Matching 2.5D face scans to 3D models — Source link ⧉

Xiaoguang Lu, Anil K. Jain, Dirk Colbry

Institutions: Michigan State University

Topics: Matching (statistics), Facial recognition system, Iterative closest point, Face (geometry) and Pattern recognition (psychology)

Related papers:

- A method for registration of 3-D shapes

- Overview of the face recognition grand challenge

- A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition

- Face recognition: A literature survey

- An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition

Matching 2.5D Face Scans to 3D Models

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

Abstract—The performance of face recognition systems that use two-dimensional images depends on factors such as lighting and subject's pose. We are developing a face recognition system that utilizes three-dimensional shape information to make the system more robust to arbitrary pose and lighting. For each subject, a 3D face model is constructed by integrating several 2.5D face scans which are captured from different views. 2.5D is a simplified 3D (x, y, z) surface representation that contains at most one depth value (z direction) for every point in the (x, y) plane. Two different modalities provided by the facial scan, namely, shape and texture, are utilized and integrated for face matching. The recognition engine consists of two components, surface matching and appearance-based matching. The surface matching component is based on a modified Iterative Closest Point (ICP) algorithm. The candidate list from the gallery used for appearance matching is dynamically generated based on the output of the surface matching component, which reduces the complexity of the appearance-based matching stage. Three-dimensional models in the gallery are used to synthesize new appearance samples with pose and illumination variations and the synthesized face images are used in discriminant subspace analysis. The weighted sum rule is applied to combine the scores given by the two matching components. Experimental results are given for matching a database of 200 3D face models with 598 2.5D independent test scans acquired under different pose and some lighting and expression changes. These results show the feasibility of the proposed matching scheme.

Index Terms—Face recognition, 3D model, multimodal, surface matching, appearance-based.

✦

---

1 INTRODUCTION

AUTOMATIC human face recognition is a challenging task that has gained a lot of attention during the last decade [1], [2]. While most efforts have been devoted to face recognition from two-dimensional (2D) images [1], a few approaches have utilized depth information provided by 2.5D range images [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. 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 [14] (see Fig. 1). Because the human face is a three-dimensional (3D) object whose 2D projection (image or appearance) is sensitive to the above changes, utilizing 3D face information can improve the face recognition performance [15], [14]. Range images captured explicitly by a 3D sensor [16], [17] contain facial surface shape information. The 3D shape of a facial surface represents the face structure, which is related to the internal face anatomical structure instead of external appearance and environment. It is also more difficult to fake a 3D face compared to a 2D face image to circumvent the face recognition system.

In this research, 3D models are used to recognize 2.5D face scans, provided by commercial 3D sensors, such as Minolta Vivid series [17]. A 2.5D scan is a simplified 3D (x, y, z) surface representation that contains at most one depth value (z direction) for every point in the (x, y) plane, associated

with a registered texture image (see Fig. 2). Each scan can only provide a single viewpoint of the object, instead of the full 3D view. In real-world scenarios, similar to the current 2D camera capture systems, 3D sensors provide only partial views of the face. However, during the training (enrollment) stage, a 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. Currently, matching 2.5D scans to 3D models has a limited range of applications, such as high security access control, due to the cost of 3D data acquisition. But, with continued progress in 3D imaging technology [18], we expect that cost-efficient nonintrusive 3D data capture will become readily available. Three-dimensional facial structure reconstruction from images has received substantial attention [19], [20], [21], [15] not only to improve the visual quality, but also for improving the metrical accuracy [22].

Face recognition based on range images has been addressed in a number of different ways. Cartoux et al. [3] used facial profiles extracted from 3D range images for face recognition. Lee and Milios [4] segmented the range image to obtain the convex regions, based on the sign of the mean and Gaussian curvatures at each point. These convex regions correspond to distinct facial features. The Extended Gaussian Image (EGI) is used to represent each convex region. A similarity metric between two regions is defined to match the facial features of the two face images. Gordon [5] explored the face feature extraction for recognition based on depth and curvature features. Achermann et al. [6] applied the eigenface and Hidden Markov Model techniques to range image-based recognition. Tanaka et al. [7] considered the face recognition problem as a 3D shape recognition problem of rigid free-form surfaces. Their method is based on the surface curvature information. Chua et al. [8] extended the use of Point

---

● *The authors are with the Department of Computer Science and Engineering, Michigan State University, 3115 Engineering Building, East Lansing, MI 48824. E-mail: {Lvxiaogu, jain, colbrydi}@cse.msu.edu.*

Fig. 1. Face appearance variations.

Fig. 2. (a) A frontal 2.5D scan viewed from different viewpoints and (b) the full 3D model.

Signature to recognize frontal face scans with different expressions, which was treated as a 3D recognition problem of nonrigid surfaces. Beumier and Acheroy [9] extracted the profiles (curves) both from depth and gray-scale image for face verification. Hesher et al. [10] applied the PCA to the range image, and estimated probability models for the coefficients. Pan et al. [11] utilized the partial directed Hausdorff distance to align and match two range images for verification. Bronstein et al. [12] proposed an algorithm based on geometric invariants, in an attempt to deal with facial expression variations for face recognition. Work by Chang et al. [23] demonstrated that face recognition systems based on either two-dimensional texture information or 2.5D range information have similar performance characteristics. However, they went on to show that significant improvements can be made if a system uses a combination of texture and shape information. They applied PCA to both 2D and 3D face data.

Different methods have been used to address face recognition based on range images, but most of them have focused on only frontal view recognition. Further, most of these methods only use the surface shape information. But the appearance or texture component also plays an important role in the face recognition process, especially when the shapes of two faces in the gallery are similar. Facial appearance in 2D images is the projection of a 3D face object with lighting effects, containing the texture information of the face. Table 1 lists a number of factors which can change the facial information. Although the 3D shape will not change due to pose and lighting variations, it can still change due to expression and the aging factor. Using

3D shape information alone cannot fully handle all the variations which a face recognition system encounters.

We propose a combination scheme, which integrates surface (shape) matching and constrained appearance-based methods for multiview face matching (see Fig. 3) under some illumination changes and some expression variations. The surface matching utilizes the 3D shape information, while the appearance-based methods explore the texture clues. Integrating these two different modalities (shape and texture) may provide a more robust face recognition system to overcome the limitations encountered in the traditional 2D image-based face recognition system under pose and lighting changes. The appearance-based stage is constrained to a small candidate list generated by the surface matching stage, which reduces the classification complexity. In the conventional appearance-based algorithms, all the subjects in the training database are used for subspace analysis and construction. When the number of

TABLE 1
Relationship between Face Variation Factors and
Facial Properties (Shape and Appearance)

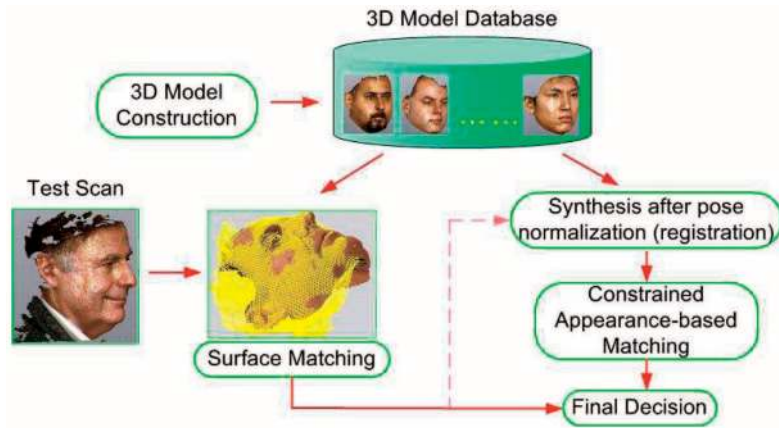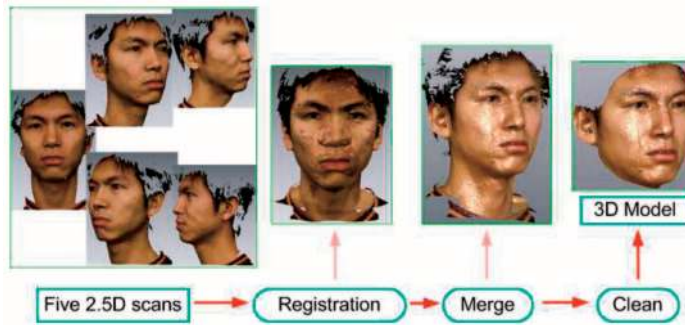| Factors	Shape (3D)	Appearance (2D)
Pose	No	Yes
Lighting	No	Yes
Expression	Yes	Yes
Aging	Yes	Yes
Makeup	No	Yes
Facial accessories	Yes	Yes

Fig. 3. Matching scheme.

Fig. 4. Three-dimensional face model construction.

subjects is large, this leads to a problem due to large interclass similarity. In our scheme, the 3D model is utilized to synthesize training samples with facial appearance variations, which are used for discriminant subspaces analysis. The matching scores obtained by the two matching components are combined to make the final decision.

In Section 2, we will present our 3D face model construction procedures. Section 3 describes the surface matching scheme. The constrained appearance-based matching component is proposed in Section 4. The integration scheme is presented in Section 5. Section 6 provides our experimental procedures for testing the system and results. Conclusions and our future work are provided in Section 7.

2 THREE-DIMENSIONAL MODEL CONSTRUCTION

Since each 2.5D scan obtained by a Minolta Vivid910 scanner can only cover a partial view of the full 3D face, the 3D face model for each subject is constructed by stitching several 2.5D scans obtained from different viewpoints, which cover the full facial area. In our current setup, five scans are used, i.e., frontal, left 30 degrees, left 60 degrees, right 30 degrees, and right 60 degrees. The 2.5 scans are first registered. Then, they are merged to resolve the ambiguity and create a surface model. Basic clean-up procedures are applied to fill holes, smooth the surface, and remove noisy points associated with hair and clothing. The end result is a smooth, full view, texture mapped mesh model of the face for each of our subjects. All the techniques used in the model construction procedures are well-studied in the computer graphics and vision research communities [24], [25], [26], [27]. For easy

manipulation, a commercial software, called Geomagic Studio [28] is currently used during our model construction process. Fig. 4 demonstrates the 3D face model construction procedure. The resulting model is highly dense, containing approximately 27,000 vertices and 50,000 polygons. It can be used to render new realistic facial appearance with pose and illumination variations.

3 SURFACE MATCHING

In order to match two facial surfaces (2.5D test scan and 3D model), we follow the coarse-to-fine strategy shown in Fig. 5.

3.1 Feature Point Extraction and Coarse Alignment

There are two objectives for extracting feature points from the 2.5D scans. First, three labeled feature points are used to calculate the rigid transformation that will align the 2.5D scan with the 3D model [29]. Second, specific feature points are needed to align a grid of control points, which are used in our fine alignment step. In the first stage, any three labeled feature points can be used as long as each point is matched with the same point on the 3D model. However, in the second stage, specific points are needed (inside corners of the eyes, outside corners of the eyes, and nose tip) to align a grid of control points to the face (see Fig. 9). These specific points can be calculated by back-projection using the rigid transformation discovered for the first objective.

Feature points are selected by determining the local shape index at each point within the 2.5D scan [30]. The shape index at point p is defined using the maximum (k_1)
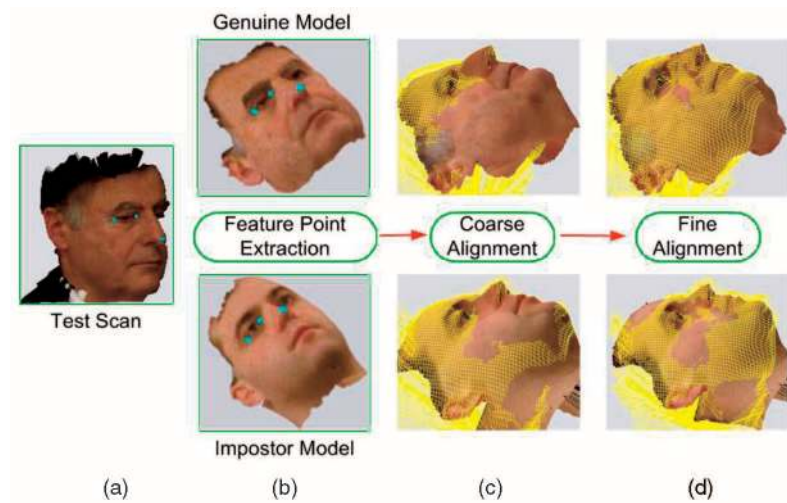
Fig. 5. Surface matching streamline. The alignment results are shown by the 3D model overlaid on the wire-frame of the 2.5D test scan.
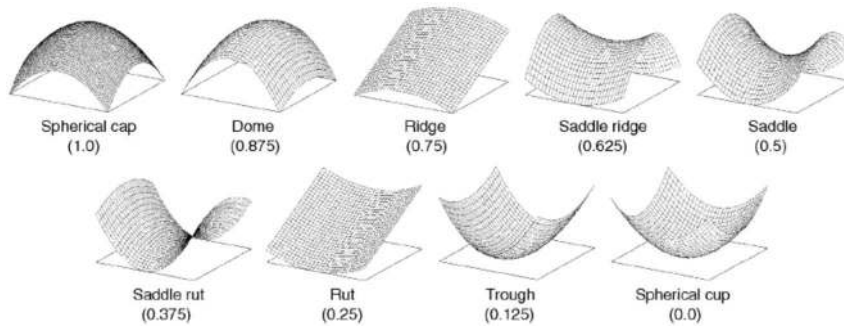

Fig. 6. Nine representative shapes on the shape index scale [30].

and minimum (k_2) local curvature (see (1)). The shape index takes a value in the interval [0, 1].

$$S(p) = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}. \qquad (1)$$

The low end of the shape index scale represents a spherical cup while the high end of the scale represents a spherical cap. In the middle of the scale (value of 0.5) is a saddle point, see Fig. 6. This shape calculation is independent of the coordinate system and, therefore, it is a potentially useful metric for finding similar points between 2.5D face scans with different poses.
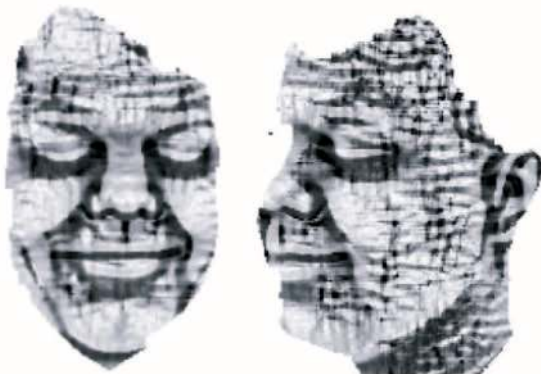

Fig. 7. Shape index images. The dark regions represent lower values and the light regions represent higher values.

Fig. 7 shows the shape index image of two facial scans. Notice that there are several consistencies (correspondences) between these two different scans of the same face. For example, the area between the eyes and the bridge of the nose is consistently trough shaped.

Using a combination of the pose invariant shape index, the 3D coordinates, and the texture image, we have developed heuristics to locate a set of candidate feature points. This set of candidate points is designed to match with similar anchor points already identified on the 3D models. Usually, a large number of (more than the minimum of three) candidate points are selected, many of them being invalid. The general strategy is to use simple heuristics to select a sufficiently large set of points that contains at least three valid feature points and then search this list to find the best three point/label pairs. For example, among all the facial scans, both frontal and profile views,[1] an easy feature point to identify is the inside edge of an eye right next to the bridge of the nose, because this point has a shape index value that is very close to zero and the area around this point has a consistent shape index value across all face images and poses, see Fig. 8 for an illustration.

Simple heuristics are also used to select the candidates for other feature points, e.g., the nose tip. The more in-depth description of the applied heuristics is provided in [31]. The false alarms generated by the simple heuristics are filtered out by the following searching algorithm.

1. The "profile" is used to describe the pose changes greater than 45 degrees from the frontal view.
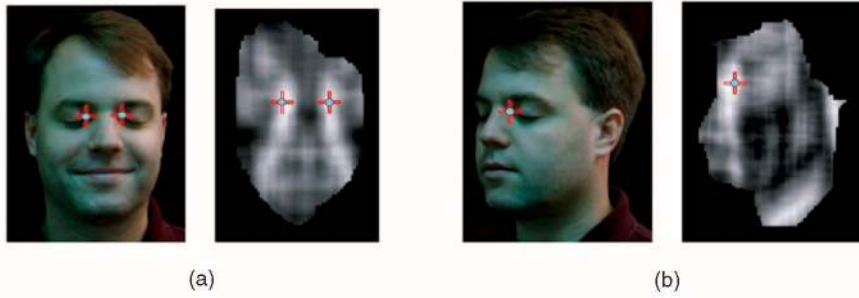
Fig. 8. Identifying feature points in the shape index space. (a) Frontal view. (b) Profile view. The shape index value is inverted for better illustration.

Once we have a list of feature point candidates, the next step is to search through all possible points with all possible labels to find the best point/label triplet that transforms the 2.5D scan onto the 3D model. The best point/label triplet is a set of three points with three labels that generates the best transformation from the 2.5D scan onto the 3D model. In order to evaluate the quality of the transformation, a uniform sample of control points is selected from the 2.5D scan and projected onto the 3D model. The distance from each point to the surface of the 3D model is summed and used as a fitness function. The transformation, which produces the smallest value of the fitness function, is determined to be the best transformation and associated feature points.

An exhaustive search of all possible sets of triplets is not practical. A set of constraints has been developed to prune the search area. We have formulated the problem into a standard Constraint Satisfaction Problem (CSP) and handled the problem using a relaxation-labeling algorithm [32]. The constraints used are based on a flexible model of the relationships between the points on human faces. For example, the distance between the inside of an eye and the outside of the eye is no less than 20 mm and no more than 50 mm; the shape index of the nose tip must be greater than 0.5, etc. Most of these constraints are conservative and are designed to prune the obviously wrong point/label choices. The complete list of the types of constraints used in our feature extractor is given in [31].

Once three points are identified, the coarse alignment transformation can be calculated. There are six parameters to be recovered for the rigid transformation, three for the 3D rotation and three for 3D translation. For simplicity, we use the algorithm in [29], where the transformation with 6 degrees of freedom is decomposed into several simpler transformations, which are easier to estimate and calculate.

3.2 Fine Alignment

The coarse alignment can only provide an approximation of the true registration due to the location displacement of the estimated feature points. But for the purpose of surface matching, the two sets of 3D points (one from 2.5D scan and one from 3D model) should be further tuned for fine registration. Because both the scan and model contain highly dense data, it is possible to find a good approximation of the closest points in each data set, which is the basis of the Iterative Closest Point (ICP) framework [33], [34], [35]. The basic Iterative Closest Point scheme is as follows:

1. Select control points in one point set.
2. Find the closest points in the other point set (correspondence).
3. Calculate the optimal transformation between the two sets based on the current correspondence.
4. Transform the points; repeat step 2, until convergence.

Starting with an initial estimate of the rigid transformation derived in the coarse alignment stage, ICP iteratively updates the transformation parameters by alternately choosing corresponding (control) points in the 3D model and the 2.5D scan and finding the best translation and rotation that minimizes an error function based on the distance between them.

Besl and McKay [33] used point-to-point distance and a close-form solution when calculating the transformation matrix during each iteration. With the KD-tree data structure [36], [37] integrated, the algorithm can be efficiently implemented. The point-to-plane distance used in [34] makes the ICP algorithm less susceptible to local minima than the point-to-point metric [38]. It also needs a fewer number of iterations to converge. But, point-to-plane distance-based ICP has to solve a nonlinear optimization problem using numerical algorithms. Since the scan and model are represented as a dense mesh, the normal for each vertex can be calculated, which makes the computation of point-to-plane distance feasible. We integrate Besl's and Chen's ICP algorithms [33], [34] in a zigzag running style, and call it the hybrid ICP algorithm. Each iteration consists of two steps, using Besl's scheme to compute an estimation of the alignment, followed by Chen's scheme for a refinement. The hybrid algorithm has the potential to combine the advantages of each individual component in terms of the robustness and efficiency.

In order to increase the number of correspondences, regions were selected within the face scans that do not vary greatly between the scans. Fig. 9 shows the grids used for control point selection for various poses. Regions around the eyes and nose were selected because these regions are less malleable than other parts of the face (such as the region around the mouth, which changes greatly with facial expression). The fine alignment results are demonstrated in Fig. 5d. Other nonuniform control point selection schemes, such as curvature-based sampling schemes, can also be applied.

3.3 Surface Matching Distance

The root mean square distance minimized by the ICP algorithm is used as the primary matching distance between a face scan and the 3D model. We use the point-to-plane distance metric MD_{ICP} defined in [34].
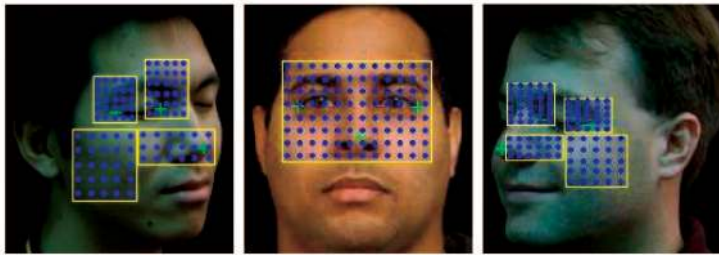
Fig. 9. Automatic control point selection, based on three feature points (shown in "+"), for a left profile, frontal, and right profile scans. About 100 control points are selected in each scan.

$$MD_{ICP} = \sqrt{\frac{1}{N_c} \sum_{i=1}^{N_c} d^2(\Psi(p_i), S_i)}, \qquad (2)$$

where $d(\cdot, \cdot)$ is the point-to-plane metric, $\Psi(\cdot)$ is the rigid transformation applied to each control point p_i in the 2.5D test scan, S_i is the corresponding tangent plane in the 3D model with regards to p_i, and N_c is the number of control points. The smaller the MD_{ICP}, the better the surface matching.

4 CONSTRAINED APPEARANCE-BASED MATCHING

In addition to the 3D shape, texture, contained in a 2D intensity image, is also an important cue for face recognition. There are a number of appearance-based algorithms for image-based face recognition [39], [40], [41]. The typical appearance-based method analyzes the intensity correlation between all the pixels in the image, which is a global characteristic of the face image. The alignment of the training and test images is important to the matching accuracy of the appearance-based algorithms [42]. The ICP registration procedure aligns the test scan and the 3D model, so the pose is already normalized. By synthesizing new appearance (image variations) from the constructed 3D model, additional training samples of the subjects can be obtained. This allows us to use the linear discriminant analysis (LDA) for appearance-based matching [40], [43]. Instead of using all the subjects in the database, the LDA is applied only to a small list of candidates, which is generated dynamically by the surface matching stage for each test scan. We call this as the constrained appearance-based matching in our framework.
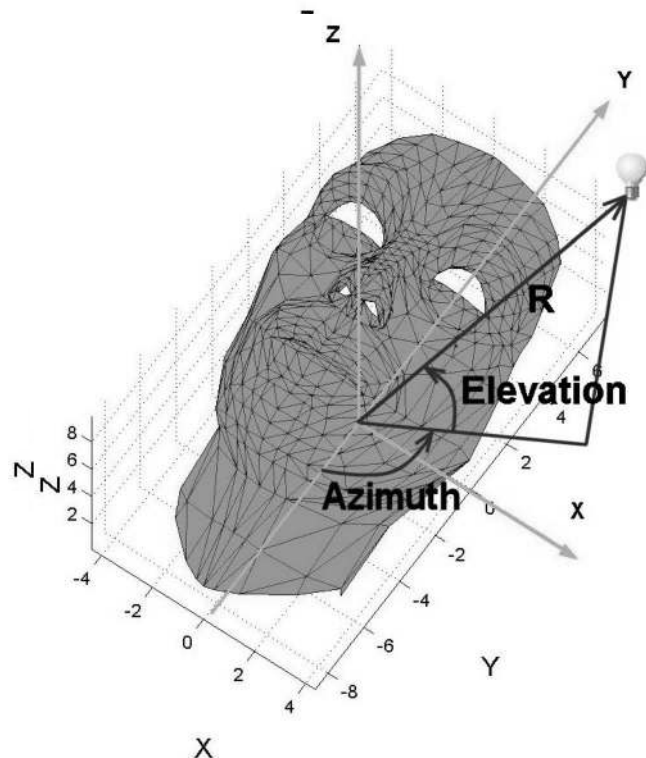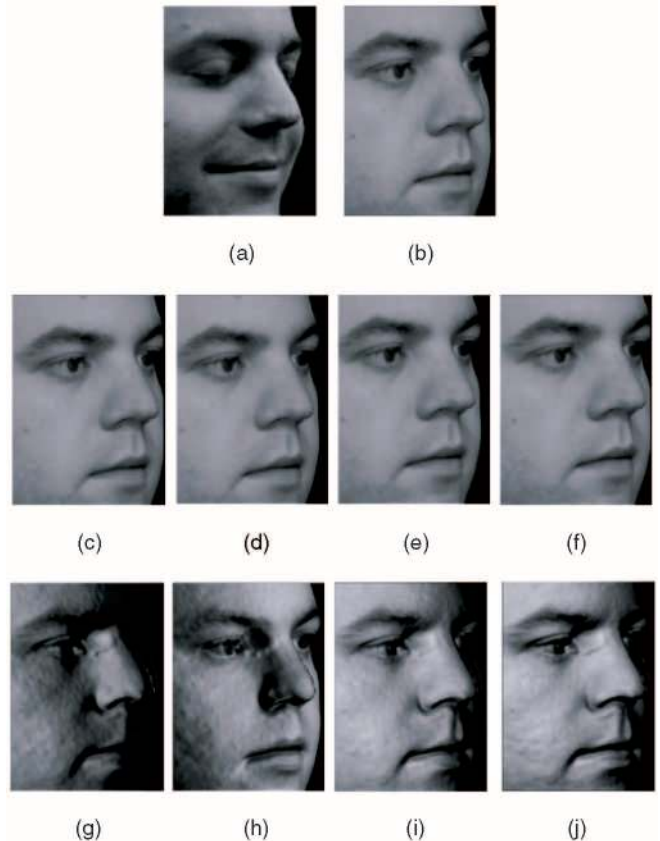

Fig. 11. Cropped synthesized training samples for discriminant subspace analysis. (a) Test (scan) image, (b) image rendered by the 3D model after pose normalization (alignment), (c)-(f) images synthesized by the 3D model with shift displacement in horizontal and vertical directions, and (g)-(j) images synthesized by the 3D model with lighting changes. Only gray scale is used for appearance-based analysis. Because the pose is normalized and feature points are known, the cropping is done automatically.

Fig. 10. Lighting simulation. The light bulb denotes the simulated light source.
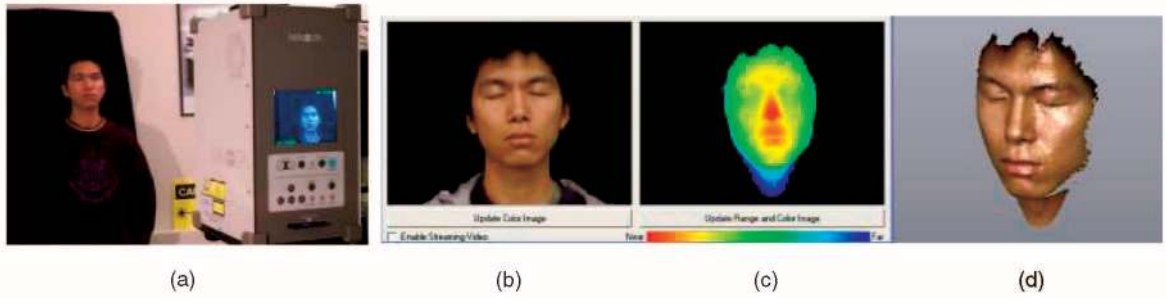
Fig. 12. An example of Minolta Vivid 910 facial scan. (a) Data capture scenario, (b) texture image, (c) range image, showing points closer to the sensor in red, and (d) 3D visualization.
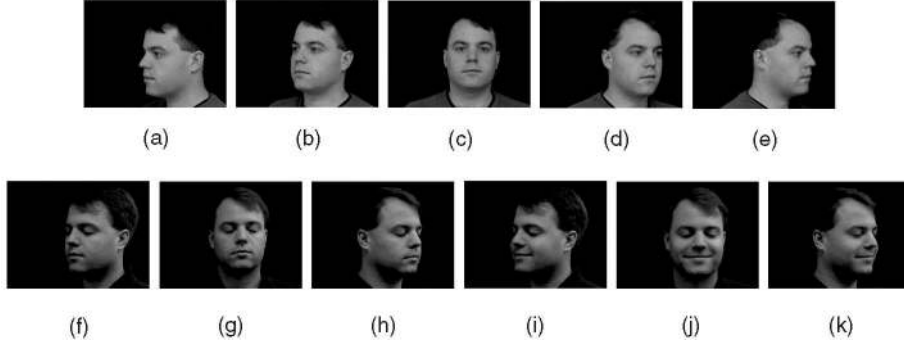

Fig. 13. An example of data collection for each individual. (a)-(e) are used for constructing the 3D model stored in the MSU gallery database. (f)-(k) are used for testing, which contains variations in pose, lighting, and expression (smiling).

4.1 Appearance Synthesis

Each subject is represented by a 3D face model with neutral expression in the database. In order to apply the subspace analysis based on the facial appearance, many training samples, which are aligned with the test sample, are needed [40], [43]. After the surface registration (pose normalization), the 3D model has been aligned with the test scan. Since the dense 3D model is available, it is easy to synthesize new appearance with lighting variations. As the alignment may not be perfect, small pose variations are also synthesized in our framework.

Synthesis of pose variations is straightforward by simply rotating and shifting the 3D model. Lighting is simulated by adding a virtual light source around the reconstructed face surface as illustrated in Fig. 10. The position of the light source is controlled by the distance R between the light source and the origin of the model coordinate system and by the azimuth and elevation angles. Different illumination variations are generated by changing the position of the light source. Phong shading technique is employed to render lighting effects on the face surface [44]. Due to the sensor noise, smoothing filters can be applied to improve the appearance synthesis results.

Based on the feature points (eye corners and the nose tip) and registration results, the critical area in the face is

TABLE 2
Test Data Distribution

	Frontal	Profile	Subtotal
Neutral	99	213	312
Smiling	98	188	286
Subtotal	197	401	598

determined, which is used to automatically crop the synthesized images. Examples of the cropped synthesized images are shown in Fig. 11. These images are used in the following discriminant subspace analysis.

4.2 Linear Discriminant Subspace Analysis

A two-dimensional face image is considered as a vector by concatenating each row (or column) of the image. Let $X = (x_1, x_2, \ldots, x_i, \ldots, x_N)$ denote the data matrix, where N is the number of face images in the training set. Each x_i is a face vector of dimension n, concatenated from a $p \times p$ face image, where n represents the total number of pixels in the face image, $n = p \times p$. The Linear Discriminant Analysis (LDA) [45], [40] representation is a linear transformation from the original image vector to a projection feature vector, i.e.,

$$Y = W_{LDA}^T X, \qquad (3)$$

where Y is the $d \times N$ feature vector matrix, d is the dimension of the feature vector, $d \ll n$, and W_{LDA} is the transformation matrix derived by

$$W_{LDA} = \arg\max_W \frac{W^T S_B W}{W^T S_W W}, \qquad (4)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix,

$$S_B = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T, \qquad (5)$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T. \qquad (6)$$
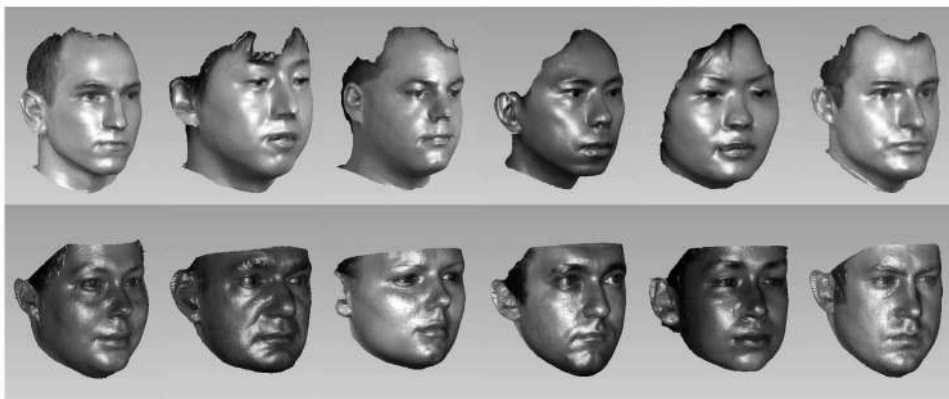
Fig. 14. Some of the 3D face models in the MSU database (top) and the USF database (bottom).

Fig. 15. Representative 2.5D test scans.

TABLE 3
Rank-One Matching Accuracy for Different Categories of Test Scans

Test scan category	Surface matching	Constrained LDA	Surface matching + constrained LDA
Frontal & Neutral	98% (2)	86% (14)	99% (1)
Profile & Neutral	96% (7)	84% (35)	98% (5)
Frontal & Smiling	68% (31)	71% (28)	77% (23)
Profile & Smiling	76% (45)	69% (59)	84% (31)

The total number of test scans in each category is listed in Table 2. The number of errors is provided in the parenthesis. The weights for the surface matching and the constrained appearance matching components are set to be equal (i.e., $\alpha = 1$ in (7)).

In (5) and (6), N_i is the number of training samples in class i, c is the number of classes (number of subjects), m is the mean vector of all the samples, i.e., $m = \frac{1}{N}\sum_{i=1}^{N} x_i$, m_i is the mean vector of samples belonging to class i, and X_i represents the set of training samples belonging to class i.

In the face recognition problem, the within-class scatter matrix S_W can be singular, due to the fact that the rank of S_W is at most $(N - c)$ and the number of training samples is generally smaller than the dimensionality of the face image (number of pixels). In such a situation, PCA [39] can be used to reduce the dimensionality of the original face image space [40] prior to applying LDA.

The projection coefficients in LDA are used as the feature representation of each face image. The matching score between the test face image and the training image is calculated as the cosine value of the angle between their coefficients vectors.

4.3 Dynamic Candidate Selection

In the conventional LDA, all the subjects in the training database (gallery) are used for subspace construction. When the number of subjects is large, the complexity of the recognition problem is increased due to large intra-class variations and large interclass similarities, resulting in a low recognition accuracy. However, if the number of subjects in the gallery is small, the appearance-based method can provide a relatively good performance. For each individual test scan, the gallery used for subspace analysis and matching is dynamically generated based on the output of the surface matching. Only a small number of subjects is selected for the appearance-based matching, so that the number of subjects to be matched to the test scan in LDA is small. In our experiments, the top M ($M = 30$) candidates in the sorted matching list are selected (rank-based selection).

5   INTEGRATION

Surface matching and appearance-based matching provide two matching metrics based on different cues. Each of them can be considered as a classifier. Since they explore different properties of the face object, namely, shape and texture, a combination of these two classifiers has the potential to outperform each individual classifier [46]. Another possibility is to combine these scores using the well-known sum rule after

TABLE 4
Matching Accuracy with Equal Weights for
ICP and LDA Components (i.e., $\alpha = 1$ in (7))

Scheme	Rank-one match accuracy
Surface matching	86%
Constrained LDA	77%
Surface matching + Constrained LDA	90%

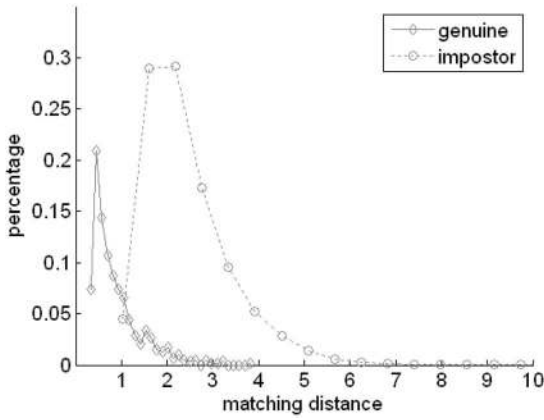*The total number of test scans is 598.*

Fig. 16. Surface matching distance distributions.

z-score normalization [47], [48]. We applied the simple weighted sum rule to integrate the surface matching and appearance-based matching distances.

$$MD_{comb} = MD_{ICP} + \alpha \cdot MD_{LDA}, \tag{7}$$

where $MD_{LDA} = (1 - MS_{LDA})/2$, normalized to $[0,1]$, MS_{LDA} is the matching score generated by the appearance-based matching component, converting the matching score (similarity) to matching distance (dissimilarity). The weighting parameter α balances the two matching components, which can be set beforehand or learned from an independent validation data set.

6 EXPERIMENTS AND DISCUSSION

6.1 Data

Currently, there is no publicly available *multiview* range (with registered texture) face database, along with expression variations. All range images (downsampled to 320×240 with a depth resolution of $\sim 0.1mm$) were collected using a Minolta Vivid 910 scanner [17]. The subject stands in front of the scanner at a distance of about 1.5 m. This scanner uses structured laser light to construct the face image in less than one second. Each point in a scan has a 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). Fig. 12 shows the data collection scenario and an example of these scans.

There are currently 100 subjects in our database (MSU database). Five scans with neutral expression for each subject were captured to construct the 3D model. For each subject, another six scans were captured for testing, including three scans with neutral expression and three with smiling expression. The scan protocol used for each subject is demonstrated in Fig. 13. For a few subjects, we had fewer than six test scans. In total, the test database consists of 598 independent scans (different from training scans) of the same 100 subjects. All the scans varied in pose and facial expression (smiling). In addition to MSU database, another 100 3D full-view face models with neutral expression captured by the Cyberware scanner [16] are available to us, which are provided by the University of South Florida (USF) [49], named USF database. By combining these two databases, we have a total of 200 3D models stored in the gallery database and 598 2.5D scans for testing. For the USF database, no 2.5D test scans are available. The test data distribution is listed in Table 2. Representative 3D models and test scans are shown in Figs. 14 and 15, respectively.

We applied the three ICP algorithms, Besl's [33], Chen's [34], and our hybrid ICP, on the entire database. The number of surface matching errors are 98 (Besl's), 88 (Chen's), and 85 (hybrid). Based on these results, we have decided to use the hybrid ICP algorithm in the following experiments.

6.2 Matching Performance

Based on the three feature points, control points are automatically sampled for the ICP registration. Fig. 9 showed the control point sampling scheme. Examples of the registration results were given in Figs. 5c and 5d. The surface matching was achieved using the distance score produced by the ICP registration. Our matching process is conducted in the identification mode. Each scan is matched to all the 3D models stored in the gallery.

Conventional appearance-based algorithms suffer from large pose changes [1], [14], and their performance depends on the quality of the alignment. In our matching scheme, after the surface matching, the test scan and 3D models are aligned,

Fig. 17. Test scans (top row), and the corresponding 3D models correctly matched. The 3D model is shown in a pose similar to the corresponding test scan.
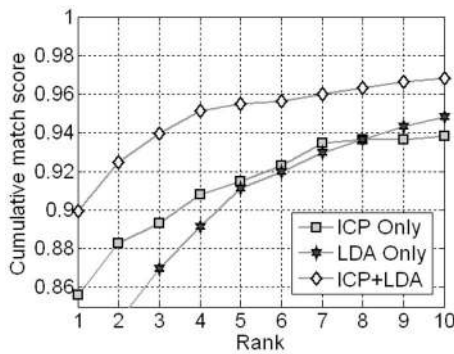
Fig. 18. Cumulative matching performance with equal weights for the surface matching (ICP) and the constrained appearance matching (LDA) components (i.e., $\alpha = 1$). The LDA component is constrained by the surface matching (ICP) component. The LDA is only applied to the top 30 candidate models selected in the surface matching stage.
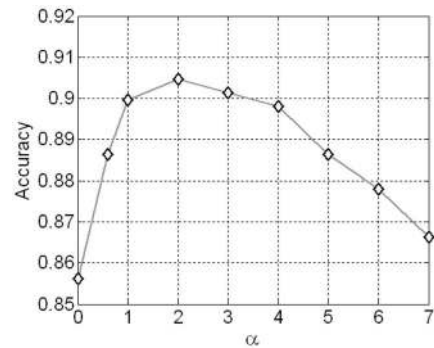

Fig. 19. Identification accuracy based on the combination strategy with respect to α, the parameter used to balance the surface matching and appearance matching. A higher accuracy is achieved at $\alpha = 2$ than the 90 percent accuracy at $\alpha = 1$.

which makes the appearance-based algorithm more feasible. In the constrained appearance-based matching stage, although the number of synthesized samples, which are used as the training samples for the appearance-based methods, can be arbitrary large in principle, in practice, we only generate a small number of samples because this synthesis process and the subsequent LDA need to be conducted online. In our experiments, four images with different shift displacements and four images with different lighting conditions were synthesized. Hence, nine images for each model are used for the LDA calculation (eight synthesized versions plus the original one, see Figs. 11b, 11c, 11d, 11e, 11f, 11g, 11h, 11i, and 11j for an example).

The LDA is only applied to the first 30 matched candidates based on the surface matching distance. By applying surface matching and constrained appearance-based scheme separately to the data set, we found that the sets of misclassified test scans are significantly different for these two matching schemes, implying that these two schemes are not highly correlated. Therefore, a suitable fusion of these two schemes has the potential to lead to an improvement in the matching accuracy.

The performance of the proposed matching scheme depends on the nature of the test scan. Our current feature extractor still has difficulties in robustly locating all the feature points in the test scans with pose variations. In order to evaluate the proposed matching scheme, we study the feature extraction and matching components separately. The coarse alignment here is performed using manually picked feature points. The matching results are summarized in Table 3 and explained below:

1. **Frontal Scan with Neutral Expression**. In this category, all test scans are frontal, with neutral expression, which is similar to the expression contained in the 3D models. The surface matching achieves 98 percent accuracy on these test scans. The constrained appearance-based method also achieves the highest accuracy (86 percent) among all the categories listed in Table 3, due to the good alignment results and very little change in the expression. A combination of surface matching and appearance-based matching gives an accuracy of 99 percent.

2. **Profile Scan with Neutral Expression**. Although both surface matching and appearance-based matching components perform a little bit worse than the frontal case, we still attain an accuracy of 96 percent for the surface matching and 98 percent for the combination scheme. The lower performance here compared to the frontal cases is due to the smaller overlap between the 2.5D test scan and 3D models.

3. **Scans with Smiling Expression**. Regardless of pose variations, the expression changes, which change the facial shape, decrease the surface matching accuracy drastically. This is mainly because our ICP-based surface matching is focused on extracting the rigid transformation parameters, while the facial expression change is a typical nonrigid transformation. Although the appearance-based method can handle the facial expression changes to some extent, its performance depends on the quality of the alignment (pose normalization), which is provided by the surface matching component. Still, surface matching and appearance-based matching augment each other and their combination leads to 81 percent accuracy.

The expression change affects both sides of the face. According to our current control point sampling scheme, the frontal case has a larger facial area whose shape is changed more by the expression than the profile views. This could be one reason for a lower surface matching accuracy in the frontal smiling category compared to the profile smiling test scans.

In all the three categories of the test scans, the combination of surface matching and appearance-based matching outperforms each individual matching component.

6.3 Performance Summary

A summary of the experimental results for the entire data set consisting of 598 test scans is given in Table 4, running in the identification mode. Out of the 60 errors over the entire test database (corresponding to 90 percent accuracy), 54 test scans contain smiling expression. As mentioned earlier, the expression change leads to nonlinear surface deformation that is not adequately handled by the rigid transform-based ICP algorithm. The surface matching distance distributions for genuine users and impostors are provided in Fig. 16. Fig. 17 shows four correctly matched examples using the combined scheme.[2]

---

2. Due to IRB regulations, we cannot show all the test subjects in the database.
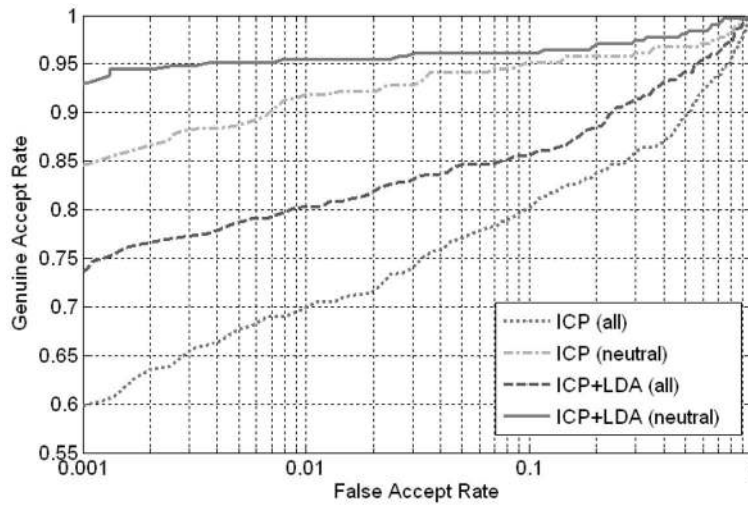
Fig. 20. ROC curves. ICP (all): surface matching on the entire test database; ICP (neutral): surface matching on the test scans with neutral expression. LDA is applied only after pose normalization by ICP rigid registration. Equal weights (i.e., $\alpha = 1$) were applied to the surface matching (ICP) and the constrained appearance-based matching (LDA) components.

The cumulative match score curves for the three different matching schemes are provided in Fig. 18. The combination of surface matching (ICP only) and constrained appearance-based matching (LDA only) consistently outperforms each individual scheme.

The performance reported in Table 4 is based on setting equal weights to surface matching and appearance-based matching distances, i.e., the value of α in (7) is set to 1. However, there may exist an optimal value of α, which minimizes the number of errors. The performance change with respect to α is shown in Fig. 19. In practice, the value of α can be learned from the validation data.

Using the matching distances computed from matching 598 test scans to 200 3D face models, the ROC curves are generated, which are provided in Fig. 20. The curves are calculated by setting the same threshold for all the users. A user-specific threshold could be computed for each user to yield better performance [50]. Note that the appearance-based matching (LDA) in Fig. 19 relies on the 3D pose alignment achieved by ICP-based registration.

In our current implementation, matching one test scan to a 3D face model takes about 20 seconds, on an average, using a Pentium 4 2.8GHz CPU. The speed bottleneck is the nearest neighbor search in ICP, because the computation required for sequential (exhaustive) search for one control point is proportional to N, where N is the number of vertices in the model. We have integrated the KD-tree structure[3] [36], [37] with the Besl's ICP algorithm [33]. The expected computation to perform the nearest-neighbor search for each control point is then proportional to $\log N$. If we use only Besl's ICP algorithm in the surface matching stage instead of the proposed hybrid ICP algorithm, the entire matching process can be achieved in approximately 4 seconds with about 2 percent decrease in the identification accuracy. Unlike the point-to-point (Euclidean) distance-based Besl's ICP algorithm, the point-to-plane distance-based Chen's ICP algorithm cannot be integrated with the KD-tree structure. The nearest-neighbor search in ICP can be implemented in parallel for each control point, so parallel computation and hardware accelerators can also be utilized. With the current

computation power, the proposed scheme would be more suitable for identification on a small database or verification applications. For identification in a large database, fast screening or indexing approaches would need to be integrated.

7 CONCLUSIONS AND FUTURE WORK

We have presented a face recognition system that matches 2.5D scans of faces with different pose and expression variations to a database of 3D neutral face models. Both shape and texture information contained in 3D models are employed. We propose a combination scheme, which integrates surface (shape) matching and a constrained appearance-based method for face matching that complement each other. The surface matching is achieved by a hybrid ICP scheme. The subsequent appearance-based identification component is constrained to a small candidate list generated by the surface matching component, which reduces the classification complexity. The registered 3D model is utilized to synthesize training samples with facial appearance variations, which are used for discriminant subspace analysis. The matching distances obtained by the two matching components are combined using the weighted sum rule to make the final decision. Regardless of the pose, lighting, and expression, given the feature points, the entire matching scheme is fully automatic, including surface registration/matching, dynamic candidate list selection, 3D synthesis, sample image cropping, LDA, and appearance-based matching.

This research is an encouraging first step in designing a system that is capable of recognizing faces with arbitrary pose. Nonrigid deformation such as expression is a challenge to the current system. More sophisticated surface matching schemes are being pursued to improve the surface matching accuracy [51] and speed. We are exploring 3D models that can be deformed to deal with nonrigid variations [52], [15]. To make the entire matching system fully automatic, a more robust and accurate feature point locator is being developed.

---

3. The KD-tree software package is provided by Guy Shechter.

ACKNOWLEDGMENTS

REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," VL Technical Report, Univ. of Maryland, ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz. Oct. 2000.

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

[3] J.Y. Cartoux, J.T. LaPreste, and M. Richetin, "Face Authentication or Recognition by Profile Extraction from Range Images," *Proc. Workshop Interpretation of 3D Scenes,* pp. 194-199, 1989.

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

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

[6] B. Achermann, X. Jiang, and H. Bunke, "Face Recognition Using Range Images," *Proc. Int'l Conf. Virtual Systems and MultiMedia,* pp. 129-136, 1997.

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

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

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

[10] C. Hesher, A. Srivastava, and G. Erlebacher, "PCA of Range Images for Facial Recognition," *Proc. 2002 Int'l Multiconf. Computer Science,* 2002.

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

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

[13] X. Lu, D. Colbry, and A. Jain, "Three-Dimensional Model Based Face Recognition," *Proc. Int'l Conf. Pattern Recognition,* pp. 362-366, 2004.

[14] Face Recognition Vendor Test (FRVT), http://www.frvt.org/, 2002.

[15] 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.

[16] Cyberware Inc., http://www.cyberware.com/, 2005.

[17] Minolta Vivid 910 noncontact 3D laser scanner, http://www.minoltausa.com/vivid/, 2005.

[18] *Proc. Fifth Int'l Conf. 3-D Digital Imaging and Modeling (3DIM),* http://www.3dimconference.org/, 2005.

[19] Q. Chen and G. Medioni, "Building 3-D Human Face Models from Two Photographs," *J. VLSI Signal Processing,* vol. 27, pp. 127-140, 2001.

[20] Z. Zhang, "Image-Based Modeling of Objects and Human Faces," *Proc. SPIE,* vol. 4309, pp. 1-15, Jan. 2001.

[21] B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju, "Model-Based 3D Face Capture with Shape-from-Silhouettes," *Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures,* pp. 20-27, Oct. 2003.

[22] M. Dimitrijevic, S. Ilic, and P. Fua, "Accurate Face Models from Uncalibrated and Ill-Lit Video Sequences," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1034-1041, 2004.

[23] 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,* Oct. 2003.

[24] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," *Proc. ACM SIGGRAPH,* July 1994.

[25] C. Dorai, G. Wang, A.K. Jain, and C. Mercer, "Registration and Integration of Multiple Object Views for 3D Model Construction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 83-89, Jan. 1998.

[26] P. Liepa, "Filling Holes in Meshes," *Proc. Eurographics/ACM SIGGRAPH Symp. Geometry Processing,* pp. 200-205, 2003.

[27] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher, "Geometric Surface Smoothing via Anisotropic Diffusion of Normals," *Proc. Visualization '02,* 2002.

[28] Geomagic Studio, http://www.geomagic.com/products/studio/, 2005.

[29] 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, Univ. of Utah, Mar. 1998.

[30] C. Dorai and A.K. Jain, "Cosmos—A Representation Scheme for 3D Free-form Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 10, pp. 1115-1130, Oct. 1997.

[31] D. Colbry, X. Lu, A. Jain, and G. Stockman, "3D Face Feature Extraction for Recognition," Technical Report MSU-CSE-04-39, Computer Science and Eng., Michigan State Univ., East Lansing, Sept. 2004.

[32] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 5, no. 3, pp. 267-287, 1983.

[33] P. Besl and N. McKay, "A Method for Registration of 3D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 239-256, Feb. 1992.

[34] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing,* vol. 10, no. 3, pp. 145-155, 1992.

[35] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *Int'l J. Computer Vision,* vol. 13, no. 1, pp. 119-152, 1994.

[36] J. Bentley, "Multidimensional Binary Search Trees Used for Associative Searching," *Comm. ACM,* vol. 18, no. 9, pp. 509-517, 1975.

[37] J.H. Friedman, J.L. Bentley, and R.A. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," *ACM Trans. Math. Software,* vol. 3, no. 3, pp. 209-226, 1977.

[38] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, "Geometrically Stable Sampling for the ICP Algorithm," *Proc. Int'l Conf. 3D Digital Imaging and Modeling,* Oct. 2003.

[39] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, Mar. 1991.

[40] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[41] M.S. Bartlett, H.M. Lades, and T.J. Sejnowski, "Independent Component Representations for Face Recognition," *Proc. SPIE,* vol. 3299, pp. 528-539, 1998.

[42] S. Shan, Y. Chang, W. Gao, and B. Cao, "Curse of Mis-Alignment in Face Recognition: Problem and a Novel Mis-Alignment Learning Solution," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 314-320, 2004.

[43] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 228-233, Feb. 2001.

[44] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer Graphics: Principles and Practice,* second ed. New York: Addison-Wesley, 1996.

[45] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification,* second ed. New York: Wiley, 2000.

[46] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226-239, Mar. 1998.

[47] K.-P. Li and J.E. Porter, "Normalizations and Selection of Speech Segments for Speaker Recognition Scoring," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* pp. 595-597, 1998.

[48] S. Bengio, J. Mariethoz, and S. Marcel, "Evaluation of Biometric Technology on XM2VTS," technical report, Dalle Molle Inst. for Perceptual Artificial Intelligence, 2001.

[49] USF HumanID 3D Face Dataset, courtesy of Sudeep Sarkar, Univ. of South Florida, Tampa, 2005.

[50] A.K. Jain and A. Ross, "Learning User-Specific Parameters in a Multibiometric System," *Proc. IEEE Int'l Conf. Image Processing,* pp. 57-60, 2002.

[51] X. Lu and A. Jain, "Deformation Analysis for 3D Face Matching," *Proc. Seventh IEEE Workshop Applications of Computer Vision,* 2005.

[52] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 135-164, 2004.

Xiaoguang Lu received the BS and MS degrees from the Department of Automation, Tsinghua University, China, in 1997 and 2000, respectively. From 2000 to 2001, he was a visiting student at Microsoft Research Asia. He is currently a doctoral student in the Department of Computer Science and Engineering at Michigan State University. He worked as a summer research intern at Siemens Corporate Research and Mitsubishi Electric Research Laboratories in 2002 and 2003, respectively. His research interests include pattern recognition, computer vision, and machine learning. He is a student member of the IEEE.

Anil K. Jain received the BTech degree from the Indian Institute of Technology, Kanpur, in 1969 and the MS and PhD degrees from The Ohio State University in 1970 and 1973, respectively. He is a University Distinguished professor in the Departments of Computer Science and Engineering at Michigan State University. He served as the department chair during 1995-1999. His research interests include statistical pattern recognition, data clustering, texture analysis, document image understanding, and biometric authentication. He received awards for best papers in 1987 and 1991, and for outstanding contributions in 1976, 1979, 1992, 1997, and 1998 from the Pattern Recognition Society. He also received the 1996 *IEEE Transactions on Neural Networks* Outstanding Paper Award. He was the Editor-in-Chief of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1991-1994). He is a fellow of the IEEE, the ACM, and the International Association of Pattern Recognition (IAPR). He has received a Fulbright Research Award, a Guggenheim fellowship, and the Alexander von Humboldt Research Award. He delivered the 2002 Pierre Devijver lecture sponsored by the International Association of Pattern Recognition (IAPR) and received the 2003 IEEE Computer Society Technical Achievement Award. He holds six patents in the area of fingerprint matching, and he is the author of a number of books: *Biometric Systems, Technology, Design, and Performance Evaluation* (Springer 2005), *Handbook of Face Recognition* (Springer, 2005), *Handbook of Fingerprint Recognition* (Springer, 2003) (received the PSP award from the Association of American Publishers), *BIO-METRICS: Personal Identification in Networked Society* (Kluwer 1999), *3D Object Recognition Systems*, (Elsevier, 1993), *Markov Random Fields: Theory and Applications* (Academic Press, 1993), *Neural Networks and Statistical Pattern Recognition* (North-Holland, 1991), *Analysis and Interpretation of Range Images* (Springer-Verlag, 1990), *Algorithms For Clustering Data* (Prentice-Hall, 1988), and *Real-Time Object Measurement and Classification* (Springer-Verlag, 1988). He is an associate editor of the *IEEE Transactions on Information Forensics and Security* and is currently serving as a member of the study team on *Whither Biometrics* being conducted by the National Academies (CSTB).

Dirk Colbry received the BME degree from The Georgia Institute of Technology in 1997. He worked for two years as a robotics engineer for FANUC Robotics North America in Rochester Hills, Michigan. In 2001, he received the MSE degree from the University of Michigan, where his research focused on artificial intelligence and intelligent planning systems. Currently, he is enrolled in the doctoral program of the Department of Computer Science and Engineering at Michigan State University (MSU). He is also earning an interdepartmental graduate specialization in cognitive science through the MSU Cognitive Science Program and has been awarded an US National Science Foundation IGERT Associate Fellowship. His research interests include computer vision, robotics, cognitive science, and artificial intelligence. He is a student member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.