

Deformation Analysis for 3D Face Matching

Xiaoguang Lu and Anil K. Jain

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

Abstract

Current two-dimensional image based face recognition systems encounter difficulties with large facial appearance variations due to the pose, illumination and expression changes. Utilizing 3D information of human faces is promising to handle the pose and lighting variations. While the 3D shape of a face does not change due to head pose (rigid) and lighting changes, it is not invariant to the non-rigid facial movement and evolution, such as expressions and aging effect. We propose a face surface matching framework to take into account both rigid and non-rigid variations to match a 2.5D face image to a 3D face model. The rigid registration is achieved by a modified Iterative Closest Point (ICP) algorithm. The thin plate spline (TPS) model is applied to estimate the deformation displacement vector field, which is used to represent the non-rigid deformation. For the purpose of face matching, the non-rigid deformations from different sources are identified, which is formulated as a two-class classification problem: intra-subject deformation vs. inter-subject deformation. The deformation classification results are integrated with the matching distances to make the final decision. Experimental results on a database containing 100 3D face models and 98 2.5D scans with smiling expressions show that the number of errors is reduced from 28 to 18.

1. Introduction

Automatic human face recognition is a challenging task that has gained a lot of attention during the last decade [28]. While most of the efforts have been devoted to face recognition from two-dimensional (2D) images [28], a few approaches have utilized depth information provided by range images [18, 16, 23, 12, 7, 21, 19]. 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 pose, lighting conditions and facial expressions [2]. Because the human face is a three-dimensional (3D) object whose 2D projection (image or appearance) is sensitive to the above changes, utilizing 3D facial information should improve the face recognition performance [8, 2]. Range images captured explicitly by a 3D sensor [1, 3] contain facial

surface shape information. The 3D shape of facial surface represents the facial structure, which is related to the internal anatomical structure instead of external appearance.

The 3D shape of a human face is invariant to the head pose and illumination changes. However, it can still change due to facial expressions and aging factor. The 3D face shape differences for the same subject but with different expressions (i.e., intra-subject difference) can even be larger than that between two different subjects (identities) with the same expression (i.e., inter-subject difference). Therefore, measuring 3D shape difference alone may lead to the ambiguity in face (identity) matching. To resolve this ambiguity, we propose to explicitly estimate and discriminate the shape deformation into two classes for the identity matching purpose, namely, intra-subject deformation and inter-subject deformation.

We use 3D models to recognize 2.5D face (range) images, provided by a range sensor, such as the Minolta Vivid series [3]. Each range image can only provide a single view point of the object, instead of the full 3D view (see Fig. 1).¹ As the 3D imaging technology is progressing quickly [5], non-intrusive 3D data capture will become readily available. The 3D facial structure reconstruction from images has received substantial attention [10, 27, 14]. In real world scenarios, 3D sensors only provide a partial view of the human face. However, during the training stage, 3D face model can be constructed by taking several scans from different viewpoints. Therefore, we address the scenario that matches a 2.5D facial scan to 3D models.

Face recognition based on range images has been addressed in a number of different ways [18, 16, 23, 12, 7, 21, 19]. We propose a matching framework, which captures both rigid and non-rigid deformation, and explicitly classify the non-rigid deformation into intra-subject or inter-subject category. The iterative closest point (ICP) algorithm [6, 11, 26] is applied to achieve the rigid registration. The non-rigid registration is performed by the thin plate spline (TPS) model [9, 15, 17], which generates the displacement vector field as the deformation representation. The displacement vector field is used as the feature representation, which is fed into the deformation classifier. The deformation clas-

¹ In this paper, the range image will be shown by its corresponding intensity image for better illustration.

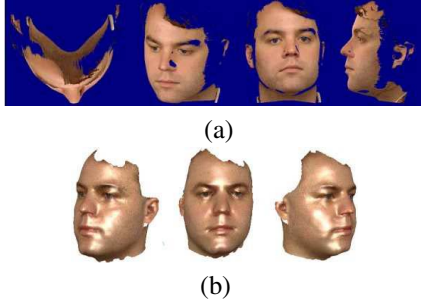


Figure 1. 2.5D scan and 3D face model. (a) One frontal 2.5D scan from different view-points; (b) full 3D model.

sification results are integrated with the matching distances obtained from rigid and non-rigid registration for the final decision making.

2. 3D Face Matching

Our face matching framework is illustrated in Fig. 2. The deformation is analyzed at two levels, rigid and non-rigid. The nonrigid registration also provides the displacement vector field as a deformation representation, which is utilized as the feature vector to classify the deformation into one of two categories, intra-subject and inter-subject classes. Results from the two levels are combined together along with the deformation classification results to make the final matching decision.

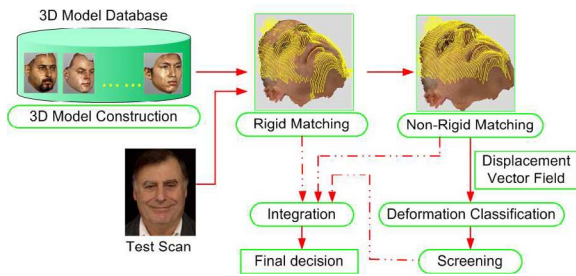


Figure 2. Matching framework with deformation analysis.

2.1. 3D Model Construction

The 3D face model for each subject is constructed by stitching several 2.5D scans obtained from different view points. See [20] for details.

3. Deformation Analysis

In the proposed framework, the deformation analysis includes three modules: rigid registration, nonrigid registra-

tion and deformation classification.

3.1. Rigid Registration

We perform the rigid registration in a coarse-to-fine mode to globally align two surfaces as best as possible. Figure 3 illustrates examples. A feature based alignment

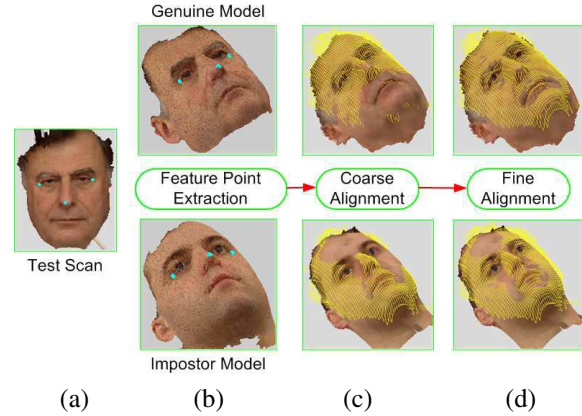


Figure 3. Rigid surface matching. The alignment results are shown by the 3D model overlaid on the wire-frame of the test scan.

is applied for coarse registration for its simplicity and efficiency. Given three corresponding points (anchor points)², the transformation can be made using a combination of rigid transformation matrices following the guidelines described in [25], by achieving a least squares fitting between the triangles formed from the two sets of three anchor points. We pick a combination of the outside corners of the two eyes and the nose tip as our three anchor points. See Fig. 4(a) for an example.

The fine rigid registration process follows the Iterative Closest Point (ICP) framework [6, 11, 26]. 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.

Figure 4(b) shows the grids used for control point selection. 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 results are demonstrated in Fig. 3(d). The root mean square dis-

² In order to evaluate the matching scheme, we study the feature extraction and matching components separately. The coarse alignment is currently performed using manually picked anchor points. Our scheme for automatic feature extraction is described in [13], which can extract anchor points with about 98% accuracy on frontal face scans.

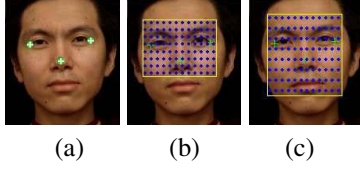


Figure 4. (a) anchor point labeling; (b) automatic control point selection based on three anchor points for rigid registration. (About 100 control points are selected in each scan.) (c) automatic control point selection for non-rigid analysis.

tance minimized by the ICP algorithm is used as the matching distance MD_{Rigid} . For details, see [19].

3.2. Non-rigid Registration

Let g_0 and g_1 denote two surfaces. A warping of g_0 to g_1 is defined as the function F such that

$$F(g_0) = g_1. \quad (1)$$

The function F is called the warping function which takes g_0 to g_1 . Given a pair of point patterns with known correspondences (landmarks) in two surfaces, $U = (u_1, u_2, \dots, u_m)^T$ and $V = (v_1, v_2, \dots, v_m)^T$, where $U \subset g_0$ and $V \subset g_1$, we need to extract correspondence between other surface points; u_k and v_k denote the locations (x, y, z) of the k -th corresponding pair and m is the total number of corresponding points. We assume that the two point pattern sets with known correspondences are aligned as close as possible using a rigid transformation model prior to non-linear warping. Thus, a warping function, F , that warps U to V subject to perfect alignment is given by the conditions

$$F(u_j) = v_j, \quad (2)$$

for $j = 1, 2, \dots, m$.

We use thin-plate splines (TPS) [9, 15] to estimate the deformation F for the point sets (U, V) . TPS represents a natural parametric generalization from rigid to mild non-rigid deformations. The thin plate spline algorithm specifies the mapping of points for a reference set to corresponding points on a target set. The interpolation deformation model is given in terms of the warping function $F(u)$, with

$$F(u) = c + A \cdot u + W^T s(u), \quad (3)$$

where $u \in g_0$; c , A and W are TPS parameters; $s(u) = (\sigma(u - u_1), \sigma(u - u_2), \dots, \sigma(u - u_m))^T$ and $\sigma(r) = |r|$.

An analytical solution of F in two dimensional cases can be derived by minimizing the integral bending energy

$$\int_{R^2} \int_{R^2} ((F_{xx})^2 + 2(F_{xy})^2 + (F_{yy})^2) dx dy. \quad (4)$$

Similar solutions can be obtained in 3D cases [9, 15].

Since TPS can align landmarks perfectly, the root mean square distance calculated on the landmarks will be zero after TPS registration, which is not appropriate to measure the difference between two surfaces after TPS. Therefore, we propose to sample 3 sets of points in g_0 to perform TPS and measure the registration error, namely, two control point sets I and II, and a validation point set. For each control point (u_k) in g_0 , the correspondence is established by finding the closest counterpart (v_k) in the other surface g_1 . These three sets are in the same topology but shift to each other in the $x - y$ plane, as shown in Fig. 5. The two control point set are used for the TPS registration, while the validation set is used to calculate the registration distance. The control point set I is used first to perform one-pass TPS, then the surfaces are deformed by the obtained TPS parameters, and the control point set II. Then use the control point set II for TPS to deform the surface and control point set I. In this way, TPS is performed iteratively to improve the registration results.

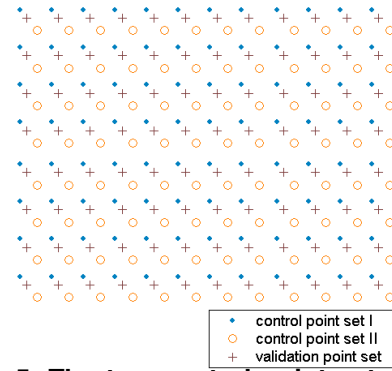


Figure 5. The two control point sets are selected for the TPS registration, while the validation set is used to calculate the registration distance. The control point set I is the same as shown in Fig. 4(c).

The nonrigid registration difference is measured by the root mean square distance between the deformed validation point set after TPS and its counterpart in the other surface. The point-to-point distance metric $MD_{Nonrigid}$ is used, defined as:

$$MD_{Nonrigid} = \sqrt{\frac{1}{N_v} \sum_{i=1}^{N_v} d^2(p_i, q_i)}, \quad (5)$$

where $d(\cdot)$ is the Euclidean distance between two 3D points; q_i in surface g_1 is the closest point to p_i in the deformed surface g_0 ; N_v is the number of points in the validation set.

After the rigid and nonrigid registration, the displacement vector for each validation point can be defined as follows: its tail is the position of the validation point after ICP rigid registration, the head is the position of its closest coun-

terpart in the other surface after TPS nonrigid registration. The displacement vectors construct a field at the validation point set to describe the deformation between the two surfaces. The displacement vector field is denoted by the set $DVF = \{dv_1, dv_2, \dots, dv_{N_v}\}$, where each dv_i is a displacement vector, represented by a triple $(\Delta x, \Delta y, \Delta z)$.

Figure 6 demonstrates an example of the registration and the resulting displacement vector field. It is shown that for the given smiling range scan, when matching it to the model with the same identity (genuine model), as expected, the deformation field magnitude is strong around the cheeks, while low around nose and eye areas. When matching the given range scan to the model with different identity (impostor model), the deformation field is more noisy and some strong deformation is obtained near the nose and eyes. Notice that, even registering a range scan to the model with different identity, the TPS has the capability to make the alignment much better than the rigid ICP registration. Therefore, matching, which is based on the registration difference only, may result in some ambiguity. The deformation classification can be helpful to solve this ambiguity.

3.3. Deformation Classification

The nonrigid deformation between two surfaces is represented by the displacement vector field (DVF). We concatenate the displacement vectors to construct a feature vector for classification. Given a $DVF = \{dv_1, dv_2, \dots, dv_{N_v}\}$, where $dv_k = (x_k, y_k, z_k)$ obtained by matching the i -th range image to the j -th model, the feature vector is constructed as $DF_{i,j} = [x_1 y_1 z_1 x_2 y_2 z_2 \dots x_{N_v} y_{N_v} z_{N_v}]^T$.

The deformation classification classifies each DF into one of two categories: intra-subject and inter-subject. Due to the large intra-class variability, this is not a linearly separable problem. Therefore, we applied a nonlinear classifier, Support Vector Machine (SVM) [24]. In order to classify the derived feature vector, each vector should have the same length for SVM. One problem of our sensor (laser-based) is that it cannot capture dark areas in the face, such as some eyebrows. Therefore, at some points in the validation set (which are automatically selected), there may be missing values in the derived displacement vector field. To handle this incomplete data problem, we replace the missing value with the sample mean computed from the displacement vectors that contain the effective values at the corresponding positions. More complicated algorithm to handle missing values, such as multiple imputation [22], could be applied.

3.4. Face Matching with Deformation Analysis

It is difficult to collect images of all types of facial expression variations for a large number of subjects. A small training set may provide only a low deformation classification accuracy. Therefore, we considered the deformation classifier as a weak classifier, which is used as a screening step to reduce the number of false alarms. In other words,

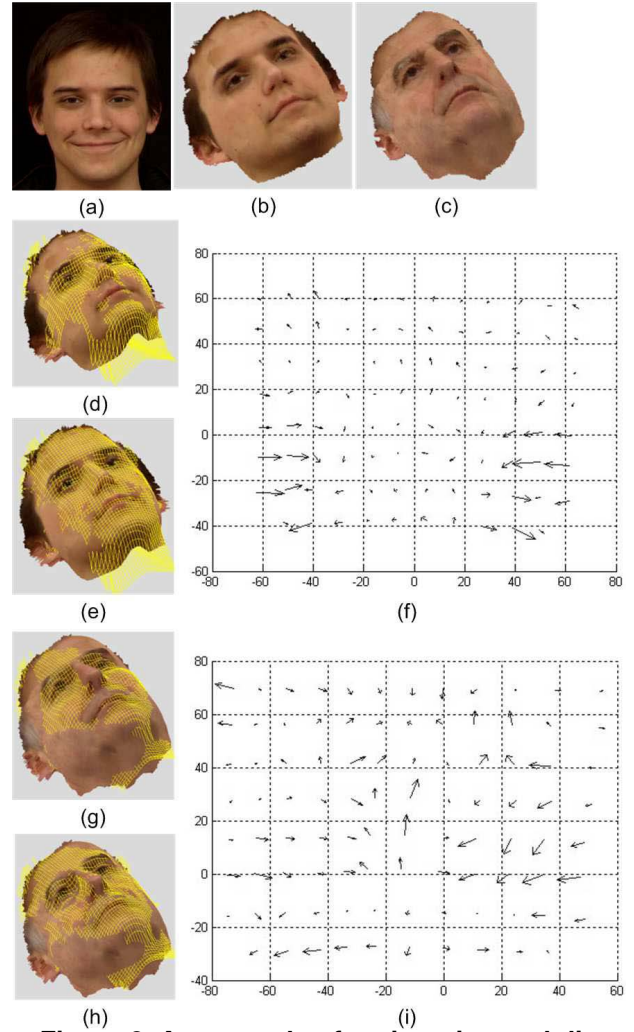


Figure 6. An example of registration and displacement vector field generation. (a) input range scan; (b) the 3D face model with the same identity as the range scan (genuine model); (c) a 3D model with different identity (impostor model); (d,g) rigid (ICP) registration results; (e,h) nonrigid (TPS) registration results; (f,i) corresponding displacement vector fields from smiling to neutral.

when the deformation between an input range image and a 3D model has a high matching score to be assigned to the inter-subject category, this image will not be forwarded proceed to the next stage (registration distance matching stage).

To integrate the registration results of rigid and nonrigid registrations, their matching distances are combined by the sum rule as follows:

$$MD_{comb} = MD_{Rigid} + MD_{Nonrigid}. \quad (6)$$

The rigid and nonrigid matching are also run in a cascade mode. Only those test scans for which the rigid matching does not have sufficient evidence to make the decision are fed to the nonrigid registration stage. Since the nonrigid registration requires additional computational burden, this cascading framework is important in real applications. In our current implementation, if the rigid matching distance (MD_{Rigid}) is below a pre-defined threshold δ , then it is considered as a good surface matching. Since the surface matching distance is measured by the root mean square distance among control points, it has a physical meaning. We choose δ equal to $1mm$. The value of δ depends on the noise level of the scans and the performance of the anchor point locator.

4. Experiments and Discussion

All range images (downsampled to 320×240 with a depth resolution of $\sim 0.1mm$) were collected using a Minolta Vivid 910 scanner [3] in our laboratory. The subject stands in front of the scanner at a distance of about 1.5m. This scanner uses structured laser light to construct the face range image in less than a second. Each point in a scan has a texture color (r, g, b) as well as a location in 3D space (x, y, z). Each facial scan has around 18,000 effective points (excluding the background).

There are currently 100 subjects in our database. Five scans with neutral expression for each subject were captured to construct the 3D model. For each subject, two additional frontal scans are captured for testing, one with neutral expression and one with smiling expression. Due to the sensor failure, four scans were not successfully captured. In total, the test database consists of 196 independent scans (different from training scans) of the same 100 people stored in the gallery database. The representative 3D models and test scans are shown in Figs. 7 and 8, respectively.

In order to learn the deformation and train the classifier, another 10 volunteers (not in the 100 subject database) were asked to provide 5 training scans for building the 3D model and two independent scan to learn the intra-subject deformation variation (one neutral and one smiling). In total, we have 20 independent scans and 10 3D models to learn the deformation. This results in 20 intra-subject and 180 inter-subject deformation field vectors. Each scan has 96 3D validation points, so the dimensionality of the concatenated feature vector is $96 \times 3 = 288$. Utilizing this small amount of training data, a SVM classifier is trained. The SVM^{light} package [4] is used as the implementation of the classifier in the experiments. In our experiments, we are working at a 1% false accept rate and 34% false reject rate. While this classification accuracy is not good, it is more important to test if this deformation classification information is helpful for the matching purpose.

A summary of the matching results is given in Table 1. Among the 30 errors (out of 196 test scans) after rigid matching, 28 correspond to test scans with smiling expression, indicating that the facial expression poses a significant difficulty when comparing 3D facial surfaces. A com-



Figure 7. Some of the 3D face models in the database.



Figure 8. Representative 2.5D test scans.

bination of rigid and nonrigid deformation improves the performance. Deformation classification screening further improves the performance. In addition, combining the surface matching with constrained appearance-based matching, which is proposed in [20], results in higher matching accuracy. Figure 9 shows some of the test scan that were correctly recognized.

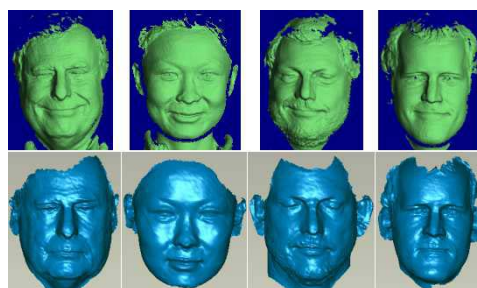


Figure 9. Test scans (top row), and the corresponding 3D neutral models correctly matched (bottom row). For illustration purpose, the texture is removed from both test scans and 3D models in order to enhance the surface property.

5. Conclusions and Future Work

Face matching is addressed in a deformation analysis framework. We have presented a face recognition system that matches 2.5D range images of faces with pose, lighting and expression variations to a database of 3D face models. The proposed matching system takes into account both rigid and non-rigid face shape variations. The rigid registration is achieved by the Iterative Closest Point algorithm. The thin plate spline model is applied to estimate the deformation displacement vector field, which is used to represent the non-rigid deformation between a range image and a 3D model. For the face matching purpose, the non-rigid deformations from two different sources are discriminated,

Matching scheme	Rank-one matching accuracy		
	Neutral test scans	Smiling test scans	Total
I	98% (# of errors = 2)	71% (# of errors = 28)	85% (# of errors = 30)
I+II	99% (# of errors = 1)	74% (# of errors = 25)	87% (# of errors = 26)
I+III+II	100% (# of errors = 0)	78% (# of errors = 22)	89% (# of errors = 22)
I+III+II+IV	100% (# of errors = 0)	82% (# of errors = 18)	91% (# of errors = 18)

Table 1. Matching accuracy. The total number of test scans is 196, including 98 scans with neutral expression and 98 scans with smiling expression. Scheme I: rigid (ICP); II: nonrigid (TPS); III: deformation classification screening; IV: appearance-based [20].

namely, intra-subject deformation vs. inter-subject deformation. The deformation classification results are integrated with the registration distances for making the final matching decision. A small training data set is used for deformation classification, which is shown to improve the matching performance.

The proposed framework is a general scheme to separate intra-subject and inter-subject deformations, not limited to any particular expression, such as smiling expression considered here. We plan to collect more data with facial shape variations to better discriminate between inter-subject and intra-subject deformation. More sophisticated surface registration techniques are being pursued to improve the non-rigid deformation estimation.

References

- [1] Cyberware Inc. <<http://www.cyberware.com/>>.
- [2] Face Recognition Vendor Test (FRVT). <<http://www.frvt.org/>>.
- [3] Minolta Vivid 910 non-contact 3D laser scanner. <<http://www.minoltausa.com/vivid/>>.
- [4] *SV M^{light} package*. <<http://svmlight.joachims.org/>>.
- [5] The 4th International Conference on 3-D Digital Imaging and Modeling (3DIM). <<http://www.3dimconference.org/>>, 2003.
- [6] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [7] C. Beumier and M. Acheroy. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.
- [8] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
- [9] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.
- [10] Q. Chen and G. Medioni. Building 3-D human face models from two photographs. *Journal of VLSI Signal Processing*, 27:127–140, 2001.
- [11] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.
- [12] C. Chua, F. Han, and Y. Ho. 3D human face recognition using point signature. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 233–238, Grenoble, Mar. 2000.
- [13] D. Colbry, X. Lu, A. Jain, and G. Stockman. 3D face feature extraction for recognition. Technical Report MSU-CSE-04-39, Computer Science and Engineering, Michigan State University, East Lansing, Michigan, September 2004.
- [14] M. Dimitrijevic, S. Ilic, and P. Fua. Accurate face models from uncalibrated and ill-lit video sequences. In *Proc. IEEE CVPR*, pages 1034–1041, Washington, DC, 2004.
- [15] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.
- [16] G. Gordon. Face recognition based on depth and curvature features. In *Proc. IEEE CVPR*, pages 108–110, 1992.
- [17] T. Hutton, B. Buxton, and P. Hammond. Automated registration of 3D faces using dense surface models. In *Proc. British Machine Vision Conference*, pages 439–448, 2003.
- [18] J. Lee and E. Milios. Matching range images of human faces. In *Proc. ICCV*, pages 722–726, 1990.
- [19] X. Lu, D. Colbry, and A. Jain. Three-dimensional model based face recognition. In *Proc. International Conference on Pattern Recognition*, pages 362–366, Cambridge, UK, 2004.
- [20] X. Lu and A. Jain. Integrating range and texture information for 3D face recognition. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, Breckenridge, CO, 2005.
- [21] G. Pan, Z. Wu, and Y. Pan. Automatic 3D face verification from range data. In *Proc. ICASSP*, volume 3, pages 193–196, 2003.
- [22] D. Rubin. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91:473–489, 1996.
- [23] H. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 372–377, 1998.
- [24] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
- [25] 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*, March 1998.
- [26] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(1):119–152, 1994.
- [27] Z. Zhang. Image-based modeling of objects and human faces. In *Proc. of SPIE*, volume 4309, pages 1–15, Jan. 2001.
- [28] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *CVL Technical Report, University of Maryland*, Oct. 2000. <<ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz>>.