# Deformation Modeling for Robust 3D Face Matching

Xiaoguang Lu, *Member*, *IEEE*, and Anil K. Jain, *Fellow*, *IEEE*

**Abstract**—Face recognition based on 3D surface matching is promising for overcoming some of the limitations of current 2D image-based face recognition systems. The 3D shape is generally invariant to the pose and lighting changes, but not invariant to the nonrigid facial movement such as expressions. Collecting and storing multiple templates to account for various expressions for each subject in a large database is not practical. We propose a facial surface modeling and matching scheme to match 2.5D facial scans in the presence of both nonrigid deformations and pose changes (multiview) to a stored 3D face model with neutral expression. A hierarchical 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 user-specific (3D) deformable model is built for each subject in the gallery with respect to 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 pose changes simultaneously. A fully automatic and prototypic deformable model based 3D face matching system has been developed. Experimental results demonstrate that the proposed deformation modeling scheme increases the 3D face matching accuracy in comparison to matching with 3D neutral models by 7 and 10 percentage points, respectively, on a subset of the FRGC v2.0 3D benchmark and the MSU multiview 3D face database with expression variations.

**Index Terms**—Deformation modeling, 3D face recognition, facial expression, deformable model, expression transfer, nonrigid.

## 1 Introduction

Automatic face recognition is a challenging task that has gained a lot of attention during the last decade [1], [2]. 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 [3]. Since the 2D projection (image or appearance) of a 3D human face is sensitive to the above changes, utilizing 3D facial information appears to be a promising avenue to improve the face recognition accuracy [4].

Face recognition based on range images has been investigated by a number of researchers [5], [6], [7], [8], [9], [10], but only a few of them have addressed the deformation (expression) issue, which is a major challenge in 3D face recognition [4], [11]. Chua et al. [12] extended the use of Point Signature to recognize frontal face scans with different expressions, which was treated as recognition of nonrigid 3D surfaces. A database of six subjects with four different expressions was used in the experiments. Chang et al. [13] presented a method to independently match multiple regions around the nose and integrate individual matching results to make the final matching decision. Their method was evaluated on a database of about 4,000 facial scans from 449 subjects. Bronstein et al. [14], [15] proposed an algorithm based on an isometric model of facial surfaces, in an attempt to derive an expression-invariant facial surface representation for 3D face recognition. However, they considered only frontal face scans, and the proposed model assumed that the mouth was closed in all facial expressions. Their experiments were conducted on a database containing 27 human subjects with eight expressions. Passalis et al. [16], [17] fitted an annotated face model to a given facial scan by using a finite-element method approximation after initial alignment, and they applied a wavelet analysis on a geometry image derived from the fitted model to extract features for matching. The FRGC v2.0 database [18] was used for evaluating the algorithms.

We address the problem of matching *multiview* 2.5D facial scans (range images) to 3D neutral face models (or 2.5D facial scans) in the presence of expression variations. A 3D deformation modeling scheme is proposed to handle the expression variations. To account for the large intrasubject difference in 3D facial shape caused by expression changes, we propose to explicitly model the 3D deformation. Deformable models have been demonstrated to be promising in handling facial variations [19], [20]. Gross et al. [21] showed that user-specific deformable models are more robust than the generic deformable model (across subjects). However, to build a user-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 (see Fig. 1).

- *X. Lu is with Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540. E-mail: Lvxiaogu@ieee.org.*
- *A.K. Jain is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824. E-mail: jain@cse.msu.edu.*

*Manuscript received 13 Aug. 2006; revised 16 Jan. 2007; accepted 21 Aug. 2007; published online 11 Sept. 2007.*
*Recommended for acceptance by P.P. Jonathon.*
*For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0601-0806.*
*Digital Object Identifier no. 10.1109/TPAMI.2007.70784.*

Fig. 1. Deformation variations for one subject with the same type of expression.

We collect data for learning 3D facial deformations from only a small group of subjects, called the control group. Each subject in the control group provides a scan with neutral expression and several scans with nonneutral expressions. The deformations (between neutral scan and nonneutral scans) learned 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. For each subject in the gallery, deformable models are built based on the deformed templates.

Our deformation transfer and synthesis falls under the performance-driven framework [22], [23], [24], [25]. Unlike previous methods designed for realistic animation, we simplify the deformation transfer problem that is suitable for 3D matching. In order to learn deformation from the control group, we need a set of fiducial landmarks. Besides the fiducial facial landmarks such as eye and mouth corners, landmarks in the facial area with little texture, for example, cheeks, are extracted in order to model the 3D surface movement due to expression changes. We have designed a hierarchical geodesic-based resampling scheme constrained by fiducial landmarks to derive a new landmark-based surface representation for establishing correspondence across expressions and subjects. Thin-plate spline (TPS) is used for transferring the landmark-based deformation. The deformation transfer is achieved by minimizing a global bending energy function [26] while preserving the facial topology. Currently, these landmarks that are used for deformation modeling in the training stage are extracted in a semiautomatic fashion (see Section 2 for details). In the testing stage, the matching system is fully automatic, where the landmarks are not needed, and the three facial anchor points needed for alignment are automatically extracted [27], [11].

During matching, the user-specific deformable model is fitted to a test scan by solving an optimization problem to yield a matching distance. To handle the head pose changes, the rotation and translation parameters are integrated into the cost function for fitting. The model and pose parameters are solved using an alternating optimization scheme.

## 2 Deformation Modeling

The proposed scheme of deformation modeling for 3D face matching is presented in Fig. 2.

### 2.1 Hierarchical Facial Surface Sampling

Human faces share a common geometric topology, which can be represented by the craniofacial (fiducial) landmarks defined in anthropometry [28]. To model the expressions across the population, we use a fiducial set of nine landmarks (that is, two inner eye corners, two outside eye corners, two mouth corners, the nasion, nose tip, and subnasal) as constraints and the first layer in the hierarchical scheme (see Fig. 3a). For those facial regions that have little texture but are important for expression modeling such as the cheeks, we extract landmarks by sampling the facial surface hierarchically based on geodesics, which have been demonstrated to be insensitive across facial expressions [14]. The second layer

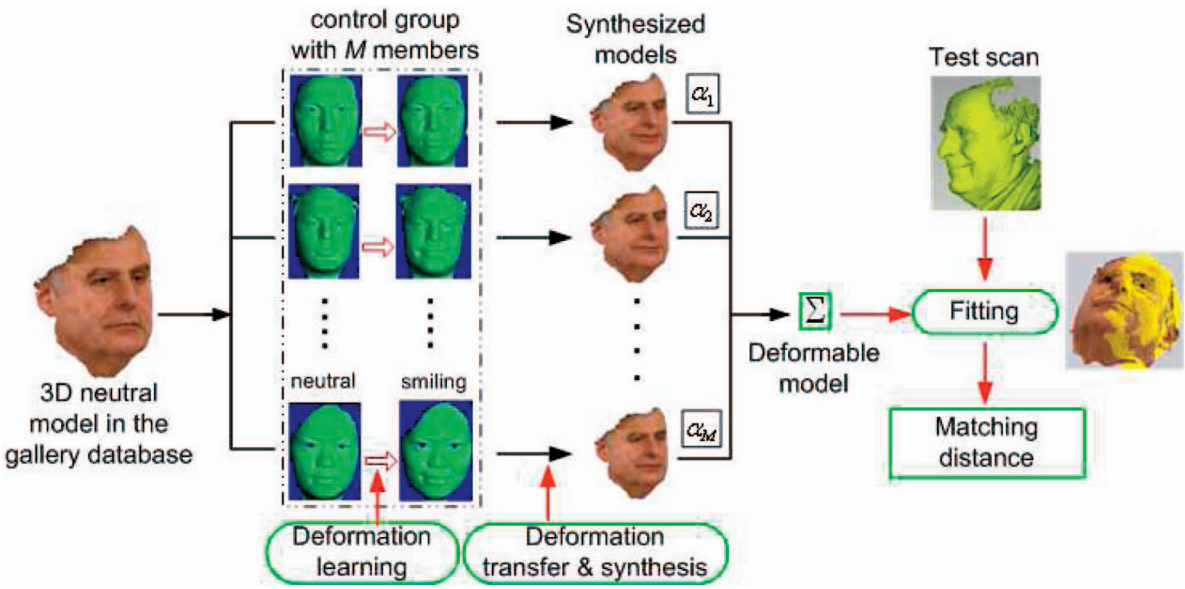


Fig. 2. Deformation modeling for 3D face matching. To match a 2.5D test scan to a 3D neutral face model in the gallery database, the deformation learned from the control group is transferred to the 3D neutral model. Each subject in the control group provides its own deformation transform. The 3D models with the corresponding deformation are synthesized. The $M$ synthesized models are combined to construct a user-specific deformable model, which is fitted to the given test scan.

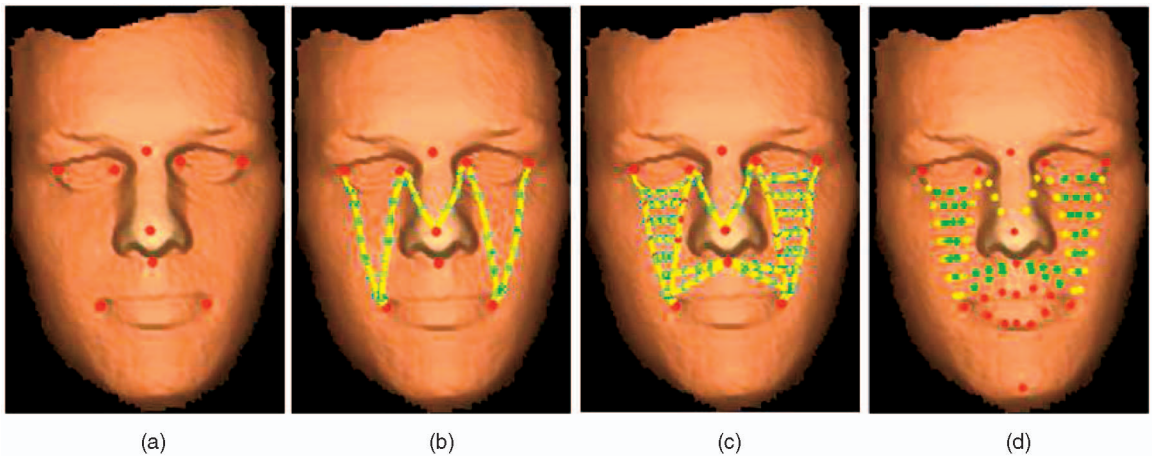


Fig. 3. Hierarchical surface sampling. (a) First layer (fiducial set). (b) Second layer. (c) Third layer. (d) Final landmark set.

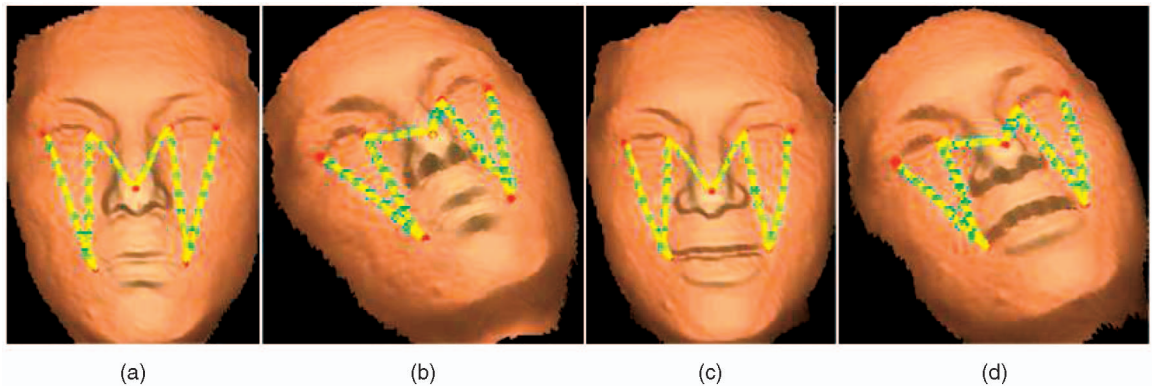


Fig. 4. Geodesic paths (yellow) across different expressions. (a) and (b) A neutral scan shown in two different views. (c) and (d) A scan of happy expression from the same subject in the same two views.

of landmarks is established based on the landmarks in the first layer. The geodesic distance and the corresponding path between two fiducial landmarks (for example, from one eye corner to one mouth corner) on the facial surface are computed based on the fast marching algorithm [29]. The derived paths encode the facial surface movement of different expressions, as shown in Fig. 4. We divide each path into $L$ segments with equal geodesic length. These points are then used as the newly extracted landmarks. Fig. 3b gives an example.

The third layer of landmarks is constructed based on the extracted landmarks obtained in the second layer by computing the geodesic paths between landmarks in the second layer and sampling the paths with equal geodesic length (see Fig. 3c). Our experiments show that three layers provide a sufficient number of landmarks for expression modeling.

The resulting landmark set includes fiducial landmarks (9 points), first-layer landmarks (34 points), second-layer landmarks (40 points), along with the chin point (1 point), and mouth contour (10 points). The chin point and the mouth contour are currently manually identified: they are not involved in the geodesic-based sampling scheme but are important for expression modeling. In total, there are 94 landmarks, as shown in Fig. 3d. To learn the 3D surface deformation, the correspondences between the landmarks need to be established [23], [24].

### 2.2 Deformation Transfer and Synthesis

The deformation is learned from a control group of $M$ subjects, who provide both neutral and nonneutral expression scans. 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. Therefore, we propose the following procedures to transfer the learned deformation to a 3D neutral model in the gallery for synthesis (see Fig. 5 for the notations and illustrations):

1. Register the nonneutral scan with the neutral scan to estimate the displacement vector of landmarks due to the expression change.
2. Establish a mapping $\phi$ from the landmark set $LS_{ne}$ of the neutral scan to that $LM_{ne}$ of the 3D neutral model.
3. Use the mapping $\phi$ to transfer the landmarks $LS_{sm}$ in the nonneutral scan to the 3D neutral model as $LS'_{sm}$.
4. Establish a mapping $\psi$ from the landmarks $LM_{ne}$ of the 3D neutral model to $LS'_{sm}$.
5. Apply $\psi$ 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.

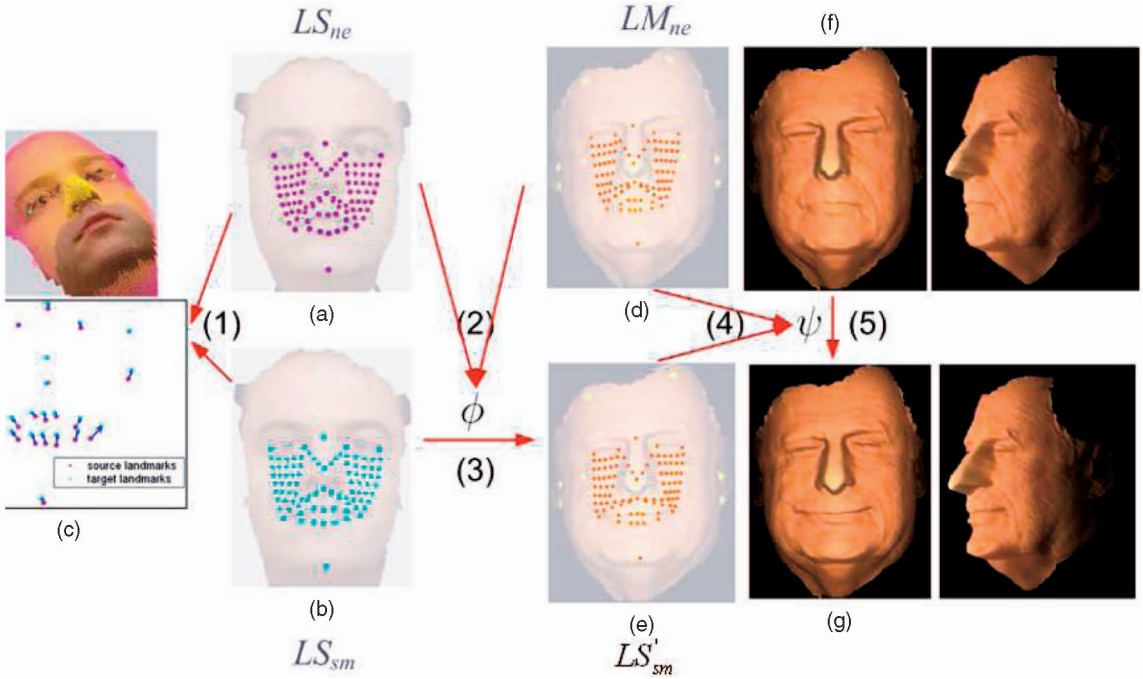


Fig. 5. Deformation transfer and synthesis. (a) Landmark set $LS_{ne}$ of the neutral scan in the control group. (b) Landmark set $LS_{sm}$ of the scan with nonneutral expression in the control group. (c) Rigid alignment between (a) and (b) by using the nose region that is invariant to expression changes, and the deformation field of the landmarks from (a) to (b) after the rigid alignment. (d) Landmark set $LM_{ne}$ of the (f) 3D neutral model in the gallery. (e) Landmark set $LS'_{sm}$ after deformation transfer. (g) Three-dimensional nonneutral model after applying deformation transfer and synthesis on (f). (h) and (i) Profile views of the model in (f) and (g), respectively.

#### 2.2.1 Thine-Plate Spine

TPS [26], [30] represents a natural parametric generalization from rigid to mild nonrigid deformations and is used for estimating the deformation $F$ between two sets of points. Given a pair of point patterns with known correspondences (landmarks) on two surfaces $U = (u_1, u_2, \cdots, u_m)^T$ and $V = (v_1, v_2, \cdots, v_m)^T$, where $U \subset g_0$, and $V \subset g_1$, we need to establish correspondences between other surface points. $u_k$ and $v_k$ denote the $(x, y, z)$-coordinates of the $k$th corresponding pair, $m$ is the total number of corresponding points, and $g_0$ and $g_1$ are two surfaces. A warping function $F$ that warps point set $U$ to $V$ subject to perfect alignment is given by the following 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 [26], [30]. In our application, the sets $U$ and $V$ correspond to 94 landmarks on a neutral scan and a nonneutral scan or a 3D neutral model, respectively.

#### 2.2.2 Deformation Transfer

The deformation transfer problem is defined as follows: Given a pair of source surfaces represented by meshes (in the control group), $S$, $S'$, and a target mesh $T$ (in the gallery) 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 the extracted landmarks. Fig. 5a shows the landmark set on the pair of face scans in the control group. The same set of landmarks is extracted on the 3D neutral model for deformation transfer (see Fig. 5d).

In order to separate nonrigid facial expressions from rigid head motion, a rigid transformation (translation and rotation) is applied to align the neutral scan and the nonneutral scan in the control group based on those landmarks that are insensitive to expression changes such as eye corners and nose tip. This normalizes the facial (geometry) position (see Fig. 5c). After the rigid alignment of neutral and nonneutral scans, the estimated displacement vectors need to be transferred to the 3D neutral model in the gallery. We establish a TPS mapping from the landmark set of the neutral scan in the control group to that in the 3D neutral model in the gallery. Since the TPS mapping contains the affine component and the distortion component, both the scale and the orientation of the motion vectors are also adjusted. The landmarks for the nonneutral scans are mapped onto the corresponding positions in the coordinate system of the 3D neutral model by applying the estimated TPS mapping.

#### 2.2.3 Deformation Synthesis

Deformation transfer establishes the new positions of the landmarks in the 3D neutral model. A TPS mapping is computed from the landmarks in the 3D neutral model to their deformed positions. The resulting mapping is used for interpolating the positions of surface points in between the

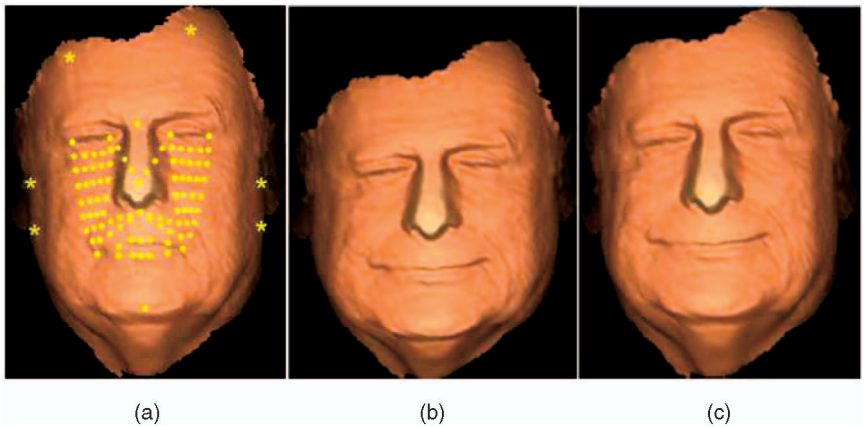


Fig. 6. Deformation synthesis. (a) Three-dimensional neutral model with landmarks. The dots are the landmarks in correspondence to those in the control group (see Fig. 5a). The star points are used for boundary constraints. (b) Synthesis without fixed-point boundary constraint, resulting in distortions. (c) Synthesis with fixed-point boundary constraints.

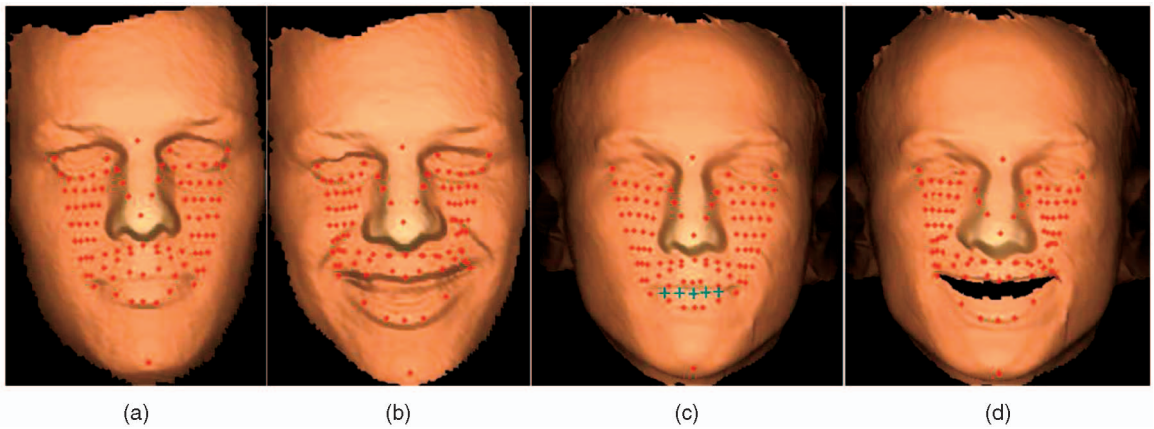


Fig. 7. Expression transfer and synthesis with the mouth open. (a) Landmark set for the neutral scan in the control group. (b) Landmark set for the scan with nonneutral expression in the control group. (c) Landmark set for a 3D neutral model in the gallery. Points marked as "+" are included to partition the mouth so that the upper and lower lips can move independently. (d) Three-dimensional nonneutral model with a synthesized expression transferred from the pair (a, b) to (c).

landmarks. For the vertices in between the convex hull spanned by the landmarks, the interpolation can be done by TPS mapping. However, for those vertices that lie outside this convex hull, an extrapolation has to be performed, leading to distortions, as shown in Fig. 6b. Therefore, we add a few additional landmarks (shown as "*" in Fig. 6a), which specify the boundary constraints. These landmarks are mapped to themselves. By computing the TPS mapping based on this augmented landmark set (dots plus stars in Fig. 6a), the interpolation can generate a better synthesis result, as shown in Fig. 6c.

#### 2.2.4 Synthesizing Open Mouth

A number of facial expressions involve an open mouth, but the templates (3D model or 2.5D scan) with a neutral expression usually do not contain any data inside the mouth. In order to model the open mouth according to expression changes, we add five landmarks to partition the mouth (labeled as "+" in Fig. 7) so that the upper and lower lips can move independently.

All the landmarks used for deformation modeling are summarized in Table 1. To ensure high modeling accuracy, the fiducial landmarks needed at the model construction stage are currently manually labeled.

### 2.3 Deformable Model Construction

Although a change in facial expression is a continuous process, a synthesized template (nonneutral model) captures only a specific instance of the expression. Further, since each single synthesized nonneutral model is obtained by transferring the deformation from one member in the control group to the neutral gallery model, it is not likely to be the true expression of the gallery model. Therefore, we learn the expression deformation from all the $M$ members in the control group. This leads to a user-specific deformable model that is a linear combination of nonneutral models, each obtained as a result of deformation transfer from one member of the control group to the neutral model.

Let $S$ represent a face surface model

$$S = (x_1, y_1, z_1, \cdots, x_n, y_n, z_n)^T,$$

where $(x_k, y_k, z_k)$ is the location of the $k$th surface vertex, $k = 1, 2, \cdots, n$, and $n$ is the total number of vertices. For each subject, let $S_{ne}$ denote the neutral model and let $S_i$ $(i = 1, 2, \cdots, M)$ denote the deformed model generated by the $i$th member in the control group. We assume that all $S_i$'s correspond to the same type of expression synthesized from $S_{ne}$. Notice that since all $S_i$s are synthesized from $S_{ne}$, the

TABLE 1
Landmarks for Deformation Modeling (in the Training Stage)

| Landmarks | Total number | Extraction | Precision requirement for modeling | Examples |
|---|---|---|---|---|
| Fiducial set and mouth contour | 20 | Manual | High | Red dots in Fig. 3d |
| Surface resampling | 74 | Automatic | High | Yellow & green dots in Fig. 3d |
| Boundary constraint | 6 | Manual[1] | Low[3] | '*' in Fig. 6a |
| Mouth splitting | 5 | Manual[2] | Low[3] | '+' in Fig. 7c |

[1] *The boundary constraints can be automatically estimated based on the fiducial points and facial geometry.* [2] *The mouth splitting points can be automatically computed using the mouth contour.* [3] *The boundary constraints and mouth splitting points are not used as correspondences between the control group and the 3D model.*

correspondence between them is automatically established. By combining all the $M$ synthesized models, we construct the deformable model for this subject 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 $\alpha_i$s are the mixing weights. The deformable model consists of two components: the first component is the subject's neutral model $S_{ne}$, and the second is the variation component representing the change in facial surface due to expression. In other words, although $S_{ne}$ controls the subject's identity, the variation component does deformation adaptation by adjusting the weights $\alpha_i$. As the number of subjects $M$ in the control group increases, the number of weights $\alpha_i$ also increases, leading to a more complex fitting problem in a high-dimensional parameter space. The principal component analysis can be applied to reformulate the deformable model and reduce the complexity by keeping only the principal modes [20].

### 2.4 Expression-Specific versus Expression-Generic Models

For each subject, we construct one deformable model for each type of expression of interest. Thus, if the control group contains $P$ different nonneutral expressions, we learn $P$ expression-specific deformable models. These expression-specific models can also be integrated into a single expression-generic deformable model by adding new linear variation components in (3). Unlike expression-specific models, the expression-generic model is a single model for multiple expressions. The integrated expression-generic model has a larger parameter space as the number of expression types increases, leading to higher model complexity. Experimental results show that the expression-generic deformable-model-based scheme gives lower matching accuracy than the expression-specific-model-based scheme, but with lower computational cost when handling multiple expressions (see Section 3 for details).

### 2.5 Deformable Model Fitting

Two types of transformations are applied to a 3D deformable model when it is matched to a given test scan with a claimed identity. The first one is the rigid transformation due to the head pose changes, which can be represented by a rotation matrix and a translation vector. The second one is the nonrigid deformation, which can be modeled by the weights $\alpha_i$ in (3). Fitting the deformable model to a given test scan is formulated as an optimization problem to minimize the cost function:

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

where $R$ and $T$ are the rotation matrix and translation vector, respectively, $S$ is the 3D deformable model, $S_t$ denotes the test scan, and $\xi(S_t|R,T)$ represents applying the rotation and translation transformations $(R,T)$ to $S_t$. To reduce the computation cost in the optimization process, we subsample the test scan surface into a number of control points (around 96 control points are sampled in our experiments) that are used for the alignment and cost function evaluation [11], as explained in the following.

We factorize the rigid and nonrigid components and use an alternating optimization scheme to solve for the unknown parameters:

1. Initialize the deformable model parameters to generate a 3D model. Estimate a coarse alignment between the model and the test scan by using three anchor points. See [27] for an automatic anchor point extraction algorithm.
2. The iterative closest point (ICP) algorithm is utilized to solve for the rotation and translation parameters $(R,T)$ [31] to achieve pose normalization while fixing $\alpha_i$s.
3. Given $R$ and $T$ obtained in step 2, minimize the cost function $E$ by solving for $\alpha_i$s.
4. Use the $\alpha_i$s computed in step 3 to generate a new instance of the 3D model. Repeat steps 2 to 4 until the convergence is reached.

In step 3, the optimization can be achieved by a gradient-based iterative approach such as the BFGS quasi-Newton method [32]. Because the cost function is evaluated based on the control points in the test scan and their closest counterparts in the deformable model, and the closest counterparts may change due to the adjustment of $\alpha_i$s, the optimization problem is highly nonlinear. Multiple iterations of cost

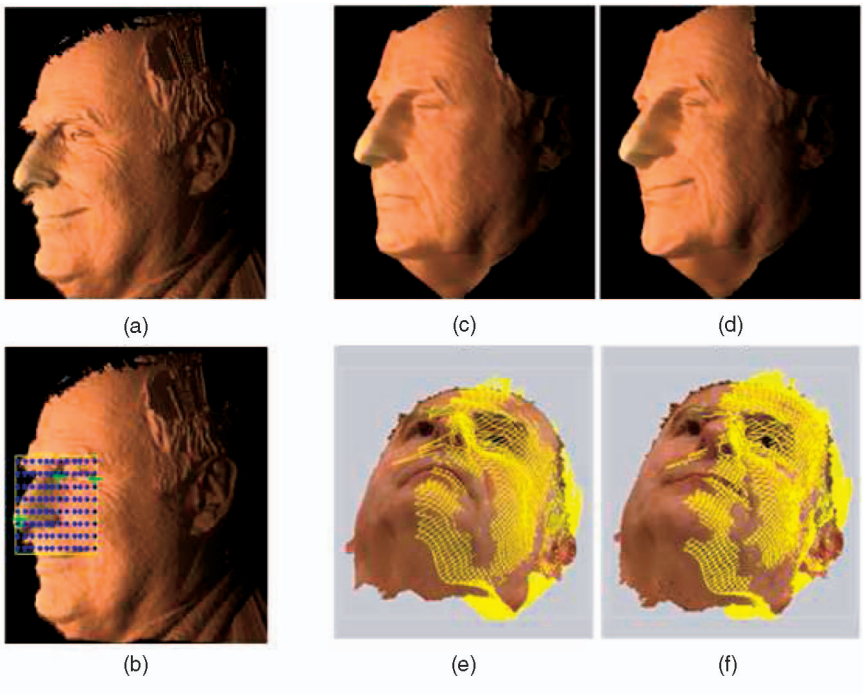


Fig. 8. Deformable model fitting. (a) Test scan. (b) Three anchor points (two eye corners and the nose tip) are automatically extracted [27] from (a), based on which a set of control points (blue dots) are sampled [11] for deformable model fitting and matching distance calculation. (c) A 3D neutral model. (d) Deformed model after fitting to (a). Registration results of (a) to models (c) and (d) are given in (e) and (f), respectively (the test scan (yellow wireframe) is overlaid on the 3D model). The matching distances are 2.7 and 1.3 mm, respectively.

function evaluation are computationally expensive due to the large number of closest point searches. However, as an approximation, by fixing the correspondence, the $\alpha_i$s can be obtained in a *noniterative* way by solving a linear least square problem as

$$\alpha_{opt} = (\tilde{S}^T \tilde{S})^{-1} (\tilde{S}^T (S_t - S_{ne})), \quad (5)$$

where $\tilde{S}$ is the matrix $[(S_1 - S_{ne}), (S_2 - S_{ne}), \ldots, (S_M - S_{ne})]$ (here, $S_i$s and $S_{ne}$ have already been transformed by $(R, T)$ using $\xi$). Experimental results show that this simplification significantly reduces the computational cost while providing competitive accuracy compared to the iterative BFGS optimization algorithm. Moreover, this linear noniterative optimization is much more efficient than iterative gradient-based algorithms as the number of parameters ($\alpha_i$s) increases. After the fitting process, the root-mean-square distance calculated by the ICP algorithm for the set of control points is used as the matching distance [11]. A model fitting example is provided in Fig. 8. In the expression-specific model-based scheme, for each subject, we match all its deformable models, one per expression, to a given test scan. The minimum of all the obtained matching distances is used as the final matching distance.

## 3 Experiments and Discussion

We collected a control group consisting of 10 subjects in our laboratory (MSU). All the range images (downsampled to $320 \times 240$ with a depth resolution of $\sim 0.1\,mm$) were collected using a Minolta Vivid 910 scanner [33]. Each 2.5D facial scan has $\sim 18,000$ effective points (excluding the background). Each subject in the control group provided seven expressions, namely, neutral, happy, angry, smile, surprise, deflated, and inflated [15].

We evaluate the proposed scheme in the identification mode, that is, by matching a test scan to all the gallery models. Both expression-specific and expression-generic deformable-model-based schemes are evaluated. The expression-specific deformation is learned by utilizing the neutral scan and a specific expression scan. The expression-generic deformable model is constructed by including all the seven expressions collected in the control group. To initialize a coarse alignment between a test scan and a gallery template (see step 1 in Section 2.5), three anchor points (two eye corners and the nose tip) are automatically extracted from a test scan [27]. The matching process is fully automatic.

### 3.1 Frontal View and Nonneutral Expression

FRGC v2.0 [18] is a large public domain face database, which contains (near) frontal 2.5D facial scans. Although no 3D models are available for subjects in this database, the proposed deformation modeling and matching scheme is still applicable by replacing a 3D full-view model in the gallery with a 2.5D frontal neutral scan. There are 4,007 2.5D face scans from 465 subjects, captured during Fall 2003 and Spring 2004 by a Minolta Vivid 900/910 series scanner. In addition to the neutral expression, subjects provided scans with several nonneutral expressions such as smiling (happiness), frowning, astonishing (surprise), and puffy cheeks. In our experiments, all the scans are downsampled to $320 \times 240$. Due to the computational cost of model fitting, only the first 100 subjects are selected from the FRGC v2.0 database. For each subject, the scan with a neutral expression and the earliest time stamp is used as the

Fig. 9. Example scans from one subject in the FRGC v2.0 database.

template to construct the gallery. The remaining scans with various expressions, including happiness, frowning, surprise, disgust, sadness, and puffy cheeks, are chosen as test scans.[1] In total, there are 100 2.5D gallery templates and 877 independent 2.5D scans for testing. Representative scans are provided in Fig. 9. The expression deformations are learned and transferred from the control group to construct a deformable model (a 2.5D deformable frontal template) for each subject in the gallery.

Two baseline algorithms used in FRGC [18], that is, ICP and PCA-based approaches, are applied. The PCA baseline utilizes both range and intensity maps to compute a combined score for matching [18]. The scheme of using only the nose region is also evaluated, where the nose region I used in [13] is applied.

The Cumulative Match Characteristic (CMC) curves from our matching algorithm are provided in Fig. 10. Based on all the computed matching distances, the Receiver Operating Characteristic (ROC) curves are generated, which are given in Fig. 11. Fig. 12 shows some of the test scans that are incorrectly matched using a rigid transformation (ICP) but are correctly matched by using the proposed deformation modeling scheme.

### 3.2 Multiview and Nonneutral Expression

To our knowledge, there is no publicly available 3D facial scan database containing a simultaneous expression and (large) pose variations. Therefore, we collected a database of 90 subjects at MSU. Five scans with a neutral expression for each subject were captured to construct the full-view 3D model [11]. In addition, for each subject, six independent scans were captured for testing, including three scans with neutral expression and three with smiling expression. For each expression, the three scans were captured at frontal, left profile ($\sim 45$ degrees left from frontal), and right profile ($\sim 45$ degrees right from frontal) positions, respectively. For a small number of subjects in the database, we had fewer than six test scans. In total, the test database consists of 533 independent scans (different from training scans used for constructing the full-view 3D model) of the same 90 subjects. The 10 subjects in the control group are different from the 90 subjects in this database. The proposed deformable model scheme is compared with rigid-transformation (ICP [31])-based matching scheme. Representative test scans are shown in Fig. 13. The CMC curves for the MSU database are provided in Fig. 14. Based on all the computed matching distances, the ROC curves are generated, which are given in Fig. 15. Two subscenarios are further evaluated: 1) the test scans are nonfrontal with a neutral expression and 2) the test scans are nonfrontal with expression variations. Corresponding CMC curves are presented in Fig. 16.

1. See supplemental materials for the complete filename lists used in the experiments.

### 3.3 Discussion

These experimental results (CMC and ROC curves) demonstrate that the proposed deformation modeling scheme improves the matching accuracy in the presence of expression variations along with large head pose changes. Fig. 17 shows examples where the proposed scheme fails to find the correct matches for FRGC experiment. One of the reasons for the matching errors is that the current fitting (optimization) process is prone to the local minimum. In addition, since our control group contains only 10 subjects,

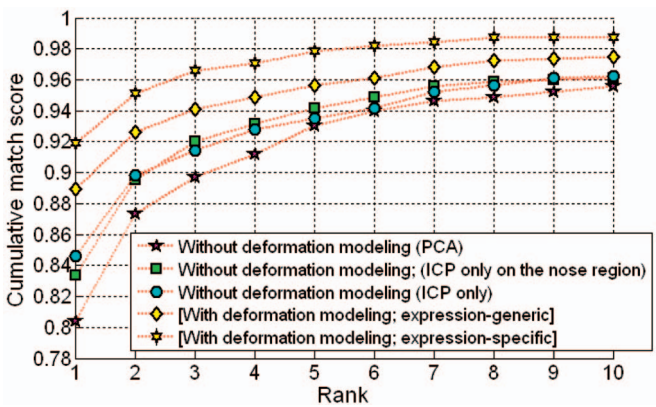


Fig. 10. CMC curves for the subset of the FRGC v2.0 database.

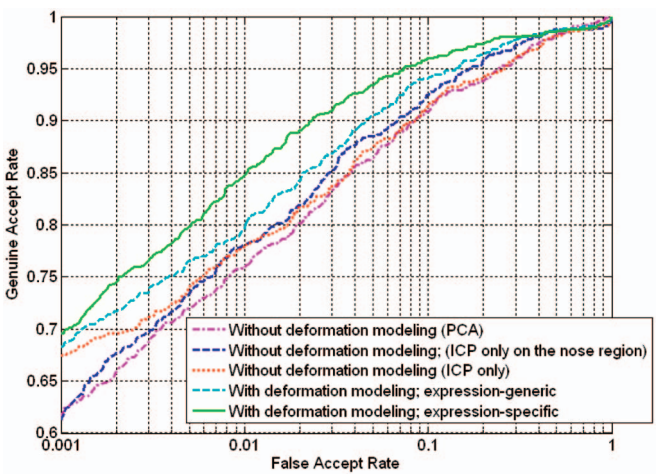


Fig. 11. ROC curves for the subset of the FRGC v2.0 database.

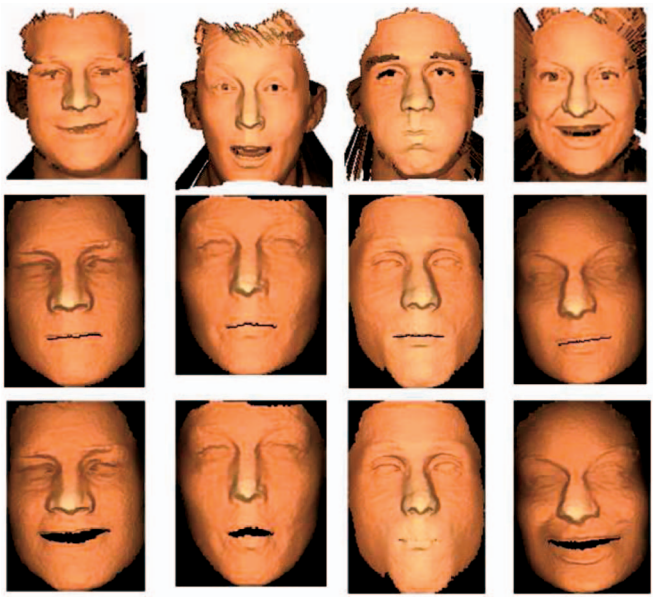

Fig. 12. (a) Examples of test scans in the FRGC database that are incorrectly identified with a rigid transformation (ICP) but are correctly identified with deformation modeling. (b) Corresponding genuine 2.5D neutral templates. (c) Corresponding genuine deformed templates after model fitting.

Fig. 13. Representative 2.5D test scans in the MSU database.

we are not able to fully learn the deformation that is generalizable across a large population. Although the expression-specific scheme has higher computational cost than the expression-generic one, it has better performance. An expression classification module could be added on top of the expression-specific scheme as a filter. The experiments on the MSU data set show higher accuracy improvement by integrating the proposed deformation modeling scheme than on the FRGC v2.0 data. A possible explanation is that there are only two expressions in the MSU data set,

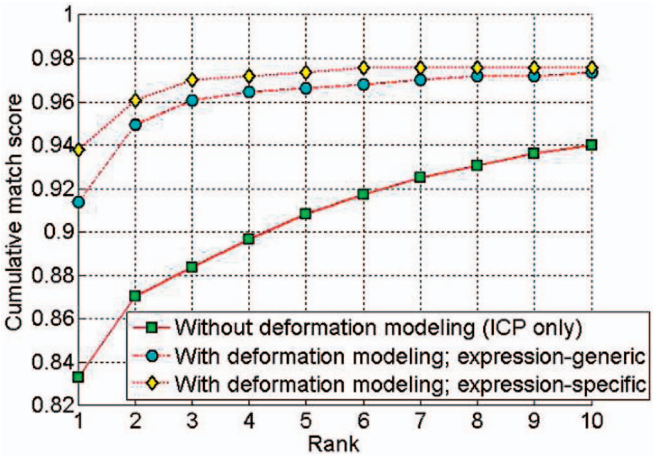


Fig. 14. CMC curves of the experiments on the MSU database.

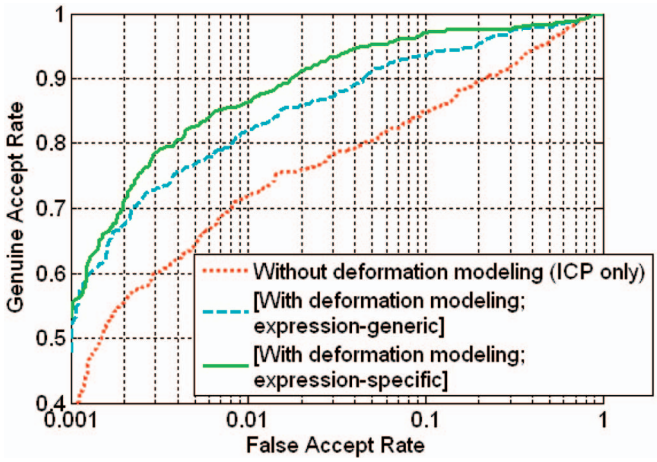


Fig. 15. ROC curves for the MSU database.

that is, neutral and smile, whereas there are seven expressions in the FRGC v2.0 data set.

The average CPU time (Pentium 4, 2.8 GHz) for matching a test scan to a model is 5 s (per expression) implemented in Matlab. The bottleneck is the time-consuming ICP process that has to be repeated, since the deformable model (template) changes as the model parameters (that is, $\alpha_i$s in (4)) are updated. It takes about 2 s to automatically extract three anchor points from a test scan [27].

## 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 pose changes in the test scans. A hierarchical surface sampling scheme is used for augmenting fiducial landmarks for analyzing 3D facial surfaces across expressions. The fiducial landmarks needed during expression learning are manually extracted. Additional landmarks (74 points) in facial surface regions with little texture are automatically extracted using the geodesic-based approach. The 3D deformation learned from a small number of subjects (control group) is transferred to the 3D neutral models in the gallery. The corresponding

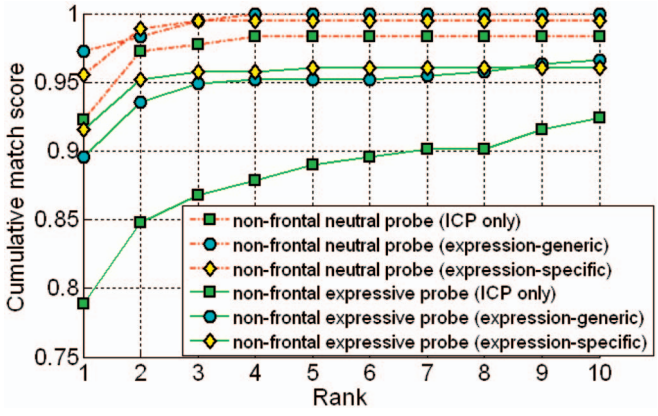


Fig. 16. CMC curves of the experiments on the MSU database in two categories: 1) nonfrontal test scans with neutral expression and 2) nonfrontal test scans with expression variations.

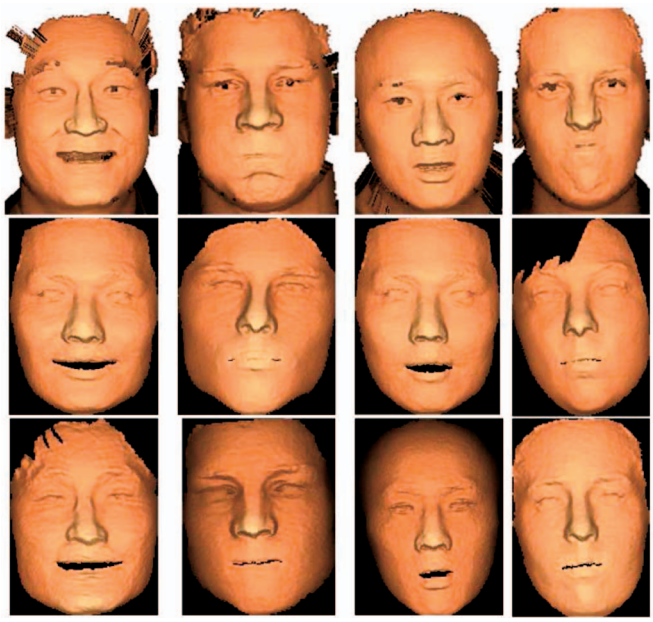

Fig. 17. Examples of incorrect matches in the FRGC database. (a) Test scans. (b) Corresponding best matched templates after model fitting. (c) Corresponding genuine templates after modeling fitting.

deformation is synthesized in the 3D neutral model to generate user-specific 3D nonneutral models. Two types of deformable models have been built: expression-specific and expression-generic. 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 gallery face models. The resulting 3D face surface matching system is more robust across expressions. As expected, the expression-specific deformation modeling gives better matching results but requires more computation, since the test scan needs to be matched to multiple deformable models.

We are currently exploring following issues:

1. fully automatic extraction of landmarks for deformation modeling (in the training stage),
2. reducing the computational cost of matching,
3. deformation learning using a larger control group, and
4. evaluating the performance on the full FRGC v2.0 database.

## Acknowledgments

An earlier version of this paper appeared in the *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2006) [34]. The authors wish to thank the anonymous reviewers for useful suggestions and all the volunteers who provided the face data.

## References

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, 2003.

[2] *Handbook of Face Recognition,* S. Li and A. Jain, eds. Springer, 2005.

[3] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report," http://www.frvt.org/FRVT2002/documents.htm, Mar. 2003

[4] K.W. Bowyer, K. Chang, and P.J. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-Modal 3D+2D Face Recognition," *Computer Vision and Image Understanding,* vol. 101, no. 1, pp. 1-15, 2006.

[5] J. Lee and E. Milios, "Matching Range Images of Human Faces," *Proc. Third Int'l Conf. Computer Vision (ICCV '90),* pp. 722-726, 1990.

[6] G. Gordon, "Face Recognition Based on Depth and Curvature Features," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '92),* pp. 108-110, 1992.

[7] H. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-Based Face Surface Recognition Using Spherical Correlation," *Proc. Third IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR '98),* pp. 372-377, 1998.

[8] C. Beumier and M. Acheroy, "Automatic 3D Face Authentication," *Image and Vision Computing,* vol. 18, no. 4, pp. 315-321, 2000.

[9] G. Pan, Z. Wu, and Y. Pan, "Automatic 3D Face Verification from Range Data," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '03),* vol. 3, pp. 193-196, 2003.

[10] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "Multi-Modal 2D and 3D Biometrics for Face Recognition," *Proc. IEEE Workshop Analysis and Modeling of Faces and Gestures (AMFG '03),* pp. 187-194, Oct. 2003.

[11] X. Lu, A.K. Jain, and D. Colbry, "Matching 2.5D Face Scans to 3D Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 1, pp. 31-43, Jan. 2006.

[12] C. Chua, F. Han, and Y. Ho, "3D Human Face Recognition Using Point Signature," *Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR '00),* pp. 233-238, Mar. 2000.

[13] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1695-1700, Oct. 2006.

[14] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Expression-Invariant 3D Face Recognition," *Proc. Fourth Int'l Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA '03),* pp. 62-70, 2003.

[15] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Three-Dimensional Face Recognition," *Int'l J. Computer Vision,* vol. 64, no. 1, pp. 5-30, 2005.

[16] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of the UR3D Algorithm Using the FRGC v2 Data Set," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments (FRGC '05),* June 2005.

[17] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 640-649, Apr. 2007.

[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," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '05),* pp. 947-954, 2005.

[19] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001.

[20] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 9, pp. 1063-1074, Sept. 2003.

[21] R. Gross, I. Matthews, and S. Baker, "Generic vs. Person Specific Active Appearance Models," *Image and Vision Computing,* vol. 23, no. 11, pp. 1080-1093, 2005.

[22] L. Williams, "Performance-Driven Facial Animation," *Proc. ACM SIGGRAPH '90,* pp. 235-242, 1990.

[23] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D.H. Salesin, "Synthesizing Realistic Facial Expression from Photographs," *Proc. ACM SIGGRAPH '98,* pp. 75-84, 1998.

[24] J. Noh and U. Neumann, "Expression Cloning," *Proc. ACM SIGGRAPH '01,* pp. 277-288, 2001.

[25] R. Sumner and J. Popovic, "Deformation Transfer for Triangle Meshes," *Proc. ACM SIGGRAPH '04,* pp. 399-405, Aug. 2004.

[26] F.L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, pp. 567-585, 1989.

[27] X. Lu and A.K. Jain, "Automatic Feature Extraction for Multiview 3D Face Recognition," *Proc. Seventh IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR '06)*, pp. 585-590, 2006.
[28] L.G. Farkas, *Anthropometry of the Head and Face*, second ed. Raven Press, 1994.
[29] R. Kimmel and J.A. Sethian, "Computing Geodesic Paths on Manifolds," *Proc. Nat'l Academy of Sciences*, vol. 95, pp. 8431-8435, 1998.
[30] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*. John Wiley & Sons, 1998.
[31] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992.
[32] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*. Academic Press, 1981.
[33] *Minolta Vivid 910 Non-Contact 3D Laser Scanner*, http://www.minoltausa.com/vivid/, 2007.
[34] X. Lu and A.K. Jain, "Deformation Modeling for Robust 3D Face Matching," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1377-1383, 2006.

**Xiaoguang Lu** received the BS and MS degrees from Tsinghua University, China, in 1997 and 2000, respectively, and the PhD degree from Michigan State University in 2006. From 2000 to 2001, he was a visiting student at Microsoft Research Asia. He was a summer research intern in Siemens Corporate Research and Mitsubishi Electric Research Laboratories in 2002 and 2003, respectively. He is currently with Siemens Corporate Research. His research interests include pattern recognition, computer vision, and machine learning, with applications to biometrics and medical image analysis. He is a member of the IEEE.

**Anil K. Jain** is a university-distinguished professor in the Department of Computer Science and Engineering, the Department of Electrical and Computer Engineering, and the Department of Statistics and Probability, Michigan State University. He was the editor in chief of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He is currently an associate editor for the *IEEE Transactions on Information Forensics and Security* and the *ACM Transactions on Knowledge Discovery in Data*. He is a member of the Biometrics Defense Support Team and the National Academies Committee on Whither Biometrics and Committee on Improvised Explosive Devices. His research interests include statistical pattern recognition, data clustering, and biometric authentication. He is the holder of six patents in fingerprint matching. He is the author of a number of books: *Handbook of Biometrics* (2007), *Handbook of Multibiometrics* (2006), *Biometric Systems, Technology, Design and Performance Evaluation* (2005), *Handbook of Face Recognition* (2005), *Handbook of Fingerprint Recognition* (2003), *BIOMETRICS: Personal Identification in Networked Society* (1999), and *Algorithms for Clustering Data* (1988). He is a coauthor of *Algorithms for Clustering Data* (Prentice Hall, 1988), which, according to Citeseer, is ranked 93 in the Most Cited Articles in Computer Science (over all times). He is a fellow of the IEEE, the AAAS, the ACM, SPIE, and the International Association of Pattern Recognition (IAPR). He has received Fulbright, Guggenheim, Alexander von Humboldt, and IEEE Computer Society Technical Achievement Awards. He is the recipient of the 1996 IEEE Transactions on Neural Networks Outstanding Paper Award and the 1987 and 1991 Pattern Recognition Society Best Paper Awards. ISI has designated him as a highly cited researcher.