Face Recognition from Unfamiliar Views: Subspace Methods and Pose Dependency

Daniel B Graham and Nigel M Allinson
University of Manchester Institute of Science and Technology
Department of Electrical Engineering and Electronics
Image Engineering and Neural Computation Laboratory

Abstract

A framework for recognising human faces from unfamiliar views is described and a simple implementation of this framework evaluated. The interaction between training view and testing view is shown to compare with observations in human face recognition experiments. The ability of the system to learn from several training views, as available in video footage, is shown to improve the overall performance of the system as is the use of multiple testing images.

1 Introduction

Recognising faces from previously unseen viewpoints is inherently more difficult than matching faces at the same view. Simple image comparisons such as correlation demonstrate that there is a greater difference between different viewpoints of the same subject than between different subjects at the same view which means that the recognition method used must take into account the non-linear variations of faces with viewpoint. In order to achieve recognition of previously unseen views, we require a method of relating the information available from the previously seen viewpoint to the information in the test (novel) viewpoint.

Previously this problem has been treated as an *image synthesis* problem whereby novel views (images) of the face are generated from one previously stored view using a variety of methods such as the optical flow method of Beymer and Poggio [1], the linear object classes of Vetter and Poggio [16] and the 3D structure estimation of Nagashima [9]. The resulting image is then matched to the stored images using a suitable image comparison technique.

Pentland et al [11] used a *view-based subspace* technique by producing separate subspaces each constructed from faces at the same viewpoint. Recognition in their work is performed by first finding the subspace most representative of the test face and then matching using a simple distance metric in this subspace. The subspace method employed here was a standard eigenspace decomposition as described in Turk and Pentland [13].

Valentin and Abdi [14] employ an *analytical subspace* method which determines whether a face has been seen previously or not (but not who) using the reconstruction quality of a novel image as a threshold. Simply, a face is *known* if it can be reconstructed (at an unfamiliar viewpoint) to within a certain accuracy, and *unknown* if the reconstruction quality is insufficient. This technique shows that the subspace here (namely a completely trained linear associator) can determine whether a novel face was used to construct the subspace or not, even when it was present only in different viewpoints. Valentin and Abdi also show that the number of previously seen views of the face and which particular viewpoints were previously seen significantly affects performance.

Finally McKenna et al [5] use a *characteristic subspace* method which describes an individual using several points in a subspace (e.g. from video footage). Here previously unseen views tend to cluster around the stored views of one particular individual and probabilistic analysis leads to recognition.

The method presented in this paper is one of a *predictive characterised subspace* - whereby the characteristic of the individual through the subspace (as in [5]) is estimated from *one or more* previously seen views. Identification is then a matter of matching an unseen view to a point on the characteristic. Obviously the performance of such a system is dependent upon the initial point (training view) chosen to characterise the individual, the method of characterisation and upon the distance between the training view and the novel view. These interactions between training and testing views will be examined , as will the relative performance of differing initial training views.

Figure 1. Pose Varying Images

2 Pose Varying Eigenspace

The use of eigenspace methods for facial image analysis has been common since early papers by Sirovich and Kirby [12] and the more often cited Turk and Pentland [13]. The majority of such systems have shown that separating the shape information (e.g. by morphing) from the texture information yields additional performance enhancements as in Costen et al [3]. The view-based eigenspaces of Moghaddam and Pentland [6] have also shown that separate eigenspaces perform better than using a combined eigenspace of the pose-varying images. This approach is essentially several discrete systems (multiple-observers) and so highly dependent upon the number of views chosen to sample the viewing sphere and of the accuracy of the alignment of the views. Producing an eigenspace from all of the different views (a pose varying eigenspace), could continuously describe an individual through an eigenspace in the form of a convex curve. This has been shown in [5] and in Murase and Nayar [8] for 3D objects. The continuous nature of the eigenspace allows us to match not just novel points in the eigenspace to a curve but to match continuous and ordered line segments (or clusters) to segments of the curve. A simple use of this technique has been presented in Graham and Allinson [4] to increase recognition rates for single image matching by estimating additional points in the eigenspace from single test images.

Our pose varying eigenspaces are constructed from images like those shown in Figure 1. The characteristic curves of these two individuals are shown in Figure 2. The images are all manually segmented and aligned (pose estimated). Note that these curves are represented here by ten 3D points corresponding to the first three eigenvalues of the images (EV1, EV2 & EV3) in an eigenspace constructed from 40 images randomly sampled from a database of 563 pose varying images of 20 people.

We have called these loops in the eigenspace *eigensignatures* as each one corresponds uniquely to a specific individual.

3 Recognition from Unfamiliar Views

Consider the pose varying eigenspace described in Section 2, where a unified pose/identity subspace is generated which captures the manner in which faces change over varying pose and quantifies the extent of that change in terms of distances in the subspace. It can be seen in Figure 2, that individuals all differ in this subspace but that each subject undergoes a characteristic motion through the subspace. As the motion we are capturing is the same in each case and the 3D structures of each subject are closely related, it is not unreasonable to assume that the general nature of these characteristic curves can be obtained, and that a curve may be estimated from a single given point.

Formally, the recognition of faces in previously unseen views requires a function Γ which maps a real point p to a virtual eigensignature Υ. This virtual eigensignature has a confidence factor $\delta(p)$ which depends upon the initial point p. Formally:

$$\Gamma(p) = \Upsilon, \delta(p) \tag{1}$$

Given further real points p_i we can generate further virtual eigensignatures Υ_i each with their own confidence factor $\delta(p_i)$. We can then combine virtual eigensignatures to produce a *refinement* of the virtual eigensignature which approaches the true eigensignature Ω. Given that the confidence factors $\delta(p_i)$ lie in the range $\{0,1\}$ we can define the weight function ω_i for each virtual eigensignature:

$$\omega_i = \frac{\delta(p_i)}{\sqrt{\sum_{i=0}^{N}(\delta(p_i))^2}} \tag{2}$$

We can combine the eigensignatures to produce:

$$\Omega \cong \sum_{i=0}^{N} \omega_i \Upsilon_i \tag{3}$$

Note that this framework is independent of the chosen representation of the eigensignatures, the confidence factors and the weight function. Additionally, the weight deduction (eqn 2) is sub-optimal, in that real points in the

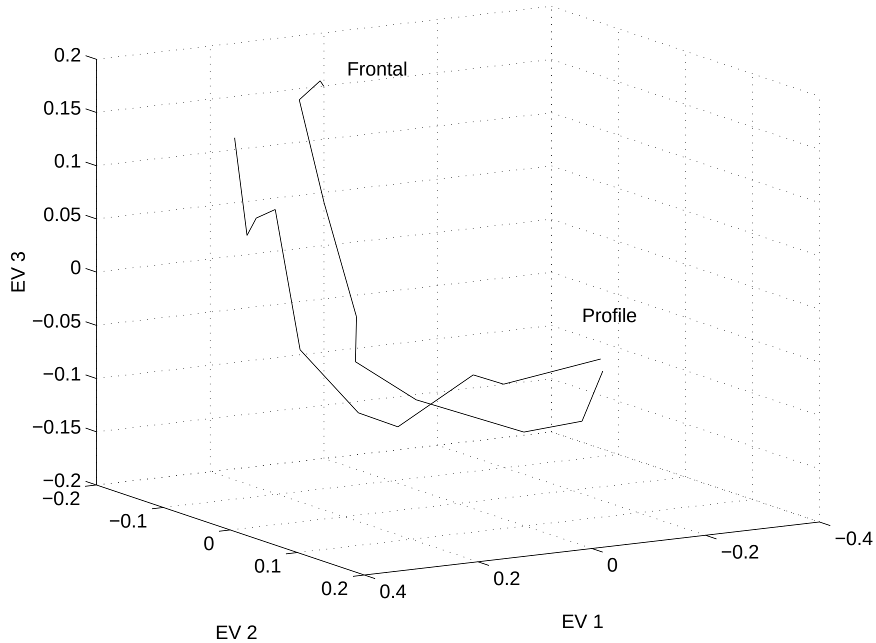

Figure 2. Eigensignatures of two people.

eigenspace ($\delta(p_i) = 1.0$) should remain in the eigensignature and not be influenced by other points. The development of an algorithm for effectively combining multiple eigensignatures will be described in a later paper.

In order to investigate the above formulation we define an eigensignature as consisting of ten points in the eigenspace sampled from profile to frontal view in $\sim 10^o$ degree steps. Virtual eigensignatures Υ_i are generated from a test point p_i using a Radial Basis Function Network (RBFN) - see Moody and Darken [7] as the mapping function Γ. The RBFN was trained on one view (p_i) to produce the full eigensignature Υ_i. The output from the RBFN thus gives ten points in the eigenspace for an individual which estimates the characteristic curve of that face in the eigenspace. Each RBFN is trained on 19 of the subjects' true eigensignatures and the remaining subjects' eigensignature was generated from the RBFN to form a virtual eigensignature - this was repeated for each of the 20 subjects (leave-one-out cross-validation) to produce 20 virtual eigensignatures. To investigate the pose dependent nature of the method an RBFN was trained using each of the ten views (producing ten virtual eigensignatures per person) and the performance of each of these eigensignatures was compared. In total 200 virtual eigensignatures were produced.

Recognition is performed by matching a test image (i.e a test point in the eigenspace) to one of the virtual eigensignatures using a nearest neighbour Euclidean distance. Different metrics in this eigenspace, such as the Mahalanobis distance, have also been investigated and found to perform similarly. It should be noted that simple point matching in this formulation describes the base-level performance achievable. Matching with multiple, ordered, points - whether *real* or *virtual* as in [4] - should improve the performance of such systems.

4 Experimental Results

4.1 Train/Test View Interaction

The performance of this approach is dependent on several factors as described in section 2. Here we establish the baseline performance of the system by matching the real eigensignatures (omitted during the RBFN training) with the virtual eigensignatures generated by the RBFN. Table one shows the percentage of correct identifications at each train/test view. It can be seen from the **Average** row that there is a clear advantage to testing at the 40^o to 50^o view. This is normally referred to as the *3/4 view* and is often reported the best performing pose in human face recognition experiments such as Bruce et al [2], Valentin et al [15] and partially in Patterson and Baddeley [10].

A similar result would be observed for an average over

		Test View									
		0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**
	0	100	90	60	40	20	20	20	15	15	15
	10	90	100	70	60	50	45	50	40	15	15
	20	35	65	100	70	30	20	30	20	20	25
Training View	**30**	35	75	85	100	70	60	25	25	20	20
	40	35	25	40	70	100	80	55	50	25	30
	50	25	30	40	65	95	100	60	55	30	30
	60	15	15	30	30	60	85	100	80	55	45
	70	10	10	25	30	45	55	90	100	75	50
	80	15	20	35	20	20	30	45	70	100	65
	90	20	10	10	20	20	15	25	30	60	100
	Average	38	44	49.5	50.5	**51**	**51**	50	48.5	41.5	39.5

Table 1. Train/Test View Interaction (Pose 0 = Profile, 90 = Frontal)

testing view to determine the relative performances of each training view but it was felt that such an interpretation would be biased in favour of the central views by the window effects of the data around the end views of 0^o and 90^o. As such it is difficult to determine the optimal training view. However, were we to assume that *all tests* were to be carried out in this pose range, we would have reason to propose the 3/4 view as the preferred training view.

It should be noted that these results do not represent a real-world situation - they merely establish the performance characteristic of a single virtual eigensignature which, as in human unfamiliar-view face recognition experiments deteriorates rapidly as the test viewpoint moves away from the training viewpoint. The following experiment describes a more useful approach whereby the information contained in multiple training images is combined to form a more accurate representation of an individual.

4.2 Multiple Training Images

Recognition of face when having only seen one previous image of that face is classed as *unfamiliar face recognition.* As the number of images increases the process tends towards *familiar face recognition.* The system presented in Section 3 provides a general purpose formulation for these two types of recognition. Section 4.1 has shown the baseline performance for unfamiliar face recognition and examined the pose dependent nature of the system. Here we examine the effect of increasing the number of training images used to form the refined eigensignature according to eqns 2 & 3. In a simple experiment we show the effect of increasing N_v (the number of virtual eigensignatures) and the pose dependency of this increase. For this evaluation we have used a confidence factor, centered around the p_i, which decays sharply with distance from p_i. Namely:

$$\delta(p_i, p_j) = \frac{1}{1 + \parallel p_i - p_j \parallel} \tag{4}$$

where p_i is the pose used to train the RBFN and p_j is the test pose.

Table 2 shows the performance of this system as N_v increases, and how this performance varies over pose. The results shown are the percentage of correct identifications at each pose for every possible combination of N_v virtual eigensignatures from ten. As in Section 4.1 it can be seen clearly that, on average, the 50^o test view outperforms all other views. There is also a clear trend of performance increasing with N_v. Furthermore we see a preference for testing at frontal views over profile views - another common observation in human face recognition experiments [2]. These differences are more pronounced for unfamiliar faces (low N_v) than for familiar faces (high N_v) - also noted in [2].

These results show the maximum performance increase obtainable with multiple training views as the multiple views are all pose-aligned on the test views. However, we would expect similar improvements in local test areas for non-aligned images due to the nature of the eigensignature combination (eqns 2 & 3) and the confidence factor (eqn 4).

4.3 Multiple Testing Images

The experiments described in Sections 4.1 & 4.2 have demonstrated the use of virtual eigensignatures for recognition, including the case where multiple training images are available. Conversely, in real world systems, the number of training images may be low and fixed whereas the number of test images may be large and variable (e.g. video monitoring). We show here the simple situation where multiple training images are used to produce a total Euclidean distance from which we again attempt recognition. There

		Test View p_j										
		0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**Average**
	1	27.0	32.0	40.5	45.0	45.5	54.5	49.5	48.0	43.5	36.0	42.1
	2	36.9	43.7	53.0	57.8	58.2	69.1	66.0	64.3	61.2	49.6	56.0
Number	**3**	44.5	52.9	61.5	65.4	68.3	78.3	75.8	74.8	73.9	59.4	65.5
of	**4**	50.3	59.7	67.2	69.2	76.5	84.7	82.2	81.2	82.8	68.3	72.2
Signatures	**5**	56.1	65.5	72.0	72.3	83.1	89.3	86.6	85.7	87.7	76.0	77.4
Combined	**6**	62.1	71.0	77.2	74.8	88.3	92.4	90.1	89.4	90.8	82.2	81.8
(N_v)	**7**	68.4	77.2	82.8	78.3	92.7	95.1	92.8	92.4	93.0	87.6	86.0
	8	75.9	84.4	87.9	84.2	95.7	97.3	94.3	94.6	94.0	92.0	90.0
	9	81.0	89.5	93.0	91.5	98.0	98.5	94.5	95.5	93.5	96.0	**93.1**
	Average	55.8	64.0	70.6	70.9	78.5	**84.4**	81.3	80.6	80.0	71.9	

Table 2. Refined Eigensignature Performance (%)

were 363 test images of the same twenty people in the database, none of which were used at any stage during the RBFN training. These images were taken at the same time and place as the previous images but were considered to lie in views intermediate to the 10^o views used in the previous experiments.

Table 3 shows the recognition improvements gained by using increasing numbers of test images for each of the virtual eigensignatures. The figures shown indicate the percentage of correct identifications for all possible combinations of N_t test images of the same subject. As would be expected, there is a clear improvement in using additional test images. For comparison, we show the performance of the system when the test images are measured against the true eigensignature (**True**). As can be seen there is little change in the performance of the system using the true eigensignature, as there is at the frontal and profile areas of the training views for increasing numbers of images. However we see a marked improvement at the 50^o training view of some 17% (compared with 2-3%), providing further evidence for the preference of this view as the best training view to use, but with the same reservations as in Section 4.1.

5 Conclusion

A novel framework for describing individuals at unfamiliar views has been described which uses a Radial Basis Function Network to characterise a subjects' pose-varying behaviour in a suitable eigenspace. A simple implementation has shown good comparisons with some reported results for human face recognition in the interaction between training and testing view, the performance differences between familiar and unfamiliar face recognition, and in the preference for the 3/4 view.

The proposed framework provides the basis for an automatic system of developing individual characteristics from video footage. The ability to use multiple training views in the characterisation stage provides a flexible means of identification. Similarly the use of multiple images in the testing stage can be used to develop a further characteristic to aid the recognition process. A combination of both of these approaches should provide a powerful means of recognition from video.

References

[1] D. Beymer and T. Poggio. "Face Recognition From One Example View." *Proceedings International Conference on Computer Vision,* Boston, MA, pp. 500-507, June, 1995.

[2] V. Bruce, T. Valentine and A. Baddeley. "The Basis of the 3/4 View Advantage in Face Recognition." *Applied Cognitive Psychology,* Vol 1, pp 109-120, 1987.

[3] N.P. Costen, I. Craw, G. Robertson, and S. Akamatsu. "Automatic face recognition: What representation?" *Computer Vision*, Bernard Buxton and Roberto Cipolla (eds), ECCV'96, vol 1064, Lecture Notes in Computing Science, pp 504-513. Springer-Verlag, 1996.

[4] D.B. Graham and N.M. Allinson. "Face Recognition Using Virtual Parametric Eigenspace Signatures." *Proceedings IEE Conference on Image Processing and its Applications.* Dublin, Ireland, pp 106-111, July 1997.

[5] S. McKenna, S. Gong and J.J. Collins. "Face Tracking and Pose Representation." *British Machine Vision Conference,* Edinburgh, Scotland, September 1996.

[6] B. Moghaddam and A. Pentland. "Face Recognition using view-based and modular eigenspaces." *SPIE* Vol 2277, pp 12-21, 1994.

	Training View p_i											
N_t	**True**	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**Average**
1	85.86	32.75	48.17	41.50	51.57	48.97	48.82	46.14	44.44	37.81	27.02	42.72
2	85.91	37.67	55.40	51.36	58.22	58.90	58.09	46.67	49.64	43.64	29.24	48.88
3	89.24	35.33	56.31	52.84	58.87	60.89	62.09	52.51	54.29	44.56	27.91	50.56
4	89.11	36.36	56.69	53.49	60.15	62.45	63.66	53.77	56.17	46.66	28.38	51.78
5	87.26	35.88	56.55	54.08	59.65	62.33	**65.47**	54.25	59.10	47.37	27.67	**52.24**
Average	87.48	35.60	54.63	50.65	57.69	58.71	**59.62**	50.67	52.73	44.01	28.04	

Table 3. Use of Multiple Test Images

[7] J. Moody and C.J. Darken. "Fast Learning in Networks of Locally-Tuned Processing Units." *Neural Computation,* vol 1, pp 281-294, 1989.

[8] H. Murase and S.K. Nayar. "Learning Object Models from Appearance." *Proceedings of the AAAI,* pp 836-843, Washington DC, July 1993.

[9] Y. Nagashima, H. Kawamura, M. Kosugi and N. Sonehara. "3D Face model reproduction using multi view images." *Proceedings of the SPIE, Visual Communications and Image Processing '91,* Vol 1606, pp 566-573, 1991.

[10] K.E. Patterson and A.D. Baddeley. "When Face Recognition Fails." *Journal of Experimental Psychology: Learning Memory and Cognition,* Vol 3(4), pp 406-417, 1977.

[11] A. Pentland, B. Moghaddam and T. Starner. "View-Based and Modular Eigenspaces for Face Recognition." *IEEE Conference on Computer Vision and Pattern Recognition,* pp 84-91, June 1994.

[12] L. Sirovich and M. Kirby. "Low Dimensional procedure for the characterisation of human faces." *Journal of the Optical Society of America,* Vol 4(3), pp 519-525, March 1987.

[13] M. Turk and A. Pentland. "Eigenfaces for Recognition." *Journal of Cognitive Neuroscience,* Vol 3(1), pp 71-86, 1991.

[14] D. Valentin and H. Abdi. "Can a Linear Autoassociator Recognize Faces From New Orientations." *Journal of the Optical Society of America A - Optics, Image Science and Vision,* Vol 13(4), pp 717-724, 1996.

[15] D. Valentin, H. Abdi and B. Edelman. "What Represents a Face: A Computational Approach for the Integration of Physiological and Psychological Data." *Perception (in press),* Volume 26, 1997.

[16] T. Vetter and T. Poggio. "Linear Object Classes and Image Synthesis from a Single Example Image." *Technical Report No. 16,* Max-Planck-Institut für biologische Kybernetik, April 1995.