

# Matching 2.5D Scans for Face Recognition

Xiaoguang Lu, Dirk Colbry, and Anil K. Jain

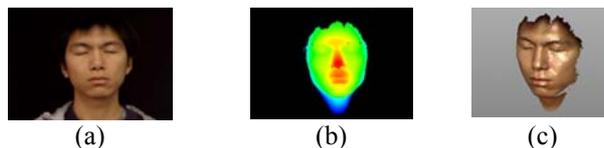
Department of Computer Science & Engineering,  
Michigan State University, East Lansing, MI 48824  
{lvxiaogu, colbrydi, jain}@cse.msu.edu

**Abstract.** The performance of face recognition systems that use two-dimensional images is dependent on consistent conditions such as lighting, pose, and facial appearance. We are developing a face recognition system that uses three-dimensional depth information to make the system more robust to these arbitrary conditions. We have developed a face matching system that automatically correlates points in three dimensions between two 2.5D range images of different views. A hybrid Iterative Closest Point (ICP) scheme is proposed to integrate two classical ICP algorithms for fine registration of the two scans. A robust similarity metric is defined for matching purpose. Results are provided on a preliminary database of 10 subjects (one training image per subject) containing frontal face images of neutral expression with a testing database of 63 scans that varied in pose, expression and lighting.

## 1. Introduction

Automatic human face recognition has gained a lot of attention during the last decade [1]. A few approaches utilized depth information provided by 2.5D range images [2-5], but most efforts have been devoted to face recognition from only two-dimensional (2D) images [1]. Current 2D face recognition systems can achieve good performance in constrained environments, however, they still encounter difficulties in handling large variations in pose and illumination [6]. Utilizing 3D information can improve face recognition performance [6, 7]. Range images captured by 3D sensor [8, 9] explicitly represent face surface shape information as well as providing registered color texture images. Face recognition based on range images has been addressed in different ways [2-4, 10, 11].

In this research 2.5D range images are used for face recognition. A 2.5D image is a simplified three-dimensional  $(x, y, z)$  surface representation that contains at most one depth ( $z$ ) value for every point in the  $(x, y)$  plane (see Figure 1). A full three-

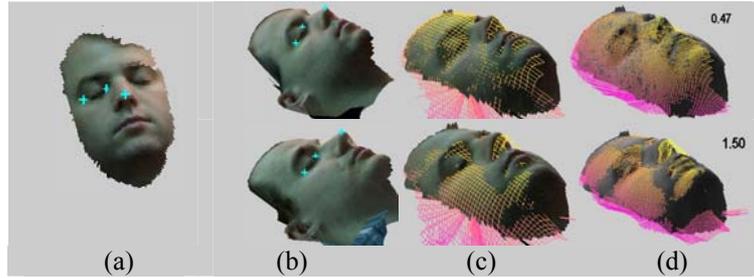


**Fig. 1.** An example for Minolta Vivid 910 facial scan. (a) Texture image; (b) range image, the more red a point, the closer it is to the sensor; (c) 3D rendering of the face scan.

dimensional model could account for a more robust recognition system, however, there is an added cost to construct and store the complete 3D face model.

## 2. Face Alignment and Matching

Our face recognition system is illustrated in Fig. 2.



**Fig. 2.** Matching a given 2.5D face scan to 2.5D face scans in a database. a) Automatic feature point detection; b) the genuine scan and an imposter scan in the database with feature points labeled. c) Coarse alignment between the test scan and the gallery scan. d) Fine iterative registration to find the closest match.

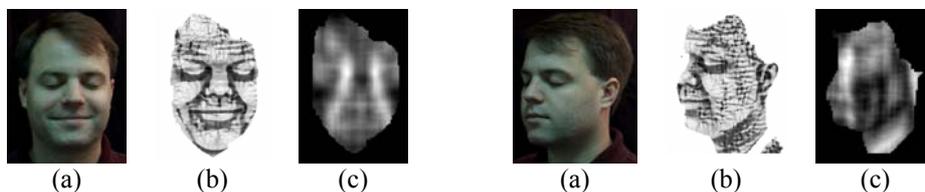
### 2.1 Automatic Feature Point Localization and Coarse Alignment

A minimum of three corresponding points is needed in order to calculate the rigid transformation between two 2.5D scans. Once the three corresponding points are known, the transformation can be made using a combination of rigid transformation matrices following the guidelines described by Weinstein [12].

Using the procedure described by Dorai and Jain [13], we determine the local shape information at each point within the 2.5D image. This shape index at point  $p$  is calculated using the maximum ( $\kappa_1$ ) and minimum ( $\kappa_2$ ) local curvature (see Equation 1). This calculation produces a shape scale between zero and one. A value near one represents a dome shape and a value close to zero represents a cup or dent shape. A value in the middle (0.5) represents a saddle point.

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

This shape calculation is independent of the coordinate system and, therefore, is a potentially useful metric for finding similar points between scans with different poses. The faces shown in Figure 3(b) are examples of 2.5D face images with the grayscale representing the shape index at a particular location, a dark point has a shape index close to zero and a lighter point has a shape index close to one.



**Fig. 3.** (a) Texture Image; (b) shape index; (c) shape index after applying 30x10 averaging mask (compared with (b), the shape index value is inverted for better illustration).

For simplicity, we pick a combination of the inside of an eye, the outside of the eye and the nose tip, which vary, based on the pose of the range image. Of these, the easiest point to identify is the inside edge of an eye right next to the bridge of the nose, because this point (and the area around this point) has a shape index value that is very close to zero. A simple averaging mask of size 30 x 10 can be used to identify this area on the shape space (See Figure 3(c)). We use the inside edge of the eye as an anchor in order to identify other feature points within the face image. Some failures in our system occur when these points are misclassified.

Once the inside eyes are found, other feature points within the face can be located using a simple model of the relative locations between the points. The outside of the eye is detected by following the rut (defined by the shape space) that consistently runs along the bottom of the eye. The mouth tips are found by looking for a valley in the shape space below the eyes. The mouth tips are not robust features because of the movement of the mouth. Despite this, the inside eye, outside eye and a mouth point form a plane that can be used to reliably locate the nose tip which is the farthest point from the plane formed by these three points.

## 2.2 Fine Alignment with Hybrid ICP

After coarse alignment the scans are not registered well due to the localization errors of the feature points. The fine alignment process can be formulated as follows: Given two partially overlapping sets of 3D points,  $P$  and  $Q$ , roughly aligned, find the best (rigid) transformation  $\Psi$  to align them so that the error function  $E = Dist(\Psi(P), Q)$  is minimized, where the  $Dist(.)$  is the distance metric. Our fine registration process follows the Iterative Closest Point (ICP) framework [14-16]. With an initial estimate of the rigid transformation, the algorithm iteratively refines the transform by alternately choosing corresponding (control) points in two scans and finding the best rigid transformation that minimizes an error function based on the distance between them. In Besl and McKay [14], the point-to-point distance is applied and a fast close-form solution is provided when calculating the transformation matrix during each iteration. The point-to-plane distance in Chen and Medioni [15] makes the ICP algorithm less susceptible to local minima than the point-to-point metric [14, 17]. This algorithm shows that if the two meshes are close to each other, the point-to-plane distance is the best approximation for the true distance between the two surfaces [18]. But calculating the point-to-plane distance takes longer than the point-to-point distance. We integrate these two classical ICP algorithms [14, 15] in a zigzag running style, which is called hybrid ICP algorithm. Besl's scheme is used to compute an

estimation of the alignment, followed by Chen’s scheme for a refinement. Since the distance metrics are utilized interactively, it is possible to achieve a better registration result than each individual. The control points are selected only within the overlapping area to avoid outliers. A scheme similar to RANSAC [19, 20] is applied. Each time a number of control points are chosen randomly and ICP is applied based on the points selected. This procedure of control point selection and applying ICP is repeated several times. We pick the alignment with the minimum registration error as the final result.

### 2.3 Matching

In our experiments, the intensity (color), shape index, along with the 3D Cartesian coordinates are used for facial scan matching. One way to define the similarity metric is to combine these attributes into a single feature vector. Each attribute has relatively independent interpretation so they are treated separately in this matching scheme. The  $N$  control points used in the fine registration stage are represented by  $p_i, i=1,2\dots N$ .

- **Surface matching.** Surface matching in fine registration stage provides the registration error, which can be used as a matching score between two facial scans.

$$E_D = \sqrt{\frac{1}{N} \sum_{i=1}^N D(\Psi(p_i), S_i)} \quad (2)$$

where  $\Psi$  is the transformation,  $D(\cdot)$  represent the point-to-plane distance between transformed point  $\Psi(p_i)$  and the tangent plane  $S_i$  at the position of its counterpart in the other scan surface [15].

- **Texture matching.** The 2.5D surface registration from two scans also registers the corresponding color texture maps. Let  $p_i (i=1,2\dots N)$  denote the control points in scan  $P$ , and  $q_i$  be the closest counterpart in scan  $Q$  after the transformation  $\Psi$  is applied to  $P$ . Let  $I(p_i)$  be the texture intensity at  $p_i$ . Then the texture vectors of  $P$  and  $Q$  are represented as  $T_P = \{I(p_i), i=1,2\dots N\}$  and  $T_Q = \{I(q_i), i=1,2\dots N\}$ , respectively. The matching score between two registered texture vectors is calculated as follows,  $MS_T = \langle T_P, T_Q \rangle / (\|T_P\| \cdot \|T_Q\|)$ , where  $T_P$  and  $T_Q$  are the two texture vectors,  $\langle \cdot \rangle$  denotes the dot-product,  $\|\cdot\|$  represents the norm of the texture vector. The  $MS_T$  is normalized to a range from 0 to 1.
- **Shape Index Matching.** Each point in the facial scan is associated with a shape index calculated by Eq. 1. Similar to the texture matching, the matching score based on the shape index attribute can be defined as,  $MS_S = \langle S_P, S_Q \rangle / (\|S_P\| \cdot \|S_Q\|)$ , where  $S_P$  and  $S_Q$  are the two shape index vectors of the control point sets. The  $MS_S$  is also normalized to range from 0 to 1.

The surface matching distance, the registered texture matching score, and the registered shape index matching score can be integrated in a cascade mode or in a combination mode. The range information is more robust than the texture information under varying pose and lighting conditions. The shape index attribute is subject to

estimation errors due to the discretization of the range data. Therefore, the surface matching distance is considered as the primary dissimilarity metric. In the cascade mode, only the test scan, whose surface matching distance is low enough, will proceed to the texture and shape index matching stages. In the combination mode, the texture and shape index matching scores are used to weight the surface matching distance. The final matching distance is computed as:

$$MD = E_D \cdot (1 - MS_T) \cdot (1 - MS_S) \quad (3)$$

The smaller the  $MD$ , the better the matching. The combined matching distance can be considered as a combination of range and texture evidence by the product rule [21].

### 3. Experiments

All range images were collected using a Minolta Vivid 910 scanner [9], which uses structured laser light to record the face scans in less than a second. Each point in a scan has a texture color (r, b, g) as well as a location in 3-space (x, y, z). This data is stored in individual ( $p \times q$ ) matrices. Some of the data points are not valid (e.g., points outside the range of the scanner) so there is also a ( $p \times q$ ) binary flag matrix that is true for all the valid points in the scan. The range image is of size  $480 \times 640$  in the fast scan mode of the Minolta sensor (with a reported depth accuracy of approximately 0.1 mm) and downsampled to  $240 \times 320$  in our experiment to reduce the computation cost. The registered color texture image has the same size as the range image. There are 10 subjects in our collected database. The frontal scan with neutral expression for each subject was used to construct the training database. The test database consists of 63 scans of the same 10 people, which varied in pose (up to left and right 60 degrees from the frontal view), facial expression and lighting.

In order to properly evaluate the performance of the system, it is tested in two experimental modes; automatic and manual. In automatic mode, the algorithm selects the feature points used in coarse registration, where as in the manual mode, the feature points are labeled by hand. The error between the manually selected feature points and the automatically selected feature points is 18.12mm with a standard deviation of 37.58mm. The error is calculated by taking the average distance between the manual labeled points and the automatically extracted points.

The fine alignment and face recognition system has been tested on both manually selected feature points and automatically selected feature points. The recognition accuracy is given in Table 1. Besides the hybrid ICP, two classical ICP algorithms are presented for comparison. Table 1 provides a detailed accuracy for the hybrid ICP, which achieves the best performance among three ICP schemes.

One error out of 15 frontal test scans was achieved when comparing frontal 2.5D face images. This is noteworthy considering that many of the test images had people with different expressions than the training images (i.e., some were smiling). When the testing database was extended to include semi-profile images, the error went up but still maintained a reasonable recognition rate of 5 errors out 63 (92% accuracy). These results assume that the coarse alignment procedure is correctly done. The combined total error of the entire automatic system is 10 out of 63 (84% accuracy).

The hybrid ICP scheme is more robust to the localization errors of feature points, which outperforms the other two classical ICP algorithms on the surface matching (range image only). The combination of surface, texture and shape index matching achieves better performance than surface matching based on the range image alone.

**Table 1.** Number of rank-one failed tests due to recognition. The combination of range, texture and shape index matching is computed as Eq. 3. The number of iterations (10,10) represents a total of 10 iterations for Besl’s ICP scheme and 10 iterations for Chen’s scheme.

Algorithm	Number of Iterations	Feature Point Extraction	Range image only	Range + Texture + Shape Index
ICP_hybrid	(10, 10)	Manual	6	5
ICP_hybrid	(10, 10)	Auto	13	10
ICP_Chen [15]	20	Manual	10	6
ICP_Chen [15]	20	Auto	16	14
ICP_Besl [14]	20	Manual	19	5
ICP_Besl [14]	20	Auto	33	15

#### 4. Conclusions and Future Work

We have presented a face recognition system that is more robust to variations in pose and lighting. This system can match 2.5D scans of arbitrary poses with lighting and expression variations to a database of neutral expression frontal 2.5D scans. Unlike the appearance-based method [22], the proposed matching scheme does not need an extensive training process based on a collected large-scale database.

A Hybrid ICP scheme is proposed to integrate two classical ICP algorithms to achieve better registration performance. Besides the range image, other attributes, such as the registered texture and shape index information, are also used to design a combined metric for the facial scan matching.

This research is an encouraging first step in designing a system that is capable of recognizing faces with arbitrary pose and illumination. There are a number of directions that we are planning to consider in our future work. The first will be to improve the automatic coarse recognition system by incorporating a more complete model of the human face. We also plan on increasing the training database to include profile images. This should make a more robust template set and increase the systems matching capabilities. With a better model we will also consider methods for matching arbitrary two-dimensional training data.

#### Acknowledgement

This research was supported by US Army contract No. DAAD05-03-C-0045. The authors would like to thank Dr. Patrick Flynn and Chris Boehnen for their help on the range data processing. Dr. George Stockman provided helpful suggestions.

## Reference

1. Zhao, W., R. Chellappa, A. Rosenfeld and P. J. Phillips. "Face Recognition: A Literature Survey," *CVL Technical Report*, University of Maryland, Oct. 2000.
2. Gordon, G. "Face Recognition Based on Depth and Curvature Features," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 108-110, 1992.
3. Lee, J. and E. Millios, "Matching Range Images of Human Faces," *Proc. ICCV*, pp. 722-726, 1990.
4. Tanaka, H., M. Ikeda and H. Chiaki. "Curvature-based face surface recognition using spherical correlation," *Proc. Third Int. Conf. on FG*, pp. 372-377, 1998.
5. Chua, C., F. Han and Y. Ho. "3D Human Face Recognition Using Point Signature," *Proc. Fourth Int. Conf. on FG*, pp. 233-238, Grenoble, March 2000.
6. "Face Recognition Vendor Test (FRVT)," <http://www.frvt.org/>.
7. Blanz, V. and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1063-1074, 2003.
8. "Cyberware Inc.," <http://www.cyberware.com/>.
9. "Minolta Vivid 910 non-contact 3D laser scanner," <http://www.minoltausa.com/vivid/>.
10. Beumier, C. and M. Acheroy, "Automatic 3D face Authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 315-321, 2000.
11. Pan, G., Z. Wu and Y. Pan. "Automatic 3D Face Verification From Range Data," *Proc. IEEE ICASSP*, pp. 193-196, 2003.
12. Weinstein, D. M. "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.
13. Dorai, C. 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.
14. Besl, P. and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. PAMI*, vol. 14, no. 2, pp. 239-256, 1992.
15. Chen, Y. and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145-155, 1992.
16. Zhang, Z., "Iterative point matching for registration of free-form curves and surfaces," *International Journal of Computer Vision*, vol. 13, no. 1, pp. 119-152, 1994.
17. Gelfand, N., 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.
18. Pottman, H. and M. Hofer. "Geometry of the squared distance function to curves and surfaces," *Vienna Institute of Technology Technical Report*, No. 90, January 2002.
19. Fischler, M. and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, 1981.
20. Chen, C. S., Y. P. Hung and J. B. Cheung, "RANSAC-based darces: a new approach to fast automatic registration of partially overlapping range images," *IEEE Trans. PAMI*, vol. 21, no. 11, pp. 1229-1234, 1999.
21. Kittler, J., M. Hatef, R. Duin and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.
22. Chang, K. I., K. W. Bowyer and P. J. Flynn. "Multi-Modal 2D and 3D Biometrics for Face Recognition," *Proc. IEEE Workshop on AMFG*, France, October 2003.