Three-Dimensional Face Recognition Using Surface Space Combinations

Thomas Heseltine, Nick Pears, Jim Austin
Department of Computer Science
The University of York
tom.heseltine@cs.york.ac.uk

Abstract

In this paper we test a range of three-dimensional face recognition systems, based on the fishersurface method developed in previous work. We show the effect of using a variety of facial surface representations and suggest a method of identifying and extracting useful qualities offered by each system. Combing these components into a unified surface subspace, we create a three-dimensional face recognition system producing significantly lower error rates than individual systems tested on the same data. We evaluate systems by performing up to 1,079,715 verification operations on a large test set of 3D face models. Results are presented in the form of false acceptance and false rejection rates, generated by varying a decision threshold applied to a distance metric in combined surface space.

1 Introduction

Despite significant advances in face recognition technology, it has yet to achieve levels of accuracy required for many commercial and industrial applications. The high error rates stem from well-known sub-problems. Variation in lighting, facial expression and orientation all significantly increase error rates. 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 depth information in the facial surface structure, not available from two-dimensional images, provides supplementary cues for recognition.

In this paper we expand on previous research [1] involving the use of facial surface data, derived 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 three-dimensional structure can be used to aid recognition. Zhao and Chellappa [2] use a generic 3D face model to normalise facial orientation and lighting direction in two-dimensional images, increasing recognition accuracy from approximately 81% (correct match within rank of 25) to 100%. Similar results are witnessed in the Face Recognition Vendor Test [3], showing that pose correction using Romdhani et al's technique [4] reduces error rates when applied to the FERET database.

Blanz et al [5] take a comparable approach, using a morphable face model to aid in identification of 2D images. Beginning with an initial estimate of lighting direction and face shape, Romdhani 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 three-dimensional geometric structure to perform recognition. Whereas Beumier and Acheroy [6, 7] 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 curvature values of vertical surface profiles. Verification tests carried out on a database of 30 people produced equal error rates (EER) between 7.25% and 9.0%. Hesher et al [8] test a different method, using PCA (principal component analysis) of depth maps and euclidean distance to perform identification with 94% accuracy on 37 face models (when trained on the gallery set). Further investigation into this approach is carried out by Heseltine et al [9], 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. However, the focus of this research has been on identifying optimum surface representations, with little regard for the advantages offered by each individual representation. We suggest that different surface representations may be specifically suited to different capture conditions or certain facial characteristics, despite a general weakness for overall recognition. For example, curvature representations may aid recognition by making the system more robust to inaccuracies in 3D orientation yet also be highly sensitive to noise. Another representation may enhance nose shape, but lose information regarding jaw structure.

In this paper we analyse and evaluate a variety of three-dimensional fishersurface [1] face recognition systems, each incorporating a different surface representation of facial structure. We propose a means of identifying and extracting components from the surface subspace produced by each system, such that they may be combined into a single unified subspace. Pentland et al [10] have previously examined the benefit of using multiple eigenspaces, in which specialist subspaces were constructed for various facial orientations, from which cumulative match scores were able to reduce error rates. Our approach differs in that we extract and combine individual dimensions, creating a single unified surface space, as applied to two-dimensional images in previous investigations [11].

3 The Fishersurface Method

In this section we provide details of the fishersurface method of face recognition. We apply PCA and LDA (linear discriminant analysis) to surface representations of 3D face models, producing a subspace projection matrix, as with Belhumier et al's fisherface approach [12], 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 Ψ and classes averages Ψ_n, as shown in equation 1.

$$\begin{aligned} &Training\ Set = \{X_1, X_2, \ldots X_c\} \\ &where\ X_n = \{\Gamma_{n1}, \Gamma_{n2}, \Gamma_{n3}, \ldots\} \\ &\Psi = \frac{1}{\sum_{m=1}^{c}|X_m|}\sum_{n=1}^{c}\sum_{i=1}^{|X_n|}\Gamma_{ni} \quad . \\ &\Psi_n = \frac{1}{|X_n|}\sum_{i=1}^{|X_n|}\Gamma_{ni} \quad . \end{aligned} \qquad \begin{aligned} S_T &= \sum_{n=1}^{c}\sum_{i=1}^{|X_n|}(\Gamma_{ni}-\Psi)(\Gamma_{ni}-\Psi)^T \quad . \\ S_B &= \sum_{n=1}^{c}|X_n|(\Psi_n-\Psi)(\Psi_n-\Psi)^T \quad . \\ S_W &= \sum_{n=1}^{c}\sum_{i=1}^{|X_n|}(\Gamma_{ni}-\Psi_n)(\Gamma_{ni}-\Psi_n)^T \quad . \end{aligned} \qquad \textbf{(1)}$$

The training set is partitioned into c classes, such that all surface vectors Γ_{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.

$$U_{fld} = \arg\max_U\left(\frac{\left|U^T U_{pca}^T S_B U_{pca} U\right|}{\left|U^T U_{pca}^T S_W U_{pca} U\right|}\right) \quad . \qquad U_{pca} = \arg\max_U\left(\left|U^T S_T U\right|\right) \quad . \qquad \textbf{(2)}$$
$$U_{ff} = U_{fld}\, U_{pca} \quad .$$

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 fisherface system [12], components of the projection matrix U_{ff} can be viewed as images, as shown in Figure 1 for the depth map surface space.

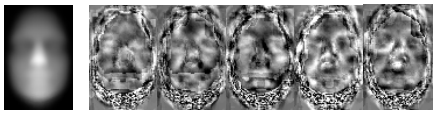

Figure 1: The average surface (*left*) and first five fishersurfaces (*right*)

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.

$$\Omega = (\Gamma - \Psi)^T U_{ff} \quad . \qquad \textbf{(3)}$$

The vector $\Omega^T=[\omega_1,\omega_2,\ldots,\omega_{c-1}]$ 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.

$$d_{euclidean} = \left\|\Omega_a - \Omega_b\right\| \quad . \qquad d_{cosine} = 1 - \frac{\Omega_a^T \Omega_b}{\left\|\Omega_a\right\|\left\|\Omega_b\right\|} \quad . \qquad \textbf{(4)}$$

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.

3 The Test 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 [13]. 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 Figure 2. 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° above and below the camera, but no effort was made to enforce precise orientation.

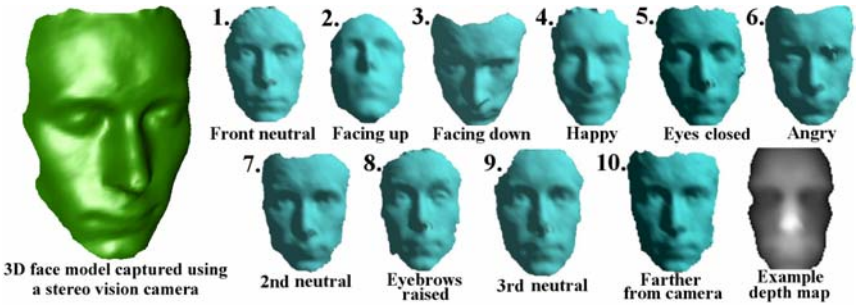

Figure 2: Example face models taken from the University of York 3D Face Database

3D models are aligned to face directly forwards before conversion into 60 by 90 pixel depth map representation. We then take a training set of 300 depth maps (50 people), used to compute the scatter matrices described in section 3. The remaining 1470 depth maps (230 people) are then separated into two disjoint sets of equal size (test set A and test set B). We use test set A to analyse the face-key variance throughout surface space, calculate discriminant weightings (see section 4) and compute the optimum surface space combinations. This leaves set B as an unseen test set to evaluate the final combined system. Both training and test sets contain subjects of various race, age and gender and nobody is present in both the training and test sets.

4 Surface Space Analysis

In this section we analyse the surface spaces produced when various facial surface representations are used with the fishersurface method. We begin by testing the variety of fishersurface systems introduced by Heseltine et al [1] on test set A, showing the range of error rates produced when using various surface representations (Figure 3). Continuing this line of research we persist with the same surface representations, referring the reader to previous work [1, 9] for implementation details, while in this paper we focus on the effect and methodologies of combining multiple systems, rather than the surface representations themselves.

Figure 3 clearly shows the choice of surface representation has a significant impact on the effectiveness of the fishersurface approach, with horizontal gradient representations providing the lowest EER (point at which false acceptance rate equals false rejection rate).

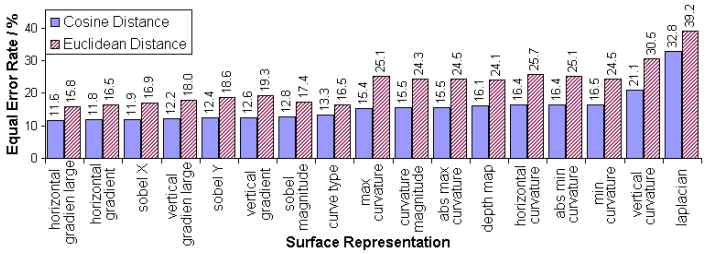

Figure 3: Equal error rates of fishersurface systems applied to test set A

However, the superiority of the horizontal gradient representations does not suggest that the vertical gradient and curvature representations are no use whatsoever. Although discriminatory information provided by these representations may not be as robust and distinguishing, they may contain a degree of information not available in horizontal gradients and could therefore still make a positive contribution to a combined surface space. We measure the discriminating ability of surface space dimensions by applying Fisher's Linear Discriminant (FLD) (as used by Gordon [14]) to individual components (single dimensions) of each surface space. We calculate the discriminant d_n, describing the discriminating power of a given dimension n, between c people in test set A.

$$d_n = \frac{\sum_{i=1}^{c}(m_i - m)^2}{\sum_{i=1}^{c}\frac{1}{|\Phi_i|}\sum_{x\in\Phi_i}(x - m_i)^2} \qquad m_i = \frac{1}{|\Phi_i|}\sum_{x\in\Phi_i} x \qquad m = \frac{1}{\sum_{i=1}^{c}|\Phi_i|}\sum_{i=1}^{c}\sum_{x\in\Phi_i} x \tag{3}$$

Where Φ_i is the set of all class i face-key vector elements in dimension n, and m and m_i the mean and class mean of nth dimension elements in test set A. Applying equation 3 to the assortment of surface space systems listed in Figure 3, we see a wide range of discriminant values across the individual surface space dimensions (Figure 4).

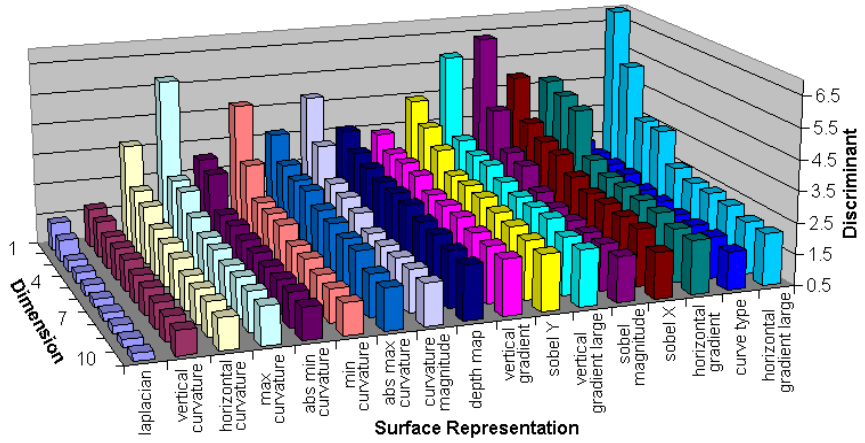

Figure 4: Top ten discriminant values of all fishersurface dimensions

It is clear that although some surface representations do not perform well in the face recognition tests, producing high EERs, some face-key components do contain highly discriminatory information. For example, we see that the min and max curvature representations contain one dimension with a higher discriminant than any horizontal gradient and curve type dimension, yet the EERs are significantly higher. We hypothesise that the reason for these highly discriminating anomalies, in an otherwise ineffective subspace, is that a certain surface representation may be particularly suited to a single discriminating factor, such as nose shape or jaw structure, but is not effective when used as a more general classifier. Therefore, if we were able to isolate these few useful qualities from the more specialised subspaces, they could be used to make a positive contribution to a generally more effective surface space, reducing error rates further.

5 Combining Systems

In this section we describe how the analysis methods discussed in section 4 are used to combine multiple face recognition systems. Firstly, we need to address the problem of prioritising surface space dimensions. Because the average magnitude and deviation of face-key vectors from a range of systems are likely to differ by some orders of magnitude, certain dimensions will have a greater influence than others will, even if the discriminating abilities are evenly matched. To compensate for this effect, we normalise moments by dividing each face-key element by its within-class standard deviation (calculated from test set A face-keys). However, in normalising these dimensions we have also removed any prioritisation, such that all surface space components are considered equal. Although not a problem when applied to a single surface space, when combining multiple dimensions we would ideally wish to give greater precedence to the more reliable components. Otherwise the situation is likely to arise when a large number of less discriminating dimensions begin to outweigh the fewer more discriminating dimensions, diminishing their influence on the verification operation and hence increasing error rates. In section 4 we showed how FLD could be used to measure the discriminating ability of a single dimension from any given face space. We now apply this discriminant value d_n as weighting for each surface space dimension n, prioritising those dimensions with the highest discriminating ability.

With this weighting scheme applied to all face-keys, we now require some criterion to decide which dimensions to combine. It is not enough to rely purely on the discriminant value itself, as this only provides an indication of the discriminating ability of that dimension alone, without any indication of whether the inclusion of this dimension would benefit the existing set of dimensions. If an existing surface space already provides a certain amount of feature specific discrimination, it would be of little benefit (or could even be detrimental) if we were to introduce an additional dimension describing a feature already present within the existing set.

Previous investigations [11] have used FLD, applied to a combined subspace in order to predict effectiveness when used for recognition. Additional dimensions are introduced if they result in an increase in discriminant value. This method has been shown to produce face space combinations achieving significantly lower error rates than individual two-dimensional systems, although Heseltine et al do note that an EER-based criterion is likely to produce a better combination, at the expense of greatly increased training time. However, with a more efficient program and greater computational

resources, we now take that approach: the criterion required for introduction of a new dimension to an existing surface space is a resultant decrease in EER (computed using test set A).

Combined surface space = first dimension of current optimum system
Compute *EER* of *combined surface space*
For each surface representation system:
    For each *dimension* of surface space:
        Concatenate *dimension* onto *combined surface space*
        Compute *EER* of *combined surface space*
        If *EER* has not decreased:
            Remove *dimension* from *combined surface space*
Save *combined surface space* ready for evaluation

Figure 5: Fishersurface combination algorithm

6 The Test Procedure

In order to evaluate the effectiveness of a surface space, we project and compare each facial surface with every other surface in the test set, no surface is compared with itself and each pair is compared only once. 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. By varying the threshold, we produce a series of FAR FRR pairs, which plotted on a graph produce an error curve as seen in Figure 8. The equal error rate (EER, the point at which FAR equals FRR) can then be taken as a single comparative value.

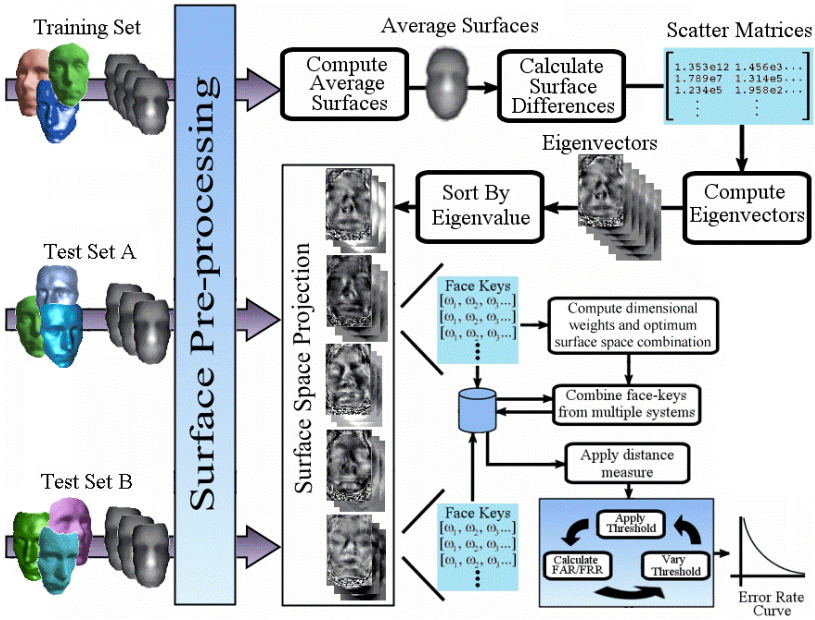

Figure 6: Flow chart of system evaluation procedure

7 Results

In this section we present the dimensions selected to form the combined fishersurface systems (Figure 7) and the error rates obtained from a range of tests sets, making a comparison to optimum individual systems in Figure 8.

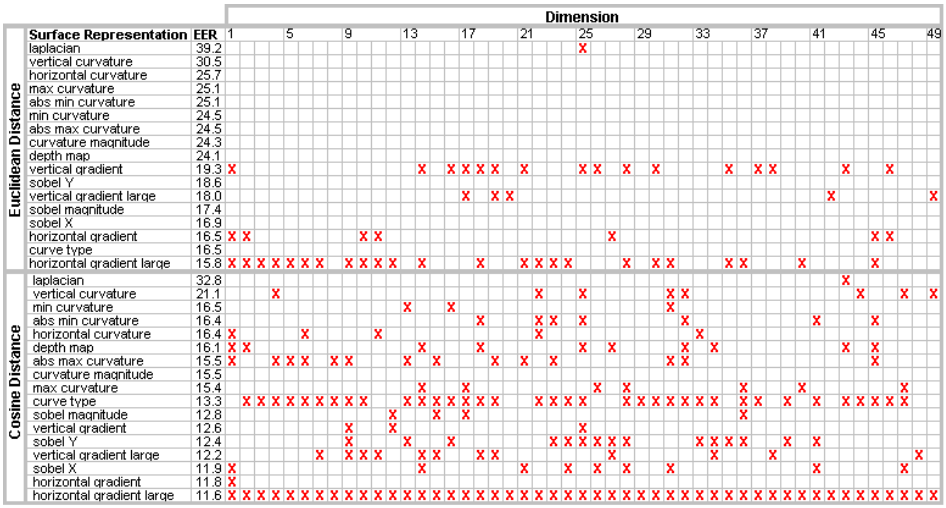

Figure 7: Face space dimensions included (**x**) in the combined fishersurface systems

We see that systems with lower EERs generally make the most contribution to the combined system, as would be expected. However, it is also interesting to note that even systems with particularly high EERs do contain some dimensions that make a positive contribution, although this is much more prominent for the cosine distance, showing that this metric is more suited to combing multiple surface spaces.

Having selected and combined the range of dimensions shown in Figure 7, we now apply these combined systems to test sets A and B using both the cosine and euclidean distance metric. We also perform an evaluation on the union of test sets A and B: an experiment analogous to training on a database (or gallery set) of known people, which are then compared with newly acquired (unseen) images.

Figure 8 shows the error curves obtained when optimum individual fishersurface systems and combined systems are applied to test set A (used to construct the combination), test set B (the unseen test set) and the full test set (all surfaces from sets A and B), using the cosine and Euclidean distance metrics. We see that the combined systems produce lower error rates than the optimum individual systems for all six experiments. As would be expected, the lowest error rates are achieved when tested on the surfaces used to construct the combination (7.2% and 12.8% EER respectively). However an improvement is also seen when applied to the unseen test set B, from 11.5% and 17.3% using the best single systems to 9.3% and 16.3% EER for the combined systems. Performing the evaluation on the larger set, providing 1,079,715 verification operations (completed in 14 minutes 23 seconds on a Pentium III 1.2GHz processor, providing a verification rate of 1251 per second), the error drops slightly to 8.2% and

14.4% EER, showing that a small improvement is introduced if some test data is available for training, as well as suggesting that the method scales well, considering the large increase in verification operations.

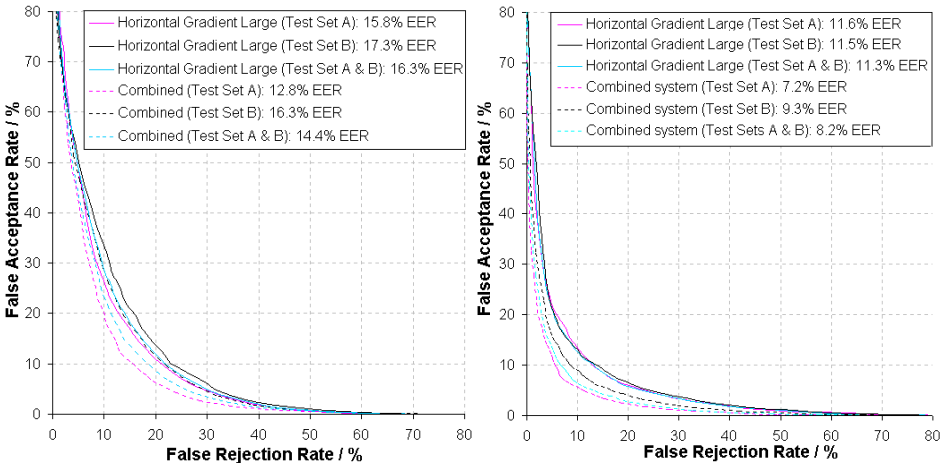

Figure 8: Error curves comparing combined (*dashed lines*) and individual (*solid lines*) systems using the Euclidean (*left*) and cosine (*right*) distance measures.

8 Conclusion

We have shown how a well-known method of two-dimensional face recognition can be applied to three-dimensional face models achieving reasonably low error rates, depending on the surface representation used. Drawing on previous work combing face recognition eigenspaces [11], we have applied the same principle to multiple three-dimensional face recognition systems, showing that the combination method is applicable to both two-dimensional and three-dimensional data. Using FLD as an analysis tool, we have confirmed the hypothesis that although some surface representations may not perform well when used for recognition, they may harbour highly discriminatory components that could complement other surface spaces.

Iteratively improving error rates on a small test set, we have built up a combination of dimensions extracted from a variety of surface spaces, each utilising a different surface representation. This method of combination has been shown to be most effective when used with the cosine distance metric, in which a selection of 184 dimensions were combined from 16 of the 17 surface spaces, reducing the EER from 11.6% to 8.2%. Applying the same combined surface space to an unseen test set of data presenting typical difficulties when performing recognition, we have demonstrated a similar reduction in error from 11.5% to 9.3% EER.

Evaluating the combined system at its fundamental level, using 1,079,715 verification operations between three-dimensional facial surfaces, demonstrates that combining multiple surface space dimensions improves effectiveness of the core recognition algorithm. Error rates have been significantly reduced to state-of-the-art levels, when evaluated on a difficult test set including variations in expression and orientation. However, we have not applied any additional heuristics, typically

incorporated into fully functional commercial and industrial systems. For example, we have not experimented with multiple facial alignments, optimising crop regions or storing multiple gallery images. All of which are known to improve error rates and can easily be applied to the combined systems presented in this paper. With these additional measures in place, it is likely that the improvements made to the core algorithm will propagate through to producing a highly effective face recognition system. Given the fast 3D capture method, small face-keys of 184 vector elements (allowing extremely fast comparisons), invariance to lighting conditions and facial orientation, this system is particularly suited to security and surveillance applications.

References

[1] Heseltine, T., Pears, N., Austin, J. *Three-Dimensional Face Recognition: A Fishersurface Approach*. In Proc. of the International Conference on Image Analysis and Recognition (2004)

[2] Zhao, W., Chellappa, R. *3D Model Enhanced Face Recognition*. In Proc. of the International Conference on Image Processing (2000)

[3] Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E., Bone, J. *FRVT 2002: Overview and Summary*. www.frvt.org/FRVT2002, March (2003)

[4] Romdhani, S., Blanz, V., Vetter, T. *Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions*. The European Conference on Computer Vision (2002)

[5] Blanz, V., Romdhani, S., Vetter, T. *Face Identification across Different Poses and Illuminations with a 3D Morphable Model*. In Proc. of the 5th IEEE Conference on Automatic Face and Gender Recognition (2002)

[6] Beumier, C., Acheroy, M. *Automatic 3D Face Authentication*. Image and Vision Computing, Vol. 18, No. 4, (2000)

[7] Beumier, C., Acheroy, M. *Automatic Face Verification from 3D And Grey Level Clues*. 11th Portuguese Conference on Pattern Recognition (2000)

[8] Hesher, C., Srivastava, A., Erlebacher, G. *Principal Component Analysis of Range Images for Facial Recognition*. In Proc. CISST (2002)

[9] Heseltine, T., Pears, N., Austin, J. *Three-Dimensional Face Recognition: An Eigensurface Approach*. In Proc. of the International Conference on Image Processing (2004)

[10] Pentland, A., Moghaddom, B., Starner, T. View-Based and Modular Eigenfaces for Face Recognition, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (1994)

[11] Heseltine, T., Pears, N., Austin, J. *Combining multiple face recognition systems using Fisher's linear discriminant*. In Proc. of the SPIE Defense and Security Symposium (2004)

[12] Belhumeur, P., Hespanha, J., Kriegman, D. *Eigenfaces vs. Fisherfaces: Face Recognition using class specific linear projection*. Proc. of the European Conference on Computer Vision (1996)

[13] The 3D Face Database, The University of York. www.cs.york.ac.uk/~tomh

[14] Gordon, G. *Face Recognition Based on Depth and Curvature Features*. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (1992)