

Norm²-based face recognition

D. B. Graham and N. M. Allinson

Department of Electrical Engineering and Electronics
University of Manchester Institute of Science and Technology
Manchester, M60 1QD, UK
danny@sound.ee.umist.ac.uk

Abstract

*Increasingly the problems of recognising faces under a variety of viewing conditions, including depth rotations, is being considered in the field. The concept of **norm-based** coding in face recognition is not new but has been little investigated in machine models. Here we describe a norm-based face recognition system which is capable of generalising from a single training view to recognise novel views of target faces. The system is based upon the characteristic nature of faces as they move through a pose-varying eigenspace of facial images and deviations from the norm of a gallery of face images. We illustrate the use of the technique for a large range of pose variation.*

1 Introduction

Recognising human face over larger pose changes has, until recently, been considered a very difficult task in automated face recognition and one often avoided. This is essentially because different faces in similar poses appear more alike than the same face at very differing views. Several researchers have shown that the identity information in human faces is obtainable and transferable to novel views. Edwards et al [2] have demonstrated that a linear discriminant analysis in an eigenspace of face shapes/textures can maintain identity information over a range of poses. Duvdevani-Bar et al [1] have shown that a neural network classifier can generalise to novel views of unseen subjects by similarity to a *chorus of prototypes* - essentially several measures of similarity to prototype (training) faces. Similarly, Graham and Allinson [3] have shown that the trajectory of a face undergoing pose changes through an eigenspace of pose-varying images (which they refer to as an *eigensignature*) is predictable and may be used to recognise faces at novel views. Finally

the Gabor-wavelet work of the Bochum/USC group [7] has shown that by using prior knowledge of faces undergoing pose changes the linear transformation of local feature vectors from one pose to another is obtainable and hence may be utilised for the recognition of novel views (as long as the linear transformation is known for the pose of the novel view).

A common feature of all of these systems is that they rely on a similarity measure to recognise a face - i.e. a face is recognised by measuring the distances between it and all members of the target gallery of faces using some system-dependent metric. An alternative to such similarity functions is suggested by several observations from face recognition by humans - namely that of *norm-based* face-recognition. In a norm-based system a face would be characterised (and hence recognised) by its distance from the norm of the gallery of faces. Rhodes and Tremewan [8] have shown that the effect of producing *caricatures* (i.e. enhancing deviations from the population norm) can increase recognition rates; that negative caricature (i.e. reducing the deviation from the norm) decreases recognition rates, but that caricatures orthogonal to the deviation from the norm dramatically reduce recognition rates. This behaviour would suggest that the human face recognition system is influenced by deviations from the norm and is dependent upon the direction of such deviation. The remainder of this paper will describe a norm-based face recognition system which can be used to recognise faces over a large range of poses and viewing conditions. The basis of the system relies on obtaining a characteristic norm vector for the face population and using it to represent a gallery of faces, i.e. each face is represented by its position relative to the norm of the population. In order to determine the effects of direction from the norm, and not merely distance, we will examine the gallery of faces in terms of *their norms* with the population norm.

2 Norm based representation

In order to recognise images of faces we require a representation method for each individual to be recognised. In the majority of face recognition systems the representation for an individual must be formed from a single training image. Recognition in this scenario then consists of matching the test image (or probe) to the set of stored representations (the gallery). In almost all recognition systems the viewing condition of the test probe is known or can easily be determined. In many cases there exists only one test condition - for example a single (and consistent) change in expression or pose. Clearly this represents a major limitation for a face recognition system. The problem of pose-varying face recognition may be phrased in terms of a set of discrete systems specific to a single pose (multiple observers) - yet such systems will fail for a pose not included in the range covered by the set of single views. Confusingly, there exists a useful and near-linear region around the frontal pose ($\pm 20^\circ$) for which most models seem to work. This restricted pose variation is often considered sufficient to illustrate pose invariance for recognition. Such limitations apply to the studies of Edwards et al [2] and Wiskott et al [9]. In human face recognition small variations are rarely considered - the *transfer* of information from one viewing condition to another is usually only examined for full, three-quarter and profile views. Attempts at such large pose-varying recognition have been given by Duvdevani-Bar et al [1], Graham and Allinson [3] and for a single pose variation by Okada et al [7]. The major drawback of these systems is that they require a discrete set of poses and, with the exception of [1] & [3], are unable to interpolate between known poses. The system presented here considers that an individual is defined by the vector from the norm of the face population - and that this vector remains (relatively) constant over all poses. So it becomes feasible to recognise test images by computing the norm vector of the probe to the population norm and then matching this vector to the gallery of similarly formed norms. This idea is based upon the concept that an individual's face exhibits a distinctive characteristic through a pose-varying eigenspace - as noted in McKenna et al [4] for pose-determination and as used for recognition in [3]. Unlike the related works [1] and [3] where identity information is extracted using neural network models, what we attempt here is a specific extraction of identity information from the eigenspace. By showing that this information is obtainable directly we illustrate further understanding

of the identity information obtainable from facial images over changing pose.

2.1 Forming the norm curve

Using principal components analysis of a set of pose-varying images provides an eigenspace suitable for the representation of faces over the range of poses used to form the eigenspace. In such eigenspaces it has been shown by Nayar et al [6] that objects exhibit continuous trajectories. With a set of pose-varying faces we can form an eigenspace and examine each individual in this eigenspace. In Figure 1 we see the distribution of 20 individuals in a total of 566 images ranging from frontal to profile views for the first three eigenvectors of the principal components analysis.

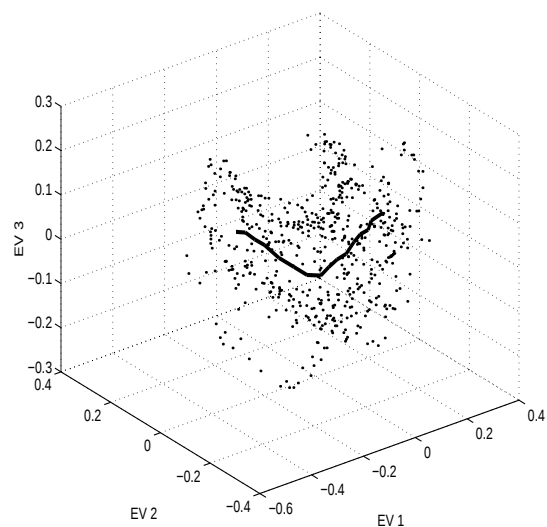


Figure 1: The population and norm in the eigenspace.

Given these images it is relatively straightforward to form an average of each person over the range of poses in the database (frontal to profile here). Taking each person in turn we can select several points along their trajectory through the eigenspace to form an estimate of the norm curve on which to base our representation of faces. This is the curve shown through the points in Figure 1. It is the discrete pose sampled representation of the norm that we use for recognition. Linear interpolation can be used to determine any point on this norm curve. Figure 2 shows these discrete points when reconstructed in the image domain using 100 eigenvectors (from 500). It can be seen that this norm provides a reasonable foundation for representing faces over the pose range considered here.

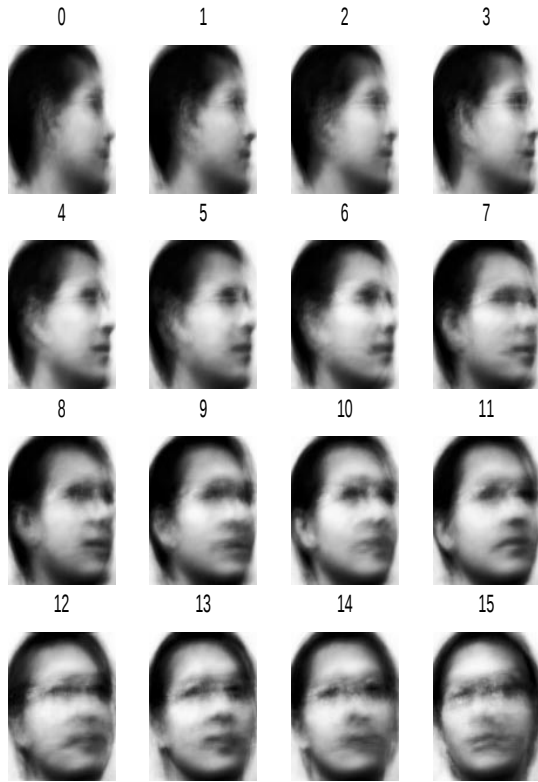


Figure 2: The population norm in image space.

2.2 Forming representations

Given this discrete norm curve we then require a method of implementing the gallery and the ability to generalise from this gallery when presented with a test probe of a novel view of a known individual. Note that what is proposed here essentially means that we consider every individual to consist of a manipulation of the norm curve and that this behaviour is simple to obtain. Several strategies will be considered for this process in the following section. Essentially, we can say that, given a gallery image (training image) it is sufficient to store the *norm vector* ($\mathbf{g}_i$) which defines the nearest point normal to the *norm curve* (hence the “Norm²” title) and the anchor point on the norm curve which defines that vector ($\mathbf{n}_i$). This representation is illustrated in Figure 3. It can be seen that it is now feasible to attempt recognition of faces in other views than the gallery using a transformation of the test probe norm $\mathbf{g}_p$. The experiments of the next section will demonstrate that this model is sufficient to recognise individuals over a large range of poses given a single training image.

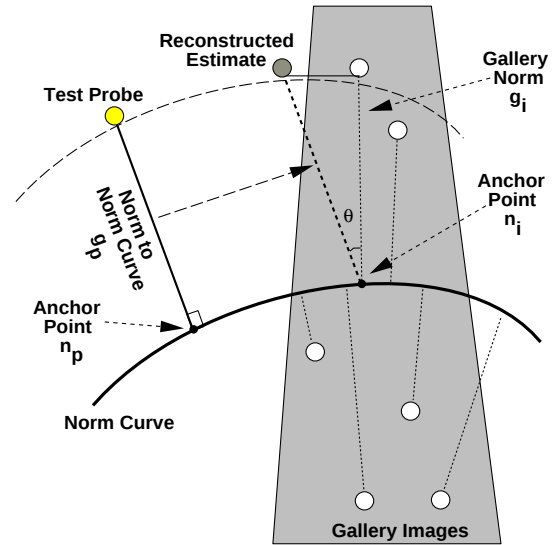


Figure 3: Norm based distance measures.

3 Experimental results

In order to test the above formulation we set ourselves a particularly difficult face recognition task. We supplied the training algorithm with a single, central (3/4) view of each of the 20 individuals in the database. Recall that for each individual we have an average of approximately 28 images covering the pose range from frontal to profile approximately evenly. We shall attempt to recognise every image of each person given only the single training image using a variety of formulations with the norm-based coding approach. While we may not expect high recognition rates from such an experiment, we can show that the norm-based method is a viable representation for recognition and that the pose-free nature of the process demonstrates that a pose-bound model is not essential for a recognition system. We shall see later an application of this approach to a publicly available face database for which comparable recognition results are available.

3.1 Norm-based measures

The main objective of this paper is to establish whether the normal distance vector from a test probe to the population norm curve may be used for recognition purposes. There are several methods of using this distance. Firstly we may use the norm as the identity vector - the vector for the probe $\mathbf{g}_p$ is compared with the gallery of normal vectors $\mathbf{g}_i$ using the value of the

angle between the two vectors,

$$d_{norm} = \theta = \cos^{-1} \frac{\mathbf{g}_p \cdot \mathbf{g}_i}{\|\mathbf{g}_p\| \|\mathbf{g}_i\|}$$

and the minimum of this distance for all members of the gallery is proposed to be the same individual as the probe.

This measure assumes that *at all points* relative to the population norm an individual will deviate in the same direction from the norm. It also ignores the magnitude of this deviation (ignoring caricature effects). To include possible caricature effects we calculate θ as above and then use the ratio of probe to gallery as a scaling factor before finding the maximum value:

$$d_{car} = \frac{\|\mathbf{g}_p\|}{\|\mathbf{g}_i\|} \cos \theta$$

Although this measure is only of relevance when caricature effects have been applied to the probe images and will not be examined here.

More interestingly we may assume that the norm vector for a probe face may be used to estimate the location in the eigenspace of any other pose by adding the norm vector for the probe to the norm curve for the population at all points on norm curve. This is essentially an attempt at reconstructing an image of a face in a different pose from the probe and then matching the reconstructions to the gallery. Recall that each image in the gallery is defined by its norm vector $\mathbf{g}_i$ and the point on the norm curve which anchors it $\mathbf{n}_i$. Recognition using this information essentially performs the nearest neighbour operation on the vectors $\mathbf{n}_i + \mathbf{g}_p$ and $\mathbf{n}_i + \mathbf{g}_i$ for all i . The measure will be referred to as d_{est} and is defined by:

$$d_{est} = \|\mathbf{n}_i + \mathbf{g}_p - (\mathbf{n}_i + \mathbf{g}_i)\| = \|\mathbf{g}_p - \mathbf{g}_i\|$$

Finally, it becomes apparent after initial experiments using these measures that the norm vector alone does not remain constant over a large enough range of poses for recognition purposes. However, there may be a relationship between the norm vectors $\mathbf{g}_i$ and the anchor points on the norm curve $\mathbf{n}_i$. Using any prior experience we have of individuals over a range of poses we may attempt to model this variation and utilise it to enhance the estimate of a probes norm at another pose. Essentially we use our experience of an individual $\mathbf{p}$ over a range of poses $\mathbf{j}$ to give us a set of anchor points $\mathbf{n}_j^p$ and norm vectors $\mathbf{g}_j^p$. We know the gallery pose for this person which is defined by $\mathbf{n}_i$ and $\mathbf{g}_i$. We can then attempt to deduce a relationship between the anchor points $\mathbf{n}_j^p$ and the difference between the

norm vector $\mathbf{g}_j^p$ and the gallery norm vector for that individual $\mathbf{g}_i$. If we assume a linear model for this relationship then we may write:

$$(\mathbf{g}_j^p - \mathbf{g}_i) = \mathbf{W} \mathbf{n}_j^p$$

To summarize our experience of the relationship between all anchor points and the norm differences we calculate $\mathbf{W}$ using:

$$\mathbf{W} = \mathbf{G} \mathbf{N}^+$$

where $\mathbf{G}$ is the matrix of differences between $\mathbf{g}_j^p$ and $\mathbf{g}_i$ and $\mathbf{N}^+$ the pseudo-inverse of the matrix of anchor points. The matrix $\mathbf{W}$ thus gives us the weight matrix of a linear associator relating these factors. We use this matrix to refine our estimate of the norm vector for a probe:

$$\widetilde{\mathbf{g}}_p = \mathbf{g}_p + \mathbf{W} \mathbf{n}_p$$

We can then perform recognition using the d_{est} measure with the refined version of the probe norm. The measure will be referred to as d_{lin} where:

$$d_{lin} = \tilde{\theta} = \cos^{-1} \frac{\widetilde{\mathbf{g}}_p \cdot \mathbf{g}_i}{\|\widetilde{\mathbf{g}}_p\| \|\mathbf{g}_i\|}$$

In the following experiments our experience of faces $\mathbf{G}$ and $\mathbf{N}$ is formed from all available subjects except the probe face and we form the weight matrix $\mathbf{W}$ using the pseudo-inverse equation given earlier.

3.2 Comparing the measures

Table 1 illustrates the performance of these norm-based recognition methods on the database described earlier. The error rates given are for the 1 from 20 recognition task (δ_{20}). The figures in brackets are the standard deviations (over test subject). We note that there is a large amount of deviation for the first two measures indicating a large amount of variation in each particular test subject's distinctiveness when using these measures. The considerable improvement obtained when using the refined version of the normal distance measure is accompanied by a reduction in the amount of variation per subject. This indicates that this measure has captured a further degree of the identity information available within this representation.

d	d_{norm}	d_{est}	d_{lin}
δ_{20}	0.40 (0.21)	0.31 (0.21)	0.13 (0.10)

Table 1: Norm-based error rates (and deviations).

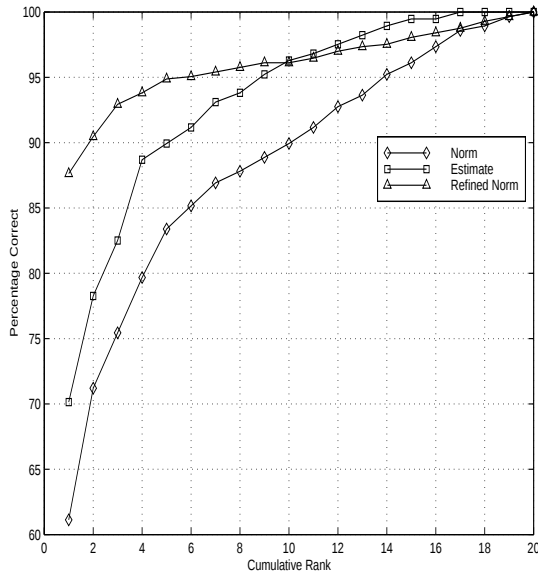


Figure 4: Cumulative rank for all measures.

Figure 4 gives the cumulative rank behaviour for the above experiments using each of the three distance measures. The areas under the curves for each measure are 0.89, 0.93 and 0.96 respectively. It can be seen from these results that the identity information available in the norm-based representation has been effectively captured - especially in the refined norm measure d_{lin} .

3.3 Discussion

Clearly the experiments reported here are demonstrating the usefulness of a population norm curve and vectors normal to this curve for recognition. The considerable decrease in error rates between the simple norm and the adjusted norm demonstrates that the information available in this representation is a persistent and characteristic measure of identity over a range of poses. The refined norm measure d_{lin} shows that 87% of the images ranging from frontal to profile are correctly recognised having previously seen only a single image of each individual. This constitutes a substantial amount of generalisation from a single image and illustrates that norm-based coding extracts an identity-specific representation from the face which can be used as the foundation for a recognition algorithm. It is also feasible that, as the norm vector remains effectively constant over a small range of poses, that this vector can be used to track faces (and hence identity) in image sequences.

4 A comparative experiment

While the results of section 2 illustrate the performance characteristics of a norm-based recognition system the experiments performed to test this recognition lack a comparison with other work in the field. As discussed earlier, few experiments have been reported which permit a comparison of techniques. One paper which does permit a comparison is the one by Duvdevani-Bar et al [1]. In their paper 15 different images of 18 subjects (5 different poses, 3 different expressions) are recognised having previously seen a single (gallery) image of the 18 individuals only. Recognition in their work is carried out by establishing a *chorus of prototypes* using 10 other individuals in the same set of poses. This produces a neural network classifier which produces a 1 for each class (individual) and 0 for the others when presented with an image of the target in any of the 15 poses. A new subject (i.e. one of the gallery images) is then characterised by its response in all 10 of the networks classes. Recognition matches a probe to the 18 characteristic responses to the gallery images. The images used are part of the Weizmann FaceBase [5]. In summary, Duvdevani-Bar et al achieved an error rate of **0.31** having trained an RBF classifier on the 10 training people and using that to recognise the remaining 18 when each of the 18 members of the gallery have only seen the image in the frontal pose with neutral expression (VP2, EX1, IL0). This is the result from the recognition of each test image matched against the full 18 members of the gallery. Matching against only 3 members of the gallery reduced the error rate to **0.08** - which compares with **0.04** of a similar experiment on the same database when human subjects were asked to recognise the same images.

When using the norm-based methods described in section 2 we require some population knowledge to form the norm-curve. For a comparative experiment we limit this norm-curve to be represented the averages of each of the 10 training image at each of the 15 viewing conditions. Additionally the 10 training individuals can be used to attempt to capture any of the systematic variations between the norm vectors over pose changes. Table 2 illustrates the comparative performance of the distance measures d_{norm} , d_{est} and d_{lin} for the recognition tasks of 1 in 18 (δ_{18}) and 1 in 3 (δ_3), with the results of Duvdevani-Bar et. al. (Duv) for comparison.

Here we see that the simple norm measure d_{norm} has a comparable performance to the Duvdevani-Bar experiment. The d_{est} measure is

d	d_{norm}	d_{est}	d_{lin}	Duv
δ_{18}	0.33	0.44	0.56	0.31
δ_3	0.10	0.13	0.16	0.08

Table 2: Error rates for the comparative experiment.

slightly worse - perhaps reflecting the non-continuous nature of the pose changes in the Weizmann Face-Base. The poor performance of the d_{lin} measure is entirely due to the fact that the weight matrix $\mathbf{W}$ is formed from only the 10 training subjects and then must generalise to the 18 test subjects. The experiment was performed in this way in order to compare the two methods as fairly as possible. Without enough experience of the variation of the norm over the viewing condition changes we cannot generalise sufficiently for novel individuals - and the attempted generalisation has a detrimental effect. It should be noted that the comparable performance of the d_{norm} measure is achieved by explicitly capturing the norm-based data and using this information for recognition purposes. It is not unreasonable to assume that such identity information extracted by both models may be similar.

5 Summary

We have demonstrated a method of representing and recognising faces over a large and continuous range of poses. The method of dealing with large pose ranges is only limited by the experience provided by the database used to construct the population norm. We have shown that the representation used in the norm-based scheme exhibits similar performance characteristics to a comparable model despite being limited by a lack of image experience. The norm-based schemes described here models the identity in an explicit manner as a direct consequence of its formation. The norm vectors used for recognition are reliable identity indicators in the eigenspaces constructed over a large range of poses. The present norm-based coding has been carried out in an pose-varying eigenspace yet there is no reason why a norm-based approach should not be used in other subspaces such as a Gabor-wavelet one [9]. Future work will examine the norm-based methods in such alternative subspaces. Of additional interest we can investigate norm-based processing in the image domain as a means of reconstructing high quality images from low quality sequences. We will also examine the nature of identity confusion and similarity within this model including possible caricature effects.

References

- [1] S. Duvdevani-Bar, S. Edelman, A. Jonathan Howell, H. Buxton. "Similarity-based method for the generalization of face recognition over pose and expression." *Proc. of the 3rd Int. Conf. on Automatic Face and Gesture Recognition*, pp 118-123, NARA, Japan, April 1998.
- [2] G.J. Edwards, C.J. Taylor and T. Cootes. "Interpreting Face Images using Active Appearance Models." *Proc. of the 3rd Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, pp 300-305, 1998.
- [3] D. B. Graham and N. M. Allinson. "Face Recognition from Unfamiliar Views: Subspace Methods and Pose Dependency." *Proc. of the 3rd Int. Conf. on Automatic Face and Gesture Recognition*, pp 348-353, NARA, Japan, April 1998.
- [4] S. McKenna, S. Gong and J.J. Collins. "Face Tracking and Pose Representation." *British Machine Vision Conference*, Edinburgh, Scotland, 1996.
- [5] Y. Moses, S. Ullman and S. Edelman. "Generalisation to novel images in upright and inverted faces." *Perception*, vol 25, pp 443-462, 1996.
- [6] S. K. Nayar, H. Murase and S.A. Nene. "Parametric Appearance Representation." in: *Early Visual Learning*, S.K. Nayar and T. Poggio (eds), Oxford University Press, February 1996.
- [7] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven and C. von der Malsburg. "The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test." in *Face Recognition: From Theory to Applications*, NATO ASI Series F, Computer and Systems Sciences, Vol. 163. H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie and T. S. Huang (eds), pp 186-205, 1998.
- [8] G. Rhodes and T. Tremewan. "Understanding face recognition: Caricature effects, inversion, and the homogeneity problem." *Visual Cognition*, vol 1, pp 275-311, 1994.
- [9] L. Wiskott, J.M. Fellous, N. Kruger and C. von der Malsburg. "Face Recognition by Elastic Bunch Graph Matching." Internal Report 96-08, *Institut für Neuroinformatik, Ruhr-Universität Bochum*, April 1996.