

# Three-Dimensional Face Recognition: A Fishersurface Approach

Thomas Heseltine, Nick Pears, Jim Austin

Department of Computer Science, The University of York, United Kingdom

**Abstract.** Previous work has shown that principal component analysis (PCA) of three-dimensional face models can be used to perform recognition to a high degree of accuracy. However, experimentation with two-dimensional face images has shown that PCA-based systems are improved by incorporating linear discriminant analysis (LDA), as with Belhumier et al's fisherface approach. In this paper we introduce the fishersurface method of face recognition: an adaptation of the two-dimensional fisherface approach to three-dimensional facial surface data. Testing a variety of pre-processing techniques, we identify the most effective facial surface representation and distance metric for use in such application areas as security, surveillance and data compression. Results are presented in the form of false acceptance and false rejection rates, taking the equal error rate as a single comparative value.

## 1 Introduction

Despite significant advances in face recognition technology, it has yet to achieve the levels of accuracy required for many commercial and industrial applications. The high error rates stem from a number of well-known sub-problems. Variation in lighting conditions, facial expression and orientation can all significantly increase error rates, making it necessary to maintain consistent image capture conditions between query and gallery images. However, this approach eliminates a key advantage offered by face recognition: a passive biometric that does not require subject co-operation. In an attempt to address these issues, research has begun to focus on the use of three-dimensional face models, motivated by three main factors. Firstly, relying on geometric shape, rather than colour and texture information, systems become invariant to lighting conditions. Secondly, the ability to rotate a facial structure in three-dimensional space, allowing for compensation of variations in pose, aids those methods requiring alignment prior to recognition. Thirdly, the additional discriminatory depth information in the facial surface structure, not available from two-dimensional images, provides supplementary cues for recognition.

In this paper we investigate the use of facial surface data, taken from 3D face models (generated using a stereo vision 3D camera), as a substitute for the more familiar two-dimensional images. A number of investigations have shown that geometric facial structure can be used to aid recognition. Zhao and Chellappa [1] use a generic

3D face model to normalise facial orientation and lighting direction prior to recognition, increasing accuracy from approximately 81% (correct match within rank of 25) to 100%. Similar results are witnessed in the Face Recognition Vendor Test [2], showing that pose correction using Romdhani et al's 3D morphable model technique [3] reduces error rates when applied to the FERET database.

Blanz et al [4] take a comparable approach, using a 3D morphable face model to aid in identification of 2D face images. Beginning with an initial estimate of lighting direction and face shape, Blanz et al iteratively alters shape and texture parameters of the morphable face model, minimising difference to the two-dimensional image. These parameters are then taken as features for identification, resulting in 82.6% correct identifications on a test set of 68 people.

Although these methods show that knowledge of three-dimensional face shape can aid normalisation for two-dimensional face recognition systems, none of the methods mentioned so far use actual geometric structure to perform recognition. Whereas Beumier and Acheroy [5, 6] make direct use of such information, testing various methods of matching 3D face models, although few were successful. Curvature analysis proved ineffective, and feature extraction was not robust enough to provide accurate recognition. However, Beumier and Acheroy were able to achieve reasonable error rates using the curvature of vertical surface profiles. Verification tests carried out on a database of 30 people produced EERs between 7.25% and 9.0% on the automatically aligned surfaces and between 6.25% and 9.5% when manual alignment was used. Chua et al [4] attempt to identify and extract rigid areas of 3D facial surfaces, creating a system invariant to facial expression. The similarity of two face models is computed by comparing a set of unique point signatures for each face. Identification tests show that the probe image is identified correctly for all people when applied to a test set of 30 depth maps of 6 different people.

Hesher et al [7] use PCA of depth maps and a euclidean distance metric to perform identification with 94% accuracy on 37 face models (when training is performed on the gallery set). Further investigation into this approach is carried out by Heseltine et al [8], showing how different surface representations and distance measures affect recognition, reducing the EER from 19.1% to 12.7% when applied to a difficult test set of 290 face models.

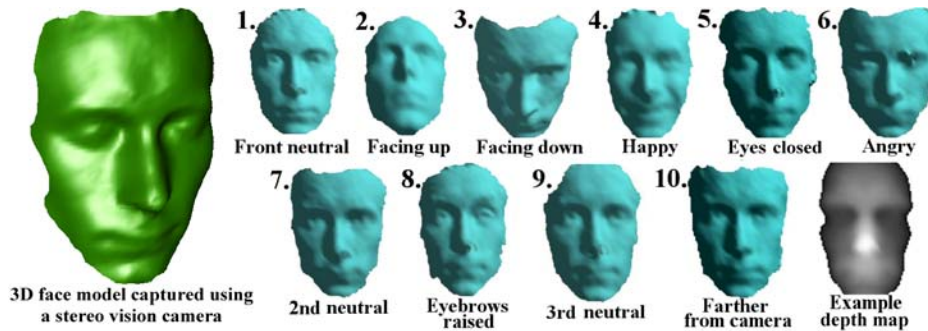
Having achieved reasonable success from the PCA-based eigensurface system in previous work [8], we now continue this line of research, experimenting with another well-known method of face recognition, namely the fisherface approach as described by Belhumeur et al [9], adapted for use on three-dimensional face data. Testing a range of surface representations and distance metrics, we identify the most effective methods of recognising faces using three-dimensional surface structure.

## 2 The 3D Face Database

Until recently, little three-dimensional face data has been publicly available for research and nothing towards the magnitude required for development and testing of three-dimensional face recognition systems. In these investigations we use a new

database of 3D face models, recently made available by The University of York, as part of an ongoing project to provide a publicly available 3D Face Database [10]. Face models are generated in sub-second processing time from a single shot with a 3D camera, using a stereo vision technique enhanced by light projection.

For the purpose of these experiments we select a sample of 1770 face models (280 people) captured under the conditions in Fig. 1. During data acquisition no effort was made to control lighting conditions. In order to generate face models at various head orientations, subjects were asked to face reference points positioned roughly  $45^\circ$  above and below the camera, but no effort was made to enforce precise orientation.



















**Fig. 1.** Example face models taken from The University of York 3D face database

3D models are aligned to face directly forwards before conversion into depth map representation. The database is then separated into two disjoint sets: the training set consisting of 300 depth maps (6 depth maps of 50 people) and a test set of the remaining 1470 depth maps (230 people), consisting of all capture conditions shown in Fig. 1. Both the training and test set contain subjects of various race, age and gender and nobody is present in both the training and test sets.

### 3 Surface Representations

It is well known that the use of image processing can significantly reduce error rates of two-dimensional face recognition methods [11, 12, 13], by removing effects of environmental capture conditions. Much of this environmental influence is not present in 3D face models, however Heseltine et al [8] have shown that such pre-processing may still aid recognition by making distinguishing features more explicit and reducing noise content. In this section we describe a variety of surface representations, derived from aligned 3D face models, which may affect recognition error rates. Pre-processing techniques are applied prior to both training and test procedures, such that a separate surface space is generated for each surface representation and hence a separate face recognition system.

**Table 1.** Descriptions of surface representations with the convolution kernels used

Horizontal Gradient	Vertical Gradient	Horiz. Gradient Large	Vert. Gradient Large
 $\begin{bmatrix} -1 & 1 \end{bmatrix}$	 $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	 $\begin{bmatrix} -1 & 0 & 0 & 0 & 1 \end{bmatrix}$	 $\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$
Applies the 2x1 kernel to compute the horizontal derivative	Applies the 1x2 kernel to compute the vertical derivative	Horizontal gradient over a greater horizontal distance	Vertical gradient over a greater vertical distance
Laplacian	Sobel X	Sobel Y	Sobel Magnitude
 $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	 $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	 $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$	
An isotropic measure of the second spatial derivative	Application of the horizontal sobel derivative filter	Application of the vertical sobel derivative filter	The magnitude of Sobel X and Y combined.
Horizontal Curvature	Vertical Curvature	Curvature Magnitude	Curve Type
			
Applies sobel X twice to calculate the 2nd horizontal derivative	Applies sobel Y twice to calculate the 2nd vertical derivative	The magnitude of the vertical and horizontal curvatures	Segmentation of the surface into 8 discrete curvature types
Min Curvature	Max Curvature	Abs Min Curvature	Abs Max Curvature
			
The minimum of the horizontal & vertical curvature values	The maximum of horizontal & vertical curvature values	The minimum of absolute horizontal & vertical curvatures	The maximum of absolute horizontal & vertical curvatures

## 4 The Fishersurface Method

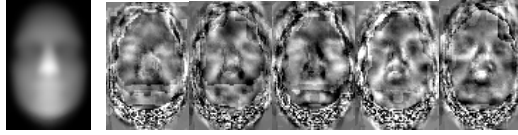
In this section we provide details of the fishersurface method of face recognition. We apply PCA and LDA to surface representations of 3D face models, producing a subspace projection matrix, as with Belhumeur et al's fisherface approach [9], taking advantage of 'within-class' information, minimising variation between multiple face models of the same person, yet maximising class separation. To accomplish this we use a training set containing several examples of each subject, describing facial structure variance (due to influences such as facial expression), from one model to another. From the training set we compute three scatter matrices, representing the within-class ( $S_W$ ), between-class ( $S_B$ ) and total ( $S_T$ ) distribution from the average surface  $\Psi$  and classes averages  $\Psi_n$ , as shown in equation 1.

$$\begin{aligned}
\text{Training Set} &= \{X_1, X_2, \dots, X_c\} \\
\text{where } X_n &= \{\Gamma_{n1}, \Gamma_{n2}, \Gamma_{n3}, \dots\} \\
\Psi &= \frac{1}{\sum_{n=1}^c |X_n|} \sum_{n=1}^c \sum_{j=1}^{|X_n|} \Gamma_{nj} \\
\Psi_n &= \frac{1}{|X_n|} \sum_{i=1}^{|X_n|} \Gamma_{ni} \\
S_T &= \sum_{n=1}^c \sum_{i=1}^{|X_n|} (\Gamma_{ni} - \Psi)(\Gamma_{ni} - \Psi)^T \\
S_B &= \sum_{n=1}^c |X_n| (\Psi_n - \Psi)(\Psi_n - \Psi)^T \\
S_W &= \sum_{n=1}^c \sum_{i=1}^{|X_n|} (\Gamma_{ni} - \Psi_n)(\Gamma_{ni} - \Psi_n)^T
\end{aligned} \tag{1}$$

The training set is partitioned into  $c$  classes, such that all surface vectors  $\Gamma_{ni}$  in a single class  $X_n$  are of the same person and no person is present in multiple classes. Calculating eigenvectors of the matrix  $S_T$ , and taking the top 250 (number of surfaces minus number of classes) principal components, we produce a projection matrix  $U_{pca}$ . This is then used to reduce dimensionality of the within-class and between-class scatter matrices (ensuring they are non-singular) before computing the top  $c-1$  eigenvectors of the reduced scatter matrix ratio,  $U_{fld}$ , as shown in equation 2.

$$\begin{aligned}
U_{fld} &= \arg \max_U \left( \frac{U^T U_{pca}^T S_B U_{pca} U}{U^T U_{pca}^T S_W U_{pca} U} \right) \\
U_{pca} &= \arg \max_U (U^T S_T U) \\
U_{ff} &= U_{fld} U_{pca}
\end{aligned} \tag{2}$$

Finally, the matrix  $U_{ff}$  is calculated, such that it projects a face surface vector into a reduced space of  $c-1$  dimensions, in which between-class scatter is maximised for all  $c$  classes, while within-class scatter is minimised for each class  $X_n$ . Like the eigenface system, components of the projection matrix  $U_{ff}$  can be viewed as images, as shown in Fig. 2 for the depth map surface space.



**Fig. 2.** The average surface (left) and first five fishersurfaces (right)

Once surface space has been defined, we project a facial surface into reduced surface space by a simple matrix multiplication, as shown in equation 3.

$$\Omega = (\Gamma - \Psi)^T U_{ff} \tag{3}$$

The vector  $\Omega^T = [\omega_1, \omega_2, \dots, \omega_{c-1}]$  is taken as a ‘face-key’ representing the facial structure in reduced dimensionality space. Face-keys are compared using either euclidean or cosine distance measures as shown in equation 4.

$$d_{euclidean} = \|\Omega_a - \Omega_b\| \quad d_{cosine} = 1 - \frac{\Omega_a^T \Omega_b}{\|\Omega_a\| \|\Omega_b\|} \tag{4}$$

An acceptance (facial surfaces match) or rejection (surfaces do not match) is determined by applying a threshold to the distance calculated. Any comparison producing a distance value below the threshold is considered an acceptance.

## 5 The Test Procedure

In order to evaluate the effectiveness of a surface space, we project and compare the 1470 face surfaces with every other surface in the test set, no surface is compared with itself and each pair is compared only once (1,079,715 verification operations). The false acceptance rate (FAR) and false rejection rate (FRR) are then calculated as the percentage of incorrect acceptances and incorrect rejections after applying a threshold. Varying the threshold produces a series of FAR/FRR pairs, which plotted on a graph produce an error curve as seen in Fig. 5. The equal error rate (EER, the point at which FAR equals FRR) can then be taken as a single comparative value.

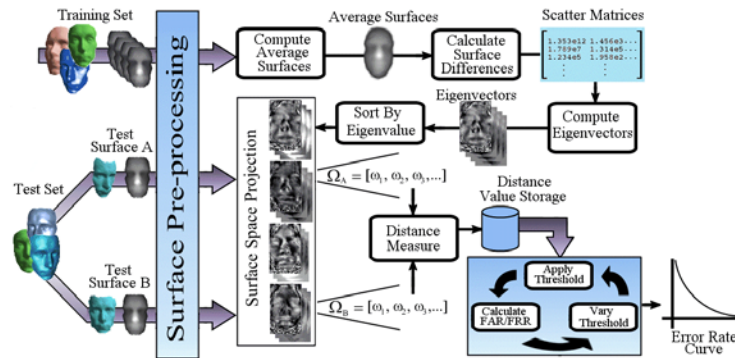


Fig. 3. Flow chart of system evaluation procedure

## 6 Results

In this section we present results gathered from performing 1,079,715 verification operations on the test set of 1470 face models, using the surface representations described in section 3. Systems are tested separately using Euclidean and cosine distance measures. In addition we provide a direct comparison to the eigensurface method [8] trained and tested using the same face models, distance metrics and the same number of ( $c-1$ ) principal components.

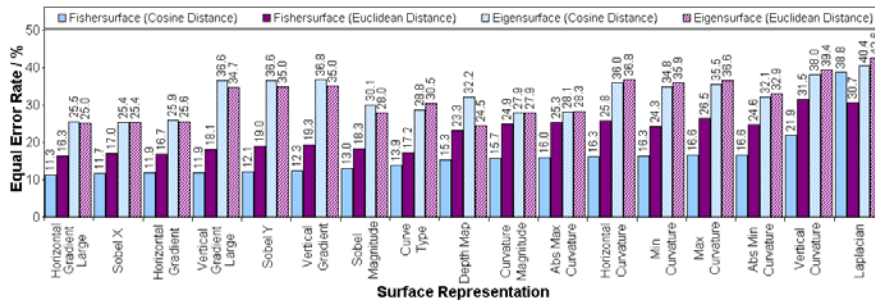


Fig. 4. EERs of fishersurface and eigensurface systems using two distance metrics

Fig. 4 shows the diversity of error for eigensurface and fishersurface methods, using cosine and Euclidean metrics for the range of surface representations. The initial depth map produces an EER of 23.3% (euclidean distance) and 15.3% (cosine distance). This trend is common for all fishersurface systems, with the cosine distance typically producing three quarters of the error produced by the euclidean distance. In all cases the EERs of the fisherface system are lower than the equivalent eigensurface method. Surface gradient representations are the most distinguishing, with horizontal derivatives providing the lowest error of 11.3% EER.

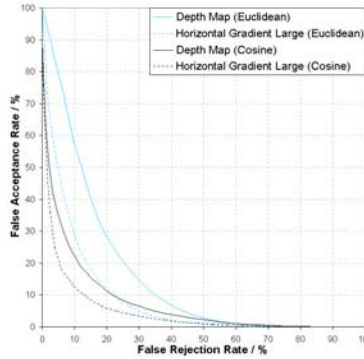


Fig. 5. Fishersurface system error curves using two distance metrics and surface representations

## 7 Conclusion

We have applied a well-known method of two-dimensional face recognition to three-dimensional face models using a variety of facial surface representations. The error rates produced using the initial depth map representation (15.3% and 23.3% EER) show a distinct advantage over the previously developed eigensurface method (32.2% and 24.5% EER). This is also the case for the optimum surface representations, producing 11.3% EER for the fishersurface system and 24.5% EER for the eigensurface method. We also note an increase in the eigensurface EERs compared to those reported in previous work [8]. This could be attributed to the different training and test data, or possibly the different number of principal components used.

Experimenting with a number of surface representations, we have discovered common characteristics between the eigensurface and fishersurface methods: facial surface gradients provide a more effective representation for recognition, with horizontal gradients producing the lowest error rate (11.3% EER). Another observation, also common to the eigensurface method is that curvature representations seem to be least useful for recognition, although this could be a product of inadequate 3D model resolution and high noise content. In which case smoothing filters and larger convolution kernels may produce better results.

The fishersurface method appears to produce better results than corresponding two-dimensional fisherface systems (17.8% EER) tested under similar conditions in previous investigations [13], although a more direct comparison is required, using a common test database, in order to draw any quantitative conclusions.

Testing two distance measures has shown that the choice of metric has a considerable effect on resultant error rates. For all surface representations, the cosine distance produced substantially lower EERs. This is in stark contrast to the eigensurface method, in which Euclidean and cosine measures seem tailored to specific surface representations. This suggests that incorporating LDA produces a surface space with predominantly radial between-class variance, regardless of the surface representation, whereas when using PCA alone, this relationship is dependant on the type of surface representation used.

In summary, we have managed to reduce error rates from 15.3% EER using initial depth maps, to an EER of 11.3% using a horizontal gradient representation. This improvement over the best eigensurface system shows that incorporation of LDA improves performance in three-dimensional as well as two-dimensional face recognition approaches. Given that the 3D capture method produces face models invariant to lighting conditions and provides the ability to recognise faces regardless of pose, this system is particularly suited for use in security and surveillance applications.

## References

1. Zhao, W., Chellappa, R.: 3D Model Enhanced Face Recognition. In Proc. of the Int. Conf. on Image Processing, Vancouver (2000)
2. Phillips, P.J., Grother, P., Micheals, R.J., Blackburn, D.M., Tabassi, E., Bone, J.M.: FRVT 2002: Overview and Summary. <http://www.frvt.org/FRVT2002/documents.htm>, (2003)
3. Romdhani, S., Blanz, V., Vetter, T.: Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions. The ECCV (2002)
4. Blanz, V., Romdhani, S., Vetter, T.: Face Identification across Different Poses and Illuminations with a 3D Morphable Model. In Proc. of the 5th IEEE Conf. on AFGR (2002)
5. Beumier, C., Acheroy, M.: Automatic 3D Face Authentication. Image and Vision Computing, Vol. 18, No. 4, (2000) 315-321
6. Beumier, C., Acheroy, M.: Automatic Face Verification from 3D And Grey Level Clues. In Proc. Of the 11th Portuguese Conference on Pattern Recognition (2000)
7. Heshner, C., Srivastava, A., Erlebacher, G.: Principal Component Analysis of Range Images for Facial Recognition. In Proc. CISST (2002)
8. Heseltine, T., Pears, N., Austin, J.: Three-Dimensional Face Recognition: An Eigensurface Approach. In Proc. of the International Conference on Image Processing (2004)
9. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Face Recognition using class specific linear projection. The European Conference on Computer Vision, (1996)
10. The 3D Face Database, The University of York. [www.cs.york.ac.uk/~tomh](http://www.cs.york.ac.uk/~tomh)
11. Adini, Y., Moses, Y., Ullman, S.: Face Recognition: the Problem of Compensating for Changes in Illumination Direction. IEEE Trans. on Pattern Analysis and Machine Intelligence, (1997) 721-732
12. Heseltine, T., Pears, N., Austin, J.: Evaluation of image pre-processing techniques for eigenface based face recognition. In Proc. of the 2nd International Conference on Image and Graphics, SPIE vol. 4875, 677-685 (2002)
13. Heseltine, T., Pears, N., Austin, J., Chen, Z.: Face Recognition: A Comparison of Appearance-Based Approaches. In Proc. VIIth Digital Image Computing: Techniques and Applications (2003)