Robust 3D Face Recognition by Local Shape Difference Boosting

Yueming Wang, Jianzhuang Liu, *Senior Member*, *IEEE*, and Xiaoou Tang, *Fellow*, *IEEE*

Abstract—This paper proposes a new 3D face recognition approach, *Collective Shape Difference Classifier* (CSDC), to meet practical application requirements, i.e., high recognition performance, high computational efficiency, and easy implementation. We first present a fast posture alignment method which is self-dependent and avoids the registration between an input face against every face in the gallery. Then, a *Signed Shape Difference Map* (SSDM) is computed between two aligned 3D faces as a mediate representation for the shape comparison. Based on the SSDMs, three kinds of features are used to encode both the local similarity and the change characteristics between facial shapes. The most discriminative local features are selected optimally by boosting and trained as weak classifiers for assembling three collective strong classifiers, namely, CSDCs with respect to the three kinds of features. Different schemes are designed for verification and identification to pursue high performance in both recognition and computation. The experiments, carried out on FRGC v2 with the standard protocol, yield three verification rates all better than 97.9 percent with the FAR of 0.1 percent and rank-1 recognition rates above 98 percent. Each recognition against a gallery with 1,000 faces only takes about 3.6 seconds. These experimental results demonstrate that our algorithm is not only effective but also time efficient.

Index Terms—3D shape matching, collective shape difference classifier, face recognition, signed shape difference map.

✦

---

1 INTRODUCTION

AUTOMATIC face recognition has apparent advantages over other biometric technologies due to the natural, non-intrusive, and high throughput properties in face data acquisition. Over the past three decades, much effort has been made on 2D face recognition using intensity images as input data. Although some 2D face recognition systems have good performance under constrained conditions, face recognition is still a great challenge due to variations in illumination, pose, and expression [2], [34] since an intensity face image is a projection of a 3D face with one dimension lost.

With the rapid development and dropping cost of 3D digital acquisition devices, 3D face data, which represents faces as 3D point sets or range data, can be captured more quickly and accurately [7]. The use of 3D information in face recognition has attracted great attention and various techniques have been presented in recent years [4], [7]. Since 3D face data contain explicit 3D geometry, more clues can be used to handle the variations of face pose and expression. Thus, 3D face recognition is expected to overcome the challenges facing 2D face recognition and improve the system performance [1].

1.1 Related Work

Although there are two extensive surveys of 3D face recognition methods published in 2005 [4] and 2006 [7],

new progress has been made since then [42], [40], [43], [45]. Here, we give a brief review of the previous work in three parts. The first part looks at methods that appeared relatively early and were tested on relatively small or nonpublic data sets. Approaches in the second part focus on handling expression variations, which is a great challenge in 3D face recognition. The last part concentrates on the work tested on the whole FRGC v2 database [38], which is the largest public data set for comparing different methods. The three parts may have some overlaps. Besides, we focus on face recognition only by 3D face shapes. For some multimodal methods mentioned, only their 3D face engines are discussed.

1.1.1 Basic Methods

The methods discussed in this section mostly handle relatively small or nonpublic data sets. Some methods work under special conditions (e.g., manually labeled landmarks). We roughly categorize them into four classes according to the types of the features.

1. *Facial curve-based*. This kind of method extracts representative facial curves from 3D face data as features to help surface matching or to replace surface matching with curve matching for recognition. The key problem is how to develop the schemes of curve extraction and matching. Using five feature points, Nagamine et al. [8] aligned face range data and found three curves at which the face surface intersects vertical and horizontal planes and a cylinder. The euclidean distance was used for curve matching. In [33], a symmetry plane was determined to help compute the bilateral profile of the face and the matching of profile was based on a modified Hausdorff distance measure. Beumier and Acheroy [32] searched for the central profile with maximal protrusion and two parallel profiles along side. Local curvature values along the profile curves were

---

- *Y. Wang is with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, and the Qiushi Academy for Advanced Studies, Zhejiang University, China. E-mail: ymingwang@gmail.com.*
- *J. Liu and X. Tang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, and Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. E-mail: {jzliu, xtang}@ie.cuhk.edu.hk.*

estimated to compare the profiles. Recently, Samir et al. [9] proposed a method which represents a face surface by the union of the level curves of a depth function. Then a metric was defined for comparing the unions based on the similarity between curves.

2. *Shape descriptor-based.* Shape descriptors are designed to be invariant to rotation to encode the local surface shapes. Chua et al. [27] used point signature, a representation of free-form surfaces for 3D face recognition in which the rigid parts of a face are extracted to deal with different facial expressions. A similar representation named local shape map was proposed in [28]. Tanaka et al. [15] adopted an extended Gaussian image (EGI) as a mediate feature after curvature-based segmentation on which principal directions are mapped as local features. The spherical correlation was used to compare EGIs. The curvature shrinks a local neighborhood to a point and can be seen as a special shape descriptor.

3. *Holistic matching-based.* Achermann et al. [17], Bronstein et al. [36], and Hesher et al. [18] explored PCA on images obtained by the projections of 3D faces. Medioni and Waupotitsch [1], Lu et al. [13], [21], Wang et al. [12], Chang et al. [5], and Maurer et al. [29] matched face surfaces by the iterative closest point (ICP) algorithm [3] or its modified versions. The Hausdorff distance was also used for holistic surface matching [16], [19], [20].

4. *Prominent regions (points)-based.* In these methods, prominent regions or points of the facial surface are detected to form feature vectors. Comparison between two faces is made based on their relationship in the feature space. Gordon [14] extracted both high level features (such as eyes and nose) and low level ones (such as distances and curvatures) to build a feature vector. By the signs of the mean and Gaussian curvatures, Moreno et al. [10] segmented the face surface into regions and lines on which features were constructed. A feature vector was formed by a feature selection procedure. Gupta et al. [11] manually located anthropometric facial fiducial points and used a feature vector based on the anthropometric distances between the points for recognition. Xu et al. [24] converted a 3D face into a regular mesh and constructed an intrinsic feature vector to encode the 3D shape. AdaBoost was used for selecting the most effective local features. Zhong et al. [25] calculated Gabor features from each divided patch of a 3D face and made a learned visual codebook (LVC) based on the results of K-means clustering on the filter response vectors. By LVCs, recognition was achieved with a nearest neighbor classifier.

1.1.2 Methods Handling Expression Variations

Expression variations cause the distortion of the facial surface and have been an impediment to achieve high performance in 3D face recognition. Motivated by the fact that distortions do not cover the whole facial surface, Chang et al. [6] selected three regions around the nose, which are believed to be relatively rigid for matching. Based on ICP, the matching scores were fused by the product rule. The results show that the scheme improves the performance compared

with the ICP baseline. Chua et al. [27] also investigated the rigid parts for recognition. However, how to find an optimal group of rigid parts with sufficient discriminating power is still a problem [7]. An alternative method is to set up an approximate expression model and, for a given nonneutral face, a virtual 3D face is generated by the model which suppresses the expression effect. Recognition is performed using the virtual 3D face. Bronstein et al. [37] assumed that facial expressions can be modeled as isometries of facial surfaces. An expression-invariant face representation named canonical form was constructed by geodesic distances and multidimensional scaling (MDS). Lu et al. [21], [22] introduced a deformable model into facial surface matching. By a control group, a neutral face was transferred to several faces which were synthesized into a deformable face before being fitted to a test scan. Using the Poisson equation and a rigid constraint, Wang et al. [40] proposed a guidance-based constraint deformation model (GCD) for deformation from a nonneutral face to a neutral one. This kind of work is interesting and improves the performance to some extent, but it is computationally demanding.

1.1.3 Methods Tested on FRGC v2

Since the Face Recognition Grand Challenge (FRGC) started [38], public 3D face databases and predefined settings of experiments have been available. There are 466 people and 4,007 3D faces in FRGC v2 for validation and the design of Experiment 3 aims to examine the performance of a 3D face recognition method which includes three settings, ROC I, ROC II, and ROC III masks. The target of FRGC is an order of magnitude increase in performance over FRVT 2002 [34], which means a verification rate of 98 percent at the fixed FAR of 0.1 percent. More and more recent work used FRGC v2 as experimental data. Such a large data set forces researchers to further improve their algorithms and the same protocol allows for comparison between different methods.

Husken et al. [30] explored 2D and 3D multimodal face recognition. Each modality uses hierarchical graph matching (HGM) to evaluate the performance. A verification rate of 89.5 percent was obtained with the FAR of 0.1 percent on the ROC III mask by their 3D verification engine. A PCA 3D face modeling method was proposed by Russ et al. [53] in which the synthesis of 3D faces by the PCA basis vectors can potentially correct some nonrigid distortion caused by facial expression. Maurer et al. [29] tried the ICP method on the whole FRGC v2 database (4,007 versus 4,007 full cross run) and obtained an 87 percent verification rate with the FAR of 0.1 percent. A log-Gabor template method proposed by Cook et al. [31] achieves a verification rate of 93.71 percent on ROC I, 92.91 percent on ROC II, and 92.01 percent on ROC III. Kakadiaris et al. [42] proposed a deformable model to fit a 3D face to aligned 3D facial scans and measured the difference between two faces. Based on the fitting results, geometry and normal map images were constructed. Then the Haar wavelets and pyramid transform were applied to extract features for comparison. Verification rates of about 97 percent were reported on ROC I, ROC II, and ROC III. Mian et al. [43] investigated fusion of 2D and 3D methods for recognition. The eyes-forehead and nose regions were segmented for a 3D classifier, with a rejection classifier constructed by a spherical face representation and SIFT for efficiency. High verification

rates of about 97-99 percent were reported at a 0.1 percent FAR when the neutral models were placed in the gallery with their R3D algorithm. However, they did not give the results of the FRGC v2 standard protocol, namely, Experiment 3. Faltemier et al. [45] introduced a matching fusion of 28 overlapped small regions of the facial surface with Borda Count and Consensus Voting methods. A rank-1 recognition rate of 97.2 percent was reported when the earliest scans were put into the gallery and a verification rate of 94.8 percent at a 0.1 percent FAR was obtained on ROC III of Experiment 3. Mian et al. [44] constructed graphs from the local feature set based on detected key points. The similarity between two faces was measured as the similarity between their graphs. Both 2D and 3D data of faces are used for recognition. For the 3D engine alone, 93.5 percent rank-1 recognition rate and 97.4 percent verification rate were achieved on the "Neutral versus All" experiment with FRGC v2. Boehnen et al. [54] recently proposed a 3D signature method for face recognition. A 3D signature is a vector with its elements being some fixed surface points. Faces are preregistered to a reference face in a canonical coordinate system, obviating the need of iterative alignment for each matching. Eight 3D signatures corresponding to eight predefined regions are constructed and used for fusion during matching. It achieves 95.5 percent rank-1 recognition rate on the FRGC v2 database with the earliest images constituting the gallery.

Although the recent algorithms for 3D face recognition have reached relatively high performance on large databases [42], [43], [45], [35], challenges still exist.

The first problem is how to efficiently extract the main facial region from the raw face data with arbitrary posture. The raw data may contain hair, shoulders, and neck as well as some artifacts such as holes and spikes [7]. It is very important since the cropping error usually cannot be rectified by later procedures. Mian et al. [43] proposed a coarse-to-fine approach to detect the nose tip for facial region cropping. Several slices of the raw data were generated on which nose tip candidates were found. A facial region was cropped by a sphere positioned at the selected nose tip. However, the face has to be placed in the top-bottom posture, and the slice computing and repeated searching on them are time consuming. Faltemier et al. [45] located the nose tip by a consensus of two or three methods, e.g., ICP and curvature analysis.

The second problem is how to quickly and precisely align two face models. The ICP algorithm is frequently used for 3D face alignment [45], [43], [29]. However, it suffers from facial surface distortion due to expression variations and noise. Besides, the iterative process makes ICP computationally expensive and the registration must be done for each model in the gallery, which is not suitable for the identification task with a large gallery.

The third problem is how to optimize the combination of small rigid facial regions for matching to reduce the effect of expressions. It has been proven that recognition using rigid facial regions can improve the performance on a data set with expression variations [45], [43], [6]. However, the selection of rigid regions was based on intuition and the optimal selection and combination are still unsolved.

The last problem is computational efficiency. More information in 3D face data leads to more computational cost. Although some effective techniques can work in

verification scenario with time cost of about 10 seconds on a common PC [42], [45], fast identification is still a problem when thousands of faces are in the gallery.

1.2 Overview of Our Method

This paper proposes a new 3D face recognition approach, called *Collective Shape Difference Classifier* (CSDC), to deal with the problems described above. The main contributions of our work are summarized as follows:

1. A fast and effective face posture alignment technique is presented to place all face models to a standard position and orientation. This alignment is self-dependent and precise for recognition. It avoids the registration between an input face against every face in the gallery so that high computational performance in face identification can be obtained. A fast face cropping method is also proposed.

2. A *Signed Shape Difference Map* (SSDM) is defined between two depth images sampled from two aligned 3D face models. The different parts of the SSDM do not contribute the same discriminability due to nonrigid distortion on the face. Besides, the shape changes caused by the same expression may have similar patterns for different persons. The characteristics of these two aspects on SSDMs are described by three features, Haar-like feature [48], Gabor feature [49], and Local Binary Pattern (LBP) [50], on different positions of SSDMs. Based on the intrapersonal and interpersonal SSDMs, the most discriminative local features are chosen by a boosting algorithm [41] and used to build three strong CSDCs.

3. During evaluation, the multiclass 3D face recognition problem is converted to a two-class classification problem, i.e., interpersonal and intrapersonal classes, similar to the case in 2D face recognition [52]. Different schemes are developed for verification and identification. Three CSDCs are fused to obtain a high verification rate, and only the most effective CSDC, Haar feature-based CSDC is used for efficient face identification.

The framework of our method is shown in Fig. 1. The requirements of an applicable 3D face algorithm are considered in our method, e.g., fast self-dependent alignment and optimally selected local regions to overcome expression variations.

Our method has been tested on the FRGC v2 database according to the standard protocol. With ROC I, ROC II, and ROC III masks in FRGC v2, verification rates of above 97.9 percent with the FAR of 0.1 percent are obtained and the rank-1 recognition rate of above *98.3 percent* is achieved. These results are comparable to (more exactly, slightly better than) the best approaches published. Besides, not only verification but also identification (1,000 models in the gallery) can be finished within 4 seconds. Thus, our work is of both high recognition performance and computational efficiency in 3D face recognition.

The paper is an extension and improvement of our previous work in [46]. The rest of this paper is organized as follows: Section 2 presents the methods of denoising and facial region cropping. Section 3 gives the detail of our fast alignment approach. In Section 4, we explain the representation of shape difference. The training and testing methods
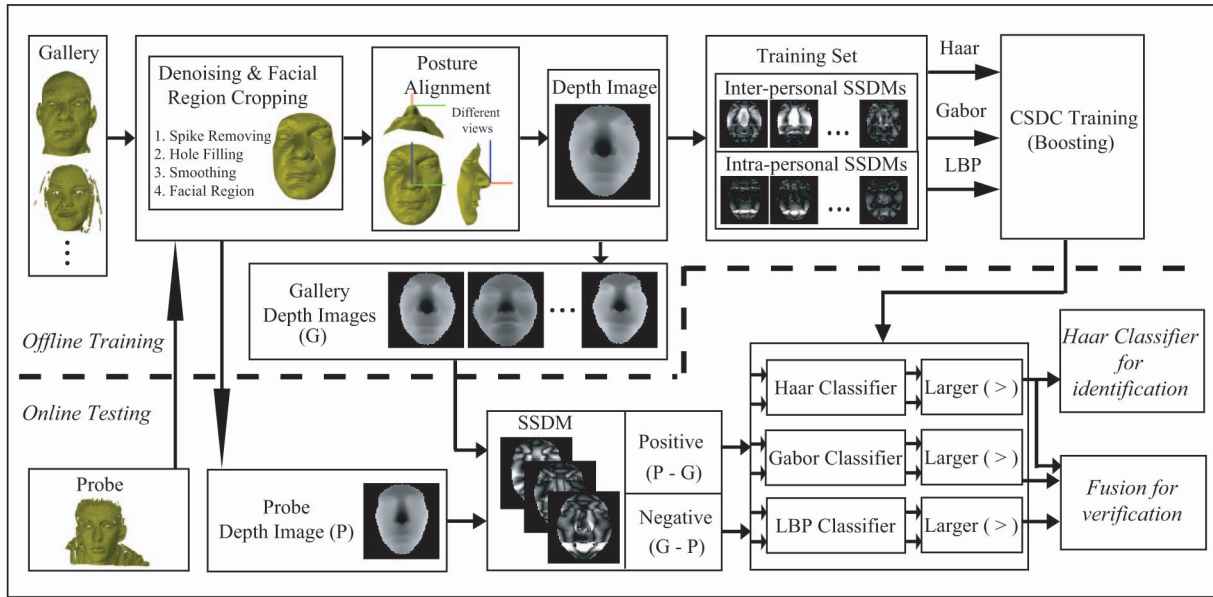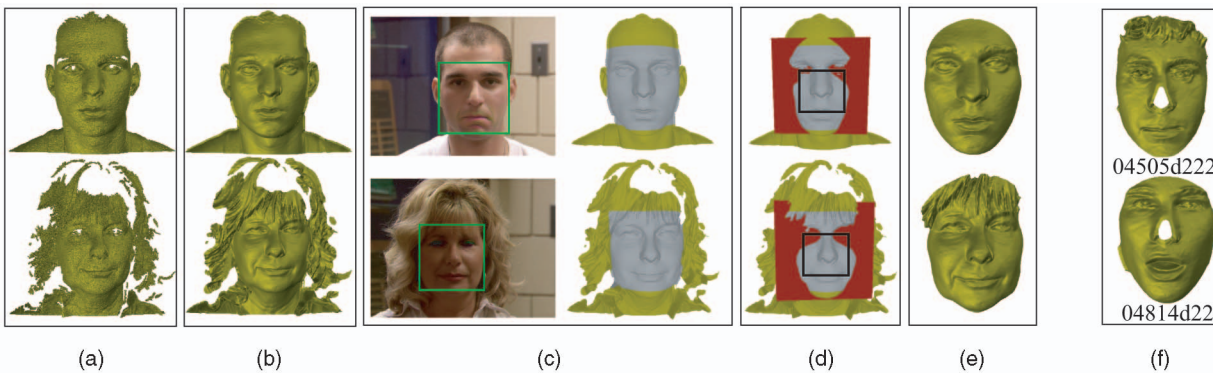
Fig. 1. The framework of our method.

Fig. 2. Denoising and facial region cropping. (a) 3D raw data. (b) Results after denoising. (c) Face detection on the texture channel. (d) Fitted planes on the 3D data corresponding to the detected 2D face regions. (e) Cropped regions. (f) Correctly cropped examples with big holes in nose regions.

for multiple features are described in Section 5. Experimental results are shown in Section 6. Finally, Section 7 draws the conclusion.

2 DENOISING AND FACE CROPPING

Many commercial 3D scanners can take a range sample and a registered 2D texture image [38]. Since the texture channel and the 3D face data correspond well, Chang et al. [6] applied a skin detection method on the texture channel to help 3D facial region extraction. We also preform 3D facial region cropping with the help of the texture channel. Unlike Chang et al.'s work, we use the face detection method on 2D images [47] to achieve 3D face extraction. It should be mentioned that we only use 2D images to assist 3D face cropping. The steps of denoising and cropping are described as follows:

1. Three Gaussian filters are designed to remove spikes, fill small holes, and smooth the data with different variances, as shown in Fig. 2b.
2. On the texture channel, a face region is detected and the corresponding 3D points of the region in the 3D raw data are labeled as a set F_c (see Fig. 2c).

3. A plane Σ_c is fitted using the points in F_c, which cuts these data into two parts. For each part, we compute a mean squared distance (MSD) from the points in the part to its centroid. The one with the smaller MSD is selected as another set F_c'. As shown in Fig. 2d, among the points in both F_c' and the middle $1/4$ area of the detected face region, the point with the largest distance to the plane Σ_c is selected as the approximate nose tip $\mathbf{p}_{nt}$.
4. By placing a sphere centered at $\mathbf{p}_{nt}$, the facial region is cropped from the denoised face data (see Fig. 2e).

This method includes the effects of removing the hair, shoulders, and neck. It is simple and fast, and works well on the FRGC v2 database. All 4,007 models in FRGC v2 are correctly handled. Even for some samples with big holes, the correct facial regions are detected, as shown in Fig. 2f. In comparison, a recent method in [45] generates 72 errors for the same database.

3 SELF-DEPENDENT ALIGNMENT

The proposed alignment is carried out by finding the facial symmetry plane. With the normal of the symmetry plane, nose tip, and the direction of nose bridge, six degrees of
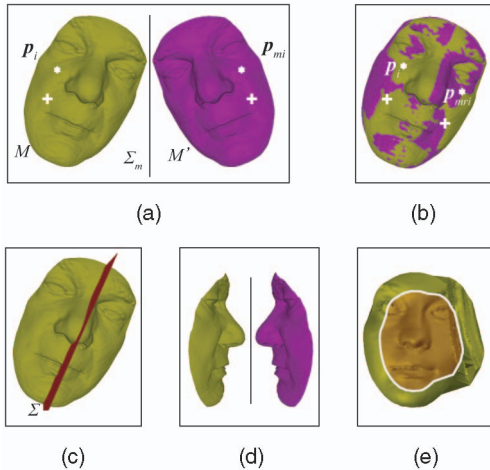
Fig. 3. Symmetry plane finding. (a) An input face M and its mirror M'. (b) Registration of M' to M. (c) The symmetry plane Σ. (d) A special mirror plane which may lead to nonconvergence in ICP. (e) Points inside the closed white curve used for registration.

Fig. 4. Symmetry plane detection for two difficult samples in the FRGC v2 data set. (a) 04339d300. (b) 04760d76.

freedom in a 3D face are fixed and thus a standard alignment posture can be obtained.

3.1 Symmetry Plane Determination

Let $M = \{\mathbf{p}_i \mid \mathbf{p}_i = (x_i, y_i, z_i)^T, 1 \leq i \leq N\}$ denote the point set of a cropped 3D facial model and Σ be the face's symmetry plane. Suppose that $\mathbf{p}_{si}$ is the mirror image of $\mathbf{p}_i \in M$ about Σ. Clearly, Σ can be determined if we know $\mathbf{p}_{si}$ for each $\mathbf{p}_i$.

With respect to a certain plane Σ_m, a mirror of M, denoted by M', is calculated with $\mathbf{p}_i$ mapped to $\mathbf{p}_{mi}$, as shown in Fig. 3a. Then we register M' to M with M fixed, which transforms $\mathbf{p}_{mi}$ to $\mathbf{p}_{mri}$. It can be seen from Fig. 3b that $\mathbf{p}_{mri}$ is exactly $\mathbf{p}_{si}$ (the mirror image of $\mathbf{p}_i$ about Σ) if the registration is correct. Thus, $\mathbf{p}_{si}$ can be obtained by the two transformations. We summarize the method of finding the symmetry plane in Algorithm 1.

Algorithm 1. Finding the symmetry plane of a face model
Input: A cropped face model M.
Procedure:
  1) Mirror M about a certain plane Σ_m to
     $M' = \{\mathbf{p}_{mi} \mid \mathbf{p}_{mi} = (x_{mi}, y_{mi}, z_{mi})^T, 1 \leq i \leq N\}$, where
     the correspondence between $\mathbf{p}_i$ and $\mathbf{p}_{mi}$ is naturally
     set up.
  2) Register M' to M, which converts M' to $M'' = \{\mathbf{p}_{mri} \mid \mathbf{p}_{mri} = (x_{mri}, y_{mri}, z_{mri})^T, 1 \leq i \leq N\}$.
  3) Obtain the symmetry plane Σ by fitting it on the set of
     points $B = \{\mathbf{p}_i^b \mid \mathbf{p}_i^b = (\mathbf{p}_i + \mathbf{p}_{mri})/2, 1 \leq i \leq N\}$.
Output: The symmetry plane Σ of M.

The three steps can be well understood by Figs. 3a, 3b, and 3c. In step 1, the plane Σ_m for the computation of M' should be carefully chosen. Consider the special pose of the input facial model and Σ_m shown in Fig. 3d. This mirror plane can lead to nonconvergent registration in step 2. Observing the shape of the face, we find that the largest variance happens in the face top-bottom direction, while the smallest variance occurs in the normal direction of the front face. Hence, by principal component analysis (PCA) on M, three eigenvectors, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, are obtained with their corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, where $\mathbf{v}_1$ approximately corresponds to the face top-bottom direction, $\mathbf{v}_3$ corresponds to
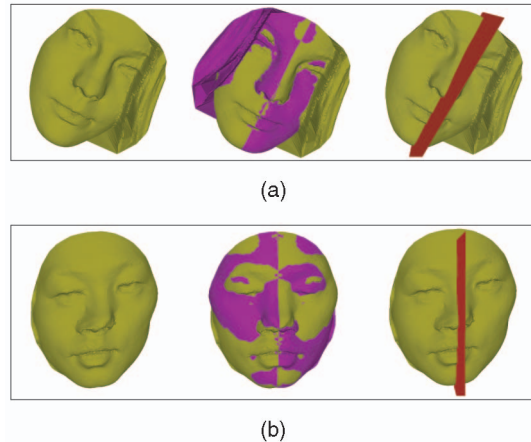
the face normal direction, and $\mathbf{v}_2$ is perpendicular to both $\mathbf{v}_1$ and $\mathbf{v}_3$. The mirror plane Σ_m is chosen as the plane passing through the centroid of M and with its normal being $\mathbf{v}_2$. Such a plane passes through M, making M' and M already roughly coincident, which leads to fast convergence in step 2.

In step 2 of Algorithm 1, we use ICP [3] for the registration between M and M'. In many previous methods, ICP is used to register an input model to every model in the gallery set for matching. Thus, the number of ICP executions is the same as the number of models in the gallery. However, in our method, we only run it once for an input face.

It should be mentioned that our algorithm is based on the fact that the human face is self-symmetric. However, the extracted facial region may not be so ideal, especially along the boundary, as shown in Fig. 3e. To guarantee better convergence and registration with ICP, the points close to the boundary are discarded in the registration. Our experiments show that this method of symmetry plane determination is sufficiently robust. Two difficult examples are given in Fig. 4 where the 3D face data are not so symmetric.

3.2 Finding the Nose Tip and Nose Bridge Direction

3.2.1 Nose Tip Location

With the symmetry plane Σ of the face model M, the central profile C is obtained by the intersection between Σ and M (see Fig. 5). The nose tip is one of the most distinctive landmarks on M and can be located from C.

As stated in Section 2, the fitted plane Σ_c of the detected facial region F_c cuts the facial data into two parts. The part with the smaller MSD contains the nose. Thus, a point, $\mathbf{p}_t$, is considered as the nose tip if it satisfies the following three conditions:
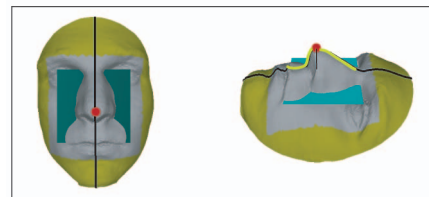

Fig. 5. Locating the nose tip on the profile by fitting a plane to the detected facial region.
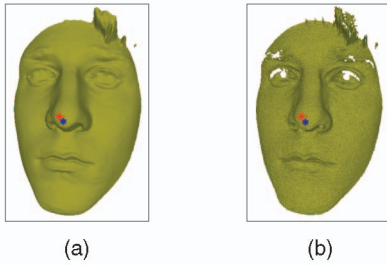
Fig. 6. Nose tip detection on sample 04531d399. The points marked with "+" are the detection results and the points "*" are the ground truth. (a) A hole filled and smoothed sample where "+" and "*" are 4.58 mm apart. (b) A noisy sample where "+" and "*" are 5.01 mm apart.

1. $\mathbf{p}_t$ is on the profile C and in the middle $100 \text{ mm} \times 100 \text{ mm}$ area of the detected facial region F_c;
2. $\mathbf{p}_t$ is in the part with the smaller MSD;
3. among all points meeting 1 and 2, $\mathbf{p}_t$ is the point with the largest distance to the plane Σ_c.

In this method, the symmetry plane limits the search for the nose tip on the central profile. The facial region avoids the effects of the hair and the bias of the face cropping. The plane Σ_c reduces the effects of pose changes. Besides, the spike removal, hole filling, and smoothing steps help accurately find the symmetry plane and Σ_c.

To evaluate the precision of our nose tip location method, we manually label the nose tips for the samples in FRGC v2 as the ground truth. Two samples with big holes on the noses (see Fig. 2f) are not included in the evaluation. In the first experimental setting (a), the spike removal, hole filling, smoothing, and cropping are applied to all 4,005 samples before the nose tip detection. To test the stability of our method to noise, in another setting (b), only the spike removal and cropping are conducted. Two examples in the two settings are shown in Fig. 6. Lu et al. [22] proposed a method of nose tip detection based on directional maximum. In their scheme, the nose tip and the profile of a face are determined with mutual aid. The nose tip is assumed to be both with the largest z value for the front pose and on the central profile, while the central profile is assumed to pass through the nose tip and is identified by a subspace method. We have implemented their method and compared it with our method.

It should be noted that the nose tip actually is a tiny region for which different people may give slightly different labeling results. Thus, the distances away from the ground truth are computed and the numbers of detected nose tips falling in different distance ranges away from the ground truth are shown in Table 1. In setting (a), our method obtains 99.75 percent nose tips within 4 mm away from the ground

truth. All results do not exceed a distance of 6 mm from the ground truth. In setting (b), there is only one sample with the result more than 12 mm away. However, the method in [22] obtains 63 and 75 nose tips that are 12 mm or more away from the ground truth in the two settings, respectively. From Fig. 6, it can be seen that 4-5 mm displacement is quite small. Thus, our method of nose tip detection is accurate and robust enough for recognition.

3.2.2 Nose Bridge Direction

Along the profile C, the nose bridge direction is a stable feature across a variety of expressions and the changes of face shapes. It can be found by the following three steps, as shown in Fig. 7:

1. A coarse direction $\mathbf{n}_p$ along the profile C is determined by PCA with the points on C, which is the eigenvector corresponding to the largest eigenvalue. Along $\mathbf{n}_p$, the points on C are sorted and the nose tip $\mathbf{p}_t$ separates C into two parts (see Fig. 7a).
2. On each side of C separated by $\mathbf{p}_t$, the points on C within a distance d_p from $\mathbf{p}_t$ form a set P_{pl} or P_{pr}. Using the least square method, two lines are fitted for P_{pl} and P_{pr} (see Fig. 7b).
3. The side with the nose bridge should result in a smaller fitting error and the direction $\mathbf{n}_y$ of the corresponding line is selected as the nose bridge direction (see Fig. 7b).

The first step is to make our method work for different face poses. In our experiments, d_p is set to 40 mm.

3.3 Standard Posture

With the nose tip $\mathbf{p}_t$, the nose bridge direction $\mathbf{n}_y$, and the unit normal $\mathbf{n}_\Sigma$ of the symmetry plane Σ known, the posture of a face model can be fixed. Next, we show how to transform the face model to a special posture and position, called the standard posture. As shown in Fig. 8e, in the standard posture, the nose tip is placed at the origin, the top-bottom direction of the face is along the y axis, the face normal

Fig. 7. Finding the nose bridge direction. (a) A coarse direction $\mathbf{n}_p$ of the profile C and the two parts of points near the nose tip $\mathbf{p}_t$. (b) Line fitting and the direction $\mathbf{n}_y$ of the nose bridge.

TABLE 1
Numbers of Detected Nose Tips Falling in
Different Distance Ranges Away from the Ground Truth

distance away from the ground truth	[0, 4) mm	[4, 6) mm	[6, 12) mm	$\geq$ 12 mm
Our method (a)	3995 (99.75%)	10	0	0
Our method (b)	3976	26	2	1
Method in [22] (a)	3937	5	0	63
Method in [22] (b)	3902	28	0	75

(a) Spike removal, hole filling, and smoothing are carried out. (b) Only spike removal is conducted. (total 4,005 samples)
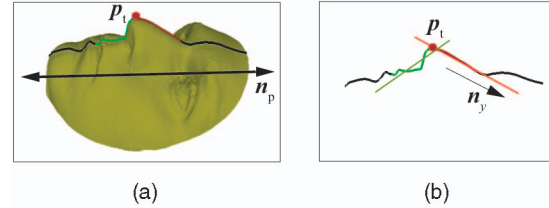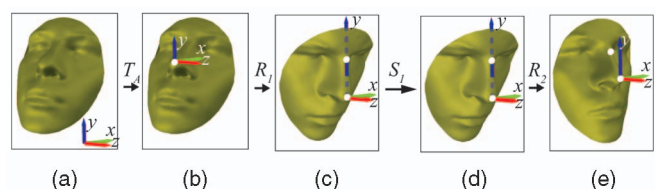

Fig. 8. Transformation of a face model to the standard posture. (a) The face model with an arbitrary posture. (b) Result after translation T_A. (c) Result after rotation R_1. (d) Result after transformation S_1. (e) Final standard posture after rotation R_2.

direction is along the z axis, and the x axis is naturally set up according to the right-hand coordinate system.

The transformation of a face to the standard posture is represented by $R_s = R_2 S_1 R_1 T_A$, and given a point $(x, y, z)^T$, its coordinate in the standard posture is computed by $(x^*, y^*, z^*, 1)^T = R_s(x, y, z, 1)^T$, where the four matrices are defined as follows:

$$T_A = \begin{bmatrix} I_{3\times3} & -\mathbf{p}_t \\ \mathbf{0} & 1 \end{bmatrix}, \tag{1}$$

$$R_1 = \begin{bmatrix} \mathbf{n}_\Sigma^T & 0 \\ \mathbf{n}_y^T & 0 \\ \left(\frac{\mathbf{n}_\Sigma \otimes \mathbf{n}_y}{|\mathbf{n}_\Sigma \otimes \mathbf{n}_y|}\right)^T & 0 \\ \mathbf{0} & 1 \end{bmatrix}, \tag{2}$$

$$S_1 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{or} \quad I_{4\times4}, \tag{3}$$

$$R_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) & 0 \\ 0 & sin(\alpha) & cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where $I_{3\times3}$ and $I_{4\times4}$ are two identity matrices, $\otimes$ denotes the cross product of two vectors, and $\alpha = -30°$. The geometric meanings of these transformations are illustrated in Fig. 8. After the transformation of $R_1 T_A$, if the majority of the data points have positive z coordinates, we set S_1 to the first matrix in (3) so that the x axis is toward the left-hand side of the face. Otherwise, S_1 is set to $I_{4\times4}$.

Our posture alignment is based on the determination of $\mathbf{p}_t$, $\mathbf{n}_y$, and $\mathbf{n}_\Sigma$. Clearly the denoising step in Section 2 is important to support the accurate detection of these features for the alignment. With the aligned samples after denoising, it can be seen that our alignment method is accurate enough for high performance from the results of our recognition experiments (see Section 6). Here we analyze the reliability of the alignment by comparing the alignment results between denoised and noisy samples (see Fig. 6 for one sample pair). The average distance between each pair is computed after the alignment. Among the 4,005 samples with full noses in FRGC v2, 4,002 sample pairs have their average distances from 0.77 mm to 2.58 mm. The other three average distances are 9.62 mm, 10.1 mm, and 24.9 mm, respectively, caused by the displacements of the detected nose tips on the noise data (see Table 1). This result shows that 99.93 percent of the faces in FRGC v2 with the original noise can be aligned reliably. It also verifies the effectiveness of our method of determining the nose bridge direction. Thus, our alignment method works well on common faces of approximate mirror-symmetry with a full nose.

4 SHAPE DIFFERENCE REPRESENTATION

Based on the aligned models, we investigate shape differences and convert 3D face recognition to a two-class classification problem, i.e., the problem of determining whether a shape difference is interpersonal or intrapersonal. To represent the shape difference between two aligned
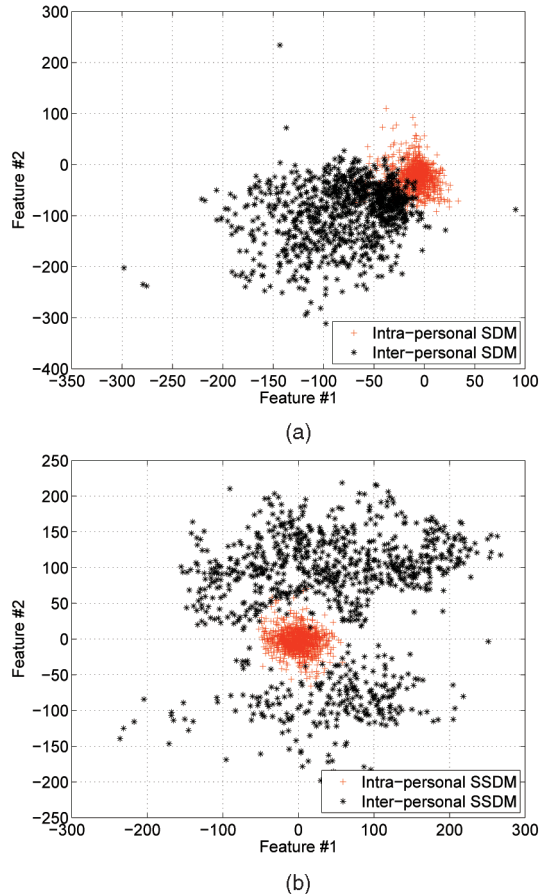

Fig. 9. The distributions of the first two features from (a) SDMs and from (b) SSDMs, respectively, where there are 2,000 randomly selected samples, with 1,000 being intrapersonal and 1,000 being interpersonal.

models, we need to generate their depth images. By a sphere with radius r centered at the nose tip, the region of interest is picked out and projected to a $w \times w$ depth image with the nose tip at the center of the image (we choose $w = r = 75$ in our experiments).

Let I be a depth image, $I(i, j)$ be the depth value at position (i, j). Here, we consider two shape difference representations: signed shape difference map (SSDM) and shape difference map (SDM). Given two depth images I_1 and I_2, their SSDM D_s and SDM D are defined as $D_s(i, j) = I_1(i, j) - I_2(i, j)$ and $D(i, j) = |I_1(i, j) - I_2(i, j)|$, respectively. The SSDM contains richer information of the difference. For example, it has the information of whether some areas of one shape are upon/below those of the other. It helps to capture not only the shape difference, but also its change pattern. Fig. 9 shows the distributions of the first two Haar features on SDMs and SSDMs (see Section 5 for the detailed discussion of the features). It can be seen that there is less overlap between intrapersonal and interpersonal features from SSDMs than that from SDMs, which makes SSDMs easier to be classified.

However, using I_1-I_2 or I_2-I_1 to compute the SSDM may give different recognition results. Our strategy is to keep consistency in training (for example, subtract a "small" expression face from a "large" expression one) and to compute SSDMs with both I_1-I_2 and I_2-I_1 during testing, as shown in Fig. 1. More detail can be found in Section 5. This scheme is not trivial and improves verification performance remarkably (see the experiments).
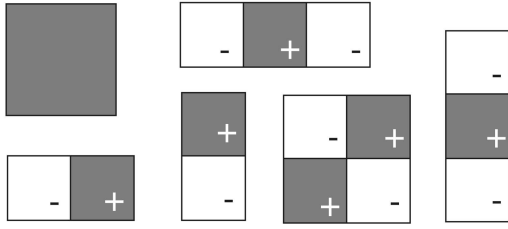
Fig. 10. Haar features used in the CSDC, where the first one denotes the average of a square region.

M-LBP = 2+8+16+128=154

Fig. 11. The computation of multiblock LBPs.

5 COLLECTIVE SHAPE DIFFERENCE CLASSIFIER

The SSDM keeps the information of the similarity between two face models and the Root-Mean-Square (RMS) is a choice for the dissimilarity measure. However, the RMS runs into trouble when there are noise and expression distortion on facial surfaces.

It is obvious that different parts of the SSDM have different contributions to recognition. Although we do not know which areas are the most discriminative across a broad range of expression variations, the boosting algorithm can help select and combine them with suitable weights. This is the main idea of the proposed classifier, called *Collective Shape Difference Classifier* (CSDC). The CSDC is a collective classifier of the form, $H_T(D_s) = \sum_{t=1}^{T} c_t(D_s)$, where D_s denotes a SSDM, $c_t(D_s)$ is a weak classifier selected based on the local features on the SSDMs during the boosting training [41], and T is the number of the weak classifiers. The output of $c_t(D_s)$ is a real value, i.e., confidence, and the final summed confidence is used as the similarity measure between the two 3D face models yielding D_s.

5.1 Local Features

Three kinds of features are used based on the SSDM which are expected to capture the different characteristics of the SSDM. First, Haar-like features [48] are used for their good computational efficiency, which measure the differences between region averages of the SSDM. To reduce the number of features, square filters other than rectangle filters are used in our framework. Besides, the averages of square regions with different sizes on the SSDM are also used, as shown in Fig. 10. Without confusion, we also call them Haar features. Different region averages reflect the local difference between two shapes and the differences of region averages encode the change pattern of the shape difference.

Gabor features are chosen as the second kind of local features to encode the characteristics of a set of spatial localities and orientations of the SSDM. The Gabor kernels are defined as [49]

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\|k_{u,v}\|^2 \|z\|^2/(2\sigma^2)} (e^{ik_{u,v}z} - e^{-\sigma^2/2}), \quad (5)$$

where $z = (x, y)$, $\|\cdot\|$ is the norm operator, u and v denote the orientation and scale, respectively, and

$$k_{u,v} = \frac{k_{max}}{f^v} e^{i\pi u/8}, \quad (6)$$

where f is the spacing factor between the kernels and k_{max} is the maximum frequency. Five scales and six orientations are used in our experiments.

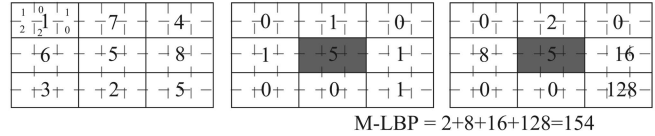The local binary patterns (LBPs) [50] are able to describe the texture characteristics of the SSDM. The LBP is extended to multiblock LBP in [51], which is used as the third kind of features constructed by comparing the average of a central rectangle with the averages of its neighbor rectangles (see Fig. 11).

Two advantages of the local features should be emphasized. 1) The positions of the local features selected from the SSDM in fact correspond to the local regions of the original 3D face. Thus, the CSDC essentially uses rigid patches on the 3D face surfaces for expression-invariant matching and the rigid patches are optimally selected. 2) The local features (Haar, Gabor, and LBP) of the SSDM also capture the characteristics of intraclass shape change of 3D faces besides the similarity of facial shapes.

5.2 Learning and Testing

In the learning of the CSDC, intrapersonal and interpersonal SSDMs are built from given 3D face models, which compose the training set Q. When computing intrapersonal SSDMs, we always subtract a face with "small" expression from another face with "large" expression so as to keep consistency for the SSDMs. The BU-3DFE database [39] is used for training in our work, which provides the expression level for each face.

Usually, the size of Q is very large, mainly due to many different pairs of interpersonal depth images. It is impractical to use all SSDMs in Q for training simultaneously. Thus, bootstrapping is used in learning by starting with all of the intrapersonal and part of the interpersonal SSDMs which form a subset Q_w of Q. Then we keep exchanging the interpersonal SSDMs between Q_w and Q so that all interpersonal samples can be used during the learning procedure. The detail of the learning is shown in Algorithm 2. The three kinds of features, Haar, Gabor, and LBP, are used for training individually and three CSDCs, H_T^{Haar}, H_T^{Gabor}, and H_T^{LBP}, are finally obtained.

Algorithm 2. Collective Shape Difference Classifier Training

Input:
1) $Q = \{(X_i, Y_i) \mid 1 \leq i \leq v\}$ and $Q_w = \{(X_i, Y_i) \mid 1 \leq i \leq m + n\}$, where X_i is the ith SSDM and $Y_i = 1$ or -1 indicates X_i is an intra-personal or inter-personal sample, Q and Q_w contain all samples and starting samples, respectively, and m and n are the numbers of the starting intra-personal and inter-personal samples, respectively.
2) T: the target number of the weak classifiers.

Initialization:

$$w_{0,i} = \begin{cases} \frac{1}{2m}, & Y_i = 1 \\ \frac{1}{2n}, & Y_i = -1 \end{cases}, 1 \leq i \leq m + n.$$

Learning:
For $t = 1, 2, \ldots, T$

1) Normalize the weights:

$$w_{t,i} = \begin{cases} \frac{1}{2} \frac{w_{t-1,i}}{\sum_i w_{t-1,i}}, Y_i = 1 \\ \frac{1}{2} \frac{w_{t-1,i}}{\sum_i w_{t-1,i}}, Y_i = -1 \end{cases}, 1 \le i \le m+n.$$

2) For each feature, train a weak classifier on Q_w and find the best weak classifier c_t with the minimum weighted error using the algorithm in [41].

3) Update the current CSDC: $H_t = H_{t-1} + c_t$.

4) If $sign(H_t)$ successfully classifies all the samples in Q_w, update Q_w by swapping 20 percent smallest weight interpersonal samples with the interpersonal samples not used in Q.

5) Update the weights:

$$w_{t,i} = \begin{cases} \frac{1}{2m} e^{-Y_i H_t(X_i)}, Y_i = 1 \\ \frac{1}{2n} e^{-Y_i H_t(X_i)}, Y_i = -1 \end{cases}, 1 \le i \le m+n.$$

Output: H_T.

We apply different schemes for face verification and identification. For verification, three CSDCs are used and the results are fused. For identification, the fusion gives trivial improvement on the recognition rate; only H_T^{Haar} can already perform very well. Thus, we just use H_T^{Haar} for identification, which also runs faster. The testing procedure is shown in Algorithm 3.

Algorithm 3. Collective Shape Difference Classifier Testing
Identification:
Let $G = \{I_1, \ldots, I_r\}$ be the gallery of depth images and I_p be a probe depth image.
For $i = 1, 2, \ldots, r$
  1) Compute the SSDMs:
     $D_s^{i+} = (I_i - I_p)$ and $D_s^{i-} = (I_p - I_i)$.
  2) Compute the score:
     $\Omega_i = \max\{H_T^{Haar}(D_s^{i+}), H_T^{Haar}(D_s^{i-})\}$.
Result: Label$(I_p) = \arg \max_{1 \le i \le r}(\Omega_i)$.

Verification:
Let I be a depth image in the gallery, I_p be a probe depth image, and ζ be a threshold.
  1) Compute the SSDMs:
     $D_s^+ = (I - I_p)$ and $D_s^- = (I_p - I)$.
  2) Compute the scores:
  $\Omega_1 = \max\{H_T^{Haar}(D_s^+), H_T^{Haar}(D_s^-)\}$.
  $\Omega_2 = \max\{H_T^{Gabor}(D_s^+), H_T^{Gabor}(D_s^-)\}$.
  $\Omega_3 = \max\{H_T^{LBP}(D_s^+), H_T^{LBP}(D_s^-)\}$.
  3) Compute the score: $\Omega = \sum_{i=1}^3 \Omega_i$.
Result:
  Accept: if $\Omega > \zeta$.
  Reject: if $\Omega \le \zeta$.

6 EXPERIMENTS

Two 3D face databases, FRGC v2 [38] and BU-3DFE [39], are used in our experiments. The BU-3DFE database includes 100 people and 2,500 models. Each person has seven kinds of expressions, one neutral and six other expressions. Each expression includes four levels of intensity from "small" to "large." FRGC v2 has 466 people and 4,007 test models, in which 41 percent of the models are nonneutral.

In our experiments, the BU-3DFE database is used to generate intrapersonal and interpersonal SSDMs for training and FRGC v2 is used for testing. Since these two databases were captured with different devices and in different environments, the experiments can well test the ability of our algorithm for practical applications. According to the expression levels in the BU-3DFE database, we put models with neutral, first level, and second level expressions into a set Set_1 (1,300 models) and the rest to Set_2 (1,200 models). The intrapersonal SSDMs are constructed by subtracting each model in Set_1 from the one of the same person in Set_2. The interpersonal SSDMs are generated between each pair of faces from different people, one from Set_1 and the other from Set_2. As a result, a total of 15,600 intrapersonal and 1,544,400 interpersonal SSDMs are obtained. All of the intrapersonal SSDMs and randomly selected 10,000 starting interpersonal SSDMs are initially input for training. Bootstrapping is used to enumerate the rest of the interpersonal SSDMs.

For testing with FRGC v2, both verification and identification are carried out. For verification, the standard protocols of FRGC v2, ROC I, ROC II, and ROC III masks are used for comparison with related algorithms. Other partition methods, "neutral versus all," "neutral versus neutral," and "neutral versus nonneutral" in [43], are also tested. Receiver Operating Characteristic (ROC) curves from the false accept rate (FAR) of 0.001 are plotted for verification comparison. As for identification, according to the routine in the previous literature, we evaluate two partition methods: one using the first session of the 466 people as the gallery and the remaining 3,541 as the probes [42], [45], and the other using the first neutral session as the gallery and the remainder as the probes [43].

The models in BU-3DFE have been preprocessed by the provider. Before testing, the models in FRGC v2 are smoothed and cropped by the preprocessing methods given in Section 2. All models in these two databases are aligned to the standard posture by our posture alignment method before training and testing. In training, total 3,000 features are selected for each CSDC.

6.1 Comparison between SSDM and SDM

First we compare the performances of SSDM and SDM. The test is carried out by a single CSDC classifier with the Haar features. The verification results are shown in Fig. 12 with the standard ROC I, ROC II, and ROC III masks of FRGC v2. About 2 percent improvement of the verification rates are obtained by SSDM with the FAR of 0.1 percent. These results indicate that the scheme with the signed shape difference map is more significant thanks to its ability to encode more information than the absolute shape difference map.

6.2 CSDCs and Their Fusion

The three individual CSDCs, H_T^{Haar}, H_T^{Gabor}, and H_T^{LBP}, and the fusion of them are tested in this section. Fig. 13 and Fig. 14 indicate the verification results by the ROC curves. The former uses the ROC I, ROC II, and ROC III masks in FRGC v2 and the latter tests the ability of the CSDCs to handle expression variations by separating the data into neutral and nonneutral parts.

From Fig. 13 and Fig. 14, we can see that the Haar features have higher performance than the Gabor and LBP features.
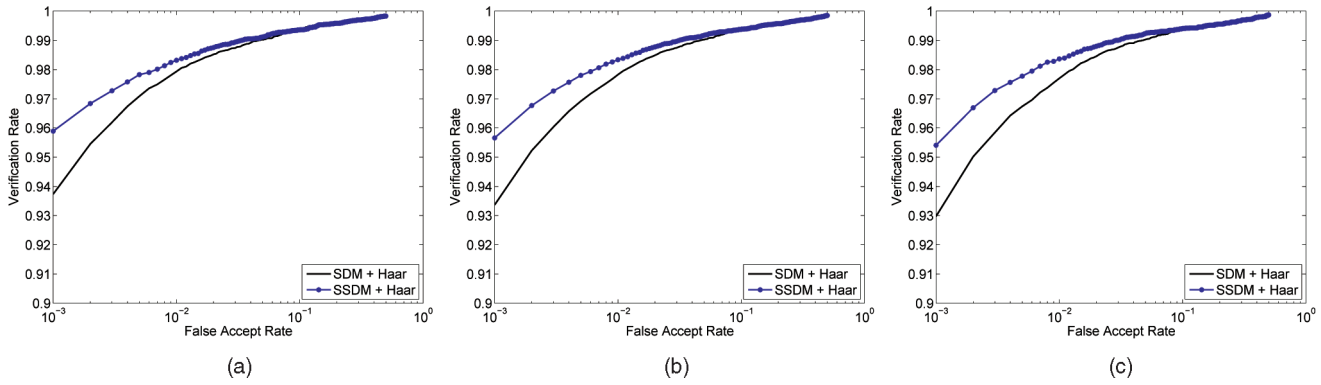
Fig. 12. ROC curves using SDMs and SSDMs in FRGC Experiment 3. (a) ROC I. (b) ROC II. (c) ROC III.
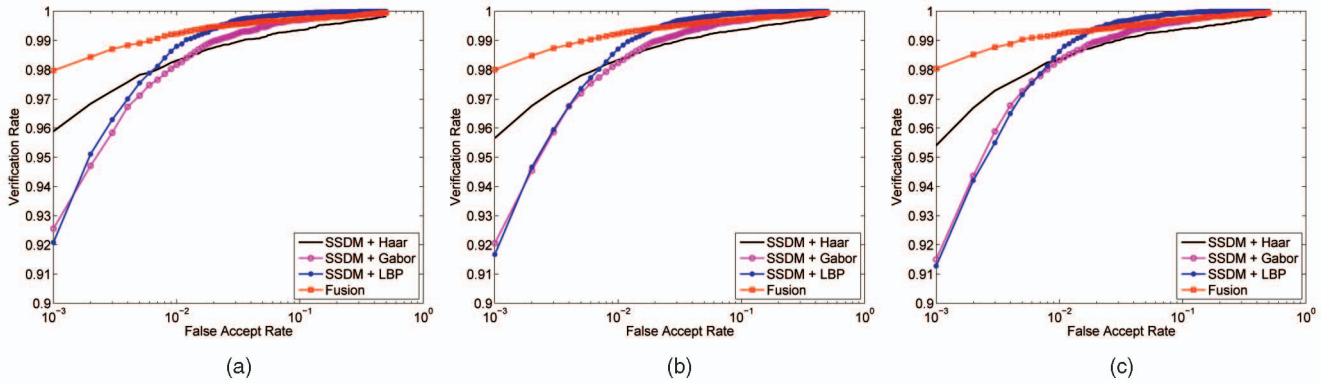

Fig. 13. ROC curves using individual features and their fusion in FRGC Experiment 3. (a) ROC I. (b) ROC II. (c) ROC III.
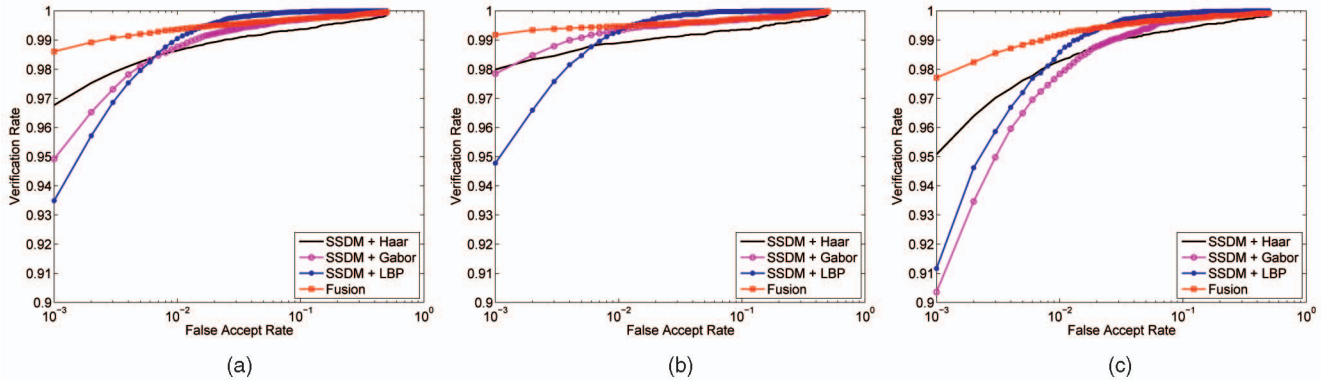

Fig. 14. ROC curves using individual features and their fusion in the experiment with neutral faces enrolled. (a) Neutral versus all. (b) Neutral versus neutral. (c) Neutral versus nonneutral.

Intuitively, Gabor and LBP are relatively "stronger" features and can encode richer information than Haar. The underlying reasons of such results are as follows: 1) As shown in Table 1, there are still small displacements among aligned samples using our pose alignment method. The Haar features using the differences between patches in the SSDMs can remove the small alignment errors to some extent in local neighborhoods. 2) With different orientations and scales in the Gabor kernels and different neighborhood sizes in the LBP computing, a huge amount of features is generated. For practical programming, two subsets are randomly selected from the huge pools of the Gabor and LBP features, which may not include all of the "stronger" features.

The fusion of the three classifiers gives higher verification rates than the individual classifiers. With the FAR of 0.1 percent in the fusion, the verification rates of 97.97 percent

for ROC I mask, 98 percent for ROC II mask, and 98 percent for ROC III mask are obtained. On the neutral and nonneutral partition with neutral faces enrolled, we have verification rates of 99.2 percent for the neutral set, 97.7 percent for the nonneutral set, and 98.6 percent for all the models. The results demonstrate that the Haar, Gabor, and LBP features can be complementary in describing the local characteristics and the combination of them achieves higher performance.

6.3 Identification

Fig. 15 shows the identification results with Cumulative Match Characteristic curves. Only H_T^{Haar} is used for identification. One curve corresponds to using the earliest scan in the gallery and the rest as probes, and the other curve corresponds to using the earliest neutral scan in the gallery and the rest as probes. The experiments give rank-1
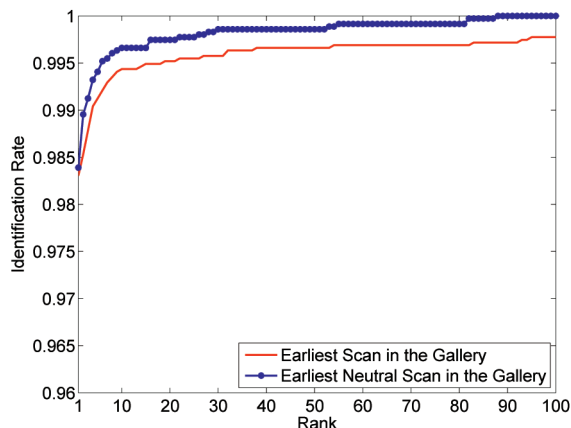
Fig. 15. Cumulative match characteristic curves (only H_T^{Haar} is used).

recognition rates of 98.3 percent and 98.4 percent, respectively. By the fusion of the three CSDCs, slight rank-1 improvements of about 0.2-0.5 percent can be obtained, but it takes much more time than the single CSDC with the Haar features because the computation of the Gabor features is slower than that of the Haar features. While the increased time cost of the fusion is trivial for verification (1 versus 1), it is expensive for identification (1 versus N). Therefore, we only use the single Haar CSDC for identification, considering the computational efficiency.

6.4 Comparison with the State-of-the-Art Methods

We compare our method with the most recent published methods that use FRGC v2 as the testing database in their experiments. The comparisons concern both verification and identification.

Table 2 shows the verification results with the FAR of 0.1 percent. The results of the other methods are quoted from their papers. For the standard protocol testings, ROC I, ROC II, and ROC III masks of FRGC v2, we obtain the verification rates of around 98 percent, which are comparable to, or even slightly better than, the best published results. In the "All versus All" experiment, our method obtains about 5 percent improvement compared with the best one in the table. Thus, our verification method can handle more general cases (such as nonneutral faces in the gallery), since we include the SSDMs between nonneutral faces for the training of the CSDCs.

The comparison of identification is shown in Table 3. Our CSDC with the Haar features obtains rank-1 recognition rates

TABLE 2
Verification Rate Comparison with the State-of-the-Art Methods for $FAR = 0.1\%$

	ROC I	ROC II	ROC III	Neutral vs. All	All vs. All
Maurer* [29]	-	-	-	92%	87%
Cook [31]	93.71%	92.91%	92.01%	95.81%	92.31%
Husken* [30]	-	-	89.5%	-	-
Faltemier [45]	-	-	94.8%	-	93.2%
Kakadiaris [42]	97.3%	97.2%	97.0%	-	-
Mian* [43]	-	-	-	98.5%	86.6%
Our method	97.97%	98.01%	98.04%	98.61%	98.13%

The mark "*" means that only the results of their 3D engines are shown from their 2D + 3D multimodal methods.

TABLE 3
Comparison of Rank-1 Recognition Rates

	Cook [31]	Faltemier [45]	Kakadiaris [42]	Mian (3D only) [43]	Our H_T^{Haar}
E-S	92.9%	97.2%	97.0%	-	98.30%
E-N-S	94.6%	98.1%	-	96.2%	98.39%

"E-S" means the earliest scan in the gallery and the remaining as probes. "E-N-S" means the earliest neutral scan in the gallery and the remaining as probes.

of better than 98 percent, outperforming all other methods in the two experiments.

6.5 Computational Performance

Although some of the previous methods can do verification nearly real-time, the high computational cost of identification is still a challenging problem since identification needs to match the probe face against every gallery face. Thus, the size of the gallery and the matching time are the main obstacle to fast identification. Most related papers do not report the time consumed by their algorithms for identification. The approach proposed by Kakadiaris et al. [42] has relatively good computational efficiency while keeping high identification performance. The time used in their metadata generation for one input model is about 15 seconds and the comparison of the metadata is at a rate of 1,000 per second on a typical PC. Mian et al. [43] proposed a rejection classification method to improve the efficiency, but the ICP algorithm used for 3D matching is computationally expensive. Faltemier et al. [45] also used ICP to match multiple regions. The reported time cost of data preprocessing is about 7.5 seconds and each matching spends about 2.3 seconds, which is slow for identification when the number of models in the gallery is large. At a late stage of the review of this paper, we were introduced the work of 3D signature [54]. A 3D signature is in a vector form extracted from preregistered faces. The matching computation of the 3D-signatures is much faster than ICP-based matching methods since it only needs to compute vector distances (102,594 matching scores per second). Except for the matching time, the time cost of face cropping, denoising, face/region alignment, and 3D signature generation is not reported in the paper [54].

Usually the running time of all steps, especially the preprocessing, depends on the numbers of points in 3D face models. We select the models with the minimum and maximum numbers of points from FRGC v2 to test our algorithm and also compute the average identification time. The consumed time on a PC with CPU P4 3.0 GHz and 2 GB RAM is shown in Table 4 (the registration in the symmetry plane determination has been accelerated compared with our previous work [46]).

The SSDM computation and the classification by H_T^{Haar} are very fast. In this experiment, since there are 466 models in the gallery, we need to compute 932 (466×2) SSDMs and 932 scores for classification for each probe (see Algorithm 3 for why it is 932 instead of 466). Thus, if the gallery has 1,000 models, the average identification time is about 3.6 seconds.

For verification, the time cost by H_T^{LBP} and H_T^{Haar} is trivial and the one matching by H_T^{Gabor} is about 3 ms. Thus, the average time cost of our algorithm is about 2.2 seconds for verification.

TABLE 4
Identification Time Used in Each Step

point number	Denoising	Facial Region Extraction	Posture Alignment	Depth Image	SSDMs (466×2)	Scores (466×2)
min. 53898	895ms	123ms	746ms	32ms	516ms	140ms
max. 197298	1844ms	167ms	976ms	32ms	516ms	140ms
av. 100474	1195ms	140ms	821ms	32ms	516ms	140ms

7 CONCLUSION

We have proposed an automatic 3D face recognition method which can obtain both high accuracy and computational efficiency. From the experimental results on the largest available public database, FRGC v2, the following conclusions can be drawn:

1. The results of verification and identification obtained by our algorithm indicate that the SSDMs between 3D faces contain rich information of the similarity/dissimilarity between 3D face shapes and the developed local features are complementary and effective. Another key component is the optimal combination of these features.

2. Two factors are important to result in the high computational performance of our algorithm. Our self-dependent alignment approach greatly reduces the time cost in face matching. The SSDMs further reduce the processing time by converting 3D data into 2D image maps.

3. Our framework is relatively simple and easy to implement. With the low computational cost and the high recognition accuracy, it is promising for use in practical 3D face recognition systems.

Although our method works well on common faces of approximate mirror-symmetry with a nose, it can fail when too many data points of the nose are missing, which causes incorrect alignment. This is the main limitation of our method. Fortunately, this is a rare case and most 3D scanners generate faces that can be well handled by our algorithm.

ACKNOWLEDGMENTS

REFERENCES

[1] G. Medioni and R. Waupotitsch, "Face Modeling and Recognition in 3-D," *Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures*, pp. 232-233, 2003.

[2] 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.

[3] P.J. Besl and N.D. 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.

[4] A. Scheenstra, A. Ruifrok, and R.C. Veltkamp, "A Survey of 3D Face Recognition Methods," *Proc. Int'l Conf. Audio- and Video-Based Biometric Person Authentication*, pp. 891-899, 2005.

[5] K. Chang, K.W. Bowyer, and P. Flynn, "Effects on Facial Expression in 3D Face Recognition," *Proc. SPIE Conf. Biometric Technology for Human Identification*, pp. 132-143, 2005.

[6] K. Chang, K.W. Bowyer, and P. 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.

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

[8] T. Nagamine, T. Uemura, and I. Masuda, "3D Facial Image Analysis for Human Identification," *Proc. Int'l Conf. Pattern Recognition*, pp. 324-327, 1992.

[9] C. Samir, A. Srivastava, and M. Daoudi, "Three-Dimensional Face Recognition Using Shapes of Facial Curves," *IEEE Trans. Pattern Analysis and Machine Integlligence*, vol. 28, no. 11, pp. 1858-1863, Nov. 2006.

[10] A.B. Moreno, A. Sánchez, J.F. Vélez, and F.J. Díaz, "Face Recognition Using 3D Surface-Extracted Descriptors," *Proc. Irish Machine Vision and Image Processing Conf.*, 2003.

[11] S. Gupta, J.K. Aggatwal, M.K. Markey, and A.C. Bovik, "3D Face Recognition Founded on the Structural Diversity of Human Faces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

[12] Y. Wang, G. Pan, Z. Wu, and Y. Wang, "Exploring Facial Expression Effects in 3D Face Recognition Using Partial ICP," *Proc. Asian Conf. Computer Vision*, vol. 3851, pp. 581-590, 2006.

[13] 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.

[14] G.G. Gordon, "Face Recognition Based on Depth and Curvature Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 808-810, 1992.

[15] H.T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-Based Face Surface Recognition Using Spherical Correlation—Principal Directions for Curved Object Recognition," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 372-377, 1998.

[16] B. Achermann and H. Bunke, "Classifying Range Images of Human Faces with Hausdorff Distance," *Proc. Int'l Conf. Pattern Recognition*, pp. 809-813, 2000.

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

[18] C. Hesher, A. Srivastava, and G. Erlebacher, "A Novel Technique for Face Recognition Using Range Imaging," *Proc. Int'l Symp. Signal Processing and Its Applications*, pp. 201-204, 2003.

[19] Y. Lee and J. Shim, "Curvature-Based Human Face Recognition Using Depth-Weighted Hausdorff Distance," *Proc. Int'l Conf. Image Processing*, pp. 1429-1432, 2004.

[20] T.D. Russ, M.W. Koch, and C.Q. Little, "A 2D Range Hausdorff Approach for 3D Face Recognition," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments*, 2005.

[21] X. Lu and A.K. Jain, "Deformation Modeling for Robust 3D Race Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1377-1383, 2006.

[22] X. Lu and A.K. Jain, "Deformation Modeling for Robust 3D Face Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1346-1357, Aug. 2008.

[23] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3D Face Recognition Combining Global Geometric Features with Local Shape Variation Information," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 308-313, 2004.

[24] C. Xu, T. Tan, S. Li, Y. Wang, and C. Zhong, "Learning Effective Intrinsic Features to Boost 3D-Based Face Recognition," *Proc. European Conf. Computer Vision*, pp. 416-427, 2006.

[25] C. Zhong, Z. Sun, and T. Tan, "Robust 3D Face Recognition Using Learned Visual Codebook," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.

[26] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* 2005.

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

[28] Z. Wu, Y. Wang, and G. Pan, "3D Face Recognition Using Local Shape Map," *Proc. IEEE Int'l Conf. Image Processing,* pp. 2003-2006, 2004.

[29] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of Geometrix ActiveID™ 3D Face Recognition Engine on the FRGC Data," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* 2005.

[30] M. Hüsken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and Benefits of Fusion of 2D and 3D Face Recognition," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* 2005.

[31] J. Cook, V. Chandran, and C. Fookes, "3D Face Recognition Using Log-Gabor Templates," *Proc. British Machine Vision Conf.,* pp. 83-83, 2006.

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

[33] G. Pan, Y. Wu, Z. Wu, and W. Liu, "3D Face Recognition by Profile and Surface Matching," *Proc. IEEE Int'l Joint Conf. Neural Network,* pp. 2168-2174, 2003.

[34] P.J. Phillips, P. Grother, R.J. Michaels, D.M. Blackburn, E. Tabassi, and J. Bone, "FRVT 2002: Overview and Summary," www.frvt.org, 2002.

[35] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P. Flynn, K.W. Bowyer, C.L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Results," www.frvt.org, 2006.

[36] 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.

[37] 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.

[38] P.J. Phillips, P. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 947-954, 2005.

[39] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D Facial Expression Database for Facial Behavior Research," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 211-216, 2006.

[40] Y. Wang, G. Pan, and Z. Wu, "3D Face Recognition in the Presence of Expression: A Guidance-Based Constraint Deformation Approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.

[41] R. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-Rated Predictions," *Machine Learning,* vol. 37, no. 3, pp. 297-336, 1999.

[42] I.A. Kakadiaris, G. Passalis, G. Toderici, 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 Integlligence,* vol. 29, no. 4, pp. 640-649, Apr. 2007.

[43] A.S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *IEEE Trans. Pattern Analysis and Machine Integlligence,* vol. 29, no. 11, pp. 1927-1943, Nov. 2007.

[44] A.S. Mian, M. Bennamoun, and R. Owens, "Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition," *Int'l J. Computer Vision,* vol. 79, no. 1, pp. 1-12, 2008.

[45] T. Faltemier, K.W. Bowyer, and P. Flynn, "A Region Ensemble for 3D Face Recognition," *IEEE Trans. Information Forensics and Security,* vol. 3, no. 1, pp. 62-73, Mar. 2008.

[46] Y. Wang, X. Tang, J. Liu, G. Pan, and R. Xiao, "3D Face Recognition by Local Shape Difference Boosting," *Proc. European Conf. Computer Vision,* pp. 603-616, 2008.

[47] C. Huang, H. Ai, Y. Li, and S. Lao, "High-Performance Rotation Invariant Multiview Face Detection," *IEEE Trans. Pattern Analysis and Machine Integlligence,* vol. 29, no. 4, pp. 671-686, Apr. 2007.

[48] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 511-518, 2001.

[49] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, Apr. 2002.

[50] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," *Pattern Recognition,* vol. 29, no. 1, pp. 51-59, 1996.

[51] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Li, "Face Detection Based on Multi-Block LBP Representation," *Proc. Int'l Conf. Biometrics,* pp. 11-18, 2007.

[52] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *IEEE Trans. Pattern Analysis and Machine Integlligence,* vol. 26, no. 9, pp. 1222-1228, Sept. 2004.

[53] T. Russ, C. Boehnen, and T. Peters, "3D Face Recognition Using 3D Alignment for PCA," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1391-1398, 2006.

[54] C. Boehnen, T. Peters, and P. Flynn, "3D Signatures for Fast 3D Face Recognition," *Proc. IAPR/IEEE Int'l Conf. Biometrics,* pp. 12-21, 2009.
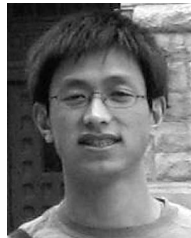
Yueming Wang received the PhD degree from College of Computer Science and Technology, Zhejiang University, P.R. China, in 2007. Currently, he is a postdoctoral fellow in the Department of Information Engineering, the Chinese University of Hong Kong. His research interests include 3D face processing and recognition, object detection, and statistical pattern recognition.

Jianzhuang Liu received the BE degree from Nanjing University of Posts and Telecommunications, P.R. China, in 1983, the ME degree from Beijing University of Posts and Telecommunications, P.R. China, in 1987, and the PhD degree from The Chinese University of Hong Kong in 1997. From 1987 to 1994, he was a faculty member in the Department of Electronic Engineering, Xidian University, P.R. China. From August 1998 to August 2000, he was a research fellow at the School of Mechanical and Production Engineering, Nanyang Technological University, Singapore. Then he was a postdoctoral fellow at The Chinese University of Hong Kong for several years. He is now an assistant professor in the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include image processing, computer vision, machine learning, and graphics. He is a senior member of the IEEE.

Xiaoou Tang received the BS degree from the University of Science and Technology of China, Hefei, in 1990, the MS degree from the University of Rochester, New York, in 1991, and the PhD degree from the Massachusetts Institute of Technology, Cambridge, in 1996. He is a professor in the Department of Information Engineering and associate dean (research) on the Faculty of Engineering of the Chinese University of Hong Kong. He worked as the group manager of the Visual Computing Group at the Microsoft Research Asia from 2005 to 2008. He received the Best Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009. He is a program chair of the IEEE International Conference on Computer Vision (ICCV) 2009 and an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* and *International Journal of Computer Vision (IJCV).* He is a fellow of the IEEE. His research interests include computer vision, pattern recognition, and video processing.

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