Three-Dimensional Model Based Face Recognition*

Xiaoguang Lu, Dirk Colbry, and Anil K. Jain
Department of Computer Science & Engineering,
Michigan State University, East Lansing, MI 48824
{lvxiaogu, colbrydi, jain}@cse.msu.edu

Abstract

The performance of face recognition systems that use two-dimensional (2D) images is dependent on consistent conditions such as lighting, pose and facial expression. We are developing a multi-view face recognition system that utilizes three-dimensional (3D) information about the face to make the system more robust to these variations. This paper describes a procedure for constructing a database of 3D face models and matching this database to 2.5D face scans which are captured from different views, using coordinate system invariant properties of the facial surface. 2.5D is a simplified 3D (x, y, z) surface representation that contains at most one depth value (z direction) for every point in the (x, y) plane. A robust similarity metric is defined for matching, based on an Iterative Closest Point (ICP) registration process. Results are given for matching a database of 18 3D face models with 113 2.5D face scans.

1. Introduction

Automatic human face recognition is a challenging task that has gained a lot of attention during the last decade [16]. While most efforts have been devoted to face recognition from two-dimensional (2D) images [16], a few approaches have utilized depth information provided by 2.5D range images [4,10-12]. Current 2D face recognition systems can achieve good performance in constrained environments, however, they still encounter difficulties in handling large amounts of facial variations due to head poses, lighting conditions and face expressions [7] (See Figure 1). Because the human face is a three-dimensional (3D) object whose 2D projection (image) is sensitive to the above changes, utilizing 3D face information can improve the face recognition performance [2, 7]. Range images captured explicitly by a 3D sensor [5, 13] present face surface shape information. These sensors also provide registered texture images. The 3D shape of facial surface represents the face structure, which is related to the internal facial anatomical structure instead of external appearance and environment. In this research, 3D models are used to recognize 2.5D face images. A 2.5D image is a simplified 3D (x, y, z) surface representation that contains at most one depth value (z direction) for every point in the (x, y) plane (see Figure 2).

Different methods have been used to address face recognition based on range images [10-12]. Most of these methods are focused on matching frontal facial range images and require that key facial feature points are known within the image to align the face images. We propose a multi-view face recognition system from range images across illumination and expression.

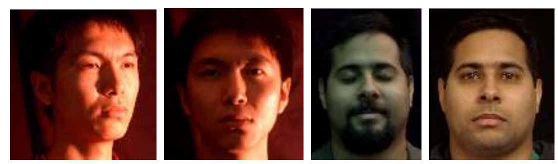
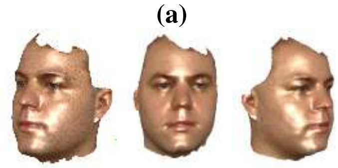


Figure 1. Face Appearance variations.

(a)



(b)

Figure 2. 2.5D scans and 3D face model. (a) one 2.5D test scan from different viewpoints; (b) full 3D model.

2. Three Dimensional Face Modeling

We considered various methods for producing a 3D model from 2.5D scans. We chose 3D VRML models because most 3D editing programs can generate output in the VRML format. By using such a flexible format we hope to test our system on face models using various types of scanners, software and algorithms. For the experiments reported here, we use a combination of five test scans generated with a Minolta VIVID 910 scanner. For each subject, the scans were stitched together using commercial software package, called Geomagic Studio[9]. All the models were also cleaned up by filling holes, smoothing

* This research was supported by US Army contract No. DAAD05-03-C-0045.

the surface and deleting noisy points associated with hair and clothing. We also used a decimate function to reduce the number of polygons in the model to a reasonable level (approximately 50,000). The end result is a smooth full view of the face for each of our subjects.

3. Face Alignment and Matching

There are four main steps in our face recognition system that matches a given 2.5D facial scan to the 3D face models stored in the database. (i) Automatic feature point detection in the 2.5D scans; (ii) rigid transform to coarsely align the 2.5D scan with the full 3D model; (iii) fine iterative registration using the Iterative Closest Point algorithm (ICP); (iv) using local feature information at the correlation points from ICP for matching.

3.1 Automatic Feature Point Detection

A minimum of three corresponding points is needed in order to calculate the rigid transformation between two sets of 3D points. We determine the local shape information at each point within the 2.5D scan [6]. The shape index at point p is calculated using the maximum (κ1) and minimum (κ2) local curvature (see Eq. 1). This produces a shape scale from zero and one.

$$S(p)=\frac{1}{2}-\frac{1}{\pi}\tan^{-1}\frac{\kappa_1(p)+\kappa_2(p)}{\kappa_1(p)-\kappa_2(p)} \qquad (1)$$

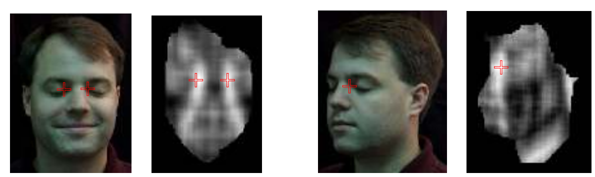
The low end of the shape index scale represents a spherical cup while the high end of the scale represents a spherical cap. In the middle of the space (value of 0.5) is a saddle point. This shape calculation is independent of the coordinate system and, therefore, it is a potentially useful metric for finding similar points between scans with different poses. The shape index algorithm used here, requires that the data points be in a uniform grid. In order to use the same algorithm for the 3D polygonal VRML face model, the polygon mesh is converted into cylindrical coordinates, which can then be stored in a two dimensional matrix of height and angle.

Figure 3. Shape index images.

Figure 3 shows the computation of the shape index; the dark regions represent lower numbers and the light regions represent higher numbers. Notice that there are several consistencies (correspondences) between these two scans of the same face. For example, the area between the eyes and the bridge of the nose is consistently trough shaped (dark) as well as the area around the mouth. We use these consistencies to help locate prominent feature points within the face regardless of the face orientation.

For coarse registration we pick a combination of the inside of one eye, the outside of that eye and the nose tip as our three feature points. These points are selected because they are relatively easy to locate in the range image and they do not change between different scans of different people in different poses. Our model of the face requires that the major axis of the scanned image be vertical (i.e., we assume that in all scans, the eyes are above the mouth). The easiest feature point to identify is the inside edge of an eye right next to the bridge of the nose, because this point has a shape index value that is very close to zero and the area around this point has a consistent shape index value across all face images and poses. A simple averaging mask of size 30×10 in the shape space is used to identify this area (see Figure 4).

Once this inside eye-point is found, other feature points within the face can be located using a simple model of the relative locations between the points as well as other heuristics based on local surface curvature.



(a) **(b)**

Figure 4. Identifying feature points using convolution in the shape space. (a) frontal view, (b) semi-profile view. Bright spots represent inside of the eye next to the bridge of the nose.

3.2 Coarse Alignment

Once the three points have been identified (inside eye, outside eye and mouth), a rigid transform can be applied [14] using a least square fitting between the triangles formed from the two sets of three points. See Figure 5 for an example of a 2.5D face image coarsely aligned to a 3D face mesh model.

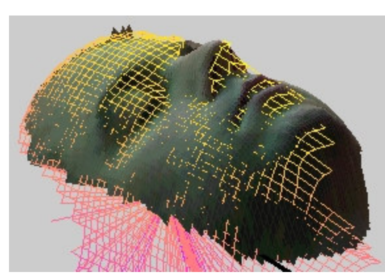


Figure 5. A 2.5D face scan (grid) coarsely aligned to a 3D textured model from the same subject.

3.3 Fine Alignment with Hybrid ICP

Our fine registration process follows the Iterative Closest Point (ICP) framework [1, 3, 15]. Starting with an initial estimate of the rigid transformation, ICP iteratively refines the transform by alternately choosing corresponding (control) points in the 3D model and the 2.5D scan and finding the best translation and rotation that minimizes an error function based on the distance between them.

Besl [1] uses point-to-point distance and a close-form solution when calculating the transformation matrix during each iteration. The point-to-plane distance used in Chen [3], makes the ICP algorithm less susceptible to local minima than the point-to-point metric [1, 8]. We integrate the two classical ICP algorithms [1, 3] in a zigzag running style, called the hybrid ICP algorithm. Each iteration consists of two steps, using Besl's scheme to compute an estimation of the alignment, followed by Chen's scheme for a refinement. The two different distance metrics are utilized together, which has the potential for a better registration result.

In order to minimize the number of outliers, regions were selected within the face scans that do not vary greatly between the scans. Figure 6 shows the grids of control points selected for various poses. Regions around the eyes and nose were selected because these regions are less malleable than other parts of the face (such as the region around the mouth, which changes greatly with facial expression.). The fine alignment result is demonstrated in Fig. 7.

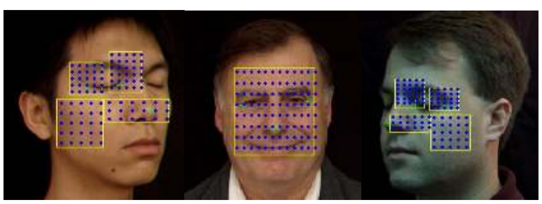


Figure 6. Control point selection for a left profile, front, and right profile scans. (There are a total of ~100 control points selected in each scan).

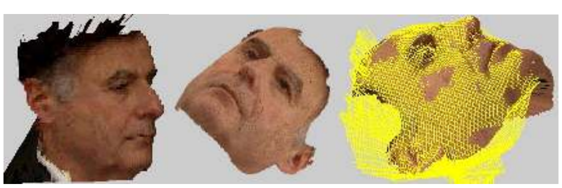


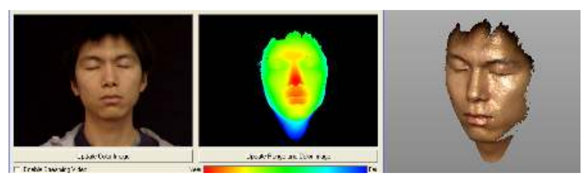
Figure 7: Matching results after fine alignment. From left to right, the 2.5D test scan, the 3D model, the 3D textured model overlaid by the wire-frame of the test scan after fine alignment.

3.4 Matching

The root-mean-square distance minimized by the ICP algorithm is used as the primary matching score of face scans. Each control point is associated with a shape index value. The cross-correlation between the shape index vectors of two sets of control points is computed as the shape index matching score, which was combined with the ICP distances after z-score normalization in order to improve the matching score. The sum rule is applied for the combination.

4. Experiments

Currently, there is no multi-view range (with registered textures) face database, which is publicly available. In our experiments, all range images (320×240 with a depth resolution of ~0.1mm) were collected using a Minolta Vivid 910 scanner [13]. This scanner uses structured laser light to construct the face image in less than a second. Each point in a scan has a texture color (r, b, g) as well as a location in 3D space (x, y, z). Each facial scan has around 18,000 effective points (excluding the background). Figure 8 shows an example of one of these scans.



(a) (b) (c)

Figure 8. An example of Minolta Vivid 910 facial scan. (a) texture image; (b) range image, showing points closer to the sensor in red; (c) 3D visualization.

There are 18 subjects in our database. Five scans with neutral expression for each subject were used to construct the 3D model. The test database consists of 113 different scans of the same 18 people. All the scans varied in pose, facial expression and lighting. Therefore, there are a total of 18 3D models in the database and 113 independent 2.5D scans for testing.

While the feature point extraction algorithm is fairly robust, there are some images where our approach cannot accurately locate the required feature points. In order to properly evaluate the performance of this matching system, we run the entire system using manually selected feature points. The error between the manually selected and automatically derived feature points is 10mm with a standard deviation of 17mm. These errors are not easy to interpret since many of the points selected by the system are reasonable even though they do not match the ground truth. In addition, there are some large outliers where the automatic point selection system fails completely due to noise in the data from ears and facial hair.

The face recognition system was run with two different matching scores. The first only uses the distance measurement produced by the ICP algorithm. The second uses the distance measurement as well as the local shape index for each of the control points.

A summary of the experimental results is shown in Table 1. The best matching occurs when both surface matching and shape index are used for the matching score.

Table 1. Matching Error Rates

	Best Match Classification
ICP Only	4.4 %
ICP + Shape Index	3.5 %

All the errors were due to changes in facial expressions. These changes are non-rigid and actually change the 3D shape of the model. Table 2 shows a summary of the errors with respect to two variations; smiling and pose.

Table 2. Distribution of Matching error from a total of 113 test scans

	Frontal	Profile	Total
w/o smile	0	0	0
w/ smile	1	3	4
Total	1	3	4

Figure 9 shows two examples of difficult 2.5D scans that are correctly matched by our algorithm to the 3D model. Figure 10 shows two examples of difficult 2.5D scans that are incorrectly matched.

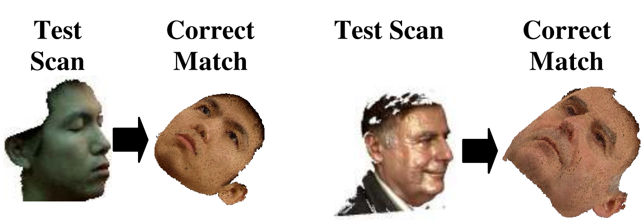


Figure 9. Examples of correctly matched test scans. Notice the changes in lighting, pose and facial expression.

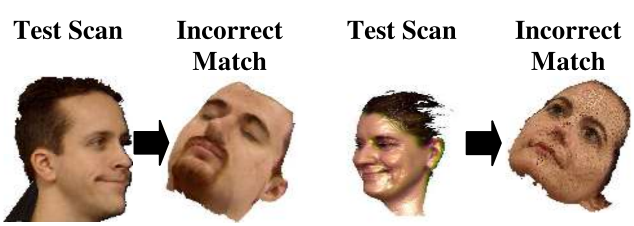


Figure 10. Examples of incorrectly matched test scans.

5. Conclusions and Future Work

We have presented a face recognition system that matches 2.5D scans of arbitrary poses and expressions to a database of 3D models. This research is an encouraging first-step in designing a system that is capable of recognizing faces with arbitrary pose and illumination. We have learned that a 3D polygonal model by itself is not enough to solve this ambitious goal; instead, additional work is required to incorporate color and texture in order to improve the system's recognition capabilities. We are also beginning to look at models that can be deformed to deal with non-rigid variations such as facial expressions. A larger database of 100 3D face models is being constructed.

6. References

[1] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. PAMI*, vol. 14, no. 2, pp. 239-256, 1992.
[2] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1063-1074, 2003.
[3] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145-155, 1992.
[4] C. Chua, F. Han and Y. Ho. "3D Human Face Recognition Using Point Signature," *Proc. Fourth Int. Conf. on FG*, pp. 233-238, Grenoble, March 2000.
[5] "Cyberware Inc.," http://www.cyberware.com/.
[6] C. Dorai and A. K. Jain, "COSMOS - A Representation Scheme for 3D Free-Form Objects," *IEEE Trans. on PAMI*, vol. 19, no. 10, pp. 1115-1130, 1997.
[7] "Face Recognition Vendor Test" http://www.frvt.org/.
[8] N. Gelfand, L. Ikemoto, S. Rusinkiewicz and M. Levoy. "Geometrically Stable Sampling for the ICP Algorithm," *Proc. International Conference on 3D Digital Imaging and Modeling*, pp. 260-267, Banff, October 2003.
[9] "Geomagic Studio," Raindrop Software http://www.geomagic.com/products/studio/.
[10] G. Gordon. "Face Recognition Based on Depth and Curvature Features," *Proc. CVPR*, pp. 108-110, 1992.
[11] J. Lee and E. Millios, "Matching Range Images of Human Faces," *Proc. ICCV*, pp. 722-726, 1990.
[12] H. Tanaka, M. Ikeda and H. Chiaki. "Curvature-based face surface recognition using spherical correlation," *Proc. Third Int. Conf. on FG*, pp. 372-377, 1998.
[13] "Minolta Vivid 910 non-contact 3D laser scanner," http://www.minoltausa.com/vivid/.
[14] D. M. Weinstein. "The Analytic 3-D Transform for the Least-Squared Fit of Three Pairs of Corresponding Points," *School of Computing Technical Report,* No. UUCS-98-005, University of Utah, Salt Lake City, UT, March 1998.
[15] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International Journal of Computer Vision*, vol. 13, no. 1, pp. 119-152, 1994.
[16] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips. "Face Recognition: A Literature Survey," *CVL Technical Report,* University of Maryland, Oct. 2000.