Towards Unconstrained Face Recognition from Image Sequences

A. Jonathan Howell and Hilary Buxton

CSRP 430

August 1996

ISSN 1350–3162

UNIVERSITY OF

Cognitive Science Research Papers

Towards Unconstrained Face Recognition from Image Sequences *

A. Jonathan Howell Hilary Buxton
School of Cognitive and Computing Sciences,
University of Sussex, Falmer, Brighton BN1 9QH, UK
{jonh,hilaryb}@cogs.susx.ac.uk

Abstract

s p p r pr s nts p r nts us n B s s Fun t on BF n t or s to t t un onstr n r o n t on pro us n o r so ut on v o n or t on Input r pr

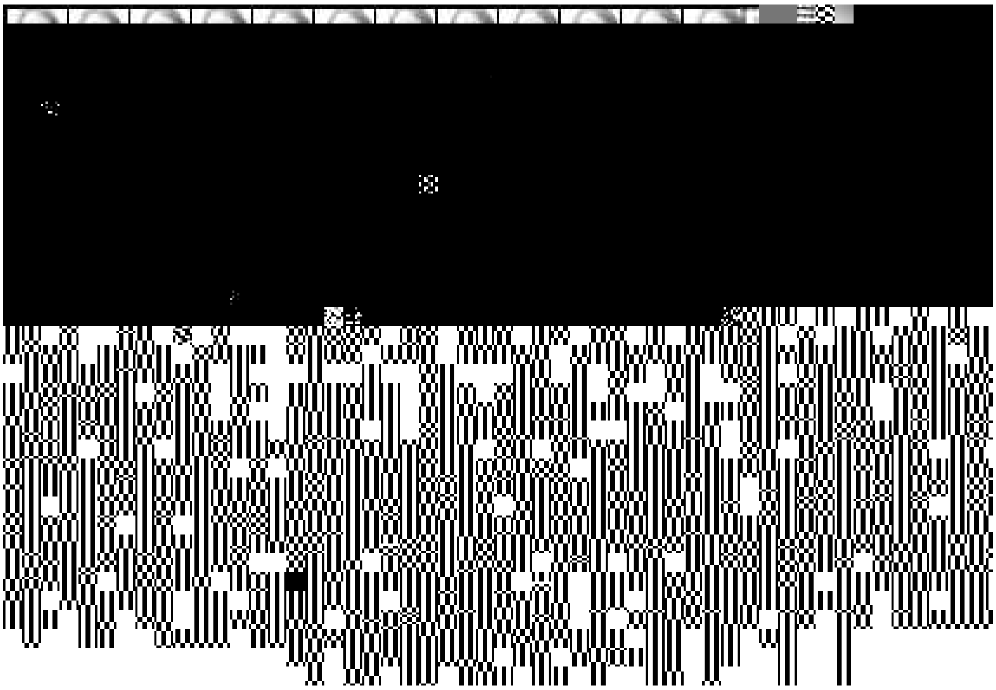


Figure . A complete secondary sequence for class *st v*, after segmentation but before pre-processing. As only front-view face detection has been implemented at this stage, some non-face frames are included and profile views, although segmented, are incorrectly scaled.

giving an activation that is related to the relative proximity of the test data to the training data. This allows a direct measure of confidence in the output of the network for a particular pattern, as very low (or no) output will occur if the pattern is more than slightly different to those trained, allowing the removal of outliers.

The weights can be adjusted using the Widrow-Hoff (Widrow & Hoff 1960) delta learning rule, however, the single layer of linear output units permits a matrix pseudo-inverse method (Poggio & Girosi 1990) for their exact calculation. The latter approach allows training of the network in a small fraction of a second. In the test phase, 500 images are processed in around two seconds, giving a single classification in around 4 ms, which is already adequate for real-time sequences.

4. 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 Westfield College (QMW), University of London, who are researching real-time face detection and tracking (McKenna & Gong 1996). We have devised two types of sequences to simulate a (fairly) unconstrained environment, termed 'Primary' and 'Secondary'.

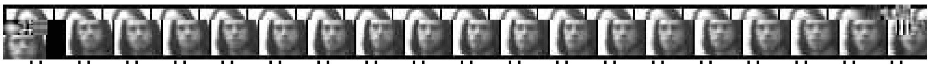
The intention is to train the network with a controlled set of data – the Primary image sequences – known to include the types of variability which we want our trained system to be tolarant to (thus including 180° range of pose angles but a blank background), and to test on totally unrelated data – the Secondary image sequences. This total separation of training and test data is to allay any fears that the system ... horizontal ... detail ... spurious ... lighting

(a)

Interval	Train/Test	Initial %	% Discarded	% After Discard
2	278/276	96	12	98
5	114/440	88	30	99
10	60/494	75	50	90
20	33/521	58	68	90
30	24/530	48	81	93
50	16/538	40	81	86

(b)

Interval	Train/Test	Initial %	% Discarded	% After Discard
2	278/276	99	2	100
5	114/440	98	7	100

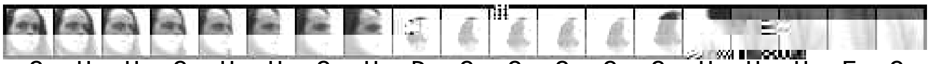


H H H H H H H H H H H H H H H H H H H

H H H H H H H H H H H H H H H H H H H

H H H H H H H H H H H H H H H H G G G

H . H H H H H H H . H H H . . . . G .



G H H G H H G H D G G G G G H H H F G

G . . . H . . . . G . . G . H H . . G

H H G G G G D G G G G G H H E H H H H

. H . . . . H . H . H H

8. Performance with More Classes

It must be emphasised that this research is at a preliminary stage but that the technique shows promise for scaling up to large databases. Although only a few individuals are shown in our image sequences, this type of network has been shown to work well with larger numbers of classes. For example, the Olivetti Research Laboratory database of faces[1] with 400 images of 40 people can be distinguished with a high level of performance (see Table 4).

Images per Person	Train/Test	Initial %	% Discarded	% After Discard
5	200/200	84	39	95
4	160/240	80	43	95
3	120/280	72	52	91
2	80/320	64	60	87
1	40/360	46	70	81

Table 4: Results for ORL Face Database, using Gabor preprocessing, averaged over two different selections.

9. Conclusion/Future Work

Several points can seen from the results:

1. The RBF network is shown to generalise well from samples in classifying faces (3-D complex shapes) in real-time sequences.
2. Gabor preprocessing is shown to give a more generally useful input representation than the DoG preprocessing, especially for the more difficult Secondary sequence.
3. The confidence measure used in discarding uncertain classifications is shown to be important for handling sequences especially where a small training set is used.

In conclusion, the locally-tuned linear Radial Basis Function (RBF) networks showed excellent performance in the simpler face recognition task when trained and tested on images from Primary sequences. This is a promising result for the RBF techniques considering the high degree of variability due to the varying views (mostly rotations) of a person's face in these data sets. The results so far from the Secondary sequences also show considerable promise, especially with the additional use of temporal coherence to improve performance. In the these sequences, the face detection scheme (McKenna & Gong 1996) currently selects and rescales faces in near face-on views but does not discard the others. It is expected that further development of this scheme will allow improvements in the reliable and consistent labelling of faces in unconstrained image sequences. It is clear that the ability of the RBF networks to give a measure of confidence, 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 refining the face detection scheme and automated on-line learning of new classes of individual. The next stage of development will integrate this refined 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 (1995) and others that using statistically based techniques is the key to good performance. The RBF techniques are mathematically well-founded, which gives a clear advantage in engineering a solution to our application problems.

[1] available via *tp* for comparative tests, further information is at: `http://www.cam-orl.co.uk/facedatabase.html`

References

Beymer, D. & Poggio, T. (1995), Face recognition from one example view, *n* 'Proceedings of International Conference on Computer Vision', Cambridge, MA, pp. 500–507.

Bishop, C. M. (1995), *ur t or s or tt rn o n t on*, Oxford University Press.

Brunelli, R. & Poggio, T. (1993), 'Face recognition: Features versus templates', *IEEE r ns t ons on tt rn An s s n n Int n* **15**, 1042–1052.

Craw, I., Costen, N., Kato, T., Robertson, G. & Akamatsu, S. (1995), Automatic face recognition: combining configuration and texture, *n* 'Proceeding of International Workshop on Face and Gesture Recognition', Zurich, Switzerland, pp. 53–58.

Daugman, J. G. (1988), 'Complete discrete 2-D gabor transforms by neural networks for image analysis and compression', *IEEE r ns t ons on A oust s p n n ro ss n* **36**(7), 1169–1179.

Girosi, F. (1992), 'Some extensions of radial basis functions and their applications in artifical intelligence', *Co put rs t t s t App t ons* **24**(12), 61–80.

Howell, A. J. & Buxton, H. (1995), 'Invariance in radial basis function neural networks in human face classification', *ur ro ss n tt rs* **2**(3), 26–30.

Howell, A. J. & Buxton, H. (1995), Receptive field functions for face recognition, *n* 'Proceeding of 2nd International Workshop on Parallel Modelling of Neural Operators for Pattern Recognition', Faro, Portugal.

Howell, A. J. & Buxton, H. (1995), A scaleable approach to face identification, *n* 'Proceedings of International Conference on Artificial Neural Networks', Vol. 2, EC2 & Cie, Paris, Fran