

# 3D FACE ACQUISITION, MODELLING AND RECOGNITION

*Jamie Cook, Jason Baker, Vinod Chandran and Sridha Sridharan*

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

## ABSTRACT

This paper presents a novel integrated 3D face verification system. The system comprises of a Structured Light Scanner acquisition stage, with 3D surface fitting and filtering. A-priori information of the human face is utilised to select and extract a region of interest from the face surface, whilst the Iterative Closest Point (ICP) algorithm is employed to establish correspondence between test and target and to compensate for the non-rigid nature of the surfaces. The verification process uses a Euclidean distance classifier across combinations of features extracted from the surfaces and results from a database of 21 subjects are presented.

## 1. INTRODUCTION

There is a growing demand for biometric identification and verification systems that are simple and secure. Features such as voice pattern, finger print, DNA, or face have been proposed and are currently being researched. Due to the ease with which images of a person's face may be acquired and the large degree of correlation to human perception, face is a preferred modality in many applications.

Traditional 2D Face Recognition Technologies (FRT) have struggled to cope with variations in lighting and pose. Three dimensional facial data can be used to overcome 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 focused 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 expression, 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.

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.

Methods for comparing faces using pure 3D data have also been proposed. Pan [10] proposed using registration error of low resolution Structured Light Scanning (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 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 measure 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.

This work presents a method for automated comparisons of human faces. An introduction to the 3D capture system and the problems encountered while establishing correspondence between two facial surfaces are identified and addressed in Section 2. The comparison of these facial structures is then discussed in Section 3. The experimentation and resulting performance of this algorithm are detailed in Section 4.

## 2. 3D FACE MODELLING

The following section introduces the techniques and challenges of modelling a face in 3D. The generic process of reconstructing a 3D representation of an object is extended to incorporate prior information of the human facial structure to provide a canonical representation of the face.

### 2.1. Face Acquisition

The acquisition of 3D models can be achieved using a variety of techniques with differing results. 3D acquisition techniques are categorised into either active energy systems or passive energy systems.

Active energy systems project an energy source onto the scene being captured and measure the disturbance for model construction. These systems include laser range finders, structured light systems and touch probes.

Passive energy systems determine 3D structure using the incident energy reflected by the scene. Passive energy systems include the Structure from X series of techniques, where X is stereo, motion, silhouette, etc.

Each type of system introduces a trade off between the complexity of the capture hardware system and the complexity of the post processing software required to generate the 3D models. In general a more complex hardware capture system has a simplified 3D reconstruction process.

A structured light 3D scanning (SLS) system was selected for use in this research because of its high density of 3D data and low cost. The SLS system includes a coded light stripe pattern projector and a high resolution colour camera.

The principle behind the SLS reconstruction is that a sequence of binary light stripe patterns are projected onto the face. A simple weighted summation of the light patterns provides a series of unique planes intersecting the face commonly called a stripe image. Each pixel in the stripe image is interpreted as rays with which the stripe value defines the location of the plane. Given the parameters relating to the

ray and plane, the 3D point can be determined by the intersection of the plane and the ray.

This approach requires the system to be calibrated, this involves determining the perspective transformation matrices (PTM) between the world coordinate system and both the camera and structured light coordinate systems. A detailed explanation of calibration and reconstruction is given by McIvor and Valkenburg [14]. Calibration is achieved using a cube with easily distinguished fiducial markings at known geometric locations. A series of a 108 such fiducial markings is used to estimate the PTM using the Least Squares Method.

The process of capturing the SLS image scans requires several successive images of the object, with varying light patterns. The subtle movement present in natural human posture has the effect of introducing a degree of noise into the final point description. The data noise can be minimised by fitting a surface model and filtering the data against the fitted model. The following section describes the process of surface filtering and noise reduction.

### 2.2. Face Modelling

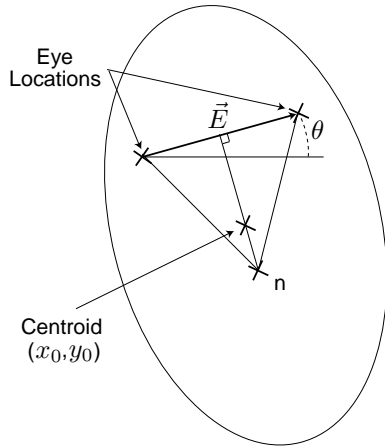
The output of the SLS system is a 3D point cloud data set, where each pixel of an image is represented as a 3D point. This generates a large number of superfluous 3D points. Given the 2D nature of the initial images, techniques commonly used for face localisation can be applied to preprocess the input to the SLS system.

Face Localisation is a heavily researched field and there are many automated algorithms such as [15] [16] which can accurately localise facial features. In the initial implementation of the system face localisation is accomplished via manual localisation of the eye positions. Further extensions to the system shall include the incorporation of an automatic face localisation subsystem.

In order to minimise the number of redundant 3D points in the reconstruction, an elliptical region of interest mask approximating the face is generated. Firstly the vector  $\vec{E}$  is defined as the vector connecting the left eye to the right eye and a third point,  $n$ , approximating the nose location is calculated by constructing an equilateral triangle with the eye locations as illustrated in Figure 1. The general equation of an ellipse is given in Equation 1.

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1 \quad (1)$$

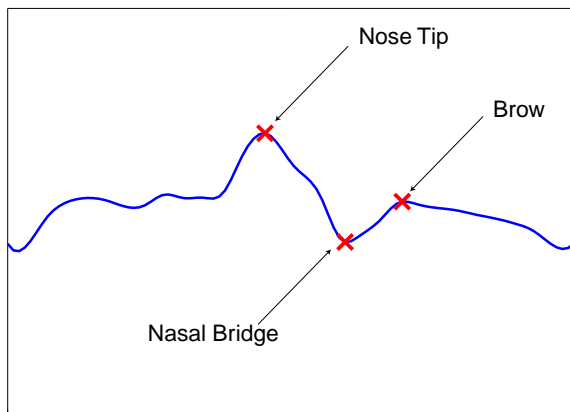
The centroid,  $(x_0, y_0)$ , of the ellipse is defined as a point lying 1/3rd the distance along the line that perpendicularly bisects  $\vec{E}$ . The parameters  $a$  and  $b$  are defined as ratios of the magnitude of  $\vec{E}$ , and the orientation,  $\theta$ , is defined as the angular distance of the vector above the horizontal. Only the image pixels within this region of interest are considered when generating the 3D point cloud data.



**Fig. 1.** Elliptical region of interest mask for face localisation.

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.

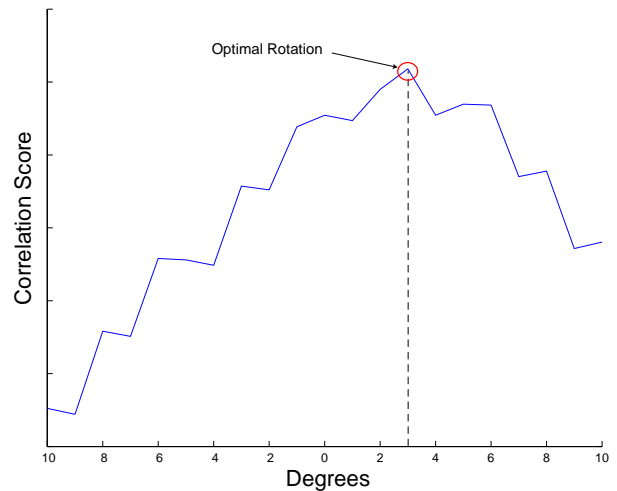
This alignment with the major axes greatly simplifies the task of locating the nose tip in the range image. An approximating B-Spline is used as an alternative surface representation from which contour lines such as that shown in Figure 2 can be easily extracted. As can be seen features such as the nose tip, nasal bridge and brows are prominent in this representation. These features are then used for coarse alignment and scale normalisation.



**Fig. 2.** Face Profile with labelled feature points

After such processing, 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 correlation 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  $-10^\circ$  to  $+10^\circ$  an orientation which maximises the symmetry can be chosen. As can be seen in Figure 3, the peak is easily detected and this step can compensate for small rotational errors which aren't apparent in other representations.



**Fig. 3.** Correlation Scores for various rotations about  $y$ -axis

This pre-processing transforms any 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.

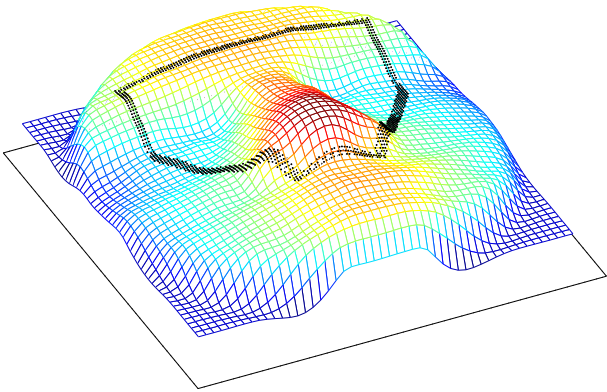
### 3. 3D FACE VERIFICATION

Verifying the identity of a claimant using 3D face information requires the comparison of a captured 3D model against a known 3D model of the claimed identity. Any 3D comparison task must first start by establishing correspondence between the 3D objects. 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.

### 3.1. 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 this subsection. The area chosen for the alignment was that surrounding the nose and eye region as they exhibit 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 registration, a region of interest is defined along the ridge of the nose and across the brow. The convex hull of this region is selected in order to encompass some of the cheek bone structure. The area of the region considered is proportional to the distance between the nose tip and nasal bridge. A down-sampled illustration of this region can be seen in Figure 4



**Fig. 4.** 3D filtered face model with highlighted region of interest.

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) [17].

The ICP algorithm can be stated as follows. Given two point clouds of data,  $A$  and  $B$ , comprising, respectively,  $M$  and  $N$  points in  $\mathbb{R}^3$ . 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  $A$ ,  $x_i^A$ ,  $i \in \{1 \dots M\}$ , the closest point,  $x_j^B$ ,  $j \in \{1 \dots N\}$ , 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. Section 2 discussed how 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.

### 3.2. Feature Extraction and Comparison

Given the aligned range images, 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.

The following features which were considered for the comparison of facial surfaces;

- distance between surfaces
- difference in surface curvature
- angular difference between surface normals

The distance between the surfaces is easily calculated based on the difference in range at a given  $(x, y)$  position. To calculate the difference in surface curvature the discrete Laplacian filter is applied to both the target and the test range image and the change in magnitude at each location is recorded. The angular divergence between surface normals  $A$  and  $B$  is calculated from  $\phi = \cos^{-1} \left( \frac{A \cdot B}{|A||B|} \right)$  and is also found at all points.

For the purpose of verification two overlapping regions of interest are considered. First is the subsection corresponding to that used in the registration process (ROI 1). The second region is specified using the greater part of the face and unlike the first region includes areas of the face which were not used directly in the registration process (ROI 2). In both cases the region is calculated for the first face and subsequently applied to both faces.

## 4. RESULTS

The proposed algorithm was tested using a data bases consisting of 21 subjects, of which 15 had at least 12 distinct scans. All images used for the scans were acquired using a 1.2 Mpixel colour camera and the structured light patterns were projected from a GRF SLS projector. Imposter

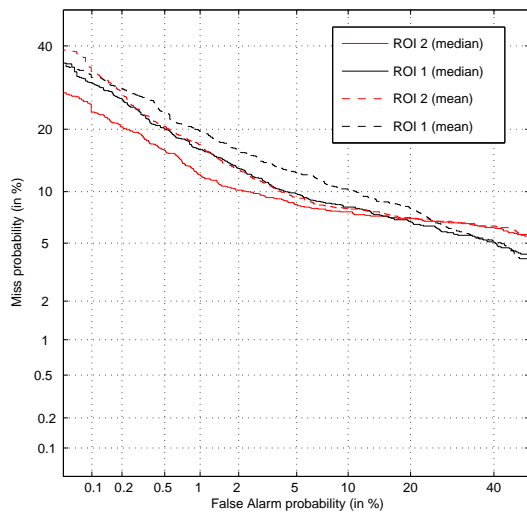
tests were generated by comparing all scans of each subject against all scans of 10 random subjects. For each subject  $n_i C_2$  target comparisons, where  $n_i$  is the number of scans taken for subject  $i$ , are conducted.

The features mentioned in the previous section were tested using both the mean and the median operator to provide averaging across the surface of interest. The use of the median operator provides a measure of robustness to outlier values, the resulting equal error rates for all testing scenarios can be found in Table 1.

Features	ROI 1		ROI 2	
	Mean	Median	Mean	Median
Surf. Depth	10.20%	8.55%	8.22%	7.86%
Curvature	18.92%	19.35%	17.35%	17.14%
Surf. Normals	12.61%	12.45%	9.64%	9.21%

**Table 1.** Equal Error Rates for Features

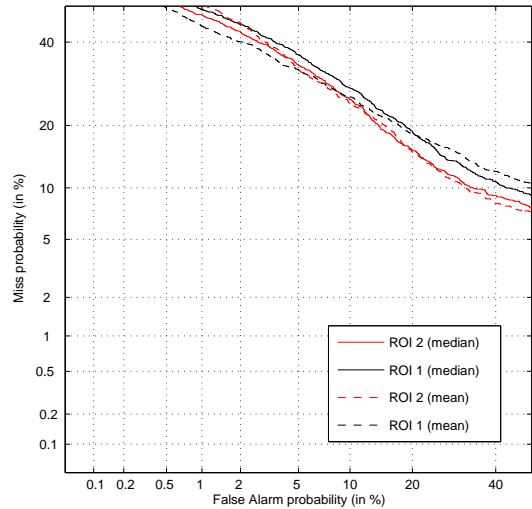
By simply using the distance between the surfaces it was possible to achieve an Equal Error Rate (EER) of (7.86%). The Detection Error Tradeoff (DET) plots for surface comparison are shown in Figure 5. From these it is evident that the use of the median operator as opposed to the mean operator for the averaging, yields a consistent performance increase.



**Fig. 5.** Physical distance between surfaces.

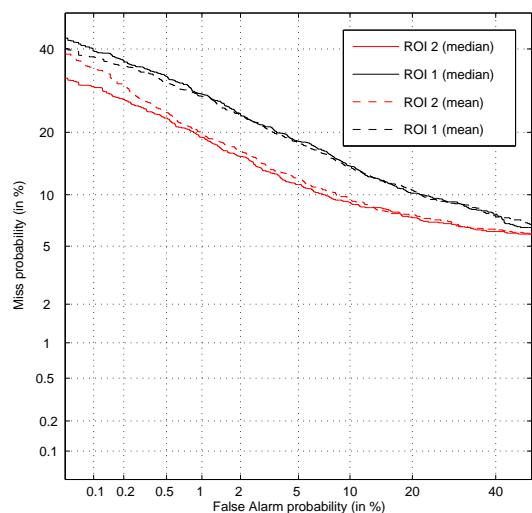
As might be expected the curvature scores failed to perform well with the given data sets. This can mainly be attributed to the lack of highly accurate 3D data, previous authors using curvature values [2], typically utilised laser range scans of the face. As can be seen in Figure 6 the EER achieved using only curvature values are significantly below

those achieved using other features. It is interesting to note however that unlike the other features, curvature obtains a best EER in ROI 1 when using the mean operator.



**Fig. 6.** Direct Comparison of Curvature.

The angular difference in surface normals gave results that were somewhat behind that of the surface difference. A best EER of 9.21% was achieved when using the median operator over the larger region of interest. Furthermore it should be noted that in all scenarios the use of the larger ROI generated a consistent improvement in recognition accuracy. This increase in discriminability can be directly related to the incorporation of surface points which were not explicitly used to establish correspondence.



**Fig. 7.** Angular Divergence of Surface Normals.

## 5. CONCLUSION

This paper has illustrated a complete approach to the face verification process. The face surface is successfully acquired using Structured Light Scanning techniques with appropriate surface filtering. The region of interest is extracted from the 3D model from which surface properties are measured and utilised in the verification process.

Experimentation has been presented illustrating the performance of 3 different surface features across varied regions of interest. The results presented have shown the validity of using face surface information to facilitate the pose and lighting invariance required for robust face verification.

Future improvements to the system include automated eye location extraction, a passive face acquisition system, and an extended database of 3D face models.

## Acknowledgements

This research was supported by the Office of Naval Research (ONR), USA, under Grant Award No: N000140310663 and by the Australian Research Council (ARC) through Discovery Grant Scheme, Project ID DP452676, 2004-6.

## 6. REFERENCES

- [1] J. Lee and E. Miliou, "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.
- [7] Volker Blanz and Thomas Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis 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] R.J. Valkenburg A.M. McIvor, "Calibrating a structured light system," Tech. Rep. 362, Industrial Research Limited, Auckland, New Zealand, 1995.
- [15] S. Baluja H. A. Rowley Sawashima and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, Jan 1998.
- [16] D. Kriegman M. Yang and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, Jan 2002.
- [17] 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.