Face recognition from 3D data using Iterative Closest Point algorithm and Gaussian mixture models

Conference Paper · October 2004

DOI: 10.1109/TDPVT.2004.1335279 · Source: IEEE Xplore

4 authors, including:

V. Chandran
Queensland University of Technology
234 PUBLICATIONS **5,277** CITATIONS

SEE PROFILE

FACE RECOGNITION FROM 3D DATA USING ITERATIVE CLOSEST POINT ALGORITHM AND GAUSSIAN MIXTURE MODELS

Jamie Cook, Vinod Chandran, Sridha Sridharan and Clinton Fookes

Image and Video Research Lab
Queensland University of Technology
2 George St. Brisbane Qld 4000, Australia
{j.cook,v.chandran,s.sridharan,c.fookes}@qut.edu.au

ABSTRACT

A new approach to face verification from 3D data is presented. The method uses 3D registration techniques designed to work with resolution levels typical of the irregular point cloud representations provided by Structured Light scanning. Preprocessing using a-priori information of the human face and the Iterative Closest Point algorithm are employed to establish correspondence between test and target and to compensate for the non-rigid nature of the surfaces. Statistical modelling in the form of Gaussian Mixture Models is used to parameterise the distribution of errors in facial surfaces after registration and is employed to differentiate between intra- and extra-personal comparison of range images. An Equal Error Rate of 2.67% was achieved on the 30 subject manual subset of the the 3d_rma database.

1. INTRODUCTION

Traditional 2D Face Recognition Technologies (FRT) have struggled to cope with variations in lighting and pose. Three dimensional facial data can be used to covercome these issues. 3D data is by definition lighting invariant and the task of pose normalisation becomes more tractable with knowledge of the physical surfaces being considered.

Early work with 3D facial recognition focussed on the use of surface curvature information and the Extended Gaussian Image (EGI), which provides a one-to-one mapping of the surface's curvature normals to the unit sphere. Lee *et, al.* [1] located convex portions of the face, deemed more stable under changes due to facial expession, and used graph matching techniques for recognition. Gordon [2] utilised face descriptors based on the nose and eyes along with a simple Euclidean distance classifier. Tanaka [3, 4] approached a solution by first generating a reduced mapping of the EGI, retaining only surface normals that lay on a ridge or valley line, then using Fischer's Spherical Correlations for identification.

A key limitation of such approaches is that the direct use of curvature data depends heavily on accurate 3D acquisition. Without highly accurate data, curvature analysis is prone to spurious results. These authors use laser range data as the basis of their experimentation, and while laser range finding gives the highest accuracy of current acquisition technology, the equipment required is very expensive compared to other acquisition methods such as Stereopsis and Structured Light Scanning (SLS).

Current research has been focussing on the use of 3D models to increase the number of training samples available to new and existing 2D image recognisers. Zhao [5, 6] uses a 3D head model and Shape from Shading (SfS) algorithms to synthesise a prototype face surface from noisy intensity images which can be artificially illuminated to generate new normalised intensity images. This idea is extended by Blanz *et al.* [7, 8] to the creation of full 3D models for any face given a single input image. Huang [9] utilises these 3D models to generate training images for a component based Support Vector Machine (SVM) classifier. Blanz and Vetter [7] use the model parameters directly to compare faces using the summation of Mahalanobis distances of both shape and texture parameters. These results have shown promise but the sheer amount of processing required to obtain the model parameters (4.5 minutes on a 2GHz P4) reduces the usefulness for real world applications.

Methods for comparing faces using pure 3D data have also been proposed. Pan [10] proposed using registration error of low resolution SLS data and Lee [11] used contour lines extracted from a normalised laser scan as the basis of their recognition. Both methods demonstrate the efficacy of their approaches but both lack comprehensive testing strategies and do not provide definitive solutions.

Recently Bronstein and Bronstein [12] have patented [13] their approach to 3D face recognition using bending-invariant canonical forms. This technique makes use of the empirical observation that while transformations of the human face are non-rigid in nature, the set of possible transformations belongs to the isometric (or length preserving) set of transformations. In other words the deformations caused

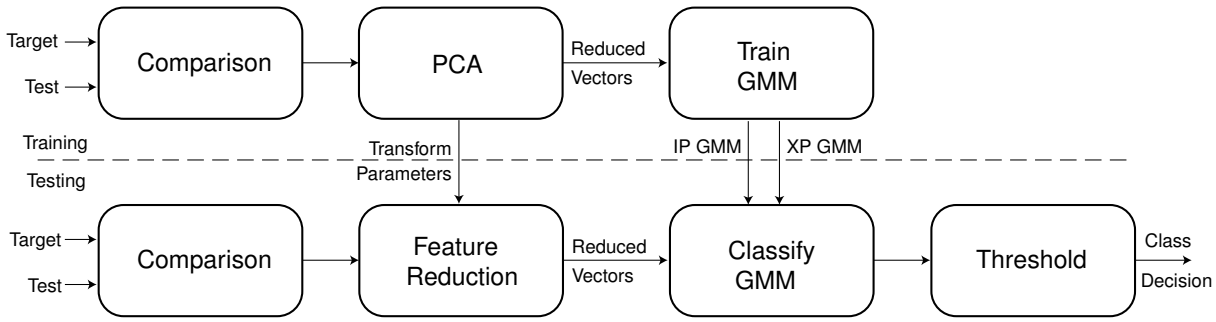

Fig. 1. Verification Methodology

by facial expression changes do not stretch or tear the facial surface.

They then combine the calculated canonical surface with intensity values and apply dimensionality reduction. A weighted Euclidean distance difference measures is then calculated in this reduced space. The results presented show that this algorithm is capable of distinguishing between identical twins however overall results across the database were not provided.

Rather than the transformation into a common coordinate system, such as the EGI, this work focusses on directly establishing correspondence between features on the facial surfaces and the statistical modelling of the differences in such features in both intra personal and extra personal comparisons.

This work presents a method for fully automated comparisons of human faces. A block diagram representing the methodology used in the verification process is shown in Figure 1. The problems encountered while establishing correspondence between two facial surfaces are identified and addressed in Section 2. A novel method for the comparison of these facial structures is then presented in Section 3. The experimentation and resulting performance of this method are detailed in Section 4.

2. 3D REGISTRATION

Any 3D comparison must start by establishing correspondence. The correspondence problem can be stated as finding pairs of features in two perspective views of a physical object such that each pair corresponds to the same point on the object.

There is still no single solution to the correspondence problem which works for all 3D data sets. A popular method, due to its generic nature and its ease of application, is the Iterative Closest Point algorithm (ICP) [14]. A block diagram representing the implementation of an ICP registration process along with the subsequent feature extraction is shown in Figure 2.

2.1. Iterative Closest Point

The ICP algorithm can be stated as follows. Given two point clouds of data, A and B, comprising, respectively, M and N points in $\Re^3$. ICP attempts to find a rotation, R, and translation, T, which minimises the average distance between corresponding closest points. At each iteration, for each point in A, x_i^A, $i \ni \{1...M\}$, the closest point, x_j^B, $j \ni \{1...N\}$, in set B is found along with the distance, d_N, between the two.

Robustness is increased by only using pairs of points whose distance are below a threshold, the Singular Value Decomposition of these points is then calculated and rotation/translation parameters are derived from this. Set B is rotated and translated accordingly and the process is repeated either until either the average error falls below a predetermined level or some maximum number of iterations is reached.

The ICP algorithm has a very generic nature which leads to problems with convergence when the initial misalignment of the data sets is large (typically over 15 degrees). The impact of this limitation in the ICP process upon facial registration can be countered through the use of pre-processing stages. Features such as nose and brow can be located and used to give a rough estimate of alignment from which we can be confident of convergence, this process is described in Section 2.2. The use of ICP in facial registration is a non-trivial application of the algorithm and the issues regarding this application are discussed in Section 2.3.

2.2. Pre-Processing

The first pre-processing step is to remove the extraneous surface regions that are not of great use in the verification process, such as the neck region. This is done using a K-means clustering approach, when there is insufficient spatial linkage between a cluster and the main point cloud it is removed from the data set.

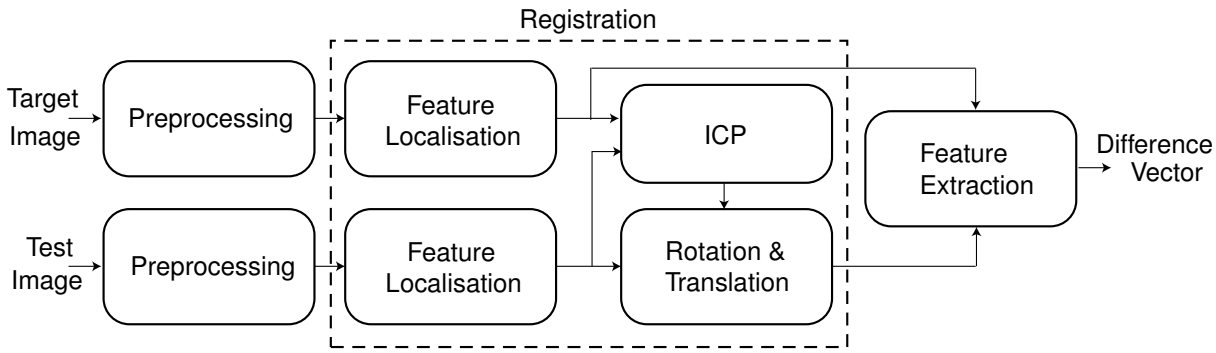

Fig. 2. Facial Comparison

After this, a plane is fitted to the data using a least-squares fitting method. The point cloud and corresponding plane are then rotated such that the major and minor axes of the data projected onto the plane correspond to the *y*- and *x*- axes respectively. Sensor noise is then detected at points of high curvature differential and such points are removed from the data set under consideration.

This alignment with the major axes greatly simplifies the task to locating the nose tip in the range image. An approximating B-Spline is used as an alternative surface representation from which contour lines such as that shown in Figure 3 can be easily extracted. As can be seen features such as the nose tip, nasal bridge and brows are prominent in this representation. These features are then used for coarse alignment and scale normalisation. After such processing,

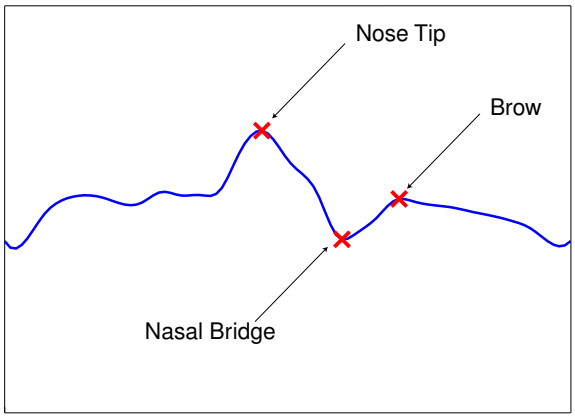

Fig. 3. Face Profile with labelled feature points

the face surface is now aligned such that the face is looking straight up out of the *x*-*y* plane and that the *y*-axis represents the vertical of the face. A final fine tuning process is then used to compensate for rotations of the face about this vertical axis. The nose tip identified in the previous step is used as a starting point and surface strips in both horizontal directions are extracted from the spline representation. These strips are compared using a cross correllation measure to determine the amount of symmetry present around the given nose tip location. By calculating this metric for a variety of rotations ranging from -10° to $+10^\circ$ an orientation which maximises the symmetry can be chosen. As can be seen in Figure 4, the peak is easily detected and this step can compensate for small rotational errors which aren't apparent in other representations. This pre-processing transforms any

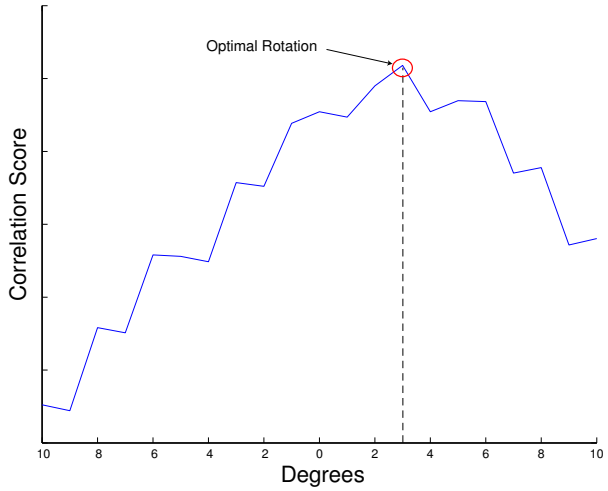

Fig. 4. Correlation Scores for various rotations about *y*-axis

given point cloud representation of a face into a common co-ordinate system. It does not, however, guarantee that two faces are directly comparable, due to the inherent limitation of common axis alignment. To confidently establish correspondence, the surfaces under comparison must be directly aligned with one another using an algorithm such as ICP.

2.3. Registration

The task of face comparison is a non-rigid registration problem due to the inherent elasticity present in human skin and the range of motion available to the human jaw. However, the problem can be simplified by considering only a small section of the face and performing alignment of these subsections. The area chosen for the alignment was that surrounding the nose as it exhibits less severe distortion due to facial expression changes than the mouth region and can be located with a high degree of accuracy.

To ensure that sufficient surface features are available for the ICP algorithm a region extending from the nose tip to the bridge of the brow is considered, this region also encompasses an equivalent area symmetrically across the nose. A surface mesh of this region can be seen in Figure 5

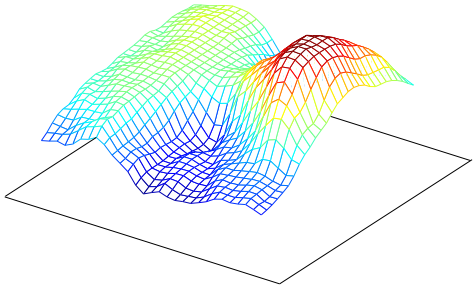

Fig. 5. Nose Region used for Registration

The data used in testing was aquired via SLS and consists of (x,y,z) co-ordinate information sampled along distinct scanlines crossing the face of interest [15]. Each scan line is uniformly spaced and parallel to the next, and the samples along the line are also uniformly spaced. However, the spacing between lines and the spacing between samples along a line is not equal which leads to a disparity between the sampling density along the orthogonal axes.

The scan lines along which data is captured have a data density of approximately twice that of a line running perpendicularly across the surface. This presents a problem for the ICP algorithm as the alignment of scanlines creates a minimum in the error landscape which is non-indicative of registration of the underlying surfaces from which the lines are sampled. To overcome this, surface fitting is employed on the data sets to give a more uniform sampling rate: first, the data is fitted using an approximating spline and then surface points are calculated across a regular grid.

The sampling grid used for a test subject is a slightly smaller subset of that used for the target; which allows for some lateral movement of the test set within the target. The results of registration can be seen in Figure 6. The displayed data has been considerably downsampled for ease of viewing.

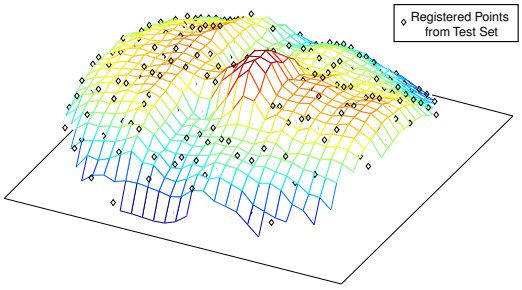

Fig. 6. Target Surface with Registered Points from Test Set

3. VERIFICATION

Currently the availability of 3D facial image databases is somewhat limited and there are no large databases readily accessible to the research community. The databases that are available (i.e. 3d_rma, xm2vts) have a relatively small number of observations for each subject, and hence it is not possible to reliably train statistical models, such as Gaussian Mixture Models (GMM), to individual subjects.

This can be overcome by restating the verification problem in the classical pattern recognition two class form; instead of training a classifier to recognise individuals it is trained to recognise a comparison observation as belonging to either the intra-personal or the extra-personal comparison class. [16].

The features which are used to compare two faces and how they are obtained from the registered images are described in Section 3.1. The process used to reduce the size of the feature space in order to obtain more accurate modelling is explained in Section 3.2 and the final classification of observations using GMMs is outlined in Section 3.3.

3.1. Feature Extraction

After the range images are aligned, it is possible to extract corresponding features from the surfaces that can be compared directly. Features such as surface depth and curvature which are extracted from point locations upon the surface are all directly comparable as the surfaces are coincident. By building a vector of differences across such features it is possible to create a comparison feature vector which comprises information on the *differences* between surfaces.

In Section 2.1, the non-rigid nature of the human face was discussed along with a method for partially circumventing this problem. For the purpose of verification a subsection of the face similar to that used in the registration pro-

cess was re-calculated for both faces; a square grid is overlaid across a region surrounding the nose the size of which is calculated using the distance between the nose tip and the nasal bridge.

The initial pre-processing stages have oriented the face such that the nose is aligned with the *z*-axis. The surface depth in this direction is re-estimated from both the target and testing image for each node on the grid which yields a 36x36 grid of depth estimates. These grids are vectorised to form a 1296 length vector and the vector difference between the two is calculated and used as the feature vector in the comparison stage.

These feature vectors are calculated for a training set of range images for both the intra- and extra-personal cases (IP and XP). However the large number of feature dimensions requires a correspondingly large number of training observations for accurate model estimation. Given the limited nature of the data available for training purposes, some form of dimensionality reduction is required to overcome this.

3.2. Dimensionality Reduction

Principal Component Analysis (PCA) [17] is a commonly used technique for dimensionality reduction. The technique works by finding an alternate set of orthonormal basis vectors which best represent the data set. It is possible to use a subset of the new basis vectors to represent the same data with a minimal reconstruction error. The transformation, C, to the new feature space can be found by minimising Equation (1).

$$tr\left(C\Sigma_Y C^{'}\right) \tag{1}$$

Where Y is the combined training data set (both IP and XP cases). For example, in Figure 7 an illustrating data set is represented in the traditional *x*-*y* plane. However it can be seen that by discarding the secondary component (*y'*) and retaining only the primary component (*x'*), less information is lost than would be by discarding either of the original dimensions. This concept can be extended to an arbitary number of dimensions and through experimentation it has been found that 50 dimensions are capable of adequately modelling our 1296 dimension feature vectors to a reasonable degree of accuracy. In this manner it is possible to maintain the information content of the feature space while projecting into a space with a dimensionality which is more applicable to standard statistical modelling.

3.3. Classification

After reduction the data is split back into IP and XP data sets, X_{ip} and X_{xp}; these sets are then used to train a Gaussian Mixture Model (GMM) for both the IP and XP classes.

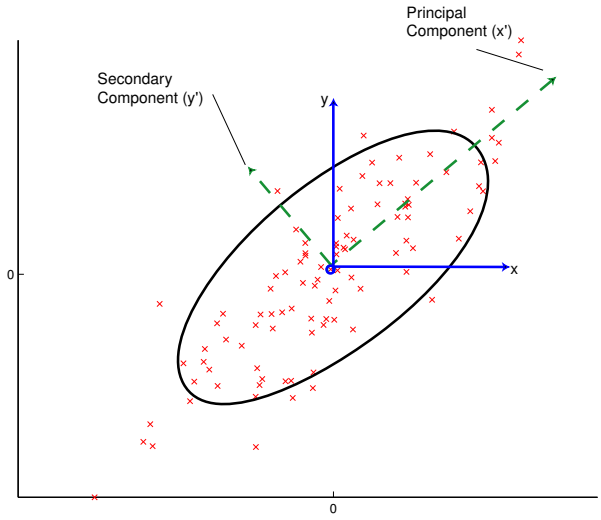

Fig. 7. Illustration of Principal Component Analysis

GMM training is achieved through the use of the Expectation Maximisation (EM) algorithm [18].

From these models the likelihoods $p(\vec{x}|IP), p(\vec{x}|XP)$ of a test observation, $\vec{x}$, can be obtained. These likelihoods are combined together to calculate the log-likelihood ratio of the intra-personal case by using Equation (2). The log-likelihoods are used in favour of standard likelihoods due to the ease with which they can be combined. The calculation of the ratio can be reduced to a simple differencing operation and multiple observations can be combined by summation as is shown in Figure 8.

$$\begin{aligned} llh(\vec{x}) &= \log\left(\frac{p(\vec{x}|IP)}{p(\vec{x}|XP)}\right) \\ &= \log(p(\vec{x}|IP)) - \log(p(\vec{x}|XP)) \end{aligned} \tag{2}$$

Where IP and XP are the intra- and extra-personal classes. The likelihood scores are then accumulated over all target images in the database as shown in Figure 8. This accumulated log-likelihood score can then be simply tested using a thresholding method as given in Equation (3).

$$C = \begin{cases} IP & llh\,(\vec{x}) > \tau \\ XP & llh\,(\vec{x}) \leq \tau \end{cases} \tag{3}$$

4. RESULTS

Testing was conducted using the 3d_rma [15] database of human faces. The database comprises data collected across two sesssions from a total of 120 subjects, each session containing 3 scans of the subject. The database is organised into two categories based on the method used to segment captured images into corresponding scan lines. In the

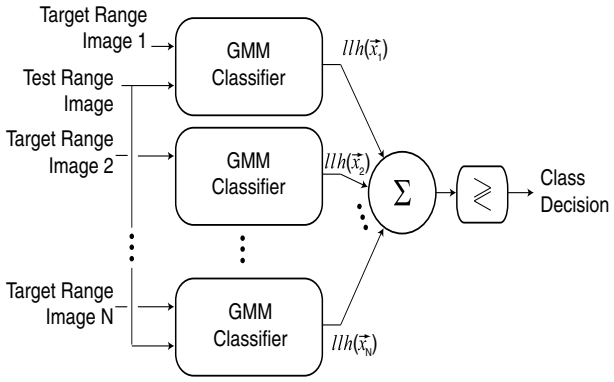

Fig. 8. Fusion of log-likelihood scores

"AUTO" set, an automatic algorithm was used for this task, however this introduces some erroneous values into many scans. The "MANUAL" set was constructed by manually refining a subset of the "AUTO" set. Experiments described in the following section were performed using the "MANUAL" subset.

The six observations available for each subject were labelled, with 1-3 corresponding to observations taken from the first session and 4-6 to those observations taken in the second session. The six observations were broken up into various combinations of training and testing sets in order to evaluate the proposed process, these are shown in Table 1 along with the corresponding Equal Error Rate and the rationale for each test case is described below.

Test Case	Training Set	Testing Set	Best EER
A	[1 2 4 6]	[3 5]	5.42%
B	[2 3 4 5]	[1 6]	6.25%
C	[1 2 3 5]	[4 6]	7.08%
D	[2 4 5 6]	[1 3]	5.42%
E	[1 2 3]	[4 5 6]	8.89%
F	[4 5 6]	[1 2 3]	8.89%
G	[1 3 5]	[2 4 6]	5.93%
H	[1 2 3 4 5]	[6]	3.33%
I	[2 3 4 5 6]	[1]	2.67%

Table 1. Test Cases used with 3d_rma

Firstly a 2-1 split was used to separate the database into training and testing sets. Four observations for each subject are used to generate the training observations and the remainder are used as the test set.

Each range image in the test set is compared against the range images from the same subject in the training set as well as against all training observations from 4 other random subjects from the database. This gives 1 comparisons of the "True Target" variety and 4 comparisons of the "Imposter Target" variety for each range image in the test set. Each comparison between two range images generates two observations due to the non-symmetric nature of the comparison algorithm. These two scores can be averaged together to improve the accuracy of the recognition process, alternatively they can be treated as independent scores in order to increase the number of test observations. In the 2-1 cases this resulted in 8 scores being generated for each comparison of the "True Target" variety.

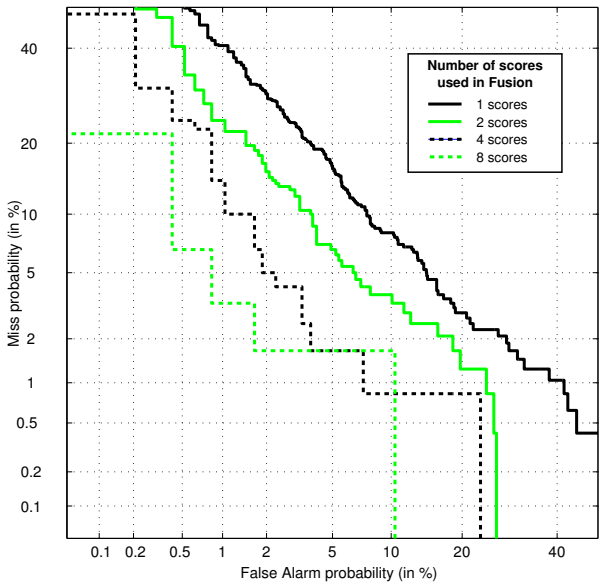

Fig. 9. Fusion of liklihood scores.

In Figure 9, increasing fusion of these scores is shown. By averaging the forward and backward comparisons 4 scores are obtained for each subject. By averaging across 2 database images 2 scores are obtained and by averaging over all 4 database images a single score is obtained. As the number of observations used in the fusion process is increased the error rates obtained from the DET plot correspondingly drop. This fusion, however, does not generate a sufficient number of scores to confidently construct the DET plot; this is evidenced by the blocky nature of both the 2 scores and 1 score plots. Hence all subsequent DET plots shown in this section have been constructed using only averaging of the forward and backward comparisons.

Testing conducted across test cases using the 4-2 split of data was broken down into two distinct sets. Test cases A and B have training samples drawn evenly from both sessions whereas C and D have the majority of training taken from a single session. In Figure 10 the combined scores from both scenarios are shown. The even distribution of training samples yields a better plot in the lower false alarm regions and has a slightly lower EER (6.25% compared to 6.46%).

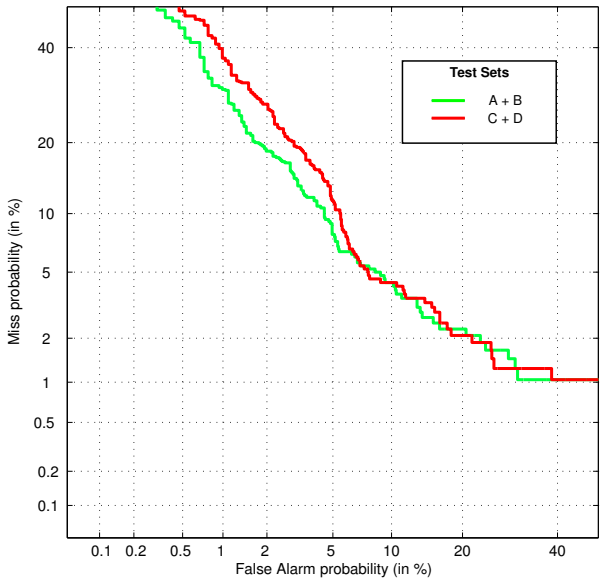

Fig. 10. DET Curves for 4-2 Split Test Sets

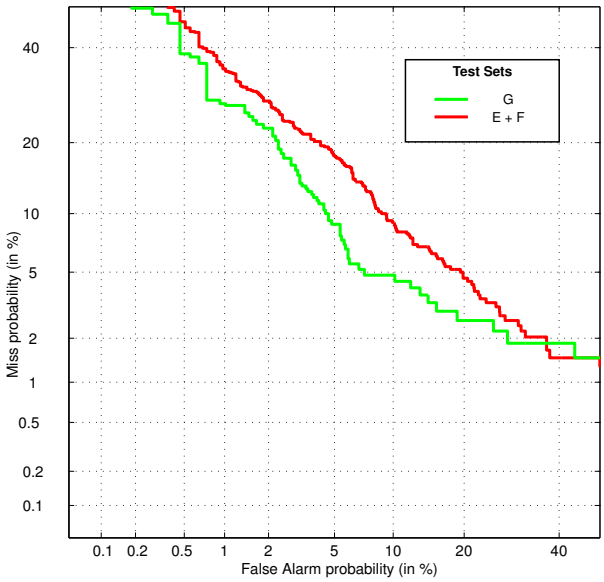

Fig. 11. DET Curves for 3-3 Split Test Sets

Secondly, an equal split of the data was used between training and testing sets, corresponding to test cases E through G. The DET curves for these test cases can be seen in Figure 11. Once again, the performance of the system when the training and testing sets have mismatched conditions (test cases E & F), is significantly below that of the case where the training data is sampled from both sessions (test case G).

Finally, to indicate the efficiency of the technique given larger amounts of training data, a "leave-one-out" approach corresponding to test cases H and I, was also trialled. As is expected, this configuration has a significant improvement in accuracy over the previous cases, as is shown in Figure 12.

5. CONCLUSIONS

In this paper a novel method for 3D face recognition using Iterative Closest Point algorithm for registering rigid portions of the face and statistical modelling of registration error over a specified region for comparison is presented. The proposed algorithm has been tested using a variety of scenarios derived from the 3d_rma database and a best Equal Error Rate of 2.67% has been achieved.

6. ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research (ONR) under Grant Award No: N000140310663 and by the Australian Research Council through the Discovery Grants Scheme, Grant DP0452676, 2004-6.

7. REFERENCES

[1] J. Lee and E. Milios, "Matching range images of human faces," *International Conference on Computer Vision*, pp. 722–726, December 1990.

[2] Gaile G. Gordon, "Face recognition based on depth and curvature features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 808–810, June 1992.

[3] H. Tanaka and M. Ikeda, "Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition," *Intl. Conf. on Pattern Recognition*, pp. 638–642, 1996.

[4] H.T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation. Principal directions for curved object recognition," *Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 372–377, April 1998.

[5] W. Zhao and R. Chellappa, "SfS based view synthesis for robust face recognition," *Intl. Conf. on Automatic Face and Gesture Recognition*, 2000.

[6] W. Zhao and R. Chellappa, "3D model enhanced face recognition," *Intl. Conf. on Image Processing*, pp. 50–53, 2000.

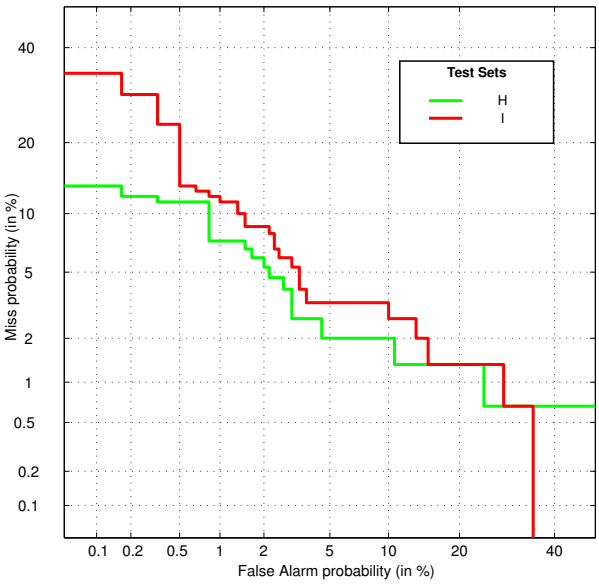

Fig. 12. DET Curves for 5-1 Split Test Sets

[7] Volker Blanz and Thomas Vetter, "Face recognition based on fiting a 3D morphable model," *IEEE Transactions on Pattern Analyssis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, September 2003.

[8] Volker Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model," *Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 192–197, May 2002.

[9] Jennifer Huang, Volker Blanz, and Bernd Heisele, "Face recognition with support vector machines and 3D head models," *Pattern Recognition with Support Vector Machines, First International Workshop, SVM 2002*, pp. 334–341, 2002.

[10] Gang Pan, Zhaohi Wu, and Yunhe Pan, "Automatic 3D face verification from range data," *International Conference on Acoustic, Speech & Signal Processing*, pp. 193–196, 2003.

[11] Y. Lee, K. Park, J. Shim, and T. Yi, "3D face recognition using statistical multiple features for the local depth information," *Intl. Conf. on Multimedia and Expo, ICME 03*, vol. 3, pp. 133–136, 2003.

[12] M. Bronstein A. Bronstein and R. Kimmel, "Expression-invariant 3D face recognition," in *AVBPA*, 2003.

[13] M. Bronstein A. Bronstein and R. Kimmel, "3D face recognition, us provisional patent no. 60 416243," 2002.

[14] P. J. Besl and N. D. McKay, "A method of registration of 3-D shapes," in *IEEE trans. on Pattern Analysis and Machine Intelligence*, Feb 1992, number 14 in 2, pp. 239–256.

[15] Charles Beumier and Marc Acheroy, "Automatic 3D face authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 315–321, 2000.

[16] C. Nastar B. Maghaddam and A. Pentland, "A bayesian similarity measure for direct image matching," in *Intl. Conf. on Pattern Recognition*, 1996.

[17] K. Fukunaga, *Introduction to Statistical Pattern Analysis*, Academic Press, 1972.

[18] N. Laird A. Dempster and D.Rubin, "Maximum likelihood from incomplete data via the EM algorithm.," in *Journal of the Royal Statistical Society*, vol. 39, no. 1 of *1–38*. 1997.