

3D Surface Reconstruction and Recognition

Daniel J. Bardsley, Li Bai
University of Nottingham Face Recognition Group,
Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, England
djb@cs.nott.ac.uk, bai@cs.nott.ac.uk
+44 115 951 3312

ABSTRACT

In this paper we propose a novel 3D face recognition system. Furthermore we propose and discuss the development of a 3D reconstruction system designed specifically for the purpose of face recognition. The reconstruction subsystem utilises a capture rig comprising of six cameras to obtain two independent stereo pairs of the subject face during a structured light projection with the remaining two cameras obtaining texture data under normal lighting conditions. Whilst the most common approaches to 3D reconstruction use least square comparison of image intensity values, our system achieves dense point matching using Gabor Wavelets as the primary correspondence measure. The matching process is aided by Voronoi segmentation of the input images using strong confidence correlations as Voronoi seeds. Additional matches are then propagated outwards from the initial seed matches to produce a dense point cloud and surface model. Within the recognition subsystem models are first registered to a generic head model, and then an ICP variant is applied between the recognition subject and each model in the comparison database, using the average point-to-plane error as the recognition metric. Our system takes full advantage of the additional information obtained from the shape and structure of the face, thus combating some of the inherent weaknesses of traditional 2D methods such as pose and illumination variations. This novel reconstruction / recognition process shows 98.2% accuracy on databases containing in excess of 175 meshes.

Keywords: 3D reconstruction, face recognition, Gabor wavelet, ICP

1. INTRODUCTION

For the past decade the majority of face recognition research has been focused on recognition from single frame, frontal view, 2D face images of the subject. Whilst there has been significant success in this area using techniques such as eigenfaces [1-3] and elastic bunch graph matching [4, 5] several issues look set to remain unsolved by such approaches. These issues include the current set of algorithms inability to robustly deal with large changes in head pose and illumination. As such an algorithm which displays properties invariant to each of these recognition issues would be of significant use. Recently, a growing body of research has focused on obtaining accurate 3D data of face surfaces with a view to use such information directly for recognition. Obtaining accurate 3D data would allow direct comparison between the shape of each subject face, thus eliminating errors associated with changes in illumination. Furthermore, the availability of true 3D data allows comparisons between models from arbitrary views, thus making such a solution far more pose invariant than current 2D solutions. Obviously the technical challenges associated with obtaining a 3D model of a face are far greater than those involved in capturing a 2D image and as such, significant improvements in recognition rates will only be achieved given a sufficiently accurate 3D capture method. Our work outlines the development of a reconstruction system designed specifically for the purpose of 3D face recognition. Since the reconstruction system has been designed with recognition in mind from the start, various assumptions about the nature of the object being reconstructed allow more accurate face models to be produced as apposed to a more generic, general purpose reconstruction system.

Following the successful reconstruction of a subjects face comparisons must be carried out between the new model and faces already present in the system database. Ideally the algorithm must perform recognition quickly, however, accuracy should be considered a higher priority than speed in a robust identification system. Prior to recognition each model must be aligned with all other models in the database. Our system

carries out this registration phase by aligning each model with a generic head immediately after reconstruction. Registration to a base model provides an approximate alignment between each of the head models to provide the ICP algorithm with an initial estimate of the transformation required to minimise the alignment error between each of the database models. Recognition is carried out by minimizing the ICP point-to-plane alignment error between the subject model and each model already stored in the database. The average point-to-plane inter-model error is then used as the recognition metric.

2. LITERATURE REVIEW

The topics of 3D reconstruction and face recognition encompass a huge body of work amounting to decades worth of research. Firstly we summarise the most important and relevant work in the field and secondly propose how our research extends some of these ideas and methodologies. A more comprehensive analysis of state-of-the-art reconstruction and recognition techniques can be found in [6], however, we summarise only the publications most applicable to our work here.

Our work implements a new algorithm for stereo vision. This involves capturing simultaneous images of a subject from two or more different angles. Range data is calculated by using the known geometry of the camera rig and solving a correspondence problem between the captured images. Stereo matching has been well studied within both mathematics and computer science. Hartley and Zisserman [7] provide an in-depth explanation of the subject in *Multiple View Geometry for Computer Vision*.

Uchida, Shibahara and Aoki [8] propose a stereo vision based method for reconstructing face models for recognition using a phase-based correspondence metric. They also utilise a rig setup incorporating 4 cameras with a small inter-pair baseline to minimise the differences between stereo images and thus ease the correlation process. Additionally they employ a course-to-fine matching strategy using image pyramids and a sub-pixel window alignment technique. The models are registered and recognised using ICP, however, the small size of the recognition database suggests the method may not be accurate enough to scale to large subject databases. We propose in this paper that reconstruction accuracy could be improved through the use of projected random light patterns and wavelet based matching in order to create a more robust correlation between stereo pairs.

An alternative to the stereo vision approach to reconstruction is to create a 3D morphable model. A method for automatically generating such a morphable model and performing recognition is analysed in [9]. The morphable model is learned from a set of 100 male and 100 female head scans such that its vector space representation of the face can represent any combination of shape and texture vectors that would also describe a realistic human face. The aim of the morphable model technique is to deduce model coefficients that best represent novel faces presented to the system. Model based recognition can then be achieved either by comparison of the model coefficients or by generating synthetic images of the model under different pose and illumination conditions and using a standard 2D recognition technique. The morphable model method has the advantage of allowing models to be constructed from a single image of a subject and thus allows for a relatively passive method of reconstruction without the need for special capture rigs or equipment. Hu, Jiang, Yan, Zhang and Zhang [10] apply the morphable model approach to recognition in their recent paper, implementing an analysis-by-synthesis framework for reconstruction and recognition. An attention worthy aspect of their work is that in addition to synthesising faces under variant pose and illumination, different expressions are also generated in 3D space using a MPEG-4 based facial animation technique, thus attempting to add a degree of expression invariance to the recognition system.

In recent research Bronstein, Bronstein and Kimmel [11-14] develop an effective expression invariant recognition system. They first reconstruct head models using a structured light 3D scanner, however, the novel aspect of their work involves the representation of faces as isometries. They show that faces represented as geodesic distances on facial surfaces were significantly less sensitive to facial expressions compared to Euclidean representations. Their most recent work embeds the faces geometric structure into a spherical representation which they show by experiment is isometry-invariant. They refer to these new invariants as spherical canonical images. A similarity measure between the spherical canonical images based

on the harmonic transform is also introduced to allow direct comparison of their facial representations. The accuracy of this method proved to be sufficient to successfully identify twins.

In this paper we extend the work of the stereo vision based reconstruction systems through the investigation of the Gabor Wavelet as a stereo correspondence metric and a matching strategy suited to facial reconstruction. We also aim to have comparable model accuracy and recognition rates to the morphable model methods, despite not being able to achieve reconstructions from singular frontal images. Whilst not formally discussed in the current paper, in future work we hope to extend our recognition system to incorporate expression invariance either via a model based animation technique or through the use of an explicitly expression invariant facial representation.

3. RECONSTRUCTION METHODOLOGY

This section defines the precise methodology used to reconstruct the recognition subject faces. The reconstruction rig used for collecting 3D data consists of six cameras, of which four are black and white and two are colour. A projector emits a random light pattern onto the subject face as the four black and white cameras capture a single frame. Immediately after the projector is turned off the colour cameras capture a single frame of the subject face. The black and white input is used for point matching across the images where as the colour cameras are used to obtain texture information. Total capture time is less than 1.5 milliseconds thus ensuring that there is minimal chance for the subject to move during the two camera capture phases. The capture rig was supplied by 3DMD [15], however, we use the raw camera data directly for our reconstructions. Figure 1 shows the raw input images of a single subject during the capture process.



Figure 1: Raw input images

3.1. Rig Calibration

Prior to performing reconstruction the rig is calibrated using a standard calibration object. A single image frame of the calibration object is captured simultaneously in all cameras. The initial estimate for each camera projection matrix is calculated using the normalised Direct Linear Transform (DLT) algorithm. This initial estimate is refined by minimising the geometric error present when calibration points are back projected into each of the calibration images.

The initial estimate of the projection matrix for each camera is calculated by utilising each correspondence between 2D image space coordinates (corners on the calibration object) and 3D world space coordinates. For each correspondence $X_i \leftrightarrow x_i$ we derive the following relationship:

$$\begin{bmatrix} 0^T & -w_i X_i^T & y_i X_i^T \\ w_i X_i^T & 0^T & -x_i X_i^T \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \quad (\text{Eq. 1})$$

From a set of n point correspondences, we obtain a $2n \times 12$ matrix A by stacking equations for each correspondence. The projection matrix is calculated by solving the set of equations $Ap = 0$. This solution can be further refined by assuming that the world points in our calibration are accurately known and minimising the geometric error present within the initial estimate of P . We may define the geometric error as:

$$\sum d(x_i, \hat{x}_i)^2 \quad (\text{Eq. 2})$$

Where x_i is the re-projected point and \hat{x}_i is the exact projection of the world point. Thus the solution to the following minimisation is the maximum likelihood estimate of P .

$$\min_p \sum_i d(x_i, PX_i)^2 \quad (\text{Eq. 3})$$

Minimising the geometric error requires the use of iterative techniques. This increases the computation time but as the calculation only occurs during calibration it is an acceptable loss in performance. The Levenberg-Marquardt minimisation technique is suitable for calculating the initial DLT estimate for P which can then be used as an initial parameterisation for calculating the maximum likelihood of the projection matrix. When used in conjunction with data normalisation and the DLT this calibration method is known as the Gold Standard algorithm for estimating P . The full details of this method are detailed by Hartley and Zisserman in Multiple View Geometry for Computer Vision [7] with the complete algorithm referenced below.

Objective:

Given $n \geq 6$ world to image point correspondences $X_i \leftrightarrow x_i$, determine the Maximum Likelihood estimate of the camera projection matrix P .

Algorithm:

Linear Solution: Compute an initial estimate of P using a linear method.

Normalisation: Use a similarity transformation T to normalise the image points, and a second similarity transform U to normalise the world space coordinates. T should be such that the RMS distance from the origin is $\sqrt{2}$ and U such that the RMS to origin is $\sqrt{3}$

DLT: Form the $2n \times 12$ matrix A as generated by stacking equation 1 for each 2D to 3D correspondence. A solution to $Ap=0$, subject to $\|p\|=1$, is obtained from the unit singular vector of A corresponding to the smallest singular value.

Minimise Geometric error: Using the linear estimate as a starting point and minimise the geometric error:

$$\min_p \sum_i d(\tilde{x}_i, \tilde{P}\tilde{X}_i)^2$$

Over \tilde{P} , using an iterative algorithm such as Levenberg-Marquardt.

Denormalisation: The camera matrix for the original (un-normalised) coordinates is obtained from \tilde{P} as:

$$P = T^{-1} \tilde{P} U$$

Using the normalised DLT algorithm we obtain a camera matrix, P , for each of the cameras in our stereo rig. Given that all the cameras are calibrated simultaneously each camera matrix will project stereo matches into the same world coordinate system. Assuming sufficient accuracy in the location of calibration points, there is no requirement to align range data taken from different stereo pairs during reconstruction. It may be beneficial at this stage to perform bundle adjustment on each of the camera matrices to minimise the projection error

between the independent stereo pairs, however, we found reconstruction to be sufficiently accurate without a bundle adjustment stage.

3.2. Gabor Jet Correspondence Metric

A novel aspect of our work involves the use of Gabor Jets as a correspondence metric. Gabor filters have been employed extensively within the face recognition field [16-18], however, our work utilises many of the well studied aspects of Gabor filters to judge correspondence between stereo pairs. The aspects of the Gabor filter which are of interest to us include their invariance to lighting conditions and small perspective changes. Pötzsch, Krüger and Malsburg [19] show Gabor jets to be robust against exactly this class of distortions making them an ideal candidate for stereo correspondence problems.

A Gabor jet is a condense and robust representation of a local grey value distribution. It is based on a Gabor wavelet transform, which is a convolution with a family of complex Gabor wavelets having the shape of plane waves restricted by a Gaussian envelope function. The wavelets are similar in the sense that they can all be generated from a mother wavelet by rotation and scaling. All complex coefficients of the transform taken at one image location form a jet. Useful properties of the Gabor filter include invariance to changes in lighting conditions and to small perspective changes. These properties are ideally suited to the stereo correspondence problem since changes in lighting and perspective are always present between images taken from different cameras aimed at the same subject.

Daugman [20] generalised the 2D Gabor function to the following form:

$$G(x, y) = \frac{1}{2\pi\sigma\beta} e^{-\pi \left[\frac{(x-x_0)^2}{\sigma^2} + \frac{(y-y_0)^2}{\beta^2} \right]} e^{i[\xi_0 x + \nu_0 y]} \quad (\text{Eq. 4})$$

Where, (x_0, y_0) is the centre of the receptive field in the spatial domain and (ξ_0, ν_0) is the optimal spatial frequency of the filter in the frequency domain. σ and β are the standard deviations of the elliptical Gaussian along x and y .

In order to perform analysis of a particular image region a family of Gabor wavelets is derived from a mother wavelet. Each of these derived filters is then convolved with the image, with the response of each filter being combined into a vector representing all of the filters. This vector of Gabor filter responses is known as a Gabor Jet. Comparisons between different Gabor jets allow a measure of similarity between the image regions to be computed. Equation 5 defines the jet similarity functions for two images (J and J'):

$$S_a(J, J') = \frac{\sum_{j=1}^{G_f} a_j a'_j}{\sum_{j=1}^{G_f} a_j^2 \sum_{j=1}^{G_f} a'^2_j} \quad (\text{Eq. 5})$$

Where $a_j, j=1, \dots, G_f$ is the magnitude of the result of the convolution between the real and imaginary part of the Gabor Filter, j , and the image.

In the described stereo vision system the initial seed points in the reference image are matched to pixels in the corresponding image first by obtaining the Gabor jet for filters centred on the reference seed pixel, this jet is then compared with the jet corresponding to each pixel on the corresponding epipolar line. The pixel with the highest similarity is then selected as a match.

3.3. Voronoi based propagation matching

Whilst attempting to correlate feature points between images in a stereo pair, various factors such as image noise, occlusion or illumination differences can lead to incorrect matches regardless of which correlation algorithm is used. For this reason it is necessary to constrain the matching process as far as possible using knowledge of the nature of the surface we are attempting to reconstruct. Common matching constraints include: similarity threshold, uniqueness, continuity, ordering, epipolar and relaxation. In order to constrain the way in which the correlation algorithm searches for an appropriate match a search strategy is required. An efficient search strategy will increase the accuracy of a correlation algorithm by reducing the potential search space, whilst usually decreasing the overall search time by requiring fewer comparisons per feature point. An efficient matching strategy is described below, which increases both accuracy and speed within the reconstruction system.

The proposed matching strategy is based on the Voronoi propagation method described by Li Tang, Tsui and Wu [21]. A number of modifications to their original design have been made in order to produce a more robust strategy. These differences include the use of Gabor Jets rather than SSD (Sum of Squares Difference) as the similarity metric and the application of the epipolar constraint during matching. We show the superiority of Gabor Jets as a correlation metric in Section 4. Furthermore Tang, Tsui and Wu opted to withdraw the epipolar constraint in order to allow their matching strategy to function on un-calibrated image pairs, however, since we have full calibration data available, it is trivial to reduce the matching search space by utilising our knowledge of the Fundamental matrix.

Initially N seed points are selected in the source image. These seed points should, ideally, be the most salient feature points in the input image since errors at this stage will produce catastrophic results later in the matching process. The candidate seed points are selected by finding corners with large eigenvalues within each source image. These candidate seed points are then matched to their corresponding locations in the image pair. Since it is imperative at this stage to correctly match the seed points, the Gabor correlation algorithm is used and performs a full epipolar line search for each of the seed points.

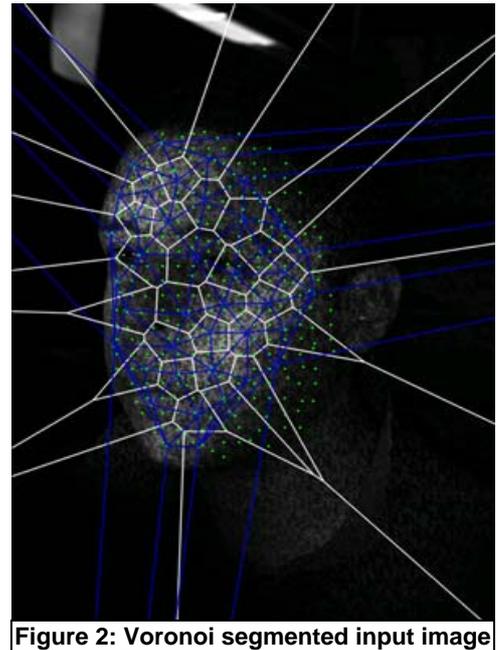


Figure 2: Voronoi segmented input image

Once the seed points have been selected and matched the Voronoi diagram of the original seed points is calculated. The Voronoi diagram of a collection of feature points is a partition of an image space into cells, each of which consists of those image points which are closer to one feature point than to any other. Voronoi diagrams are involved in situations where a space needs to be partitioned into “regions of influence”. Once the Voronoi diagram has been calculated, matches are propagated from the seed points towards boundaries of the Voronoi cells until all of the matched regions are merged together. Matches are still confined to the appropriate epipolar line, however, the search space is constrained to within a small range of disparities around the seed cell disparity, thus within each cell, matches form a smooth surface. Figure 2 shows an example source image with the corresponding Voronoi segmentation of the seed points.

This method of propagation inherently enforces a continuity constraint into the matching process. This makes the assumption that object surfaces will be smooth and continuous. This assumption is not always valid for real world objects and will certainly break down at large discontinuities in the image, however, it is a suitable constraint given the advantages in speed that can be obtained through its use. Furthermore, additional processing steps could be employed and the constraint dynamically withdrawn at image locations where it does not hold true. Propagation provides a convenient method of producing dense correlation maps whilst

also reducing the computational cost of the matching process. The reduction in computation stems from the fact that once the match for the initial seed point has been calculated the search for points within the same cell can be guided by the relative position of the matched seed point. This reduces the search space by an order of magnitude from a full epipolar line search to a small localised area.

Matches propagate outwards from the initial seed points in each cell in a standard breadth first search pattern. As a match for each pixel is found its neighbours are then added to the queue of pixels waiting to be matched. Pixels with high match strengths are used to produce an initial estimate for the position of neighbouring pixels, resulting in a smooth surface whilst only requiring a small number of comparisons between candidate matches. The algorithm cycles until every pixel within the given Voronoi cell has been matched to its corresponding point. The entire process is then repeated for each initial seed point until a dense disparity map has been produced and each of the matches can be projected into the required 3D world coordinates. The Voronoi propagation method proves suitable for facial reconstruction since propagation performs best in situations where no large discontinuities are present. As each side of the face is dealt with by an independent stereo pair no major discontinuities are encountered thus matching via propagation is an ideal method. The validity of these claims is demonstrated by the reconstruction results.

4. RECONSTRUCTION RESULTS

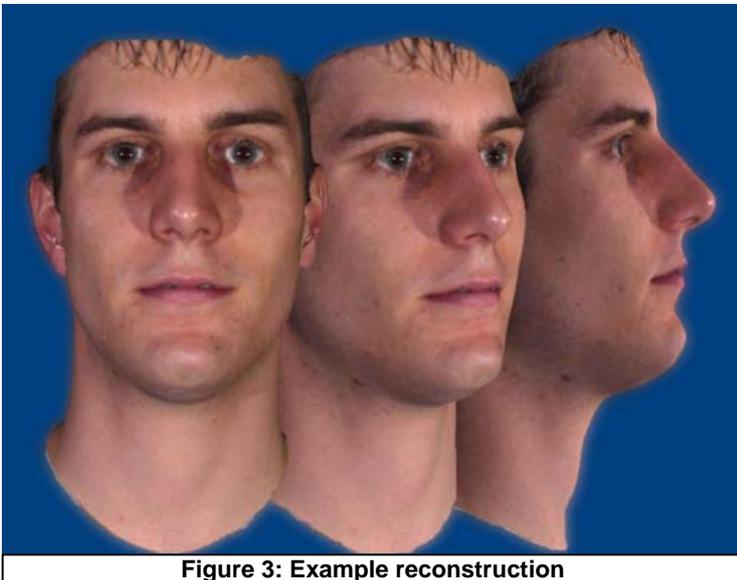


Figure 3: Example reconstruction

Using the methods outlined in the previous sections we are able to obtain high quality models of recognition subject faces. The design of the camera rig allows almost full coverage of the subject face making it highly suitable for recognition. Due to the nature of the structured light projection during the capture phase, hair is not reconstructed. This is of no concern to us since the exclusion of subject hair is not relevant to the recognition process. Figure 3 shows a single example of one of the reconstruction subjects from our recognition database.

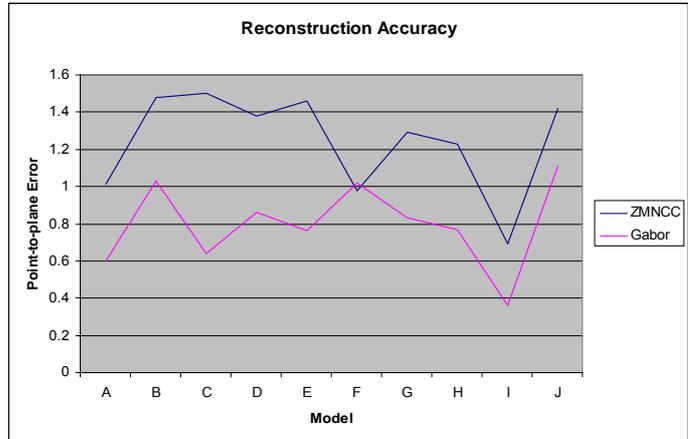
It should be noted that some portions of non-face objects are also reconstructed. Currently these portions of the model must be manually removed prior to recognition, however, basic skin detection algorithms would probably be sufficient to automatically

cut the unwanted parts of the model from the initial reconstruction.

In order to analyse the efficiency of the Gabor wavelet as a similarity metric we compare it to two other popular stereo matching algorithms. In all cases we use the Voronoi propagation method, with the same input parameters for each algorithm. We also use identical local window sizes and the same initial seed points for each model reconstructed. To demonstrate the effectiveness of the Gabor wavelet we compare the accuracy of the reconstructed models against models produced using both SSD and ZMNCC (Zero Mean Normalised Cross Correlation). In the absence of true measurements taken from each subjects face we use reconstructions produced by the 3DMD [15] scanner as the “ground truth” data. 3DMD claim accuracy to within 0.5 of a millimetre RMS in their reconstructions, therefore they are sufficiently accurate for our testing purposes. In order to measure the accuracy of each of the reconstructed models we use the same methodology as is used in the recognition phase of the system. All models are manually aligned by selecting the tip of the nose and the centre of the eyes in each of the reconstructions and then performing a least squares minimisation to register the models. We then use the average point-to-plane ICP error to determine the level of similarity between the models. Lower values are considered a better match than higher values.

Thus, in order to measure reconstruction accuracy we compare models produced via SSD, ZMNCC and Gabor wavelets with the “ground truth” 3DMD model and deduce the quality of the reconstructed model via the ICP point-to-plane error metric.

The graph on the right shows the error rates of reconstructions from 10 randomly selected models from our database. Reconstructions were carried out using Gabor, ZMNCC and SSD however the SSD reconstruction results are omitted as the reconstructions produced using this method were very poor. Upon closer analysis of the reasons behind the poor SSD results we find that illumination differences between each image in the stereo pair produces erroneous matches when SSD is used as the algorithm does not contain any mechanism to compensate for variations in illumination. Thus it becomes obvious that any algorithm capable of performing accurate matches on our input data displays a degree of illumination invariance. This claim is backed up by the ZMNCC results since this algorithm is specifically designed to compensate for constant illumination variations between input images and performs much better than SSD on our test data. This also demonstrates the illumination invariance of the Gabor transform and thus the differences between ZMNCC and Gabor correlation metrics are largely the result of each of the algorithms ability to compensate for the perspective variations between each of the images in a stereo pair. As can be seen from the reconstruction accuracy of the two analysed correlation metrics, Gabor outperforms ZMNCC in almost all cases.



5. RECOGNITION METHODOLOGY

Once the reconstruction phase is complete we begin the recognition process. This phase consists of first registering the newly scanned model to a generic head model. A generic model is used to ensure that each of the scans is aligned with a consistent model. For the highest accuracy alignment it would be best to use the average face of the database models as the generic head. Alignment to the generic head is carried out by manually selecting the tip of the nose and the centres of the eyes both on the subject scan and the generic head. A least-squares minimisation is then carried out to find the rotation and translation required to align the subject and the generic head. Figure 4 shows a screenshot of a scan (green points) registered to the generic model (blue points). Each model is registered to the generic model immediately after being scanned, therefore, each of the models in the database is approximately aligned to every other model. This initial registration phase is required since the ICP algorithm requires a reasonable initial estimate of the inter-model alignment to function correctly.

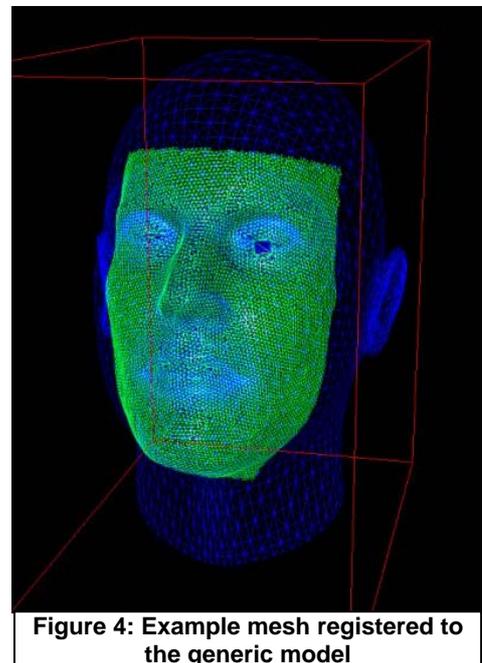


Figure 4: Example mesh registered to the generic model

The ICP algorithm is primarily used for aligning range data of an object where several scans of the object have been taken (for example from different views) and need to be stitched together to form the complete model. We apply the same technique here, however, face models from different subjects are aligned to each other using ICP and then the resultant alignment error between the models is used as a recognition metric. Many variants of the ICP algorithm exist and have been well studied. We pick a variant of the ICP algorithm

which has proved useful within production environments [22] and has been shown to be robust against scanned data containing many kinds of surface feature [23]. Specifically, the features of this variant include: random point sampling, matching selected points to the closest corresponding point with a normal within 45 degrees of the source normal, constant point weighting, rejection of edge vertices, the point-to-plane error metric and the “select-match-minimise” iteration method. In order to obtain an accurate match between different subjects we iterate the ICP algorithm 20 times between each of the models in the database. The similarity between models is measured by calculating the average point-to-plane error between each of the models after they have been aligned as closely as possible. Testing was carried out using both the RMS and point-to-point error as the recognition metric, however, this led to much lower recognition accuracy.

6. RECOGNITION RESULTS

The table below shows the overall accuracy of the recognition system along with the number of subjects, meshes and comparisons. The 5 unmatchable meshes in the database are not considered in the overall accuracy of the system since these subjects only had one head scan in the database and hence no correct match is possible. The subjects in the database were not asked to perform any specific expression during the capture stage, nor were they asked to look in any specific direction. This led to a database containing many different expressions and varied head poses. The accuracy of the results demonstrates the pose invariance of the system and, to a lesser degree, some expression invariance. All the incorrect matches are shown to be the result of large changes in subject expression, however, this was to be expected since we were not aiming to incorporate any explicit expression invariance within the system. Figure 5 shows an example of one of the incorrectly matched models, where variation in expression has caused errors in recognition.



Figure 5: Erroneously matched models

| nSubjects | nMeshes | nComparisons | Good Matches | Unmatchable | Bad Matches | Accuracy |
|-----------|---------|--------------|--------------|-------------|-------------|----------|
| 58 | 175 | 30450 | 167 | 5 | 3 | 98.24% |

In all the system produced 3 incorrect matches over 170 queries. In each case where an incorrect match occurred the subject was captured in an expression varying greatly from the other scans of the same subject in the database. Future revisions of the recognition system will likely incorporate a greater degree of expression invariance, either through expression synthesis or an expression invariant model representation. Despite these shortcomings the overall accuracy of the system is comparable to other current 3D recognition technology.

7. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a complete 3D reconstruction system using a 6 camera capture rig. Stereo correlation was carried out using the Gabor wavelet as the similarity metric. We show that this correspondence metric is relatively robust to changes in illumination levels across stereo pairs and that it is a suitable choice for a similarity measure when used for small baseline stereo applications. Furthermore we introduce a Voronoi cell propagation matching strategy and show that this technique provides suitable

constraints for increasing both the accuracy and speed of the dense stereo correlation process. We also discuss a number of reasons why this matching strategy is particularly suitable for facial reconstruction. Finally we demonstrate how, when used in conjunction with the ICP algorithm, our reconstructed models can be used as the basis for an accurate 3D recognition system.

Whilst we present a complete reconstruction and recognition system, much work can be done to improve on reconstruction accuracy, reconstruction speed and recognition rates. The primary focus of future work will be the introduction of systems for dealing with expression variance on scanned faces. We propose the use of expression invariant representations of face models or expression synthesis in the 3D domain followed by traditional 3D recognition techniques. The introduction of such a system would almost certainly produce 100% accuracy rates on our current 3D database, hence future work should include the expansion of our current set of scans to include more subjects with even greater expression and pose variation. Other potential improvements to the recognition system include the use of texture data in the recognition process or the combination of more traditional 2D methods backed up by the additional 3D data. Finally, we also wish to investigate the use of the 3D Gabor wavelet as a recognition metric on key feature points of the face model. 2D Gabor wavelets have proved a fruitful area of research in the 2D domain and as such the 3D Gabor wavelet may show potential for 3D recognition.

In addition to improving the recognition subsystem we propose a number of improvements to the reconstruction system. It may be possible to increase matching speed and accuracy through the use of model guided matching. The current implementation of the system allows for the reconstruction of non-face models (despite being tailored to face model reconstruction), by further constraining the matching process to only reconstruct face objects it would be possible to improve both model accuracy and reconstruction speed. This approach seems to have some interesting parallels with model based reconstruction and it would be perhaps worth investigating such an approach.

A final important aspect of our work which we have not considered in detail is the selection of a surface meshing algorithm. Our current implementation uses the well researched PowerCrust [24] algorithm to construct surface meshes from point clouds, however, recent work produced by Song Yi [25] for the University of Nottingham suggest that a Bezier surface approach to meshing may be far superior to conventional triangulation approaches.

The availability of a number of commercial 3D scanners has the potential to make our work on model reconstruction somewhat redundant, however, the benefits of developing a reconstruction system tailored to facial 3D reconstruction should allow higher recognition rates than when using a generic commercial scanner. Furthermore the proprietary and non-modular nature of most commercial scanners prevents their use in other areas of research where certain aspects of our work may otherwise be useful. As an example of this our work has been used within research relating to augmented reality for car navigation systems and surface modelling using Bezier surfaces. Such cross project collaboration would not have been possible without the availability of an open, modular reconstruction system such as the one presented here. A closed, proprietary reconstruction system would not have been suitable for the purposes of these other projects and as such our work has wider applicability than that which is presented here, thus justifying the development of an open reconstruction project.

It is hoped that the future work outlined above will eliminate the current shortcomings of the reconstruction / recognition system. Despite the continuing work, the system is at a stage where we have demonstrated accurate reconstruction and recognition results which are comparable with current commercial systems. The implementation of the future work discussed in this section should provide a truly state of the art system, with a high degree of pose, illumination and expression invariance suitable for 3D face recognition within the security industry.

8. REFERENCES

1. Matthew A Turk, Alex P. Pentland. *Face recognition using eigenfaces*. in *Proc. of Computer Vision and Pattern Recognition*. 1991.
2. Tat-Jun Chin, David Suter, *A Study of the Eigenface Approach for Face Recognition*. 2004: Monash University, Technical Report: MECSE-6-2004.
3. Zoran Biuk, Sven Loncaric. *Face Recognition from Multi-Pose Image Sequence*. in *2nd International Symposium on Image and Signal Processing and Analysis*. 2001. Pula, Croatia.
4. Bolme, David S., *Elastic Bunch Graph Matching*, in *Computer Science*. 2003, Colorado State University: Fort Collins. p. 98.
5. Malsburg, Laurenz Wiskott and Jean-Marc Fellous and Norbert Kruger and Christoph von der. *Face Recognition by Elastic Bunch Graph Matching*. in *Proc. 7th International Conference on Computer Analysis of Images and Patterns*. 1997. Kiel: Springer-Verlag.
6. L. Akarun, B. Gokberk, A.A. Salah, *3D Face Recognition for Biometric Applications*. 2005: Department of Computer Engineering, Bogazic University, Bebek, Istanbul, Turkey.
7. Richard Hartley, Andrew Zisserman, *Multiple View Geometry in Computer Vision*. 2003: Cambridge University Press.
8. Naohide Uchida, Takuma Shibahara, Takafumi Aoki, Hiroshi Nakajima, Koji Kobayashi, *3D Face Recognition Using Passive Stereo Vision*. IEEE, 2005.
9. Volker Blanz, Thomas Vetter, *Face Recognition Based on Fitting a 3D Morphable Model*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. **25**(9).
10. Yuxiao Hu, Dalong Jiang, Shuicheng Yan, Lei Zhang, Hongjiang Zhang, *Automatic 3D Reconstruction for Face Recognition*. 2004, Microsoft Research.
11. Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. *Expression-Invariant 3D Face Recognition via spherical embedding*. in *Proc. IEEE ICIP*. 2005.
12. M. Bronstein, A. Bronstein, R. Kimmel, *Three-dimensional Face Recognition*. 2004: Technical Report CIS-2004-04 Department of Computer Science, Technion, Israel.
13. M. Bronstein, A. Bronstein, E. Gorden, R. Kimmel. *Fusion of 3D and 2D Information in Face Recognition*. in *IEEE ICIP*. 2004.
14. M. Bronstein, A. Bronstein, R. Kimmel. *Expression Invariant 3D face Recognition*. in *AVBPA. Lecture Notes on Computer Science*. 2003.
15. 3dMD, *3dMDface™ System including 3dMDpatient*. p. www.3dmd.com. Last Visited: 15/01/07
16. LinLin Shen, Li Bai. *Gabor Feature Based Face Recognition Using Kernel Methods*. in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*. 2004.
17. Peng Yang, Shiguang Shan, Wen Gao, Stan Z. Li, Dong Zhang. *Face Recognition Using Ada-Boosted Gabor Features*. in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*. 2004.
18. Yanwei Pang, Lei Zhang, Mingjing Li, Zhengkai Liu, and Weiyang Ma, *A Novel Gabor-LDA Based Face Recognition Method*. 2004.
19. Malsburg, M. Potzsch and N. Kruger and C. von der, *Improving Object Recognition by Transforming Gabor Filter Responses*. *Computation in Neural Systems*, 1996. **7**(2).
20. Daugman, J.G, *Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile*. *Vision Research*, 1980. **20**: p. 847-856.
21. Li Tang, Tsui, Wu, *Dense Stereo Matching Based on Propagation with a Voronoi Diagram*. 2003.
22. Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, Duane Fulk. *The Digital Michelangelo Project: 3D scanning of large statues*. in *SIGGRAPH*. 2000.
23. Szymon Rusinkiewicz, Marc Levoy. *Efficient Variants of the ICP Algorithm*. in *Third International Conference on 3D Digital Imaging and Modeling*. 2001.
24. Nina Amenta, Sunghee Choi and Ravi Kolluri, *The power crust, unions of balls, and the medial axis transform*. *Computational Geometry: Theory and Applications*, 2001. **19**(2-3): p. 127-153.
25. Yi Song, Li Bai, *3D Surface Modelling and Warping for Face Recognition*. *ICP 2005*, 2005.