Deformation Modeling for Robust 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

Human face recognition based on 3D surface matching is promising for overcoming the limitations of current 2D image-based face recognition systems. The 3D shape is invariant to the pose and lighting changes, but not invariant to the non-rigid facial movement, such as expressions. Collecting and storing multiple templates for each subject in a large database (associated with various expressions) is not practical. We present a facial surface modeling and matching scheme to match 2.5D test scans in the presence of both non-rigid deformations and large pose changes (multiview) to a neutral expression 3D face model. A geodesic-based resampling approach is applied to extract landmarks for modeling facial surface deformations. We are able to synthesize the deformation learned from a small group of subjects (control group) onto a 3D neutral model (not in the control group), resulting in a deformed template. A person-specific (3D) deformable model is built for each subject in the gallery w.r.t. the control group by combining the templates with synthesized deformations. By fitting this generative deformable model to a test scan, the proposed approach is able to handle expressions and large pose changes simultaneously. Experimental results demonstrate that the proposed matching scheme based on deformation modeling improves the matching accuracy.

1. Introduction

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 [1]. Since human face is a three-dimensional (3D) object whose 2D projection (image or appearance) is sensitive to the above changes, utilizing 3D facial information can improve the face recognition performance [5].

Face recognition based on range images has been investigated by a number of researchers [13, 10, 21, 4, 17], but only a few of them have addressed the deformation (expression) issue. Chua et al. [8] extended the use of Point Signature to recognize frontal face scans with different expressions, which was treated as a 3D recognition problem of non-rigid surfaces. Bronstein et al. [7] proposed an algorithm based on geometric invariants, in an attempt to deal with facial expression variations in 3D face recognition, again for frontal face scans, and the proposed algorithm assumed that the mouth was closed in all facial expressions.

We address the problem of matching *multiview* 2.5D facial scans (range images) in the presence of expression variations to 3D face models (or 2.5D facial scans) with neutral expression. To account for the large intra-subject difference in 3D shapes caused by expression changes, we propose to explicitly model the 3D deformation. Gross et al. [11] showed that person-specific deformable model is more robust than the generic deformable model (across subjects). However, to build a person-specific deformable model, a large number of training samples for a user are needed; collecting and storing 3D data of each subject in a large gallery with multiple expressions is not practical. Further, it is difficult to collect face scans to cover all possible variations even for the same type of expression, because the expression deformation is a continuous facial movement.

We collect data on 3D facial deformations from only a small group of subjects, i.e., the control group. The extracted deformations from the control group are transferred to and synthesized for all the 3D neutral face models in the gallery, yielding deformed templates with synthesized expressions. Multiple deformed templates for the same subject based on members in the control group are combined to build deformable models for each subject in the gallery.

Our deformation transfer and synthesis falls under the performance-driven framework [22, 19, 16, 20]. Unlike previous methods designed for realistic animation, we simplify the deformation transfer problem and provide a reasonable approximation for 3D matching. Besides the fiducial fa-

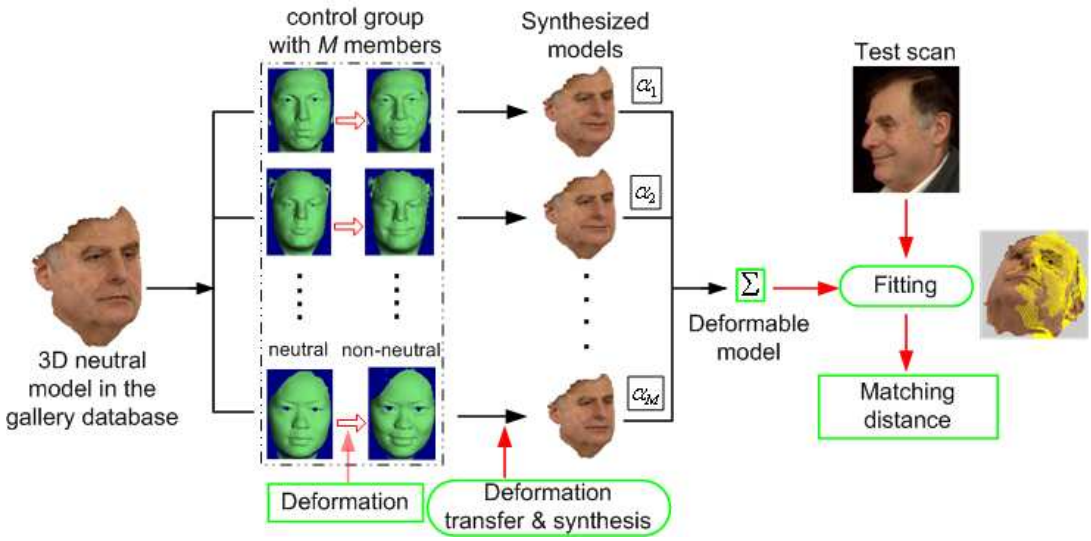


Figure 1. Deformation modeling for 3D face matching.

cial landmarks, such as eye and mouth corners, a geodesic-based surface resampling approach is applied to extract landmarks in the facial area with little texture, e.g., cheeks. We use the thin-plate-spline (TPS) mapping to transfer the landmark-based deformation. The deformation synthesis is also driven by TPS to interpolate the new positions of surface vertices in-between the landmarks.

In matching a test scan to a 3D face model in the gallery, the person-specific deformable model is fitted to the test scan by solving an optimization problem to yield a matching distance. The proposed scheme is designed to handle both pose and expression changes simultaneously.

2. Deformation Modeling for Matching

The proposed scheme of deformation modeling for 3D face matching is presented in Fig. 1.

2.1. Landmark Extraction

To derive the 3D surface deformation, we use facial landmarks to establish the correspondence [19, 16]. We manually labeled the fiducial landmarks in 3D facial surfaces, i.e., the nose tip, eye corners, and mouth corners, along with the mouth contour. For those facial regions that have little texture but are important for expression modeling in 3D, such as the cheeks, we extract landmarks by resampling the facial surface based on geodesics, which has been demonstrated to be insensitive across expressions [7]. The geodesic distance and the corresponding path between two fiducial landmarks (e.g., from one eye corner to one mouth corner) on the facial surface are computed based on the fast marching algorithm [12]. The derived paths encode the facial surface movement according to different expressions. We divide each path into L segments with equal geodesic length (L is 8 in our experiments). The segmenting vertices are then used as the newly extracted landmarks as shown in Fig. 2. Note that each landmark is represented by (x, y, z) coordinates.

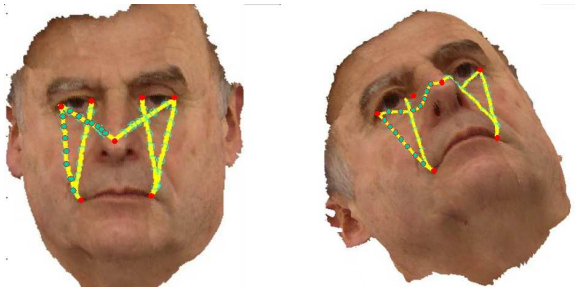


Figure 2. Examples of landmark extraction using surface resampling along the geodesic paths. Two paths are overlaid with resampled landmarks (green dots) in this example for illustration.

2.2. Deformation Transfer and Synthesis

The deformation is learned from a control group of M subjects, who provide both neutral and non-neutral expression scans. The learned deformation is transferred to a 3D neutral model in the gallery for synthesis, according to the following procedure, which is illustrated in Fig. 3.

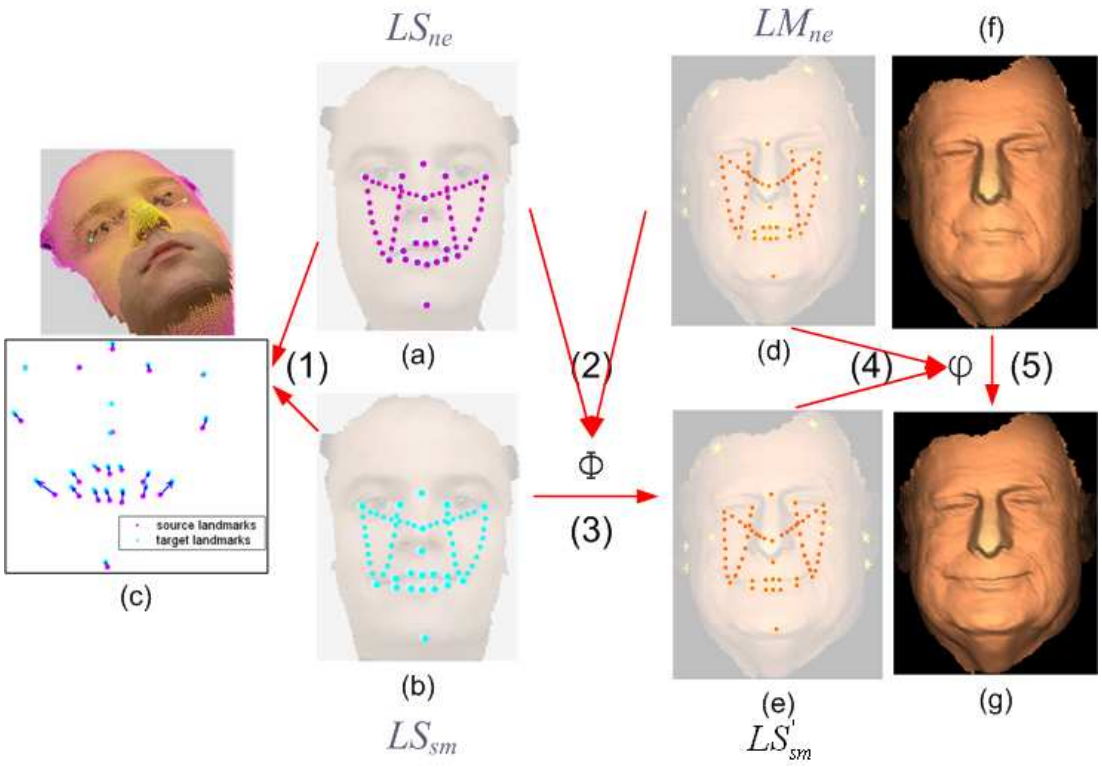


Figure 3. Deformation transfer and synthesis. (a) Landmark set (LS_{ne}) of the neutral scan in the control group. (b) Landmark set (LS_{sm}) of the scan with non-neutral expression in the control group. (c) Deformation field of the landmarks from (a) to (b) after the rigid alignment. (d) Landmark set (LM_{ne}) of the 3D neutral model (f) in the gallery. (e) Landmark set (LS'_{sm}) after deformation transfer. (g) 3D model after applying deformation transfer and synthesis on (f).

(1) Register the non-neutral scan with the neutral scan to estimate the displacement vector of landmarks due to the expression change.

(2) Establish a mapping ϕ from the landmark set (LS_{ne}) of the neutral scan to that (LM_{ne}) of the 3D model;

(3) Use ϕ to transfer the landmarks (LS_{sm}) in the non-neutral scan to the 3D model as LS'_{sm}.

(4) Establish a mapping φ from the landmarks (LM_{ne}) of the 3D neutral model to LS'_{sm}.

(5) Apply φ to other vertices in the 3D neutral model to move them to the new positions caused by the expression.

We use TPS as the mapping and interpolation tool for deformation transfer and synthesis.

2.2.1 Thin-Plate-Spline

Given a pair of point patterns with known correspondences (landmarks) on two surfaces, $U = (u_1, u_2, \cdots, u_m)^T$ and $V = (v_1, v_2, \cdots, v_m)^T$, we need to extract correspondence between other surface points; u_k and v_k denote the (x, y, z) coordinates of the k-th corresponding pair and m is the total number of corresponding points. A warping function, F, that warps U to V subject to perfect alignment is given by the conditions

$$F(u_j) = v_j, \tag{1}$$

for $j = 1, 2, \cdots, m$. The interpolation deformation model is given in terms of the warping function $F(u)$, with

$$F(u) = c + A \cdot u + W^T s(u), \tag{2}$$

where $u \in g_0$; c, A and W are TPS parameters; $s(u) = (\sigma(u - u_1), \sigma(u - u_2), \cdots, \sigma(u - u_m))^T$ and $\sigma(r) = |r|$. An analytical solution of F can be obtained for 3D points [6, 9].

2.2.2 Deformation Transfer

The deformation transfer can be formulated as follows: given a pair of source surfaces represented by meshes, S and S', and a target mesh T, generate a new mesh T' such

that the relationship between T and T' is similar to the relationship between S and S'. Our deformation transfer is based on extracted landmarks. Figure 3(a) shows the protocol of landmark labeling on the face scans. The same set of landmarks is labeled on the 3D neutral model for deformation transfer (see Fig. 3(d)).

In order to separate non-rigid facial expressions from rigid head motion, a rigid transformation (translation and rotation), is applied to align the neutral scan and the non-neutral scan based on those landmarks that move very little due to expression changes, such as eye corners and nose tip. This normalizes the facial (geometry) position, (see Fig. 3(c)). After the rigid alignment of neutral and non-neutral scans, the estimated displacement vectors need to be transferred to the 3D neutral model. Since facial geometry and aspect ratios are different between the scans in the control group and the 3D models in the gallery, source displacements cannot be simply transferred without adjusting the direction and magnitude of each motion vector. We establish a TPS mapping from the landmark set of the neutral scan in the control group to that in the 3D model. Since the TPS mapping contains the affine component and the distortion component, both the scale and orientation of the motion vectors are also adjusted. The landmarks for the non-neutral scans are mapped onto the corresponding positions in the coordinate system of the 3D model by applying the estimated TPS mapping.

2.2.3 Deformation Synthesis

The estimated TPS is applied as an interpolation approach to obtain an approximate movement for the surface vertices in-between the landmarks. For the vertices in-between the convex hull spanned by these points, the interpolation can be done by TPS. However, for those vertices which lie outside this convex hull (e.g., vertices in-between the dots and stars in Fig. 4(a)), an extrapolation has to be performed, leading to distortions, such as in Fig. 4(c). Therefore, we add a few additional landmarks, which specify the boundary constraints. These landmarks are mapped to themselves. By computing the TPS based on this augmented landmark set (dots plus stars in Fig. 4(a)), the interpolation can generate a better synthesis result as shown in Fig. 4(c).

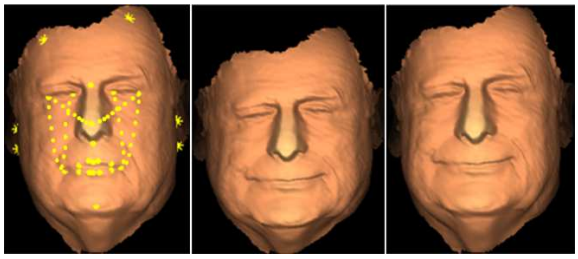
2.3. Deformable Model Construction

Facial expression change is a continuous motion process, while a synthesized template (model) captures only a single frame. Further, since each single synthesized template is obtained by transferring the deformation from one member in the control group, it is not likely to be the true expression of the gallery model. Therefore, a more general expression deformation is learned from all M members in the control group. This leads to a person-specific deformable model that is a linear combination of multiple deformed templates (models), each obtained as a result of deformation transfer from the members of the control group.

We use a shape vector S_0 to represent each surface model: $S_0 = (x_1, y_1, z_1, \cdots, x_n, y_n, z_n)^T$, where each triple (x_k, y_k, z_k) is the location of the surface vertex k, and n is the total number of vertices. For each subject, let S_{ne} denote the original neutral model and S_i be the deformed template with the same type of expression synthesized from S_{ne}. Notice that since all S_i's are synthesized from S_{ne}, the correspondence between them is automatically established. The deformable model for this subject is constructed as

$$S = S_{ne} + \sum_{i=1}^{M} \alpha_i \cdot (S_i - S_{ne}), \tag{3}$$

where M is the total number of synthesized templates from S_{ne} and α_i's are the weights. The deformable model consists of two components, the first component is the neutral model S_{ne} and the second is the variation component represented as a linear combination of model differences. S_{ne} is used to control the identity, whereas the variation component is used for the deformation adaptation by adjusting the weights α_i.



(a) (b) (c)

Figure 4. Deformation synthesis. (a) 3D neutral model with landmarks. The dots are the landmarks in correspondence to those in the control group (see Fig. 3(a)). The star points are used for boundary constraints. (b) Synthesis result without fixed-point boundary constraints. (c) Synthesis result with fixed-point boundary constraints.

In principle, the synthesized models of different expressions can be integrated into a single deformable model by adding new linear components in Eq. 3, but this increases the complexity of the model. Therefore, currently for each subject, we construct one deformable model for each type of expression.

2.4. Robust 3D Face Matching

Two types of transformations are applied to a 3D deformable model, when it is matched to a given test scan with a claimed identity: (i) rigid transformation due to the head pose changes, which can be represented by a rotation matrix and a translation vector; (ii) non-rigid deformation, which can be modeled by the weights α_i in Eq. 3. Fitting the deformable model to a given test scan is formulated as an optimization problem to minimize the cost function

$$\begin{aligned} E &= \|(R \cdot S + T) - S_t\|^2 \\ &= \|R \cdot (S_{ne} + \sum_{i=1}^{M} \alpha_i \cdot (S_i - S_{ne})) + T - S_t\|^2, \end{aligned} \quad (4)$$

where R and T are the rotation matrix and translation vector, respectively; S is the 3D deformable model (weights α_is are embedded) and S_t denotes the test scan. To reduce the computation cost in the optimization process, we subsample the test scan surface.

We factorize the rigid and nonrigid components and solve for them using the following iterative procedure:

1. Initialize the deformable model parameters to generate a 3D model; estimate a coarse alignment between the model and the test scan using three pairs of points.
2. The iterative closest point (ICP) algorithm is utilized to solve for the rotation and translation parameters (R, T) [3], while fixing α_i's.
3. Given R and T obtained in step 2, minimize the cost function E by solving for α_i's.
4. Use the α_i's computed in step 3 to generate a new instance of the 3D model; repeat steps 2 to 4 until the convergence is reached.

After the fitting process, the root-mean-square distance calculated by the ICP algorithm is used as the matching distance. A fitting example is provided in Fig. 5. Since each subject has multiple deformable models for different expressions, for each subject, we match all its deformable models of different expressions to a given test scan. The minimum of all the obtained matching distances is used as the final matching distance.

3. Experiments and Discussion

The proposed matching scheme is evaluated on three databases. Since there is no publicly available 3D facial scan database containing simultaneous expression and (large) pose changes, we collected two databases (I and II) in our lab. All the range images (downsampled to 320×240 with a depth resolution of $\sim 0.1mm$) were collected using a Minolta Vivid 910 scanner [2]. To build the 3D gallery models, for each subject, five scans (different from the scans

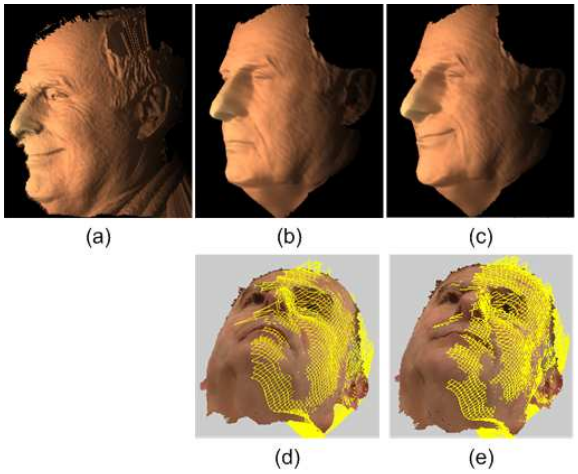


Figure 5. Deformable model fitting. (a) Test scan. (b) 3D neutral model. (c) Deformed model after fitting to (a). Registration results of (a) to models (b) and (c), are given in (d), (e), respectively (the test scan (yellow wire-frame) is overlaid on the 3D model); the matching distances are 2.7 and 1.3, respectively.

used to evaluate the matching performance) with neutral expression were captured at different viewpoints and stitched to construct the full view 3D neutral model using a commercial software [15]. Database III is built from FRGC Ver.2.0 benchmark [18], where both gallery models and test scans are 2.5D frontal scans. We evaluate the proposed scheme on these three databases in an identification mode, matching a test scan to all the gallery models. In Step 1 of the modeling fitting procedures (see Sec. 2.4), we use three feature points (two eye corners and the nose tip) to initialize a coarse alignment [15]. In order to evaluate the proposed deformation modeling scheme without introducing feature extraction errors, three manually labeled feature points are applied. An automatic feature point extraction algorithm has been proposed in [14].

3.1. Experiment I

Database I contains range images of 10 subjects at 3 different poses (frontal, left 30 degrees, left 60 degrees) with 7 different expressions, which are neutral, happy, angry, smile, surprise, deflated, inflated [7]. The data collection protocol for one subject is provided in Fig. 6. In total, there are 210 ($3 \times 7 \times 10$) scans and 10 3D gallery models. Five subjects are randomly chosen as the control group and the remaining 5 subjects are used as the gallery. There are 105 ($5 \times 7 \times 3$) test scans in total. For the subjects in the control

	Mean	Std
Without deformation modeling	87.6%	3.6%
With deformation modeling	92.1%	3.4%

Table 1. Identification accuracy of 10-fold cross-validation in experiment I.

group, only frontal scans are used for deformation modeling. The recognition accuracy based on 10-fold cross validation is provided in Table 1.

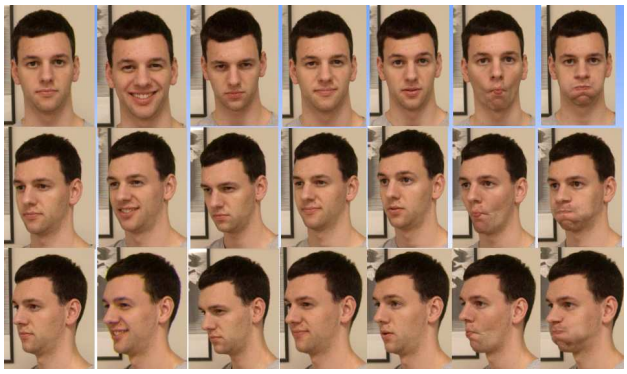


Figure 6. Data collection for experiment I (7 expressions at 3 poses).

3.2. Experiment II

The control group is composed of the 10 subjects in database I. Another 90 subjects formed the gallery. For each subject in the control group, only frontal scans are used to learn and transfer the deformation. Another six scans were captured for each subject in the gallery for testing at different viewpoints, including 3 scans with neutral expression and 3 scans with smiling expression. So, there are a total of 90 3D models stored in the gallery and 533 independent 2.5D scans for testing (for a few subjects fewer than 6 test scans are available). The representative test scans are shown in Fig. 7. The CMC curves are provided in Fig. 8.

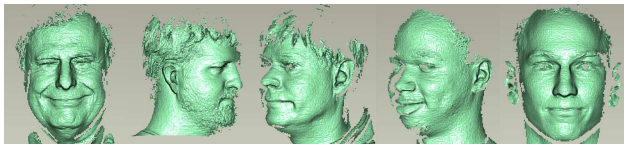


Figure 7. Test scan examples in database II.

3.3. Experiment III

Preliminary experiments on a subset of FRGC Ver2.0 dataset are conducted. FRGC dataset contains only (near)

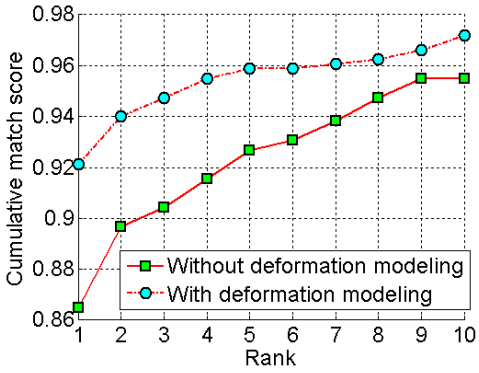


Figure 8. CMC curves of experiment II.

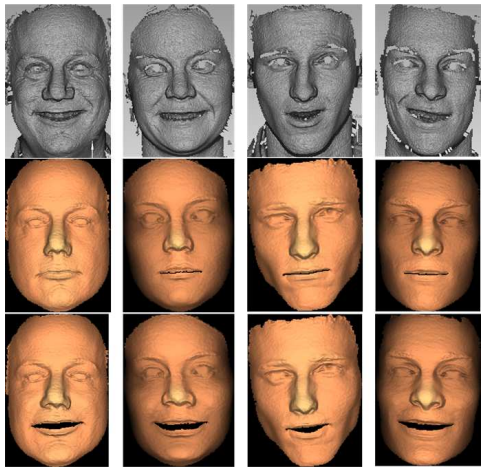


Figure 9. Examples of test scans (top row) that are incorrectly identified without deformation modeling but correctly identified with deformation modeling. Middle row: corresponding genuine 2.5D neutral templates; bottom row: deformed templates after modeling fitting.

frontal 2.5D facial scans and no 3D models are available. But, the proposed deformation modeling scheme is still applicable. 50 subjects are randomly selected. For each subject, the scan with neutral expression and the earliest time stamp is used as the template; another 3 scans of the subject (one neutral, one smiling, and one surprise) are chosen as test scans. In total, there are 50 2.5D gallery scans and 150 independent 2.5D test scans for testing. The 10 subjects in database I formed the control group, based on which the expression deformation is learned and the deformable model (a 2.5D frontal template) is constructed for each subject. The rank-1 identification accuracy is 97% with deformation modeling scheme integrated, compared to 81% without de-

formation modeling. Fig. 9 shows examples that are incorrectly matched without deformation modeling but correctly matched by integrating the proposed deformation modeling scheme.

3.4. Discussion

These experimental results demonstrate that the proposed deformation modeling scheme improves the matching accuracy in the presence of expression variations along with large pose changes. One possible reason for the matching errors is that the current model fitting process is subject to local minimum. In addition, in our experiments, the control group only contains 10 subjects, which is not adequate to cover all variations of the same type of expression across a large population.

Each fitting (matching) of the deformable model to a test scan takes ~ 5 seconds on a Pentium 4 2.8GHz CPU. Current Matlab®-based implementation along with the algorithm is being optimized to reduce the computational cost for practical applications.

4. Conclusions and Future Work

We have proposed a framework for robust 3D face matching in the presence of nonrigid deformation (due to expression changes) and large pose changes simultaneously in the test scan. Landmarks in facial surfaces in regions with little texture are automatically extracted using the geodesic-based approach. 3D deformation learned from a small control group is transferred to the 3D models with neutral expression in the gallery. The corresponding deformation is synthesized in the 3D neutral model to generate a deformed template. A person-specific deformable model is built by combining the deformed templates from each member in the control group. The matching is performed by fitting the deformable model to a given test scan, which is formulated as a minimization of a cost function. Experimental results demonstrate the capabilities of the proposed scheme to learn and synthesize the deformation on new face models and to make the 3D face surface matching system more robust across expression and pose.

Landmark labeling is needed in deformation modeling. Currently, fiducial landmark labeling is done manually. Although this is conducted in the offline training stage, it would be more desirable to make it a fully automatic process in many applications. Reducing the computational cost is also being pursued.

References

[1] *Face Recognition Vendor Test (FRVT).* <http://www.frvt.org/>.

[2] *Minolta Vivid 910 non-contact 3D laser scanner.* <http://www.minoltausa.com/vivid/>.

[3] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. PAMI*, 14(2):239–256, 1992.

[4] C. Beumier and M. Acheroy. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.

[5] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. PAMI*, 25(9):1063–1074, 2003.

[6] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. PAMI*, 11:567–585, 1989.

[7] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant 3D face recognition. In *Proc. AVBPA*, pages 62-70, Guildford, UK, 2003.

[8] 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.

[9] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.

[10] G. Gordon. Face recognition based on depth and curvature features. In *Proc. CVPR*, pages 108–110, 1992.

[11] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(11):1080–1093, 2005.

[12] R. Kimmel and J. A. Sethian. Computing geodesic paths on manifolds. *Proc. Natl. Acad. Sci. USA*, 95:8431–8435, 1998.

[13] J. Lee and E. Milios. Matching range images of human faces. In *Proc. ICCV*, pages 722–726, 1990.

[14] X. Lu and A. K. Jain. Automatic feature extraction for multiview 3D face recognition. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Southampton, UK, 2006.

[15] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Trans. PAMI*, 28(1):31–43, 2006.

[16] J. Noh and U. Neumann. Expression cloning. In *Proc. ACM SIGGRAPH*, pages 277–288, 2001.

[17] G. Pan, Z. Wu, and Y. Pan. Automatic 3D face verification from range data. In *Proc. ICASSP*, volume 3, pages 193–196, 2003.

[18] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. CVPR*, pages 947–954, San Diego, CA, 2005.

[19] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. Salesin. Synthesizing realistic facial expression from photographs. In *Proc. ACM SIGGRAPH*, pages 75–84, 1998.

[20] R. Sumner and J. Popovic. Deformation transfer for triangle meshes. In *Proc. ACM SIGGRAPH*, pages 399–405, Aug. 2004.

[21] 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.

[22] L. Williams. Performance-driven facial animation. In *Proc. ACM SIGGRAPH*, pages 235–242, 1990.