3D modelling and recognition

RODRIGUES, Marcos <http://orcid.org/0000-0002-6083-1303>, ROBINSON, Alan, ALBOUL, Lyuba <http://orcid.org/0000-0001-9605-7228> and BRINK, Willie

3D Modelling and Recognition

M. Rodrigues, A Robinson, L. Alboul, W. Brink
Geometric Modelling and Pattern Recognition Research Group
Materials and Engineering Research Institute
Sheffield Hallam University, Sheffield, UK
M.Rodrigues@shu.ac.uk   http://www.shu.ac.uk/gmpr

Abstract

3D face recognition is an open field. In this paper we present a method for 3D facial recognition based on Principal Components Analysis. The method uses a relatively large number of facial measurements and ratios and yields reliable recognition. We also highlight our approach to sensor development for fast 3D model acquisition and automatic facial feature extraction.

Keywords: 3D face recognition, modelling, 3D scanning, image processing, pattern recognition

1 Introduction

In this paper we highlight methods for full 3D-3D acquisition, modelling and recognition. Our aims are to develop robust methods and procedures for real time capture of the 3D geometry of a human face, process the 3D model to extract relevant features and proceed to identification based on such features.

Much research has been undertaken in the area of 2D face recognition (e.g. see survey in [1]) while 3D face recognition is incipient. It is often said [2] that 3D face recognition has the potential for greater accuracy than 2D techniques, as 3D is invariant to pose and illumination and incorporates important face descriptors embedded within the 3D features. Thus, using 3D the facial descriptors can be enhanced with added accuracy. The challenges to improved 3D face recognition [2] including real time applications reflect the shortcomings of current methods:

1. the need for fast and accurate 3D sensor technology,
2. improved algorithms to take in consideration variations in size and facial expression (or a model incorporating non-rigid motion of the face), and
3. improved methodology and datasets allowing algorithms to be tested on large databases, thus removing bias from comparative analyses of the algorithms.

Examples of some approaches to modelling and recognition include methods based on feature selection [1], estimating local curvatures and eigenvector decomposition [3]. Dynamically deformable models were first proposed by Terzopoulos et al. [4] and have attracted considerable interest. However, work on this topic (see [5]) yielded limited and mixed results. In that work, a wire frame deformable model is used to obtain subject-specific 3D face representations. When a new face is to be recognised, the 3D pose of the face is estimated and the faces of all persons in the database are projected to this view using their 3D representations. It then becomes a 2D face recognition problem but without lighting variations. Such an approach is also referred to as 3D wire frame models (WFM) of the face, and has been used for face synthesis [6, 7]. Non-parametric [7] and parametric [8] deformable models have been used for feature detection but again with limited results and are mostly dependent on user intervention.

Addressing the above challenges for improved 3D facial recognition, this paper describes work on current sensor technology and automatic extraction of facial features together with a method for face identification with some preliminaries results on recognition.

2 Fast 3D Model Acquisition Using Uncoded Structured Light

Capture of the 3D surface model is achieved using GMPR's patented scanning device, which uses the structured light method [9,10,11] . By using a dense pattern of stripes of uniform width and only one or two brightness levels, the density of vertices is very high, and a coloured texture map can also be produced, as depicted in Figure 1. This shows the original bitmapped image (detail of the eye region), and arbitrary poses of shaded polygons, the 3D colour-mapped model, and a wire-frame model.

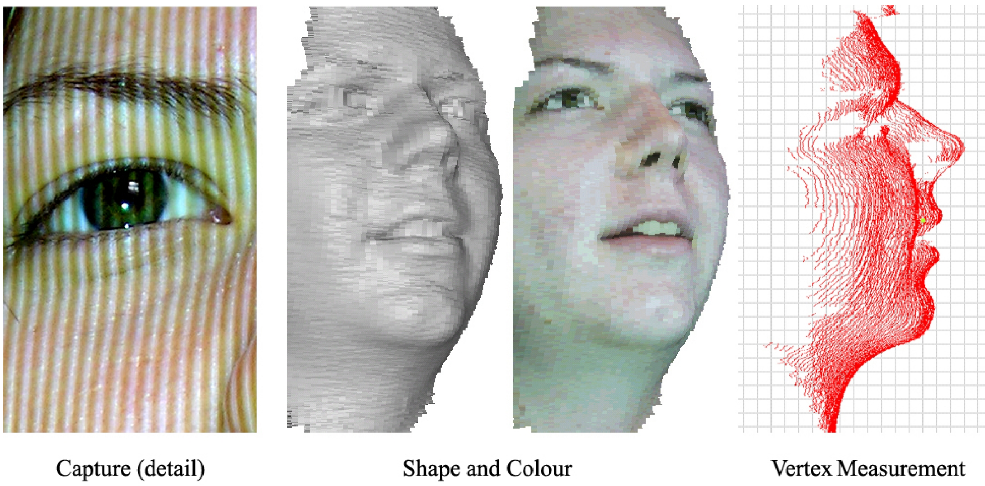

Fig 1: Multiple stripe patterns are processed to recover 3D geometry.

Our method projects a pattern of evenly-spaced stripes onto the subject, and records the deformation of the stripes in a video camera placed in a fixed geometric relationship to the stripe projector. Using dense, uncoded stripes presents greater feature correspondence problems, but provides greater resolution of measurement and allows an accurately-coloured texture map. Solutions to the uncoded stripe problem are given in [15]. Fig 2a shows a detail from one such video frame, clearly showing the deformed stripes. The advantage of this over stereo vision methods is that the stripe pattern provides an explicitly connected mesh of vertices (Fig 2c), so that the polyhedral surface can be rendered without the need for surface reconstruction algorithms. Also, a smoothly undulating and featureless surface (such as in Fig 2a) can be more easily measured by structured light than by stereo vision methods. High-speed acquisition at video rates opens up the possibility of animated 3D models, which is the subject of further work by our group [11].

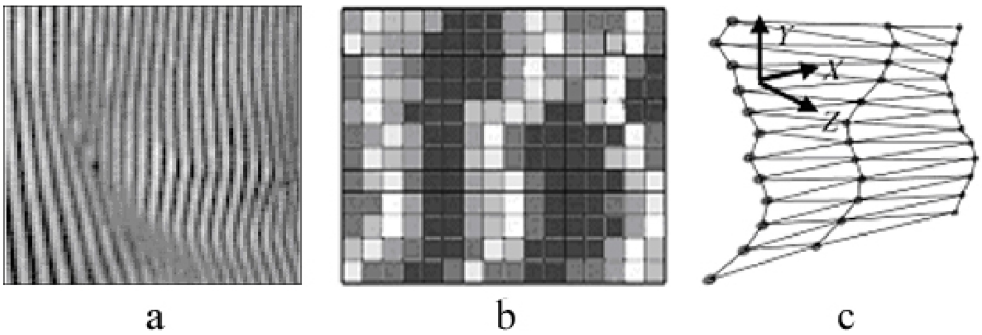

Fig 2: (a) Part of image showing stripe deformation. (b) Detail showing pixel array (c) resulting mesh of vertices.

Once the surface shape has been modelled as a polygonal mesh, we return to the video image, take the colour of the reflected white stripe at each pixel that maps to a vertex, and colour the vertex (or triangle) accordingly. The final model therefore contains the (x, y, z) coordinates and their corresponding RGB (red, green, blue) values for each vertex, and the face can be visualised as in Fig 1.

The method described above uses a standard video image of 768×576 pixels at a frame rate of 1/25th second; our latest camera has a maximum resolution of 2208 x 3000 pixels and a maximum frame rate (at lower resolution) of 1/271 seconds. The resolution and accuracy of the face data is directly related to the pixel size of the image, so that the maximum number of vertices captured in the highest resolution camera is currently 2208×750. The lower resolution in one dimension is due to the spacing between stripes, but we are developing a new method to increase this resolution threefold, to give a resolution of $> 2000 \times 2000$ vertices. The accuracy of measurement between two adjacent vertices will be > 0.2 pixels. These figures mean in practice that in the new system if a human face occupies the whole of the viewing volume, the resolution per vertex will be approximately 0.1mm, and the accuracy 0.05mm.

2.1 Image Processing

The method starts by acquiring 2D images of the subject. Noise removal functions are required in order to recover a more complete 3D model. Figure 3 shows on the left a 3D model from a noisy 2D image. Undesirable holes are present from such un-processed 2D image. On the right, noise in the 2D image was removed using a combination of median filters and weighted mean filters and the recovered model has no holes except for occluded areas. As any occlusion leaves holes in the model, we deal with this by a hole filling procedure applied to the reconstructed 3D model. The hole filling algorithm uses a bilinear interpolation which is very appropriate for the application given the intrinsic continuous, smooth nature of the human face.

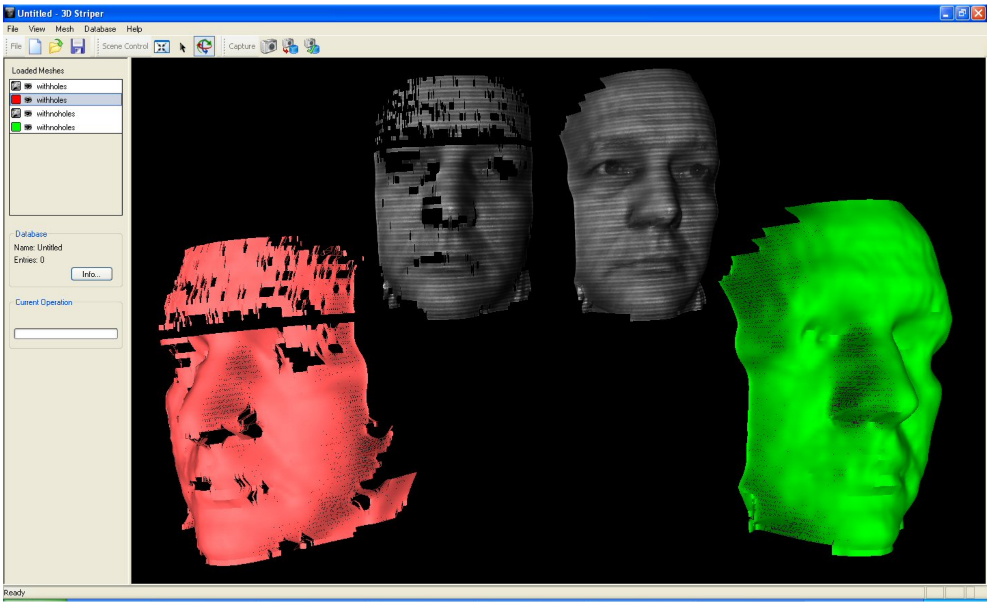

Fig 3: The effects of noise in the input image. On the left, the 3D model is recovered from a noisy 2D image, resulting in a large number of holes. On the right, the same 2D image with noise removal prior to 3D recovery. Any remaining small holes due to occlusion are filled in 3D by a bilinear interpolation.

3 Feature Extraction

Once a 3D face model is acquired, measurements are performed over a number of key feature points such as position of the eyes, tip of nose, and so on. The process of extracting 3D features is automatic; all that is required is to provide the position of the eyes in the original 2D image. At the moment, such 2D co-ordinates are determined manually, and the next implementation will include an automatic eye detection function in 2D. Several research groups worldwide including [12,13] have developed automatic eye detection algorithms and it is a matter of choosing the most appropriate method given our camera and image parameters.

Once the eyes are tagged in 2D and the 3D face model is constructed, there is an one-to-one relationship between the tagged eye positions in 2D and their counterparts in 3D. Figure 4 shows a number of points automatically determined. Only 12 points are highlighted in the model for clarity but we determine a total of 84 distances and ratios and we are continuously improving automatic feature point detection. Automatic detection is based on determining a number of key planes on the face model, followed by point projections on such planes, then measuring the various inter-point distances and geodesic distances on the mesh together with a number of horizontal and vertical ratios. The recognition engine then takes these 84 distances and performs both identification and verification as required.

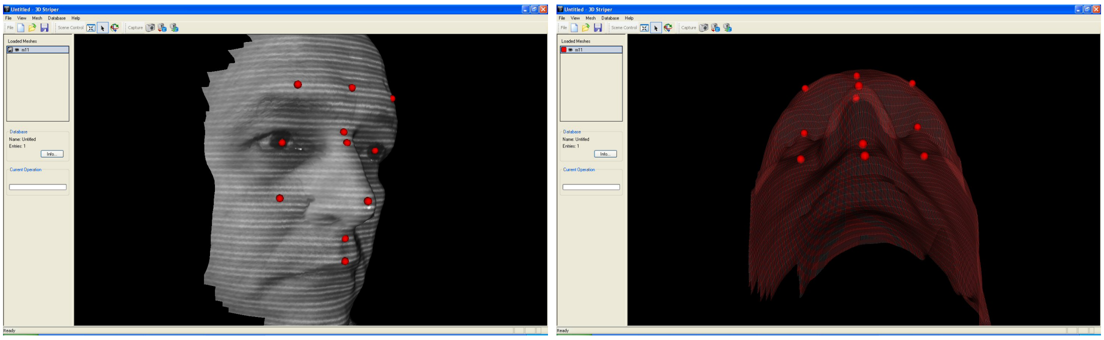

Fig 4: Left: from the tagged eye positions in 2D, their 3D counterparts are determined together with a large number of ancillary points (only 12 points are shown for clarity). Right: cutting planes and finding nearest points to such planes allows the determination of a stable set of points.

4 Recognition

The above set of 3D measurements uniquely identify a face. As mentioned above, each face is made out of a vector with 84 entries. These data are organised into a matrix where each column is a new face while each row describes a particular measure of that face. Once a face exists in the database, i.e. a person is enrolled, recognition can take place in the form of identification which is a one-to-many mapping to retrieve the closest face to the input vector, or verification which is a one-to-one mapping to verify the claimed identity. Identification and verification are performed based on Principal Components Analysis (PCA).

The purpose of PCA is to derive new variables in decreasing order of importance that are a linear combination of the original variables and are uncorrelated. Geometrically, we can think of PCA as a rotation of the original coordinate axis to a new set of orthogonal axes that are ordered according to the amount of variation of the original data they account for. There are a number of ways in which a set of principal components can be derived. We choose to highlight the Hotelling approach here, and more details can be found in the literature (e.g. [14]).Let $x_1, \ldots, x_p$ be our set of original variables (the 3D facial measurements) and let $\xi_i, i = 1, \ldots, p$ be a linear combination of these variables.

$$\xi_i = \sum_{j=1}^{p} a_{ij} x_j \qquad \text{or} \qquad \xi = \mathbf{A}^{\mathbf{T}} \mathbf{x} \tag{1}$$

where ξ and $\mathbf{x}$ are vectors and $\mathbf{A}$ is the matrix of coefficients. $\mathbf{A}$ is the the orthogonal transformation that when applied to the vector $\mathbf{x}$ yields new variables ξ_j that have stationary values of their variance. For instance, considering the first variable ξ_1:

$$\xi_1 = \sum_{j=1}^{p} a_{1j} x_j \tag{2}$$

We choose $\mathbf{a}_1 = (a_{11}, a_{12}, \ldots, a_{1p})^T$ to maximise the variance ξ_1,subject to the orthogonal constraint $\mathbf{a}_1^T \mathbf{a}_1 = |\mathbf{a}_1|^2 = 1$. The variance of ξ_1 is

$$\text{var}(\xi_i) = \mathbf{a}_1^T \mathbf{\Sigma} \mathbf{a}_1 \tag{3}$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\mathbf{x}$. For a non-trivial solution, $\mathbf{a}_1$ must be an eigenvector of $\mathbf{\Sigma}$, so now $\mathbf{\Sigma}$ has p eigenvalues $\lambda_1, \ldots, \lambda_p$ which can be ordered such that

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0.$$

Since we wish to maximise the variance we choose the largest eigenvalue λ_1 and $\mathbf{a}_1$ its corresponding eigenvector. Continuing the argument, $\mathbf{a}_2$ is also an eigenvector of $\mathbf{\Sigma}$, orthogonal to $\mathbf{a}_1$, and the kth principal component $\xi_k = \mathbf{a}_k^T \mathbf{x}$, where $\mathbf{a}_k$ is the eigenvector corresponding to the kth largest eigenvalue of $\mathbf{\Sigma}$ and variance equal to the kth largest eigenvalue. Thus, in Equation 1, $\mathbf{A} = [\mathbf{a}_1|\dots|\mathbf{a}_p]$ is the matrix whose columns are the eigenvectors of $\mathbf{\Sigma}$.

Thus, in the case of facial data, each face is treated as a vector of N values (as we have 84 measures so $N = 84$). A facial database, or a set of facial data, is represented by a two-dimensional matrix $N \times M$ where M is the number of face vectors. Let this set of data, called the training or enrolment set, be referred to as $\tau_{\mathbf{1}}, \tau_{\mathbf{2}}, \dots, \tau_{\mathbf{M}}$ for a set containing M vectors. It is straightforward to calculate the average vector $\bar{\tau}$ of this set using

$$\bar{\tau} = \frac{1}{M} \sum_{i=1}^{M} \tau_i \tag{4}$$

and subsequently to calculate the difference δ from the mean $\bar{\tau}$ for each vector

$$\delta_{\mathbf{i}} = \tau_i - \bar{\tau} \tag{5}$$

where $(i = 1, 2, \dots, M)$ resulting in $\delta_{\mathbf{1}}, \delta_{\mathbf{2}}, \dots, \delta_{\mathbf{M}}$ vectors. We now have a set of difference vectors in a matrix of dimension $N \times M$ and ready to perform eigenvector decomposition.

In this context, finding the principal components is equivalent to finding the set of orthonormal vectors $\mathbf{u_n}$ and their associated eigenvalues λ_i which best describe the distribution of the face vector data. We can obtain $\mathbf{u_n}$ and λ_i by applying eigen analysis to the covariance matrix of the data, C, which measures the tendency of two face vectors to vary together for each combination of two vectors in the data set. The covariance matrix C can be obtained by

$$C = \frac{1}{M} \sum_{n=1}^{M} \delta_{\mathbf{n}} \delta_{\mathbf{n}}{}^T \tag{6}$$

which equates to $AA^T = C$. Once the eigenvectors of the matrix C are evaluated, we can use the associated eigenvalues to rank eigenvectors by their usefulness in characterising the variation among face vectors: the eigenvector with the highest eigenvalue is the most useful and so on. Based on this, a number M' of eigenvalues can be used as coefficients to uniquely identify a face vector where $M' < M$.

Thus, once we have M eigenvectors of AA^T we can safely discard some, giving us M' eigen face vectors. We propose that a suitable number M' in the context of facial data is between 10 and 15. This set of eigen face vectors can be accurately described as a linear combination of all face vectors in the training/enrolment set. The coefficients of the eigen face vectors in this linear combination are called the weights, ω. To obtain the weight vectors for any face vector $\tau_{\mathbf{i}}$, simply calculate for $k = 1, 2, \dots, M'$:

$$\omega_k = \mathbf{u}_i^T (\tau_{\mathbf{i}} - \bar{\tau}) \tag{7}$$

The weight vectors are put into a matrix of weights $\mathbf{\Omega^T} = [\omega_{\mathbf{1}}, \omega_{\mathbf{2}}, \dots, \omega_{\mathbf{M'}}]$ and this is the basis for recognition. In order to proceed, we need to determine which face vector in the training set is most similar to an input vector. First, we calculate the weight vectors for all known vectors as well as for the unknown vector. Then the best match is the face vector whose weight vector has the smallest distance from the weight vector of the unknown face vector, i.e. the $\tau_{\mathbf{k}}$ which minimises the distance measure

$$\tau_{\mathbf{k}} \rightarrow ||\mathbf{\Omega} - \mathbf{\Omega_k}||_{\min} \tag{8}$$

The distance measures that can be used are, for instance, the Mahalanobis or Euclidean distances. Given two vectors X_1, X_2, the distances are defined as follows:

- Euclidean distance: $d_{12}^2 = (X_1 - X_2)(X_1 - X_2)^T$
- Mahalanobis distance: $d_{12}^2 = (X_1 - X_2)^T V^{-1} (X_1 - X_2)$
  where V is the sample covariance matrix.

The method highlighted above picks the vector in the training/enrolment set with the shortest distance from the input vector. By looking at the identity of the picked vector, the assumption is that the input vector has the same characteristics, i.e. is the same person if we are verifying identity. If we are performing identification, then a ranking method can be used to pick say the 5 closest vectors ranking them by shortest distances or by setting some minimum threshold to decide whether the input vector is the same or not. Such thresholds must be obtained through experimentation.

5 Results and Conclusion

We have tested our methods on synthetically generated data (400 entries) and on a small database of 3D facial data (40 entries). Synthetic data were generated from small random variations on the measurements from a real set of facial data. It is clear that such method has severe limitations in representing real faces as randomness means that some synthetic faces were quite asymmetric and almost impossible to occur as most faces have a good degree of symmetry. Nevertheless, the purpose was to investigate whether a given vector could be recognised from a database or not. For synthetic data results ranged from 98.4% to 100% accuracy in identifying a given vector in the database. For real data, accuracy was lower at 92%.

We believe that the poorer performance of the acquired compared with the synthetic data is related to the accuracy limitations of using a standard video camera with an LCD projector, and to the measurement of the calibration constants. For these reasons we are currently acquiring new data using a high resolution camera (2208×3000 pixels), an optical projector that can project 250 stripes onto the face, all mounted on an optical bench. The results with this new system will enable us to judge how important such precise measurement is to the effectiveness of the recogniser, and this will in turn determine the likely cost of the final application.

We are working on improving all those aspects with special emphasis on camera calibration and controlling environmental variables such as the ones impairing the quality of the acquired 2D bitmapped image. The results above on a small database of real facial measurements are promising but need to be tested on large scale model acquisition. We are acquiring a larger database of high quality models and results will be reported in the near future.

References

[1] T. Nagamine, T. Uemura, and I. Masuda. "3D facial image analysis for human identification", *ICPR 1992*, Netherlands, pp. 324–327.

[2] K.W. Bowyer, K. Chang, P. Flynn, "A Survey of 3D and Multi-Modal 3D+2D Face Recognition", *ICPR 2004*, Cambridge (UK), 2005.

[3] C. Hesher et al., "A novel technique for face recognition using range images", *Intl Symp on Sig Proc Apps, 2003*, Paris,

[4] D. Terzopoulos, A. Witkin, and M. Kass, "Symmetry-Seeking Models for 3D Object Reconstruction", *IJCV* 1(3) (1987) 211–221.

[5] M. W. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model", *Pattern Recognition* 36 (2003) 1835–1846.

[6] M. Rydfalk, *CANDIDE, a parameterised face*, Internal Report, Department of Electrical Engineering, Link oping University, Sweden, (October 1987).

[7] K. Aizawa, H. Harashima, T. Saito, "Model-based Analysis synthesis Image Coding for a Persons Face", *Image Commun.* 1(2) (1989) 139–152.

[8] A.L. Yuille, P.W. Halliman, D.S. Cohen, "Feature Extraction from Faces using Deformable Templates", *Int. J. Comput. Vision*, 8(2) (1992) 133–144.

[9] A. Robinson and M. Rodrigues, *Fast 3D Acquisition with the GMPR Scanner*, http://www.shu.ac.uk/research/meri/gmpr/projects/scanner.html

[10] M.A. Rodrigues, Alan Robinson, Lyuba Alboul. "Apparatus and Methods for Three Dimensional Scanning, Multiple Stripe Scanning Method", *UK Patent Application* No. 0402565.6, February 5th, 2004.

[11] A. Robinson, M.A. Rodrigues, L. Alboul, "Producing Animations from 3D Face Scans", *Game-On 2005, 6th Annual European GAME-ON Conference*, De Montfort University, Leicester, UK, Nov 23–25, 2005.

[12] Peng Wang, Matthew B. Green, Qiang Ji and James Wayman, "Automatic Eye Detection and Its Validation", *IEEE Workshop on Face Recognition Grand Challenge Experiments* (with CVPR), San Diego, CA, June 2005

[13] John Cowel, *Automatic eye detection in facial images with unconstrained backgrounds*, http://www.cse.dmu.ac.uk/ jcowell/Research/ifmip-035.pdf

[14] A. Web, *Statistical Pattern Recognition*, Arnold Publishers, London, 1999.

[15] A. Robinson, L. Alboul and M.A. Rodrigues, "Methods for Indexing Stripes in Uncoded Structured Light Scanning Systems", *Journal of WSCG*, 12(3), pp. 371-378, 2004.