A Scaleable Approach to Face Identification

A. Jonathan Howell and Hilary Buxton

CSRP 376

June 1995

ISSN 1350-3162

UNIVERSITY OF

Cognitive Science
Research Papers

A Scaleable Approach to Face Identification

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

June 1995

Abstract

This paper describes a novel approach to solving two problems inherant in neural networks. The 'face unit' network system avoids the unmanagability of neural networks above a certain size by using small, individual networks for each class, and allows the addition of new data to the database without complete re-training of the system.

1 Introduction

Recognising objects and, in particular, the difficult subproblem of recognising human faces is the subject of a great deal of research in computer vision. However, it is only recently that work on biologically-motivated, statistical approaches to face recognition has begun to deliver real solutions.

2 he 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. Unlike a back-propagation network, for instance, this gives the RBF an activation that is related to the relative proximity of the test data to the training data, which gives a direct measure of confidence in the output of the network for a particular pattern. If the pattern is more than slightly different to those trained, very low (or no) output will occur.

3 he 'Face Unit' Concept

The concept of *face recognition units* was suggested in the perceptual frameworks for human face processing proposed by Hay & Young (1982) and Bruce & Young (1986). Each unit here produces a positive signal only for the particular person it is trained to recognise. For each individual, an RBF network is trained to discriminate between that person and others selected from the data set. Rather than using all the data available to train the network against an individual, the strategy adopted was to use only negative data that was most similar (using an Euclidean distance metric) to the positive data. Note that we assume similiarity leads to confusability, so the inclusion of this type of negative evidence in the training should improve descrimination. It was anticipated that this data was that with which the network would have the most 'trouble' when learning to discriminate 'for' and 'against' the individual, since it would be the most ambiguous. Unlike earlier tests which had only positive output signals (one per class), here two outputs are used for each 'face unit' network: 'yes' for the current class and 'no' for all other classes.

The reduction in the size of the network plus the use of negative knowledge, allows a more efficient coding of the information with greatly reduced training times. Furthermore, people can be added to the data set of a trained set of networks by the creation of a new 'face unit' network for each new individual to be added without retraining the original database, as the reorganised scheme is completely modular.

4 Method

4.1 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 used in ten different positions in 10° steps from face-on to profile of the left side, 90° in all.

The data set of ten faces (100 images in all) was gathered using a video camera and frame grabber, giving 8-bit grey-scale 384×287 images. 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 (1995).

4.2 Invariance Data Sets

Two additional data sets were created from the original data to test the RBF network's generalisation abilities: One data set to test scale-invariance was produced with five copies of each image: one at the standard sampling 'window' size, and four re-scaled at ±12.5% and ±25% of the standard surface area. The other data set, which tested offset-invariance, was produced also 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. The random selection of data from this set effectively doubles the variation in data, eg the scale of a test scale-invariance image could be up to ±50% that of a training image.

4.3 Types of 'Face Unit' Networks

For the training of 'face unit' networks, the term 'pro' is used to denote hidden units or evidence *for* the class, whilst 'anti' denotes that *against* the class. This evidence was selected according to Euclidean vector distance comparisons with images of the same pose angle of face with 'anti' evidence taken from the class that was the closest (most confusable) to the 'pro' class.

Two types of network layout were used: one where equal numbers of 'pro' and 'anti' hidden units were used, and one where two 'anti' were used for every 'pro'. The latter was used to show whether it would give better negative descrimination, which is important where there are large number of potential classes in large datasets. 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. To give an optimal spread of the image data for training, fixed selections of pose angle were used for each size of network. For instance, the 5+5 and 5+10 networks used poses 1, 3, 5, 7 and 9, where the pose range was 0–9.

Two strategies were investigated for the selection of 'anti' evidence: *Multiple* best negative networks used whichever 'anti' image was closest for each pose angle, so that several 'anti' person-classes could be used. *Single* best negative networks used an average of all vector distances over all pose angles to select one 'anti' person-class to represent all negative evidence. It was anticipated that the latter method would be superior, as a more coherant 3-D class boundary would be given by a single negative person-class for all pose angles. Fig. 1 shows how the images used for training were selected in an actual test for a 5+10

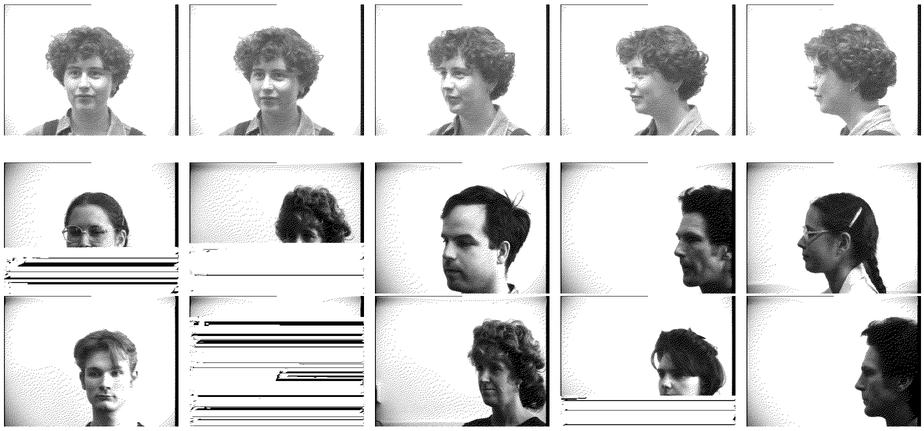

Figure 1: Example of 'pro' (top line) and 'anti' (middle and bottom lines) evidence used for a 5+10 'face unit' network

multiple best negative 'face unit' network. This shows how the same person is not necessarily used for all 'anti' views.

4.4 Adding 'Face Units'

To add new person-classes to the dataset, it would be necessary to save vector difference information after the initial selection of "anti" evidence. On the addition of a new person-class, vector differences would be calculated for the new class, saved and compared with the existing values. Any 'face unit' where the new class was closer than existing 'anti' evidence would need to be re-trained. All other 'face units' would not require further training. In the worse case, this would mean the entire system of 'face unit' networks being re-trained, but it anticipated that this would be unusual, especially as the number of classes became large.

4.5 Use of Confidence Measures

The statistical nature of the RBF network's output allows a 'confidence' measure based on the level of output. Initial tests used a 'winner take all' strategy, where input was classified according to the output node with the highest value. Subsequent examination of results showed that when the network correctly classified an image, the output values tended to be more disparate that when it incorrectly classified an image, with the correct output unit much larger than all others. The largest and second largest output values[1] are most different in correct classifications and least in incorrect classifications. This allows the use

[1]in the case of 'face unit' networks, there are only two output values, but this behaviour is also apparent with other RBF networks with larger numbers of outputs.

of a threshold based on the relationship of these two values to reject as 'uncertain' results below this threshold, leaving a smaller, but more accurate, set of classifications.

The initial approach taken was to use a threshold based on the ratio of the two output values, eg, if the two values were 0.2 and 0.5, the ratio between them would be 2.5. For comparison, further tests have been made using a threshold based on the absolute difference of the two outputs.

5 Results

In all these tests the network had a 100% success at classifying training images once trained, which is *not* included in the test results. These give performance values for the classification of test images only, which were all those images not used for training.

The 'Hidden Units' column indicates the number of hidden units in the network which is the number of 'pro' and 'anti' training images. '% Correct' is the average classification performance for all the face unit networks without any discarding strategy. 'Min. Pro' and 'Min. Anti' is the minimum performance found in all the face units, the maximum always being 100%. 'Max. % Correct' is the maximum average classification performance found using a discard stategy, with the ratio and percentage discarded in the 'Ratio' and '% Discard' columns. Tests where the threshold was so high that all of either the 'pro' or 'anti' results had been discarded for an individual face unit network were ignored.

5.1 Multiple Best Negative Classes, Ratio Threshold

Test 1: Equal 'pro' to 'anti' training

Hidden Units	Ave. % Correct	Min. 'Pro'	Min. 'Anti'	Max. Ave. % Correct	Min. 'Pro'	Min. 'Anti'	Ratio	% Discard
6+6	83	75	37	88	100	27	1.8	26
5+5	83	80	44	89	80	57	1.9	40
4+4	82	50	57	85	40	54	1.3	18
3+3	75	71	23	78	80	9	1.7	47
2+2	73	50	38	83	50	21	1.4	44
1+1	60	0	9	64	0	4	1.5	36

Test 2: 1 'pro' to 2 'anti' training

Hidden Units	Ave. % Correct	Min. 'Pro'	Min. 'Anti'	Max. Ave. % Correct	Min. 'Pro'	Min. 'Anti'	Ratio	% Discard
6+12	89	50	83	96	67	86	1.8	23
5+10	83	40	80	89	25	81	1.9	32
4+8	80	17	67	83	0	70	1.3	15
3+6	72	14*	70	78	0	76	1.7	38
2+4	66	0*	88	68	0	91	1.4	26
1+ 2	58	0*	91	57	0*	92	1.2	8

5.3 Multiple Best Negative Classes, Difference Threshold

Test 5:

5.5 Scale Variance Data

Test 9: Equal 'pro' to 'anti' training

Hidden Units	Ave. % Correct	Min/Max 'Pro'	Min/max 'Anti'
10+10	53	13/85	15/99
20+20	48	3/67	37/99
30+30	54	5/65	43/98

Test 10: 1 'pro' to 2 'anti' training

Hidden Units	Ave. % Correct	Min/Max 'Pro'	Min/max 'Anti'
10+20	51	3/33	70/99
20+40	48	0/37	59/99
30+60	54	0/55	61/98

5.6 Remarks

- The use of 'face unit'

6 Conclusion/Future Work

In summary, the RBF network 'face unit' organisation has proved to give a flexible, scaleable architecture which can perform at a high level in terms of both classification, generalisation over varying views, and speed of training. It is also a highly modular architecture that allows us to add more data and create as many new face units as are required. In particular, these studies showed that negative evidence plays a crucial role in shaping the discrimination between individuals and that this showed up particularly in the correct "no" responses of trained units. Multiple views of different people were more effective in improving performance than taking the same number of views of just one confusable person even though we might have expected a clearer decision boundary for the latter. It is also clear from these studies that the use of a confidence measure to discard some possible classifications is an effective strategy for improving the classification and generalisation performance on this dataset. This strategy would be most effective in studies of face recognition from image sequences we have planned for the near future. This extension of the work will exploit motion segmentation and look at a range of representations of the face data (Psarrou et al. 1995). We are interested in tracking the faces and gathering enough information to classify them accurately with good generalisation to other image sequences containing familiar people.

One disadvantage of our current scheme is the need to try all candidate face units during recognition of test data. This could be improved by parallel implementation or an indexing scheme to find the right face unit or set of face units in a hierarchical organisation of the units themselves. The work of Rao & Ballard (1995) is particularly interesting in this respect as they claim real-time indexing is possible using convolutions for distance computations to identify likely candidates. Another promising approach uses Gabor wavelet representations (Daugman 1988) which can be used for segmentation and tracking of faces using transforms of the data and may allow indexing in a similar way. Although such processing schemes are capable of multiscale face recognition and are robust to some changes in expression and orientation, we feel that a better strategy is to characterise the degrees of freedom in the input data required for the application. Systematic training can then be used to engineer a solution that copes with the dataset as required since typical 'mugshot' recognition, for example, is a very different task from active surveillance and recognition of moving, emotive people. What is required here is to explicitly address the need for invariance to scale, orientation, motion and expression in recognition performance or conversely characterise the need to estimate these measures if they are of interest for a particular application.

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.

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

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

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.

Hay, D. C. & Young, A. (1982), The human face, *in* H. D. Ellis, ed., 'Normality and Pathology in Cognitive Functions', Academic Press.

Howell, A. J. & Buxton, H. (1995), Invariance in radial basis function neural networks in human face classification, Technical Report CSRP 365, School of Cognitive and Computing Sciences, University of Sussex.

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.

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), 'Regularization algorithms for learning that are equivalent to multilayer networks', *Science* **247**, 978–982.

Psarrou, A., Buxton, H. & Gong, S. (1995), Recurrent nets for trajectory prediction and moving face recogntion, (preprint).

Rao, R. P. N. & Ballard, D. H. (1995), Natural basis functions and topographic memory for face recognition, (preprint).

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