and Scale Invariance Properties of the RBF Network

Two further data sets were created to test the RBF network’s generalisation abilities:

- A shift-varying data set with five 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 ±10 pixels from the centre (see Figure 2).
- A scale-varying data set with five copies of each image: one at the standard sampling ‘window’ size, and four re-scaled at ±12.5% and ±25% of its surface area, ranging from 87×87 to 111×111 (see Figure 3).

Inherent Invariance - Training with Original Images Only

These experiments used only the original from each group of five for training, using all the varied ones (and the remainder of the original ones not used for training) for testing. This gives a measure of the intrinsic invariance of the network to shift and scale, *ie.*the invariance not developed during training by exposure to examples of how the data varies.

(a)

Network	Pre-processing	Initial %	% Discarded	% After Discard
100/400	DoG	14	84	21
100/400	Gabor	35	82	47
50/450	DoG	22	82	56
50/450	Gabor	37	77	53

(b)

Network	Pre-processing	Initial %	% Discarded	% After Discard
10+20	DoG	51	30	51
10+20	Gabor	57	38	52
6+12	DoG	54	32	53
6+12	Gabor	57	38	57

Table 2: Effect of pre-processing methods on **shift-varying** dataset (the original from each group of five used for training) (a) Standard RBF Networks (b) Face Unit RBF Networks

(a)

Network	Pre-processing	Initial %	% Discarded	% After Discard
100/400	DoG	58	63	78
100/400	Gabor	77	46	95
50/450	DoG	58	67	85
50/450	Gabor	75	52	94

(b)

Network	Pre-processing	Initial %	% Discarded	% After Discard
10+20	DoG	69	40	69
10+20	Gabor	83	36	88
6+12	DoG	69	44	66
6+12	Gabor	80	42	88

Table 3: Effect of pre-processing methods on **scale-varying** dataset (the original from each group of five used for training) (a) Standard RBF Networks (b) Face Unit RBF Networks

Some the increase in performance seen in Tables 2 and 3 going from 100/400 to 50/450 can be accounted for by the 50 original images which are not used for training being used for testing. Since these will already be allowed for with y-axis generalisation, they don't give a real increase in shift or scale invariance.

Learnt Invariance - Training with Shift and Scale Varying Images

These experiments again used a fixed selection of positions for training examples, using all five versions of each original image. This gives the network information about the shift and scale variance during training to help in learning this kind of invariance.

(a)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	72	46	94
Gabor	85	35	98

(b)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	84	32	93
Gabor	90	24	97

Table 4: Effect of pre-processing methods on **shift-varying** dataset (full groups of five used for training) (a) Standard 250/250 RBF Network (b) 30+60 Face Unit RBF Network

(a)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	83	34	98
Gabor	90	26	97

(b)

Pre-processing	Initial %	% Discarded	% After Discard
DoG	91	24	97
Gabor	93	20	98

Table 5: Effect of pre-processing methods on **scale-varying** dataset (full groups of five used for training) (a) Standard 250/250 RBF Network (b) 30+60 Face Unit RBF Network

Observations

Several points can seen from the results:

- The RBF network is shown to be able to generalise well in a non-trivial task classifying y-axis rotated faces (3-D complex shapes).
- Gabor preprocessing is shown to give a more generally useful input representation than the DoG preprocessing.
- Not suprisingly, the multi-scale Gabor preprocessing is shown to give greater scale invariance than the DoG preprocessing.
- The Gabor preprocessing is also shown not to fail catastrophically on the tougher shift invariance tests, unlike the DoG preprocessing.
- The RBF network is shown to have an inherent scale invariance on these tasks that does not need to be explicitly learnt from examples.
- In contrast, RBF networks do not have an inherent shift invariance, but this can be learnt from appropriate training data.
- The 'face unit' RBF network is shown to be superior to the standard network in terms of lower discard proportions for a particular level of generalisation performance.

Conclusion/Future Work

In summary, the locally-tuned linear Radial Basis Function (RBF) networks showed themselves to perform well in the face recognition task. This is a promising result for the RBF techniques considering the high degree of variability introduced by the varying views (y-axis rotation) of a person's face in these data sets. By centering our sampled faces on the nose of the profile views, we can regard the partial occlusion as simply missing features from the other side of the face. This is in accord with known results from Ahmad & Tresp (1993) who trained a variety of nets to recognise stationary hand gestures from computer-generated 2-D views (polar coordinates) of fingertips. They obtained good generalisation for 3-D orientation and showed that RBF nets were able to cope well even when much of the data was missing. Although their standard test data was handled well by a BP net, it performed badly with missing features and suffered a serious falling off in performance as more elements were lost. They showed, however, that a Gaussian RBF net (of the kind we used in our studies) could cope well, having a success rate of over 90% even with 50% of the features missing. This behaviour is very useful for coping with occlusion and other factors which lead to incomplete visual data.

We are now testing to see if the degree of view, scale and shift invariance that can be learnt by the RBF nets is sufficient to cope with data isolated from real-time video by a general purpose motion tracker. We are also studying invariance to facial expression and refining an automated 'face-finder' routine. This is necessary for the next stage of development in which people are to be identified in natural image sequences with the usual variations in illumination as well as position, scale, view and facial expression. The statistical nature of the information successfully captured by RBF nets to do the classification task may also be effective for the face localisation task. It is clear from the work of Turk & Pentland (1991) and Bishop (1995) and others using statistically based techniques that this is the key to good performance and the RBF techniques are mathematically well-founded, which gives a clear advantage in engineering a solution to our application problems. Future work will tackle the full unconstrained recognition task by tracking faces in real-time and gathering enough information to classify them accurately with good generalisation to other image sequences containing familiar people.

References

Ahmad, S. & Tresp, V. (1993), Some solutions to the missing feature problem in vision, *in* S. J. Hanson, J. D. Cowan & C. L. Giles, eds, 'Advances in Neural Information Processing Systems', Vol. 5, Morgan Kaufmann, pp. 393–400.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.

Bruce, V. (1988), *Recognising Faces*, Lawrence Erlbaum Associates.

Bruce, V. & Young, A. (1986), 'Understanding face recognition', *British Journal of Psychology* **77**, 305–327.

Bruce, V., Burton, A. M. & Hancock, P. J. (1995), Missing dimensions of facial distinctiveness, *in* T. Valentine, ed., 'Cognitive and Computational Aspects of Face Recognition: Explorations in face space', Routledge, pp. 138–158.

Chen, S., Cowan, C. F. N. & Grant, P. M. (1991), 'Orthogonal least squares learning algorithm for radial basis function networks', *IEEE Transactions on Neural Networks* **2**, 302–309.

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.

Edelman, S., Reisfeld, D. & Yeshurun, Y. (1992), Learning to recognize faces from examples, *in* '2nd European Conference on Computer Vision', Genoa, Italy, pp. 787–791.

Girosi, F. (1992), 'Some extensions of radial basis functions and their applications in artifical intelligence', *Computers Math. Applic.* **24**(12), 61–80.

Hertz, J. A., Krogh, A. & Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Addison-Wesley.

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*), Receptive field functions for face recognition, *in* 'Proc. 2nd Int. Workshop on Parallel Modelling of Neural Operators for Pattern Recognition (PAMONOP)', Faro, Portugal.

Howell, A. J. & Buxton, H. (1995*c*), A scaleable approach to face identification, *in* 'Proc. Int. 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.

Moody, J. & Darken, C. (1989), 'Fast learning in networks of locally-tuned processing units', *Neural Computation* **1**, 281–294.

Musavi, M. T., Ahmad, W., Chan, K. H., Faris, K. B. & Hummels, D. M. (1992), 'On the training of radial basis function classifiers', *Neural Networks* **5**, 595–603.

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.

Poggio, T. & Girosi, F. (1990*a*), Networks for approximation and learning, *in* 'Proceedings of the IEEE', Vol. 78, pp. 1481–1497.

Poggio, T. & Girosi, F. (1990*b*), 'Regularization algorithms for learning that are equivalent to multilayer networks', *Science* **247**, 978–982.

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.

Widrow, B. & Hoff, M. (1960), Adaptive switching circuits, *in* '1960 IRE WESCON Convention Record', Vol. 4, IRE, New York, pp. 96–104.