

# 3D Face Recognition by Local Shape Difference Boosting

Yueming Wang<sup>1</sup>, Xiaoou Tang<sup>1</sup>, Jianzhuang Liu<sup>1</sup>, Gang Pan<sup>2</sup>, and Rong Xiao<sup>3</sup>

<sup>1</sup> Dept. of Information Engineering, The Chinese University of Hong Kong  
ymingwang@gmail.com, {jzliu,xtang}@ie.cuhk.edu.hk

<sup>2</sup> College of Compute Science, Zhejiang University  
gpan@zju.edu.cn

<sup>3</sup> Microsoft Research Asia

**Abstract.** A new approach, called *Collective Shape Difference Classifier* (CSDC), is proposed to improve the accuracy and computational efficiency of 3D face recognition. The CSDC learns the most discriminative local areas from the *Pure Shape Difference Map* (PSDM) and trains them as weak classifiers for assembling a collective strong classifier using the real-boosting approach. The PSDM is established between two 3D face models aligned by a posture normalization procedure based on facial features. The model alignment is self-dependent, which avoids registering the probe face against every different gallery face during the recognition, so that a high computational speed is obtained. The experiments, carried out on the FRGC v2 and BU-3DFE databases, yield rank-1 recognition rates better than 98%. Each recognition against a gallery with 1000 faces only needs about 3.05 seconds. These two experimental results together with the high performance recognition on partial faces demonstrate that our algorithm is not only effective but also efficient.

## 1 Introduction

With explicit shape information, three dimensional face recognition has been expected to overcome the problems, such as the variations of pose, lighting and expression [1,2], facing traditional 2D face recognition. Various techniques for 3D face recognition have been presented to make use of the shape clues [2]. However, as a relatively new research topic, a number of challenges still exist that limits the performance of current 3D face recognition algorithms both in accuracy and speed.

The first challenge is how to automatically extract the facial region from the raw data captured by a 3D scanner, which may contain hair, shoulder, and neck. Chang *et al.* [3] designed a skin model for the 2D color face image to help the 3D face extraction. Accurate registration between the image pixels and the 3D points are needed and the 3D facial region is found according to the skin detection. The requirement of an additional 2D image limits the usage of the approach.

The second challenge is the precise and fast alignment between two face models. A popular algorithm is the *iterative closest point* (ICP) [4] which is widely

used during the matching stage in 3D face recognition [5,7,1,3]. The ICP algorithm iteratively minimizes the mean square distance (MSD) metric and has relatively good recognition performance. However, it suffers from facial surface distortion due to expression variation and noise [3]. Furthermore, the iterative process makes the ICP computationally expensive and the registration must be done for each model in the gallery. Consequently, it is not suitable for a recognition task with a large dataset.

The third problem is how to measure the similarity between two given facial shapes. Existing features include curvature [8], profile [9], surface descriptors [16], *Point Signature* [7], and *Spherical Face Representation* [17]. Some recent methods treat the aligned face model as a point set. The similarity is calculated via *Hausdorff distance* [5] or *Root-Mean-Square* (RMS) of the closest distances [3,10]. This kind of distances based on averaging provides a plain dissimilarity measure. The performance may decrease seriously in the presence of intra-personal variation.

Expression variation is the fourth challenge. To reduce the effect of expressions, one approach is to choose only rigid regions for matching [7,3]. Another is to map or deform the original facial data to a middle model in which the distortion is reduced [10,11,6]. However, in the former case, there may not be such parts of the face that is shape invariant with sufficient discriminating power [2]. The latter work is interesting and improves the performance to some extent, but it is computationally demanding and more details are needed to understand the underlying mechanism.

The last challenge concerns computational efficiency. More information in 3D facial model leads to more computational cost. To our knowledge, there are no effective algorithms that can run in real time or close to real time in 3D face recognition when thousands of faces are in the gallery.

This paper proposes a new 3D face recognition approach named *Collective Shape Difference Classifier* (CSDC) to deal with the problems described above. The experimental results on FRGC v2 dataset [13], including 466 persons and 4007 models, achieve the rank-1 recognition rate *98.22%*, which outperforms the best published approaches and its speed is nearly real-time. The main contributions of our work are listed as follows:

(1) A fast and effective face posture alignment technique is presented to place all face models to a standard position and orientation. We show that this feature-based alignment step is enough for our later processing and no other iterative registration is needed. Also, the alignment is self-dependent which greatly reduces the time cost in face matching. This technique is to overcome the second and fifth challenges.

(2) A *Pure Shape Difference Map* (PSDM) is defined between two depth images sampled from two aligned 3D face models. The PSDM removes the alignment error in the pitch angle elegantly so as to encode more shape difference information between two faces. The scheme is used to tackle the second and the third challenges.

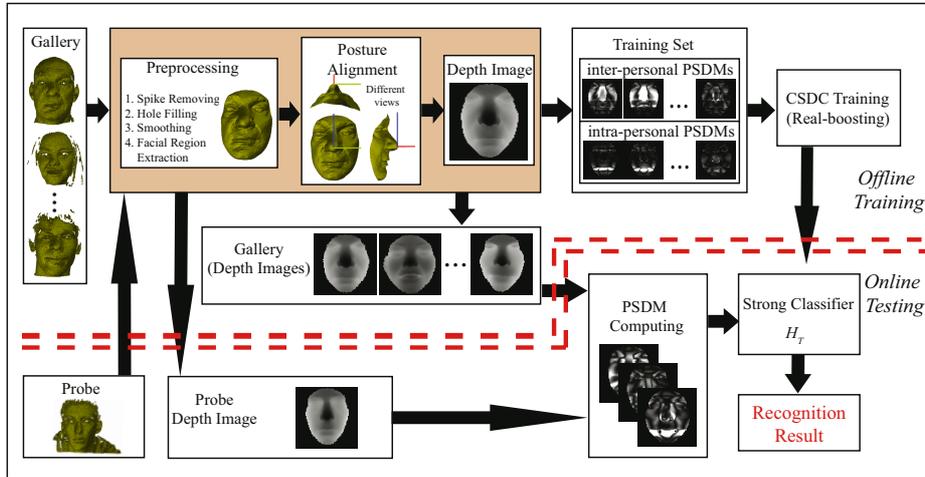


Fig. 1. The framework of our method

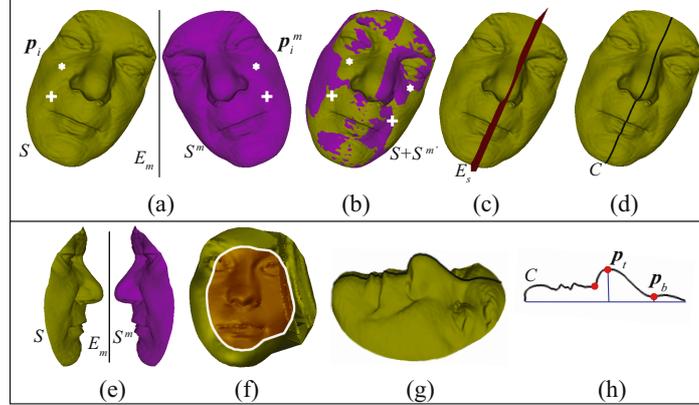
(3) The PSDM helps convert the multi-class face recognition problem to a 2-class classification problem, i.e., inter-personal and intra-personal classes, similar to the case in 2D face recognition [19,20]. Clearly, the different parts of PSDM do not contribute the same discriminability due to non-rigid distortion on the face. From the training inter-personal and intra-personal PSDM sets, the real-boosting [12] is used to choose the most discriminative local shape difference to build a strong classifier, namely, the CSDC. This part is to conquer the third, fourth, and fifth challenges.

Moreover, we also introduce a scheme for facial region extraction which addresses the first difficulty. Summarizing the above techniques, we show the whole framework in Fig. 1. Besides the accuracy and speed, another advantage of our method is its robustness on the partially missing face data. The specific techniques are discussed in the rest of this paper.

## 2 Posture Alignment

It is not clear how to define exact alignment of poses for two given 3D faces with rather different shapes. ICP uses the minimum of the *mean square distance* as a measure for alignment. We prefer to the coincidence of the prominent features such as the nose tips and the normals of the 3D face symmetry planes for alignment due to consideration of algorithmic efficiency and the fact that these two features (the nose tip and the normal) are relatively stable. By the two features, five out of the six degrees of freedom in face model can be fixed and we only need another point to determine the pitch angle. The top of the nose bridge is used for this task.

More detailed geometric definitions of the nose tip and the top of the nose bridge are necessary. Let the central profile curve  $C$  be the intersection of the



**Fig. 2.** Symmetry plane and profile finding. (a) An original model  $S$  and its mirror  $S^m$ . (b) Registration of  $S^m$  to  $S$ . (c) The symmetry plane  $E_s$ . (d) The profile  $C$ . (e) A special case of the mirror plane. (f) Points inside the closed white curve used for registration. (g) The horizontally placed profile. (h) The profile  $C$ ,  $p_t$  and  $p_b$ .

symmetry plane and the facial surface. Then, the nose tip  $p_t$  is defined as the point on the central profile  $C$  with the maximum distance to the line  $l_c$  passing through the two endpoints of  $C$ . The definition of the top of the nose bridge  $p_b$  is given in Section 2.2.

If the input facial surface only covers the facial region as shown in Fig. 2, these features can be found reliably by our method described in Section 2.1 and 2.2. However, failure can appear when the data contain the neck or much hair. Thus, the facial model should cover only the main facial region which does not exceed much of the forehead and the chin. Fortunately, it is not a difficult requirement. The facial region extraction introduced in Section 4.1 works well for this. The rest of this section discusses how to find the symmetry plane, the central profile, the nose tip and the top of the nose bridge, and how to align the facial model.

## 2.1 Facial Central Profile Finding

Let  $S = \{p_i \mid p_i = (x_i, y_i, z_i), 1 \leq i \leq n\}$  denote the point set of a 3D facial model, and  $S^m = \{p_i^m \mid p_i^m = (x_i^m, y_i^m, z_i^m), 1 \leq i \leq n\}$  be its mirror set with respect to some plane  $E_m$ , where the correspondence is naturally set up. Then we register  $S^m$  to  $S$  using the ICP algorithm [4] with  $S$  fixed (see Fig. 2(b)). After the registration,  $S_m$  becomes another set  $S^{m'} = \{p_i^{m'} \mid p_i^{m'} = (x_i^{m'}, y_i^{m'}, z_i^{m'}), 1 \leq i \leq n\}$ . The facial symmetry plane  $E_s$  is defined as the best fitting plane of the set of points  $B = \{p_i^b \mid p_i^b = (p_i + p_i^{m'})/2, 1 \leq i \leq n\}$ . The central profile  $C$  can easily be found by computing the intersection between  $S$  and  $E_s$ .

There are two issues that need to be addressed in the implementation of our technique. The first is that the mirror plane  $E_m$  should be chosen carefully. In some cases, an arbitrary  $E_m$  may cause the ICP to be nonconvergent, such as the

case shown in Fig. 2(e), because the initial poses of the two faces are changed too much. To deal with this problem, we propose the following scheme: 1) Perform PCA on  $S$  to obtain three new principal directions (eigenvectors)  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_3$  with their corresponding eigenvalues  $\lambda_1 \geq \lambda_2 \geq \lambda_3$ . Roughly speaking,  $\mathbf{v}_1$  is in the direction passing through the nose bridge,  $\mathbf{v}_3$  is perpendicular to the front face, and  $\mathbf{v}_2$  is perpendicular to both  $\mathbf{v}_1$  and  $\mathbf{v}_3$ . 2)  $E_m$  is chosen as the plane passing through the centroid of  $S$  and with its normal being  $\mathbf{v}_2$ . Such a mirror plane  $E_m$  passes through  $S$ , making  $S^m$  and  $S$  already quite coincident, which leads to fast convergence of the ICP.

The second issue comes from the fact that our method is based on the mirror symmetry of human faces. However, the extracted facial region may not be so ideal, especially along the boundary. One example is given in Fig. 2(f). To guarantee the better convergence and alignment with ICP, a simple strategy is used where the points close to the boundary are discarded in the registration. When the facial model is represented as a 3D mesh, the inner points can easily be determined.

## 2.2 The Standard Coordinate Frame

According to definition, the nose tip  $\mathbf{p}_t$  can be obtained from the central profile  $C$  by

$$\mathbf{p}_t = (x_t, y_t, z_t) = \arg \max_{\mathbf{p}_i^c \in C} \text{dist}_1(\mathbf{p}_i^c, l_e), \quad (1)$$

where  $\text{dist}_1(\cdot, \star)$  is the Euclidean distance from a point to a line segment, and  $l_e$  is the line passing through the two endpoints of  $C$ .

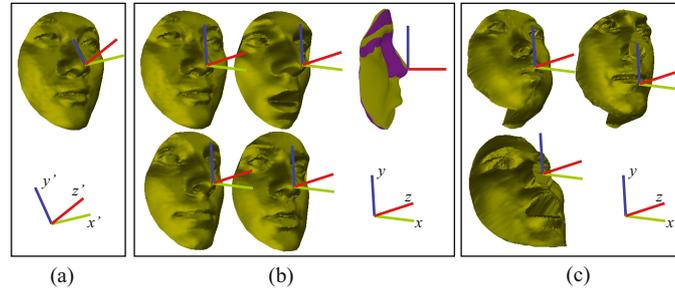
Along the profile  $C$  with  $C$  placed horizontally as shown in Fig. 2(h), we find a local minimum point on each side of  $\mathbf{p}_t$ , which is closest to  $\mathbf{p}_t$ , denoted as  $\mathbf{p}_1^*$  and  $\mathbf{p}_2^*$ . Then the top of the nose bridge  $\mathbf{p}_b$  is defined as

$$\mathbf{p}_b = \arg \max_{\mathbf{p}^* \in \{\mathbf{p}_1^*, \mathbf{p}_2^*\}} \text{dist}_2(\mathbf{p}^*, \mathbf{p}_t), \quad (2)$$

where  $\text{dist}_2(\cdot, \cdot)$  denotes the Euclidean distance between two points. Figs. 2(g) and (h) show the geometric relations of these new terms. It is worth noting that in the implementation for robustly finding the local minimum  $\mathbf{p}_1^*$  or  $\mathbf{p}_2^*$ , we use its six nearest points along  $C$ , three on its one side and three on its other side, to detect if it is a local minimum.

Now let the normal of the symmetry plane  $E_s$  be  $\mathbf{v}_x'$ . We define a candidate frame  $(\mathbf{v}_x', \mathbf{v}_y', \mathbf{v}_z')$  with  $\mathbf{p}_t$  being the origin and  $\mathbf{v}_y' = \mathbf{p}_b - \mathbf{p}_t$ ,  $\mathbf{v}_z' = \mathbf{v}_x' \otimes \mathbf{v}_y'$ , where  $\otimes$  denotes the cross product of two vectors. The geometry of this frame is illustrated in Fig. 3(a). When we find the majority of the data points have positive  $z$  coordinates, we multiply all the  $x$  and  $z$  coordinates of the data points by  $-1$  so that  $\mathbf{v}_x'$  is in the direction towards the left-hand side of the face.

Finally, the standard coordinate frame is defined as  $(\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z)$ , which is obtained by rotating the candidate frame counterclockwise around the  $x$  axis by an angle  $\alpha$  ( $\alpha = 30^\circ$  in our experiments), as shown in Fig. 3(b). Compared with the candidate frame, in this standard frame, we can reduce the number of coincident points when the 3D face data are projected to the  $x - y$  plane



**Fig. 3.** (a) The candidate Frame. (b) Models in the standard frame with small pitch angle variation. (c) Some failure cases.

for constructing the PSDM (see the next section). Now we can align all models into this standard frame. Since different people have different nose shapes, the models, after alignment, may have small variation in the pitch angle (see Fig. 3 (b)). The effect of this pitch angle error can be removed with the PSDM, as described in the next section.

It should be mentioned that when there are many missing data and/or large distortion in a face model, it is possible for our method to align incorrectly. Three such examples are shown in Fig. 3(c). However, our method can obtain very good results for almost all the models in FRGC v2 and BU-3DFE databases.

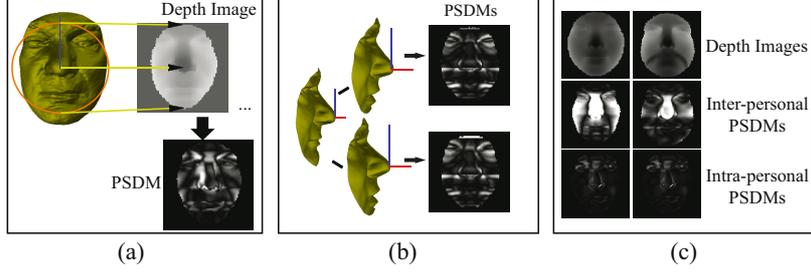
### 3 Discriminative Local Shape Difference Boosting

Based on the aligned models, we investigate the shape differences and convert the 3D face recognition to a 2-class classification problem, i.e., the problem of determining a shape difference is inter-personal or intra-personal. In this section, we design a shape difference representation method, called *Pure Shape Difference Map* (PSDM), and a classifier, called *Collective Shape Difference Classifier* (CSDC). The PSDM aims to depict the shape difference/similarity between two models with reduced alignment error, and the CSDC can choose the most discriminative local patches from the PSDM to make the recognition decision.

#### 3.1 Pure Shape Difference Map

Before obtaining the PSDM between two face models in the standard coordinate frame, we need to generate their depth images. By a sphere with radius  $r$  centered at the nose tip, the *region of interest* (ROI) is picked out and projected to a  $w \times w$  image with the nose tip at the center of the image (we choose  $w = r = 75$  in our experiments), as shown in Fig. 4(a). The positions that the projected face surface does not cover is set to a special value  $\varphi$ .

The difference image between two depth images can reflect the shape similarity between two 3D face models. However, the accuracy may be affected by the small alignment error in the pitch angle, which results from the small position change



**Fig. 4.** (a) PSDM computing. (b) Two very similar PSDMs obtained with small alignment error in the pitch angles. (c) Examples of inter-personal and intra-personal PSDMs.

of the top of the nose bridge  $p_b$  among a group of 3D models. Suppose that each 3D model has a ground truth posture in the standard coordinate frame. Then we can see that the pitch angle variation of a face model brings approximately the same depth error on the same row of this depth image. Consequently, we can remove its effect by constructing a map of the shape difference, i.e., the PSDM.

Let  $I_1$  ( $I_2$ ) be a depth image and  $I_1(i, j)$  ( $I_2(i, j)$ ) be the depth value at the position  $(i, j)$ ,  $\delta_i^1$  ( $\delta_i^2$ ) be the depth error on row  $i$ , and the ground truth depth images corresponding to  $I_1$  and  $I_2$  be  $I_1^*$  and  $I_2^*$  respectively. Then,

$$I_1(i, j) = I_1^*(i, j) + \delta_i^1, \quad I_2(i, j) = I_2^*(i, j) + \delta_i^2. \quad (3)$$

The signed difference image  $D_s$  of  $I_1$  and  $I_2$  and the signed difference image  $D_s^*$  of  $I_1^*$  and  $I_2^*$  are defined as

$$D_s(i, j) = \begin{cases} I_1(i, j) - I_2(i, j), & \text{if } I_1(i, j) \neq \varphi \text{ and } I_2(i, j) \neq \varphi \\ \xi, & \text{otherwise} \end{cases}, \quad (4)$$

$$D_s^*(i, j) = \begin{cases} I_1^*(i, j) - I_2^*(i, j), & \text{if } I_1^*(i, j) \neq \varphi \text{ and } I_2^*(i, j) \neq \varphi \\ \xi, & \text{otherwise} \end{cases}, \quad (5)$$

where  $\xi$  is a special value denoting that  $D_s(i, j)$  or  $D_s^*(i, j)$  is invalid at the positions where at least one of the two depths under study equals  $\varphi$ . The PSDM,  $D_{ps}$  is defined as

$$D_{ps}(i, j) = \begin{cases} |D_s(i, j) - \frac{1}{k} \sum_{l=q}^{q+k-1} D_s(i, l)|, & \text{if } D_s(i, j) \neq \xi \text{ and} \\ & D_s(i, l) \neq \xi, q \leq l \leq q+k-1 \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where  $\{q, q+1, \dots, q+k-1\}$  denotes  $k$  consecutive pixel positions on row  $i$  (In our experiments, we choose  $q = 28$  and  $k = 21$ ). Since  $D_s(i, j) = D_s^*(i, j) + \delta_i^1 - \delta_i^2$ , we have

$$D_{ps}(i, j) = \begin{cases} |D_s^*(i, j) - \frac{1}{k} \sum_{l=q}^{q+k-1} D_s^*(i, l)|, & \text{if } D_s(i, j) \neq \xi \text{ and} \\ & D_s(i, l) \neq \xi, q \leq l \leq q+k-1 \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

It can be seen from Eq. (7) that the alignment errors  $\delta_i^1$  and  $\delta_i^2$  are removed in  $D_{ps}(i, j)$  (see Fig. 4(b)). Since the PSDM encodes the difference between two depth images with their alignment errors removed, we call it the *Pure Shape Difference Map* and use it as a critical representation for the recognition. Some examples of inter-personal and intra-personal PSDMs are shown in Fig. 4(c).

### 3.2 Collective Shape Difference Classifier

The PSDM keeps the information of the similarity between two face models and the Root-Mean-Square (RMS) is a choice for dissimilarity measure. However, the RMS runs into trouble when the noise and distortion of the facial surface occur. It is obvious that different parts of the PSDM have different contributions to recognition. Although we do not know which areas are the most discriminative across a broad range of distortion, the boosting algorithm can help select and combine them with suitable weights. This is the main idea of the *Collective Shape Difference Classifier* (CSDC).

The CSDC is a collective classifier of the form,  $H_T(D_{ps}) = \sum_{t=1}^T c_t(D_{ps})$ , where  $c_t(D_{ps})$  is a weak classifier selected based on the simple features on the PSDMs during the real-boosting training [12], and  $T$  is the number of the weak classifiers. The output of  $c_t(D_{ps})$  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  $D_{ps}$ .

In the learning of the CSDC, the intra-personal and inter-personal PSDMs are built from the given 3D face models, which compose the training set  $Q$ . Usually, the size of  $Q$  is very large mainly due to many different pairs of inter-personal depth images. It is impractical for a common PC to use all PSDMs in  $Q$  for training simultaneously. Thus, bootstrapping is used in learning by starting with all intra-personal and part of inter-personal PSDMs which form a subset  $Q_w$  of  $Q$ . Then we keep exchanging the inter-personal PSDMs between  $Q_w$  and  $Q$  so that all inter-personal samples can be used during the learning procedure. The detail of the learning is shown in the training part of Algorithm 1. Besides, two types of features are used in constructing the weak classifiers. One is the mean values of the rectangle patches in the PSDMs with different sizes. The other is Haar-like features that are frequently used in face detection [18]. Both use integrate images for better computational efficiency [18]. The testing part of Algorithm 1 shows how the CSDC carries out the recognition.

## 4 Experiments

Two 3D face databases, FRGC v2 [13] and BU-3DFE [14], are used to test our algorithm. The BU-3DFE database includes 100 persons and 2500 models. Each person has 7 kinds of expressions, 1 neutral and 6 other expressions. FRGC v2 has 466 persons and 4007 models.

Half of the 2400 models with non-neutral expressions in BU-3DFE are randomly selected and used together with the 100 neutral models to build the PSDM

---

**Algorithm 1.** Collective Shape Difference Classifier Training and Testing

---

**Training Procedure:****Input:**

1.  $Q'$  and  $Q'_w$ : containing all and starting samples corresponding to  $Q$  and  $Q_w$ .
2.  $T$ : the target number of the weak classifiers.

**Initialization:**  $w_{0,i}$ **Learning:**For  $t = 1, 2, \dots, T$ 

1. Normalize the weights  $w_{t,i}$ .
2. Train weak classifiers on  $Q'_w$  and find the best weak classifier  $c_t$ .
3. Update the current collective classifier  $H_t = H_{t-1} + c_t$ .
4. If  $\text{sign}(H_t)$  successfully classifies all samples in  $Q'_w$ , update  $Q'_w$  by swapping 20% smallest weight inter-personal samples with those never used in  $Q'$ .
5. Update the weights  $w_{t,i}$ .

**Output:**  $H_T$ .**Testing Procedure:**Let  $G = \{g_1, \dots, g_r\}$  be the gallery set and  $M_p$  be a probe.For  $i = 1, 2, \dots, r$ 

1. Compute  $D_{ps}^i$  between  $g_i$  and  $M_p$ ,
2. Compute the score  $\Omega_i = H_T(D_{ps}^i)$ .

**Recognition result:**  $\text{Label}(M_p) = \arg \max_{1 \leq i \leq r} (\Omega_i)$ 

---

training set. The PSDMs are computed from pairs of these training models. Thus, 1200 intra-personal PSDMs and 118,800 inter-personal PSDMs are computed. These samples train the CSDC which is used for the recognition experiments. The remaining half of the models with non-neutral expressions in BU-3DFE are employed to determine the number of features in the CSDC, with the 100 neutral models forming the gallery set. For FRGC v2, the first session of the 466 persons are used as the gallery set, and the remaining 3541 models are used for testing. Note that we do not use any models in FRGC v2 for training. Besides, experiments on partial faces are also designed to test our algorithm.

Before testing, the models in FRGC v2 are smoothed and cropped by the methods given in Section 4.1. The models in BU-3DFE have been preprocessed by the providers. All models in these two databases are aligned to the standard coordinate system by our posture alignment method. The two parameters  $m$  and  $n$  in Algorithm 1 are 1200 and 10000, respectively.

#### 4.1 Preprocessing

This subsection briefly discusses the preprocessing steps including face denoising and facial region extraction. The raw face data are assumed to be stored with known adjacency relations among the 3D points and the faces are placed

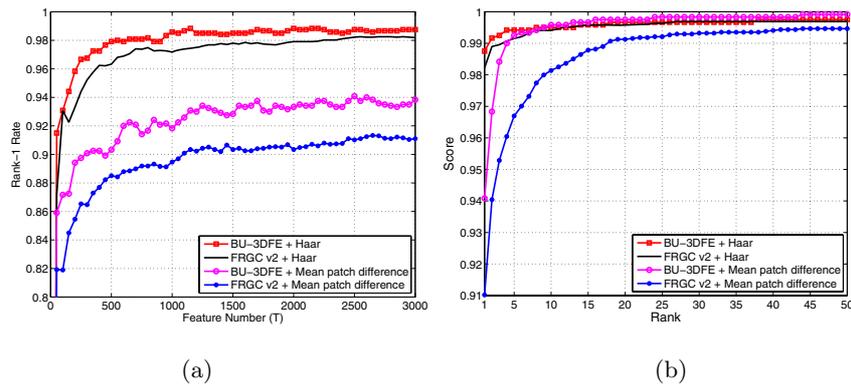
roughly in the common top-down posture (the front direction of the face can be arbitrary), as those in the FRGC v2 and 3D-BUFE databases. Many commercial 3D scanners can generate such data [2].

With this assumption, three Gaussian filters are designed to remove spikes, fill holes, and smooth the data with different variances. After that, the facial region is extracted based on the rough detection of the nose tip as follows: 1) 2/3 points close to the centroid of the denoised face in top-down direction are used to fit a plane  $E_1$ . Among the removed 1/3 points, 2/3 are at the bottom of the face and 1/3 at the top. The plane  $E_1$  cuts the face data to two parts. The one with the smaller variance is selected as the candidate facial region. 2) With the points in the candidate facial region, we fit another plane  $E_2$  and again select the part with the smaller variance. Among the points in this part, the point with the largest distance to the plane  $E_2$  is selected as the approximate nose tip  $\mathbf{p}_t'$ . 3) By placing a sphere centered at  $\mathbf{p}_t'$ , the facial region can be cropped from the original denoised face (the sphere radius = 95 is selected in our experiments). This method is simple and fast, but works very well on the FRGC v2 database.

#### 4.2 Effects of the Number and Type of Features

The number  $T$  of the weak classifiers in the CSDC balances the recognition accuracy and the running time both in training and testing. The maximum value  $T = 3000$  is trained in our experiment. As shown in Fig. 5(a), the rank-1 rates keep increasing, but the curves become flat when  $T > 2500$ . Thus, we set  $T = 2500$  in the subsequent experiments where the rank-1 rates on both databases exceed 98.2%.

As for the type of features, the mean values of rectangle patches give worse rank-1 rate than the Haar-like features by about 4% drop on the two sets. This result indicates that not only the shape difference but also its change patterns encode the similarity when intra-personal variations occur. With Haar-like features,



**Fig. 5.** (a) Rank-1 recognition rates against the feature number. (b) Cumulative match characteristic curves ( $T = 2500$ ).

recognition rates better than 99% are achieved after rank-3 in the *Cumulative Match Characteristic* (CMC) curves. (see Fig. 5(b)).

### 4.3 Comparison with Other Methods

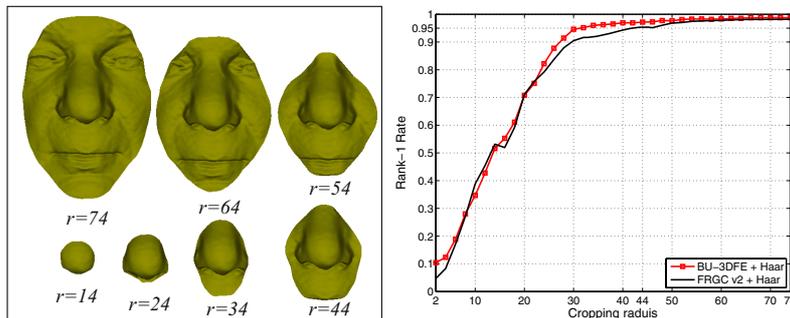
We compare our CSDC with the ICP and the state of the art methods including the ARMS [3], the AFM [15], the GCD [10], and the *R3D* [17]. Each method uses all data in FRGC v2 database. For the AFM, the GCD, and our method, the first data session of each subject is used as the gallery set (total 466 faces) and the rest as probes (3541). The ARMS is with a superset, 449 vs. 3939 [3], and *R3D* chooses a neutral model for each person to compose a gallery and the remaining models are used as probes. The rank-1 rates obtained by the methods are given in Table 1, which clearly indicate that our CSDC performs better than the others. Note that the result of GCD is obtained from the authors of [10], and the others except ICP are quoted from the original papers.

**Table 1.** Comparison with five other works on FRGC v2

	ICP	ARMS	AFM	GCD	<i>R3D</i>	CSDC
Rank-1 Rate	75.66%	91.9%	97%	87.74%	96.2%	98.22%

### 4.4 Evaluation on Partial Faces

Two kinds of partial 3D faces are generated. One is obtained by shrinking the radius  $r$  used in projection for generating depth images (see Fig. 3.1), the other is by removing one or more quadrants of the face region. Some examples are shown in Fig. 6 and Fig. 7. Clearly, with such large parts of the face missing, our pose alignment may fail. However, here our purpose is to find out which parts of the face contribute more to the recognition, based on the partial faces that are aligned well.



**Fig. 6.** Radius shrinking

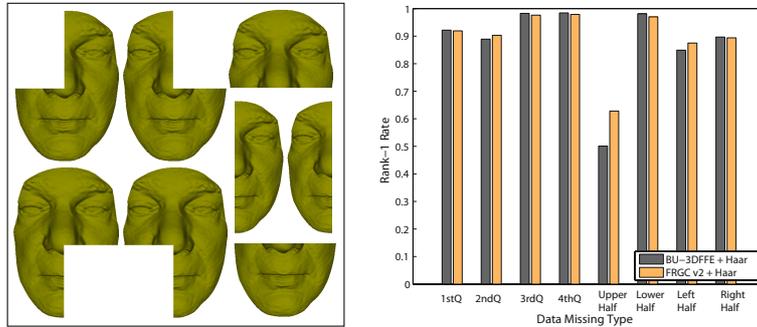


Fig. 7. Quadrant missing

**Radius Shrinking.** Totally 37 cropping radii, from 2 to 74 with step = 2, are tested in the experiment. Fig. 6 shows that the rank-1 rates are better than 95% for all  $r \geq 44$  on both databases. It is rather surprising that such a small central patch as  $r = 44$  still results in very good recognition rates by our CSDC. This is an important finding from our experiments and more attention should be paid to the nose region.

**Quadrant Missing.** We evaluate 8 kinds of quadrant missing. The results are illustrated in Fig. 7. The most important conclusion from this experiment is that the upper half of the face is more important than the lower half in recognition, with the average rank-1 rate 97% versus 56%.

#### 4.5 Computational Performance

Although some of the previous methods can do verification nearly real-time, the computational performance of recognition is still a challenging task since the recognition must match the probe face against every gallery face. Thus the size of the gallery and the matching time are the main obstacle of fast recognition.

Usually, the running time of all steps, especially the preprocessing, depends on the number 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 recognition time. The consumed time on a PC with CPU P4 3.0GHz and 2GB RAM is shown in Table 2.

The PSDM computation and the classification by the CSDC are very fast. In this experiment, since there are 466 models in the gallery, we need to compute

Table 2. Time used in each step

points number	Denosing	Facial Region Extraction	Posture Alignment	Depth Image	PSDMs ( <i>466 times</i> )	Scores ( <i>466 times</i> )
min. 53898	895ms	35ms	780ms	32ms	289ms	70ms
max. 197298	1844ms	130ms	1962ms	32ms	289ms	70ms
av. 100474	1195ms	71ms	978ms	32ms	289ms	70ms

466 PSDMs and 466 scores for classification (see Algorithm 1) for each probe. If the gallery has 1000 models, the average recognition time is about *3.05* seconds which is nearly real-time. Our method is several orders faster than existing methods.

## 5 Conclusion and Limitation

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 rank-1 rate better than 98% obtained by the CSDC indicates that the shape differences of 3D models are effective and the local area is critical. For most cases, our pose alignment is good enough.

(2) The low computational cost together with the accuracy makes our method practical for real time 3D face recognition system.

(3) Our method is robust on faces with large missing regions if the faces are aligned well. An important finding is that even a small nose area can give very good recognition results by the CSDC.

Although the CSDC works very well on common faces of approximate mirror-symmetry with a nose, it can fail when the data 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 scanners can generate faces that can be handled by our algorithm.

## Acknowledgement

This work was supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK 414306). Gang Pan was supported by Natural Science Foundation of China (No.60503019) and 863 Program of China (2008AA01Z149).

## References

1. Medioni, G., Waupotitsch, R.: Face recognition and modeling in 3D. In: IEEE Int'l Workshop on AMFG (2003)
2. Chang, K., Bowyer, K., Flynn, P.: A Survey of Approaches and Challenges in 3D and Multi-Modal 2D+3D Face Recognition. *Computer Vision and Image Understanding* 101(1), 1–15 (2006)
3. Chang, K.I., Bowyer, K., Flynn, P.J.: Adaptive Rigid Multi-Region Selection for Handling Expression Variation in 3D Face Recognition. In: IEEE Workshop on FRGC (2005)
4. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. *IEEE Trans. on PAMI* 14(2), 239–256 (1992)

5. Russ, T.D., Koch, M.W., Little, C.Q.: A 2D range Hausdorff approach for 3D face recognition. In: IEEE Workshop on FRGC (2005)
6. Lu, X., Jain, A.K.: Deformation modeling for robust 3D face matching. In: IEEE Conf. on CVPR (2006)
7. Chua, C.S., Han, F., Ho, Y.K.: 3D Human Face Recognition Using Point Signature. In: Int'l Conf. on FG (2000)
8. Gordon, G.G.: Face recognition from depth maps and surface curvature. In: SPIE Conf. on Geometric Methods in Computer Vision (1991)
9. Wu, Y.J., Pan, G., Wu, Z.H.: Face Authentication based on Multiple Profiles Extracted from Range Data. In: Int'l Conf. on AVBPA (2003)
10. Wang, Y.M., Pan, G., Wu, Z.H.: 3D Face Recognition in the Presence of Expression: A Guidance-based Constraint Deformation Approach. In: IEEE Conf. on CVPR (2007)
11. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Three dimensional face recognition. *Int'l Journal of Computer Vision* 64(1), 5–30 (2005)
12. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. In: Annual Conf. on Computational Learning Theory (1998)
13. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the Face Recognition Grand Challenge. In: IEEE Conf. on CVPR (2005)
14. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.: A 3D facial expression database for facial behavior research. In: Int'l Conf. on FG (2006)
15. Kakadiaris, I.A., Passalis, G., Toderici, G., et al.: Three-Dimensional Face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. on PAMI* 29(4), 640–649 (2007)
16. Moreno, A.B., Sanchez, A., Velez, J.F., et al.: Face recognition using 3D surface-extracted descriptors. In: Irish Machine Vision and Image Processing Conference (2003)
17. Mian, A., Bennamoun, M., Owens, R.: An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face recognition. *IEEE Trans. on PAMI* 29(11), 1927–1943 (2007)
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conf. on CVPR (2001)
19. Wang, X., Tang, X.: A unified framework for subspace face recognition. *IEEE Trans. on PAMI* 26(9), 1222–1228 (2004)
20. Wang, X., Tang, X.: Unified subspace analysis for face recognition. In: IEEE International Conference on Computer Vision (2003)